

Universidad Cenfotec

Maestría en Tecnología de Bases de Datos

Documento Final de Proyecto de Investigación Aplicada 2

Validación de un Modelo de Construcción de Almacén de Datos para la Universidad de Costa Rica

Zúñiga Chaves, Gosvami

Agosto, 2016

página en blanco

Declaratoria de derechos de autor

Yo, Gosvami Zúñiga Chaves, número de identificación 1-1092-0732, estudiante de la Universidad Cenfotec, de la carrera Maestría en Tecnologías de Bases de Datos, autorizo a la Universidad Cenfotec y a sus estudiantes activos a que el presente documento pueda ser consultado única y exclusivamente con fines académicos.

_____

Gosvami Zúñiga Chaves

Dedicatoria

A mi familia, por inculcarme siempre el deseo de superación y en instruirnos en la importancia del estudio como meta personal. Todos los años de esfuerzos, no han sido en vano.

A María, por acompañarme desde hace más de 10 años en esta travesía llena de aventuras y sorpresas que es la vida, y ahora al doble, con nuestro hijo Isaac.

No estaría aquí si no fuera por su apoyo.

Agradecimientos

A mis compañeros de maestría, que al final, este proyecto refleja el trabajo de 2 años en múltiples cursos, de aprovechar el conocimiento de otros para mejorar, y que el camino no termina aquí, sino que apenas empieza.

A mis compañeros de trabajo, en especial a los ingenieros Róger Tenorio y Carlos Granados, por todos los consejos, ayudas y recomendaciones para hacer de este un mejor proyecto, y de mí, un mejor profesional.

Hoja de Aprobación del Proyecto

	Firma	Fecha
Máster José Cabezas Jaikel Profesor tutor	_____	_____
Máster Ignacio Trejos Zelaya Rector	_____	_____
Ph.D.(in fieri) Luis Naranjo Zeledón Profesor lector	_____	_____

Tabla de Contenidos

RESUMEN EJECUTIVO **11**

CAPÍTULO 1. INTRODUCCIÓN **12**

1.1 Antecedentes del Problema 14
1.2 Definición y Descripción del Problema 14
1.3 Justificación 15
1.4 Viabilidad 16
1.5 Objetivos 17
1.6 Alcances y Limitaciones 18
1.7 Estado de la Cuestión 19

CAPÍTULO 2. MARCO TEÓRICO **25**

2.1 Características de un Data Warehouse 27
2.2 Caracterizar el modelo dentro de un DW 29
2.3 Evaluación de las instituciones de educación superior públicas 32
2.4 Técnicas de construcción de un DW para instituciones universitarias públicas 34

CAPÍTULO 3. MARCO METODOLÓGICO **38**

3.1 Tipo de Investigación 39
3.2 Alcance Investigativo 39
3.3 Enfoque 40
3.4 Diseño 42
3.5 Población y Muestreo 42
3.6 Instrumentos de Recolección de Datos 43
3.7 Estrategia de Desarrollo de la Propuesta 43

CAPÍTULO 4. PROPUESTA DEL MODELO PARA CONSTRUIR UN DW PARA LA UNIVERSIDAD DE COSTA RICA **45**

4.1 Uso del modelo 47
4.2 Fase de Proyecto 50
4.3 Fuente de datos 54
4.4 Diseño 59
4.5 Visualización 65
4.6 Implementación 69
4.7 Modelo en paralelo 71
4.8 Finalizar el modelo 73

CAPÍTULO 5. APLICACIÓN DEL MODELO EN LA ELABORACIÓN DEL SISTEMA DE INFORMACIÓN ESTUDIANTIL **74**

5.1 Ejecución del modelo 75
5.2 Tiempo de construcción 76
5.3 Demostración de los datos en el DW 78
5.4 Satisfacción del usuario 81

CAPÍTULO 6. CONCLUSIONES Y RECOMENDACIONES 83

6.1 Conclusiones 84
6.2 Recomendaciones 91
6.3 Reflexiones Finales 92

BIBLIOGRAFÍA 93

REFERENCIAS 94

ABREVIATURAS Y ACRÓNIMOS 97

GLOSARIO 98

APÉNDICES 99

APÉNDICE 1. REVISIÓN SISTEMÁTICA PARA EL ESTADO DE LA CUESTIÓN 100

1. Formulación de la pregunta 102
2. Selección de fuentes 104
3. Selección de los estudios 106
4. Extracción de la información 111
5. Resumen de resultados 119
5.1. Situación actual 120

APÉNDICE 2. PLANTILLA DEL MODELO 121

Introducción 123
1. Fase de Proyecto 125
2. Fase de Fuente de Datos 126
3. Fase de Diseño 127
4. Fase de Visualización 128
5. Fase de Implementación 129

APÉNDICE 3. MODELO APLICADO 130

Introducción 133
1. Proyecto 134
2. Fuente de Datos 144
3. Diseño 165
4. Visualización 180
5. Implementación 186

APÉNDICE 4. CONSULTA DE SATISFACCIÓN DEL USUARIO 195

Cuestionario aplicado 197
Funcionario 1 199
Funcionario 2 201
Funcionario 3 203
Funcionario 4 205
Tabulación de respuestas 207

Índice de Figuras

FIGURA 1. PROCESO DE DISEÑO EN CINCO PASOS. 23
FIGURA 2. ARQUITECTURA PROPUETA DE UN DW EN 3 CAPAS. 28
FIGURA 3. EJEMPLO DE MATRIZ DE PROCESOS PARA VENTA DE PRODUCTOS. 36
FIGURA 4. ONTOLOGÍA DE UN MODELO DE CREACIÓN DE UN DW PARA LAS IES. 41
FIGURA 5. SECUENCIA DE EJECUCIÓN DE LAS FASES DEL MODELO 48
FIGURA 6. DIAGRAMA DE PROCESOS PARA LA CONSTRUCCIÓN DEL DW. 49
FIGURA 7. ARQUITECTURA DE UN DW EN 3 CAPAS. 51
FIGURA 8. DIAGRAMA DE IMPLEMENTACIÓN DE ABAJO HACIA ARRIBA. 52
FIGURA 9. DIAGRAMA DEL MODELO LÓGICO. 60
FIGURA 10. EJEMPLO DEL MODELO DIMENSIONAL. 61
FIGURA 11. JERARQUÍA DE ACCESO DEL PÚBLICO META POR COMPLEJIDAD DE VISUALIZACIÓN. 66
FIGURA 12. SECUENCIA DE EJECUCIÓN EN PARALELO PARA DIFERENTES FUENTES DE DATOS. 72
FIGURA 13. SECUENCIA DE EJECUCIÓN EN PARALELO CON ESPECIALIZACIÓN POR FASES. 73
FIGURA 14. CONSULTA DE ESTUDIANTES POR NIVEL ACADÉMICO Y SEXO UTILIZANDO EL ROL DE AUTORIDAD UNIVERSITARIA. 79
FIGURA 15. CONSULTA DE ESTUDIANTES POR NIVEL ACADÉMICO Y SEXO UTILIZANDO EL ROL PÚBLICO. 79
FIGURA 16. CONSULTA DE CURSOS POR CUPOS MATRICULADOS UTILIZANDO EL ROL DE AUTORIDAD UNIVERSITARIA. 80
FIGURA 17. CONSULTA DE TÍTULOS OTORGADOS UTILIZANDO EL ROL DE AUTORIDAD UNIVERSITARIA. 80

Índice de Tablas

TABLA 1. PALABRAS CLAVES UTILIZADAS EN LA REVISIÓN SISTEMÁTICA. 20
TABLA 2. CRITERIOS DE EVALUACIÓN PARA MEDIR LO ADECUADO DEL MODELO AL NEGOCIO. 42
TABLA 3. EJEMPLO DE MATRIZ DE PROCESOS. 53
TABLA 4. EJEMPLO DE PERMISOS DE ACCESO A DIMENSIONES Y METAS POR ROL DE USUARIO. 67
TABLA 5. EJEMPLO DE MATRIZ DE ACCESO A LAS DIMENSIONES Y VARIABLES POR ROL DE USUARIO. 68
TABLA 6. EJEMPLO DE MATRIZ DE ACCESO PARA METAS POR ROL DE USUARIO. 68
TABLA 7. TIEMPO DE EJECUCIÓN DEL MODELO EN LA PRIMERA ITERACIÓN. 77
TABLA 8. COSTOS TEÓRICOS DE INVERSIÓN EN LA EJECUCIÓN DEL PROYECTO. 78
TABLA 9. EVALUACIÓN PARA CUANTIFICAR LO ADECUADO DEL MODELO AL NEGOCIO 90

Resumen Ejecutivo

La construcción de un almacén de datos en la Universidad de Costa Rica se propone ante el imperativo de contar con información para la toma de decisiones y rendición de cuentas por parte de la Universidad de una manera más eficiente. Para lograrlo, se requiere la unificación de datos desde múltiples fuentes y, para este fin, se necesita formalizar su construcción.

Para normar la construcción del almacén de datos, se decide crear un modelo para estandarizar su construcción. Para generar este modelo, se revisan las mejores prácticas y las técnicas documentadas en estudios relacionados con el ámbito de las instituciones de educación superior. De la investigación se obtiene el tipo de implementación recomendada, la arquitectura, la matriz de procesos, el tipo de diseño y técnicas específicas de diseño.

El modelo incorpora estas buenas prácticas, a la vez que incluye la clasificación de usuarios y complejidad de uso para definir la visualización de los datos, diagrama de procesos, procesamiento independiente y, en general, elementos los cuales faciliten la aplicación del modelo en un entorno real, todo, dentro de un patrón iterativo de ejecución.

Para probar el modelo, se realizó una primera iteración con datos reales de estudiantes y así corroborar el cumplimiento de los objetivos planteados y si el modelo es adecuado para la Universidad de Costa Rica. Luego de esta primera iteración, se aplicó una tabla para validar el modelo, la cual incluye una serie de rubros para cuantificar los resultados. Los resultados indican que los usuarios están satisfechos con la información brindada -luego de esta primera iteración- y que el modelo es aceptado y puede ser implementado en el ámbito de la Universidad.

página en blanco

Capítulo 1. Introducción

1.1 Antecedentes del Problema

La Universidad de Costa Rica, como parte de sus políticas dentro del Plan Estratégico Institucional para la mejora de la gestión universitaria, promueve la creación de un Sistema de Información Institucional. La propuesta refiere la creación de un almacén de datos (*data warehouse*) previo al sistema de información institucional, pero contemplando desde ya, las necesidades detectadas en estudios previos[1] realizados a lo interno de la Universidad.

Los factores analizados hasta el momento relacionados con el *data warehouse* (DW) muestran las necesidades inmediatas contempladas en su construcción. Los factores detectados son:

- Volumen de información estudiantil en el rango de millones de registros.
- No hay almacenamiento histórico en el sistema transaccional.
- Información vital en la mayoría de procesos en donde se utilizan múltiples indicadores. Se realizó un conteo y solo la información sobre matrícula de estudiantes se requiere en seis diferentes escenarios en esta Oficina y se desagregan en más de 100 indicadores.
- No existen históricos sistematizados.
- No hay posibilidad de tomar decisiones a partir de los datos.

1.2 Definición y Descripción del Problema

La propuesta de un modelo de DW específico surge como requerimiento para sentar las bases de un proyecto más amplio por parte de la Oficina de Planificación Universitaria, que plantea la construcción de un sistema de apoyo para toma decisiones a los altos mandos.

El punto de partida ha sido el análisis de la situación con respecto a los avances en temas de almacenes de datos en la Universidad. Las consultas se han realizado a múltiples instancias y, en muchos casos, hay esfuerzos independientes para cubrir la necesidad del manejo de los datos al utilizar bases de datos propias o almacenando los históricos de

[1] Referencia a encuesta de la Oficina de Planificación Universitaria a las autoridades universitarias. Los resultados de la encuesta son privados pues fueron realizados para uso interno.

forma manual, para luego transformarlos de datos a información útil para las unidades de trabajo. Los esfuerzos son importantes, pero en ninguno de los casos, se ha mostrado un modelo específico para la construcción de un DW que pueda ser ejecutado a nivel universitario o como resultado de un proceso investigativo.

El principal problema es el faltante de un modelo el cual sirva como base al proyecto del DW universitario y, por tanto, es necesario formalizar este proceso para asegurar una ejecución adecuada de los objetivos planteados para su construcción.

1.3 Justificación

La Universidad de Costa Rica, como institución de educación superior pública, requiere información constante para 1) rendición de cuentas y para 2) toma de decisiones, para lo cual se cuenta con soluciones individuales muchas veces de esfuerzos empíricos por ordenar y sistematizar los datos.

Las solicitudes de información a la Oficina de Planificación Universitaria son permanentes y no se cuenta con un modelo para almacenar o analizar los datos solicitados. La información suele ser solicitada de manera repetida a las fuentes originales, con la consecuente desventaja de recibir información variable para un mismo indicador en un mismo periodo. Las variaciones no son imputables a las unidades pues su función es mantener sistemas transaccionales y, por el contrario requieren una solución para resguardar su información histórica.

Lo primordial en una institución de educación superior pública con financiamiento gubernamental es optimizar el uso de los recursos. No existe el concepto de ahorro o ganancia y los recursos deben ser invertidos en su totalidad para ser considerados como una institución eficiente. Además, los objetivos y las políticas institucionales se deben cumplir con los recursos otorgados, ser ejecutados en su totalidad en el tiempo planificado y rendir cuentas de lo alcanzado a los órganos supervisores de la Universidad, a otras instituciones gubernamentales y a la sociedad.

Bajo esta premisa, los indicadores clave deben responder a esta optimización de funciones y, por ende, desligados de lo económico en muchos casos, como la

comparación de estudiantes por profesor o el aprovechamiento del espacio físico. Esto no quiere decir que los costos e inversiones serán eliminados, pero sí difieren del concepto de ganancia utilizado en el ámbito comercial.

Para demostrar el uso de este modelo, se analizan los datos de los estudiantes en donde los indicadores son conteos de personas, cursos y créditos en el tiempo para determinar la duración de un estudiante en una carrera específica o el tiempo de permanencia promedio antes de abandonar los estudios por un período prolongado. Los cálculos se realizan en función de variables como por ejemplo el área de procedencia, la carrera que cursa y el tipo de beca asociada.

Para obtener esta información, se construirá un DW con los datos requeridos de los estudiantes al utilizar el modelo propuesto en este trabajo.

1.4 Viabilidad

La viabilidad se define desde cuatro puntos de vista, el técnico, el operativo, el económico y el político. A partir de estos cuatro puntos se determina que el proyecto a desarrollar es viable.

1.4.1 Punto Vista Técnico

Se cuenta con el *software* y el *hardware* para desarrollo. El recurso cuenta con el conocimiento técnico necesario.

1.4.2 Punto de Vista Operativo

La construcción del modelo cuenta con el apoyo de recursos de la administración.

1.4.3 Punto de Vista Económico

La solución propuesta es independiente de un *software* específico y se puede gestionar la infraestructura virtualizada.

1.4.4 Punto de Vista Político

El Consejo Universitario de la Universidad de Costa Rica, en la sesión 4194 del 5 de junio de 1996 expresa la necesidad de “promover un sistema de información gerencial integrado que establezca los requerimientos y los procedimientos con los usuarios

potenciales, a partir de la información básica que se genera en las distintas dependencias académicas y administrativas, sobre la gestión de la Institución" (p. 20).

Asimismo, dentro del Plan Estratégico Institucional 2013-2017 se establece la necesidad de potenciar la transparencia y calidad de la gestión por medio de un sistema información integral. El objetivo general de este proyecto se enmarca dentro del eje sobre gestión institucional, en la política institucional 6.1.1., que manifiesta la necesidad de "Diseñar un sistema de información integral que apoye la toma de decisiones de las autoridades universitarias, la transparencia y la rendición de cuentas." (Universidad de Costa Rica, 2006).

Por lo tanto, desde los puntos de vista técnico, operativo, económico y político, se cuenta con la viabilidad para el proyecto.

1.5 Objetivos

Los objetivos se plantean con la taxonomía original de Bloom (Bloom, 1990) que representa el aprendizaje a partir de 6 niveles cognitivos.

1.5.1 Objetivo General

- Validar un modelo para la creación de un depósito de datos para la Universidad de Costa Rica.

1.5.2 Objetivos Específicos

1. Identificar las características esenciales de un modelo para generar un depósito de datos.
2. Definir técnicas comunes en los casos de estudio, a partir de las fuentes primarias.
3. Preparar un modelo de creación de depósitos de datos enfocado a las instituciones de educación superior pública, a partir de las teorías y técnicas investigadas.
4. Probar el modelo en la construcción de un sistema de información estudiantil.

1.6 Alcances y Limitaciones

Los alcances se consideran los entregables del proyecto y las limitaciones el marco de referencia que indica hasta donde se considera el proyecto como aceptado y completo.

1.6.1 Alcances

Disponer de una lista extensa de variables de las cuales se extrae aquellas que cumplan con el propósito del proyecto. Se cuenta con un listado de las variables específicas que permiten definir un modelo específico para un sistema de información estudiantil en la Universidad.

Realizar una revisión de las propuestas existentes en publicaciones reconocidas, para extraer las características de cada propuesta y proponer un nuevo modelo que sea independiente de una solución de inteligencia de negocios.

Obtener un modelo de construcción de DW enfocado al sector de educación superior pública al utilizar como parámetro las necesidades de información para los datos de un sistema de información estudiantil en la Universidad de Costa Rica.

Presentar un sistema informático funcional, el cual incluirá información real, dirigido a autoridades universitarias, a manera de una herramienta técnica pertinente, como base para la toma de decisiones en el ámbito de las autoridades universitarias.

Validar lo adecuado del modelo, con base en una rúbrica definida previo a su ejecución, otorgando un valor cuantitativo a una serie de características esperadas dentro del modelo.

1.6.2 Limitaciones

El modelo no explica cómo funciona un DW o cómo se debe construir un modelo dimensional. La solución técnica es responsabilidad del encargado de BI o el desarrollador. No obstante, el modelo se generó con una visión general para que una persona sin conocimiento en este campo, entienda cómo funciona, pueda ejecutarlo y adaptarlo a su propio ambiente de trabajo.

El sistema informático será funcional, pero se presentará en un ambiente de desarrollo. La puesta en producción quedará sujeta a la adquisición de licencias y aprobación para su uso dentro de la Universidad por parte de las autoridades pertinentes.

La capacitación en el uso del sistema será otorgada una vez implementado el sistema en producción y estará a cargo de la oficina respectiva.

La minería de datos está fuera del rango de alcance del proyecto. El término minería es mencionada en el proyecto, pues suele ser una herramienta complementaria a los DW y a los sistemas analíticos, sin embargo no se ahonda en ningún proceso de análisis.

No se presenta un procedimiento para el aseguramiento de la calidad de los datos. El aseguramiento de la calidad está fuera del rango de alcance de este proyecto, pero esto no implica que los datos utilizados no fueron sometidos a un proceso de aseguramiento. El aseguramiento de calidad fue aplicado durante la extracción de los datos de las fuentes y una posterior revisión al ser ingresados al DW.

El modelo se enfoca en el sistema de información estudiantil, para la Universidad de Costa Rica como institución de educación superior pública. El modelo debe responder de forma primaria a este requisito para ser aprobado.

1.7 Estado de la Cuestión

La revisión de fuentes relacionadas con el tema se efectuó utilizando una revisión sistemática de acuerdo con los lineamientos propuestos por Biolchini (2005) donde se propone una serie de pasos por seguir para que la investigación sea replicable en cuanto a las fuentes bibliográficas.

El procedimiento se inicia con los objetivos y las áreas a trabajar en la investigación. Se formula y define el problema y sus principales características. A continuación, se realiza la selección de fuentes como bases de datos y bibliotecas. Luego, se realiza la selección de los estudios a los cuales se les aplican criterios de inclusión y exclusión para obtener una lista de la cual se revisa y extrae información sobre la situación actual del tema. Finalmente, se presenta un resumen de los resultados obtenidos.

La revisión sistemática con todos los puntos detallados, se ubica como **Apéndice 1**. En el apéndice se incluyen todos los apartados aplicados para obtener el listado de fuentes, su posterior selección y análisis. En este apartado, solo se menciona la pregunta propuesta y los resultados obtenidos a partir de la revisión sistemática.

1.7.1 Formulación de la pregunta

La pregunta a responder es, ¿Cuál es el estado del arte de modelos de creación de depósitos de datos relacionados con la educación superior? El objetivo de la pregunta es ubicar investigaciones referentes a modelos para crear depósitos de datos con énfasis en el sector educativo.

Las palabras claves utilizadas para la selección de estudios primarios de investigación (Tabla 1) se relacionan con el objetivo general del proyecto, sobre la construcción de un modelo de almacenes de datos en el área de la educación superior.

Área de estudio	**Palabras claves**	**Términos relacionados**
Diseño de un modelo	modelo de diseño *design model* *design modelling* *modelling* *design*	*build modelling*
Depósito de datos	*Data warehouse* *DW*	*DWH*[2] *EDW*[3]
Educación superior	*high education*	sector educativo *educational environment*

Tabla 1. Palabras claves utilizadas en la revisión sistemática.

Fuente: Elaboración propia.

[2] Data Warehouse
[3] Enterprise Data Warehouse

1.7.2 Criterios de revisión

Los criterios para considerar un estudio dentro del trabajo también se relacionan con el objetivo general del proyecto y de sus alcances. La lista incluye los criterios que cada estudio debe cumplir al menos en uno de los puntos. Los estudios deben considerar o examinar:

1. Trabajos teóricos sobre construcción de un depósito de datos.
2. Técnicas recomendadas de modelado en un entorno educativo.
3. Problemas y soluciones al diseño y modelado de un DW en educación superior.
4. Trabajos prácticos los cuales demuestren la factibilidad al aplicar técnicas de modelado de datos relacionados al entorno educativo.
5. Recomendaciones y buenas prácticas en el modelado y mantenimiento de un depósito de datos.

1.7.3 Resumen de resultados

Los estudios obtenidos de la revisión sistemática se someten a un proceso de extracción de las ideas principales y se documentan en el resumen de resultados. De esta manera obtenemos el estado del arte sobre modelos de construcción de un DW para instituciones de educación superior.

El primer estudio menciona los problemas encontrados al modelar e implementar un depósito de datos para el sistema nacional de información de la educación superior en Croacia, para mejorar la planificación y operación de las actividades educativas (Baranović, 2003). La mejora propuesta por el estudio es crucial para este proyecto al ser un sistema de información que responde a las necesidades de la UCR y, de manera directa a la Oplau. La oportunidad de mejorar estos procesos de planificación, y en especial, lo referente a la información estudiantil, le da un apoyo adicional al proyecto y sus repercusiones en la Universidad.

El estudio propone una estructura para análisis de datos rápidos y eficientes por parte de las autoridades (decanos, vicerrectores) y una interfaz web intuitiva para que los usuarios obtengan sus reportes con un servicio de alta disponibilidad. La implementación se divide en cuatro elementos:

1. La arquitectura del DW.
2. El modelo ETL para el DW.
3. La seguridad de la información.
4. La presentación de los datos.

La arquitectura del DW 1) se inicia desde un sistema transaccional/relacional, se realiza la copia de los datos a una base de datos multidimensional y se transforma en un cubo. La base de datos multidimensional utiliza el esquema de estrella. A partir de los estudios primarios, el esquema en estrella es propuesto por Baranović (2003), Hua-long Zhao (2008) y Kimball & Ross (2011) y solo en Ningning (2010) el esquema de copo de nieve. En el modelo ETL 2) se hacen cargas nocturnas para el SA y del OLAP. Enfatiza en la protección de la comunicación 3) en el manejo de las cuentas de acceso y en donde la cantidad de funciones es similar a la cantidad total de usuarios, por lo específico de los permisos de lectura.
Por último 4) se establecen 3 categorías de consultas a los usuarios. Las categorías utilizadas son consultas predefinidas, consultas personalizadas y consultas de agregación personalizadas. Las consultas predefinidas son SQL o MDX con un filtro automático desde un *script* de acuerdo con la función y permisos del usuario. Las consultas detalladas son SQL parametrizados desde listas o campos de texto y se realizan al DW; son consultas atípicas por el grado de detalle para un sistema analítico. Las agregaciones son *scripts* en DHTML a los cubos OLAP que permiten *drill-down* y la aplicación de filtros.

El estudio sobre el análisis de la carrera seleccionada por los estudiantes (Hua-long Zhao, 2008) concentra la investigación en el diseño de la estructura del modelo de datos como punto crucial de un DW. A partir del modelo se define el tipo de análisis posible y la complejidad del mantenimiento del sistema, al incluir procesos de minería de datos. Esta idea la reitera Ningning (2010) y la complementa con la descripción de 5 pasos para la creación del DW. Los pasos mencionados son:

1. Modelado de datos -*blueprint* del DW-.
2. Extracción de los datos como una carga completa o de manera incremental.
3. Administración de los datos eliminando redundancias, al realizar conversión de tipos, agregación de datos y aseguramiento de la calidad.
4. Carga del DBMS con procesos y accesos masivos.
5. Acceso a los datos por medio de funciones (roles).

En general los DW son determinados a partir del objeto de estudio y deben girar alrededor de este objeto.

La mejora en el rendimiento de las consultas (Ningning, 2010) se realiza agregando datos calculados directo en la tabla de hechos. Por ejemplo, se agregan datos de repetición, promedio de notas, cantidad de cursos con más de 90 o cantidad de cursos perdidos, para esto se crean múltiples tablas de hechos como la sumatoria de cursos perdidos y ganados sin necesidad de calcularlos en tiempo real.

La solicitud de estas variables mencionadas como agregaciones pre calculadas, deben ser incluidas posterior a un análisis de la información disponible -la relación de los datos deseados contra los posibles- Además, se resalta que las métricas no evalúan el rendimiento a partir de ganancias o retribuciones (Selmoune 2008), por el contrario miden la proporción de éxito y fracaso. Pero, ¿Qué determina el éxito o fracaso -en el marco de este proyecto- de un estudiante? El estudio propone un proceso de 5 pasos de diseño (Figura 1).

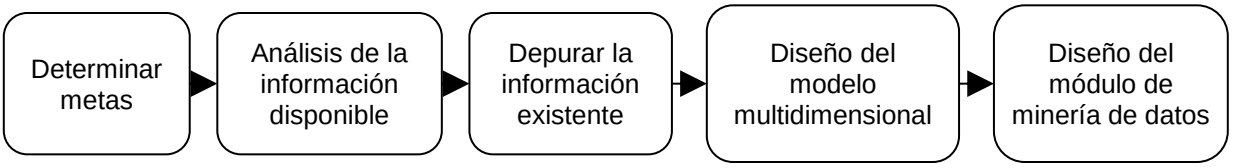


Figura 1. Proceso de diseño en cinco pasos.
Fuente: Selmoune 2008.

Las metas se discuten con los diferentes participantes y se definen los análisis pertinentes. El análisis de la información muestra cuáles datos hay para cumplir con las metas definidas y cuáles metas se descartan porque no hay datos. El diseño multidimensional mide la relación entre variables, para determinar el éxito o fracaso de la medida. Se recomienda crear múltiples tablas de hechos para estas medidas, como por ejemplo una tabla de hechos para medir la aprobación de los exámenes, una para la aprobación por profesor y otra para medir el ausentismo. La minería de datos, aunque está fuera de los alcances de este proyecto, es parte de los pasos de diseño propuestos, y un proyecto que se puede ejecutar a posterior.

Por último, el estudio analizado de Kimball (Kimball & Ross, 2011) trata sobre herramientas para el DW. El libro contiene una sección específica sobre DW para el sector educativo. En él se pone énfasis en el sector de educación superior y menciona algunas estrategias recomendadas a partir de su conocimiento experto en trabajo de campo. Se mencionan cuatro estrategias:

1. La matriz de procesos.
2. El rastreo de hitos (*accumulating snapshot*).
3. Las tablas sin medidas asociadas (*factless tables*).
4. El manejo de eventos que no ocurrieron.

Capítulo 2. Marco Teórico

La propuesta de un modelo para la construcción de un DW enfocado en las instituciones de educación superior (IES), requiere la definición de un marco teórico como fundamento para establecer un procedimiento formal de investigación apoyada en trabajos previos relacionados con el tema. La búsqueda de fuentes primarias derivadas del marco conceptual, junto a otras fuentes de información secundaria que amplíen los conceptos tratados, permiten la generación de nuevo conocimiento para crear un modelo con características específicas a la Universidad de Costa Rica, la cual es la institución a la que se enfoca este trabajo.

El capítulo inicia con la definición de un DW, concepto elemental en este trabajo y el que define el norte en cuanto al modelado, técnicas de construcción y a su capacidad en el análisis de los datos. El DW se acompaña -aunque no es sinónimo- de datos OLAP. Para una mejor comprensión, se contrasta la presentación de los datos relacionales contra un formato analítico, como preámbulo a la definición del concepto OLAP. La arquitectura de un DW se representa con técnicas de diseño y construcción.

Luego se define de manera general lo entendido como modelo y sus características deseables. La caracterización del modelo es acorde con las necesidades del proyecto, por tanto, se establece su relación con la construcción de un DW y con las IES. El modelado incluye la mención de herramientas, fases recomendadas, arquitectura, implementación y técnicas esperadas.

El objetivo del proyecto es enfocarse en las IES públicas y realizar una definición general. Las características propias de este mercado es lo que diferencia esta propuesta de otros modelos de construcción. Las IES presentan similitudes en requerimientos de información con la empresa privada, pero se hace énfasis en las diferencias con estos mercados. Las particularidades mencionadas de las IES incluyen referencias a los beneficiarios, a saber qué promueven y a que responden. También, se incluye como evaluarlas y ejemplos ya aplicados.

De manera complementaria, se detallan características especiales de un DW para las IES públicas, incluyendo técnicas y recomendaciones obtenidas de las fuentes primarias y las consultadas para este marco teórico.

2.1 Características de un Data Warehouse

El DW es una colección de datos integrados con el elemento de tiempo, enfocado en la toma decisiones, el cual almacena información histórica, agregada, extraída de múltiples fuentes que pueden ser heterogéneas y distribuidas (Hüsemann, Lechtenbörger & Vossen, 2000, p.1). Además, el DW debe cumplir con una serie de requerimientos para ser considerado útil para el negocio, los cuales, de acuerdo con Kimball (2011) son los siguientes: 1) que la información sea de fácil acceso y comprensible para el usuario; 2) Los datos guardados deben representar una única verdad; 3) Los datos se consideran no volátiles, por lo tanto las actualizaciones y borrados suelen ser restringidas. Asimismo, el DW debe ser adaptable y permeable a los cambios, sin modificar o invalidar los datos existentes al hacerlo. Es necesario que el negocio comprenda su importancia y se le brinde la divulgación necesaria.

Los *data mart* (DM) son un subconjunto del ámbito de un DW y responden a un solo proceso del negocio. Los DM suelen contener datos con la máxima granularidad, presentados en un modelo dimensional y deben unirse entre sí por medio de las dimensiones comunes. Se utiliza la noción de Kimball para establecer la existencia de múltiples DM, que al unirlos, generan el DW. Se utilizan dos conceptos, 1) DM de primer nivel que se deriva de una única fuente de datos y 2) DM consolidados o de segundo nivel, donde las medidas provienen de diversos procesos del negocio (Kimball, 2011, p.396-402). El uso de uno u otro depende de las necesidades, del objeto analizado y contemplar que la consolidación implica trabajos adicionales de homologación.

La divulgación se realiza cuando los usuarios pueden disponer de los datos y esto se da al generar los datos para el procesamiento analítico. Los OLAP permiten realizar análisis de datos multidimensionales a partir de los DW y proveen la capacidad para realizar cálculos complejos, análisis de tendencias y modelado de datos. El procesamiento analítico permite a los usuarios efectuar sus propios análisis de los datos en múltiples dimensiones, al propiciar un mejor entendimiento para la toma de decisiones (Olap.com, 2016).

La construcción de un DW implica el uso de una arquitectura adecuada para implementar el modelo. La primera consideración es la separación de los sistemas transaccionales de los utilizados para el DW. La separación es necesaria por 1) la diferencia en cargas de

trabajo 2) la infraestructura tecnológica, 3) el uso de las bases de datos especializadas para la manipulación de datos OLAP, 4) la conservación integral de los datos y 5) aspectos de seguridad en el acceso a la información.

La arquitectura del DW más aplicada consiste en 3 capas de acceso -3 tier- (Sharma, 2009, p.3); en la Figura 2 se muestra 1) la capa de datos, 2) la capa de aplicación y 3) la capa de presentación (Paoletti, 2012).

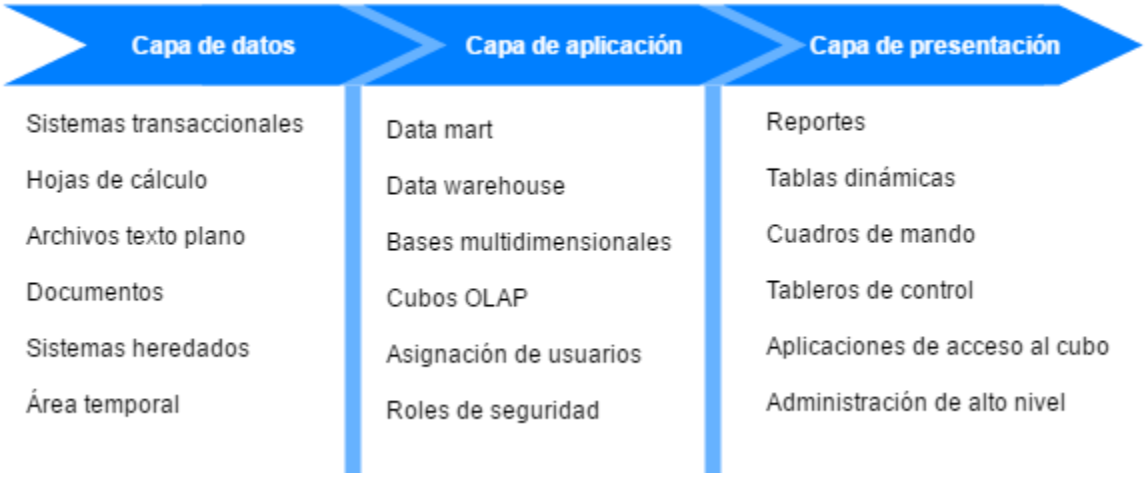


Figura 2. Arquitectura propuesta de un DW en 3 capas.
Fuente: Elaboración propia con base en Moura (2015)

La capa de datos es donde residen las bases de datos fuentes y se les aplican los procesos de extracción y transformación ETL. Los autores difieren en donde colocar el área temporal (*staging area*), algunos lo hacen en la capa de datos y otros en la de aplicación. En este proyecto se asume el *staging area* (SA) dentro de la capa de datos por considerarlo ligado al proceso de ETL y tan solo como una etapa transitoria hacia el DW o DM. La capa de aplicación incluye los DM, el DW, las bases multidimensionales y los cubos OLAP, incluye los accesos de aplicaciones para consultas directas, la asignación de usuarios y funciones para el control de la seguridad en el acceso a los datos. Por último, la capa de presentación, que permite a los usuarios de alto nivel acceder a los datos de manera práctica. En esta capa se incluyen las aplicaciones de acceso a los cubos, generación de reportes, disponibilidad de cuadros de mando, de tableros de control y de la administración del conjunto de datos.

La definición anterior de la arquitectura ayuda a comprender el flujo de los datos durante el ciclo de vida del DW y a definir el diseño necesario para la infraestructura del DW. La técnica de diseño y construcción propuesta por Selmoune & Alimazighi (2008, p.2) consta de 5 pasos que consisten en:

1. Determinar las metas que se discuten con los diferentes participantes y se definen los análisis pertinentes.
2. Analizar la información disponible al mostrar los datos existentes para cumplir con las metas definidas y cuáles metas se descartan por no existir fuentes de donde extraerlas.
3. Depurar la información existente.
4. Diseñar el modelo multidimensional que refleje la relación entre variables por medio de tablas de hechos.
5. Crear un módulo de minería de datos para hallar patrones entre las variables.

Para extender esta propuesta y como parte del proyecto, se requiere la definición de lo que se entiende por modelo y sus alcances.

2.2 Caracterizar el modelo dentro de un DW

Se entiende por modelo una "propuesta, normalmente de carácter teórico-práctico, que tiene una serie de características que se consideran dignas de emular. Generalmente, el modelo ilustra una situación deseable para ser analizada y puesta en práctica en un contexto educativo similar, o bien adaptarla a otras características del entorno." (RIACES, 2004). El modelo debe responder a las necesidades del objeto de estudio, conservando la generalización necesaria para ser aplicado en contextos similares y que pueda ser reproducible.

La definición del modelo, unido a lo teorizado sobre almacenes de datos, permite establecer modelos específicos. A continuación, se definen el modelado dimensional, el diseño del DW en 3 fases, el tipo de implementación de abajo hacia arriba (*bottom-up*), los metadatos y las características de las herramientas de modelado.

El diseño del DW propuesto por Selmoune & Alimazighi (2008, p.2) en 5 etapas, incluye el diseño del modelo dimensional. Una base de datos dimensional es una base de datos

desnormalizada que contiene dimensiones comunes y tablas de hechos con medidas categorizadas por sus dimensiones (Rainardi, 2008, p.30). La construcción del modelo dimensional se recomienda utilizando dos fases, 1) la creación de un modelo lógico, donde se reflejan las necesidades del negocio, y 2) el modelo físico en donde se crean las estructuras y se logra la optimización de las consultas para los análisis (CA, 2008, p.7).

La propuesta de Castro (2011, p.10) es diseñar el DW en 3 fases:

1. Modelo de negocios.
2. Modelo dimensional.
3. Modelo físico.

La primera fase, es decir el modelo de negocios, requiere un análisis estratégico en donde se identifiquen y prioricen los procesos cruciales del negocio. Además, se debe crear el modelo del negocio, al identificar las medidas, dimensiones, granularidad, definiciones, reglas, verificar fuentes de los datos y documentar los metadatos. La segunda fase corresponde al modelo dimensional, en donde se convierte las medidas en tablas de hechos y establece su relación con las dimensiones.

En esta segunda etapa se establece el tipo de esquema a utilizar, ya sea el modelado en estrella o copo de nieve. En el desarrollo de este proyecto, se utiliza el esquema en estrella. La selección de este tipo de esquema se establece a partir de lo obtenido de los artículos consultados en el estado de la cuestión, en donde 3 de 4 fuentes primarias proponen el uso de este modelo en ambientes similares relacionados con las IES. Además, el esquema en estrella es (Castro, 2011, p.23) fácil de visualizar, la llave primaria se define en un solo nivel de profundidad, permite respuestas rápidas a las consultas y reduce la cantidad de uniones entre tablas.

La tercera etapa de diseño establece el modelo físico en donde se transforma el modelo dimensional a un modelo físico de implementación, se define la estrategia de implementación para tablas, realiza el dimensionamiento de la base de datos, define nomenclatura estándar para los objetos, propone las estrategias de indexado y particionamiento y determina la arquitectura de almacenamiento.

Luego de realizar el diseño del DW, se requiere establecer el modelo de implementación apropiado. Las dos implementaciones revisadas son la *top-down* y la *bottom-up*.

- Top-down: Requiere un esfuerzo inicial en planeamiento y diseño, al requerir información de cada grupo de trabajo, departamentos y procesos de negocio involucrados en la implementación del DW. Además, supone una inversión de tiempo y dinero a mediano plazo con recursos altamente capacitados. Los beneficios de una implementación *top-down* permiten una mejor consistencia de los datos, refuerza las reglas del negocio y facilita el consumo posterior de datos por parte de los DM. La recomendación es que este tipo de implementación se realice en una organización con un departamento centralizado de TI.

