

Universidad CENFOTEC

Maestría en Tecnologías de Bases de Datos

Tema:

Propuesta de un modelo de inteligencia de negocios para el Área de Cuenta Individual y Control de Pagos en la Gerencia de Pensiones de la Caja Costarricense del Seguro Social

Elaborado por:

Guideon Gabriel Bojorges Bolaños

Fecha: Agosto, 2019

Declaratoria de Derechos de autor

©2019, Bojorges Bolaños Guidion Gabriel

La presentación de este documento está sometida a procesos de confidencialidad. La vigencia de esta cláusula es por un periodo de: 5 año(s), a partir de la fecha de realizado el documento. Una vez pasado dicho periodo, el trabajo se pondrá a disposición de estudiantes, personal docente o administrativo, en la biblioteca de la Universidad, tanto para estudio, consulta o ejemplificación en clases relacionadas con el tema.

Dedicatoria y agradecimientos

Primeramente, a agradecer a Dios por darme las fuerzas para terminar el presente trabajo. A mí madre por apoyarme y darme ese sustento motivacional a lo largo de mi carrera. Agradecer también a mi compañero de trabajo el Ing. Pablo Rodriguez Guzmán y al Máster Eithel Corea Baltodano por brindarme la oportunidad de desarrollar el proyecto dentro de la institución y brindarme muchas facilidades para poder realizar y concluir el trabajo de la mejor forma.

Tabla de Contenido

Abstract 1

Capítulo 1. Introducción 3

1.1 Generalidades 3

1.2 Antecedentes del Problema 4

1.3 Definición y Descripción del Problema 5

1.4 Justificación 5

1.5 Viabilidad 6

1.5.1 Punto de Vista Técnico. 6

1.5.2 Punto de Vista Operativo. 7

1.5.3 Punto de Vista Económico. 8

1.6 Objetivos 9

1.6.1 Objetivo General. 10

1.6.2 Objetivos Específicos. 10

1.7 Alcances y Limitaciones 10

1.7.1 Alcances. 10

1.7.2 Limitaciones 11

1.8 Marco de Referencia Organizacional y Socioeconómico 12

1.8.1 Historia 12

1.8.3 Misión, Visión y Valores. 15

1.9 Estado de la Cuestión 16

Capítulo 2. Marco Conceptual 41

Capítulo 3. Marco Metodológico 48

3.1 Tipo de Investigación 48

3.2 Alcance Investigativo 48

3.3 Enfoque 49

3.4 Diseño 53

3.5 Población y Muestreo 53

3.6 Instrumentos de Recolección de Datos 54

3.7 Técnicas de Análisis de Información 54

3.8 Estrategia de Desarrollo de la Propuesta 54

Capítulo 4. Análisis del Diagnóstico 56

Capítulo 5. Propuesta de Solución 74

Capítulo 6. Conclusiones y Recomendaciones 115
6.1 Conclusiones 115
6.2 Recomendaciones 116
Capítulo 7. Reflexiones Finales 118
Capítulo 8. Trabajos a Futuro 119
Glosario 120
Referencias 122
Bibliografía 123
Apéndices 124

Índice de Figuras

Figura 1. Organigrama de Gerencia de Pensiones. Fuente: Despacho de Gerencia (2019). 14
Figura 2. Selección de estudio. Fuente: Propia. 26
Figura 3. Parámetros de búsqueda Springer. Fuente: Propia. 33
Figura 4. Conceptos comunes. Fuente: Propia. 41
Figura 5. Mapa mental de conceptos. Fuente: Propia. 42
Figura 6. Modelo estrella definido por Ralph Kimball. Fuente: (ALVI, 2019) 43
Figura 7. Vista de componentes Integration Services. Fuente: Propia. 44
Figura 8. Herramienta SSAS modelo tabular. Fuente: Propia. 46
Figura 9. Vista de Dashboard en Power BI. Fuente: Propia. 47
Figura 10. Ontología de modelo de inteligencia de negocios aplicado al problema de la analítica de datos ejecutados de forma manual. Fuente: Propia. 50
Figura 11. Caso uso en la generación de indicadores. Fuente: Propia. 64
Figura 12. Diagrama Ishikawa. Fuente: Propia. 69
Figura 13. Diagrama Ishikawa deficiencias generando indicadores. Fuente: Propia. 69
Figura 14. Imagen de servidor de prueba OLTP y Analisys Services. Fuente: Propia. 72
Figura 15. Consulta Transact-SQL de Cotizantes con edad menor a 18 y mayores a 65 años. Fuente: Propia. 77
Figura 16. Consulta SQL Trabajadores independientes que no aparecen cotizando para el seguro de pensiones IVM. Fuente: Propia. 78
Figura 17. Tabla Patronos en StagingArea. Fuente: Propia. 80
Figura 18. Diseño tabla hechos StagingArea. Fuente: Propia. 82

Figura 19. Modelo estrella DataWarehouse. Fuente: Propia. 84
Figura 20. Vista detalle tabla hechos. Fuente: Propia. 85
Figura 21. Proyecto Integration Services. Fuente: Propia. 88
Figura 22. Flujo de control carga de dimensiones StagingArea. Fuente Propia. 89
Figura 23. Estructura procedimiento almacenado inicializa dimensiones StagingArea. Fuente: Propia. 89
Figura 24. Vista Parcial de la carga StagingArea. Fuente: Propia. 90
Figura 25. Vista parcial del procedimiento carga hechos StagingArea. Fuente: Propia. 91
Figura 26. SQL de carga hechos OLTP a StagingArea. Fuente: Propia. 92
Figura 27. Carga de indicador pensionado por RNC o Ley8769. Fuente: Propia. ... 93
Figura 28. Flujo de control carga dimensiones Data Warehouse. Fuente Propia. ... 94
Figura 29. Vista parcial carga dimensión estado de cotizaciones y clase de seguro. Fuente: Propia. 94
Figura 30. Carga dimensión patronos Data Warehouse. Fuente: Propia. 95
Figura 31. Carga dimensión personas Data Warehouse. Fuente: Propia. 96
Figura 32. Consulta SQL carga hechos Data Warehouse. Fuente: Propia. 97
Figura 33. Inserción de valores desconocidos en dimensión estado de cotizaciones. Fuente: Propia. 98
Figura 34. Procesamiento de carga en tabla de hechos, de enero 2019 a junio 2019. Fuente: Propia. 99
Figura 35. Vista diagrama de modelo en Analisys Services. Fuente: Propia. 100
Figura 36. Procesamiento de datos en Analisys Services. Fuente: Propia. 101
Figura 37. Vista el cubo analítico OLAP. Fuente: Propia. 103
Figura 38. Diseño de Dashboard Power BI. Fuente: Propia. 104

Figura 39. Dashboard de Asalariados reportados con montos inferiores a la Base Mínima Contributiva. Fuente: Propia. 105
Figura 40. Dashboard Cotizantes reportados con salarios mayores a N millones. Fuente: Propia..... 106
Figura 41. Dashboard Cotizantes que registran cuotas posteriores a una fecha de defunción. Fuente: Propia. 108
Figura 42. Dashboard Trabajadores independientes que no aparecen cotizando para el seguro de pensiones IVM. Fuente: Propia. 109
Figura 43. Filtros a nivel de página para el indicador de los TI que no cotizan IVM. Fuente: Propia..... 110

Índice de tablas

Tabla 1. Costo teórico del diseño. Fuente: Propia. 9
Tabla 2. Etapas de una revisión sistemática. (Kitchenham, 2004) 18
Tabla 3. Plantilla de Bolchini para Revisión Sistemática (Biolchini, 2005) 19
Tabla 4. Palabras clave de la revisión sistemática. Fuente: Propia. 22
Tabla 5. Formulario para la extracción de información. Fuente: Propia. 29
Tabla 6. Porcentajes de evaluación del enfoque. Fuente: Propia. 52
Tabla 7. Resultado entrevista personal de TI 56
Tabla 8. Resultado entrevista al personal del Área de Cuenta Individual y Control de Pagos 58
Tabla 9. Tareas y tiempos en la generación de indicadores. Fuente: Propia. 66
Tabla 10. Costo aproximado en la generación de indicadores actualmente. Fuente: Propia. 68
Tabla 11. Plataforma tecnológica actual Gerencia Pensiones. Fuente: Administradores de Base de Datos. 71

Tabla 12. Etapas del desarrollo del modelo inteligencia de negocios. Fuente: Propia. 75

Tabla 13. Detalle de cualidades en tabla hechos. Fuente: Propia. 86

Tabla 14. Descripción sobre las dimensiones. Fuente: Propia. 87

Tabla 15. Ventajas del modelo propuesto vs actual. Fuente: Propia. 111

Tabla 16. Análisis de variables impacto modelo propuesto de inteligencia de negocios. Fuente: Propia. 113

Tabla 17. Clasificación del modelo según impacto de variables. Fuente: Propia. .. 114

Abstract

El Área de Cuenta Individual y Control de Pagos es un departamento que pertenece a la dirección de Administración de Pensiones. Está dirección administrativa es una de las cinco direcciones que componen la Gerencia de Pensiones de la Caja Costarricense del Seguro Social.