- Bottom-up: Implica la creación de DM antes de crear una infraestructura global de los datos. Esto no elimina el desarrollo de una solución global, esto indica que será construida de una manera incremental. Los beneficios incluyen un desarrollo más rápido, diseño sencillo y suele ser más económico comparado al *top-down* al requerir menos equipo y recursos. Los peligros son la redundancia o inconsistencias de los datos entre los DM, retrabajo a lo largo del proceso de implementación y complejidad en la integración de los datos conforme avanza el proyecto. (Ballard, 1998, p.18)

La implementación *bottom-up* es la utilizada en este proyecto y forma parte del modelo propuesto. Esta propuesta de implementación es recomendada en entornos donde cada departamento cuenta con su unidad de TI. También, es más rápido su desarrollo y los entregables están disponibles en un tiempo adecuado para mantener el apoyo de la administración al proyecto. Además, los costos de infraestructura y recursos se incrementan de manera paulatina conforme al crecimiento del proyecto.

La aplicación del diseño *bottom-up* no suprime la creación de un marco general (Malinowski & Zimányi, 2008, p.247). En este marco general se establece como debe ser administrada la información desde un punto de vista macro de la organización y sus relaciones, pero su desarrollo se da a partir de DM individuales generados de manera iterativa e incremental.

La metadata es requerida como parte del modelo y permite que los datos sean comprendidos, utilizados de manera amplia y correcta. Si el diseño del DW es iterativo, de igual manera serán los metadatos (Ballard, 1998, p.53). Los metadatos incluyen la documentación del diseño de los procesos, del proceso de desarrollo y mantiene un registro de los cambios y mejoras. La recolección de los metadatos se puede realizar de manera automática con las herramientas de modelado de datos, las ETL o por un proceso manual.

La documentación recomendada de los metadatos contiene (Ballard, 1998, p.66): un listado de los objetos disponibles en el DW con los esquemas, dimensiones, hechos y medidas. Los esquemas deben tener un nombre, una descripción de su contenido y su función. El nombre debe ser descriptivo y suele ser el mismo nombre de la tabla de hechos. Para las dimensiones, atributos y medidas se debe incluir un nombre, una definición y las acepciones con las cuales se conoce en la organización un mismo objeto. Los atributos incluyen el tipo de dato, dominio y reglas para determinar los valores válidos. Los hechos deben incluir la frecuencia de actualización de los datos, la granularidad, las dimensiones asociadas con especial énfasis en la dimensión de tiempo y cómo se relaciona. Para finalizar con este apartado, se mencionan las características necesarias para una herramienta de modelado que incluya el manejo de los metadatos.

La selección de una herramienta de modelado debe considerar aspectos como el manejo de notación específica para el modelo dimensional, permitir el modelado a partir de ingeniería inversa e ingeniería directa (a partir de DDL), mapeo de los datos e inclusión automatizada de metadatos a partir de los objetos existentes (Ballard, 1998, p.155). La propuesta requiere el uso de una aplicación la cual refleje de manera completa el modelo dimensional y relacional del DW. Estas características definen el tipo de herramienta a utilizar en el modelado.

2.3 Evaluación de las instituciones de educación superior públicas

En Costa Rica, la universidad pública se puede definir como “aquella institución pública autónoma con una misión orientada hacia la educación superior universitaria que financia, con recursos del Estado, actividades las cuales complementan su labor docente y que le

permiten atender necesidades específicas de la sociedad…" (Granados, 2015). Tal como lo expresa la definición anterior, las universidades públicas no solo se deben enfocar en la docencia, sino también en realizar aportes a la sociedad que respondan a los recursos otorgados.

De manera complementaria, Quesada (citado en Granados, 2015) indica que "las Instituciones de Educación Superior con financiamiento público, deben rendir cuentas al Estado, acerca de su desempeño en la ejecución de la labor que les ha sido confiada" (p.25). Con esta afirmación se muestra la necesidad de que las IES, además de cumplir con las actividades sustantivas a las cuales se deben, evalúen su desempeño y velen por el buen uso de los recursos, incluida la optimización de los gastos de la institución.

La evaluación y la optimización de recursos requiere de mecanismos para la medición que ofrezcan un seguimiento a partir de la planificación en el uso de los recursos, su ejecución y su cumplimiento. Estos procesos de seguimiento suelen ir de la mano de una lista de indicadores definidos por la misma institución y por organismos externos que, en Costa Rica, es la Contraloría General de la República la llamada a solicitar y aprobar estos indicadores ("Acerca de la CGR", 2016).

Los indicadores son la principal herramienta de medición, ya que responden tanto a políticas institucionales como gubernamentales. Además, los indicadores son una fuente valiosa para la toma de decisiones (Grad, 2009). Es en este contexto donde la disponibilidad, velocidad y pertinencia de la información cobra relevancia. Por ejemplo, en el INFOACES (Universitat Politècnica de València, 2013) se menciona la importancia de contar con sistemas de información cualitativos, objetivos, actualizados, los cuales respondan a las políticas y objetivos estratégicos de las IES (p.12), y los esfuerzos de mejora por parte de Universidades estatales como la Universidad Nacional de Costa Rica en materia de indicadores para mejorar su gestión (p.17). Este proyecto no va a definir indicadores para las IES, pero es importante entender su relevancia y conocer la diversidad de fuentes requeridas para su recopilación.

La recopilación de indicadores se realiza a partir de las fuentes primarias, lo cual implica una estandarización de los datos para unificarlos en el contexto de indicadores institucionales. Esta estandarización requiere la extracción de los datos a partir de las

fuentes, convertirlos a un lenguaje común con valores, métricas normalizadas y su posterior resguardo en un repositorio único para su consulta por parte de las diferentes unidades de trabajo que así lo soliciten. Estos procesos son homólogos a la primera fase de un DW, cuando se ejecuta el proceso de ETL.

Además, con base en la definición de un DW en el apartado anterior, se explica porque un DW es la solución adecuada a la necesidad de información para una IES que utiliza o utilizará los indicadores como una fuente de información para la toma de decisiones y para la rendición de cuentas. Esta necesidad se demuestra en Baranovic (2003), Zhao (2008), Ningning (2010) y Selmoune & Alimazighi (2008) en donde concuerdan que un DW es la herramienta adecuada para planificar, evaluar y mejorar el trabajo en las IES.

2.4 Técnicas de construcción de un DW para IES públicas

Las universidades estatales son entidades sin fines de lucro que no utilizan el concepto de rentabilidad, sino de aprovechamiento y eficiencia en el uso de los recursos (Selmoune & Alimazighi 2008, p.1). Esta diferenciación requiere el análisis de los datos con un modelo distinto, con la inclusión de técnicas específicas para este contexto. Las técnicas abarcan desde la etapa de diseño hasta la implementación del DW.

Las técnicas vinculadas al diseño del DW se refieren a la manera adecuada de brindar la información evaluando al público meta y se recalca la obligación de vincular los análisis al plan estratégico institucional y a la misión de la institución. Además, se recomienda la categorización de la información de acuerdo con la complejidad de la información solicitada. Asimismo, se mencionan técnicas de construcción del DW como la matriz de procesos y otras específicas a la creación de las tablas de hechos, como las *factless fact tables*, uso de datos agrupados y finalizar con el *dimensional drill across*.

Para responder de manera adecuada a los objetivos, se toma en cuenta lo mencionado en el apartado 2.3, en donde se recalca que las universidades públicas deben rendir cuentas sobre su gestión a múltiples actores. En este sentido, tal como lo indica Malinowski & Zimányi (2008, p.257) en su caso de estudio, la información se debe otorgar de manera diversificada de acuerdo con el público meta al cual está dirigido, en este caso

a lo 1) interno de la institución para toma de decisiones, a 2) las instituciones de control gubernamentales y 3) a la sociedad.

Análogamente, el DW debe responder a esta necesidad de información, estar alineado al plan estratégico y enfocado a la misión institucional. Este análisis permite definir parámetros para solicitar la información de las diferentes fuentes y conocer en donde se puede proveer información para: toma de decisiones, investigación y promoción de la transparencia universitaria.

Como apoyo para diversificar la información ofrecida se utiliza una categorización en donde la información se divida por la complejidad en la visualización de los datos. Como parte de este modelo Baranović (2003, p.126) señala la conveniencia de utilizar una separación en 3 categorías: datos estáticos[4], datos dinámicos[5] y acceso a los datos por consultas complejas (MDX). Esta categorización simplifica el entendimiento de los datos, tomando en cuenta necesidades específicas de todos los actores involucrados.

La complejidad en la construcción de un DW para universidades radica en la cantidad de diferentes actividades en las que se involucra la institución. “Las universidades participan en docencia con sus estudiantes y profesores, manejan recurso humano, se dedican a labores de investigación, tienen residencias estudiantiles, invierten en propuestas innovadoras, velan por la salud de los estudiantes -incluyendo servicios de emergencias- y organizan eventos” (Kimball, p.325-326). Para administrar esta diversidad de negocios en el DW y para mantener la coherencia con las necesidades institucionales, se recomienda el uso de la técnica de matriz de procesos. La técnica de la matriz de procesos no es de aplicación exclusiva a un DW del sector educativo, pero es esencial para tener el panorama institucional. La generación de los diferentes DM debe estar alineada a la visión empresarial del DW.

Tal como lo explica Ross (2005) la matriz se crea a partir de los procesos esenciales de la institución. La matriz se compone de filas y columnas, en donde las filas representan los procesos del negocio y las columnas las dimensiones. El cruce de un proceso con una

[4] Consultas predefinidas.
[5] Consultas predefinidas con la posibilidad de filtrar por una cantidad limitada de categorías.

dimensión debe ser anotado en la matriz (process-centric rows, parra. 1). En la Figura 3 se muestran las relaciones de los procesos con los atributos.

	Date	Customer	Product	Sales Rep	Deal	Warehouse	Shipper
Quoting	X	X	X	X	X		
Ordering	X	X	X	X	X		
Shipping to Customer	X	X	X	X	X	X	X
Shipment Invoicing	X	X	X	X	X	X	X
Receiving Payments	X	X		X			
Customer Returns	X	X	X	X	X	X	X

Figura 3. Ejemplo de matriz de procesos para venta de productos.
Fuente: Kimball (2011, p.168)

Lo usual es completar una fila -proceso- a la vez, por lo tanto la técnica también permite priorizar los esfuerzos en el desarrollo y que la alta gerencia entienda las necesidades pendientes de forma gráfica y ágil.

Las tablas de hecho son las tablas centrales de los diferentes esquemas del DW y contienen los resultados considerados estratégicos para el negocio. El sector educativo requiere la aplicación de técnicas específicas para elaborar las tablas de hecho y así responder a los análisis solicitados. La primera técnica propuesta es la *factless table*. La característica principal es que la tabla de hechos no contiene medidas, por lo tanto su función será el conteo de eventos y ocurrencias (Kimball 2011, p.329-332). Este comportamiento es especialmente útil en el entorno educativo, pues variables como la cantidad de cursos ofrecidos, estudiantes matriculados, títulos otorgados son solicitados de manera constante y para los cuáles la cantidad lograda es el objetivo.

El uso de datos agregados es una técnica recomendada por Ninging (2010) cuando los datos se suelen consultar agrupados. Por ejemplo, para cálculos de cursos y de rendimiento académico, se almacena por estudiante la cantidad de cursos matriculados, la cantidad de cursos perdidos, la nota ponderada del período, el rendimiento promedio y

si la situación histórica del estudiante amerita una atención individualizada. De igual manera se utilizan valores agregados para determinar el promedio de aprobación de una materia o de un grupo en particular.

Otra técnica, el *dimensional drill across*, permite compartir una dimensión en diferentes esquemas (IBM, 1998, p. 153) si se cumplen las siguientes características: 1) comparten los mismos atributos, 2) la granularidad es afín y 3) su semántica es equivalente. La existencia de una sola verdad para los atributos de un DW es una característica explicada en el apartado 2.1. La restricción es la de cumplir las 3 características mencionadas. Si dos dimensiones no tienen el mismo significado, se deben nombrar diferente para evitar confusiones en el usuario final. Esta técnica permite relacionar distintos procesos del negocio y analizarlos mediante las dimensiones comunes.

Capítulo 3. Marco Metodológico

El proyecto incluye componentes de investigación los cuales requieren una base formal para constituir el aparato crítico y dejar claro el enfoque investigativo utilizado. Esta base incluye describir el tipo de investigación realizado, el diseño y el alcance del proyecto. También se detalla la población involucrada, así como los instrumentos de recolección de datos y las estrategias de desarrollo.

3.1 Tipo de Investigación

El tipo de investigación en este proyecto es investigación aplicada. El objetivo principal en este tipo de investigación, tal como lo define Padrón (citado en Vargas Cordero, 2009) es resolver un problema práctico en un entorno específico, acotado con el campo de estudio seleccionado (p. 159). En este proyecto el problema se establece en la falta de un modelo específico para generar un DW en un entorno dentro de una universidad estatal, en donde sus características particulares dificultan la aplicación de modelos construidos con un enfoque empresarial.

En este contexto y a partir de casos particulares de estudio obtenidos de las fuentes primarias, investigación sobre modelos aplicados en otros campos y trabajos teóricos sobre generación de modelos para un depósito de datos, pues se va a proponer un modelo de creación de DW enfocado a instituciones de educación superior.

3.2 Alcance Investigativo

El alcance investigativo de este proyecto es de carácter exploratorio. La investigación exploratoria se efectúa sobre un tema poco estudiado, en donde los alcances se limitan a la observación y al registro objetivo de lo encontrado (Barrantes, 1999, p. 131). Este proyecto considera las fuentes primarias, los casos aplicados de modelos y los estudios teóricos sobre modelado para concebir un modelo para un DW dirigido a las instituciones universitarias.

Por otra parte y de acuerdo con lo establecido en el tipo de investigación, se propone una solución aplicada, en donde los modelos observados y documentados, serán utilizados para generar un nuevo modelo con las características específicas requeridas para una institución universitaria.

3.3 Enfoque

El enfoque de la investigación será un abordaje alternativo. El abordaje alternativo considera que no se puede separar lo cualitativo y cuantitativo, pues estos conceptos por su naturaleza, nunca han estado desligados (Chavarría, 2011), por lo tanto, no tiene sentido interpretarlos como enfoques separados.

El abordaje alternativo plantea tres dimensiones las cuales deben ser explicadas, para establecer de manera correcta, el enfoque de la investigación desde esta nueva perspectiva. Las dimensiones propuestas son la epistemológica, ontológica y axiológica.

Epistemología

La dimensión epistemológica implica que la investigación se hará desde una perspectiva de ejecutor, para lo cual se aplicará una revisión de las diversas fuentes primarias y de apoyo, se extraerán los datos comunes y de contraste para establecer la situación actual. A partir de estos datos, se determinarán las estrategias apropiadas para generar un nuevo modelo de creación de un depósito de datos. El proyecto va a exigir involucramiento del investigador.

Ontología

Tal como lo define Giménez et al. (2008) la ontología es “el conjunto de objetos que pueden ser representados ... y que se asocian mediante definiciones con los nombres de las entidades en el universo de discurso (por ejemplo, clases, relaciones, funciones u otros objetos) con cierto texto expresado en lenguaje natural que las describe” (p. 2). Lo que procede es partir de una base de conocimiento ya existente, como teorías establecidas para la creación de un depósito de datos o modelos aplicados en otros campos de estudio. A partir de estos conocimientos generales y del estudio de casos particulares, se va a proponer un nuevo modelo de creación de DW para instituciones de educación superior.

La Figura 4 representa el uso de teorías ya existentes para justificar el modelo obtenido durante la investigación y desarrollo del proyecto.

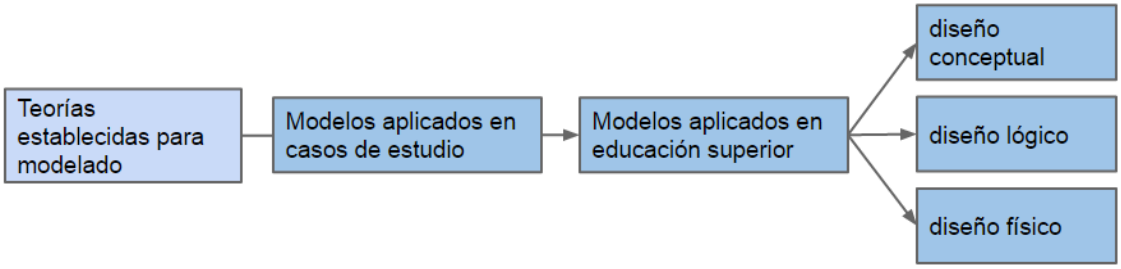


Figura 4. Ontología de un modelo de creación de un DW para las IES.
Fuente: Elaboración propia.

Axiología

La dimensión axiológica establece las medidas para determinar lo adecuado del modelo al negocio, para esto, se fija una escala de valores donde se perfila lo adecuado del modelo para cumplir con los requerimientos del usuario, a partir de los criterios de evaluación. Los criterios y la escala de valores se definen en la Tabla 2.

Criterio	Valor	Valor mínimo	Valor adecuado	Valor óptimo
Satisfacción de usuarios.	35%	10% si la evaluación arroja una satisfacción mayor a 50.	25% si la evaluación arroja una satisfacción mayor a 70.	35% si la evaluación arroja una satisfacción mayor a 90.
Total de indicadores solicitados vs. los indicadores implementados.	20%	10% si se cumplen en una proporción mayor a 50.	15% si se cumple en una proporción superior a 70.	20% si se cumplen en una proporción de 90.
Tiempo de respuesta máximo a cualquier consulta a la base de datos.	10%	0% si tarda menos de 8 segundos en responder.	5% si tarda menos de 5 segundos.	10% si tarda menos de 3 segundos.
Cantidad de problemas encontrados en el periodo de prueba (pulgas).	15%	5% si se reportan menos de 15 problemas.	10% si se reportan menos de 10 problemas.	15% si se reportan menos de 5 problemas.

Cumplimiento en el plazo establecido para la construcción del DW.	20%	Un 2% si se tarda un 1 mes más de lo planificado.	Un 5% si se tardan 2 semanas más de lo esperado.	Un 10% si se entrega en el tiempo establecido.

Tabla 2. Criterios de evaluación para medir lo adecuado del modelo al negocio.
Fuente: Elaboración propia.

La ponderación total es de 100, en donde la nota mayor indica que tan adecuado resulta el modelo para la construcción del DW en la universidad. Para considerar el modelo como adecuado se recomienda un valor arriba del 80%, un modelo aceptable arriba de un 50%, y el modelo no es recomendable si se obtiene una nota inferior al 50%.

3.4 Diseño

El diseño de la investigación define las estrategias básicas para generar información confiable e interpretable (Bautista, 2011, p. 150). La investigación-acción resuelve problemas cotidianos en la búsqueda de mejores prácticas a problemas concretos. (Álvarez-Gayou, 2003). Las estrategias estarán dirigidas a resolver un problema puntual, en este caso, proponer un modelo de creación de un DW para una IES, delimitado a la Universidad de Costa Rica.

3.5 Población y Muestreo

La población será el personal de la Oficina de Planificación Universitaria y de la Oficina de Registro quienes serán los beneficiarios iniciales del proyecto. El personal será encuestado como parte de los criterios de medición definidos en la axiología del enfoque alternativo para obtener un valor para para el grado de satisfacción del usuario y determinar si la solución es adecuada.

El muestreo será intencionado por lo limitado de la población sujeto de esta medición. La medición será una encuesta que pueda establecer el criterio de satisfacción. La encuesta será aplicada a los usuarios que interactúan con la aplicación y a los usuarios indirectos, que obtienen información ya generada por otro usuario. La medición de usuarios directos

e indirectos es apropiada al evaluar un modelo de construcción y no un sistema específico informático.

3.6 Instrumentos de Recolección de Datos

La recolección de los datos se da por medio de entrevistas, encuestas y la investigación documental. Los tres instrumentos son utilizados en diferentes etapas del proyecto para cumplir los objetivos planteados.

La investigación documental se requiere para discernir algunos de los requerimientos iniciales de información a partir de las diferentes publicaciones y documentación con información histórica solicitada por diferentes entes gubernamentales y no gubernamentales, como la Contraloría General de la República (CGR), Comisión Nacional de Rectores (CONARE), Plan Nacional de la Educación Superior (PLANES) y Banco Mundial (BM).

Las entrevistas se dividen en dos tipos, la de contextualización y la semiestructurada con un enfoque guiado. La entrevista de contextualización permite al investigador una visión general de los requerimientos, acercamiento rápido a las personas y como un abordaje inicial al problema. La entrevista semiestructurada se aplica durante las diferentes iteraciones del desarrollo de la propuesta, como un método de afinamiento y para corroborar el cumplimiento de las expectativas de la población en torno al modelo.

Las encuestas serán utilizadas para responder al criterio de satisfacción del usuario. El criterio es uno de los establecidos para medir lo adecuado del modelo propuesto. La construcción de esta medida está planteada en el enfoque alternativo.

3.7 Estrategia de Desarrollo de la Propuesta

El proyecto consta de 2 etapas. La primera etapa consiste en generar un modelo para la construcción de un DW para una IES, en específico, la Universidad de Costa Rica, en donde los ingresos no se obtienen a partir de la ganancia y el objetivo es la optimización de recursos recibidos. Este modelo debe responder primeramente a las necesidades de la Universidad en cuanto al tratamiento de los datos y luego intentar generalizarlo a otras IES. La segunda etapa es la aplicación de este modelo construyendo un sistema de

información estudiantil, para el cual se requiere obtener un producto funcional. A partir de esta división, las estrategias de desarrollo tendrán un tratamiento diferente según la etapa a trabajar.

La etapa de especificación del modelo es un proceso lineal. La generación del modelo será a partir de las fuentes primarias, la información recolectada durante la investigación del marco teórico y nuevas propuestas a partir de la experiencia profesional en procesos similares. El resultado será un procedimiento en el cual se describen las etapas para generar un DW acorde a las necesidades de la institución, al aplicar modelos formales de construcción.

La construcción del sistema de información es un desarrollo informático y se considera que una estrategia iterativa puede ser más provechosa con respecto a otros enfoques, ya que, por experiencia personal en procesos similares, siempre es necesario revisar y validar los procedimientos con más de una iteración.

Es bueno recordar que el proceso iterativo no es lo mismo a un proceso incremental. En el incremental se divide el proyecto en varios segmentos (Ortiz, 2012a) y cada uno se completa antes de iniciar el siguiente y así logra que el sistema se complete brindando al usuario los módulos esenciales mientras se construyen los menos prioritarios. En el iterativo se crea todo el sistema, pero sin esperar que cumpla todos los requerimientos desde un inicio y conforme avanzan las iteraciones, se van completando las solicitudes. La aplicación evoluciona con los cambios y recomendaciones de los usuarios encontradas durante el uso del sistema (Ortiz, 2012b).

El proceso iterativo se recomienda cuando los usuarios no tienen claro lo que necesitan o requieren y los requisitos pueden ser modificados a corto plazo del desarrollo. Se requiere que el usuario tenga alta disponibilidad, acceso para la discusión de los avances y la retroalimentación. Las características de este proyecto se consideran idóneas para aplicar este enfoque de desarrollo.

Capítulo 4. Propuesta del modelo para construir un DW para la Universidad de Costa Rica

El proyecto plantea el uso de un modelo de construcción de DW que sea adecuado con las necesidades de las instituciones de educación superior públicas, en específico a la Universidad de Costa Rica, como una propuesta formal a partir de la investigación bibliográfica, de recomendaciones de profesionales y de la propia experiencia en el campo de la inteligencia de negocios.

El modelo se formula como una serie de procesos requeridos para la consecución de un DW, al considerar características específicas que poseen las instituciones como la UCR, donde la variedad de actividades en las cuales se ve involucrada, como la docencia, la investigación, el vínculo hacia la sociedad, el manejo de recurso humano, la atención de la salud y actividades culturales son parte de la diversidad presentada. Aunado a esto, se da un manejo económico sin fines de lucro, con financiamiento estatal y bajo la supervisión de entes gubernamentales.

La propuesta abarca desde el inicio de la construcción del DW, en sus etapas de factibilidad del proyecto con el patrocinador, las prioridades y su alineación con los objetivos de la institución. Continúa con la propuesta de arquitectura más adecuada, pasando a la etapa de diseño para establecer las metas, análisis requeridos, el modelo lógico, dimensional y físico del DW. Se menciona la importancia del manejo de los metadatos y cuáles de estos son obligatorios para el buen manejo de la información.

De forma paralela, se recomiendan algunas técnicas para el modelado de los datos en el ámbito de las IES. Las técnicas mencionadas no son exclusivas para las IES, pero a partir de la investigación realizada, se determinó su uso recurrente en este contexto y, por lo tanto, se optó por incluirlas como parte del modelo para facilitar su comprensión y aplicación en la construcción del DW.

Para finalizar el modelo, se categorizan las diferentes variables, se divide el público meta para brindar una experiencia pertinente con las necesidades de cada usuario, sin olvidar la importancia de mantener la privacidad y la confidencialidad de datos sensibles para la institución.

El diagrama de procesos muestra la lista de los procesos con los insumos requeridos de entrada, los actores involucrados y las salidas esperadas para la aplicación del modelo.

4.1 Uso del modelo

El modelo consta de un diagrama SIPOC que define el orden de ejecución de los procesos requeridos para llevar a cabo la construcción de un DW. El diagrama se divide en 5 fases, en donde cada fase se considera una entidad con elementos o procesos comunes. Las fases son:

1. Proyecto (P)
2. Fuente de datos (F)
3. Diseño (D)
4. Visualización (V)
5. Implementación (I)

Cada fase se identifica con una letra dentro del diagrama y muestra la cantidad de procesos asociados. Al ser una propuesta de desarrollo iterativa, requiere la ejecución de las fases más de una vez durante la vida del proyecto.

Para la primera iteración, se espera que se ejecuten todos los procesos de manera consecutiva, al iniciar con la fase de Proyecto. La fase P se refiere a la etapa de viabilidad del proyecto, de la arquitectura y de la matriz de procesos. Esta etapa no requiere formar parte de las iteraciones sucesivas, a menos que la matriz de procesos se modifique.

Las fases F, D, V, e I, deben ser ejecutadas de manera consecutiva para cada nueva fuente o mejoras en una fuente existente, pero cada fase puede ser revisada por separado. En otras palabras una misma fase puede ser repetida inmediatamente después de finalizada, o esperar a que termine el flujo en la etapa de implementación para volver a iniciar el ciclo. La figura muestra cómo se deben aplicar las fases para cumplir con el modelo.

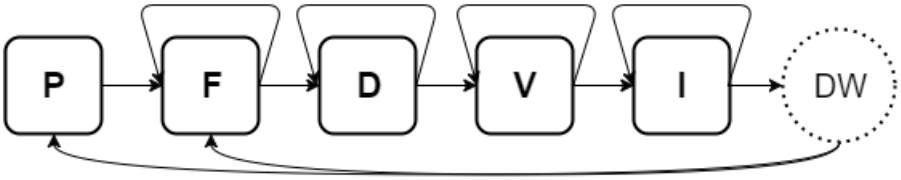


Figura 5. Secuencia de ejecución de las fases del modelo.
Fuente: Elaboración propia.

Tal como se aprecia en la Figura 5, se permite regresar al grupo F o al P una vez finalizada la implementación de la fuente en el DW, pero cada grupo puede ser iterado consigo mismo si es necesario.

El diagrama de procesos contiene por cada línea, la descripción resumida del proceso, las entradas o insumos esperados, los actores involucrados, la salida esperada y el tipo de flujo de información.

El detalle de cada proceso se especifica en el modelo, en donde se indican las entradas, salidas del proceso y una explicación ampliada de lo que se espera cumplir dentro de cada proceso de acuerdo a las tareas por ejecutar. Para algunos procesos se incluyen ejemplos de plantillas o resultados para especificar los requerimientos al nivel deseado, siempre en el contexto de las IES públicas.

A continuación, la figura 6 muestra el diagrama de procesos, el detalle del modelo para cada grupo de procesos con las entradas y salidas esperadas.

Código	Descripción	Insumos	Fuente	Equipo de BI			Cliente	Salida
			Unidad	Encargado de BI	Desarrollador de BI	Patrocinador	Usuario solicitante	
P-01	Definición de la factibilidad del proyecto	Solicitud de generación del proyecto		●		●	●	Documento del acuerdo para la elaboración del proyecto
P-02	Definición de la arquitectura del proyecto	Documento del acuerdo para la elaboración del proyecto		●	●			Documento de diseño de la arquitectura
P-03	Crear la matriz de procesos para tener una visión macro de la institución	Documento del acuerdo para la elaboración del proyecto, diseño de la arquitectura, políticas y objetivos institucionales, reunión con el patrocinador y el solicitante		●		●	●	Matriz de procesos
P-04	Definición de las fuentes prioritarias por atender	Matriz de procesos, documento con la definición de la arquitectura, reunión del patrocinador y el encargado de BI		●		●	●	Listado de fuentes prioritarias
F-01	Levantamiento de requerimientos a partir de las políticas y objetivos institucionales y necesidades del usuario solicitante	Listado de fuentes, políticas institucionales, reunión con el usuario solicitante, matriz de procesos		●			●	Lista de metas y análisis requeridos
F-02	Revisión de la información disponible para definir la viabilidad de las metas establecidas	Lista de metas y análisis requeridos	●	●			●	Lista de metas y análisis viables, Lista de variables disponibles
F-03	Establecer controles para la calidad de la información	Lista de metas y análisis viables, lista de variables disponibles	●	●	●			Metodología de extracción de datos, documento formal para establecer protocolo, lista de datos de control, procedimiento de aseguramiento de los datos
F-04	Depuración de la información disponible	Lista de metas y análisis viables, variables disponibles, metodología de extracción de datos, documento formal para establecer protocolo, lista de datos de control, procedimiento de aseguramiento de los datos			●			Datos asegurados provenientes de la fuente
D-01	Crear el modelo lógico con las necesidades del negocio y los procesos cruciales	Lista de fuentes, lista de metas y análisis viables, lista de variables disponibles, matriz de procesos	●	●	●	●	●	Modelo lógico, metadatos de las fuentes, aprobación del modelo lógico por parte del patrocinador y del usuario participante
D-02	Crear el modelo dimensional seleccionando el tipo de esquema a utilizar y proponer cómo interactúan las variables de las dimensiones con la tabla de hechos	Modelo lógico, Lista de fuentes, Lista de metas y análisis viables, variables disponibles, matriz de procesos		●	●			Modelo dimensional
D-03	Crear el modelo físico con las estructuras de tablas, nomenclatura, indexado, particionamiento y optimización	Modelo lógico, Modelo dimensional, lista de metas y análisis viables, variables disponibles, datos asegurados de la fuente			●			Modelo físico
D-04	Manejo de metadatos con procesos de desarrollo, métodos de recolección, información sobre la lógica del modelo dimensional, y sobre las estructuras de tablas y relaciones	Modelo lógico, modelo físico, lista de procesos de desarrollo, información sobre las estructuras de la base de datos		●	●			Metadatos recopilados en documentos y en la base de datos
V-01	Establecer los diferentes públicos meta de las variables y análisis, y en que formato y nivel de agregación se van a presentar los datos	Lista de metas y análisis viables, variables disponibles, modelo lógico, modelo físico		●	●		●	Lista de metas, análisis y complejidad de visualización por tipo de usuario
V-02	Validación de los conceptos de visualización propuestos por parte del usuario solicitante	Lista de metas y análisis viables, Lista de metas, análisis y complejidad de visualización por tipo de usuario	●	●		●	●	Valoración de la completitud y correctitud del acceso de datos para cada tipo de usuario y nivel de agregación permitido
I-1	Creación de la estructura para el DW	Modelo físico			●			Estructura del DW para la fuente
I-2	Carga de los datos de la fuente al DW	Estructura del DW para la fuente, Metodología de extracción de datos, documento formal para establecer protocolo, lista de datos de control, procedimiento de aseguramiento de los datos			●			Datos cargados en el DW
I-3	Validar las metas y análisis del usuario contra datos reales obtenidos del DW	Lista de metas, análisis y complejidad de visualización por tipo de usuario, datos cargados al DW, lista de datos de control	●	●	●	●	●	Resultados de consistencia de los datos

Figura 6. Diagrama de procesos para la construcción del DW. Fuente: Elaboración propia.

4.2 Fase de Proyecto

Los procesos relacionados con el proyecto son ejecutados al momento de solicitar la generación del DW por parte de las autoridades y constituyen la etapa inicial del modelo. Los procesos en esta etapa se consideran de ejecución única sin importar la cantidad de fuentes involucradas, ya que incluyen elementos generales al resto del proyecto y engloban los requisitos institucionales del DW.

4.2.1 Factibilidad del proyecto (P-01)

Entrada	Solicitud de la alta jerarquía para generar un DW para la institución.
Salida	Documento en donde se ratifique la necesidad del DW y se acuerde disponer de recursos e infraestructura y la disponibilidad de los datos por las diferentes fuentes. Este documento puede estar supeditado a políticas institucionales.

El inicio del proyecto se dicta a partir de la necesidad de la institución y del compromiso de la oficina designada para realizarlo. Se debe contar con un patrocinador con poder de decisión para brindar recurso humano y técnico para desarrollar el DW.

Además, se debe asegurar el acceso a los datos en las diferentes instancias. Esto se logra por acuerdos entre las oficinas o por mandato superior. En todos los casos es necesario generar un documento de compromiso que asegure la continuidad para otorgar la información, en donde se indiquen las variables por solicitar, el formato y medios de transmisión de los datos y la periodicidad en la cual se requiere la información. Por parte de la unidad encargada del DW, es necesario incluir el fin que se le da a la información y cómo se asegura la confidencialidad de los datos una vez salen de la fuente.

El documento de compromiso asegura la continuidad del proceso. Al ser un documento firmado por las autoridades de la institución, las jefaturas están en la obligación de facilitar los datos. En contraparte, la Oficina fuente lo utiliza como garantía de que sus datos serán manipulados de acuerdo con los controles registrados en el documento.

4.2.2 Arquitectura del Proyecto (P-02)

Entrada	Documento del acuerdo para la elaboración del proyecto.
Salida	Diseño de la arquitectura.

La arquitectura del proyecto de la Figura 7 se establece en 3 capas, la capa de datos, la de aplicación y la de presentación. La capa de datos es donde residen las bases de datos fuentes y se les aplican los procesos de ETL. Además, en esta capa reside el SA para un intercambio rápido entre la fuente de datos y este espacio de almacenamiento del DW. La capa de aplicación incluye los DM, el DW, las bases multidimensionales, los cubos OLAP y la seguridad de acceso a nivel granular y de dimensiones. La capa de presentación incluye las aplicaciones de acceso a los cubos, la generación de reportes, los cuadros de mando y la administración de los datos desde un alto nivel.

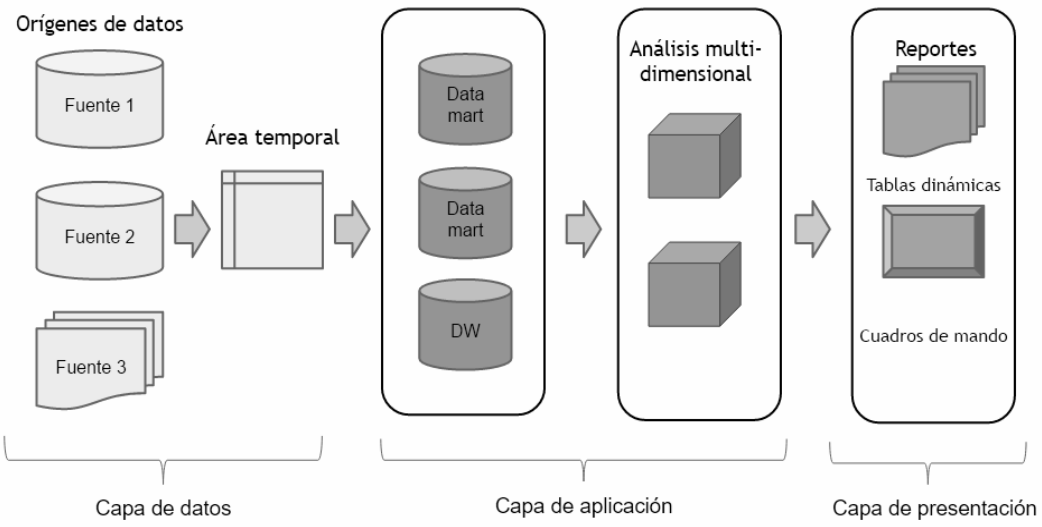


Figura 7. Arquitectura de un DW en 3 capas.

Fuente: Elaboración propia.

En este apartado también se establece el tipo de implementación aplicada en el DW. La implementación será de abajo hacia arriba (*bottom-up*) Figura 8, lo cual implica la creación de DM individuales para luego consolidarlos por medio de las dimensiones comunes. Este tipo de implementación es la adecuada en un entorno con jefaturas desconcentradas como es el caso de la UCR y en donde el manejo de la información se realiza de manera casi independiente de acuerdo a las necesidades de cada unidad de

trabajo. Asimismo, se obtienen resultados visibles de forma expedita, lo que promueve la inversión y el interés en el proyecto.

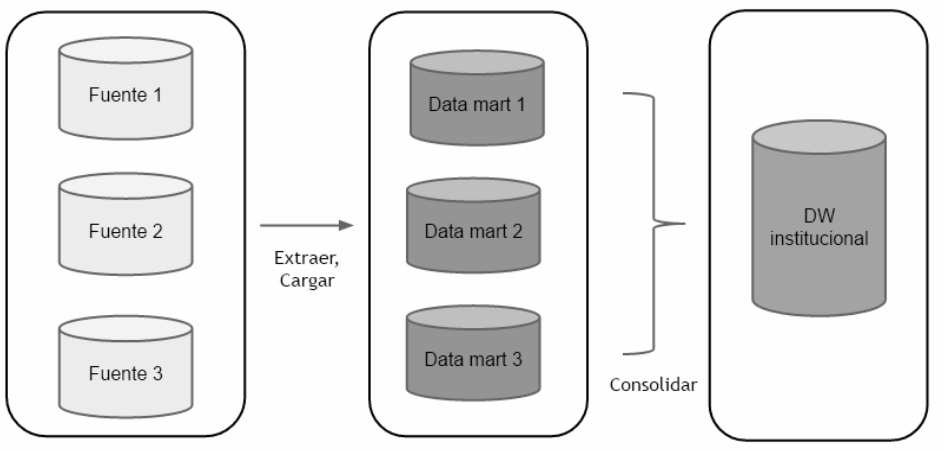


Figura 8. Diagrama de implementación de abajo hacia arriba.
Fuente: Elaboración propia basado en Kimball (2011).

4.2.3 Matriz de Procesos (P-03)

Entrada	Documento del acuerdo para la elaboración del proyecto. Diseño de la arquitectura. Políticas y objetivos institucionales. Reunión con el patrocinador y el solicitante.
Salida	Matriz de procesos.

La matriz es un requisito indispensable dentro de la arquitectura *bottom-up* y es la que brinda el panorama general de las necesidades a partir de los procesos ejecutados dentro de la Universidad. En esta etapa tan prematura del proyecto, es un esfuerzo sin réditos a corto plazo, pero permite visualizar algunos puntos como:

- La visión general que tiene el patrocinador, el solicitante y las autoridades acerca de los procesos esenciales de la institución.
- La fuente primaria de los datos para cada proceso.
- Las posibles variables y medidas que se pueden extraer de cada proceso.
- La confluencia de información entre fuentes heterogéneas las cuales actualmente no se relacionan de manera explícita o sistemática.

Para elaborar la matriz, se definen en cada fila los procesos esenciales de la institución. Para el caso de la Universidad, se recomienda agrupar los procesos por programa de acuerdo a las actividades esenciales: Docencia, Investigación, Extensión, Vida Estudiantil y Administración y el quehacer de las Sedes Regionales. Esta recomendación es para comodidad de lectura y facilitar el entendimiento por parte de las autoridades al revisar la matriz a partir de un formato ya conocido. Otra ventaja de esta agrupación es que todas las universidades estatales del país utilizan esta división, en caso de ser necesario una comunicación entre instituciones en donde es necesario encontrar un patrón en común a pesar de su independencia organizacional.