Esta Gerencia se creó de conformidad con el artículo 15° de la Ley Constitutiva, el 23 de mayo de 1995 y surge debido a la necesidad de independizar la administración de las pensiones del seguro de Invalidez, Vejez y Muerte de las prestaciones médico-hospitalarias a cargo del seguro de Enfermedad y Maternidad.

Para cumplir su cometido se requiere que dicha gerencia tenga la suficiente capacidad administrativa y resolutiva en el desarrollo de la gestión encomendada para lo cual adoptará mecanismos eficientes que le permitan utilizar racionalmente los recursos financieros provenientes de las obligaciones que aportan los trabajadores, los patronos y el estado, así como el manejo directo de su cartera de inversiones, que le aseguren mayores rendimientos del mercado financiero con los menores riesgos posibles.

La Gerencia de Pensiones brinda servicio a una cantidad diversa de población, desde niños huérfanos hasta adultos mayores dado a los diferentes riesgos que abarca el sistema de pensiones (IVM-RNC). Toda la información que se genera a partir de las solicitudes de los usuarios es almacenada en su sistema transaccional lo cual le permite llevar un seguimiento y generar ciertas estadísticas.

Por su parte el Área de Cuenta Individual y Control de Pagos, dentro de esta gestión en Gerencia de Pensiones, se encarga del registro de las cuotas aportadas por los diferentes regímenes (Ministerio de Hacienda, Magisterio, entre otros) los estudios de cotización en el nivel nacional, la distribución a las entidades pagadoras

y generación de las planillas de pago del Régimen de Invalidez, Vejez y Muerte (IVM), Régimen No Contributivo (RNC) y el Fondo de Retiro de Empleados de la CCSS (FRE) del sistema de pensiones.

Es a razón de lo anterior, que el proyecto en sí propone un diseño de inteligencia de negocios, el cual le permita al Área de Cuenta Individual y Control de Pagos centralizar la información en un único repositorio y contar con una herramienta especializada para visualizar esta información, que le brinde apoyo en el estudio de los datos, mediante la aplicación de métodos de análisis que le ayudan a transformar el dato en información y la información en conocimiento para la creación de estrategias efectivas en la toma de decisiones.

Palabras Clave: Inteligencia, modelo, negocios, análisis, herramientas, plataforma, Power BI.

Capítulo 1. Introducción

1.1 Generalidades

La Gerencia de Pensiones como un ente público de la República de Costa Rica debe garantizar la protección económica y prestaciones sociales a la población asegurada del régimen IVM y del Régimen No Contributivo, es por esto que la administración debe apoyarse en la utilización de estrategias efectivas que generen confianza en la toma de decisiones y le permitan asegurar la sostenibilidad del sistema de pensiones.

El Área de Cuenta Individual y Control de Pagos como parte fundamental de esa garantía, debe contar con recursos analíticos suficientes para lograr estrategias que apoyen la dirección del negocio. Los procesos esenciales que se ejecutan dentro del área como lo es el estudio de cuotas para determinar los derechos de los asegurados para el Régimen de IVM, la generación de las planillas, así como la distribución de estas a las entidades pagadoras, son algunas de las actividades sustantivas que generan información valiosa, a fin de lograr análisis avanzados en función de sustentar la necesidad del negocio en la toma de decisiones.

Los encargados de la plataforma tecnológica de la Gerencia de Pensiones no tienen definido en el corto plazo llevar a cabo una implementación de inteligencia de negocios dentro del área.

El proyecto al tratar con datos públicos y por lineamientos de la Caja Costarricense del Seguro Social en esta materia, prohíbe la exposición de los datos de cualquier persona mediante un contrato de confidencialidad que es obligatorio para la gestión que involucre acceso a la información, por cuanto aquella información sensible que se utilice producto del desarrollo del modelo de inteligencia

de negocios será previamente transformada a fin de resguardar los datos de los usuarios.

1.2 Antecedentes del Problema

Actualmente en el Área de Cuenta Individual y Control de Pagos no es posible generar reportes avanzados y que requieran cruzar información desde diferentes perspectivas y que involucren distintas líneas de tiempo de forma eficiente. Esto por cuanto no se cuenta con un modelo de inteligencia de negocios que apoye este proceso dentro de sus actividades.

Para generar está información recurren a reportes estáticos que se encuentran alojados dentro de la aplicación o bien dentro una ruta del servidor de reportes “Reporting Service”, una vez obtenida la información deben cotejarla manualmente contra archivos en Excel, con el fin de validar los datos y generar los informes para presentarlos a las autoridades correspondientes. Dependiendo la cantidad y el cotejo de información que deban realizar las tareas pueden durar desde 1 hora hasta varias, sino días dependiendo de las capacidades de las personas.

El Área de Cuenta Individual y Control de Pagos necesita mejorar está gestión a fin de lograr una mayor eficacia en cuanto a la cantidad de tareas y tiempo que demoran en la generación de esta información. Asimismo, desean lograr un mayor alcance de análisis con las diferentes variables que conocen y además lograr oportunidades de estudio con nuevas variables que un modelo de inteligencia de negocios podría aportarles.

1.3 Definición y Descripción del Problema

El Área de Cuenta Individual y Control de Pagos como cualquier otro departamento necesita comprender, explorar y generar conocimiento a partir de sus datos.

Durante el desarrollo del trabajo, se logró constatar que en el Área de Cuenta Individual y Control de Pagos de la Gerencia de Pensiones, no existe un modelo de inteligencia de negocios que les permita generar conocimiento más preciso y tener un estudio más profundo sobre sus datos, en su lugar utilizan para crear estudios estadísticos reportes estáticos obtenidos directamente de su sistema transaccional o del servidor de reportes “Reporting Services”, los cuales en su mayoría están segmentados en muchos lugares del sistema o del servidor de reportes. Al ejecutar el proceso de esta forma, ya sea de forma diaria, semanal o mensual se incurre en una tarea manual sumamente compleja, imprecisa, y costosa en recurso y procesamiento.

1.4 Justificación

Al desarrollar un modelo de inteligencia de negocios en el Área de Cuenta Individual y Control de Pagos, permitió tener al alcance una tecnología de vanguardia para el tratamiento y procesamiento de datos, con el fin de que, a partir de esta, se pueda asegurar una comprensión más amplia sobre la información que generan sus procesos, de modo que las estrategias que se desarrollen basadas en los estudios analíticos permitan mejorar su efectividad.

La investigación realizada, expone una solución especializada para ejecutar el análisis de datos de una forma especializada mediante la utilización de herramientas tecnológicas para este fin, mediante un modelo de inteligencia orientado al análisis que genere conocimiento y este conocimiento pueda ser

utilizado para trazar metas y objetivos de mayor impacto en los servicios que presta el Área de Cuenta Individual y Control y Pago. Esta área al tener una mayor visión sobre la información que administra puede optimizar los recursos económicos y esfuerzos para fortalecer programas, desarrollar nuevos o bien ajustar reglamentos.

El diseño del modelo de inteligencia de negocios permitió al usuario generar la información a partir de una estructura previamente perfilada en sus datos, desde un solo punto y en la forma y características que el mismo desea. De esta forma se mejora el uso del tiempo y se obtiene información relevante para mejorar la eficiencia el estudio de la información que generan sus procesos.

1.5 Viabilidad

1.5.1 Punto de Vista Técnico.

El Área de Cuenta Individual y Control de Pagos cuenta con las aplicaciones requeridas para diseñar el modelo de inteligencia de negocios como lo es: Microsoft SQL Server, Microsoft SQL Server Analysis Services, Microsoft SQL Server Integration Service en sus versiones del 2016, suscripciones de office 365. Para la herramienta Power BI los funcionarios de tecnologías de la Gerencia de Pensiones están gestionando la licencia en versión Pro ante la Dirección de Tecnologías de Información que es el órgano regulador del software institucional, pero para efectos de la propuesta del modelo de Inteligencia de Negocios se puede utilizar la versión de office 365 que ya poseen.