En las columnas se colocan las dimensiones (entidades, variables o características que permiten filtrar o agrupar los datos), procurando indicar los puntos en donde se cruzan los procesos con estas entidades. La unificación de las dimensiones se considera más adelante durante el modelado, para permitir la comunicación entre diferentes fuentes de datos.

Dimensiones / Procesos	Fecha	Estudiante	Unidad	Carrera	Presupuesto
Matrícula estudiantil	●	●	●	●	
Ejecución presupuestaria	●		●		●
Graduaciones	●	●	●	●	

Tabla 3. Ejemplo de matriz de procesos.

Fuente: Elaboración propia.

Se ejemplifica en la matriz de la Tabla 3 como diferentes procesos se comunican utilizando las dimensiones comunes para dar una nueva visión estratégica a los datos, tanto en este caso con las dimensiones de fecha y de unidad que corresponden a los tres procesos, las de estudiante y carrera que relacionan matrícula con graduaciones.

4.2.4 Fuentes de Datos (P-04)

Entrada	Matriz de procesos. Diseño de la arquitectura. Reunión del patrocinador y el encargado de BI.
Salida	Listado de fuentes según prioridad.

Una vez establecida la matriz de procesos, se requiere establecer prioridades en el avance del proyecto. Esta necesidad surge al utilizar la arquitectura de abajo hacia arriba en la construcción del DW. Por esta razón, el solicitante del DW, el patrocinador y el encargado de BI deben reunirse y establecer el orden para gestionar, extraer, verificar y disponer de los datos para cada fuente de los datos y construir de manera iterativa los DM, de tal manera se obtenga un DW vinculado a nivel institucional.

El resultado de esta reunión debe ser una lista con las fuentes de los datos, en orden de prioridad, al tomar en cuenta factores como:

- Mayor cantidad de análisis estratégicos de una misma fuente.
- Importancia a nivel institucional de la fuente.
- Madurez de la fuente con respecto a la sistematización de sus procesos internos.
- Nivel de confianza con las jefaturas para facilitar la comunicación ante nuevos requerimientos.

Estos factores son útiles al momento de seleccionar la prioridad de las fuentes, en especial las primeras fuentes, ante la incertidumbre en el uso de una nueva manera de tratar sus datos y la creencia en la pérdida de gobernabilidad de los mismos. Para estos casos la apertura y la confianza de las jefaturas de estas unidades de trabajo es trascendental para el éxito del proyecto.

4.3 Fuente de datos

Los procesos relacionados con la fuente de datos son iterativos, y pueden discurrir de manera paralela con otras fases para distintas fuentes de datos. La ejecución simultánea de procesos para distintas fuentes es posible de acuerdo a los recursos asignados al proyecto.

4.3.1 Levantar requerimientos (F-01)

Entrada	Listado de fuentes. Políticas y objetivos institucionales. Reunión con el usuario solicitante. Matriz de procesos.
Salida	Lista de metas y análisis requeridos.

El levantamiento de requerimientos se aplica desde una perspectiva de desarrollo iterativo, similar a la ejecución aplicada en una metodología ágil. Este levantamiento no se realiza con un documento de diseño, sino como una historia por parte del usuario. En la historia se documentan los requerimientos y cómo es su relación con el todo el sistema.

Para este proyecto, la recolección de requerimientos se realiza aplicando las siguientes técnicas:

- Reuniones de intercambio de ideas.
- Investigación documental.
- Entrevistas (contextualización y semiestructurada).

Las técnicas serán aplicadas durante las diferentes iteraciones del desarrollo de la propuesta -no necesariamente todas en cada ciclo-, como un método de afinamiento y para corroborar el cumplimiento de las expectativas del usuario en torno al modelo.

La interacción con el usuario para establecer las metas y análisis requeridos, es el inicio del vínculo para etapas posteriores de verificación e iteraciones del sistema.

El vínculo con la fuente permitirá una comunicación fluida para el resto de los procesos y las iteraciones requeridas. Esta comunicación con la fuente se establece al mostrar a la unidad los beneficios del proyecto en la disminución de solicitudes directas de información, en tiempo de procesamiento de datos, pero sin perder la gobernabilidad de los datos. El éxito del proyecto está supeditado con el nivel de apertura y apoyo que se reciba de la unidad fuente de los datos.

4.3.2 Revisión de la información disponible en la fuente (F-02)

Entrada	Lista de metas y análisis requeridos.
Salida	Lista de metas y análisis viables. Lista de variables disponibles.

El análisis requerido en este proceso consiste en corroborar que las metas solicitadas por el usuario son factibles a partir de la información disponible en la fuente. Para este momento del proyecto, ya debe existir algún tipo de vínculo con la unidad a partir del contacto con la fuente durante el levantamiento de requerimientos.

La revisión de los datos revela la existencia de la información necesaria para generar los análisis solicitados. En caso de no contar con las variables requeridas, se debe notificar al usuario la imposibilidad de mostrar este tipo de análisis. No concretar un análisis en esta etapa es positivo, al no crear expectativas equivocadas en el usuario. Inclusive, podría generar la inquietud por iniciar la recolección de la información, para en una siguiente iteración, adicionar el análisis al DW ya existente.

Luego de la revisión de los datos existentes en la fuente, se genera un listado de metas posibles y un listado con las variables disponibles en la fuente para posteriores referencias.

4.3.3 Establecer controles para la calidad de la información (F-03)

Entrada	Lista de metas y análisis viables. Lista de variables disponibles.
Salida	Metodología de extracción de datos. Documento formal para establecer protocolo. Lista de datos de control. Procedimiento de aseguramiento de los datos.

Los controles se establecen en conjunto con la unidad garante de los datos. Los temas por los cuales se debe llegar a acuerdos incluyen:

- Método de extracción de datos desde la fuente.
- Periodicidad de extracción de los datos.
- Responsable en la fuente.
- Datos de control.
- Compromiso de disponibilidad de los datos.
- Consistencia de los datos históricos.

La metodología de extracción es definida con la unidad, debe ser lo más automatizada y sistematizada posible de acuerdo con las posibilidades actuales de la fuente. La periodicidad será acorde con los requerimientos de los análisis solicitados del usuario.

Los datos de control abarcan todas las variables, medidas disponibles de la fuente y algunas combinaciones habituales a la unidad. El formato de los datos de control está supeditado a la disponibilidad de la fuente, pero al menos se requieren datos a nivel macro y en caso de inconsistencias, información detallada para ubicar los datos irregulares. El detalle puede no estar disponible de inmediato, pero sí el acuerdo para ejecutar una posible solicitud de datos adicionales de control.

La fuente, al ser la responsable de la información, indica el momento histórico a partir del cual los datos son consistentes y pueden ser asegurados. Luego de fijar el rango de validez de los datos, los encargados del proyecto establecen el periodo histórico adecuado para responder a las metas solicitadas.

Los acuerdos generados deben ser documentados de manera formal en un protocolo de acceso, presentados a cada jefatura y al patrocinador para su aprobación. De esta forma, se establece el compromiso de la fuente por brindar los datos y mantenerlos disponibles según los requerimientos del proyecto.

El procedimiento para el aseguramiento de los datos se recomienda, sea ejecutado por alguien con conocimiento en ingeniería de procesos de calidad y con bases sobre el manejo de datos dimensionales.

Con el fin de satisfacer al cliente, prevenir errores y reducir costos se solicita que el procedimiento de aseguramiento contenga:

- Control de calidad.
- Evaluación de la calidad.
- Ajuste de la calidad.

Este modelo no pretende explicar con detalle los pasos para realizar un aseguramiento de la calidad, por lo tanto, la institución requiere una inversión en capacitación o en recurso para suplir esta necesidad.

4.3.4 Depuración de la información disponible (F-04)

Entrada	Lista de metas y análisis viables. Variables disponibles. Metodología de extracción de datos. Documento formal para establecer protocolo. Lista de datos de control. Procedimiento de aseguramiento de los datos.
Salida	Datos asegurados provenientes de la fuente.

Al establecer protocolos para la recolección de los datos, los datos de control y un procedimiento de aseguramiento, se pueden ejecutar acciones sobre los datos extraídos para asegurar su consistencia respecto a la información brindada por la fuente de los datos. La revisión de la consistencia se ejecuta en el paso inmediato luego de la extracción por lotes de los datos de la fuente.

Los datos, una vez disponibles, deben ser revisados a profundidad al buscar:

- Inconsistencias contra los datos de control.
- Variables faltantes para realizar cálculos, filtrados y agrupaciones.
- Descripciones faltantes en las variables.
- Variables desconocidas o fuera de contexto.

Se deben aplicar los protocolos ya establecidos de revisión, someter a consideración las inconsistencias encontradas durante la ejecución de este proceso y que deban ser incluidas en los protocolos existentes.

Los controles de calidad y la depuración de los datos están ligados a la fuente y no al DW, en consecuencia, la revisión de los datos se ejecuta en el SA. Por esto, esta etapa y la 4.3.3 (F-03) de establecimiento de controles de calidad, se pueden ejecutar en paralelo a las de diseño del modelo, pues no interfieren en el resultado sino hasta la carga de datos al DW.

4.4 Diseño

La etapa de diseño transforma los requerimientos de alto nivel en un modelo físico de implementación y define la metadata obligatoria para el DW. La transformación se lleva a cabo en 3 etapas, modelo lógico, dimensional, físico y así convertir los datos en un formato adecuado para consultas analíticas. Los metadatos se definen durante el diseño y se actualizan en cada etapa del modelado y construcción del DW.

4.4.1 Modelo lógico (D-01)

Entrada	Lista de fuentes. Lista de metas y análisis viables. Lista de variables disponibles. Matriz de procesos.
Salida	Modelo lógico. Metadatos de las fuentes. Aprobación del modelo lógico por parte del patrocinador y del usuario participante.

El modelo lógico identifica y prioriza los procesos cruciales del negocio de manera descriptiva y cumple la función de marco de referencia una vez iniciado el desarrollo. La identificación de los procesos se realiza con la matriz de procesos y las metas, para luego detallar el modelo con las medidas, dimensiones, granularidad, definiciones, reglas y

metadatos. El diagrama del modelo lógico de la Figura 9 consiste en insumos de entrada, procesos, resultados a corto, mediano y largo plazo.

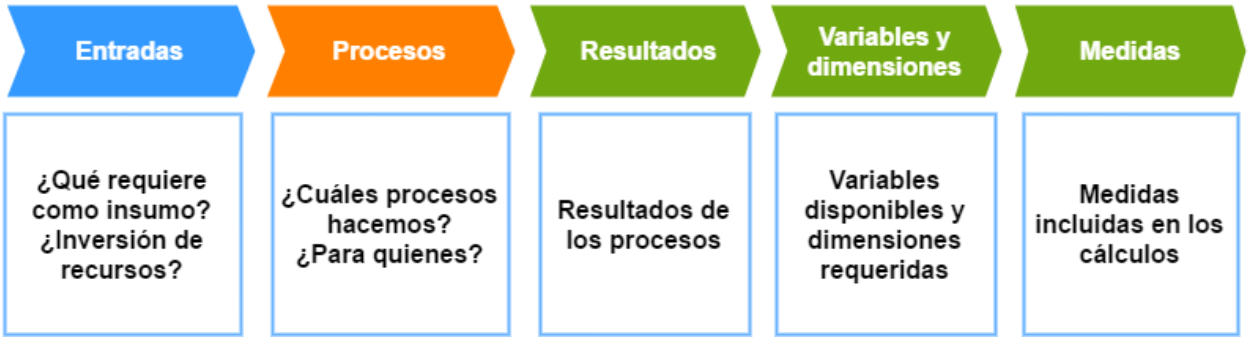


Figura 9. Diagrama del modelo lógico. Fuente: Elaboración propia basada en University of Wisconsin-Extension Program Development and Evaluation[6].

El diagrama se puede llenar en el siguiente orden:

1. La lista de fuentes requeridas.
2. Los procesos a partir de la matriz de procesos.
3. Los resultados buscados de la lista de metas y análisis viables.
4. La lista de variables disponibles y las dimensiones requeridas.

El llenado del modelo lógico se realiza en este orden para facilitar la ubicación de datos específicos luego de establecer las metas y análisis para cada fuente.

4.4.2 Modelo dimensional (D-02)

Entrada	Modelo lógico. Lista de fuentes. Lista de metas y análisis viables. Variables disponibles. Matriz de procesos.
Salida	Modelo dimensional.

[6] http://fyi.uwex.edu/programdevelopment/logic-models/bibliography

El modelo dimensional convierte el conjunto de medidas en tablas de hechos y establece su relación con las dimensiones. La aplicación de esta fase en el diseño, permite conceptualizar la cantidad de tablas de hecho necesarias dentro del DM y las dimensiones compartidas entre las tablas hechos o con otros DM.

Las dimensiones incluyen las variables asociadas y los identificadores únicos para cada entidad. La tabla de hechos incorpora las medidas y las relaciones con las dimensiones. El modelo resultante contiene tres tipos de objetos: 1) las dimensiones, 2) las tablas de hechos y 3) los conectores, como se aprecia en el ejemplo de la Figura 10 del modelo dimensional.

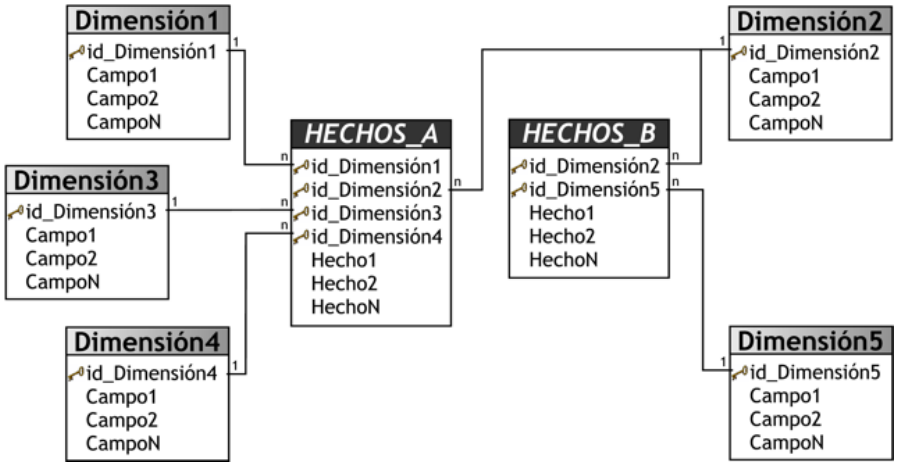


Figura 10. Ejemplo del modelo dimensional.
Fuente: DataPrix Datawarehouse manager[7].

El modelo dimensional simplifica el modelo de tal manera resulta atinado en esta etapa la aplicación de técnicas específicas para tratar los hechos; estas técnicas se presentan con la finalidad de facilitar la construcción y posterior implementación de la solución, cumpliendo con los objetivos trazados. Las técnicas no son exclusivas ni las únicas aplicables en un DW enfocado a las IES, pero se ha constatado que suelen estar asociadas en este tipo de aplicación.

Las técnicas sugeridas son:

1. Tablas de hechos sin medidas (*factless*).

[7] http://www.dataprix.com/datawarehouse-manager

2. Datos agrupados.
3. Dimensiones compartidas (*dimensional drill-across o conformed*).

Las tablas *factless* son en esencia tablas de hecho sin medidas explícitas. Lo que corresponde es el conteo a partir de la información relacionada por medio de las dimensiones. La técnica se utiliza en tablas relacionadas con la docencia, como total de cursos, grupos, estudiantes y graduaciones. De igual manera puede ligarse en muchos ámbitos de las IES, como el área de investigación, con la cantidad de proyectos en ejecución o de publicaciones científicas. En general es una técnica que demuestra que las tablas de hechos contienen mucha información, aparte de las medidas.

Los datos agregados se utilizan para el cálculo de total de cursos, rendimiento académico o total de materias perdidas contra cursos aprobados, o nota ponderada. La información de un estudiante no se requiere a nivel de curso, sino en forma general para conocer el panorama. Los datos consolidados facilitan la consulta de la información y aumentan la velocidad de acceso.

Las dimensiones compartidas son dimensiones en diferentes esquemas los cuales comparten características. Las tres características que deben cumplir son: 1) comparten los mismos atributos, 2) la granularidad es afín y 3) su semántica es equivalente. Las dimensiones compartidas cumplen 2 funciones: la existencia de una sola verdad para los atributos de un DW y permitir relacionar distintos procesos del negocio.

Las técnicas mencionadas son recomendaciones, no aplican en todos los contextos, pero se deben repasar para cada tabla de hechos y estudiar si es adecuado a su aplicación.

4.4.3 Modelo físico (D-03)

Entrada	Modelo lógico. Modelo dimensional. Lista de metas y análisis viables. Lista de variables disponibles. Datos asegurados de la fuente.

Salida	Modelo físico.

El modelo físico es la tercera fase del diseño, donde se crean las estructuras de las tablas y el dimensionamiento de la base de datos. Además, se define la estrategia de implementación para tablas, tipos de datos para las variables, jerarquías, nomenclatura estándar para los objetos, propone las estrategias de indexado, particionamiento y almacenamiento.

El modelo físico contiene siete rubros:

1. Nomenclatura de los objetos.
2. Estructuras de tablas.
3. Tipos de datos.
4. Índices.
5. Particionamiento.
6. Almacenamiento.
7. Dimensionamiento de la base datos.

Lo primero es definir una nomenclatura estándar. Existen muchas metodologías para este punto como el uso de prefijos-sufijos, abreviaciones, separar por guion bajo o *camel-case*. Sumado a cualquiera de estas metodologías, se sugieren cuatro acciones para facilitar el uso del DW:

- Utilizar nombres descriptivos.
- Incluir un identificador en las tablas para cada fuente de datos en el SA.
- Evitar identificar dimensiones y hechos con alguna fuente en el DW.
- Identificar claramente las dimensiones de las tablas de hechos en el DW.

La estructura de las tablas se establece con base en el modelo dimensional, se agregan las consideraciones técnicas del motor de base de datos y se respeta la nomenclatura definida. Esto incluye agregar el tipo de datos para cada variable de las tablas e incluir las llaves sustitutas para las dimensiones.

En los rubros 4, 5 y 6 sobre índices, particionamiento y almacenamiento se requiere la participación activa del administrador de base de datos de la institución, para definir la mejor estrategia de acuerdo al motor de base de datos utilizado.

El dimensionamiento de la base de datos considera el espacio requerido tanto en el SA como en el DW. El espacio utilizado en el SA se obtiene al sumar los tipos de variables en la fuente por la periodicidad con la cual se van a recolectar los datos. El espacio para el DW incluye los tipos de la estructura de tablas sumados al espacio de los índices, el tipo de particionamiento, las tablas agregadas. La suma de los resultados del SA y el DW es el espacio total requerido. A este resultado se agrega una estimación de crecimiento a un año y a cinco años. La estimación se obtiene de la cantidad de espacio requerido por la cantidad de veces por año con la cual se recolectan los datos.

4.4.4 Metadatos (D-04)

Entrada	Modelo lógico. Modelo dimensional. Modelo físico. Diseño de la arquitectura. Matriz de procesos. Lista de metas y análisis viables. Protocolos de extracción de datos.
Salida	Lista de procesos de desarrollo. Lista de objetos de la base de datos.

La metadata es requerida como parte del modelo, permite que los datos sean comprendidos, utilizados de manera amplia y correcta. Los metadatos son divididos en dos grupos, lo referente a los objetos y la documentación de los procesos.

Listado de objetos disponibles

La información de los objetos debe estar disponible de manera permanente para consulta por los usuarios del DW. A continuación, se mencionan las listas de objetos con el detalle esperado.

- **Esquemas**: Los esquemas deben tener un nombre, una descripción de su contenido, su función y suelen tener el mismo nombre que la tabla de hechos.
- **Dimensiones, hechos:** nombre, una definición y las acepciones con los cuales se conoce en la organización un mismo objeto.
- **Medidas**: nombre, definición, acepciones con los cuales se conoce, frecuencia de actualización de los datos, granularidad, las dimensiones asociadas con especial énfasis en la dimensión de tiempo y cómo se relaciona.
- **Atributos**: nombre, el tipo de dato, dominio y reglas para determinar los valores válidos.

La extracción de los metadatos de los objetos en la base de datos se realiza por medios automatizados, ya sea por medio de una aplicación especializada o un *script* que extraiga los datos. La extracción de los metadatos de los objetos se recomienda automatizarla al ser información cambiante en el tiempo y su mantenibilidad suele ser complicada.

Listado de procesos de desarrollo

La documentación relacionada con el desarrollo del DW es obtenida a lo largo de los procesos aplicados. Los documentos esperados luego de aplicar los procesos son el diseño de la arquitectura, la matriz de procesos, el listado de metas, análisis viables para cada fuente, metodología de extracción de datos, modelo lógico y el modelo dimensional.

4.5 Visualización

La etapa de visualización se revisa desde dos aristas, la de usabilidad y la de seguridad. El fin es cumplir con las expectativas de los usuarios brindando la información de la manera más directa y sencilla posible. Para cumplir este objetivo se establece el público meta de los análisis, el formato de presentación, el nivel de desagregación mínimo posible y la complejidad de acceso a los datos acorde al tipo de usuario. La segregación por roles establece parámetros de seguridad de la información y delimita los alcances del usuario con el almacén de datos. La delimitación no es solo por seguridad, sino también para evitarle al usuario el uso de herramientas que simplemente no son relevantes en su contexto; esto disminuye la posibilidad de fallos y facilita interpretar la información. La división de las funciones crea un balance entre experiencia y satisfacción.

4.5.1 Público meta (V-01)

Entrada	Lista de metas y análisis viables. Lista de variables disponibles. Modelo lógico. Modelo físico.
Salida	Lista de metas, análisis y complejidad de visualización por tipo de usuario.

La presentación de los datos debe separarse de acuerdo con el público meta de los análisis por seguridad y por facilidad para el usuario. La separación toma en cuenta factores como la granularidad permitida, la confidencialidad de los datos, la complejidad del análisis y el cumplimiento de objetivos específicos para cada sector. Las IES públicas y en el caso particular de la Universidad Costa Rica, el público meta se representa como una jerarquía como la mostrada en la Figura 11.

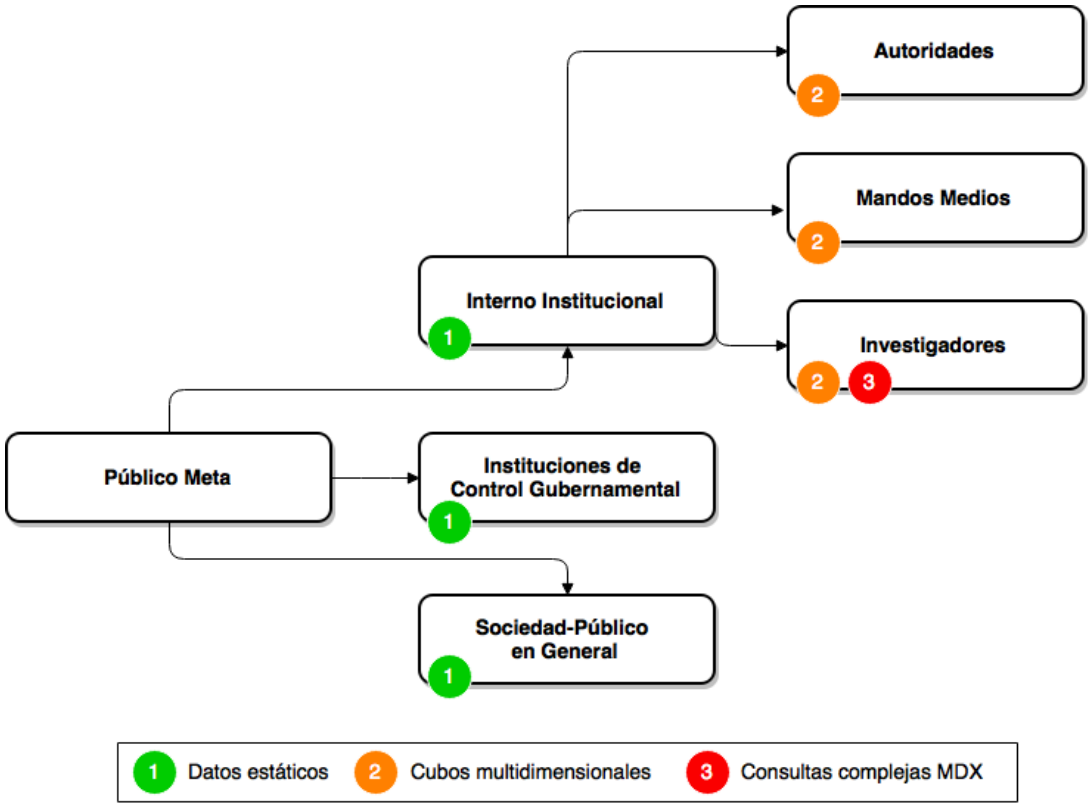


Figura 11. Jerarquía de acceso del público meta por complejidad de visualización.

Fuente: Elaboración propia.

Las diferencias de los usuarios a nivel interno son por el nivel de granularidad permitido. Para las autoridades, el acceso es completo. Los mandos medios pueden revisar el detalle, pero solo del área a la cual pertenece la jefatura, por ejemplo, administración puede analizar datos de recursos humanos o financieros, pero se impide el acceso a los datos del resto de instancias. La jerarquía de acceso se aplica de acuerdo con la estructura organizativa institucional. Los datos para los investigadores se ofrecen a nivel de área funcional, como escuelas u oficinas que posean datos públicos, pero no podrán ver el detalle que pueda ser usado para inferir la situación de un individuo o divulgar información sensible de la institución, como proyecciones o escenarios financieros.

Una vez establecida la jerarquía de acceso, se indican en la tabla, los accesos de cada público meta a las metas y análisis viables. La tabla incluye: información de la meta, el público meta, las dimensiones con permiso de acceso según el público meta y el nivel de granularidad permitido. La Tabla 4 es un ejemplo de cómo documentar el acceso de los datos para los usuarios.

Meta o análisis	Público permitido	Dimensiones con permiso de acceso	Nivel de granularidad máximo permitido
Proyectos de investigación en ejecución.	Todo público.	Autoridades y mandos medios: -Investigadores	Autoridades y mandos medios: Por investigación
		Resto del público: -Tipo de investigación -Fecha semestral	Resto del público: Por unidad
Consultas en salud preventiva.	Autoridades. Mandos medios.	Todas las dimensiones disponibles para la tabla de hechos en salud.	Máxima granularidad disponible.
...	...	...	...

Tabla 4. Ejemplo de permisos de acceso a dimensiones y metas por rol de usuario.

Fuente: Elaboración propia.

Otro formato para documentar los niveles de acceso por funciones, es una matriz de dimensiones y roles como la Tabla 5, en donde las variables de la dimensión se relacionan con uno o varios roles. Puede darse la condición de que no existan roles asociados con una variable. De manera similar, la Tabla 6 muestra los accesos permitidos a las metas por función de usuario.

		Granularidad permitida por roles					
Dimensiones con acceso	Variables en la dimensión	Todo público	Instituciones de gobierno	Interno general	Investigadores	Mandos medios	Autoridades
Fecha	Año	x	x	x	x	x	x
	Mes	x	x	x	x	x	x
	Día	x	x	x	x	x	x
Proyectos	Área de impacto	x	x	x	x	x	x
	Proyecto				x	x	x
	Actividad sustantiva	x	x	x	x	x	x
	Subactividad		x	x	x	x	x
Tipos de atención en salud	Procedimiento					x	x
...	...	...	...	...	...	...	...

Tabla 5. Ejemplo de matriz de acceso a las dimensiones y variables por rol de usuario.

Fuente: Elaboración propia.

	Acceso a metas por roles					
Metas y análisis	Todo público	Instituciones de gobierno	Interno general	Investigadores	Mandos medios	Autoridades
Investigadores	x	x	x	x	x	x
Proyectos	x	x	x	x	x	x
Publicaciones	x	x	x	x	x	x
Atenciones en salud	x	x	x	x	x	x
Incapacidades				x	x	x
...	...	...	...	...	...	...

Tabla 6. Ejemplo de matriz de acceso para metas por rol de usuario.

Fuente: Elaboración propia.

4.5.2 Validar conceptos de visualización (V-02)

Entrada	Lista de metas y análisis viables. Lista de metas, análisis y complejidad de visualización por tipo de usuario.
Salida	Valoración del acceso de datos para cada tipo de usuario y nivel de agregación permitido.

La validación se realiza a partir de los datos suministrados por la lista de metas con los permisos de visualización por usuario. La valoración debe ser ejecutada por todos los actores involucrados en el proyecto y dar su aprobación para continuar con la implementación del proyecto.

El resultado esperado es una lista de modificaciones a la lista de permisos del punto anterior con las correcciones en los accesos en caso de ser necesario.

4.6 Implementación

La aplicación del modelo deriva en la construcción física del DW. La implementación se ejecuta en tres pasos: 1) la creación de la estructura del DW en la base de datos usando el modelo físico, 2) la carga de los datos desde la fuente al DW, y 3) validar que el DW muestra la información real esperada para los usuarios.

4.6.1 Estructura del DW (I-01)

Entrada	Modelo físico.
Salida	Estructura del DW para la fuente. Script de creación de la estructura.

La estructura del DW se genera a partir del modelo físico utilizando la sintaxis propia del motor de base de datos seleccionado. Los tipos de datos y tablas se adaptan a las particularidades del gestor y se ejecutan los comandos necesarios para la persistencia de los objetos en la base de datos.

El resultado de este proceso es el *script* de creación de los objetos, de reglas y el esquema funcional en el motor de base de datos.

4.6.2 Carga de los datos (I-02)

Entrada	Estructura del DW para la fuente. Metodología de extracción de datos. Documento formal para establecer protocolo. Lista de datos de control. Procedimiento de aseguramiento de los datos.
Salida	Datos cargados en el DW.

La carga de datos es un proceso delicado y que consume mucho tiempo de desarrollo, donde la documentación recopilada de los procesos anteriores es esencial para mantener el norte en la ejecución. Lo primero es utilizar la metodología establecida para extraer los datos. La metodología fue avalada por la unidad y se determinó como la mejor forma de obtener los datos, por lo tanto debe ser acatada al igual que el protocolo para la periodicidad de la información.

La carga de datos se realiza con herramientas específicas de ETL. La herramienta es capaz de transformar los datos de la fuente al DW con parámetros estandarizados, asignando una periodicidad de ejecución automatizada. Los controles aplicados para asegurar la información son ejecutados antes, durante y después de trasladar la información al DW.

4.6.3 Validar datos del DW (I-03)

Entrada	Lista de metas, análisis y complejidad de visualización por tipo de usuario. Datos cargados al DW. Lista de datos de control.
Salida	Resultados de consistencia de los datos. Aprobación del patrocinador y del usuario.

El punto anterior aplica valores de control, pero en esta etapa se vuelven a revisar y verificar. Los datos de control deben ajustarse a los rangos especificados, ser corroborados en conjunto con el encargado de BI, el usuario y el patrocinador del proyecto.

4.6.4 Finalizar el DW

Si los datos son consistentes, la iteración del modelo se da por finalizada. En este punto se puede 1) iterar de nuevo la fuente para agregar nuevos datos, 2) revisar la implementación en caso de dudas, 3) agregar una nueva fuente de datos o 4) revisar la matriz de procesos en el proyecto.

4.7 Modelo en paralelo

La flexibilidad del modelo permite dos variantes para procesar las fuentes en forma paralela. La aplicación de estas variantes al modelo, dependen de la cantidad de los recursos asignados al proyecto. Los equipos pueden dividirse para dedicarse a: una sola fuente a lo largo de una iteración -Figura 12- o una sola fase para todas las fuentes -Figura 13-

Cualquiera de las dos variantes, requiere que la matriz de procesos solicitada en el apartado 4.2.3 esté definida y que se ejecute la primera iteración de la fuente F_N antes de iniciar con la fuente F_{N+1}. Esta restricción se da por los requerimientos técnicos de las dimensiones compartidas. Los atributos de las dimensiones compartidas se definen en la fase de fuentes de datos, -clasificada como F_N- y no es hasta finalizar este grupo de procesos, en donde se tendrá certeza de los atributos de las dimensiones que puedan estar relacionadas con una nueva fuente a tratar. En la Figura 12 se explica la ejecución en paralelo de los grupos para múltiples equipos de trabajo con una fuente por equipo.

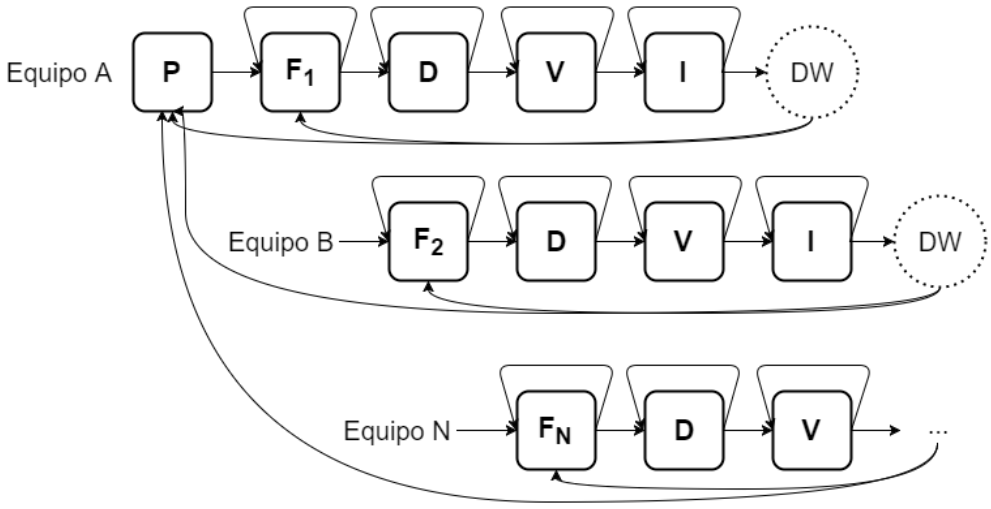


Figura 12. Secuencia de ejecución en paralelo para diferentes fuentes de datos. Fuente: Elaboración propia.

La segunda variante se refiere a la especialización de cada equipo en una fase específica. Por ejemplo, el contar con cuatro equipos de trabajo, en donde un equipo se encargue de todo lo relacionado con el análisis de la fuente, el otro con el diseño, el tercer equipo con la visualización y el último con la implementación. La Figura 13 muestra esta segunda variante de ejecución en paralelo para la especialización de los equipos por fases.

La ejecución de los procesos en paralelo es posible por 1) la modularización de los procesos en grupos comunes, 2) la existencia de entradas y salidas predefinidas para cada proceso y 3) la ejecución escalonada de los grupos para compartir las salidas de los procesos.

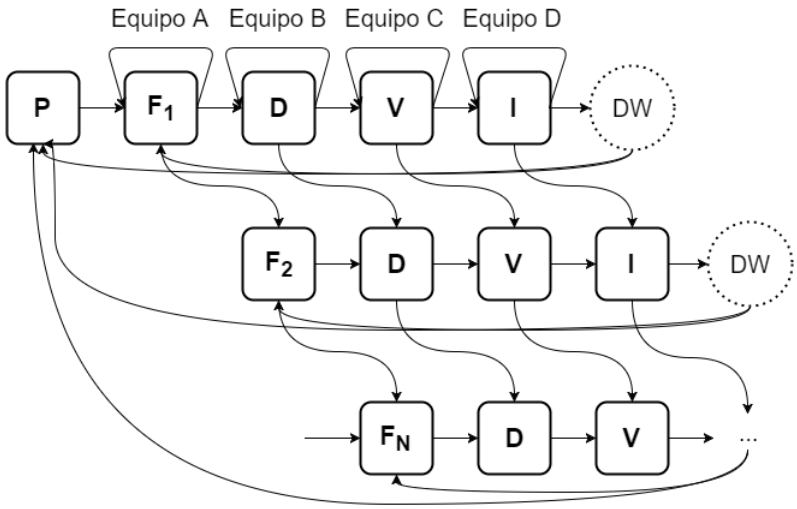


Figura 13. Secuencia de ejecución en paralelo con especialización por fases.
Fuente: Elaboración propia.

Las dos variantes que se presentan son ejemplos de cómo aprovechar el modelo más allá de la ejecución secuencial. Tal como se mencionó, la aplicación de las variantes depende enteramente de la disponibilidad de recursos para crear múltiples equipos de trabajo.

La ventaja de la primera variante es que no existe una restricción en el mínimo de equipos de trabajo, al iniciar con dos equipos hasta N equipos, en donde N es la cantidad de fuentes por agregar al DW. En contraposición, la segunda variante permite una especialización de cada equipo de trabajo, una mejor segregación del acceso a la información y promueve la uniformidad para todas las fuentes. El uso de una u otra variante, depende enteramente de las necesidades y posibilidades de la institución.

4.8 Finalizar el modelo

El uso de este modelo pretende establecer un método formal y documentado para la construcción del DW de la Universidad de Costa Rica en sus diferentes etapas, incluyendo la incorporación de diversas fuentes de información de manera paulatina de acuerdo al avance del proyecto.

Para ejemplificar la efectividad del modelo, se ejecuta una iteración secuencial considerando un solo equipo de trabajo, sobre una de las fuentes, para corroborar su aplicación real y como el modelo cumple con los requisitos solicitados por los usuarios.

Capítulo 5. Aplicación del modelo en la Elaboración del Sistema de Información Estudiantil

Las iniciativas para la construcción de un DW requieren un esfuerzo a largo plazo que para una institución pública, suelen ser procesos complicados ante la imposibilidad de disponer de los recursos de manera constante y supeditados a normativas de contratación gubernamentales. Para solventar esta necesidad, se visualizó una oportunidad de mejora al crear un modelo formal y reproducible para la construcción de un DW para una IES, enfocado a la Universidad de Costa Rica.

Este capítulo demuestra lo adecuado del modelo, al generar el DW con los datos de los estudiantes obtenidos de la Oficina de Registro. Asimismo, se mencionan los procesos clave de este modelo para la construcción del almacén. El **desarrollo completo del modelo** se puede consultar en el **apéndice 3** de este trabajo.

5.1 Ejecución del modelo

El modelo de construcción del DW se divide en 5 fases:

1. Proyecto.
2. Fuente de datos.
3. Diseño.
4. Visualización.
5. Implementación.

Cada fase consta de múltiples procesos con documentos de entrada y salida. Se realiza un chequeo de cada proceso antes de ejecutar el siguiente. Con la ayuda del apéndice 2 -el cual es un esquema de chequeo de los documentos requeridos en cada fase- se verifica que se completó el proceso en cuestión y se dispone de los documentos de salida obligatorios. En cada proceso se generan documentos de los cuales dependen procesos subsiguientes, por lo tanto cada proceso se debe finalizar antes de continuar el flujo.

Para facilitar la tarea de revisión e integración, el modelo solicita pocos documentos de entrada para cada proceso. A su vez, las salidas de documentos suelen ser pocas; en más de la mitad de procesos, se solicita un solo documento de salida. Esto también simplifica el avance entre procesos.

5.2 Tiempo de construcción

El tiempo de construcción varió en cada fase, de manera dependiente de los tiempos de aprobación y la dedicación del equipo de desarrollo. Para esta primera iteración los tiempos de duración se explican en la Tabla 7.

Fase	Tiempo de duración	Observaciones
Fase 1. Proyecto	3 semanas (para revisión y aprobación).	La arquitectura se definió en el modelo, no durante la ejecución, por lo tanto, este tiempo no se contempla. Para la matriz de procesos, ya existía un documento que incluye lo estratégico de la institución, lo que también reduce el tiempo. La arquitectura de despliegue estaba definida de previo. La duración la determinó el tiempo de revisión y aprobación.
Fase 2. Fuente de datos	1 mes (reuniones con la fuente).	Incluye la metodología y protocolos de extracción, y los datos de control con la fuente, los procedimientos de aseguramiento y depuración de la información disponible. Esta etapa estaba adelantada en un 50%, por lo que fue factible aplicarlo en este proyecto, sino, el tiempo se hubiera extendido al menos otro mes.
Fase 3. Diseño	1.5 meses.	Con el modelo lógico y dimensional aprobado, y la inclusión de metadatos en la base datos.
Fase 4. Visualización	2 semanas (para revisión y aprobación).	Las dimensiones y métricas se definen en la fase previa, por lo que esta es de revisión y aprobación.

Fase 5. Implementación	2 semanas para la estructura del DW y la inclusión de los datos. 1 semana para la validación de los datos ingresados.	El modelo físico incluye la mayor parte de la información para esta etapa, por lo que la creación de la estructura es muy rápida, y la inclusión de los datos se rige por la metodología de extracción contra los datos de control definidos para la fuente.

Tabla 7. Tiempo de ejecución del modelo en la primera iteración.

Fuente: Elaboración propia.

El tiempo total de ejecución para la primera iteración fue de 4 meses. De este tiempo se excluye:

1. El tiempo requerido para definir la arquitectura, al ser parte del modelo propuesto.
2. La comunicación con la fuente de datos ya estaba adelantada, lo que redujo el tiempo para definir metodologías, procedimientos y datos de control a la mitad.
3. la definición del esquema de funciones y permisos en la fase de visualización que también forma parte del modelo.

5.3 Costo teórico de desarrollo de la iteración

La prueba de una iteración del modelo requiere recurso humano y de infraestructura. La Tabla 8 presenta los costos teóricos del proyecto, en el caso de haber efectuado una prueba real en un entorno en donde se requiere el pago de los profesionales involucrados y de los equipos e infraestructura necesarios.

Recursos	**Costo**	**Cantidad**	**Inversión**	**Notas**
Desarrollador de BI (en meses)	$1,200	4	$4,800	Profesional (Bachillerato en campos relacionados).
Ejecutivo de BI (en meses)	$1,500	4	$6,000	Profesional (Licenciado en campos relacionados).
Apoyo de BI (en meses)	$1,500	0.5	$750	Profesional (Licenciado en campos relacionados).

Equipo				
Servidor de alto rendimiento para virtualización	$5,000	1	$5,000	De acuerdo a estándares de la institución.
Computadoras para desarrollo	$1,500	3	$4,500	De acuerdo a estándares de la institución.
Software				
Windows Server 2012 R2 Standard	$150	1	$150	Licencia académica.
MSSQL Server 2012 Standard	$400	3	$1,200	Licencia académica.
Total, de costo teórico			**$21,200**	

Tabla 8. Costos teóricos de inversión en la ejecución del proyecto.

Fuente. Elaboración propia.

La Tabla 8 no contempla el costo en horas del patrocinador, de la fuente, del cliente y solo estima el tiempo requerido para la ejecución de la primera iteración del modelo.

El costo total teórico de la ejecución es de $21.200 para los rubros de: recurso humano directo, equipo computacional y licencias de aplicaciones. El costo de la infraestructura se realiza desde una perspectiva de ambiente de desarrollo, acorde al alcance del proyecto.

5.4 Demostración de los datos en el DW

Los datos se muestran utilizando la herramienta Excel. La conexión se realiza por medio de cubos OLAP a la base de datos en SQL al utilizar usuarios con roles definidos durante la fase de visualización.

A modo de ejemplo, se muestran resultados para dos tablas de hechos, la de matrícula y la de graduación, al utilizar tres roles distintos: público en general, investigadores y

autoridades. Se realizan tres consultas de datos con la máxima granularidad disponible por función.

Consulta 1.

El total de estudiantes por escuela distribuidos por nivel académico y por sexo para los últimos cinco años. Para esta consulta la desagregación es la misma siendo un rol de investigador o autoridad universitaria Figura 14 de acuerdo con el apéndice 3, en la fase 3, proceso V-02 de validación de conceptos. La Figura 15 es con un rol público.

	2011		Total 2011	2012		Total 2012	2013		Total 2013
Rótulos de fila	Femenino	Masculino		Femenino	Masculino		Femenino	Masculino	
POSGRADO	**2831**	**2534**	**5365**	**2577**	**2263**	**4840**	**2342**	**2067**	**4409**
ESTUDIOS DE POSGRADO	**2831**	**2534**	**5365**	**2577**	**2263**	**4840**	**2342**	**2067**	**4409**
ESPECIALIDAD DE POSGRADO EN ODONTOLOGIA	2	2	4	2	2	4	1		1
ESPECIALIDADES DE POSGRADO EN MICROBIOLOGIA	52	22	74	35	13	48	23	18	41
ESTUDIOS DE POSTGRADO	5	4	9	4	2	6	2	2	4
PROGR. DE POSGRADO EN C.S. DEL MOVIMIENT	25	30	55	25	31	56	27	31	58
PROGR. DE POSGRADO EN ENS. DEL CASTELANO	24	10	34	18	4	22	22	8	30
PROGR. DE POSGRADO EN ESPAÑOL COMO SEG.	13	5	18	14	4	18	13	3	16
PROGR. DE POSGRADO EN EVALUACIÓN DE PROG	25	10	35	29	7	36	27	12	39
PROGR. DE POSGRADO EN GESTION INTEGRADA	10	6	16	5	4	9	4	8	12
PROGR. DE POSGRADO EN GOBIERNO Y POLITIC	12	14	26	12	11	23	10	9	19
PROGRAMA DE DOCTORADO EN CIENCIAS	8	9	17	9	9	18	10	9	19
PROGRAMA DE DOCTORADO EN EDUCACIÓN	19	18	37	15	12	27	22	7	29
PROGRAMA DE DOCTORADO EN ESTUDIOS DE LA	21	17	38	21	19	40	17	18	35
PROGRAMA DE DOCTORADO EN INGENIERIA		1	1		1	1		1	1
PROGRAMA DE POSGRADO EN ADM. PÚBLICA	83	53	136	83	51	134	61	48	109
PROGRAMA DE POSGRADO EN ADM. UNIVERSITAR	15	5	20	15	5	20	10	4	14
PROGRAMA DE POSGRADO EN ADM. Y DIRECCIÓN	154	178	332	143	186	329	134	171	305
PROGRAMA DE POSGRADO EN ANTROPOLOGÍA	24	20	44	27	17	44	24	13	37
PROGRAMA DE POSGRADO EN ARQUITECTURA	33	46	79	36	29	65	33	30	63
PROGRAMA DE POSGRADO EN ARTES	43	26	69	32	28	60	29	23	52

Figura 14. Consulta de estudiantes por nivel académico y sexo utilizando el rol de autoridad universitaria. Fuente: Elaboración propia con datos de la Oficina de Registro.

	2011		Total 2011	2012		Total 2012	2013		Total 2013
Rótulos de fila	Femenino	Masculino		Femenino	Masculino		Femenino	Masculino	
POSGRADO	**2831**	**2534**	**5365**	**2577**	**2263**	**4840**	**2342**	**2067**	**4409**
ESTUDIOS DE POSGRADO	2831	2534	5365	2577	2263	4840	2342	2067	4409
PREGRADO Y GRADO	**17573**	**17400**	**34973**	**18242**	**17768**	**36010**	**18974**	**18441**	**37415**
ARTES Y LETRAS	1852	1428	3280	1922	1469	3391	2006	1536	3542
CIENCIAS AGROALIMENTARIAS	696	770	1466	748	776	1524	769	806	1575
CIENCIAS BASICAS	613	1008	1621	645	1042	1687	694	1094	1788
CIENCIAS SOCIALES	8789	7097	15886	9089	7203	16292	9404	7482	16886
INGENIERIA	1669	4347	6016	1741	4409	6150	1784	4487	6271
N/A	200	134	334	195	103	298	133	97	230
SALUD	2346	1116	3462	2433	1213	3646	2551	1286	3837
SEDES REGIONALES	1408	1500	2908	1469	1553	3022	1633	1653	3286
Total general	**20404**	**19934**	**40338**	**20819**	**20031**	**40850**	**21316**	**20508**	**41824**

Figura 15. Consulta de estudiantes por nivel académico y sexo utilizando el rol público. Fuente: Elaboración propia con datos de la Oficina de Registro.

Consulta 2.

Los 10 cursos con mayor cantidad de cupos matriculados por estudiantes distribuidos por sede para el último año. La consulta de la Figura 16 solo la puede realizar el rol de autoridad universitaria, ya que la dimensión de cursos está restringida de acuerdo con el apéndice 3, en la fase 3, en el proceso V-02 de validación de conceptos.

Rótulos de fila	2015
CIUDAD UNIVERSITARIA RODRIGO FACIO	**18426**
B 0106	1449
EG0124	3184
EG0125	2989
FS0210	1423
MA0001	1821
MA1004	1490
QU0100	1901
QU0101	1258
QU0102	1298
XE0156	1613

Figura 16. Consulta de cursos por cupos matriculados utilizando el rol de autoridad universitaria. Fuente: Elaboración propia con datos de la Oficina de Registro.

Consulta 3.

Títulos otorgados por grado académico distribuidos por año de ingreso a la U para los últimos cinco años. Para la consulta en la Figura 17, la desagregación es la misma al ser rol de investigador o autoridad universitaria. El rol público no tiene acceso al año de ingreso del estudiante.

Rótulos de fila	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	Total general
Bachillerato	**225**	**425**	**643**	**1112**	**1685**	**2701**	**2402**	**1977**	**1503**	**807**	**13480**
2011	86	171	258	483	727	885	29	4			2643
2012	58	107	164	296	418	769	773	21			2606
2013	35	68	109	151	267	519	787	759	10	1	2706
2014	20	40	67	117	172	322	490	700	767	14	2709
2015	26	39	45	65	101	206	323	493	726	792	2816
Diplomado	**3**	**7**	**18**	**17**	**21**	**36**	**61**	**59**	**64**	**40**	**326**
2011	1	2	7	7	10	17	15	8			67
2012	1	2	1	4	4	5	20	18	3		58
2013		1	6	1	4	2	11	19	29	10	83
2014	1		1	3		6	6	9	23	13	62
2015		2	3	2	3	6	9	5	9	17	56
Doctorado		**4**		**5**	**1**	**3**	**1**			**1**	**15**

Figura 17. Consulta de títulos otorgados utilizando el rol de autoridad universitaria. Fuente: Elaboración propia con datos de la Oficina de Registro.

5.5 Satisfacción del usuario

El uso de la consulta se aplica para validar la satisfacción del usuario con respecto a la información brindada luego de la construcción del DW. Las preguntas se enfocan a determinar si el modelo logra responder a las necesidades de los usuarios luego de una primera iteración. La consulta aplicada y las respuestas obtenidas se pueden verificar en el **apéndice 4.**

Se realizaron 7 preguntas de respuesta única y una para comentarios. Para las preguntas de respuesta única, existen 6 cuyos rubros de evaluación son “Sí”, “No” y “Parcialmente”, y para una pregunta solo se permite “Sí” o “No” como respuestas válidas. Para las preguntas con tres rubros se otorgan dos puntos por un “Sí”, un punto por “Parcialmente” y cero si la respuesta es “No”. Para la pregunta con solo dos valores, se otorgan dos puntos para un “Sí” y cero para un “No”.

El valor máximo posible para cada cuestionario es de 14 puntos, considerando como una puntuación aceptable obtener al menos 10 de los puntos posibles.

El muestreo es de tipo intencionado para valorar la satisfacción del usuario. Se aplicó el cuestionario a 4 funcionarios que han sido usuarios del sistema de información, durante el corto período de funcionamiento.

La puntuación obtenida en la consulta por funcionario, como se puede ver en el **apéndice 4,** fue de: primer funcionario con 13 puntos, segundo con 12 puntos, tercero con 12 puntos y el cuarto con 10 puntos. Esto es un valor promedio de 11,75 puntos en la consulta.

- En relación con la completitud de la información que corresponde a las preguntas 1 a la 3, se obtiene que, para los 24 puntos tabulados, la información brindada y las variables disponibles cumplen con las expectativas del usuario de manera total o parcial. El único punto donde se obtuvo un no, fue un usuario al cual no se le brindó información adicional, pero es porque no fue necesario.
- Para el tiempo de respuesta de la pregunta 4, todos los usuarios opinan que el tiempo de respuesta para brindar la información es el esperado.

- En torno a la confiabilidad de los datos en la pregunta 6, se obtienen 7 puntos de 8 posibles, esto indica que los usuarios perciben los datos como confiables, excepto uno, que consideró los datos como parcialmente confiables.
- La satisfacción en tiempo de respuesta, cantidad de información disponible, en la pregunta 7, considerada una respuesta de validación y para dar una perspectiva general sobre la satisfacción del usuario, se obtiene un promedio de 75% de satisfacción al relacionar el tiempo de respuesta con la cantidad de información ofrecida.
- Los comentarios de los usuarios reflejan la necesidad de variables adicionales, que se deben considerar en una próxima iteración, pero también mencionan una velocidad de respuesta más rápida de lo esperado, además de presentar los resultados de manera clara al usuario.

Los resultados arrojan que la nota más baja obtenida en el cuestionario fue de 71, y el más alto de 92, para una nota promedio de 83,3. En general, a partir de los criterios de validación del punto 3.3., se considera que los usuarios están satisfechos con el sistema.

Capítulo 6. Conclusiones y Recomendaciones

6.1 Conclusiones

La propuesta de un modelo para la creación de un almacén de datos en la Universidad de Costa Rica, surgió de la necesidad institucional de obtener información rápida, resumida e integrada de las diferentes instancias universitarias. Los factores detectados que justificaron su elaboración:

- Volumen de información.
- No hay almacenamiento histórico en el sistema transaccional.
- Solicitudes de información constantes sobre los mismos datos.
- Almacenamiento histórico no sistematizado.
- Toma de decisiones ralentizada por la demora en obtener la información.

La falta de una estructura previa para generar el DW universitario, propició la creación de un modelo de construcción de un DW para la Universidad de Costa Rica, conceptualizado de manera formal a partir de técnicas y recomendaciones derivadas de la investigación. El modelo se utilizará para sentar las bases de un plan a nivel universitario, que incluya la información de diferentes instancias. Para este proyecto, se aplicó el modelo de construcción a la información estudiantil disponible en la Oficina de Registro, fuente detectada como prioritaria en el punto 4.2.4 del modelo.

Los hallazgos de la investigación permitieron crear un conjunto de procesos replicables, con una línea de trabajo definida, para normar la construcción del DW para cada una de sus fuentes. El conjunto de procesos derivó en un modelo formal de construcción de un DW específico para la Universidad de Costa Rica.

El modelo se generó a partir de las características esenciales para crear un depósito de datos utilizando referencias de autores relacionados con las IES y los DW. A partir de estas referencias, se recopilaron las mejores prácticas y recomendaciones para lograr un modelo de calidad, al tomar en cuenta aspectos como: uso extendido, complejidad de la implementación y casos de éxito documentados en ambientes comparables.

De las técnicas recopiladas, se destacan cinco incluidas en el modelo final.

1. Arquitectura en 3 capas.
2. Implementación de abajo-arriba.
3. Matriz de procesos.

4. Diseño de 3 fases con el modelo dimensional.
5. Generación de metadatos.

La separación de los sistemas transaccionales del DW es necesaria por **la diferencia de cargas de trabajo y el tipo de infraestructura necesaria**. La infraestructura incluye el uso de bases de datos especializadas para datos analíticos, la conservación de datos masivos y aspectos de seguridad en el acceso a información sensible. Para realizar esta separación, se aplicó en el modelo una arquitectura de 3 capas, tal como se recomendó en el apartado 2.1 del marco metodológico, al utilizar la capa de datos, la capa de aplicación y la capa de presentación. Uno de los beneficios de un modelo específico como el propuesto, es que algunos procesos se definen desde la conceptualización del modelo. El apartado 4.2.2 incluye la documentación requerida para definir la arquitectura sin necesidad de recurrir a una nueva investigación.

De manera similar, el tipo de implementación del DW ya está definido en el modelo. La implementación ***bottom-up*** se aplica por las **características heterogéneas** de las unidades administrativas dentro de la Universidad, la autonomía para definir sus sistemas informáticos y el crecimiento paulatino de la infraestructura requerida. Además, el desarrollo es más rápido, por consiguiente, los entregables están disponibles en menor tiempo, esto mantiene el apoyo a nivel administrativo por medio de recursos.

El uso de la implementación *bottom-up* está asociada con una falta de visión a nivel de negocio, por generar soluciones individuales que se integran en un DW común. La técnica para disminuir el impacto de la construcción progresiva de la solución, es la matriz de procesos. La matriz de procesos permitió establecer un **marco general a nivel institucional** antes de ejecutar cualquier propuesta de implementación de las diversas fuentes. Inclusive, es la cual estableció el listado de fuentes y una prioridad para su integración al DW. La matriz también permitió visualizar las **dimensiones comunes** entre las diversas fuentes, requisito esencial para la integración de los datos. La matriz de procesos es la visión general que se requiere en un proceso de construcción de un DW, con los beneficios de la implementación de *bottom-up*.

El diseño propiamente del DW se realizó con una metodología de 3 capas, que incluye el modelo del negocio, el modelado dimensional, y la generación a nivel físico del DW. El

diseño incluye el tipo de esquema a utilizar, la definición de las dimensiones, medidas y cómo se relacionan. También incluye el dimensionamiento de la base de datos, la nomenclatura estándar de objetos, estrategias de indexado, particionamiento y la arquitectura de almacenamiento. De esta manera, no solo se elaboran los diagramas lógico, dimensional y físico, sino que se definen elementos importantes del DW para los procesos posteriores del modelo.

La integración de los elementos mencionados se documenta en los metadatos. Los metadatos permiten que **los datos sean comprendidos y utilizados** de manera amplia y correcta, por lo tanto, se incluye la información de los objetos generados como los esquemas, dimensiones, hechos, atributos, complementado con la documentación de los procesos, como la matriz de procesos, la arquitectura del DW y los protocolos de extracción de los datos.

De las técnicas implementadas y recomendaciones obtenidas de la investigación, se establecieron 17 procesos, agrupados en 5 fases temáticas.

Los procesos abarcan desde la propuesta para crear el DW hasta la implementación de cada una de las fuentes. Se **documentó cada proceso** y se revisó para no sobrecargar procesos con documentación innecesaria. Los procesos se definieron pensando en una ejecución lineal, pero de manera modular, para **que cada proceso tenga independencia** en cuanto a elementos de entrada y salida. La modularidad es la cual permite la iteración del modelo para las diferentes fuentes o de ser necesario, para la misma fuente.

Una vez establecidos los procesos y su coherencia en la ejecución, se verificó el modelo para permitir una iteración constante, que fuera aplicable para todas las fuentes de datos, actuales o a futuro. La posibilidad de utilizar **el mismo conjunto de procesos** para cada una de las fuentes era una característica esencial para el éxito de este modelo. El uso del modelo para las diferentes fuentes es posible, pero fue necesario crear agrupaciones para facilitar su uso en un entorno real.

Las agrupaciones se definieron a partir del conocimiento profesional, al mantener una **coherencia temática** con los procesos incluidos en cada uno. Las cinco fases quedaron definidas de la siguiente manera:

1. Proyecto (P)
2. Fuente de datos (F)
3. Diseño (D)
4. Visualización (V)
5. Implementación (I)

La agrupación de los procesos por temas facilitó la ejecución y el seguimiento del modelo, con ayuda de la plantilla incluida -**apéndice 2**- para el control de documentos de entrada y salida.

El cumplimiento de cada documento de salida en los procesos, es esencial para continuar con los procesos subsiguientes. Para facilitar el seguimiento, además de contar con la plantilla del modelo en el apéndice 2, se creó un diagrama SIPOC para representar los procesos. El objetivo fue crear un **diagrama fácil de interpretar y de seguir**, con la mayor cantidad de información disponible. Para lograrlo, se utilizó una versión modificada del SIPOC para mostrar las fases, los procesos, los documentos de entrada, salida y los actores involucrados en cada proceso **y mostrarlo en una sola hoja**. El resultado se puede revisar en el apartado 4.1 o en la cuarta página del apéndice 2.

El uso del modelo también se explicó en el apartado 4.1. En este caso el modelo permite iterar para cada una de las fuentes. El modelo propuesto es iterativo y no incremental. El modelo iterativo propone que **todas las funciones estén disponibles desde la primera iteración**, y se mejoren en las siguientes iteraciones. Con referencia a lo anterior, en el apartado 4.7 se estableció la posibilidad de ampliar la funcionalidad del modelo y pasar de una ejecución lineal a un **procesamiento en paralelo**, respetando el orden de los procesos ya establecido. Inclusive, se proponen dos variantes para ejemplificar su factibilidad y disponer de múltiples grupos de trabajo al aplicar el modelo de manera diferida.

En ese mismo sentido, resultó de vital importancia definir los accesos a la información por roles de usuario. El acceso se dio con la separación explícita por niveles de acceso y agregando complejidad a la consulta de acuerdo al usuario. Los accesos y su validación se realizaron de manera rápida con dos matrices: una para validar las dimensiones con sus atributos y otra para las medidas (apartado 4.5.2). Cada matriz muestra por rol a cual

tiene acceso. La validación de las matrices de acceso con las jefaturas resultó sencilla y se resolvió en reuniones personales de pocos minutos, lo recomendado para este tipo de puestos. Con las matrices, se explicó los límites que se pueden establecer de acceso a las variables y a las medidas del DW.

En este propósito y luego de la definición del modelo, el objetivo del proyecto fue probar el modelo al construir un sistema de información estudiantil. El **detalle de la aplicación del modelo** se puede ver en el **apéndice 3**.

El modelo se caracteriza por solicitar **pocos documentos en cada proceso** para facilitar la tarea de revisarlos e integrarlos. A su vez, las salidas de documentos suelen ser pocas; en más de la mitad de procesos, se solicita un solo documento de salida. Esto también simplifica el avance entre procesos.

Con relación a este tema, se considera que la carga de trabajo adicional al implementar este modelo es baja, con respecto a una construcción no formal, pero con los beneficios de **seguimiento y trazabilidad** que el modelo permite. A lo largo del modelo, se ejecutan 17 procesos, se consultan 85 documentos y se generan 25 documentos nuevos en una iteración. La cantidad de documentos nuevos se reduce luego de la primera iteración, tal como se explica en el apartado 4.1, en donde la primera fase solo debería ejecutarse una vez durante el proyecto. Además, existen algunos documentos intrínsecos al crear un DW, como el *script* de creación de la base de datos o los ETL. Por lo tanto, al excluir esta documentación y los de la primera fase, la cantidad de documentos nuevos generados durante la ejecución del modelo se reduce a 20 luego de la primera iteración.

Para brindar un punto objetivo de análisis del modelo, se consideró la satisfacción del usuario por medio de un muestreo dirigido. Se aplicó un cuestionario a los usuarios que solicitaron información del DW luego de su construcción. El cuestionario se evaluó en base a 14 puntos, de los cuales, luego de la tabulación -**apéndice 4**- en promedio se obtuvo 11,75 puntos. A partir de los criterios de validación del apartado 3.3., se considera que los **usuarios están satisfechos** con el sistema con una nota de 83,3. El punto débil detectado a partir de los resultados del cuestionario, es la cantidad limitada de información disponible, inquietud comprensible al ejecutar una sola iteración. Los puntos fuertes fueron la velocidad de respuesta del sistema y la confianza en los datos ofrecidos.

Sobre la base de las consideraciones anteriores y en concordancia al marco metodológico, se enumeran los resultados que dictan cuando se considera que el modelo es adecuado a partir de la Tabla 2 de evaluación del apartado 3.3. Los criterios y la escala de valores establecen, de acuerdo con una ponderación, si el modelo se considera adecuado o no según los requisitos iniciales. La Tabla 9 muestra los **criterios de validez y el valor logrado** en cada criterio.

Criterio	**Valor**	**Valor mínimo**	**Valor adecuado**	**Valor óptimo**	**Valor logrado**
Satisfacción de usuarios.	35%	10% si la evaluación arroja una satisfacción mayor a 50.	25% si la evaluación arroja una satisfacción mayor a 70.	35% si la evaluación arroja una satisfacción mayor a 90.	**25%** con una nota promedio de 83 de satisfacción de los usuarios.
Total de indicadores solicitados vs. los indicadores implementados	20%	10% si se cumplen en una proporción mayor a 50.	15% si se cumple en una proporción superior a 70.	20% si se cumplen en una proporción de 90.	**15%** 14 indicadores solicitados y 4 no implementados, resulta en un 71% de indicadores.
Tiempo de respuesta máximo a las consultas solicitadas por los usuarios a la base de datos.	10%	0% si tarda menos de 8 segundos en responder.	5% si tarda menos de 5 segundos.	10% si tarda menos de 3 segundos.	**0%** La consulta de los datos por medio de los cubos en SQL, permiten un tiempo de respuesta de hasta 15 segundos.
Cantidad de problemas encontrados	15%	5% si se reportan menos de 15	10% si se reportan menos de 10	15% si se reportan menos de 5	**10%** Se reportaron más de 5 observaciones que

en el periodo de prueba (pulgas).		problemas.	problemas.	problemas.	requieren cambios en el DW.
Cumplimiento en el plazo establecido para la construcción del DW.	20%	Un 10% si se tarda un mes más de lo planificado.	Un 15% si se tardan 2 semanas más de lo esperado.	Un 20% si se entrega en el tiempo establecido.	**20%** La construcción se realizó acorde a la duración estimada de 4 a 5 meses.

Tabla 9. Evaluación para cuantificar lo adecuado del modelo al negocio.

Fuente: Elaboración propia.

El punto débil detectado a partir de los criterios de evaluación es el tiempo de respuesta de las consultas. Este tiempo difiere del mencionado en la satisfacción del usuario, ya que el tiempo que el usuario requería para obtener la misma información podría tardar horas o días, pues siempre requería una solicitud formal. El tiempo mencionado en la tabla de evaluación del modelo, se refiere a las consultas a la base de datos, que deben rondar pocos segundos y al no cumplir con las expectativas, recibe un puntaje de 0.

Lo positivo de la aplicación del modelo fue el tiempo que tardó la ejecución de la primera iteración, que se conservó dentro del rango de tiempo de duración del proyecto. Además, el total de indicadores solicitados contra los implementados cumple de manera apropiada las expectativas, así como la cantidad de problemas encontradas durante el proceso de ejecución del modelo.

De la Tabla 9 y a partir de los valores establecidos en el marco metodológico para considerar el modelo adecuado, se determinó que **el modelo es aceptado** con un valor de 70 a partir de los criterios definidos en la tabla de evaluación del modelo al negocio.

En resumen, el modelo permitió formalizar la construcción del DW aplicando técnicas probadas en ámbitos similares, al ofrecer los beneficios de una ejecución iterable, modular y aplicable a múltiples fuentes, con flexibilidad en la ejecución.

6.2 Recomendaciones

La construcción del DW es útil para ordenar los datos de la institución, pero es solo un paso en la sistematización de la información. Para la consulta de los usuarios, se deben crear los cubos para análisis analítico. Los cubos le van a permitir a los usuarios un cruce de información que no estaba disponible con la información dispersa en la fuente y en el futuro, el cruce con otras fuentes de datos.

En este mismo orden, se debe implementar un modelo de minería de datos con la información sobre estudiantes. La información recopilada en un DW no es esencial para aplicar minería sobre los datos, pero si existe de previo, facilita el acceso de la información y garantiza en buena medida, la calidad de los datos ofrecidos. Las oportunidades de ubicar nuevos patrones para la toma de decisiones debe ser suficiente aliciente para un proyecto de minería sobre estos datos históricos.

Los datos actuales en el DW sobre estudiantes, presentan múltiples posibilidades de indicadores como: tasa de cobertura, tasa de absorción, incremento en nuevos cupos, permanencia estudiantil y tiempo de graduación. Lo que se requiere es un estudio de cómo se deben realizar estos cálculos y la pertinencia de cada uno. Los indicadores mencionados requieren análisis adicional, pero son derivados de la información ya disponible.

Es necesario definir de manera formal, los procedimientos para el aseguramiento de la calidad de la información. Su aplicación debe ser desde la etapa de diseño y listar los datos de control a utilizar y cómo se van a aplicar contra los datos fuente, para determinar la validez de la información. Se debe cumplir con el aseguramiento de calidad para brindar soluciones estables, permitiendo el uso del tiempo en tareas de análisis y no en revisión de los datos.

El uso del modelo en paralelo no fue probado en la práctica, pero el documento en el apartado 4.7, explica de forma amplia como se debe ejecutar y las posibilidades de utilizarlo. El aprovechar las iteraciones del modelo con múltiples equipos trabajando sobre distintas fuentes, implicaría una reducción significativa en el tiempo para los entregables, pues la metodología está diseñada para que los procesos no interfieran unos con otros si se siguen las recomendaciones mencionadas en el mismo apartado.

La aplicación del modelo debe estar fundado en un sistema de indicadores que definan las prioridades de información y lo esencial para la toma de decisiones, tal como se realizó a partir del Plan Estratégico Institucional. Se recomienda utilizar esta línea estratégica como base para futuras revisiones del modelo.

6.3 Reflexiones Finales

Resulta adecuado establecer una metodología para el acercamiento a las diferentes fuentes, de manera que no sea una carga de trabajo, sino que se le brinde un valor agregado a la unidad el facilitar la información. De esta manera la cooperación será más sencilla y más rápida. Lo mejor es tener patrocinadores en las diferentes oficinas fuentes para establecer una jerarquía de orden al solicitar y facilitar el flujo de la información.

La existencia de un DW en tiempo real será necesario cuando el DW sea un producto maduro dentro de la IES. Será un valor agregado, pero aplicable en ámbitos limitados, y aprovechado por pocos usuarios. Es por esto que debe conservarse como una mejora, pero no se considera prioritaria. Un ejemplo de aplicación puede ser la ocupación real de los cursos durante el periodo de matrícula, analizar la tendencia para cursos de alta saturación, de baja matrícula y optimizar el uso de los espacios disponibles.

El modelo está construido en función de los requerimientos de la Universidad de Costa Rica, pero debe ser extrapolable a otras instituciones de educación superior públicas en Costa Rica. El modelo se propone como procesos individuales, modificables a partir de los requisitos de cada institución, en especial los que se caracterizaron desde el modelo, como por ejemplo la arquitectura o el de visualización.

El reto es convertir los datos de los estudiantes en información para la toma de decisiones. Lograr que la Universidad sea una institución fundamentada en datos, implica que cualquier usuario tomador de decisiones debe disponer de la información que requiera de manera inmediata. No es solo tener muchos datos o reportes disponibles, sino ofrecer la oportunidad al usuario de mezclar datos y obtener sus propias conclusiones. Este tipo de oportunidades colocaría a la Universidad en un lugar privilegiado en manejo de la información, si se logra aplicar en el corto o mediano plazo.

Bibliografía

Baranović, M., Madunić, M., & Mekterović, I. (2003). Data warehouse as a part of the higher education information system in Croatia. IEEE Transactions On Image Processing. http://dx.doi.org/10.1109/iti.2003.1225332

Hua-long Zhao,. (2008). Application of OLAP to the analysis of the curriculum chosen by students. 2008 2Nd International Conference On Anti-Counterfeiting, Security And Identification. http://dx.doi.org/10.1109/iwasid.2008.4688341

Kimball, R., & Ross, M. (2011). *The data warehouse toolkit: the definitive guide to dimensional modeling*. Indianapolis, Ind.: John Wiley & Sons.

Ningning, G. (2010). Proposing Data Warehouse and Data Mining in Teaching Management Research. 2010 International Forum On Information Technology And Applications, 1, 436,439. http://dx.doi.org/10.1109/ifita.2010.286

Selmoune, N., & Alimazighi, Z. (2008). A decisional tool for quality improvement in higher education. 2008 3Rd International Conference On Information And Communication Technologies: From Theory To Applications, 1,6, 7-11. http://dx.doi.org/10.1109/ictta.2008.4530368

Referencias

Acerca de la CGR. (2016). *Contraloría General de la República de Costa Rica*. Recuperado el 21 marzo 2016, desde http://www.cgr.go.cr/conozcanos/acerca-cgr

Álvarez-Gayou, J. (2003). *Cómo hacer investigación cualitativa.* México: Paidós.

Ballard, C. (1998). *Data modeling techniques for data warehousing*. [Estados Unidos]: IBM.

Barrantes, R. (1999). *Investigación: un camino al conocimiento, un enfoque cualitativo y cuantitativo*. San José, Costa Rica: EUNED.

Bautista, N. (2011). *Proceso de la investigación cualitativa: epistemología, metodología y aplicaciones*. Bogotá: Manual Moderno.

Biolchini, J., Mian, P., Natali, A., & Travassos, G. (2005). Systematic Review in Software Engineering: Relevance and Utility. Technical Report ES-679/05, PESC-COPPE/UFRJ.

Bloom, B. (1990). Taxonomía de los objetivos de la educación. Buenos Aires: El Ateneo.

Castro, E. (2011). *Modelado de Data Warehouse*. [Presentación].

Chavarría, C., (2011). *La dicotomía cuantitativo/cualitativo: falsos dilemas en investigación social.* Actualidades en Psicología. Volumen 25, 1-35.

Giménez, S., Motz, R., Carpani, F., Gayoso, D., & Colombatto, C. (2008). Diseño Multidimensional guiado por Ontología. In *Congresso Ibero-Americano em Engenharia de Software*. Pernambuco, Brasil: Maria Lencastre, João Falcão e Cunha, Antonio Vallecillo.

Grad, H. (noviembre, 2009). Evaluación institucional del sistema universitario español: Una visión de la calidad de la evaluación basada en indicadores de rendimiento. *Boletín de Psicología,* (97), 55-69. Obtenido desde http://www.uv.es/seoane/boletin/previos/N97-4.pdf

Granados, C. E.(2015). *Caracterización de la Información Requerida para la Rendición de Cuentas en las Universidades Estatales* (Tesis de maestría sin publicar). Instituto Centroamericano de Administración Pública. San José, Costa Rica.

Hüsemann, B., Lechtenbörger, J., & Vossen, G. (2000). *Conceptual Data Warehouse Design*. Lecture, Stockholm, Suecia.

Malinowski, E., & Zimányi, E. (2008). *Advanced data warehouse design*. Berlin: Springer.

Moura, J. (2015). *Data Warehouse Architecture Overview*. [Presentación].

Olap.com,. (2016). *What is the Definition of OLAP? OLAP Definition*. Recuperado el 26 de febrero de 2016, desde http://olap.com/olap-definition/

Ortiz, M. (2012a). *Modelo Incremental ~ Ingeniería de Software*. *Isw-udistrital.blogspot.com*. Recuperado el 26 de abril de 2016, desde http://isw-udistrital.blogspot.com/2012/09/ingenieria-de-software-i.html

Ortiz, M. (2012b). *Modelo Iterativo ~ Ingeniería de Software*. *Isw-udistrital.blogspot.com*. Recuperado el 26 de abril de 2016, desde http://isw-udistrital.blogspot.com/2012/09/ingenieria-de-software-continuacion.html

Paoletti, L. (2012). *Data Warehousing: "Conceptual Architecture" - Design*. *Tom's IT Pro*. Recuperado el 26 de febrero de 2016, desde http://www.tomsitpro.com/articles/data_warehsouing-business_intelligence-data_warehouse_conceptual_design,2-271.html

Rainardi, V. (2008). *Building a data warehouse with examples in SQL Server*. Berkeley, CA: Apress.

RIACES (2004). *Glosario internacional de evaluación de la calidad y acreditación*. Madrid: Agencia Nacional de Evaluación de la Calidad y Acreditación.

Ross, M. (2005). *The Matrix: Revisited*. *Kimball Group*. Recuperado el 7 de abril de 2016, desde http://www.kimballgroup.com/2005/12/the-matrix-revisited/

Sharma, P. (2009). Advanced Applications of Data Warehousing Using 3-tier Architecture. *DJLIT*, *29*(2), 61-66. http://dx.doi.org/10.14429/djlit.29.242

The Benefits of Data Modelling in Data Warehousing. (2008). CA Technologies.

Universidad de Costa Rica. (2016). *Plan de desarrollo institucional 2013-2017*. *Oficina de Planificación Universitaria.* Recuperado el 26 de enero de 2016, desde http://www.oplau.ucr.ac.cr/index.php/plan-estrategico-institucional

Universitat Politècnica de València. (2013). *Hacia un Sistema Integral de Información para la Educación Superior de América Latina* (pp. 12,17). Valencia, España. Obtenido desde http://www.infoaces.org/web-proyecto/documentacion-infoaces.html

Vargas Cordero, Z. (2009). La Investigación aplicada: Una forma de conocer las realidades con evidencia científica. *Rev. Educación*, *33*(1), 155-165. http://dx.doi.org/10.15517/revedu.v33i1.538

Abreviaturas y Acrónimos

BD	Base de datos.
BI	Inteligencia de Negocios.
DM	Data mart.
DW	Almacén de datos o Data Warehouse.
ETL	Extraer, transformar y cargar o Extract, transform, load.
GB	Gigabyte.
IES	Institución de Educación Superior.
MDX	Lenguaje y expresiones de consulta para análisis multidimensionales.
OLAP	Procesamiento Analítico en Línea.
OPLAU	Oficina de Planificación Universitaria de la Universidad de Costa Rica.
SA	Área de intercambio, área temporal o staging area.
SIPOC	Diagrama Supplier–Inputs-Process-Outputs-Customers o de procesos de gestión
SQL	Lenguaje de consultas estructurado.
TI	Tecnologías de la Información.
UCR	Universidad de Costa Rica.

Glosario

Granularidad

Nivel de detalle del dato.

Llave sustituta

La llave primaria de las dimensiones en un data warehouse está determinada por un identificador de tipo numérico por razones de eficiencia y de mantenimiento. Esta llave viene a reemplazar la llave primaria de la fuente.

Medidas

Es lo que se va a analizar, son numéricas y suelen poder agregarse según el nivel de detalle.

Nota ponderada

La nota promedio de un periodo que contabiliza el peso de los créditos de cada curso para el cálculo, además que descarta algunos cursos no ponderables según políticas de la universidad.

Tabla de hechos

Los hechos son los indicadores del negocio y esta tabla es la que los reúne.

Apéndices

Apéndice 1. Revisión sistemática para el estado de la cuestión

Índice

1. Formulación de la pregunta 102
2. Selección de fuentes 104
3. Selección de los estudios 106
4. Extracción de la información 111
5. Resumen de resultados 119
5.1. Situación actual 120

Revisión Sistemática para el Estado de la Cuestión

1. Formulación de la pregunta

En la etapa de planificación se definen los objetivos y las áreas a trabajar en la investigación, y se determinan de manera certera el problema y sus principales características.

1.1. Objetivo de la pregunta

Con esta revisión sistemática se pretende ubicar investigaciones referentes a modelos para crear depósitos de datos con énfasis en el sector educativo, relacionados con la educación superior.

1.2. Amplitud y calidad de la pregunta

Para medir la amplitud y calidad se debe establecer el problema a tratar, y luego, delimitarlo por medio de palabras claves. Se define un punto de referencia y el público meta, además de indicar los métodos aplicados para contabilizar los resultados obtenidos.

1.2.1. Problema

Se debe conocer el estado del arte en cuanto a modelos de creación de depósitos de datos, enfocados al sector educativo, en especial, a lo que se refiere a educación superior.

1.2.2. Pregunta

¿Cuál es el estado del arte de modelos de creación de depósitos de datos relacionados con la educación superior?

1.2.3. Palabras clave y sinónimos

La tabla 1 presenta una lista de palabras claves y términos relacionados que sirvan como insumo para la selección de estudios primarios de investigación.

Área de estudio	Palabras claves	Términos relacionados
Diseño de un modelo	modelo de diseño design model design modelling modelling design	build modelling
Depósito de datos	Data warehouse DW	DWH EDW
Educación superior	high education	sector educativo educational environment

Tabla 1. Palabras claves para seleccionar los estudios primarios.

1.2.4. Intervención

Se hace una revisión de los modelos de creación de un depósito de datos enfocado al sector educativo, en especial de la educación superior, y buscan casos de éxito en la aplicación de modelos con este enfoque.

1.2.5. Control

No se ha realizado una revisión relevante previa, por lo que los resultados que arroje esta revisión sistemática, determinarán las fuentes de información primarias a utilizar en esta investigación.