Adicionalmente se cuenta con las capacidades técnicas para diseñar el modelo basado en los siguientes puntos:

- Experiencia laboral de más de 5 años, con la administración y desarrollo de base de datos relacionales, como SQL Server.
- Participación en proyectos de migración de datos.

- Desarrollo de estrategias para mejorar rendimientos de consultas en las bases de datos.
- Aplicación de mejores prácticas en el diseño y programación de las bases de datos.
- Capacitaciones en SQL Server y Power BI.
- Lecturas de documentos de apoyo para obtener la certificación de Microsoft Certified Solutions Associate (MCSA).
- Conocimiento de soluciones de inteligencia de negocios obtenido a través de los diferentes cursos impartidos en la maestría en Tecnologías de Bases de datos impartido por la Universidad Cenfotec.

Además, se cuenta anualmente con un programa de cursos en el centro de trabajo para reforzar conocimientos en el manejo de los datos y procesos relacionados.

1.5.2 Punto de Vista Operativo.

La propuesta del modelo de inteligencia de negocios se pudo realizar sin alterar el funcionamiento normal del Área de Cuenta Individual y Control de Pagos, esto por cuanto se puede construir a partir en un ambiente de pruebas con las siguientes características:

- Requerimiento mínimo de intervención por parte de los usuarios dado a que una vez definida la necesidad de los indicadores, se puede trabajar en la construcción del modelo de base de datos analítico.
- El diseño del modelo de inteligencia de negocio conlleva una serie de características técnicas que pueden ser consultadas en internet sin necesidad del usuario final.
- Los jefes del Área de Cuenta Individual y Control de Pagos apoyan el desarrollo del diseño para el modelo de inteligencia de negocio.

- Existe apoyo y anuncia por parte del jefe de informática de la Gerencia de Pensiones para desarrollar el proyecto de inteligencia de negocios.
- Se puede trabajar con una versión de respaldo anterior de una base de datos de producción en un ambiente de pruebas como parte de la investigación del diseño de inteligencia de negocios.
- La visualización de la información con la herramienta de Power BI puede realizarse desde cualquier lugar y momento en que el usuario necesite accederla.
- Se cuenta con el espacio físico personal adecuado para desarrollar la investigación.
- Se cuenta con acceso un computador y acceso a internet para aclarar cualquier termino producto del desarrollo de la investigación.

1.5.3 Punto de Vista Económico.

El recurso humano destinado a realizar una propuesta de un modelo de inteligencia de negocios no necesita de una inversión monetaria dado a que este costo es un servicio que no se va a cobrar.

El costo teórico se puede visualizar en la siguiente tabla basada en el salario promedio de un consultor de inteligencia de negocios en Costa Rica actualmente de ¢1.800.000,00:

Costo hora	Total de horas	Rubro	Costo Total
¢7.500,00	80	Identificación de los de los indicadores	¢600.000,00
¢7.500,00	40	Análisis de los indicadores	¢300.000
¢7.500,00	40	Perfilado de datos (depuración)	¢300.000
¢7.500,00	24	Diseño de Staging Area	¢180.000,00
¢7.500,00	40	Diseño de Data Warehouse	¢300.000
¢7.500,00	80	Procesos ETL de OLTP a Staging Area	¢600.000,00
¢7.500,00	80	Procesos ETL de Staging Area a Data Warehouse	¢600.000,00
¢7.500,00	80	Creación del cubo multidimensional	¢600.000,00
¢7.500,00	24	Ejecución de pruebas con herramienta Power BI	¢180.000,00
¢7.500,00	80	Creación de Dashboard en Power BI	¢600.000,00
Totales			
	568 horas		¢4.260.000,00

Tabla 1. Costo teórico del diseño. Fuente: Propia.

1.6 Objetivos

Se utiliza la taxonomía de Benjamín Bloom de 1956 dado a que, según la situación para desarrollar el tema, es la que más se ajusta al enfoque para lograr cada uno de los objetivos específicos además de ser un estándar en el nivel académico costarricense.

1.6.1 Objetivo General.

Proponer un modelo de inteligencia de negocios que permita generar información con múltiples dimensiones y en diferentes temporalidades para el análisis avanzado de la información.

1.6.2 Objetivos Específicos.

- Comprender el comportamiento de los datos en el Área de Cuenta Individual y Control de Pagos.
- Identificar los indicadores del Área de Cuenta Individual y Control de Pagos que generen alto impacto en la organización.
- Crear un sistema de almacenamiento de datos (Data Warehouse).
- Desarrollar los ETL para la extracción, carga y transformación de los datos al Data Warehouse y Staging Area.
- Construir mediante Power BI los principales Dashboard que el Área de Cuenta Individual y Control de Pagos necesita.

1.7 Alcances y Limitaciones

1.7.1 Alcances.

- El proyecto abarca solamente el Área de Cuenta Individual y Control de Pagos perteneciente a la Dirección de Administración de Pensiones de la Gerencia de Pensiones.
- Creación de un almacén de datos (Data Warehouse) y Staging Area en un ambiente de pruebas.
- Extracción, transformación y carga de la información del sistema OLTP, hacia el Staging Area y Data Warehouse.

- Procesos de optimización y mejores prácticas en el diseño del Data Warehouse.
- Creación del cubo analítico OLAP en ambiente pruebas.
- Brindar una charla básica sobre el uso general del Power BI a los usuarios expertos, jefa de Cuenta Individual y Control de Pagos y al director de la dirección involucrada.
- Entregar documentación digital relacionada con el uso de la herramienta Power BI para que los encargados puedan profundizar en el uso de la herramienta y puedan generar más reportes según requieran.

1.7.2 Limitaciones

- La adquisición de la licencia de Power BI PRO queda a responsabilidad de la organización.
- No se brindarán consultorías sobre la adquisición de licencias adicionales de software relacionado al proyecto.
- La adopción del diseño queda a criterio del negocio.
- Capacitación técnica de la implementación.
- Para el desarrollo de los Dashboard se utilizarán los indicadores más relevantes que el negocio defina de modo que el diseño conserve armonía y no se sobre cargue de información en la visualización de estos.
- No se desarrollará un modelo de minería de datos, aunque con el análisis realizado y la estructura creada se podría implementar fácilmente de acuerdo con las necesidades del negocio.
- La capacitación del resto del personal sobre la herramienta de Power BI es responsabilidad del negocio.

- La charla que se brindará es introductoria y no contempla el uso avanzado de la herramienta Power BI.

1.8 Marco de Referencia Organizacional y Socioeconómico

1.8.1 Historia

El primero de noviembre de 1941, mediante Ley N°17 nace la Caja Costarricense de Seguro Social, creada con la idea de establecer los seguros sociales y brindar protección económica a la mayor cantidad posible de trabajadores, durante la administración del Dr. Rafael Ángel Calderón Guardia.

En 1960 que el Seguro de I.V.M. amplió su cobertura a empleados del comercio, escuelas de enseñanza particular, consultorios profesionales y trabajadores municipales pagados por planillas de jornales. En 1962 se amplió a trabajadores manuales ocasionales (construcción), a los pagados por planillas de jornales en obras públicas, ferrocarriles y ya para 1971 cubre en general a todos los obreros del país.

El 12 de mayo de 1961 por Ley Nº 2738, se faculta a la C.C.S.S a la Universalización de los Seguros Sociales.

En 1975 se extiende el Seguro de Invalidez, Vejez y Muerte a los trabajadores del campo (agrícola) y la C.C.S.S. se hace cargo del Sistema de Pensiones del Régimen No Contributivo, esto con el fin de dar protección a los de más bajos recursos.

Con el paso del tiempo se determina que el crecimiento del Seguro de Enfermedad y Maternidad – hoy Seguro de Salud – no fue paralelo al crecimiento del Seguro de Pensiones, este último no alcanzó la proyección social y el desarrollo

de la autonomía que le permitiera responder a las nuevas necesidades de los asegurados.

Se puede afirmar desde un punto de vista organizativo que durante casi cinco décadas el Seguro de Invalidez, Vejez y Muerte funcionó como un anexo al Seguro de Salud y en consecuencia se adaptó a sus procedimientos, sistemas y formas, pero no siempre eran los más apropiados para la gestión de los pensionados.

El 23 de mayo de 1995 la Junta Directiva de la Institución, presidida por el Doctor Álvaro Salas Chaves, en la sesión 6926 acordó crear la Gerencia de Pensiones, con el propósito de unir a todas las áreas involucradas e integrar el Sistema de Pensiones de la Caja Costarricense del Seguro Social.

Con el nacimiento de la Gerencia de Pensiones todas aquellas unidades que venían trabajando con pensiones dispersas, en diversas independencias, se integran y ahora trabajan bajo una sola dirección, para el bienestar del asegurado y del pensionado, por lo que se convierte en una marcada filosofía de servicio.

1.8.2 Tipo de Negocio y Mercado Meta.

La Gerencia de Pensiones es la encargada de garantizar la protección económica y las prestaciones sociales a la población asegurada en el Régimen de Invalidez, Vejez y Muerte (IVM) y a los beneficiarios de Régimen No Contributivo (RNC), mediante una administración efectiva que genere rentabilidad financiera que asegure la sostenibilidad del sistema de pensiones. La población que atiende va desde huérfanos, indigentes, personas con problemas de parálisis cerebral profunda (PCP) así como personas adultas mayores del seguro de Vejez, discapacitados del seguro de Invalidez y personas de diferentes edades por el seguro de Muerte (esposa, hijos). La Gerencia actualmente está conformada por 5 direcciones que

brindan los diferentes servicios dentro de la institución a fin de garantizar la sostenibilidad de régimen de pensiones y prestaciones sociales. La siguiente figura ilustra el organigrama de la Gerencia de Pensiones:

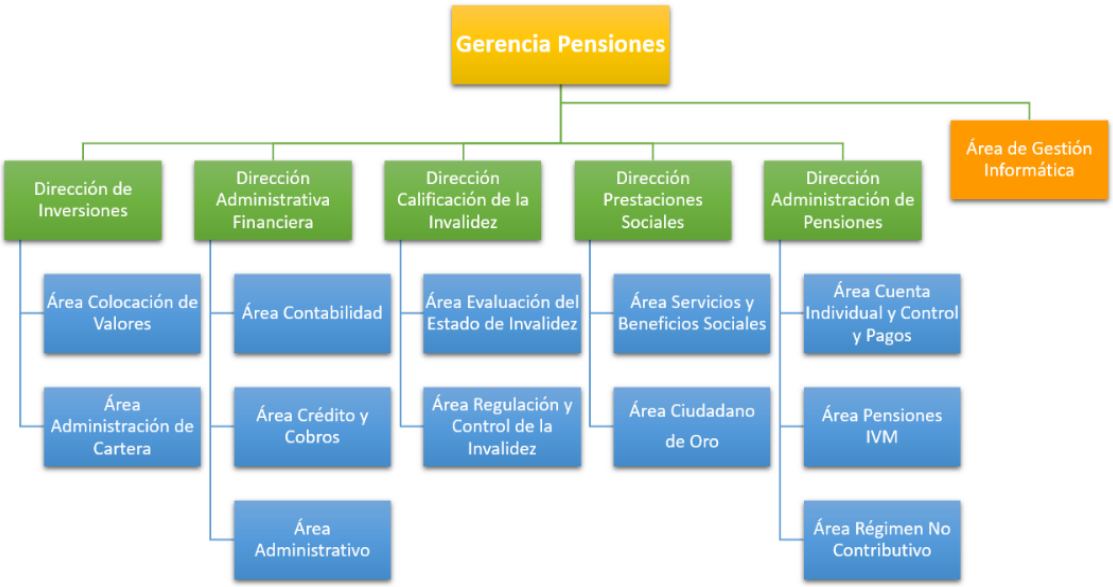


Figura 1. Organigrama de Gerencia de Pensiones. Fuente: Despacho de Gerencia (2019).

Como se puede observar en la imagen anterior el Área de Cuenta Individual y Control de Pagos pertenece a la Dirección de Administración de Pensiones, está área en realidad está compuesta por dos subáreas las cuales se denominan:

- Subárea Administración de la Cuenta Individual.
- Subárea Pagos y Control de Pagos.

Los procesos sustantivos de cada subárea están marcados según un alcance establecido por la administración.

La subárea de Cuenta Individual realiza entre otros procesos:

- Registrar las cuotas aportadas por otros regímenes.
- Realizar estudios correspondientes para reconocer y acreditar las cuotas pagadas por los asegurados y no incluidas en el sistema.

- Cuantificar la deuda patronal, mediante estudios específicos de las empresas, instituciones públicas y privadas.

La subárea de Pagos y Control realiza entre otros procesos:

- Generar la planilla ordinaria y adicional.
- Distribuir la planilla y los medíos magnéticos o digitales a las entidades pagadoras del sistema.
- Registrar los datos y digitar los cálculos de pensión con base en las tasas y porcentajes definidos.

1.8.3 Misión, Visión y Valores.

Misión:

"Proteger económicamente y otorgar prestaciones sociales a la población asegurada, por medio del Seguro de Pensiones y del Régimen No Contributivo, para mejorar la calidad de vida de los pensionados y la solidez económico financiero del sistema."

Visión:

"Seremos el Sistema de Pensiones con la mayor y más amplia cobertura de la población económicamente activa, con recursos humanos comprometidos, y el desarrollo de procesos eficientes, eficaces, de calidad y con un enfoque de servicio al usuario".

Valores:

Calidad: es un proceso de mejoramiento continuo, en la cual todas las áreas de la Gerencia participan en el desarrollo de los servicios, con la finalidad de

satisfacer oportunamente las necesidades y los requerimientos de los usuarios y lograr una mayor productividad.

Oportunidad: las necesidades de los usuarios se deben atender en el lugar y en el momento conveniente.

Humanización: es la orientación de los funcionarios de la Institución, que se manifiesta con el buen trato, la cortesía, la amabilidad, la consideración y el respeto en la prestación de los servicios, que permita la solución efectiva de las necesidades y los requerimientos individuales y colectivos de la población.

Transparencia: las actuaciones de los funcionarios en cualquier asunto

1.8.4 Políticas Institucionales.

No existen políticas relacionadas al desarrollo de la investigación, solamente existe por normativa la conformación de un contrato de confidencialidad que es obligatorio para la gestión que involucre acceso a la información, el cual será adjuntado al presente trabajo en los apéndices.

1.9 Estado de la Cuestión

El estado de la cuestión, a veces llamado estado del arte es un estudio que se hace para determinar que tanto se ha desarrollado un tema. Según Aristóteles, la Ciencia busca satisfacer la curiosidad innata del ser humano, en cambio el arte busca la aplicación del conocimiento. Así es como por ejemplo el arte según su concepción o enfoque abarca temas como la medicina, ingenierías, música, poesía y escultura, entre otros.

Para determinar el estado de la cuestión es necesario realizar un proceso de revisión sistemático de modo que permita extraer hallazgos a partir criterios de

búsqueda (inclusión y exclusión), escogencia de repositorios y patrones de calidad que generen un conocimiento robusto en procura de fortalecer el trabajo investigativo.

1.9.1 Revisión sistemática

Elaborar una revisión de la literatura existente mediante el enfoque de una revisión sistemática permite identificar, evaluar e interpretar todos los estudios importantes o significativos (llamados estudios primarios) para una pregunta de investigación en particular. De este modo se encuentra con varias razones que remarcan la utilidad de usar este enfoque:

- Resume la evidencia existente concerniente a un tratamiento o tecnología.
- Identifica brechas en la investigación actual para sugerir áreas de ulterior investigación.
- Proporciona un marco para posicionar apropiadamente nuevas actividades de investigación.
- Define la elaboración de un marco teórico o un marco conceptual.
- Enriquece el contenido investigativo, al aportar diferentes enfoques y perspectivas del tema.

El método y las etapas de revisión sistemática que se utilizará será el sugerido por Barbara Kitchenham (2004) y como guía para desarrollar el protocolo de investigación se utilizará la plantilla definida por Biolchini, Gomes, Cruz y Horta (2005) en el contexto de la ingeniería de software.