1.2.6. Efecto o resultados esperados

Se espera establecer el estado del arte en lo que se refiere a modelos de creación de depósitos de datos enfocados a instituciones de educación superior o al menos del sector educativo. Además, encontrar algunos ejemplos de casos de éxito aplicando un modelo establecido en el área de la educación superior.

1.2.7. Métricas

Se va a realizar una comparación de los modelos encontrados. Además, se va a señalar el grado en que los ejemplos aplicados pueden ser repetidos a

partir de la descripción incluida en el documento, para determinar su completitud.

1.2.8. Población

La población es la lista de publicaciones recolectadas que estén relacionadas con el objetivo planteado.

1.2.9. Aplicación

Personal que esté relacionado con el área de la inteligencia de negocios. Los que se encuentren desarrollando un proyecto de análisis de datos para la educación superior.

1.2.10. Diseño experimental

El meta-análisis de la revisión se va a enfocar en los modelos de creación de depósitos de datos para la educación superior para conocer las tendencias actuales, y con mayor relevancia la existencia de documentos sobre casos de éxito en su aplicación.

2. Selección de fuentes

A continuación, se presentan los criterios para definir las fuentes de los datos en los idiomas establecidos. Luego, se muestran los métodos de búsqueda con las palabras clave ya definidas y la lista de fuentes de datos. A partir de estos datos se define una evaluación de las fuentes y una comprobación de las referencias.

2.1. Criterios para definir las fuentes de los datos

Las características de las fuentes para ser tomadas en cuenta en la revisión.
Deben ser accesibles desde la web.

2.2. Idiomas de los estudios

El idioma de los estudios será en inglés, y se obtendrán a partir de las palabras clave definidas en el apartado 1.2.3. en las fuentes de datos definidas. Si se obtiene una fuente primaria en español, podrá ser utilizada, siempre y cuando cumpla con los criterios mínimos establecidos para definir las fuentes de los datos.

2.3. Identificación de las fuentes

Se seleccionan las fuentes para ejecutar la revisión.

2.3.1. Métodos de búsqueda

La búsqueda se realizará primero en un buscador web general. Los resultados serán filtrados en primera instancia a partir de la dirección web y que concuerden con las fuentes confiables establecidas.

2.3.2. Cadenas de búsqueda

La generación de una sola cadena no es algo obligatorio, y se puede establecer una mejor guía el ejecutar varias consultas con palabras similares, donde los resultados coincidentes sean los más relevantes por contener una mayor cantidad de palabras claves. Pero para efectos de la revisión sistemática es necesario definir una solo cadena de búsqueda que pueda ser reproducible por cualquier investigador y que no esté supeditada a subjetividades del autor.

La cadena se muestra con una serie de palabras clave y operadores lógicos, los cuáles serán modificados de acuerdo a la fuente de datos para empatarlos con su formato de consulta. El orden de las cadenas sí es relevante, y la primera tiene más peso. Si se considera que la primera cadena retorna la lista de fuentes primarias necesarias, no ejecutarán las siguientes cadenas.

La lista de cadenas de búsqueda por aplicar será:

1. `Data warehouse AND modelling AND higher education`
2. `(modelo OR creación) AND ((depósito de datos) OR (data warehouse)) AND (educación superior)`
3. `(design model OR build modelling) AND (data warehouse OR dw OR edw) AND in high education`
4. `(design model OR build modelling) AND (data warehouse OR dw OR edw) AND educational`

2.3.3. Listado de fuentes

La lista de fuentes está ordenada por prioridad, de acuerdo a su relevancia con los campos de tecnologías de información.

1. IEEE digital library
2. ACM digital library
3. Web of Knowledge
4. ScienceDirect
5. Scopus
6. Cordis

2.4. Selección de fuentes luego de la evaluación

Si la fuente cumple con todos los criterios de selección, debe ser incluida en la lista de fuentes primaria.

2.5. Comprobación de las fuentes

Se incluyeron las principales fuentes de datos, pero se había dejado por fuera ScienceDirect que tiene relevancia en el campo de investigación de tecnología y se elimina Elsevier, que no es en sí la fuente para obtener los documentos.

3. Selección de los estudios

En el punto 2 se definieron las fuentes. Ahora es necesario describir los procesos y criterios para la selección y evaluación de los documentos.

3.1. Definición de los estudios

Explica cómo se van a seleccionar los documentos.

3.1.1. Criterios de inclusión y exclusión de estudios

La inclusión de resultados se dará a partir de:

- Una revisión del título
- Lectura del resumen y que tenga relación directa con las palabras claves.
- No sobrepasar los primeros 15 resultados.

La exclusión de resultados se aplicará cuando los documentos sean:

- Enfocados a la enseñanza o dirigido a profesores (teaching)
- Que sea una revisión de políticas institucionales y no un estudio técnico.
- La solución a un sistema específico.
- Que no provenga de un journal o conferencia documentada como de prestigio y clasificada dentro del cuartil 1 o 2.

3.1.2. Tipos de estudio

El tipo de estudio estará determinado por la completitud de los criterios antes expuestos, y no supeditados a una discriminación en este sentido.

3.1.3. Procedimiento para la selección de los estudios

De acuerdo a la figura 1, se sigue un proceso iterativo, donde se revisan en cada fuente los criterios de inclusión y exclusión para encontrar los estudios primarios de una fuente.

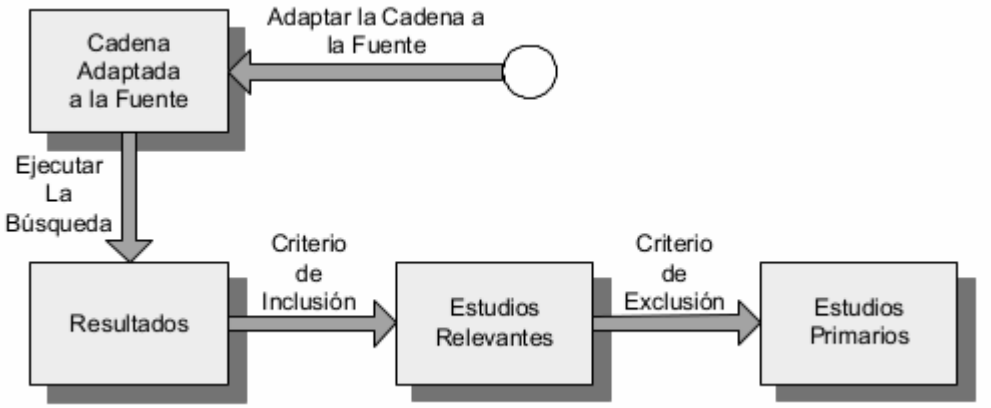


Figura 1. Procedimiento de selección estudios[8]

Lo primero es ejecutar la consulta en el formato requerido por la fuente. Se aplican los criterios de inclusión a los documentos encontrados, para luego

[8] Blanco, Lasheras, Valencia-García. 2008. Ontologías de seguridad: Revisión sistemática y comparativa.

aplicar los criterios de exclusión. De esta manera podemos extraer los estudios primarios de la fuente de datos. Este proceso se ejecuta para cada una de las fuentes definidas en la lista de fuentes del apartado 2.3.3.

3.2. Ejecución de la revisión

Luego de evaluar lo planificado, se inicia la ejecución de la revisión sistemática. A partir de ahora, las búsquedas deben ejecutarse según los criterios establecidos. Al final, tenemos que extraer la información relevante a la pregunta planteada a partir de los estudios obtenidos.

3.2.1. Selección inicial de estudios

IEEE Digital Library

Se ejecutó la cadena “Data warehouse AND modelling AND higher education” con la búsqueda básica[9] del sitio y no requirió adaptación.

- La búsqueda devolvió 7 resultados.
- Luego de aplicar los criterios de inclusión, quedaron 4 documentos.
- Luego de aplicar los criterios de exclusión se conservan los 4, pues el criterio de exclusión del ranking no es aplicable a las conferencias y cumplen el resto de requisitos.

Baranovic, M.; Madunic, M.; Mekterovic, I., "Data warehouse as a part of the higher education information system in Croatia," *Information Technology Interfaces, 2003. ITI 2003. Proceedings of the 25th International Conference on* , vol., no., pp.121,126, 16-19 June 2003 http://dx.doi.org/10.1109/ITI.2003.1225332 No fue citado.
Hua-long Zhao, "Application of OLAP to the analysis of the curriculum chosen by students," *Anti-counterfeiting, Security and Identification, 2008. ASID 2008. 2nd International Conference on* , vol., no., pp.97,100, 20-23 Aug. 2008 http://dx.doi.org/10.1109/IWASID.2008.4688341

[9] IEEE Xplore Digital Library. http://ieeexplore.ieee.org/Xplore/home.jsp

Gao Ningning, "Proposing Data Warehouse and Data Mining in Teaching Management Research," *Information Technology and Applications (IFITA), 2010 International Forum on* , vol.1, no., pp.436,439, 16-18 July 2010 http://dx.doi.org/10.1109/IFITA.2010.286 No fue citado.
Selmoune, N.; Alimazighi, Z., "A decisional tool for quality improvement in higher education," *Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008. 3rd International Conference on* , vol., no., pp.1,6, 7-11 April 2008 http://dx.doi.org/10.1109/ICTTA.2008.4530368

Web of Science

Se ejecutó la cadena “Data warehouse AND modelling AND higher education” con la búsqueda básica[10] del sitio. Al obtener los resultados, se especifica que el dominio es “Science Technology” y se excluye “Social Sciences”.

- La búsqueda devolvió 4 resultados.
- Luego de aplicar los criterios de inclusión, quedaron 0 documentos.

ACM Digital Library

Se ejecutó la cadena “Data warehouse AND modelling AND higher education” con la búsqueda avanzada[11] del sitio seleccionando las búsquedas por abstract.

- La búsqueda devolvió 1 resultado.
- Luego de aplicar los criterios de inclusión, quedaron 0 documentos.

ScienceDirect

Se ejecutó la cadena “Data warehouse AND modelling AND higher education” con la búsqueda básica[12], avanzada y experto del sitio seleccionando todas las búsquedas.

[10] Web of Science. http://apps.webofknowledge.com/
[11] ACM Digital Library. http://dl.acm.org/dl.cfm

- Para la búsqueda experto no se encontraron coincidencias. Para la búsqueda básica devuelve 5440 resultados.
- Luego de aplicar los criterios de inclusión, quedó 1 documento relacionado.
- Luego de aplicar los criterios de exclusión, quedaron 0 documentos.

Scopus

Se ejecutó la cadena <"Data warehouse" AND modelling AND "higher education"> con la búsqueda básica[13].

- La búsqueda devolvió 6 resultados.
- Luego de aplicar los criterios de inclusión, quedaron 4 documentos.
- Luego de aplicar los criterios de exclusión, quedaron 4 documentos, pero 3 de ellos ya estaban incluidos en la búsqueda en la IEEE, por lo que no serán revisados de nuevo.

Mattingly, K. D., Rice, M. C., & Berge, Z. L. (2012). Learning analytics as a tool for closing the assessment loop in higher education. Knowledge Management and E-Learning, 4(3), 236-247.

3.2.2. Evaluación de la calidad de los estudios

Todos los documentos se consideran de calidad pues cumplieron con los requisitos de inclusión y de exclusión.

3.2.3. Revisión de la selección

La revisión de estudios determina que la selección de fuentes es correcta y se debe proceder a extraer la información relevante a la investigación.

Se considera que se debe incluir en el estudio, el libro de Ralph Kimball sobre herramientas para Data Warehouses. La inclusión procede, ya que el autor es un

[12] Science Direct. http://www.sciencedirect.com/
[13] Scopus Elsevier. http://www.scopus.com/

exponente indiscutible en el diseño de data warehouses, con múltiples publicaciones en el tema, y el libro contiene un capítulo específico acerca de la creación de un modelo de data warehouse para el sector educativo, utilizando de ejemplo una institución de educación superior.

Kimball, R., & Ross, M. (2011). *The data warehouse toolkit: the definitive guide to dimensional modeling*. John Wiley & Sons.

4. Extracción de la información

La recopilación de las fuentes fue completada y validada, y se procede a extraer la información relevante. A continuación, se definen los criterios de extracción y los resultados obtenidos.

4.1. Criterios de inclusión y exclusión de la información

Los criterios para considerar la información de una fuente deben ser relevantes al estudio:

- Trabajos teóricos sobre construcción de un depósito de datos.
- Técnicas recomendadas de modelado en un entorno educativo.
- Trabajos prácticos que demuestren la factibilidad al aplicar técnicas de modelado de datos relacionados al entorno educativo.
- Recomendaciones y buenas prácticas en el modelado y mantenimiento de un depósito de datos.
- Problemas y soluciones al diseño y modelado de un data warehouse en educación superior.

4.2. Formulario de extracción de datos

La tabla 2 presenta el formato que debe utilizarse para cada fuente de datos al extraer la información.

Título original	***Título original del texto consultado***
Publicación	*Nombre de la conferencia o publicación incluyendo la fecha y páginas.*
Autor	*Nombre de los autores.*
Institución que respalda al autor	*Nombre de la Universidad, instituto o empresa patrocinadora de la investigación.*
Resumen	*Breve descripción del objetivo de la publicación.*
Datos relevantes	*Información relacionada con el título de mi investigación extraída de la publicación.*

Tabla 2. Formato esperado para la extracción de información.

Fuente: Elaboración propia.

4.3. Ejecutar la extracción

Se presenta la información extraída de cada fuente utilizando la tabla 2 presentada en el apartado anterior.

Título original	**Data Warehouse as a Part of the Higher Education Information System in Croatia**
Publicación	25th Int. Conf. Information Technology Interfaces ITI 2003, June 16-19, 2003.
Autor	Mirta Baranovic. Mirjana Madunic, Igor Mekterovic.
Institución que respalda al autor	Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia
Resumen	Menciona los problemas encontrados al modelar e implementar un depósito de datos para el sistema nacional de información de la

	educación superior en Croacia, para mejorar la planificación y operación de las actividades educativas.
Datos relevantes	Objetivos: Proveer una estructura para análisis de datos rápidos y eficientes por parte de las autoridades (decanos, vice), y un front-end web intuitivo para que los usuarios obtengan sus reportes con un servicio de alta disponibilidad. Implementado: 1. Arquitectura del DW. 2. Modelo ETL para el DW. 3. Seguridad de la información. 4. Presentación de los datos. 1. Arquitectura: Sist Transaccional/Relacional -> Copia del Relacional -> Copia de la BD Multidimensional -> BD Multidimensional -> BD OLAP. Esquema en estrella. Ejemplos: Exam taking process y Exam efficiency by course. 2. Modelo ETL: carga del SA y del OLAP se hace en cargas nocturnas. Identificar solo los datos modificados o nuevos. a. Si un elemento es borrado, y luego se crea un nuevo elemento con el mismo id pero diferentes atributos. Que hacer? Solución: actualizar la Fact y las dimensiones. b. Brinda datos de tiempos de ejecución con cantidades de registros. c. Brinda 12 pasos a seguir para la tarea automatizada. d. Eliminar temporalmente índices, llaves primarias y foráneas para incrementar el rendimiento. 3. Seguridad: La protección de la comunicación se realiza por https. El manejo de las cuentas de acceso se realizan a partir del sistema relacional. La cantidad de roles existentes es casi la cantidad de usuarios. Todos tienen permisos de lectura muy específicos. 4. Presentación: se establecen 3 categorías de consultas: predefinidas, detalladas personalizadas y de agregación personalizadas. Predefinidas-> con SQL o MDX y el conjunto reducido de manera automática con un script según el rol y

	permisos del usuario. No indica si los permisos son en BD. Detalladas-> SQL parametrizado desde listas o campos de texto. Se realizan al DW, y son consultas atípicas por el grado de detalle para un sistema analítico. Agregación->Scripts DHTML a los cubos OLAP. puede hacer drill-down y filtros. Consideraciones técnicas: se usa SQL server 2000 para la base de datos análitica y programado en ASP para el acceso a los datos, Visual Basic 6 para la aplicación de administración.

Tabla 3. Baranovic 2003.

Fuente: Elaboración propia.

Título original	**Application of OLAP to the Analysis of the Curriculum Chosen by Students**
Publicación	Anti-counterfeiting, Security and Identification, 2008. ASID 2008. 2nd International Conference on , vol., no., pp.97,100, 20-23 Aug. 2008
Autor	Hua-long Zhao
Institución que respalda al autor	Department of Geography, Dezhou University, China
Resumen	Análisis para la toma de decisiones acerca del currículum seleccionado por los estudiantes. Diseño del depósito de datos para modelar un cubo multidimensional OLAP y aplicar procesos de minería de datos.
Datos relevantes	1. El diseño de la estructura del modelo de datos es el punto crucial en la creación de un DW. Define el tipo de análisis posible y la complejidad del mantenimiento del sistema.

	2. Ejemplo práctico de creación de un esquema de estrella para un DW. El modelo estrella permite el análisis requerido de la información del estudiante por medio del modelo multidimensional. Consideraciones técnicas: SQL Server Analysis Services.

Tabla 4. Zhao 2008.

Fuente: Elaboración propia.

Título original	**Proposing Data Warehouse and Data Mining in Teaching Management Research**
Publicación	Information Technology and Applications (IFITA), 2010 International Forum on , vol.1, no., pp.436,439, 16-18 July 2010
Autor	Gao Ningning
Institución que respalda al autor	Information Engineering College, Central University for Nationalities, Beijing, China
Resumen	Aplicación y puesta en práctica creando un depósito de datos, análisis de datos con redes neuronales, teorías de aproximación (Rough Set theory) para predecir los estudiantes que pueden dejar un curso.
Datos relevantes	Importancia del ETL y del diseño previo de un DW. Menciona 5 pasos para crear un DW: 1. Modelo de datos. El blueprint del DW. 2. Extracción de los datos. Puede ser full refresh o incremental 3.Administración de los datos. Eliminar redundancias, conversión de tipos, agregar datos, asegurar consistencia. 4.Administración del almacenamiento. El DBMS debe soportar los procesos y accesos masivos. 5.Acceso a los datos. Control de roles de acceso. Los DW son determinados a partir de un objeto de estudio. El DW debe girar alrededor de este objeto.

	- Uso de esquema snowflake para mejorar el rendimiento de las consultas. - Se agregan datos calculados directo en la tabla de hechos. Como repetición y escalación, promedio de notas, cantidad de cursos con más de 90, cantidad de cursos perdidos. - Se crean múltiples tablas de hechos como estudiantes que incluye la dimensión curso, pero también la sumatoria de cursos perdidos y ganados sin necesidad de calcularlos en tiempo real.

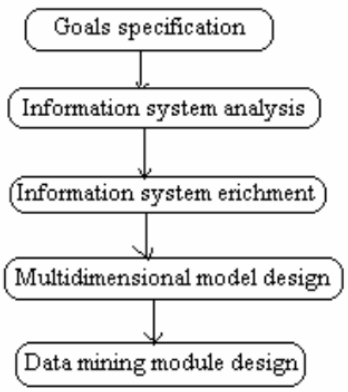
Tabla 5. Ningning 2010.

Fuente: Elaboración propia.

Título original	**A decisional tool for quality improvement in higher education**
Publicación	Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008. 3rd International Conference on , vol., no., pp.1,6, 7-11 April 2008
Autor	Nazih Selmoune, Zaia Alimazighi
Institución que respalda al autor	Laboratoire des Systèmes Informatiques (LSI) USTHB Algiers, Algeria
Resumen	Diseño e implementación de una herramienta para toma decisiones para mejorar los servicios de una institución universitaria sin fines de lucro. Esto apoyado en los depósitos de datos, y reglas de asociación.
Datos relevantes	Uso de toma decisiones donde no se evalúa el rendimiento a partir de ganancias o retribución. Medidas y análisis disponibles en este contexto. ¿Que determina los factores de éxito y fracaso de un estudiante? 5 pasos para el proceso de diseño:

Goals: discusión con los diferentes participantes y determinar los análisis pertinentes.

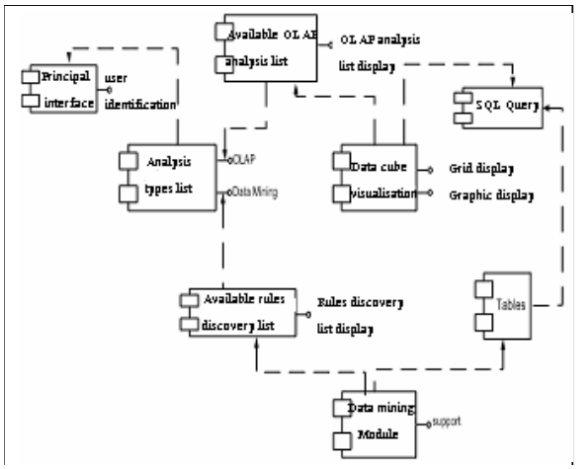
Information analysis: falta información de ausentismo, notas parciales, situación social. Por lo tanto, algunas las metas propuestas no se pueden cumplir por que no hay datos.



Diseño multidimensional: Las medidas son ratio de exito y fracaso. Se crean múltiples tablas de hechos. Fact de exito por examen, Fact de exito por profesor, Fact de ausentismo.

Arquitectura de solución clásica: fuentes->ETL->DW->Cubos y DM.

Diagrama de componentes:



Fue probado en el departamento de Ciencias de la Computación como útil para análisis de los datos existentes.

DM muestra resultados dispares en cursos relacionados.

Tabla 6. Selmoune 2008.

Fuente: Elaboración propia.

Título original	**The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling**
Publicación	John Wiley & Sons. Capítulo 13. Educación. pág 325.
Autor	Ralph Kimball, Margy Ross
Institución que respalda al autor	Kimball Group.
Resumen	Uso de applicant pipeline como un accumulating snapshot fact tables para monitorear el avance de los estudiantes en cada hito de su admisión a carrera. El uso de tablas *factless* con casos de estudio para estas tablas de hecho especiales, y discutir el análisis de eventos que no ocurrieron.
Datos relevantes	En el capítulo: 1. Ejemplo de una matriz de procesos involucrados en una institución de educación superior. 2. Rastreo de los hitos logrados para admisión con tablas de hecho accumulating snapshot. 3. Factless fact table para eventos, administración de los cursos, control de asistencia de los estudiantes a los cursos. 4. Manejo de eventos que no existieron junto a los datos existentes. Complementarlo con los capítulos de recursos humanos (9) y de suministros (5). También manejar los subsidios como un tipo de análisis financiero (10).

Tabla 7. Kimball 2011.

Fuente: Elaboración propia.

5. Resumen de resultados

La lista inicial de fuentes incluye la recomendación incluida en el apartado 3.2.3 del apéndice sobre la revisión de selección. Esto fue una recomendación no obtenida de las búsquedas, sino ingresada de manera heurística a la lista por considerarse fuente primaria.

La columna de resultados indica la cantidad encontrada luego de ingresar la cadena de búsqueda. La tercera columna muestra los resultados que permanecen luego de aplicar los criterios de inclusión. La columna de exclusión, la cuarta, luego de aplicar los criterios, donde se eliminó uno de los resultados. Finalmente, las 6 fuentes primarias, donde se excluyen tres de Scopus, pues eran los mismos artículos encontrados en la IEEE.

Fuente	Resultados	Inclusión	Exclusión	Primarias
IEEE Digital Library	7	4	4	4
Web of Science	4	0	0	0
ACM Digital Library	1	0	0	0
ScienceDirect	5440 (15)	1	0	0
Scopus	6	4	4	1
Recomendaciones	1	1	1	1
Total	**34**	**10**	**9**	**6**

Tabla 8. Resultados de las fuentes consultadas.

Fuente: Elaboración propia.

Nota: Los totales iniciales para la consulta de ScienceDirect con 5440 resultados, se contabilizan como 15 al limitar los resultados por criterio de inclusión.

5.1. Resumen de la situación actual

Primero. (Baranovic, Madunic & Mekterovic, 2003)

Proveer una estructura para análisis de datos por parte de las autoridades.

- Front-end intuitivo con alta disponibilidad para acceso a los reportes de los usuarios.
- 3 categorías de consultas.
- Cubos multidimensionales. SQL.

Segundo. (Hua-long Zhao, 2008)

- Diseño de una estructura del modelo de datos que determina los análisis posibles y la complejidad del mantenimiento.
- Modelo Estrella. Analysis Services. SQL.

Tercero. (Ningning, 2010)

- Importancia del diseño previo del modelo de datos con 5 pasos a seguir.
- El DW debe girar alrededor del objeto de estudio.
- Agregar datos calculados a la tabla de hechos (repetición, promedio de notas, cantidad de cursos con mas de 90, cantidad de cursos perdidos).
- Modelo Snowflake.

Cuarto. (Selmoune & Alimazighi, 2008)

- Uso de toma decisiones donde no se evalúa el rendimiento a partir de ganancias.
- Diseño de 5 pasos.
- Medidas y análisis disponibles en el contexto.

Quinto. (Kimball & Ross, 2011)

- Matriz de procesos.
- Rastreo de hitos logrados (accumulating snapshot).
- Tablas de hechos factless para eventos, no medidas.
- Manejo de eventos que no ocurrieron.

Apéndice 2. Plantilla del modelo

Plantilla del modelo de construcción para el DW

Autor:
Gosvami Zúñiga Chaves

Dirigido a:
Oficina de Planificación Universitaria

Fecha: Agosto, 2016

Introducción

Este documento es un esquema con la lista de los procesos, para cada uno de los grupos definidos en el modelo de construcción del DW. Incluye los documentos de entrada y salida para cada proceso.

El documento incluye:

1. Diagrama de fases y procesos.
2. Entradas y salidas de las fases:
    - Proyecto.
    - Fuentes.
    - Diseño.
    - Visualización.
    - Implementación.

Código	Descripción	Insumos	Fuente	Equipo de BI			Cliente	Salida
			Unidad	Encargado de BI	Desarrollador de BI	Patrocinador	Usuario solicitante	
P-01	Definición de la factibilidad del proyecto	Solicitud de generación del proyecto		●		●	●	Documento del acuerdo para la elaboración del proyecto
P-02	Definición de la arquitectura del proyecto	Documento del acuerdo para la elaboración del proyecto		●	●			Documento de diseño de la arquitectura
P-03	Crear la matriz de procesos para tener una visión macro de la institución	Documento del acuerdo para la elaboración del proyecto, diseño de la arquitectura, políticas y objetivos institucionales, reunión con el patrocinador y el solicitante		●		●	●	Matriz de procesos
P-04	Definición de las fuentes prioritarias por atender	Matriz de procesos, documento con la definición de la arquitectura, reunión del patrocinador y el encargado de BI		●		●	●	Listado de fuentes prioritarias
F-01	Levantamiento de requerimientos a partir de las políticas y objetivos institucionales y necesidades del usuario solicitante	Listado de fuentes, políticas institucionales, reunión con el usuario solicitante, matriz de procesos		●			●	Lista de metas y análisis requeridos
F-02	Revisión de la información disponible para definir la viabilidad de las metas establecidas	Lista de metas y análisis requeridos	●	●			●	Lista de metas y análisis viables, Lista de variables disponibles
F-03	Establecer controles para la calidad de la información	Lista de metas y análisis viables, lista de variables disponibles	●	●	●			Metodología de extracción de datos, documento formal para establecer protocolo, lista de datos de control, procedimiento de aseguramiento de los datos
F-04	Depuración de la información disponible	Lista de metas y análisis viables, variables disponibles, metodología de extracción de datos, documento formal para establecer protocolo, lista de datos de control, procedimiento de aseguramiento de los datos			●			Datos asegurados provenientes de la fuente
D-01	Crear el modelo lógico con las necesidades del negocio y los procesos cruciales	Lista de fuentes, lista de metas y análisis viables, lista de variables disponibles, matriz de procesos	●	●	●	●	●	Modelo lógico, metadatos de las fuentes, aprobación del modelo lógico por parte del patrocinador y del usuario participante
D-02	Crear el modelo dimensional seleccionando el tipo de esquema a utilizar y proponer cómo interactúan las variables de las dimensiones con la tabla de hechos	Modelo lógico, Lista de fuentes, Lista de metas y análisis viables, variables disponibles, matriz de procesos		●	●			Modelo dimensional
D-03	Crear el modelo físico con las estructuras de tablas, nomenclatura, indexado, particionamiento y optimización	Modelo lógico, Modelo dimensional, lista de metas y análisis viables, variables disponibles, datos asegurados de la fuente			●			Modelo físico
D-04	Manejo de metadatos con procesos de desarrollo, métodos de recolección, información sobre la lógica del modelo dimensional, y sobre las estructuras de tablas y relaciones	Modelo lógico, modelo físico, lista de procesos de desarrollo, información sobre las estructuras de la base de datos		●	●			Metadatos recopilados en documentos y en la base de datos
V-01	Establecer los diferentes públicos meta de las variables y análisis, y en que formato y nivel de agregación se van a presentar los datos	Lista de metas y análisis viables, variables disponibles, modelo lógico, modelo físico		●	●		●	Lista de metas, análisis y complejidad de visualización por tipo de usuario
V-02	Validación de los conceptos de visualización propuestos por parte del usuario solicitante	Lista de metas y análisis viables, Lista de metas, análisis y complejidad de visualización por tipo de usuario	●	●		●	●	Valoración de la completitud y correctitud del acceso de datos para cada tipo de usuario y nivel de agregación permitido
I-1	Creación de la estructura para el DW	Modelo físico			●			Estructura del DW para la fuente
I-2	Carga de los datos de la fuente al DW	Estructura del DW para la fuente, Metodología de extracción de datos, documento formal para establecer protocolo, lista de datos de control, procedimiento de aseguramiento de los datos			●			Datos cargados en el DW
I-3	Validar las metas y análisis del usuario contra datos reales obtenidos del DW	Lista de metas, análisis y complejidad de visualización por tipo de usuario, datos cargados al DW, lista de datos de control	●	●	●	●	●	Resultados de consistencia de los datos

Diagrama de fases y procesos del modelo de construcción del DW.

1. Fase de Proyecto

P-01 Factibilidad del Proyecto

Entrada	Solicitud de la alta jerarquía para generar un DW para la institución.
Salida	Documento donde se ratifique la necesidad del DW y se acuerde disponer de recursos e infraestructura y la disponibilidad de los datos por las diferentes fuentes. Este documento puede estar supeditado a políticas institucionales.

P-02 Arquitectura del Proyecto

Entrada	Documento del acuerdo para la elaboración del proyecto.
Salida	Diseño de la arquitectura.

P-03 Matriz de Procesos

Entrada	Documento del acuerdo para la elaboración del proyecto. Diseño de la arquitectura. Políticas y objetivos institucionales. Reunión con el patrocinador y el solicitante.
Salida	Matriz de procesos.

P-04 Fuentes de Datos

Entrada	Matriz de procesos. Diseño de la arquitectura. Reunión del patrocinador y el encargado de BI.
Salida	Lista de fuentes según prioridad.

2. Fase de Fuente de Datos

F-01 Levantar requerimientos

Entrada	Lista de fuentes según prioridad. Políticas y objetivos institucionales. Reunión con el usuario solicitante. Matriz de procesos.
Salida	Lista de metas y análisis requeridos.

F-02 Revisión de la información disponible en la fuente

Entrada	Lista de metas y análisis requeridos.
Salida	Lista de metas y análisis viables. Lista de variables disponibles.

F-03 Establecer controles para la calidad de la información

Entrada	Lista de metas y análisis viables. Lista de variables disponibles.
Salida	Metodología de extracción de datos. Documento formal para establecer protocolo. Lista de datos de control. Procedimiento de aseguramiento de los datos.

F-04 Depuración de la información disponible

Entrada	Lista de metas y análisis viables. Variables disponibles. Metodología de extracción de datos. Documento formal para establecer protocolo. Lista de datos de control. Procedimiento de aseguramiento de los datos.

Salida	Datos asegurados provenientes de la fuente.

3. Fase de Diseño

D-01 Modelo lógico

Entrada	Lista de fuentes según prioridad. Lista de metas y análisis viables. Lista de variables disponibles. Matriz de procesos.
Salida	Modelo lógico. Metadatos de las fuentes. Aprobación del modelo lógico por parte del patrocinador y del usuario participante.

D-02 Modelo dimensional

Entrada	Modelo lógico. Lista de fuentes según prioridad. Lista de metas y análisis viables. Variables disponibles. Matriz de procesos.
Salida	Modelo dimensional.

D-03 Modelo físico

Entrada	Modelo lógico. Modelo dimensional. Lista de metas y análisis viables. Lista de variables disponibles. Datos asegurados de la fuente.

Salida	Modelo físico.

D-04 Metadatos

Entrada	Modelo lógico. Modelo dimensional. Modelo físico. Diseño de la arquitectura. Matriz de procesos. Lista de metas y análisis viables. Protocolos de extracción de datos.
Salida	Lista de procesos de desarrollo. Lista de objetos de la base de datos.

4. Fase de Visualización

V-01 Público meta

Entrada	Lista de metas y análisis viables. Lista de variables disponibles. Modelo lógico. Modelo físico.
Salida	Lista de metas, análisis y complejidad de visualización por tipo de usuario.

V-02 Validar conceptos de visualización

Entrada	Lista de metas y análisis viables. Lista de metas, análisis y complejidad de visualización por tipo de usuario.
Salida	Valoración del acceso de datos para cada tipo de usuario y nivel de agregación permitido.

5. Fase de Implementación

I-01 Estructura del DW

Entrada	Modelo físico.
Salida	Estructura del DW para la fuente. Script de creación de la estructura.

I-02 Carga de los datos

Entrada	Estructura del DW para la fuente. Metodología de extracción de datos. Documento formal para establecer protocolo. Lista de datos de control. Procedimiento de aseguramiento de los datos.
Salida	Datos cargados en el DW.

I-03 Validar datos del DW

Entrada	Lista de metas, análisis y complejidad de visualización por tipo de usuario. Datos cargados al DW. Lista de datos de control.
Salida	Resultados de consistencia de los datos. Aprobación del patrocinador y del usuario.

Apéndice 3. Modelo aplicado

DW para la Universidad de Costa Rica
Enfocado al Sistema de Información Estudiantil

Autor:
Gosvami Zúñiga Chaves

Dirigido a:
Oficina de Planificación Universitaria

Fecha: Agosto, 2016

Índice

INTRODUCCIÓN 133
1. PROYECTO 134
2. FUENTE DE DATOS 144
3. DISEÑO 165
4. VISUALIZACIÓN 180
5. IMPLEMENTACIÓN 186

Introducción

Este documento desarrolla una primera iteración del modelo para la obtención de un DW, enfocado en extraer la información acerca de los estudiantes disponible en la Oficina de Registro de la Universidad de Costa Rica. Se ejecutan las cinco fases de acuerdo al uso del modelo.

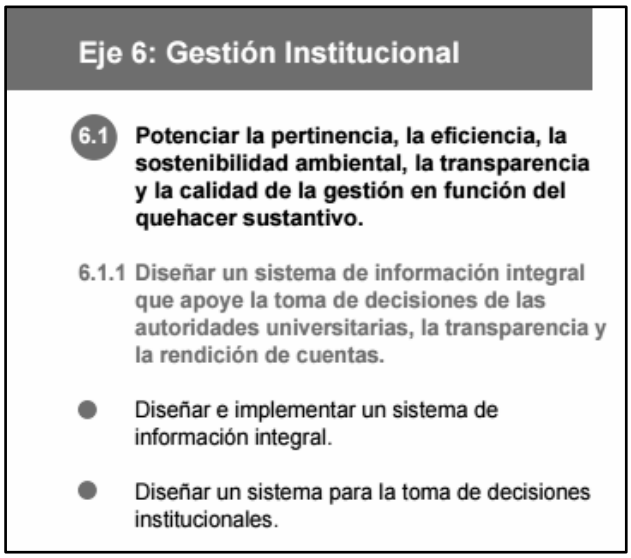
Los procesos en cada uno de los puntos se muestran con el mínimo de documentación solicitada a modo de ejemplo, y como insumo para establecer la pertinencia del modelo. La documentación puede ser ampliada de acuerdo a los requerimientos de la institución.

1. Proyecto

P-01 Factibilidad del Proyecto

Documento: Ratificación del proyecto

El proyecto del DW surge de la solicitud del Consejo Universitario por recopilar datos y brindar información centralizada para la toma de decisiones.



La viabilidad se refleja en el Plan Estratégico Institucional 2013-2017 en el eje sobre gestión institucional, en la política institucional 6.1.1., que manifiesta la necesidad de “Diseñar un sistema de información integral que apoye la toma de decisiones de las autoridades universitarias, la transparencia y la rendición de cuentas.”

La jefatura de la Oficina de Planificación patrocina el proyecto y dispone los recursos necesarios para su consecución, delegado por la Rectoría para su realización.

> **ARTÍCULO 9.** Todas las unidades deben proporcionar la información que la Oficina de Planificación requiera para cumplir con sus funciones.

La disponibilidad de los datos se da por mandato del Consejo Universitario descrito en el capítulo 2 del artículo 9 del reglamento de la Oficina de Planificación Universitaria -aprobado en sesión 3137-07 del 31 de octubre de 1984, y publicado en el Alcance a La Gaceta Universitaria 15-84, del 20 de noviembre de 1984- que dicta que la Oplau puede solicitar datos para realizar los estudios requeridos de acuerdo a las necesidades institucionales.

Además, las dependencias firmaron un acuerdo para establecer el cronograma de trabajo y la periodicidad en la entrega de los datos.

UNIVERSIDAD DE COSTA RICA

OPLAU
Oficina de Planificación

21 de marzo de 2012
OPLAU-113-2012

OFICINA DE REGISTRO E INFORMACION U.C.R.
21 MAR 2012
RECIBIDO

Oficina de Registro e Información

Estimado señor:

En la reunión del día jueves 8 de marzo del presente año, se llegaron a los siguientes acuerdos, para el envío de información por parte de su oficina a la OPLAU:

1. Archivos de matrícula y de notas en formato txt, según las siguientes fechas:

Archivo de Matricula

Ciclo lectivo	III ciclo año anterior	I ciclo año actual	II ciclo año actual
Fecha de entrega	10 de julio	10 de julio	1 de diciembre

Archivo de notas

Ciclo lectivo	III ciclo año anterior	I ciclo año actual	II ciclo año actual
Fecha de entrega	30 de agosto	30 de agosto	30 de enero

Las variables que debe contener cada uno de los archivos mencionados anteriormente son las siguientes:

- carne (carne del estudiante) que matriculo el curso)

Carta de entrega de los datos con el detalle y periodicidad requeridos.

Fuente. Oficina de Planificación, UCR.

P-02 Arquitectura del Proyecto

Documento: Diseño de la arquitectura

Tipo de arquitectura

La arquitectura será de 3 capas, capa de datos, de aplicación y de presentación. Figura 1.

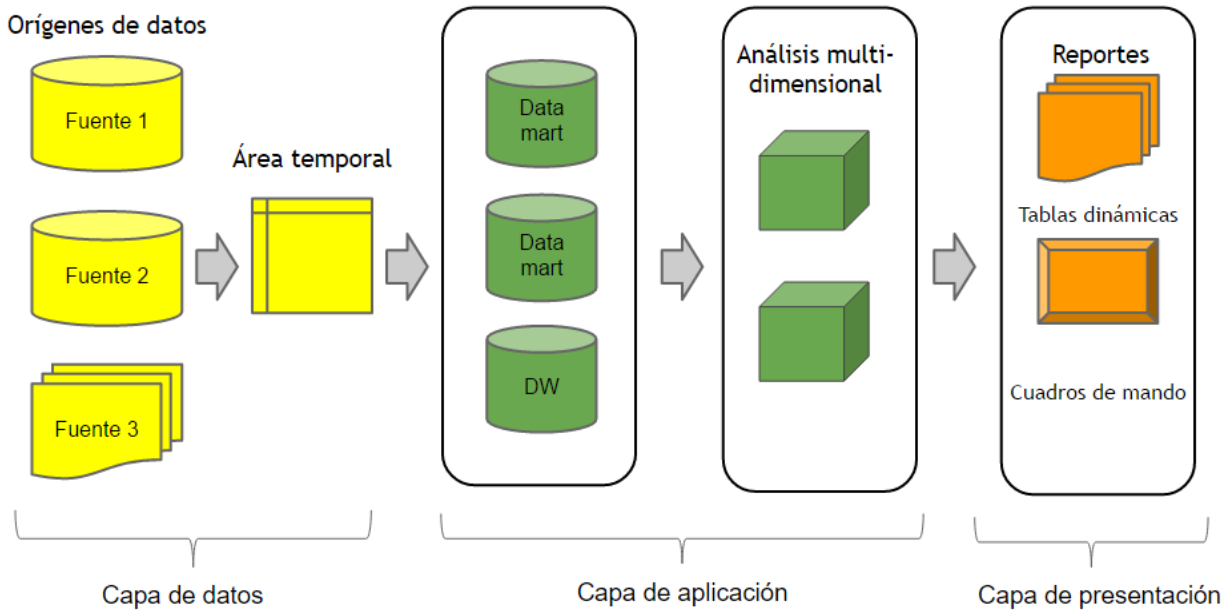


Figura 1. Arquitectura de un DW en 3 capas.