Etapa 1 **Planificación de la revisión**	Identificar la necesidad de la revisión. Desarrollo de un protocolo de revisión.
Etapa 2 **Desarrollo de la revisión**	Identificación de la investigación Selección de los estudios primarios Evaluación de la calidad del estudio Extracción y monitoreo de datos Síntesis de datos
Etapa 3 **Publicación de resultados**	Informe y validación de resultados

Tabla 2. Etapas de una revisión sistemática. (Kitchenham, 2004)

Planificación de la revisión	Formulación de la pregunta • Foco de la pregunta • Amplitud y calidad de la pregunta
Desarrollo o ejecución de la revisión	Selección de fuentes • Definición del criterio de selección de fuentes • Lenguaje de estudio • Identificación de fuentes • Selección de fuentes después de la evaluación • Comprobación de las fuentes Selección de los estudios • Definición del criterio de inclusión y exclusión de estudios

	• Definición de tipos de estudio • Procedimiento para la selección de los estudios
Análisis de resultados	Resumen de resultados

Tabla 3. Plantilla de Bolchini para Revisión Sistemática (Biolchini, 2005)

Una vez conocidas las etapas y los métodos que se utilizarán para confeccionar la revisión sistemática, en las siguientes secciones del documento se iniciarán con la etapa de planificación de la revisión para posteriormente continuar con el desarrollo del protocolo basado en las fuentes seleccionadas y finalizar con el resumen de resultados.

1.9.1.1 Planificación de la Revisión

En esta etapa se identifica la necesidad de revisión, basado en los objetivos.

Se destacan las siguientes características de esta planificación:

- Qué fuentes se utilizarán para identificar los estudios primarios.
- Si hubo restricciones.
- Cuáles son los criterios de inclusión y exclusión.
- Qué criterios se utilizarán para evaluar la calidad de los estudios primarios.
- Cómo se extraerán y se resumirán la información de los estudios.

Formulación de la pregunta

Foco de la pregunta

En esta revisión sistemática se pretende localizar documentos de trabajos que realicen contribuciones importantes al tema y estén centrados en el desarrollo

de soluciones de inteligencia de negocios que traten aspectos de diseño, calidad y seguridad.

Amplitud y calidad de la pregunta

Para definir la amplitud y calidad de la pregunta, la fundamentación se realizará en función de la respuesta a una serie de apartados en los que se analiza el problema a tratar, por lo que se propone utilizar la pregunta de investigación y el conjunto de palabras clave identificadas, así como los resultados que se esperan obtener y cómo serán analizados.

Problema

El proceso para generar informes estadísticos, realizar exploración sobre sus datos y ejecutar análisis sobre la información que se genera en el Área de Cuenta Individual y Control de Pagos, es una tarea compleja que se realiza de forma manual que consume mucho tiempo, es imprecisa y costosa en recurso y procesamiento.

La propuesta de un modelo de inteligencia de negocios permitirá al Área de Cuenta Individual y Control de Pagos de la Gerencia de Pensiones comprender, explorar y generar conocimiento sobre sus datos, de una forma sencilla y de fácil acceso. Al contar con una plataforma orientada al análisis, se dispondrán también de muchos beneficios entre los que se destaca la ventaja competitiva para determinar por ejemplo patrones o tendencias, productividad en la toma de decisiones basado en datos integrados y consistentes, así como también generar insumos para el control de procesos mediante la comparativa de datos históricos para ajustar estrategias de producción o presupuestos.

Debido a lo mencionado anteriormente el problema se centra en el estudio de los trabajos realizados en el ámbito del desarrollo de propuestas de modelos de

inteligencia de negocios aplicados por departamentos o empresas con necesidades similares.

Pregunta de investigación

¿Qué trabajos aplicados al desarrollo de propuestas de modelos de inteligencia de negocios se han realizado en los departamentos o áreas de las empresas u organizaciones en el nivel nacional e internacional?

Palabras clave y sinónimos

Para obtener las distintas fuentes de los trabajos primarios, es necesario definir palabras clave presentes en los trabajos por localizar. Para determinar los conceptos relacionados y sus palabras claves, así como conocer la correcta traducción al inglés de cada término, se revisaron varios artículos de inteligencia de negocios, los cuales fueron encontrados en una búsqueda inicial.

A continuación, se detalla un cuadro en el que aparecen agrupados por área las palabras clave y los conceptos relacionados que se utilizarán en esta revisión.

AREA	PALABRAS CLAVE
INTELIGENCIA DE NEGOCIOS Y DATA WAREHOUSE	Data Warehouse Business Analytics Business Intelligence BI Solutions Design Data Warehouse Decision support Multidimensional modeling
VISUALIZACIÓN DE DATOS EN POWER BI	Data Visualization Dashboards Power BI Power BI Models

	Power Gateway Power BI Mobil

Tabla 4. Palabras clave de la revisión sistemática. Fuente: Propia.

Intervención

En el contexto de la revisión sistemática planificada se van a observar las propuestas existentes de modelos de inteligencia de negocios, extrayendo lo más relevante y procediendo a un posterior análisis de estas.

Control

Todos los trabajos primarios incluidos son derivados de la aplicación de los criterios definidos para la revisión sistemática.

Resultado

Los resultados esperados de esta revisión sistemática son conocer las propuestas existentes en cuanto a modelos de inteligencia de negocios, para poder analizarlas y conocer qué comparten, además de identificar necesidades de investigación.

Medida de salida

Número de iniciativas encontradas basadas en el área que se centran.

Población

Publicaciones presentes en los repositorios de las fuentes de datos seleccionadas que estén relacionadas con el objetivo de esta revisión.

Aplicación

Los beneficiarios de la revisión sistemática serán las personas (académicos, investigadores, profesionales, etc.) relacionados directamente con el desarrollo de modelos de inteligencia de negocios, que estén interesados en conocer los trabajos relevantes existentes que hayan tratado el problema de la complejidad en el desarrollo de los informes estadísticos para la toma de decisiones y la necesidad de

contar con herramientas para mejorar el uso del tiempo mediante la integración de un modelo de inteligencia de negocios a sus procesos de trabajo.

SELECCIÓN DE FUENTES

En este apartado se establece cuáles serán las fuentes principales para llevar a cabo la revisión sistemática, para posteriormente utilizar los elementos definidos en la planificación para aplicar el procedimiento de obtención de estudios primarios en cada una de las fuentes seleccionadas.

Definición del criterio de selección de fuentes

Para el criterio de la selección de las fuentes se utilizarán fuentes "arbitradas".

Esto significa que un artículo que aparezca indexado en dicho repositorio tiene que haber pasado un proceso de arbitraje previo lo cual le da mayor credibilidad.

Asimismo, para los libros u otras fuentes, se utilizarán criterios como el ranking de la universidad de donde procede el autor y el índice H que el autor posea en su perfil.

Otros requisitos exigidos serán la accesibilidad vía Web y la inclusión de motores de búsqueda que permitan consultas avanzadas.

Lenguaje de estudio

El lenguaje de los estudios primarios será en inglés y estos serán extraídos mediante consultas en que las palabras clave están en inglés. El informe de la revisión sistemática es en español.

Identificación de fuentes

Se identifican las fuentes sobre las que se ejecutará la revisión sistemática planificada, al definir el método de selección de fuentes y la lista de fuentes consideradas, así como establecer la cadena de búsqueda que usará en la ejecución de la revisión.

Lista de fuentes

La lista de fuentes sobre la cual se ejecutará la revisión sistemática es la siguiente:

- Springer
- Repositorio Librarika de la Universidad Cenfotec

Adicionalmente para verificar la calidad de cada uno de los artículos seleccionados se utilizará la herramienta vía Web del sitio https://www.scimagojr.com.

Cadenas de búsqueda

Con el apoyo de los operadores lógicos AND y OR sobre las palabras clave y conceptos relacionados, que se identificaron se procede a establecer las cadenas de búsqueda utilizar en la presente revisión:

(((“Business Intellicence”) AND (“Data Warehouse” OR “Business Analytics” OR “Data Mining” OR “BI solutions” OR “Design Data Warehouse”)) OR ((“Power BI”) AND (“Data Visualization” OR “Dashboards” OR “Models” OR “Power Gateway” OR “Mobil”)))

Selección de fuentes después de la evaluación

Posterior a la ejecución de la revisión sobre las fuentes seleccionadas se realizará una fase de refinado en la que se identificarán e incluirán los estudios

primarios relevantes que a juicio del investigador sean necesarios para completar el estudio y estos no hayan podido ser obtenidos en las búsquedas realizadas.

Comprobación de las fuentes

Las fuentes arbitradas identificadas para el desarrollo de la revisión sistemática son las siguientes:

- Springer: Este sitio contiene artículos de diferentes áreas de estudio y en este caso para el tema que fue seleccionado para este foro es la mejor elección para obtener información debido a la facilidad de encontrar gran cantidad de artículos en el contexto de data mining relacionado a estudiantes y universidades.
- Repositorio Librarika de la Universidad Cenfotec: Universidad aprobada por el Consejo Nacional de Enseñanza Superior Universitaria Privada (CONESUP), además la carrera de Tecnologías de bases de datos está acreditada por el Sistema Nacional de Acreditación Superior (SINAES).

Selección de estudios

Se utilizarán las palabras claves definidas en apartado anterior, por cada fuente seleccionada se realizará un proceso iterativo formado por las etapas de búsqueda, extracción y visualización de la información. De tal forma que los resultados obtenidos se les aplica el criterio de inclusión para obtener los estudios relevantes, posterior a esos resultados se aplica el criterio de exclusión de forma que se obtengan los estudios primarios, los cuales se pasan almacenar y analizar a más profundidad extrayendo la información más relevante de cada uno de ellos con base en un formulario definido.

El proceso de selección sigue la siguiente forma:

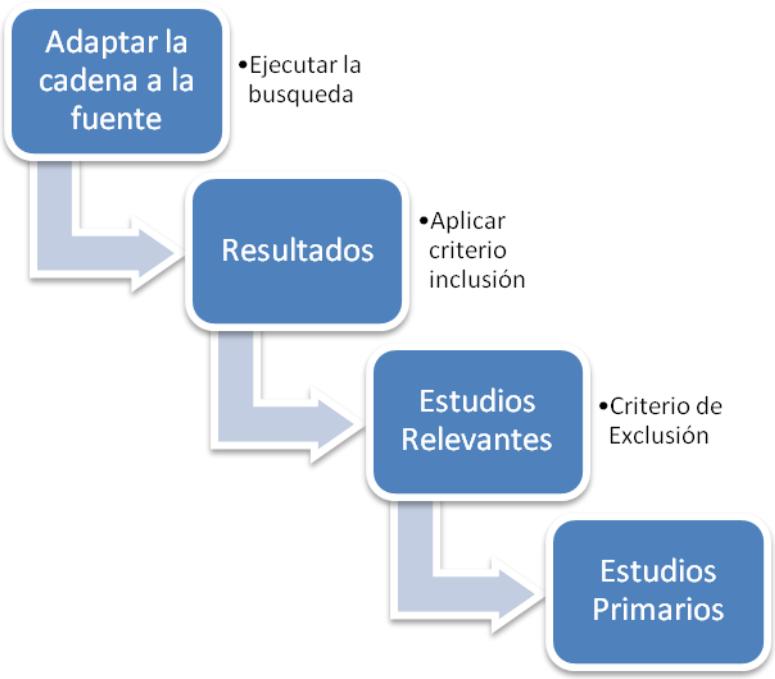


Figura 2. Selección de estudio. Fuente: Propia.

Definición del criterio de inclusión y exclusión de estudios

Como criterio de inclusión se realizará un análisis sobre el título, las palabras claves y abstract de cada documento. De forma que se eliminen aquellos resultados que no realizan aportaciones sobre diseños de modelos de inteligencia de negocios dentro de las organizaciones o empresas.

El criterio de exclusión actúa sobre el subconjunto de documentos obtenidos en la etapa anterior, de modo que se aplicará principalmente sobre la lectura y análisis del abstract del documento y sus conclusiones, al tener que profundizar en otras partes del documento, que permita ver la relación real que presenta con los objetivos buscados y si realmente es importante para la revisión, seleccionarlo como estudio primario.

Definición de tipos de estudio

Los tipos de estudios primarios que serán seleccionados durante la ejecución de la revisión sistemática son los artículos presentes en las fuentes seleccionadas, que cumplan con los criterios establecidos y hayan sido obtenidos como producto al aplicar el procedimiento de selección.

Ejecución de la revisión

Ejecución de la selección en la fuente Librarika de la Universidad Cenfotec

La búsqueda fue realizada mediante las opciones disponibles de búsqueda por textos con opción “Default“ y tipo “All”, no sé establecen más criterios dado a que la herramienta no lo permite.

La consulta realizada se adapta a la siguiente cadena: “Inteligencia de negocios” o “Data Warehouse”.

La ejecución de la búsqueda arroja 9 resultados, tras aplicar el criterio de inclusión quedan 5 documentos relevantes, a los cuales se les aplica el criterio de exclusión y quedan 3 documentos que serán considerados estudios primarios.

Estos son:

- Creación de un modelo de inteligencia de negocios en la atención de emergencias del Benemérito Cuerpo de Bomberos de Costa Rica (Ceciliano Villalobos, 2017).
- Desarrollo de una solución de inteligencia de negocios de arquitectura híbrida para el departamento de ventas de BTicino Costa Rica (Mata Solis, 2016).
- Diseño y especificación de la Fábrica de Información Corporativo del Banco C.R (Vargas Paniagua, 2013).

Evaluación de la calidad de los estudios

La universidad Cenfotec está debidamente aprobada para operar por el Consejo Nacional de Enseñanza Superior Universitaria Privada (CONESUP), además la carrera de Maestría en Tecnologías de bases de datos está acreditada por el Sistema Nacional de Acreditación Superior (SINAES), por lo que los trabajos de graduación presentados por los graduados deben cumplir con una serie de filtros y evaluaciones para poder ser aceptados como proyectos finales.

Revisión de la selección

La selección de los estudios primarios se validó, de modo que se tiene la seguridad de que no se haya dejado atrás ningún estudio relevante en esta fuente.

Extracción de la información

Definición del criterio de inclusión y exclusión de información

La extracción de la información relevante de los estudios primarios se realizará según los objetivos, centrándolos principalmente en estudiar las aportaciones de interés que realizan sobre el diseño de modelos de inteligencia de negocios en las empresas u organizaciones, considerando las siguientes:

- Modelos de inteligencia de negocios en las organizaciones o empresas.
- Modelos de inteligencia de negocios aplicadas a un dominio específico.
- Trabajos teóricos sobre modelos de inteligencia de negocios.

Formulación para la extracción de información

El formulario utilizado para la extracción de la información consta de 3 partes principales: Identificación, descripción, aspectos por destacar.

A continuación, se detalla el formulario:

Identificación	Título Publicación Autores Referencia
Descripción	Área Resumen
Aspectos para destacar	

Tabla 5. Formulario para la extracción de información. Fuente: Propia.

Extracción de resultados objetivos y subjetivos

Identificación	
Titulo	Desarrollo de una solución de inteligencia de negocios de arquitectura híbrida para el departamento de ventas de BTicino Costa Rica.
Publicación	Tesis, Universidad Cenfotec. Librarika.
Autores	Ing. Minor Mata Solís Ing. José Ahías López Portillo
Referencia	[1]
Descripción	
Área	Inteligencia de negocios, Data WareHouse, Power BI
Resumen	Diseñar una plataforma de inteligencia de negocios de arquitectura hibrida para el departamento de ventas de Bticino Costa Rica, con el fin de minimizar costos y maximizar funcionalidades en una única plataforma centralizada, accesible desde cualquier dispositivo.
Aspectos a destacar	

- Descripción del problema similar en cuanto a la ejecución de análisis los cuales toman gran cantidad de tiempo y el análisis no es sencillo de realizar.
- Aprovechar la información para proveer una herramienta que apoye la tome de decisiones tanto estratégica como operativa.
- Ofrecer a los expertos del negocio y tomadores de decisión herramientas automatizadas que faciliten la extracción, procesamiento, visualización, predicción y análisis de los datos.
- Romper mitos sobre inversión de grandes cantidades de dinero para poder utilizar inteligencia de negocios.
- Énfasis en la herramienta para visualizar los datos y no solo en el repositorio.

Identificación	
Titulo	Creación de un modelo de inteligencia de negocios en la atención de emergencias del Benemérito Cuerpo de Bomberos de Costa Rica
Publicación	Tesis, Universidad Cenfotec. Librarika.
Autores	Ing. Cristian Ceciliano Villalobos
Referencia	[2]
Descripción	
Área	Inteligencia de negocios, Data WareHouse, Power BI
Resumen	Implementar un modelo de inteligencia de negocios el cual permite centralizar la información de bomberos en un único repositorio al

	permitir aplicar métodos de análisis para obtener conocimiento y basar sus decisiones estratégicas.