Fuente: Elaboración propia.

El detalle de cada capa en la arquitectura definida:

Capa de datos

- Donde residen las bases de datos fuentes.
- Procesos de ETL.
- Se ubica el SA para un intercambio rápido entre la fuente de datos y nuestro espacio de almacenamiento.

Capa de aplicación

- Los DM.
- El DW.
- Las bases multidimensionales.
- Los cubos OLAP.
- La seguridad de acceso a nivel granular y de dimensiones.

Capa de presentación

- Las aplicaciones de acceso a los cubos.
- Generación de reportes.
- Cuadros de mando.
- Administración de los datos de alto nivel.

Tipo de implementación

La implementación será de abajo hacia arriba (bottom-up), lo que implica la creación de DMs que se integran paulatinamente en un solo conjunto de datos relacionados a partir de las dimensiones comunes.

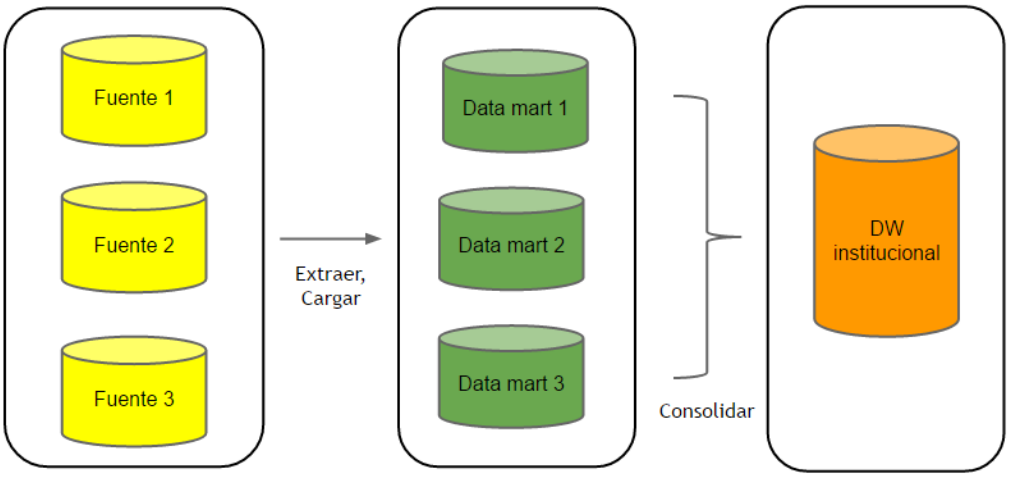


Figura 2. Implementación de abajo hacia arriba.

Fuente: Elaboración propia.

Arquitectura de despliegue

La distribución de los diferentes nodos para la implementación se muestra en el esquema de la figura 3. Los objetos dentro del esquema representan los elementos esperados para el despliegue de la solución.

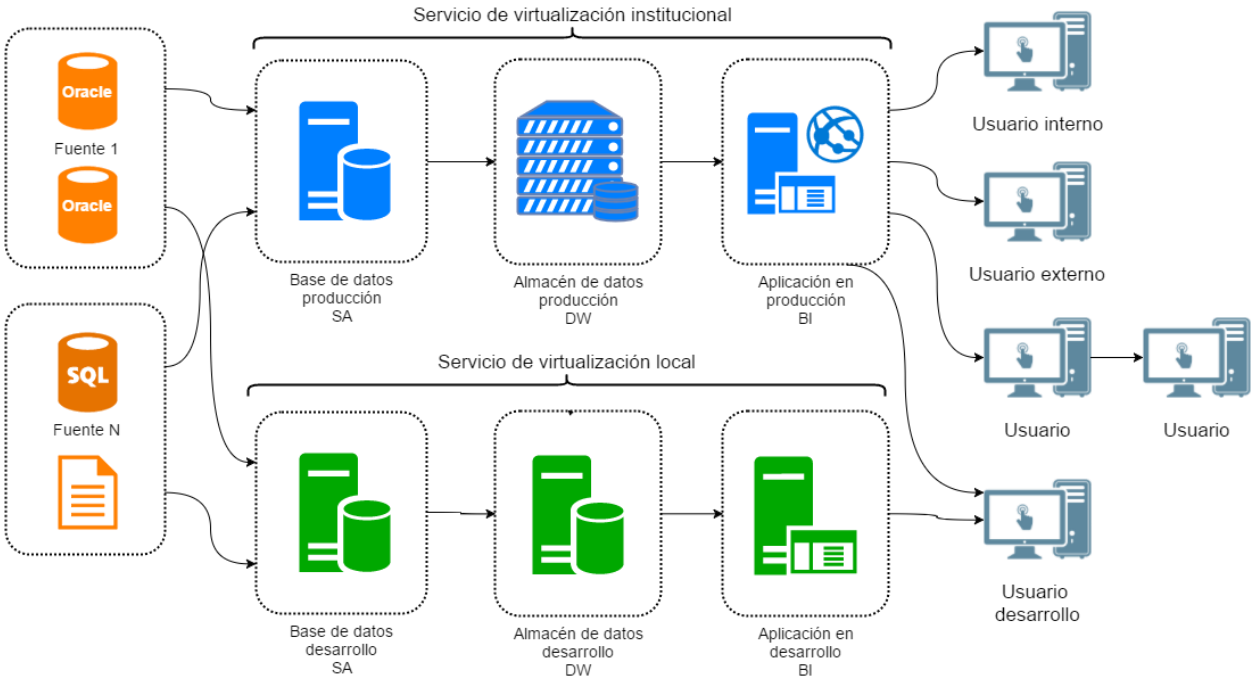


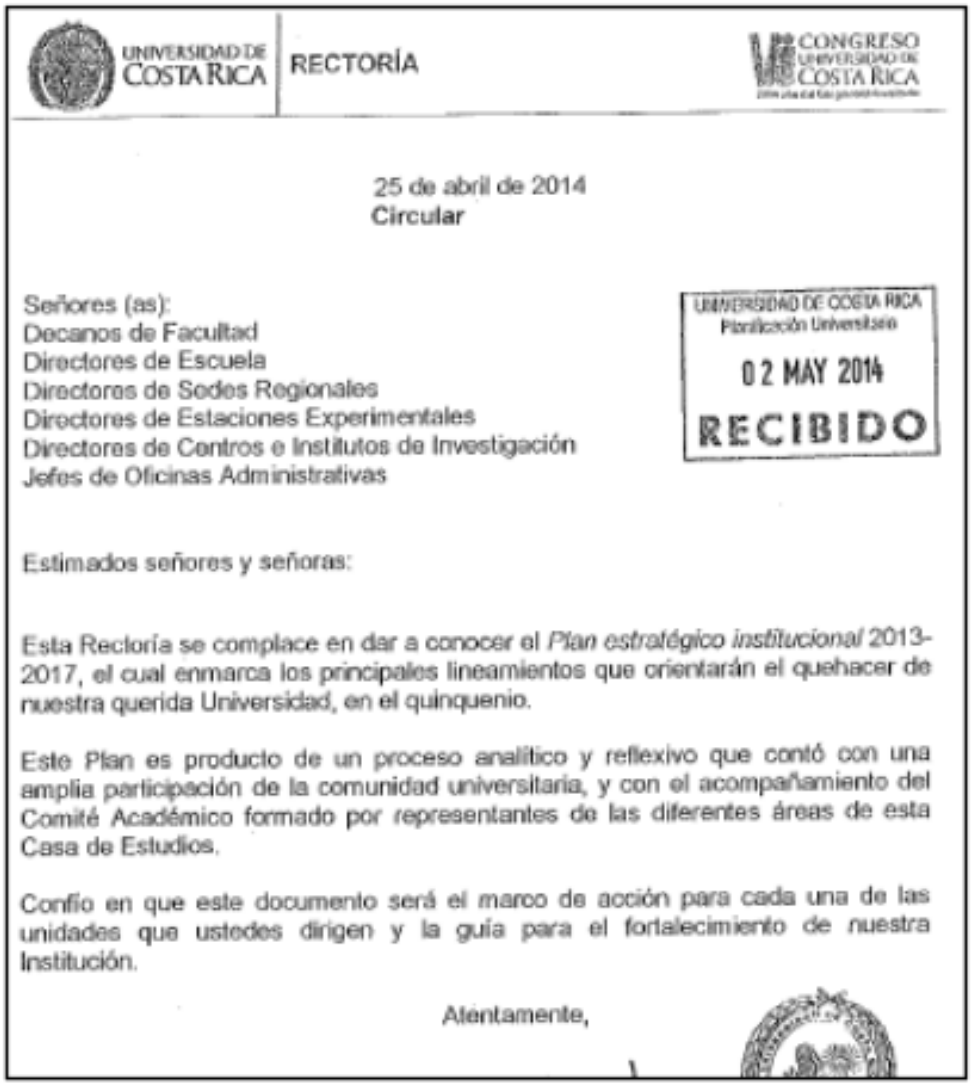
Figura 3. Arquitectura de despliegue.

Fuente: Elaboración propia.

La arquitectura consta de fuentes heterogéneas, con un ambiente de desarrollo local, y un ambiente de producción. El ambiente de producción y el de desarrollo aprovechan servicios virtualizados para separar las funciones, y así brindar un crecimiento de acuerdo a las necesidades del sistema. Las extracción y preparación de las fuentes se da en el ambiente de desarrollo, para luego aplicar los ETL depurados en el ambiente de producción. Los ambientes no poseen las mismas características de hardware, pero si utilizan las mismas versiones de los sistemas operativos y de las aplicaciones para disminuir los problemas de incompatibilidad.

Documento: Políticas y objetivos institucionales

El plan estratégico fue aprobado por la Rectoría como hace constar el documento. Disponible en línea: http://www.oplau.ucr.ac.cr/index.php/plan-estrategico-institucional



UNIVERSIDAD DE COSTA RICA | RECTORÍA

CONGRESO UNIVERSIDAD DE COSTA RICA

25 de abril de 2014
Circular

UNIVERSIDAD DE COSTA RICA
Planificación Universitaria
02 MAY 2014
RECIBIDO

Señores (as):
Decanos de Facultad
Directores de Escuela
Directores de Sedes Regionales
Directores de Estaciones Experimentales
Directores de Centros e Institutos de Investigación
Jefes de Oficinas Administrativas

Estimados señores y señoras:

Esta Rectoría se complace en dar a conocer el *Plan estratégico institucional* 2013-2017, el cual enmarca los principales lineamientos que orientarán el quehacer de nuestra querida Universidad, en el quinquenio.

Este Plan es producto de un proceso analítico y reflexivo que contó con una amplia participación de la comunidad universitaria, y con el acompañamiento del Comité Académico formado por representantes de las diferentes áreas de esta Casa de Estudios.

Confío en que este documento será el marco de acción para cada una de las unidades que ustedes dirigen y la guía para el fortalecimiento de nuestra Institución.

Atentamente,

Documento de aprobación del Plan Estratégico Institucional 2013-2017.
Fuente. Oficina de Planificación, UCR.

Documento: Matriz de procesos

La matriz se realiza a partir de los procesos claves que se ejecutan en la administración y en el plan estratégico institucional 2013-2017, el cual menciona los objetivos estratégicos institucionales para el quinquenio. De las acciones del plan, se extrae la tabla 1 con las fuentes de datos.

Fuentes mencionadas	
CEA	Centro de Evaluación Académica
OAF	Oficina de Asuntos Financieros
OAICE	Oficina de Asuntos Internacionales y Cooperación Externa
OBAS	Oficina de Becas y Asuntos Socioeconómicos
ORH	Oficina de Recursos Humanos
ORI	Oficina de Registro e Información
OSUM	Oficina de Suministros
VAS	Vicerrectoría de Acción Social
VD	Vicerrectoría de Docencia
VI	Vicerrectoría de Investigación
Vicerrectorías	Investigación, Acción Social, Vida Estudiantil, Docencia y Administración

Tabla 1. Fuentes mencionadas en la matriz de procesos.

Fuente: Oficina de Planificación Universitaria, UCR.

La matriz de la tabla 2 incluye los procesos, dimensiones y fuentes de los datos. A continuación, se muestra la matriz obtenida.

			Fuentes	ORI ORH OAF	ORI	ORI	ORI ORH CEA	ORI	ORI VD	ORI	ORI	CEA
	Ejes	Dimensiones ? Procesos ?	Fecha	Unidad	Carrera	Plan de estudio	Nivel académico	Estudiante	Curso	Ubicación geográfica	Colegio	Acreditación
1	Oferta académica de grado y posgrado	Actualización de planes de estudio	x	x	x	x	x					
2		Carreras innovadoras	x	x	x		x					
3		Acreditación académica	x	x	x		x					x
4	Formación del personal académico	Formación con alto nivel para relevo generacional	x	x			x					
5		Programa de actualización del personal docente	x	x			x					
6		Evaluación docente	x	x			x					
7		Estabilidad laboral docente	x	x			x					
8	Generación de conocimiento	Actividades de generación del conocimiento	x	x			x					
9		Proyectos y actividades interdisciplinarias	x	x								
10		Redes académicas nacionales e internacionales	x	x								
11		Trabajos finales de graduación vinculados a centros de investigación	x	x				x				
12	Vinculación con la sociedad	Espacios de intercambio universidad-sociedad	x	x					x			
13		Difusión del conocimiento	x	x								
14		Integración con las necesidades de desarrollo del país	x	x						x		
15		Emprendedurismo e innovación de proyectos hacia la sociedad	x	x								
16	Posicionar a la Universidad en el escenario nacional e internacional	Redes académicas nacionales e internacionales	x	x	x		x	x		x		
17		Movilidad docente, estudiantil y administrativo	x	x			x	x		x		
18	Equidad en el ingreso y permanencia estudiantil	Equidad de ingreso	x	x	x			x		x	x	
19		Movilidad estudiantil internacional	x	x			x	x				
20		**Permanencia y culminación de estudios**	x	x	x		x	x				
21		Capacidad de admisión a carrera	x	x	x		x	x	x			
22	Pertinencia, eficiencia, sostenibilidad, transparencia y calidad de la gestión	Infraestructura física y tecnológica	x	x								
23		Modelo de gestión del recurso humano	x	x								
25		Gestión ambiental	x	x								
26		Modelo de gestión del riesgo en desastres y atención de emergencias	x	x								
27		Cultura de planificación en los procesos	x	x								

Tabla 2. Matriz de procesos.

Fuente: Oficina de Planificación Universitaria, UCR.

P-04 Fuentes de Datos

Documento: Listado de fuentes por prioridad

La lista de once fuentes extraídas de la matriz de procesos se prioriza de acuerdo a factores como:

- Mayor cantidad de análisis estratégicos de una misma fuente.
- Importancia a nivel institucional de la fuente.
- Madurez de la fuente con respecto a la sistematización de sus procesos internos.
- Nivel de confianza con las jefaturas para facilitar la comunicación ante nuevos requerimientos.

A partir de la discusión a lo interno de la relevancia de cada fuente y de los factores clave, se finaliza con la tabla 3 con la priorización de las fuentes.

1	ORI	Oficina de Registro e Información	La Oficina de Registro está relacionada con 8 de las 18 dimensiones en la matriz, además, es la que resguarda la información de los estudiantes, pilar fundamental de la Universidad. Por otra parte, la información está sistematizada por medio de los sistemas de matrícula, de admisión, expediente académico y el Sistema de Aplicaciones Estudiantiles (SAE). La sistematización facilita automatizar la extracción de los datos, y, por ende, su revisión. Además, la ORI, por la naturaleza de sus funciones, recibe múltiples solicitudes de información que podrían ser solventadas con la construcción de un DW, facilitado por las buenas relaciones con las autoridades de la ORI. Esta revisión de factores permite establecer como primera fuente, la ORI.
2	OAF	Oficina de Asuntos Financieros	En segundo lugar, el presupuesto universitario, al ser suministrado por el Fondo Especial para la Educación Superior, requiere un especial cuidado y se somete de

			manera constante al escrutinio de las autoridades universitarias, así como entes gubernamentales y sociales. Es por su relevancia, que se opta por ser la siguiente en prioridad.
3	ORH	Oficina de Recursos Humanos	El recurso humano, ya sea docente o administrativo, es el rubro más importante de la Universidad en materia presupuestaria y, por tanto, debe ser manejado de manera estratégica para lograr calidad con equidad.
4	OBAS	Oficina de Becas y Asuntos Socioeconómicos	Las becas estudiantiles son las que permiten la inclusión de sectores vulnerables a la educación superior, y es un rubro solicitado de manera recurrente por las autoridades para fines estratégicos y de transparencia.
5	CEA	Centro de Evaluación Académica	Es la que detalla la situación de los docentes y de su nivel académico, además de corroborar las jornadas y el tipo de trabajo que desempeñan.
6	OAICE	Oficina de Asuntos Internacionales y Cooperación Externa	Controla la movilidad estudiantil, docente y administrativa.
7	VI	Vicerrectoría de Investigación	Cuenta con la mayor cantidad de proyectos y publicaciones en la Universidad.
8	VAS	Vicerrectoría de Acción Social	Los proyectos son la vinculación con la sociedad y como la universidad le retribuye.
9	VD	Vicerrectoría de Docencia	Posee pocos proyectos inscritos.
10	VVE	Vicerrectoría de Vida Estudiantil	Posee pocos proyectos inscritos.
11	VRA	Vicerrectoría de Administración	Es una dependencia de apoyo que tiene a cargo otras entidades que ya fueron revisadas, como ORH, OAF y OSUM.

Tabla 3. Lista de fuentes por prioridad.

Fuente: Elaboración propia.

2. Fuente de Datos

F-01 Levantar requerimientos

Documento: Bitácora de reuniones

En la bitácora se incluyen reuniones relacionadas con la obtención de los datos desde la fuente ORI, con acuerdos logrados en diferentes momentos y avances en las relaciones, que finalizan con la obtención de datos de manera sistemática. Algunos acuerdos son de años atrás, pues la solicitud estaba relacionada con otro proceso, pero sirvieron para facilitar los datos.

8 de marzo de 2012

Información generada.

- Admisión:
    - Admisión Estudiantes Primer Ingreso por PAA
    - Traslado de carrera
    - Primer ingreso por otra modalidad (convenios – SEP)
    - Ingreso a programas específicos

Si un estudiante es admitido – matrícula – retira se considera de nuevo como NO CONSOLIDADO (¿en qué momento se podría tener el total de matriculados descontando los que llegan a ser no consolidados?)

Admisión por tipo de colegio: la BD es transaccional por lo que se pueden generar inconsistencias en el tiempo porque el SAE no guarda históricos.

MATRÍCULA - Se definieron 29 tipos distintos de datos estadísticos (debería especificarse el momento en que se genera la información –fecha y hora-):

1. Estudiantes en prematrícula por recinto – contando estudiante 1 vez.
2. Estudiantes en prematrícula por recinto y carrera – contando estudiante 1 vez.
3. Estudiantes en prematrícula por recinto – contando todas las carreras en que tiene

proyecto el estudiante.

4. Estudiantes en prematrícula por recinto y carrera – contando todas las carreras en que tiene proyecto el estudiante.
5. Estudiantes matriculados por recinto – contando estudiante 1 vez.
6. Estudiantes en matriculados por recinto y carrera – contando estudiante 1 vez.
7. Estudiantes en matriculados por recinto – contando todas las carreras en que matriculó el estudiante.
8. Estudiantes en matriculados por recinto y carrera – contando todas las carreras en que matriculó el estudiante.
9. Cursos ofrecidos en matrícula ordinaria por recinto.
    a. Matrícula ordinaria Web (inclusiones se incorporan después)
    b. Medicina (mes antes)
    c. Suma total de a y b
    d. Programas cuatrimestrales del SEP.
10. Cursos grupos ofrecidos en matrícula ordinaria por recinto.
11. Cursos grupos ofrecidos en matrícula ordinaria + medicina por recinto
12. Cursos grupos ofrecidos en matrícula ordinaria + medicina + inclusiones por recinto
13. Cupos ofrecidos en matrícula ordinaria por recinto y unidad académica
14. Cupos ofrecidos en matrícula ordinaria + Medicina por recinto y unidad académica
15. Cupos ofrecidos en matrícula ordinaria + Medicina + Inclusiones por recinto y unidad académica
16. Estudiantes con inclusión sin Medicina por recinto y carrera – contando al estudiante 1 vez
17. Cupos ofrecidos en inclusión por recinto y carrera
18. Estudiantes con cita sin matrícula ordinaria por recinto
19. Estudiantes con cita, sin matrícula ordinaria y no ingresaron al sistema por recinto
20. Estudiantes con cita, sin matrícula ordinaria, ingresaron al sistema e indicaron demanda insatisfecha por recinto
21. Estudiantes con cita, sin matrícula ordinaria, ingresaron al sistema y no indicaron demanda insatisfecha por recinto
22. Oferta vs demanda por recinto
23. Oferta vs demanda por unidad académica
24. Cursos-grupos con renuncia

25. Estudiantes con renuncia
26. Cupos matriculados en matrícula ordinaria por recinto y unidad académica
27. Cupos sobrantes de matrícula ordinaria por recinto y UA
28. Estadística por Sexo – Estudiantes matriculados por recinto, por carrera base
29. Estadística por Sexo – Estudiantes matriculados por recinto y carrera, por carrera base

EL momento más exacto para todos estos datos es el final del ciclo lectivo.

- Permanencia:
    - Actualización de información del estudiante (e-matrícula)
    - Matrícula
    - Reconocimientos
    - Expediente
- Graduación:
    - Ordinaria
    - Extraordinaria (procesos de graduación diferentes a actos de graduación -hay más actos que procesos-)
    - Honor (Recinto – carrera – género)
    - Especial ante Rectoría

3 de agosto de 2015

Resultados:

- Distribución de matrícula y de notas.
- Se determina la existencia de una variable no identificada previamente, la modalidad de estudio.
- Se establecen las diferencias en matrícula con retiros y en cuáles casos se aplican las diferencias.
- En notas se identifican las siglas y número de grupo de cursos de equivalencia, de verano y de trabajo comunal.
- Se menciona la factibilidad de modificar la entrega de los datos en CSV y realizar una conexión directa por medio de un webservice o vistas.
- Obtener los datos de notas del 2000 al 2015.

- Definir la condición del curso, y la llave primaria de matrícula. Año, período, recinto y sigla de matrícula del curso, grupo, modalidad de estudio, carné y secuencia (un valor automático donde el mayor es el último ingreso).

5 de octubre de 2015

Matrícula, notas, graduación y primer ingreso.

Resultados:

- Vista de promedio ponderado por estudiante para notas.
- Datos de primer ingreso por estado del estudiante, solo tipo 4, que son los que completaron todo el proceso y fueron admitidos a carrera.
- Vista de colegio, con modalidad (técnico, científico, regular,...) y por tipo (privado, público,...).
- Carreras históricas no son fidedignas, se deben conservar con los nuevos períodos.
- Listados de carreras, área, facultad, escuelas y recintos.

15 de octubre de 2015

Estudiantes graduados. Datos antes del 1992 no son válidos. Del 1995 tienen algunos faltantes. En general, la tabla histórica no es 100% fiable.

9 de diciembre de 2015

- El indicador asignado contiene valores ‘A’, ‘O’ y ‘R’. El valor ‘A’ son los asignados, ‘O’ los retiros, IT, etc. Si se requiere solo la matrícula consolidada, se filtra por ‘A’, si se quieren todos son los ‘A’ y los ‘O’.
- Los datos de notas se remontan al año 1972, la inclusión de nueva información es constante, pero lenta, pues es un proceso manual con los datos que se tienen en papel.

10 de marzo de 2016

Consejo de Rectoría.

Se solicita la inclusión de rangos de edad, y de tiempo del estudiante en carrera.

La información de estudiantes de tiempo completo por 16 créditos.

La matrícula histórica debe incluir un resumen institucional.

Documento: Lista de variables y análisis requeridos

Luego de reuniones con el usuario solicitante, se establece una lista de variables y análisis requeridos cuya fuente es la ORI. La tabla 4 indica las variables, la unidad de medida a utilizar, y la granularidad y detalle necesario para cada una.

	Variables y análisis	**Unidad de medida**	**Granularidad**
1	Créditos (aprobados, matriculados y retirados)	Creditos	Matriculados, aprobados, por escuela
2	Diplomas otorgados	Diplomas otorgados	Nivel académico, área, por escuela
3	Estudiantes admitidos	Estudiantes admitidos	Tipo de colegio y por carrera
4	Estudiantes de atención especial	Estudiantes de atención especial	Escuela
5	Estudiantes admitidos de primer ingreso	No. de estudiantes	Año, zona geográfica de procedencia, por tipo de colegio
6	Estudiantes físicos graduados	Estudiantes físicos	Nivel académico, sexo, por escuela y carrera
7	Estudiantes físicos matriculados	Estudiantes físicos	Nivel académico, sexo, especialidad, edad, estado civil, nacionalidad, por escuela y carrera
8	Estudiantes matriculados de primer ingreso	Estudiantes físicos	Nivel académico, sexo,beca vigente, tipo de beca, edad, por escuela y carrera
9	Grupos	Grupos	Nivel académico, por escuela
10	Cursos (aprobados y matriculados)	Cursos	Aprobados y matriculados,

			especialidad, ciclo, por escuela.
11	Inscritos en el proceso de admisión	Participantes	Inscrito, realizado, elegible, concursante, matriculado y no consolidado
12	Solicitudes de ingreso a carrera	Solicitudes	Solicitudes a carrrera, las 3 opciones posibles
13	Cupos matriculados	Cupos	Grado académico por escuela
14	Estudiantes equivalentes de tiempo completo	Créditos	Grado académico por escuela
15	Tasa rendimiento académico	Tasa rendimiento	Institucional

Tabla 4. Variables y análisis requeridos.

Fuente: Estudio de variables[14], Oficina de Planificación Universitaria, UCR.

[14] Documento interno de la Oplau no publicado.

F-02 Revisión de la información disponible en la fuente

Documento: Lista de variables y análisis viables

La lista de variables se revisa contra los datos disponibles en la fuente en la tabla 5. Este procedimiento se realiza a partir de reuniones con la persona responsable en la fuente de datos y una revisión documental de los diccionarios de datos existentes[15].

	Variables y análisis	**Unidad de medida**	**Granularidad**	**Viable**
1	Créditos (aprobados, matriculados y retirados)	Creditos	Matriculados, aprobados, por escuela	**Sí**
2	Diplomas otorgados	Diplomas otorgados	Nivel académico, área, por escuela	**Sí**
3	Estudiantes admitidos de primer ingreso.	Número de estudiantes	Tipo de colegio, zona geográfica de procedencia y por carrera	**Sí**
4	Estudiantes de atención especial	Estudiantes de atención especial	Escuela	No se dispone del cálculo por el momento.
5	Estudiantes físicos graduados	Estudiantes físicos	Nivel académico, sexo, por escuela y carrera	**Sí**
6	Estudiantes físicos matriculados	Estudiantes físicos	Nivel académico, sexo, especialidad, edad, estado civil, nacionalidad, por escuela y carrera	**Sí**
7	Estudiantes matriculados de primer ingreso	Estudiantes físicos	Nivel académico, sexo,beca vigente, tipo de beca, edad,	Parcial. Sin datos de becas que se obtiene

[15] No se incluyen los diccionarios de datos en este trabajo a solicitud de la fuente.

			por escuela y carrera	de otra fuente de datos.
8	Grupos	Grupos	Nivel académico, por escuela	**Sí.**
9	Cursos (aprobados y matriculados)	Cursos	Aprobados y matriculados, especialidad, ciclo, por escuela.	**Sí.**
10	Inscritos en el proceso de admisión	Participantes	Inscrito, realizado, elegible, concursante, matriculado y no consolidado	No disponible por el momento.
11	Solicitudes de ingreso a carrera	Solicitudes	Solicitudes a carrrera, las 3 opciones posibles	No disponible por el momento.
12	Cupos matriculados	Cupos	Grado académico por escuela	**Sí.**
13	Estudiantes equivalentes de tiempo completo	Créditos	Grado académico por escuela	**Sí.** Es un campo calculado.
14	Tasa rendimiento académico	Tasa rendimiento	Institucional	Parcial. Se debe establecer la metodología de cálculo.

Tabla 5. Variables y análisis viables.

Fuente: Elaboración propia a partir del estudio de variables[16], Oficina de Planificación Universitaria, UCR.

[16] Documento interno de la Oplau no publicado.

Documento: Lista de variables disponibles

En la revisión de la fuente de datos, se genera la tabla 5 con las variables disponibles a partir de reuniones con la persona responsable en la fuente de datos y una revisión documental de los diccionarios de datos existentes[17]. Los datos extraídos son un insumo para determinar el contenido de las dimensiones y de las tablas de hecho para esta fuente de información mostrado en la tabla 6.

Variables	**Tipo de dato**	**Descripción de variable**	**Histórico disponible**	**Observaciones**
Carrera	Cadena	Carrera universitaria.	Para todo el registro existente de datos.	Sigla y nombre.
Ciclo o periodo	Entero	Los ciclos comprenden Primer semestre de marzo a julio. Segundo semestre de agosto a diciembre. Verano de enero a febrero.	No aplica.	Medida de tiempo.
Año	Entero	Año matriculado, graduado o admitido.	No aplica.	Medida de tiempo.
Carné	Cadena(8)	Identificación única del estudiante luego de consolidar matrícula. El carné permanece no importa el tiempo de ausencia de la Universidad.	Para todo el registro existente de datos.	No se solicita el nombre del estudiante.
Curso	Cadena(6)	El curso matriculado o cursado durante un periodo.	Para todo el registro existente de datos.	No se solicita el nombre del curso.
Número de grupo	Entero	El número de grupo para el curso.	Para todo el registro existente de datos.	

[17] No se incluyen los diccionarios de datos en este trabajo a solicitud de la fuente.

Recinto	Cadena	El recinto donde el estudiante está inscrito, o que pertenece el curso.	Para todo el registro existente de datos.	Recinto y escuela son dependientes.
Créditos	Entero	Valor asignado a cada curso de acuerdo al tiempo estimado de dedicación del estudiante.	Para todo el registro existente de datos.	Es una medida. Esta relacionada con curso.
Escuela	Cadena	La escuela donde el estudiante está inscrito, o que pertenece el curso.	Para todo el registro existente de datos.	Recinto y escuela son dependientes.
Año de ingreso del estudiante a la universidad	Entero	Año de consolidación de matrícula una vez admitido en la universidad.	Para todo el registro existente de datos.	
Sexo del estudiante	Entero	Sexo o género del estudiante. Masculino o femenino.	Para todo el registro existente de datos.	
Modalidad de estudio	Cadena	Si el curso matriculado o cursado es regular, por tutoría o por suficiencia.	Para todo el registro existente de datos.	Solo aplica en matrícula y en notas.
Tipo de estudiante	Cadena	Si se considera un estudiante de pregrado, posgrado, visitante, programas especiales o de extensión. El valor no es histórico, se actualiza con cada cambio de estado del estudiante.	Para todo el registro existente de datos.	Solo aplica en matrícula y en notas.
Nota alfa-numérica	Cadena	Nota obtenida por el estudiante en el curso. Con valores numéricos del 0 al 10 en incrementos de 0.5 y con valores alfanuméricos como	No aplica.	Es una medida en notas.

		AP=aprobado, MA=matriculado, PE=perdido, RJ=retiro justificado.		
Nota numérica	Número con 1 decimal	Nota obtenida por el estudiante en el curso. Con valores numéricos del 0 al 10 en incrementos de 0.5.	No aplica.	Es una medida en notas.
Área	Cadena	El área de estudio a la que pertenece una escuela.	Para todo el registro existente de datos.	
Facultad	Cadena	Engloba a escuelas regidas por un mismo principio de una misma área.	Para todo el registro existente de datos.	
Título	Cadena	Certificado obtenido al graduarse de una carrera universitaria.	Para todo el registro existente de datos.	
Fecha de juramenta-ción	Fecha	La fecha en la que ejecuta la ceremonia de entrega del título. Es la fecha que se considera para el cálculo de graduados por año.	No aplica.	

Tabla 6. Variables disponibles en la fuente de datos ORI.
Fuente: Elaboración propia.

La tabla 6 aplica para la información de matriculados, de notas y de graduados. La disponibilidad de información histórica para cada rubro se muestra en la tabla 7.

	Históricos disponibles
Matriculados	2008 en adelante.
Notas	1990 en adelante (se puede consultar información anterior si fuera necesario)
Graduados	1990 en adelante (se puede consultar información anterior si fuera necesario)
Admitidos	2008 en adelante.

Tabla 7. Información histórica de hechos.
Fuente: Oficina de Registro e Información, UCR.

F-03 Establecer controles para la calidad de la información

Documento: Metodología de extracción de datos

Periodicidad de la extracción de datos

La frecuencia para la recolección de los datos fue establecida al final de cada ciclo lectivo en reunión efectuada desde el año 2012[18]. Desde entonces se ha consultado a la fuente por las fechas de recolección de los datos, y se reitera que las fechas seleccionadas son las adecuadas. El protocolo establece que los datos serán extraídos los días:

Matrícula

Ciclo lectivo	III ciclo Año anterior	I ciclo Año actual	II ciclo Año actual
Fecha de entrega	10 de julio	10 de julio	1 de diciembre

Notas

Ciclo lectivo	III ciclo Año anterior	I ciclo Año actual	II ciclo Año actual
Fecha de entrega	30 de agosto	30 de agosto	30 de enero

Graduación

Ciclo lectivo	Anual
Fecha de entrega	30 de enero

Admisión

Ciclo lectivo	Anual
Fecha de entrega	10 de julio

[18] Documento de bitácora de reuniones con fecha de 8 de marzo de 2012.

Método de extracción de los datos

El método de extracción de los datos se realizaba por medio de archivo de texto en formato csv desde el año 2012. A partir de octubre del 2015 se modifica el protocolo, tal como se muestra en la figura 4, y se obtiene un acceso por medio de vistas a la base de datos, para extraer la información de matrícula, notas, graduación y admisión, además de los datos requeridos para las dimensiones, que en este caso se registran en la fuente como catálogos. El protocolo está oficializado por las partes como se puede revisar en el documento de protocolo formal de extracción.

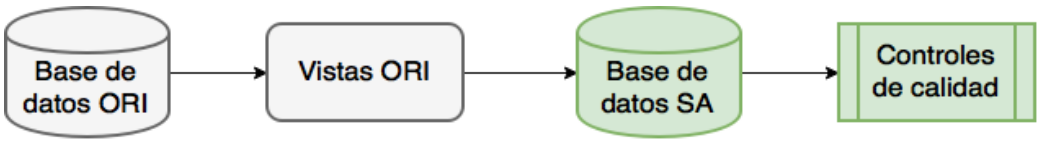


Figura 4. Método de extracción de datos de la fuente ORI.

Fuente. Elaboración propia.

Documento: Protocolo formal de extracción

El primer documento establece el protocolo y la información requerida para el DW con respecto a la fuente establecida, en este caso, la Oficina de Registro e Información. El segundo documento menciona la frecuencia para la recolección de los datos.



UNIVERSIDAD DE COSTA RICA

OPLAU
Oficina de Planificación Universitaria

13 de octubre de 2015
OPLAU-842-2015

Oficina de Registro e Información

Estimado señor:

Con el fin de contar con su apoyo y autorización, para obtener algunos datos que son registrados por la Oficina que usted dirige, seguidamente describo algunos antecedentes y las necesidades que se tienen actualmente, a saber:

En el 2012, por medio del oficio OPLAU-113-2012, se gestionó la información semestral y anual referente a la matrícula, notas, graduación y admisión de los estudiantes de la Institución. Durante estos años, se han recibido los datos por medio de archivos planos, los cuales es necesario agregar a otros archivos con datos históricos.

Actualmente, la Oplau se encuentra desarrollando un modelo para el almacenamiento y disponibilidad de esta información. Como parte del proceso, se considera pertinente obtener los datos directamente de la base de datos.

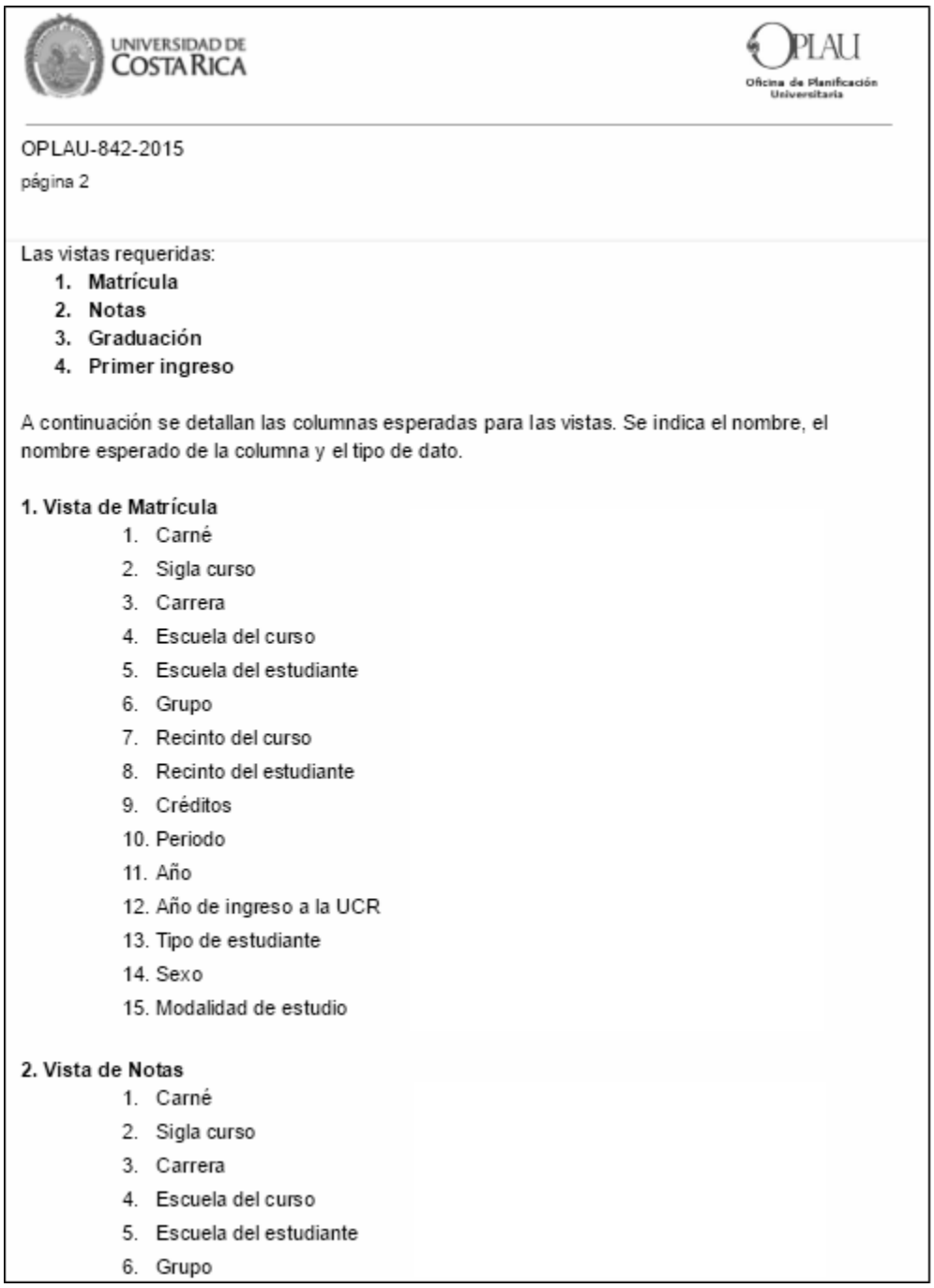
Con el propósito de analizar esta posibilidad, el Bach. Gosvami Zuñiga, de la Sección de Tecnologías de la Información de la Oplau, se reunió, el 5 de octubre del presente año, con la Jefa de la Unidad de Desarrollo Informático de la Oficina de Registro, y acordaron que no existe inconveniente para el cambio en la metodología de acceso a la información por parte de la Oplau.

La entrada a la base de datos se va a realizar por medio de accesos limitados, gestionados por el administrador de la base de datos, donde solo se pueda leer la información pertinente para la Oplau, dejando por fuera información sensible que no se requiere para los cálculos internos.

Lo más apropiado para recolectar la información es un acceso por medio de vistas, en las que se establece el contenido a mostrar, dejando de lado los procedimientos para realizar los enlaces y cálculos para obtener la información en el formato solicitado. Las vistas deben ser construidas a partir de las secuencias de comandos con las cuales se han generado los datos hasta ahora, de acuerdo a los requerimientos solicitados y sus respectivas modificaciones

Documento donde se establece el método de extracción y las variables esperadas.

Fuente: Oficina de Planificación, UCR.



UNIVERSIDAD DE COSTA RICA

OPLAU Oficina de Planificación Universitaria

OPLAU-842-2015

página 2

Las vistas requeridas:

1. **Matrícula**
2. **Notas**
3. **Graduación**
4. **Primer ingreso**

A continuación se detallan las columnas esperadas para las vistas. Se indica el nombre, el nombre esperado de la columna y el tipo de dato.

1. Vista de Matrícula

1. Carné
2. Sigla curso
3. Carrera
4. Escuela del curso
5. Escuela del estudiante
6. Grupo
7. Recinto del curso
8. Recinto del estudiante
9. Créditos
10. Periodo
11. Año
12. Año de ingreso a la UCR
13. Tipo de estudiante
14. Sexo
15. Modalidad de estudio

2. Vista de Notas

1. Carné
2. Sigla curso
3. Carrera
4. Escuela del curso
5. Escuela del estudiante
6. Grupo

Continuación del documento donde se establece el método de extracción y las variables esperadas. Fuente: Oficina de Planificación, UCR.

UNIVERSIDAD DE COSTA RICA

OPLAU
Oficina de Planificación

21 de marzo de 2012
OPLAU-113-2012

OFICINA DE REGISTRO E INFORMACION U.C.R.
21 MAR 2012
RECIBIDO

Oficina de Registro e Información

Estimado señor:

En la reunión del día jueves 8 de marzo del presente año, se llegaron a los siguientes acuerdos, para el envío de información por parte de su oficina a la OPLAU:

1. Archivos de matrícula y de notas en formato txt, según las siguientes fechas:

Archivo de Matricula

Ciclo lectivo	III ciclo año anterior	I ciclo año actual	II ciclo año actual
Fecha de entrega	10 de julio	10 de julio	1 de diciembre

Archivo de notas

Ciclo lectivo	III ciclo año anterior	I ciclo año actual	II ciclo año actual
Fecha de entrega	30 de agosto	30 de agosto	30 de enero

Las variables que debe contener cada uno de los archivos mencionados anteriormente son las siguientes:

- carne (carne del estudiante) que matriculo el curso)

Documento para la entrega de los datos con el detalle y periodicidad requeridos.

Fuente: Oficina de Planificación, UCR.

Documento: Lista de datos de control

Acuerdo para solicitar información o reportes de datos para comparación.

Los datos de control requeridos (separados por período-año) para matrícula y notas:

- Total de cupos.
- Total de cupos por nivel académico.
- Total de estudiantes.
- Total de estudiantes por nivel académico.
- Total de estudiantes por recinto universitario.

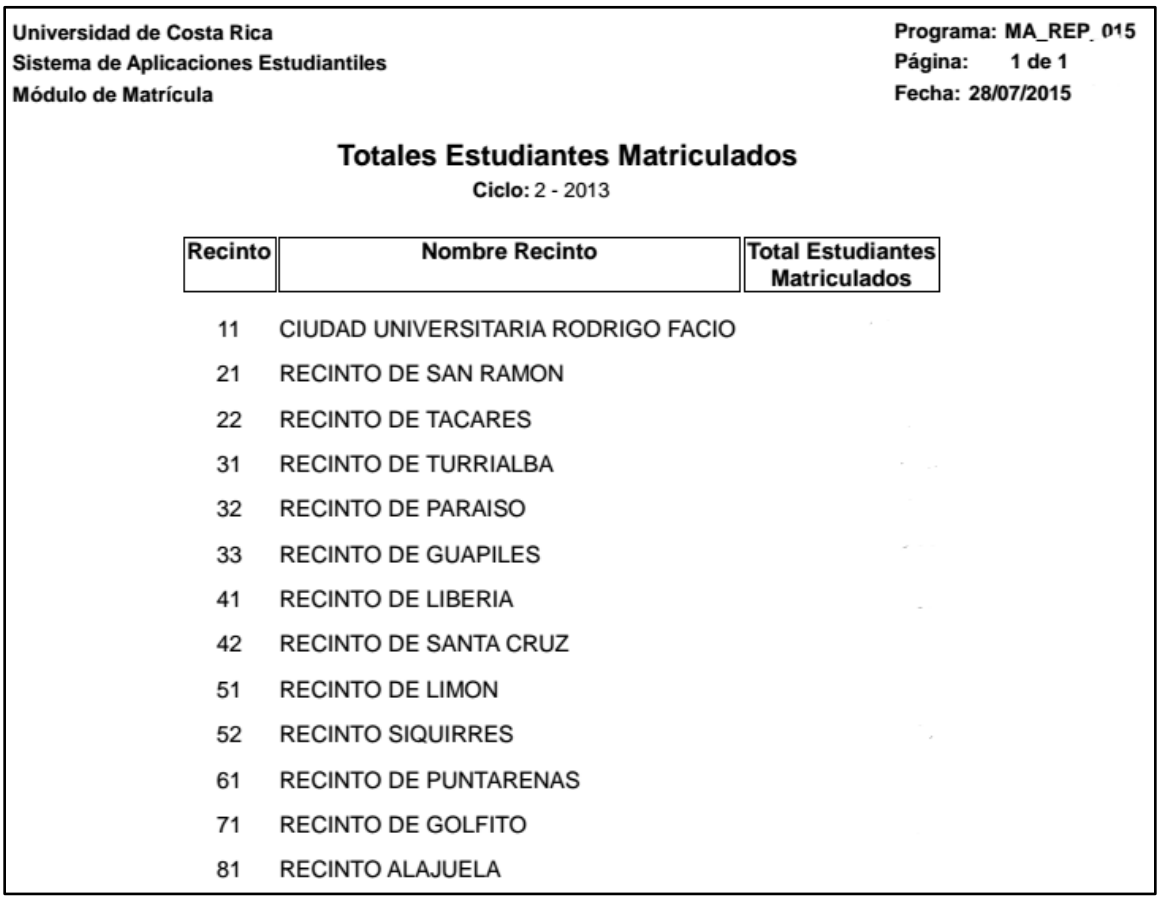
Los datos de control requeridos (separados por año) para graduación:

- Total de estudiantes graduados.
- Total de estudiantes graduados por recinto universitario.
- Total de títulos otorgados.

Los datos de control requeridos (separados por año) para admisión:

- Total de estudiantes admitidos de primer ingreso.
- Total de estudiantes admitidos de primer ingreso por recinto universitario.

La figura 5 es una muestra de los reportes suministrados por la Oficina de Registro referente al total de estudiantes matriculados por recinto universitario para efectos de control.



Universidad de Costa Rica
Sistema de Aplicaciones Estudiantiles
Módulo de Matrícula

Programa: MA_REP_015
Página: 1 de 1
Fecha: 28/07/2015

Totales Estudiantes Matriculados

Ciclo: 2 - 2013

Recinto	Nombre Recinto	Total Estudiantes Matriculados
11	CIUDAD UNIVERSITARIA RODRIGO FACIO	
21	RECINTO DE SAN RAMON	
22	RECINTO DE TACARES	
31	RECINTO DE TURRIALBA	
32	RECINTO DE PARAISO	
33	RECINTO DE GUAPILES	
41	RECINTO DE LIBERIA	
42	RECINTO DE SANTA CRUZ	
51	RECINTO DE LIMON	
52	RECINTO SIQUIRRES	
61	RECINTO DE PUNTARENAS	
71	RECINTO DE GOLFITO	
81	RECINTO ALAJUELA	

Figura 5. Muestra de reporte para datos de control.

Fuente. Oficina de Registro e Información, UCR.

Documento: Procedimiento de aseguramiento de los datos

La lista de documentos en los que se incluye el aseguramiento de los datos:

- DataMart Graduados ORI.
- DataMart Primer Ingreso ORI.
- DataMart Notas ORI.
- DataMart Matrícula ORI.
- DataMart Matrícula ORI Controles.

Los documentos no están disponibles para su publicación, por lo tanto, se omiten de esta documentación.

F-04 Depuración de la información disponible

La aplicación de los procedimientos mencionados en el apartado de aseguramiento implican que los datos han sido revisados para ubicar:

- Inconsistencias contra los datos de control.
- Variables faltantes para realizar cálculos, filtrados y agrupaciones.
- Descripciones faltantes en las variables.
- Variables desconocidas o fuera de contexto.

Las inconsistencias y variaciones encontradas durante la ejecución de este proceso, se deben documentar e incluirlas en los protocolos existentes.

Documento: Datos asegurados

Luego de una revisión de los datos, se detectan valores fuera del rango que deben ser modificados para calzar en los parámetros de búsqueda. Estos serán restaurados luego de un análisis de la mejor técnica con el menor riesgo para la consistencia de los datos. Al ser pocos datos erróneos, será un cambio manual.

Datamart Graduados

- Se ubican 3 títulos donde la escuela no corresponde con el nombre del título. Los títulos se ubican en la escuela de estudios generales (código 0001), que inicialmente no tenían área asociada, y por eso fue sencillo ubicarlos. Se opta por mantenerlos en el área y facultad de la escuela de estudios generales para no modificar el código del título original. Los códigos de los títulos son:
    a. 0001090200201 BACHILLER EN CIENCIAS DE LA EDUCACION CON ENFASIS EN ADMINISTRACION EDUCATIVA (PLAN ESPECIAL)
    b. 0001120100201 BACHILLER EN BIBLIOTECOLOGIA CON ENFASIS EN BIBLIOTECAS EDUCATICAS (PLAN ESPECIAL)
    c. 0001030100101 DIPLOMADO EN CITOTECNOLOGIA (PLAN ESPECIAL)

Datamart Primer Ingreso

- Los estudiantes con carné nulos. Cantidad: 2 del año de ingreso 2006.

```
SELECT [IDENTIFICACION_ESTUDIANTE]
      ,[TIPO_IDENTIFICACION]
      ,[CARNE]
      ,[ANNO_SOLICITUD]
      ,[ANNO_INGRESO]
  FROM [siucr_SA].[dbo].[primer_ingreso_auto_produccion]
  where carne is null
```

Datamart Notas

- Se encontró que una nota númerica estaba NULL. Esto provocó un error en la salida del SSIS. Para encontrar las filas con problemas, se redirigió el error a un archivo plano y se determinó la fila y el carné con el error. Se hizo un SELECT en el SA para ver los datos, y visualmente se encontró el nulo y se corrigió.

Datamart Matrícula

- Algunos de los códigos de escuela que no coinciden es porque el archivo de origen usa el código 1,2,3... y en el catálogo de escuelas viene 0001,0002,0003... Hay que convertirlos a como están en el catálogo con la longitud de 4 caracteres (Ej.Update en SQL).
- Actualizar créditos a 0 para grupos > 50 con el resto de columnas repetidas. Problema encontrado durante el primer ciclo del 2015 y aplica retroactivo. Se modificaron 119517 registros en total (119445 + 72)
- Actualizar variable de sexo en blanco a partir de los históricos disponibles del estudiante y asignar un dato válido.

```
UPDATE dbo.[FactMatricula]
SET id_sexo = 2
WHERE id_estudiante IN (SELECT id_estudiante
                    FROM DimEstudiante
                    WHERE carne IN ('Carne');
```

3. Diseño

D-01 Modelo lógico

Documento: Modelo lógico

La identificación de los procesos se realizó con la matriz de procesos y las metas, para luego detallar el modelo con la lista de variables disponibles. La figura 6 muestra la fuente ORI como la entrada única de datos y los cuatro procesos definidos en la matriz de procesos relacionados con esta fuente. Los resultados, dimensiones y medidas se extraen de la lista de variables disponibles.



Figura 6. Modelo lógico.
Fuente: Elaboración propia.

Documento: Modelo dimensional

En la figura 7 del diagrama dimensional de estrella se aplica la técnica de dimensiones compartidas para las tres tablas de hechos propuestas en las figuras 8, 9 y 10. Aun cuando los datos de matrícula, notas y graduaciones no son equiparables en una sola tabla de hechos, se comparten seis de las dimensiones para mostrar una única verdad y unificar los procesos del negocio.

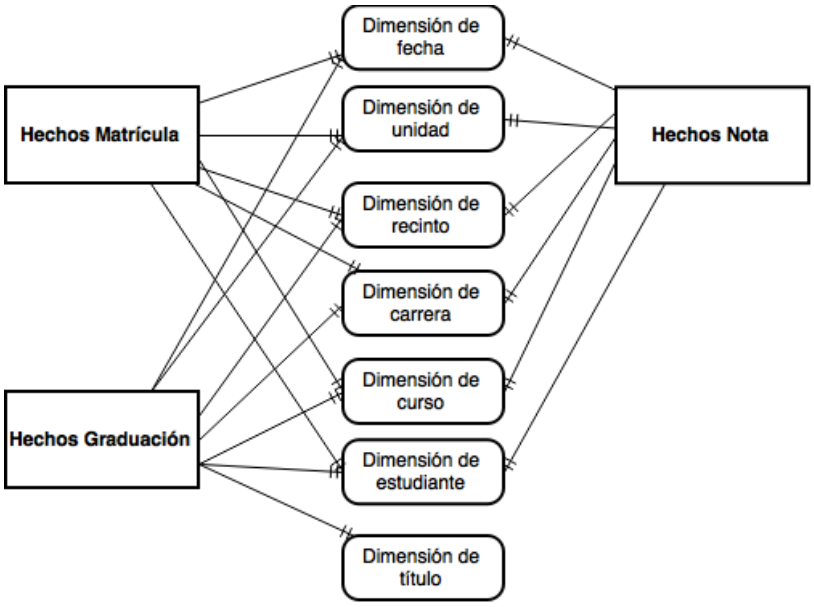


Figura 7. Diagrama dimensional de estrella.

Fuente: Elaboración propia.

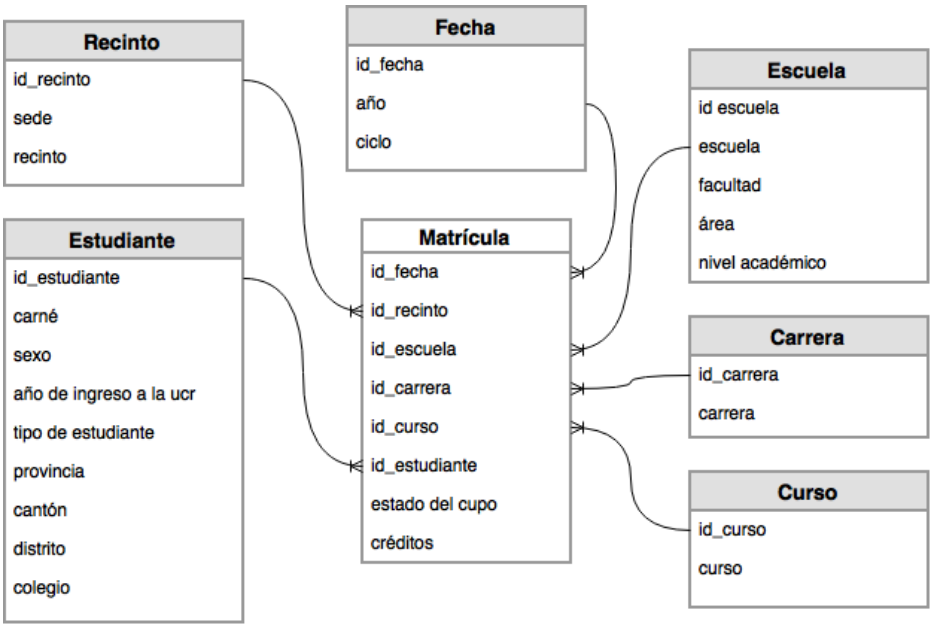


Figura 8. Diagrama dimensional para matrícula.

Fuente: Elaboración propia.

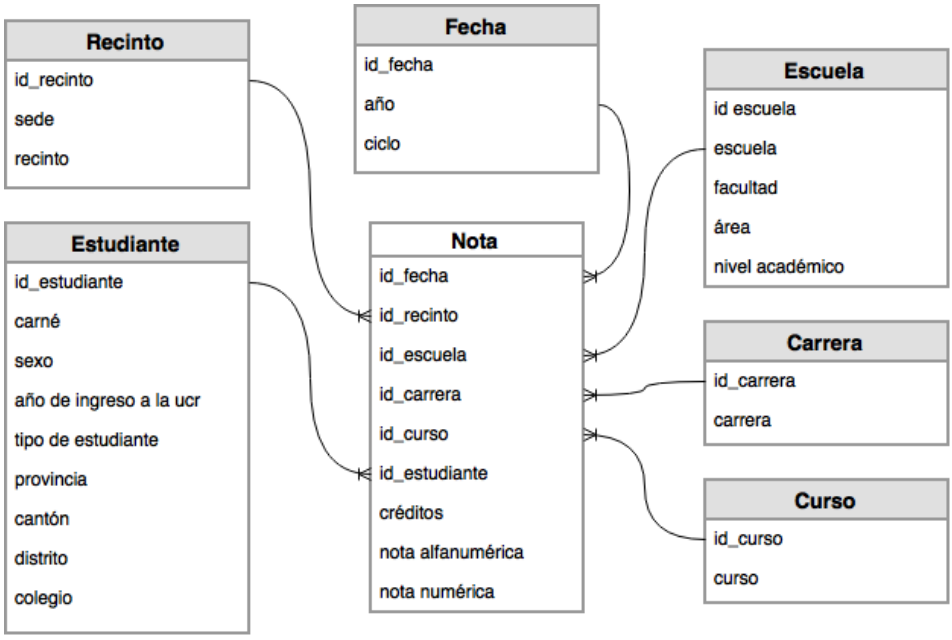


Figura 9. Diagrama dimensional para notas.

Fuente: Elaboración propia.

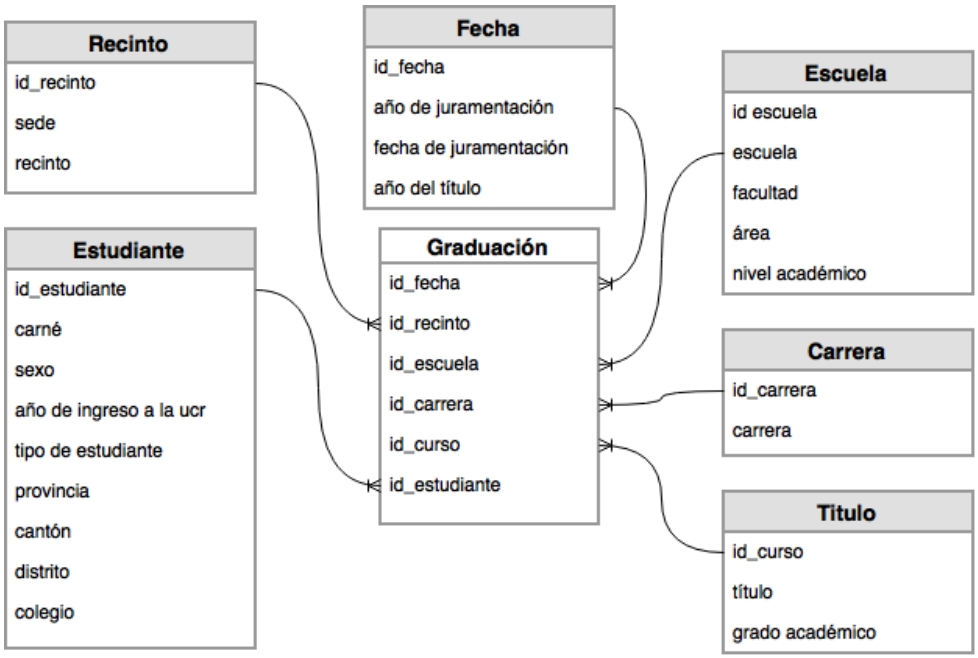


Figura 10. Diagrama dimensional para graduación.

Fuente: Elaboración propia.

En este caso, la tabla de hechos de graduación es una tabla factless. Esta tabla carece de medidas, pero ofrece toda la información requerida como cantidad de títulos, de estudiantes graduados y de graduados por escuela.

D-03 Modelo físico

Documento: Modelo físico

Nomenclatura de objetos

La nomenclatura utilizada propone el uso de nombres descriptivos, identificación de dimensiones y hechos, y el uso de nombres no relacionados con una fuente específica, en especial las dimensiones.

Las dimensiones se identifican con el prefijo *Dim* y sin espacios en blanco o guiones entre palabras, solo el uso de mayúscula en la primera letra de cada palabra (CamelCase). No se utilizan acentos del español en los nombres. Ejemplo de uso:

- DimEscuela

Las tablas de hechos se identifican con el prefijo *Fact*. Al igual que las dimensiones, se utiliza sin espacios en blanco o guiones entre palabras, solo el uso de mayúscula en la primera letra de cada palabra. No se utilizan acentos del español en los nombres. Ejemplo de uso:

- FactGraduacion

La llave primaria y la llave foránea se identifica por el prefijo *id*. Las columnas utilizan guiones bajo para separar las palabras y siempre en minúscula.

- id_curso

Lo recomendado para las variables es iniciar con un prefijo que identifique el uso y luego el nombre genérico relacionado con la tabla. Por ejemplo, en este caso, se identifica la relación entre las 3 variables y su uso:

- id_curso
- sigla_curso
- nombre_curso

Si la variable no está relacionada directamente con la dimensión, es recomendable un nombre descriptivo de su función. Por ejemplo:

- creditos_matriculados
- estado_del_curso

Estructura de tablas

La estructura de las tablas de acuerdo a la nomenclatura definida con el tipo de datos y las llaves sustitutas para las dimensiones. En este caso ninguna variable se permite en nulo.

DimCarrera

id_carrera	int	llave primaria
sigla_carrera	char(6)	llave del negocio
nombre_carrera	varchar(150)	

DimCurso

id_curso	int	llave primaria
sigla_curso	char(6)	llave del negocio
nombre_curso	nvarchar(150)	

DimEscuela

id_escuela	int	llave primaria
sigla_escuela	char(4)	llave del negocio
nombre_escuela	varchar(150)	
nivel_academico	varchar(20)	
sigla_area	char(1)	llave del negocio
nombre_area	varchar(150)	
sigla_facultad	char(1)	llave del negocio
nombre_facultad	varchar(150)	

DimEstudiante

id_estudiante	int	llave primaria
carne	char(6)	
anno_ingreso_ucr	int	

id_sexo	int	
sexo	varchar(50)	
codigo_tipo_estudiante	int	llave del negocio
tipo_estudiante	varchar(50)	
Codigo_provincia_procedencia	varchar(1)	llave del negocio
Nombre_provincia_procedencia	varchar(150)	
codigo_canton_procedencia	varchar(4)	llave del negocio
nombre_canton_procedencia	varchar(150)	
codigo_distrito_procedencia	varchar(4)	llave del negocio
nombre_distrito_procedencia	varchar(150)	
Codigo_colegio	int	llave del negocio
nombre_colegio	varchar(150)	
tipo_colegio	varchar(150)	
modalidad_colegio	varchar(150)	

DimRecinto

id_sede_recinto	int	llave primaria
num_sede	int	llave del negocio
nombre_sede	varchar(150)	
num_recinto	int	llave del negocio
nombre_recinto	varchar(150)	

DimTitulo

Id_titulo	int	llave primaria
codigo_titulo	char(13)	
nombre_titulo	varchar(250)	
codigo_grado_academico	int	llave del negocio
nombre_grado_academico	varchar(250)	

DimFecha

id_anno_ciclo	int	llave primaria
anno	int	
num_ciclo	int	
Nombre_ciclo	varchar(20)	

FactMatricula

id_anno_ciclo	int
id_estudiante	int
id_curso	int
id_sede_recinto_estudiante	int
id_sede_recinto_curso	int
id_escuela_estudiante	int
id_escuela_curso	int
Id_carrera_estudiante	int
numero_grupo	int
creditos_matriculados	int
estado_del_curso	varchar(150)

FactNota

id_anno_ciclo	int
id_estudiante	int
id_curso	int
id_sede_recinto_estudiante	int
id_sede_recinto_curso	int
id_escuela_estudiante	int
id_escuela_curso	int
Id_carrera_estudiante	int
numero_grupo	int
creditos_matriculados	int
nota_ordinaria_num	numeric(10, 2)
nota_ordinaria_alf	varchar(10)

FactGraduacion

Id_sede_recinto_titulo	int
id_carrera_titulo	int
id_escuela_titulo	int
Id_estudiante	int
id_titulo	int

Id_anno_ciclo int

Estrategias de implementación en base de datos

La decisión sobre el manejo de índices, particionamiento y almacenamiento recae, por el momento, en el personal administrador de la base de datos. La decisión se origina de la situación actual de contar con un motor de base de datos optimizado para datos relacionales. El servicio por ahora no se adecua a los requerimientos de un sistema de BI, pero la cantidad de datos se mantiene por debajo de 10 GB -como se demuestra en la sección de dimensionamiento-, tamaño que un motor de base de datos relacional puede administrar.

Dimensionamiento

Para las unidades relacionadas, se proyecta la cantidad de datos con los que se va a contar por año y al cabo de 5 años en la tabla 8.

Nombre	**Oficina fuente**	**Formato origen**	**Periodos disponibles**	**Registros por año**	**Registros al último corte**
Matrícula	ORI	BD y texto	2007-2015	330.000	2.421.000
Notas	ORI	BD y texto	1990-2015	330.000	6.572.000
Graduación	ORI	BD y texto	1990-2015	5.500	116.000

Tabla 8. Datos existentes. Fuente: Elaboración propia.

La proyección de espacio utilizado y de crecimiento de usuarios a 5 años en la tabla 9.

Proyectado	**Usuarios**	**Espacio a utilizar**
1er año	25	1,0 GB
3er año	100	1,5 GB
5to año	250	2,0 GB

Tabla 9. Datos existentes. Fuente: Elaboración propia.

D-04 Metadatos

Documento: Metadatos

Lista de esquemas

SI_DW	El esquema de base de datos SI_DW almacena los datos relacionados con las dimensiones y tablas de hecho referentes a la Universidad de Costa Rica.

Lista de dimensiones y hechos

Nombre	**Tipo**	**Definición**	**Acepciones**
Fecha	Dimensión	La unificación de fechas por año y ciclo	Período de matrícula Fecha de juramentación
Unidad	Dimensión	Unidad organizativa de la institución.	Escuela Unidad ejecutora
Recinto	Dimensión	Establecimiento universitario con el conjunto de servicios esenciales para impartir lecciones.	Recinto del estudiante Recinto de la carrera Recinto del curso
Carrera	Dimensión	Preparación del estudiante en un área de conocimiento, tomando los cursos dispuestos en un plan de estudios.	Carrera base Carrera del estudiante Carrera del título
Curso	Dimensión	Modalidad de enseñanza que imparte conocimiento durante un período corto de tiempo, un ciclo. Requiere ser aprobado para demostrar su aprovechamiento	-
Estudiante	Dimensión	Individuo inscrito en la Universidad y validado para atender diferentes cursos, y así para obtener una titulación o certificación.	-
Título	Dimensión	Certificación expedida por la institución para acreditar al estudiante la finalización	Diploma Título de grado Título de posgrado

		satisfactoria de una carrera.	
Matrícula	Hechos	Procedimiento que realiza el estudiante para inscribirse en un curso en un año y ciclo específico.	-
Nota	Hechos	Valor final obtenido de aprovechamiento de cada curso del estudiante. No es unificable con matrícula pues cada uno incluye y excluye algunos cursos por tipo.	-
Graduación	Hechos	Acto donde se otorga el título al estudiante por cumplir los requisitos solicitados para finalizar la carrera.	-

Lista de medidas

Nombre	**Definición**	**Acepciones**	**Frecuencia**	**Granularidad**	**Dimensiones asociadas**
Créditos	Valor asignado a cada curso basado en el tiempo de dedicación del estudiante.		Por ciclo	Por cupo	Fecha Recinto Escuela Carrera Curso Estudiante Grupo (dimensión degenerada)
Estudiantes físicos matriculados	Total de individuos no repetidos.		Por ciclo	Por estudiante	Fecha Recinto Escuela Carrera Curso Estudiante
Estudiantes equivalentes de tiempo completo	Cálculo para equiparar el uso real de los cupos si todos los estudiantes cursaran igual cantidad de créditos.		Por ciclo	Por cupo	Fecha Recinto Escuela Carrera Curso Estudiante Grupo (dimensión degenerada)

Total de cursos	Cantidad de cursos diferentes impartidos.		Por ciclo	Por curso	Fecha Recinto Escuela Carrera Curso Estudiante
Total de grupos	Cantidad de grupos diferentes impartidos.		Por ciclo	Por grupo	Fecha Recinto Escuela Carrera Curso Estudiante Grupo (dimensión degenerada)
Total de estudiantes graduados	Total de individuos no repetidos en un periodo.		Por año	Por estudiante	Fecha Recinto Escuela Carrera Curso Estudiante
Total de títulos	Cantidad de títulos otorgados en un período.		Por año	Por título	Fecha Recinto Escuela Carrera Curso Estudiante Título
Nota numérica	Valor obtenido de aprovechamiento del curso		Por ciclo	Por cupo	Fecha Recinto Escuela Carrera Curso Estudiante Grupo (dimensión degenerada)
Nota alfanumérica	Valor obtenido de aprovechamiento del curso con caracteres de identificación especiales.		Por ciclo	Por cupo	Fecha Recinto Escuela Carrera Curso Estudiante Grupo (dimensión degenerada)

Lista de atributos

Nombre	Tipo de dato	Dominio y reglas de validación
Año	entero	Año > 1940 y Año <= Año actual
Ciclo	entero	Ciclo = 1,2,3
Nombre del ciclo	cadena	1=primer ciclo, 2=segundo ciclo, 3= verano
Código de sede	entero	Código >= 10 y código <= 99
Nombre de sede	cadena	Contra la lista oficial de nombres de las sedes.
Código recinto	entero	Código >= 10 y código <= 99
Nombre de recinto	cadena	Contra la lista oficial de nombres de las sedes.
Carné	cadena	6 caracteres, tamaño fijo.
sexo	cadena	Masculino, Femenino
Año de ingreso a la UCR	entero	Año > 1940 y Año <= Año actual
Tipo de estudiante	cadena	0=NO CONSOLIDADO, 1=Estudiante de pregrado y grado, 2=Estudiante de postgrado, 3=Estudiante de programas especiales, 4=Estudiante de extensión docente, 5=Estudiantes visitantes, 9=NO SE CONOCE
Provincia	cadena	Contra la lista oficial.
Cantón	cadena	Contra la lista oficial.
Distrito	cadena	Contra la lista oficial.
Colegio	cadena	-
Modalidad de colegio	cadena	1=Dato no disponible, 2=DEL EXTERIOR, 3=EDUCACION ABIERTA, 4=PRIVADO, 5=PUBLICO, 6=SUBVENCIONADO
Tipo de colegio	cadena	1=ACADEMICO, 2=ARTISTICO, 3=CIENTIFICO, 4=Dato no disponible, 5=DEL EXTERIOR, 6=EDUCACION ABIERTA, 7=TECNICO
Código de escuela	cadena	4 dígitos, tamaño fijo. Contra la lista oficial.
Nombre de escuela	cadena	Contra la lista oficial.

Facultad	cadena	Contra la lista oficial.
Área	cadena	Contra la lista oficial.
Nivel académico	cadena	‘PREGRADO Y GRADO’ y ‘POSGRADO’
Código de carrera	cadena	6 dígitos. Tamaño fijo.
Nombre de carrera	cadena	Contra la lista oficial.
Sigla del curso	cadena	6 dígitos. Tamaño fijo.
Estado del cupo	cadena	‘consolidado’ y ‘retirado’.
Código del título	cadena	13 dígitos. Tamaño fijo.
Nombre del título	cadena	-
Grado académico	cadena	0=Profesorado, 1=Diplomado, 2=Bachillerato, 3=Licenciatura, 4=Especialidad, 5=Maestría, 6=Doctorado, 8=9=SIN DEFINIR
Fecha de juramentación	fecha	Entre 1/1/1940 hasta 31/12/año en curso
Año de juramentación	entero	Año > 1940 y Año <= Año actual
Año del título	entero	Año > 1940 y Año <= Año actual

La figura 11 y 12 es un ejemplo de los metadatos extraídos de las tablas utilizando un método automatizado desde el motor de base de datos.

Tabla 3 de 17	**B.D.** RegistroMatricula_DW	Fecha de creación: Feb 10 2016 10:50AM				
Nombre de tabla: DimEscuela						
Nombre de campo	**Llave primaria**	**Tipo de columna**	**Valor por Default**	**Longitud**	**Valores nulos**	**Descripción de la columna**
id_escuela	X	int identity		4	NO	Llave primaria autonumerada
sigla_escuela		char		4	NO	Identificador original de la e
nombre_escuela		varchar		150	NO	Nombre de la escuela
nivel_academico		varchar		20	NO	Los valores son 'PREGRADO Y GR
sigla_area		char		1	NO	Identificador de las posibles
nombre_area		varchar		150	NO	Nombres de las 7 áreas en las
sigla_facultad		char		1	NO	Identificador de la facultad e
nombre_facultad		varchar		150	NO	Nombre de la facultad donde se

Figura 11. Metadatos de objetos de la dimensión de escuela.

Fuente: Elaborado a partir de script[19] automatizado para obtener el diccionario de datos.

[19] http://consejosdelguru.blogspot.com/2011/09/generar-diccionario-de-datos-sql-server.html

Tabla 4 de 17	B.D. RegistroMatricula_DW	Fecha de creación: Sep 14 2015 10:37AM				
Nombre de tabla: DimEstudiante						
Nombre de campo	**Llave primaria**	**Tipo de columna**	**Valor por Default**	**Longitud**	**Valores nulos**	**Descripción de la columna**
id_estudiante	X	int identity		4	NO	Llave primaria autonumerada
carne		char		6	NO	identificador único del estudi
anno_ingreso_ucr		int		4	NO	Año en que matricula por prime
id_sexo		int		4	NO	valores posibles '1=Masculino'
sexo		varchar		50	NO	valores posibles '1=Masculino'
codigo_tipo_estudiante		int		4	NO	valores posibles '0=NO CONSOLI
tipo_estudiante		varchar		50	NO	valores posibles '0=NO CONSOLI
fila_actual		bit		1	NO	requerido para dimensión tipo

Figura 12. Metadatos de objetos de la dimensión de estudiante.

Fuente: Elaborado a partir de script[20] automatizado para obtener el diccionario de datos.

Listado de procesos de desarrollo

La documentación del desarrollo del DW se obtiene a lo largo de los procesos aplicados. Los documentos elaborados durante los procesos anteriores y considerados en este rubro son:

- Documento: diseño de la arquitectura
- Documento: matriz de procesos
- Documento: listado de metas y análisis viables
- Documento: Metodología de extracción de los datos
- Documento: Modelo lógico
- Documento: Modelo dimensional

[20] http://consejosdelguru.blogspot.com/2011/09/generar-diccionario-de-datos-sql-server.html

4. Visualización

V-01 Público meta

Documento: Análisis y complejidad de visualización por tipo de usuario

La presentación de los datos debe separarse de acuerdo al público meta de los análisis por seguridad y por facilidad para el usuario. La separación toma en cuenta factores como la granularidad permitida, la confidencialidad de los datos, la complejidad del análisis y el cumplimiento de objetivos específicos para cada sector.

Roles de usuario

El primer paso es ubicar los roles para cada usuario. Se determinan 6 roles, divididos por internos y externos.

1. Público en general
2. Instituciones gubernamentales
3. Comunidad universitaria
4. Investigadores
5. Mandos medios
6. Autoridades

Los roles externos son para el acceso al público e instituciones de control gubernamentales. Los internos, por su parte, tienen un acceso general dirigido a la población universitaria como estudiantes, docentes y administrativos. Para información más detallada, se contemplan los roles de investigadores, mandos medios (directores, decanos, jefaturas) y autoridades (vicerrectorías y rectoría).

Complejidad de acceso

El diagrama de la figura 13 muestra la segregación por roles de acuerdo a la complejidad para obtener y manipular la información. En este caso, para los roles externos y el nivel general de acceso interno, la visualización de los datos se establece como estática. Los datos estáticos contemplan reportes predefinidos, tablas de datos, resúmenes de resultados y gráficos de pocas variables. Para los usuarios internos, se permite el manejo de los datos por cubos multidimensionales, para brindar al usuario la libertad de obtener sus propios análisis. En el caso de los investigadores, se contempla un nivel extra de complejidad, para que realicen las consultas con un lenguaje más avanzado y obtener datos específicos de acuerdo a sus necesidades.

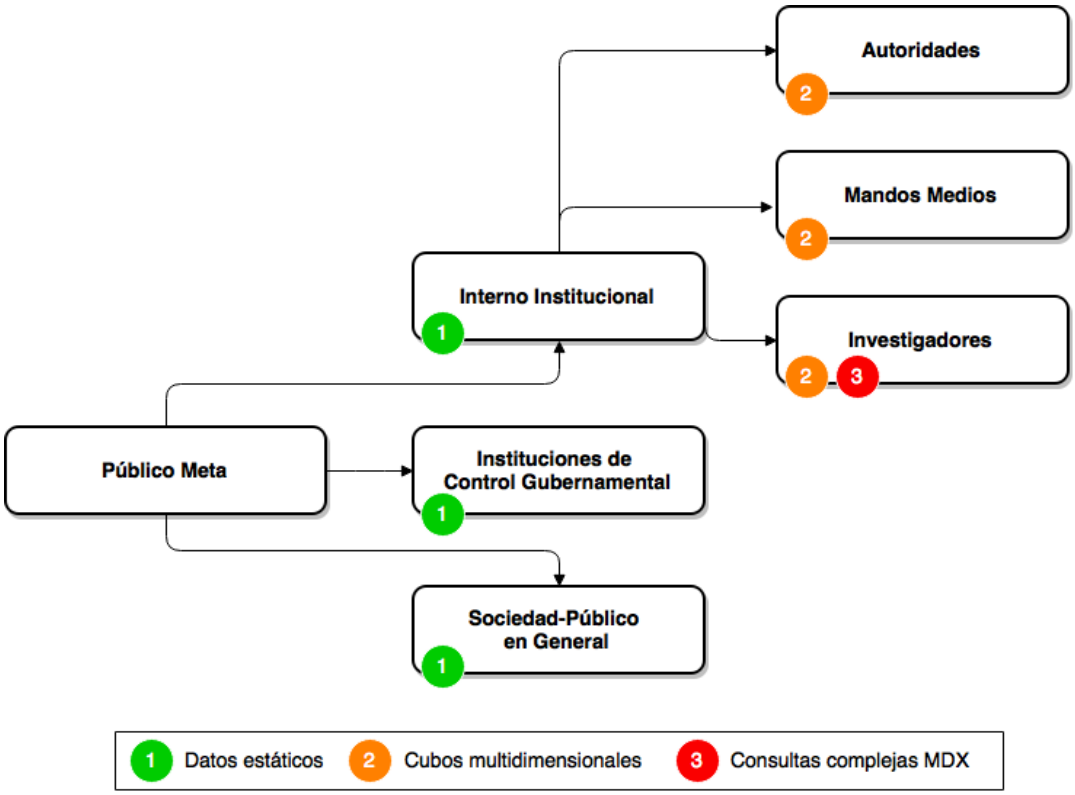


Figura 13. Jerarquía de acceso del público meta por complejidad de visualización.