Aspectos por destacar

- Descripción del problema similar en cuanto a la ejecución de análisis los cuales toman gran cantidad de tiempo y el análisis no es sencillo de realizar.
- La importancia de analizar la información de diferentes formas y combinaciones.
- La realización de comparativas en diferentes líneas del tiempo utilizando como criterio el tipo de emergencia atendida.
- Incorporación de minería de datos para la predicción de emergencias.
- La incorporación del modelo de inteligencia de negocios permite realizar trazos y dar seguimiento a un patrón que permita evidenciar su comportamiento.
- Utilización de Power BI para mostrar resultados.
- Eliminación de duplicidad de información y de funciones en cuanto a quien consulta la información y de donde se obtiene.

Identificación	
Titulo	Diseño y especificación de la Fábrica de Información Corporativa del Banco de Costa Rica
Publicación	Tesis, Universidad Cenfotec. Librarika.
Autores	Ing. Edius Vargas Paniagua

Referencia	[3]
Descripción	
Área	Inteligencia de negocios, Data Warehouse
Resumen	Construcción de un Data Warehouse para obtener decisiones más acertadas y mejora continua de sus servicios.
Aspectos a destacar	

- Permite dar orientación a los esfuerzos de integración de datos y reportería analítica para la toma de decisiones. La organización reconoce la necesidad de establecer un proceso más estructurado para soportar estos procesos que antes se realizaban de forma manual.
- La propuesta permite al negocio y a TI, comprensión adecuada que necesitan para coordinar esfuerzos en pro de la mejora continua de la plataforma de Inteligencia de Negocios.
- Madurez al personal técnico y del negocio en temas de integración y calidad de datos.
- Motivación al negocio de gestionar diversas formas de apoyo al proyecto como la inclusión de recurso humano dedicado, designación de un dueño del producto y patrocinador de alto nivel, así como la visión de optar por mejores recursos tecnológicos en futuros cercanos.
- Instrumentos definidos para la evaluación de la calidad de datos han producido mejoras en la información que se integra al Data Warehouse, al permitir asimismo señalar deficiencias en los sistemas orígenes con la subsecuente corrección de la causa raíz de los problemas de calidad.

- El proyecto aporta insumos básicos iniciales para responder a las necesidades de generar información para análisis del negocio, que podrán ser mejorados constantemente en beneficio de un manejo más efectivo los recursos destinados a este tipo de proyectos y de un mejor manejo de la información a nivel del BCR.

Ejecución de la selección en la fuente Springer

Selección de estudios iniciales

Se realiza la búsqueda avanzada mediante la opción que presenta la herramienta y se configura de esta forma, al utilizar las cadenas de búsqueda definidas previamente:

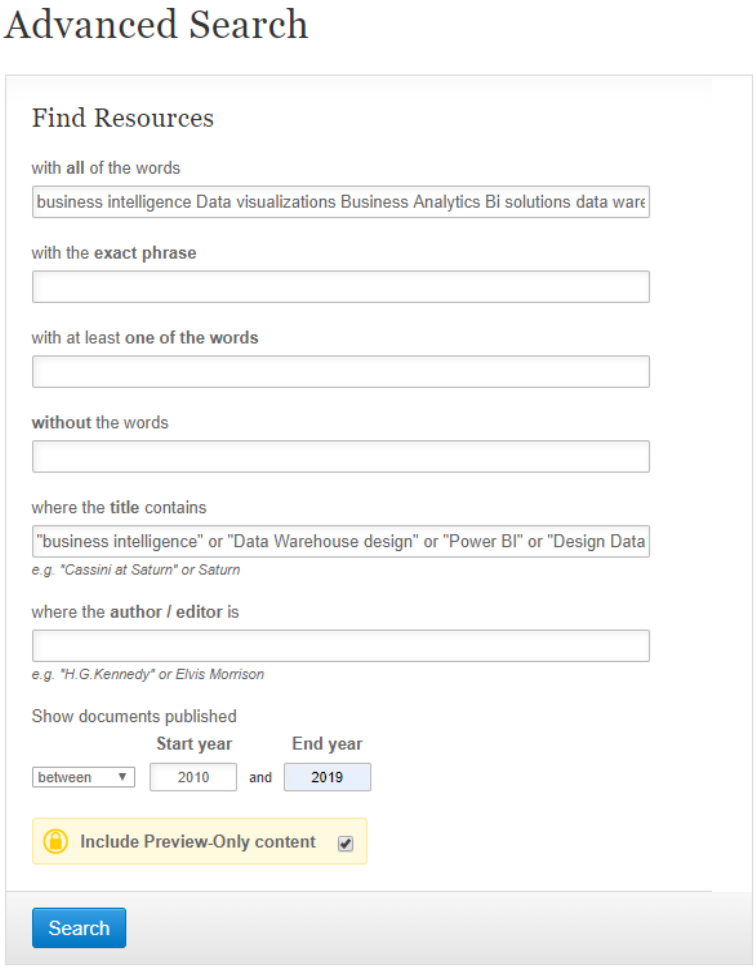


Figura 3. Parámetros de búsqueda Springer. Fuente: Propia.

Debido a la gran cantidad de información que existe en el sitio, se establece un periodo de 9 años hacia atrás (valorando más los trabajos recientes en función de mejores técnicas y formas de aplicar inteligencia de negocios), al realizar la búsqueda está inicialmente arroja un aproximado de 845 resultados.

Se procede a aplicar otros filtros propios del tema de interés y objetivos de la investigación, como lo son artículos y en idioma inglés. Al aplicar esta búsqueda se reducen los estudios a inicial con los criterios anteriores se tiene un resultado aproximado de 85 publicaciones, al aplicar los criterios de inclusión quedan aproximadamente 5 estudios, posteriormente, al aplicar los criterios de inclusión quedan 3 estudios que serán tomados como primarios.

Los cuales son:

- A business intelligence based solution to support academic affairs: case of Taibah University.
- The role of user requirements in data repository design
- An approach for incorporating quality-based cost–benefit analysis in data warehouse design

Evaluación de la calidad de los estudios

Se utiliza el sitio www.scimagojr.com, el cual provee información importante para determinar el factor de impacto del artículo, su índice H, cuartiles y el factor de la colaboración internacional.

Extracción de resultados objetivos y subjetivos

Identificación	
Titulo	An approach for incorporating quality-based cost–benefit analysis in data warehouse design

Publicación	Articulo, Information Systems Frontiers, july 2008, Volume 10, Issue 3, pp 361–373.
Autores	Lila Rao Kweku-Muata Osei-Bryson
Referencia	[1] https://doi.org/10.1007/s10796-008-9077-4
Descripción	
Área	Data Warehouse, Decision Support
Resumen	Análisis del costo-beneficio incluido en la adquisición de la solución de almacenamiento, garantías, y mantenimiento de los activos.
Aspectos a destacar	
• Perspectiva de la evaluación del costo-beneficio en el desarrollo de un Data Warehouse. • Identificar el conjunto de datos con el potencial de producir mayor valor neto para la organización. • Determinar el valor del conjunto de datos que se almacenan en un Data Warehouse mediante la aplicación de técnicas.	
Calidad del articulo	
Vinculo de verificación	https://www.scimagojr.com/journalsearch.php?q=15572&tip=sid&clean=0
Valor H Index	51
Años de cuartiles 1 representativos	2010,2014,2016
Factor de impacto	Más de dos citas por documento desde 2008

Factor colaboración internacional	Más de 3.6 desde 2017

Identificación	
Titulo	The role of user requirements in data repository design
Publicación	Artículo, International Journal on Software Tools for Technology Transfer, February 2018, Volume 20, Issue 1, pp 19–34
Autores	Ilyès Boukhari Stéphane Jean Idir Ait-Sadoune Ladjel Bellatreche
Referencia	[2] https://doi.org/10.1007/s10009-016-0443-0
Descripción	
Área	Data Warehouse
Resumen	La importancia de los requerimientos en las otras fases del repositorio de datos.
Aspectos que destacar	
• Marco de integración de requisitos basado en ontologías • Modelo genérico para unificar los vocabularios utilizados y los idiomas requeridos • Enfoque para verificar la consistencia de los requerimientos. • Requisitos que ayudan a definir estructuras de acceso adecuadas, como índices y vistas materializadas para optimizar las consultas SQL	
Calidad del articulo	

Vinculo de verificación	https://www.scimagojr.com/journalsearch.php?q=19271&tip=sid&clean=0
Valor H Index	47
Años de cuartiles 1 representativos	1999,2001,2003,2007
Factor de impacto	Más de 1.5 citaciones por documento desde 2005
Factor colaboración internacional	Más de 2 citaciones por documento desde 2017

Identificación	
Titulo	A business intelligence based solution to support academic affairs: case of Taibah University
Publicación	Artículo, Springer US, Wireless Networks pp 1–8.
Autores	Wadii Boulila Muhib Al-kmali Mohammed Farid Hamzah Mugahed
Referencia	https://doi.org/10.1007/s11276-018-1880-3
Descripción	
Área	Business Inteligence, Decisión Support, Data Warehouse
Resumen	Solución de BI para apoyar los asuntos académicos en la Universidad de Taibah.
Aspectos que destacar	
• Recopilación de datos de diferentes fuentes mediante las tres operaciones (Extraer, Transformar y Cargar).	