Fuente: Elaboración propia.

Seguridad de acceso por granularidad

La seguridad de la información se aplica al restringir la ejecución de consultas sobre dimensiones, variables o metas de acuerdo al rol. En la tabla 10 se expone la vinculación

de dimensiones y variables contra roles, indicando con una marca las variables permitidas para cada rol de usuario.

		Granularidad permitida por roles					
Dimensiones con acceso	Variables en la dimensión	Todo público	Instituciones de gobierno	Interno general	Investigadores	Mandos medios	Autoridades
Fecha	Año	●	●	●	●	●	●
	Ciclo	●	●	●	●	●	●
	Año juramentación	●	●	●	●	●	●
Escuela	Área	●	●	●	●	●	●
	Facultad				●	●	●
	Escuela				●	●	●
	Nivel académico	●	●	●	●	●	●
Recinto	Sede	●	●	●	●	●	●
	Recinto	●	●	●	●	●	●
Carrera	Carrera				●	●	●
Curso	Curso					●	●
Estudiante	Carné						●
	Año ingreso a la U				●	●	●
	Sexo	●	●	●	●	●	●
	Provincia	●	●	●	●	●	●
	Cantón				●	●	●
	Distrito				●	●	●
	Colegio						●
	Tipo de colegio	●	●	●	●	●	●
	Modalidad de colegio	●	●	●	●	●	●
Título	Título					●	●
	Grado académico	●	●	●	●	●	●

Tabla 10. Matriz de acceso a las dimensiones y sus variables.

Fuente: Elaboración propia.

La dimensión de escuela se filtra por la estructura organizacional para los mandos medios, de acuerdo al mando que ejerce el usuario. Por ejemplo, el decano de la facultad de ciencias básicas puede analizar datos de las escuelas de física, matemática o química,

pero se impide el acceso a los datos del resto de escuelas, como medicina o música. Este filtro queda sin efecto para las variables que se definieron como públicas.

Seguridad de acceso por meta

El otro tipo de restricción se aplica a nivel de la meta o el análisis. En este caso se limita el acceso por roles impidiendo la ejecución de una meta de acuerdo a las restricciones indicadas en la tabla 11.

	Acceso a metas por roles					
Metas y análisis	Todo público	Instituciones de gobierno	Interno general	Investigadores	Mandos medios	Autoridades
Créditos matriculados	●	●	●	●	●	●
Estudiantes matriculados	●	●	●	●	●	●
Cupos matriculados	●	●	●	●	●	●
Estudiantes equivalentes de tiempo completo	●	●	●	●	●	●
Estudiantes graduados	●	●	●	●	●	●
Títulos otorgados	●	●	●	●	●	●
Créditos aprobados y retirados					●	●
Estudiantes matriculados de primer ingreso					●	●
Cursos matriculados					●	●
Grupos matriculados					●	●
Tasa de rendimiento académico					●	●

Tabla 11. Permisos para metas por tipo de usuario.

Fuente: Elaboración propia.

V-02 Validar conceptos de visualización

Documento: Valoración del acceso de datos por tipo de usuario y nivel de agregación permitido

La validación se realiza sobre las dimensiones y roles de los usuarios para la figura 14 de valoración de accesos, y contra la lista de metas viables en la figura 15 de valoración de permisos por meta.

Documento para validar el acceso a las dimensiones y variables por rol de usuario

Validado por: _____

Fecha: 14/7/2016

		Granularidad permitida por roles							
Dimensiones	Variables en la dimensión	Todo público	Instituciones de gobierno	Interno general	Investigadores	Mandos medios	Autoridades	Validado	Observaciones
Fecha	Año	•	•	•	•	•	•		
	Ciclo	•	•	•	•	•	•		
	Año juramentación	•	•	•	•	•	•		
Escuela	Área	•	•	•	•	•	•		
	Facultad				•	•	•		
	Escuela				•	•	•		
	Nivel académico	•	•	•	•	•	•		
Recinto	Sede	•	•	•	•	•	•		
	Recinto	•	•	•	•	•	•		
Carrera	Carrera				•	•	•		
Curso	Curso					•	•		
Estudiante	Carné						•		
	Año ingreso a la U				•	•	•		
	Sexo	•	•	•	•	•	•		
	Provincia	•	•	•	•	•	•		
	Cantón				•	•	•		
	Distrito				•	•	•		
	Colegio						•		
	Tipo de colegio	•	•	•	•	•	•		
	Modalidad de colegio	•	•	•	•	•	•		
Título	Título					•	•		
	Grado académico	•	•	•	•	•	•		

Figura 14. Valoración de los accesos por usuario.
Fuente: Elaboración propia.

Documento para validar el acceso a las metas por rol de usuario

Validado por: _____

Fecha: 15/7/2016

	Acceso a metas por roles							
Metas y análisis	**Todo público**	**Instituciones de gobierno**	**Interno general**	**Investigadores**	**Mandos medios**	**Autoridades**	**Validado**	**Observaciones**
Créditos matriculados	•	•	•	•	•	•		
Estudiantes matriculados	•	•	•	•	•	•		
Cupos matriculados	•	•	•	•	•	•		
Estudiantes equivalentes de tiempo completo	•	•	•	•	•	•		
Estudiantes graduados	•	•	•	•	•	•		
Títulos otorgados	•	•	•	•	•	•		
Créditos aprobados y retirados					•	•		
Estudiantes matriculados de primer ingreso					•	•		
Cursos matriculados					•	•		
Grupos matriculados					•	•		
Tasa de rendimiento académico					•	•		

Figura 15. Valoración de permisos para metas por tipo de usuario.

Fuente. Elaboración propia.

5. Implementación

I-01 Estructura del DW

Documento: Script de creación de la estructura en la BD

Extracto del *script* de creación de la base de datos en el motor **SQL Server 2012**.

```
USE [SI_DW]
GO
/****** Object:  Table [dbo].[DimCarrera]   Script Date: 10/7/2016 2:12:36 p. m. ******/
CREATE TABLE [dbo].[DimCarrera](
	[id_carrera] [int] IDENTITY(1,1) NOT NULL,
	[sigla_carrera] [char](6) NOT NULL,
	[nombre_carrera] [varchar](150) NOT NULL,
	[fila_actual] [bit] NOT NULL,
 CONSTRAINT [PK_DimCarrera] PRIMARY KEY CLUSTERED
(
	[id_carrera] ASC
)) ON [PRIMARY]

GO
/****** Object:  Table [dbo].[DimCurso]     Script Date: 10/7/2016 2:12:36 p. m. ******/
CREATE TABLE [dbo].[DimCurso](
	[id_curso] [int] IDENTITY(1,1) NOT NULL,
	[sigla_curso] [char](6) NOT NULL,
	[nombre_curso] [nvarchar](150) NOT NULL,
 CONSTRAINT [PK_DimCurso] PRIMARY KEY CLUSTERED
(
	[id_curso] ASC
)) ON [PRIMARY]
....
ALTER TABLE [dbo].[FactNota]  WITH CHECK ADD  CONSTRAINT [FK_FactNota_DimSedeRecinto1]
FOREIGN KEY([id_sede_recinto_curso]) REFERENCES [dbo].[DimSedeRecinto]
([id_sede_recinto])
GO
ALTER TABLE [dbo].[FactNota] CHECK CONSTRAINT [FK_FactNota_DimSedeRecinto1]
GO
```

Documento: Estructura del DW en la BD

Se adjuntan los tres diagramas en las figuras 16, 17 y 18, generados a partir de las tablas de hechos y las dimensiones comunes del DW.

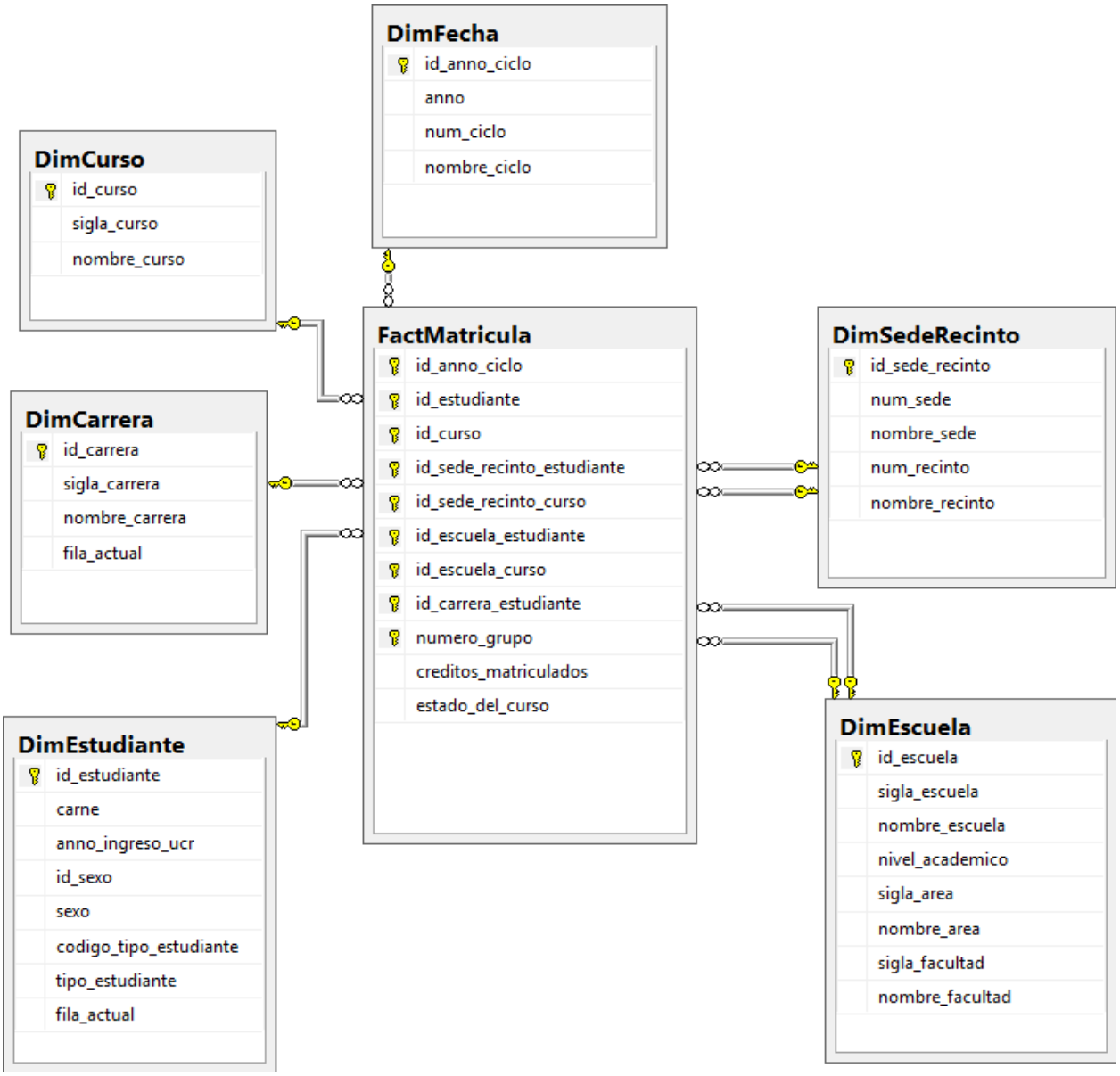


Figura 16. Diagrama de matrícula para el DW.
Fuente: Elaboración propia.

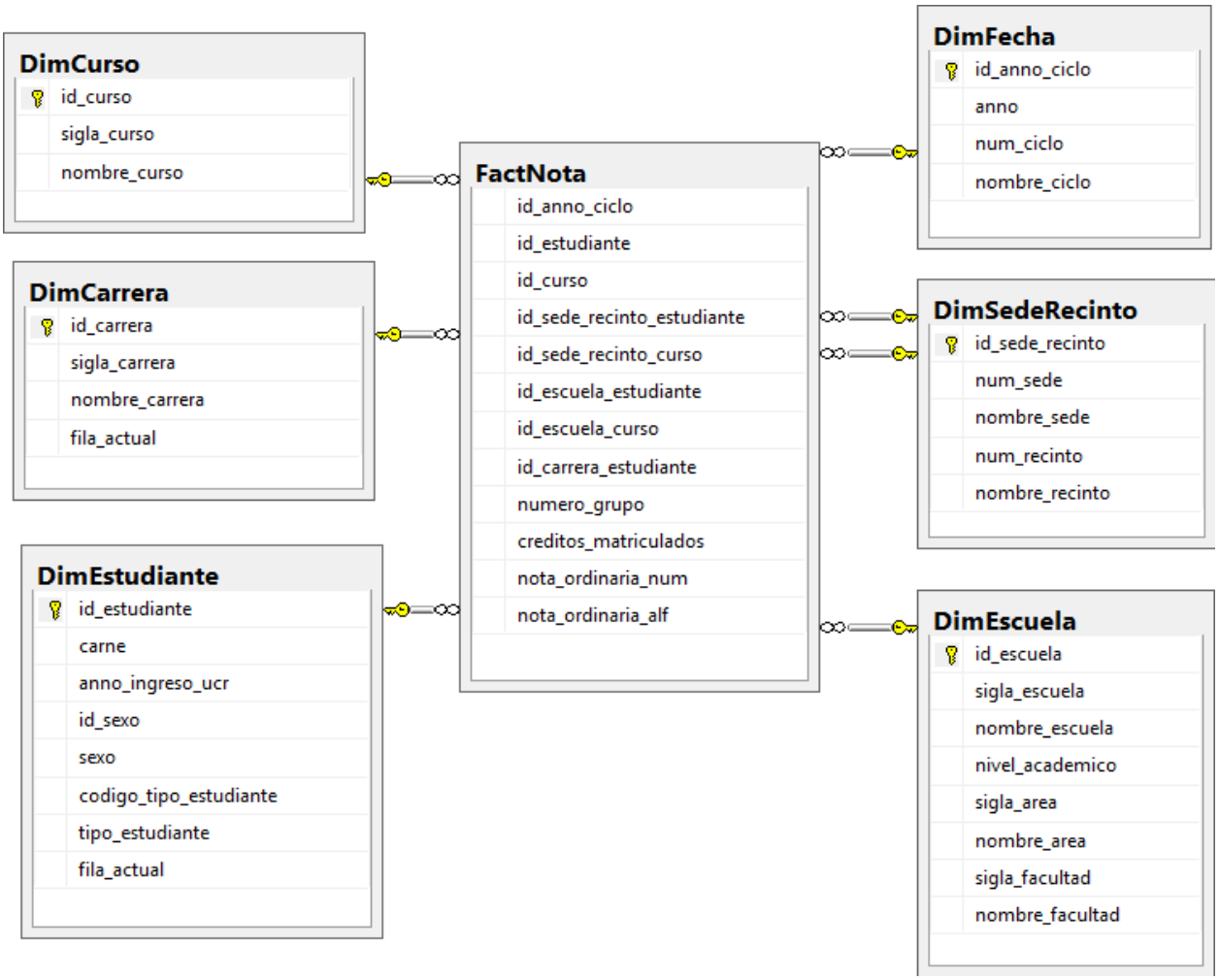


Figura 17. Diagrama de notas para el DW.
Fuente: Elaboración propia.

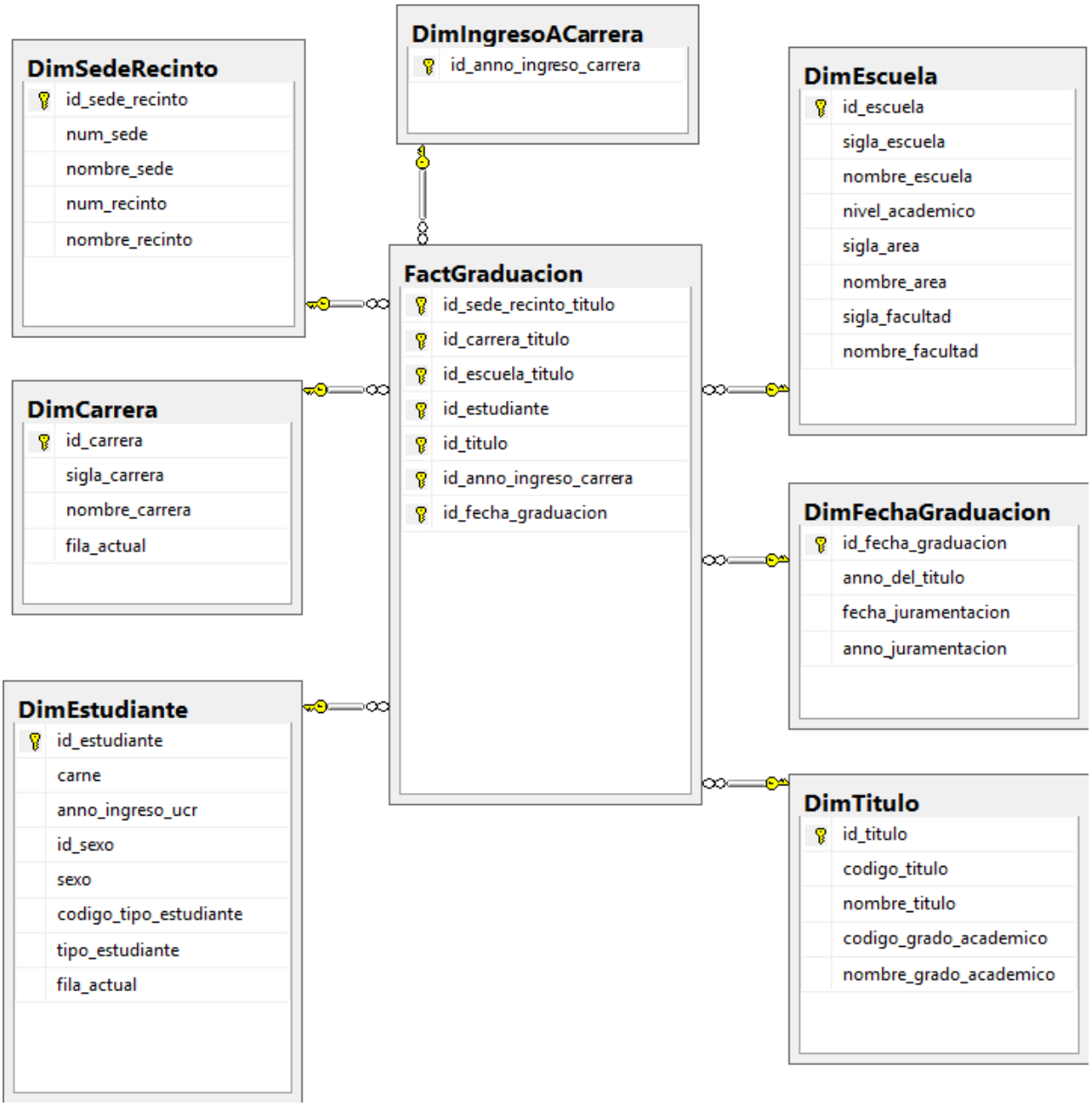


Figura 18. Diagrama de graduación para el DW.
Fuente: Elaboración propia.

I-02 Carga de los datos

Documento: Carga de datos al DW

La carga de los datos se realiza por medio de ETL, con la herramienta **SQL Server Data Tools**. Las figuras 19 y 20 muestran los procesos de carga de las tres tablas de hechos y de las dimensiones, incluyendo un ejemplo del detalle de la carga de la dimensión de estudiante en la figura 21 y de la dimensión de graduación en la figura 22.

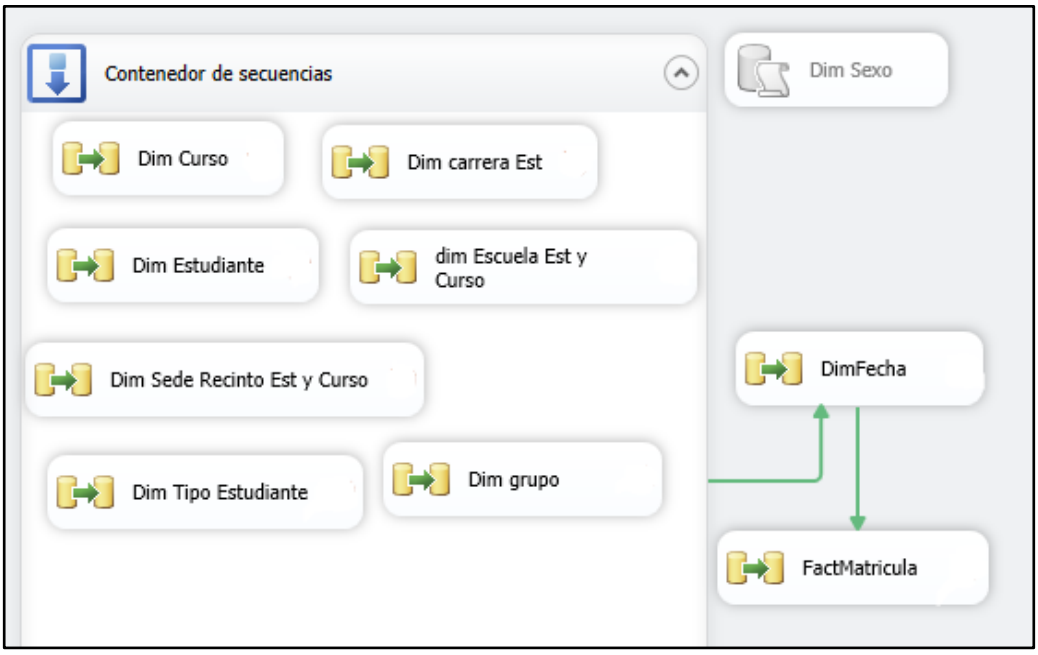


Figura 19. Carga de datos a las dimensiones y la tabla de hechos de matrícula.
Fuente: Elaboración propia.

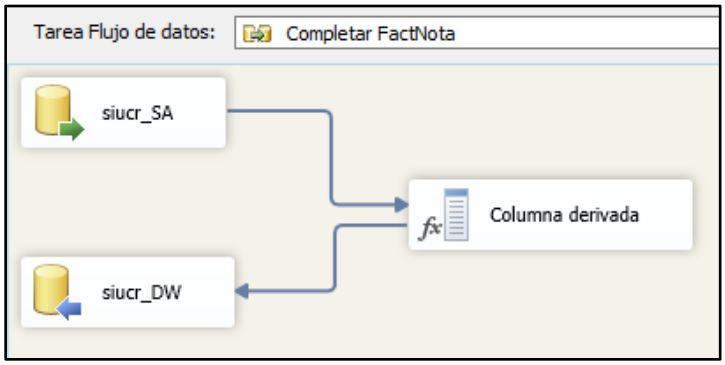


Figura 20. Carga de datos a la tabla de hechos de nota.
Fuente: Elaboración propia.

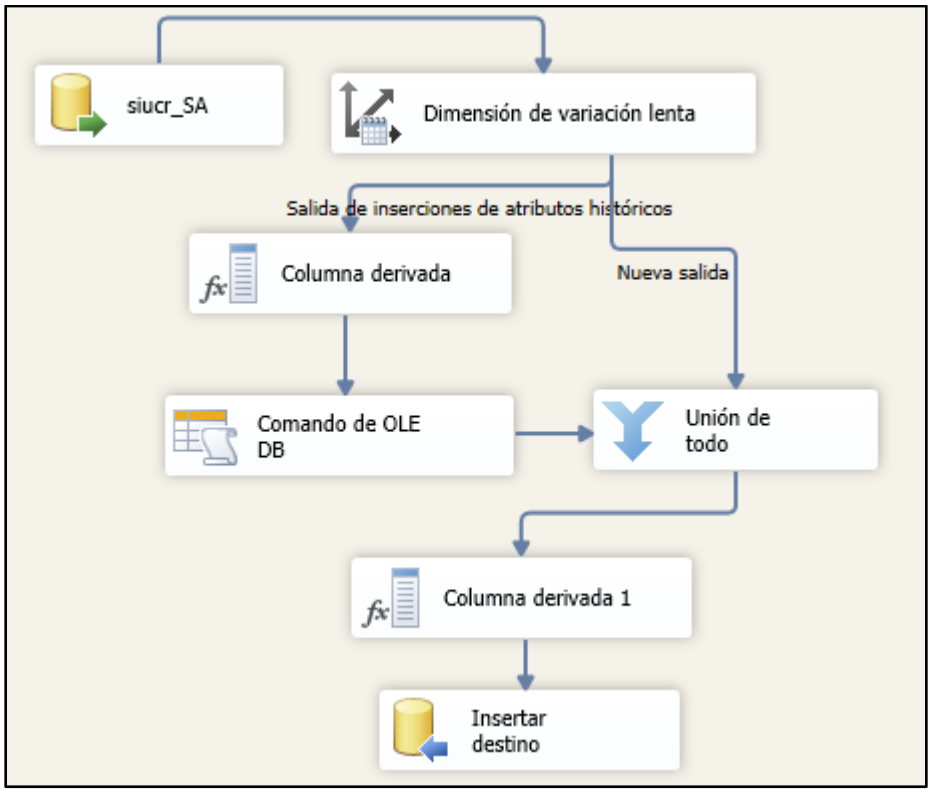


Figura 21. Carga de dimensión de estudiante del tipo de variación lenta.
Fuente: Elaboración propia.

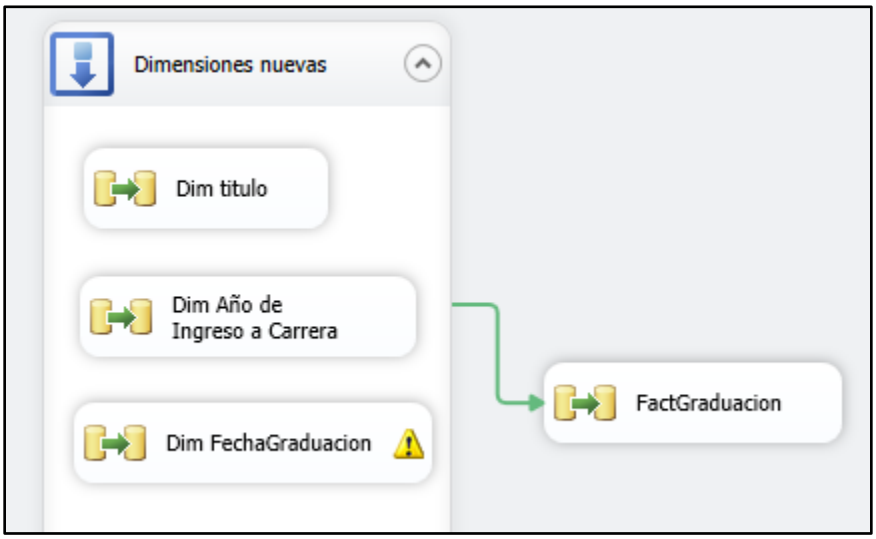


Figura 22. Carga de dimensiones y de la tabla de hechos de graduación.
Fuente: Elaboración propia.

I-03 Validar datos del DW

Documento: Resultados de consistencia de los datos

Validación de los datos de control contra la información obtenida del DW.

Totales Estudiantes Matriculados

Ciclo: 1 - 2013

Recinto	Nombre Recinto	Total Estudiantes Matriculados
	Total General:	**39.201**

Figura 23. Muestra de reporte de datos de control en el primer ciclo para el año 2013.
Fuente: Oficina de Registro e Información, UCR.

Totales Estudiantes Matriculados

Ciclo: 2 - 2013

Recinto	Nombre Recinto	Total Estudiantes Matriculados
	Total General:	**36.596**

Figura 24. Muestra de reporte de datos de control en el segundo ciclo para el año 2013.
Fuente: Oficina de Registro e Información, UCR.

Total de Estudiantes Físicos	Rótulos de columna	
Rótulos de fila	I Semestre	II Semestre
2013	39199	36594

Figura 25. Resultados del total de estudiantes físicos para el año 2013 para validación. Fuente: Elaboración propia.

La comparación de los datos de control de las figuras 23 y 24 contra los datos obtenidos en la figura 25, muestran una similitud importante, donde la diferencia de los datos es de 2 estudiantes. La base de datos de la fuente no almacena históricos, por lo que esta diferencia en los resultados es aceptable.

Se aplica otra comparación de los datos de control contra la información obtenida. En la figura 28, la información está dividida por nivel académico, por “pregrado y grado” y “posgrado”. Para este caso, la variación de los datos de pregrado y grado de la figura 27 es de 117 estudiantes, una variación importante que debe ser consultada con la fuente. Lo que respecta a los datos de posgrado en la figura 26, el valor es el mismo que los de la figura 28, por lo que para esta segregación la validación se da por aprobada.

Universidad de Costa Rica
Panorama Cuantitativo Universitario

Cuadro D5: Estudiantes físicos de posgrado, según materias matriculadas en el año, por especialidad. 2011

Especialidad	Materias Matriculadas	Total 1/		Menos de 5		de 5 a 9	
		abs.	%	abs.	%	abs.	%
Sistema de Estudios de Posgrado		4 042	100,00	2 780	68,78	1 057	26,15

Figura 26. Panorama Cuantitativo Universitario. Docencia 2011. Página 8. Fuente: Sitio oficial de la Oficina de Planificación Universitaria[21], UCR.

[21] *http://oplau.ucr.ac.cr/index.php/descargas/category/46-2011-panorama-unificado*

Unidades / Materias Aprobadas	Total 1/ abs.	Total 1/ %	Menos de 5 abs.	Menos de 5 %	de 5 a 9 abs.	de 5 a 9 %
Universidad de Costa Rica	36 417	100,00	11 317	31,08	11 446	31,43

Figura 27. Panorama Cuantitativo Universitario. Docencia 2011. Página 7.

Fuente: Sitio oficial de la Oficina de Planificación Universitaria[22], UCR

Anno	Nivel academico del estudiante	Estudiantes fisicos
2011	POSGRADO	4.042
	PREGRADO Y GRADO	36.300

Figura 28. Resultado de los cubos procesados en el DW.

Fuente: Elaboración propia con datos del DW.

Para efectos de esta documentación, no se presentan todas las validaciones realizadas a los datos, sino una muestra de cómo se aplican de forma manual las comparaciones de los datos agrupados.

[22] *http://oplau.ucr.ac.cr/index.php/descargas/category/46-2011-panorama-unificado*

Apéndice 4. Consulta de satisfacción del usuario

Índice

CUESTIONARIO APLICADO 197
FUNCIONARIO 1 199
FUNCIONARIO 2 201
FUNCIONARIO 3 203
FUNCIONARIO 4 205
TABULACIÓN DE RESPUESTAS 207

Cuestionario aplicado

Consulta al usuario sobre el almacén de datos con datos estudiantiles

La consulta se efectúa con el fin de valorar la satisfacción del usuario, y determinar el cumplimiento de los objetivos propuestos en la construcción del almacén de datos con respecto a la información relacionada de los estudiantes. La consulta consta de 8 preguntas.

Fecha: _____

Puesto dentro de la oficina: _____

1. Se brindó toda la información solicitada:

○ Sí.

○ Parcialmente.

○ No.

2. Se brindó información adicional para corroborar los datos (históricos, datos relacionados):

○ Sí.

○ Parcialmente.

○ No.

3. Las variables disponibles (desagregación de los datos) cumplen por completo lo solicitado:

○ Sí.

○ Parcialmente.

○ No.

4. Los datos se brindaron en el tiempo esperado:

○ Sí.

○ No.

5. Considera que el formato para brindar la información es el adecuado:

○ Sí.

○ Parcialmente.

○ No.

6. Considera que los datos son confiables y consistentes:

○ Sí.

○ Parcialmente.

○ No.

7. Se cumplió con la expectativa en cuanto a velocidad de respuesta y variedad de la información:

○ Sí.

○ Parcialmente.

○ No.

8. Comentarios (sobre la solución ofrecida o información que debe estar disponible):

_____

_____

_____

_____

_____

Funcionario 1

Consulta al usuario sobre el almacén de datos con datos estudiantiles

La consulta se efectúa con el fin de valorar la satisfacción del usuario, y determinar el cumplimiento de los objetivos propuestos en la construcción del almacén de datos con respecto a la información relacionada de los estudiantes. La consulta consta de 8 preguntas.

Fecha: 27/7/16

Puesto dentro de la oficina: Profesiona de Evaluación.

1. Se brindó toda la información solicitada:
☒ Completamente.
○ Parcialmente.
○ No está disponible.

2. Se brindó información adicional para corroborar los datos (históricos, datos relacionados):
☒ Sí.
○ Parcialmente.
○ No.

3. Las variables disponibles (desagregación de los datos) cumplen por completo lo solicitado:
○ Completamente.
☒ Parcialmente.
○ Faltan muchas variables.

4. Los datos se brindaron en el tiempo esperado:
☒ Sí.
○ No.

5. Considera que el formato para brindar la información es el adecuado:
☒ Sí.
○ Parcialmente.
○ No.

6. Considera que los datos son confiables y consistentes:
☒ Sí.
○ Parcialmente.
○ No.

7. Se cumplió con la expectativa en cuanto a velocidad de respuesta y variedad de la información:

Ø Sí.
O Parcialmente.
O No.

8. Comentarios (sobre la solución ofrecida o información que debe estar disponible):

La propuesta de solución cuenta con un alto nivel de usabilidad y ofrece una respuesta oportuna y fiable a la búsqueda de información institucional que normalmente está dispersa entre muchas oficinas de la Universidad

Funcionario 2

Consulta al usuario sobre el almacén de datos con datos estudiantiles

La consulta se efectúa con el fin de valorar la satisfacción del usuario, y determinar el cumplimiento de los objetivos propuestos en la construcción del almacén de datos con respecto a la información relacionada de los estudiantes. La consulta consta de 8 preguntas.

Fecha: 28-07-2016

Puesto dentro de la oficina: Profesional B

1. Se brindó toda la información solicitada:
- [X] Sí.
- [ ] Parcialmente.
- [ ] No.

2. Se brindó información adicional para corroborar los datos (históricos, datos relacionados):
- [ ] Sí.
- [ ] Parcialmente.
- [X] No.

3. Las variables disponibles (desagregación de los datos) cumplen por completo lo solicitado:
- [X] Sí.
- [ ] Parcialmente.
- [ ] No.

4. Los datos se brindaron en el tiempo esperado:
- [X] Sí.
- [ ] No.

5. Considera que el formato para brindar la información es el adecuado:
- [X] Sí.
- [ ] Parcialmente.
- [ ] No.

6. Considera que los datos son confiables y consistentes:
- [X] Sí.
- [ ] Parcialmente.
- [ ] No.

7. Se cumplió con la expectativa en cuanto a velocidad de respuesta y variedad de la información:

- [x] Sí.
- [ ] Parcialmente.
- [ ] No.

8. Comentarios (sobre la solución ofrecida o información que debe estar disponible):

Me parece excelente la información recibida pues que inmediata, con la estructura solicitada y además clara y concisa.

Funcionario 3

Consulta al usuario sobre el almacén de datos con datos estudiantiles

La consulta se efectúa con el fin de valorar la satisfacción del usuario, y determinar el cumplimiento de los objetivos propuestos en la construcción del almacén de datos con respecto a la información relacionada de los estudiantes. La consulta consta de 8 preguntas.

Fecha: 28/7/2016.

Puesto dentro de la oficina: Profesional C.

1. Se brindó toda la información solicitada:
O Sí.
⊗ Parcialmente.
O No.

2. Se brindó información adicional para corroborar los datos (históricos, datos relacionados):
⊗ Sí.
O Parcialmente.
O No.

3. Las variables disponibles (desagregación de los datos) cumplen por completo lo solicitado:
✓ Sí.
O Parcialmente.
O No.

4. Los datos se brindaron en el tiempo esperado:
✓ Sí.
O No.

5. Considera que el formato para brindar la información es el adecuado:
✓ Sí.
O Parcialmente.
O No.

6. Considera que los datos son confiables y consistentes:
✓ Sí.
O Parcialmente.
O No.

7. Se cumplió con la expectativa en cuanto a velocidad de respuesta y variedad de la información:

O Sí.

→ ⊗ Parcialmente.

⊗ No.

8. Comentarios (sobre la solución ofrecida o información que debe estar disponible):

No se pudo contar con información sobre las Carreras ofrecidas. al. I-2016, se recomienda. contar con esta información.

Funcionario 4

Consulta al usuario sobre el almacén de datos con datos estudiantiles

La consulta se efectúa con el fin de valorar la satisfacción del usuario, y determinar el cumplimiento de los objetivos propuestos en la construcción del almacén de datos con respecto a la información relacionada de los estudiantes. La consulta consta de 8 preguntas.

Fecha: 29-07-2016

Puesto dentro de la oficina: Asistente de investigación

1. Se brindó toda la información solicitada:

○ Sí.

X Parcialmente.

○ No.

2. Se brindó información adicional para corroborar los datos (históricos, datos relacionados):

X Sí.

○ Parcialmente.

○ No.

3. Las variables disponibles (desagregación de los datos) cumplen por completo lo solicitado:

○ Sí.

X Parcialmente.

○ No.

4. Los datos se brindaron en el tiempo esperado:

X Sí.

○ No.

5. Considera que el formato para brindar la información es el adecuado:

X Sí.

○ Parcialmente.

○ No.

6. Considera que los datos son confiables y consistentes:

○ Sí.

X Parcialmente.

○ No.

7. Se cumplió con la expectativa en cuanto a velocidad de respuesta y variedad de la información:

○ Sí.

X Parcialmente.

○ No.

8. Comentarios (sobre la solución ofrecida o información que debe estar disponible):

Hubo información que no estaba disponible para un porcentaje de las observaciones, se nos informó sobre esto y se tuvo en consideración. Por otro lado, la información fue de utilidad y la atención fue muy buena.

Tabulación de respuestas

La tabla 1 muestra la tabulación de las respuestas en orden de los cuestionarios aplicados. Para las preguntas de respuesta única se otorgan dos puntos por un “Sí”, un punto por “Parcialmente” y cero si la respuesta es “No”.

Núm. pregunta	Funcionario 1	Funcionario 2	Funcionario 3	Funcionario 4
#1	2	2	1	1
#2	2	0	2	2
#3	1	2	2	1
#4	2	2	2	2
#5	2	2	2	2
#6	2	2	2	1
#7	2	2	1	1
Total	**13**	**12**	**12**	**10**

Tabla 1. Tabulación de cuestionario de respuesta única. Fuente Elaboración propia.

Pregunta 8 – Comentarios

Se agregan las respuestas abiertas de la pregunta 8.

Funcionario 1	La propuesta de solución cuenta con un alto nivel de usabilidad y ofrece una respuesta oportuna y fiable a la búsqueda de información institucional que normalmente está dispersa entre muchas oficinas de la Universidad.
Funcionario 2	Me parece excelente la información recibida pues fue inmediata, con la estructura solicitada y además clara y concisa.
Funcionario 3	No se pudo contar con información sobre las carreras ofrecidas al I-2016, se recomienda contar con esta información.
Funcionario 4	Hubo información que no estaba disponible para un porcentaje de las observaciones, se nos informó sobre esto y se tuvo en consideración. Por otro lado, la información fue de utilidad.

Tabla 2. Respuestas abiertas. Fuente Elaboración propia.