• Proponiendo una solución multidimensional que describa los procesos académicos. • Visualización de resultados a través de un conjunto de paneles e informes. Utilizando las herramientas de datos de SQL Server.

Calidad del articulo	
Vinculo de verificación	https://www.scimagojr.com/journalsearch.php?q=17569&tip=sid&clean=0
Valor H Index	79
Años de cuartiles 1 representativos	1999,2001,2003,2005,2007,2009
Factor de impacto	Dos citaciones por documento desde 2014.
Factor colaboración internacional	Más de 1.8 citaciones pode documento desde 2014.

Análisis de resultados

Números de estudios obtenidos	854
Números de estudios primarios seleccionados	6
Búsqueda, selección y extracción invalida de la revisión sistemática	Se invalidan o excluyen aquellos estudios que a pesar de contar con las palabras claves en su título y abstract, no aporten mayor relevancia a la investigación realizada.
Publicaciones positivas / negativas	Se toma la decisión de abarcar las publicaciones positivas, las cuales

	cumplen con la revisión sistemática y además aportan relevancia a la investigación.
Variaciones de inter- revisión	Los conflictos en medio de las revisiones y los resultados de la revisión sistemática son resueltos por la utilidad que representen para el desarrollo del tema de investigación.
Aplicación de resultados	Se aplicarán de acuerdo con cada una de las fases en la propuesta de la solución para el modelo inteligencia de negocios que se plantea.
Recomendaciones	Comprender los requerimientos del negocio es fundamental para desarrollar un repositorio de datos que forma parte de un modelo de inteligencia de negocios. Es importante siempre aplicar procesos de calidad de datos en las fases de carga y transformación. Las herramientas de visualización son una parte importante que puede fortalecer la adopción de la inteligencia de negocios en las empresas.

Este trabajo ha permitido analizar el campo de la inteligencia de negocios aplicada al diseño, calidad y seguridad al realizar una revisión del estado del arte y una comparación formal de las principales propuestas de diseño de inteligencia de negocios detectadas. Para el análisis del estado del arte se ha utilizado la técnica de revisión sistemática, al definir y aplicar un protocolo formal para la revisión que permita centrar la pregunta de investigación y el tipo de resultados (estudios primarios) que queremos obtener y posteriormente analizar.

Una vez que se han analizado los estudios primarios detectados y se ha realizado su comparación, se puede decir que la mayoría de los trabajos identificados tienen un dominio importante en cuanto a la identificación adecuada de los requerimientos con enfoques consistentes para elaborar diseños de soluciones de inteligencia de negocios. Lo anterior soportado adicionalmente con proyecciones sobre el costo-beneficios del valor de los datos que se almacenan en el repositorio central para elaborar análisis e informes. Asimismo, de acuerdo con las lecturas de los estudios primarios encontrados que están focalizados en resolver una problemática en específico de cierta área o departamento, se puede concluir que la inclusión o propuestas de estas soluciones motiva al negocio a empoderarse de la tecnología, de tal forma que los proyectos reciben apoyo económico, humano e infraestructura. Lo que a su vez estimula dentro del negocio: orientación, madurez y comprensión al personal técnico y de negocios sobre temas de integración de datos, calidad y reportes.

Capítulo 2. Marco Conceptual

La inteligencia de negocios se entiende como:

“Se entiende por Business al conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización” (Díaz, 2012).

En otras palabras, tecnologías que facilitan la obtención rápida y sencilla de datos provenientes de los sistemas operativos del negocio para su análisis e interpretación.

Como resultado de la revisión sistemática se ha desarrollado un conjunto de conceptos comunes que definen el contexto sobre el trabajo que se está desarrollando, producto de este análisis se obtuvo la siguiente grafica de palabras las cuales serán desarrolladas en este marco a fin de aclarar términos utilizados a lo largo del documento. Asimismo, se incluirán conceptos que a criterio del investigador sean necesarios para desarrollar el tema. A continuación, se muestra una serie de palabras que se relacionan con el término de inteligencia de negocios en el sitio Web Tableau en español (Tableau software, 2019).



Figura 4. Conceptos comunes. Fuente: Propia.

El Wordcloud se generó mediante la herramienta llamada TagCrowd, en la cual se excluyeron preposiciones, las palabras “inteligencia, negocios” y palabras

mal interpretadas. El resultado es el que muestra la figura 4, está muestra la palabra más frecuente en conjunto con la frecuencia de su aparición en el texto analizado. Se puede denotar un conjunto de 15 palabras muy frecuentes asociadas con el término de inteligencia de negocios, en las cuales se destaca “datos” con 31, seguido de “soporte y rendimiento”, con 7 apariciones y por último decisiones con 6. Se podría decir con el análisis del texto del sitio Web de Tableau entonces que el término de inteligencia de negocios posee características muy claras divididas en 4 ejes:

- Datos.
- Soporte.
- Rendimiento.
- Decisiones.

Del análisis anterior, se extienden más conceptos que a su vez están relacionados. Para lo cual se define un mapa mental jerárquico que brinde una guía de conceptos del cómo están organizados estos. Representados en la siguiente imagen:

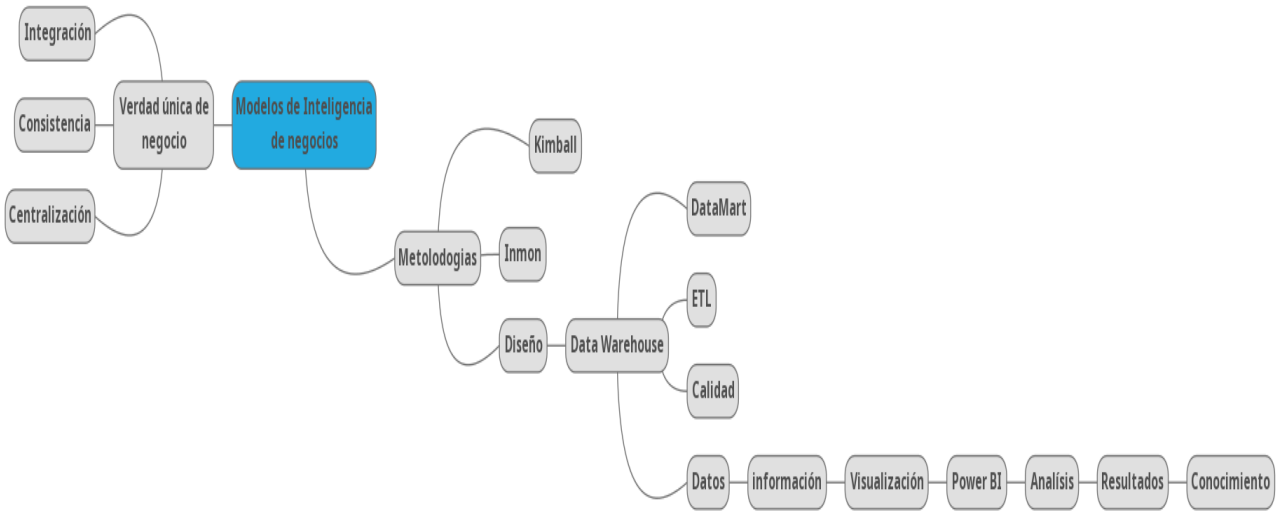


Figura 5. Mapa mental de conceptos. Fuente: Propia.

Este modelo de inteligencia de negocios debe definirse basado en una metodología o filosofía de desarrollo existente, las más conocidas y utilizadas en la actualidad son las propuestas presentadas por Ralph Kimball y Bill Inmon. Estos autores presentan en su contexto diferentes formas de desarrollar el modelado de datos para lograr soluciones de negocio efectivas, cada una conserva sus ventajas y desventajas. Una vez seleccionado el diseño se puede optar por modelar el repositorio central o Data Warehouse de la información basado en los requerimientos definidos por el negocio.

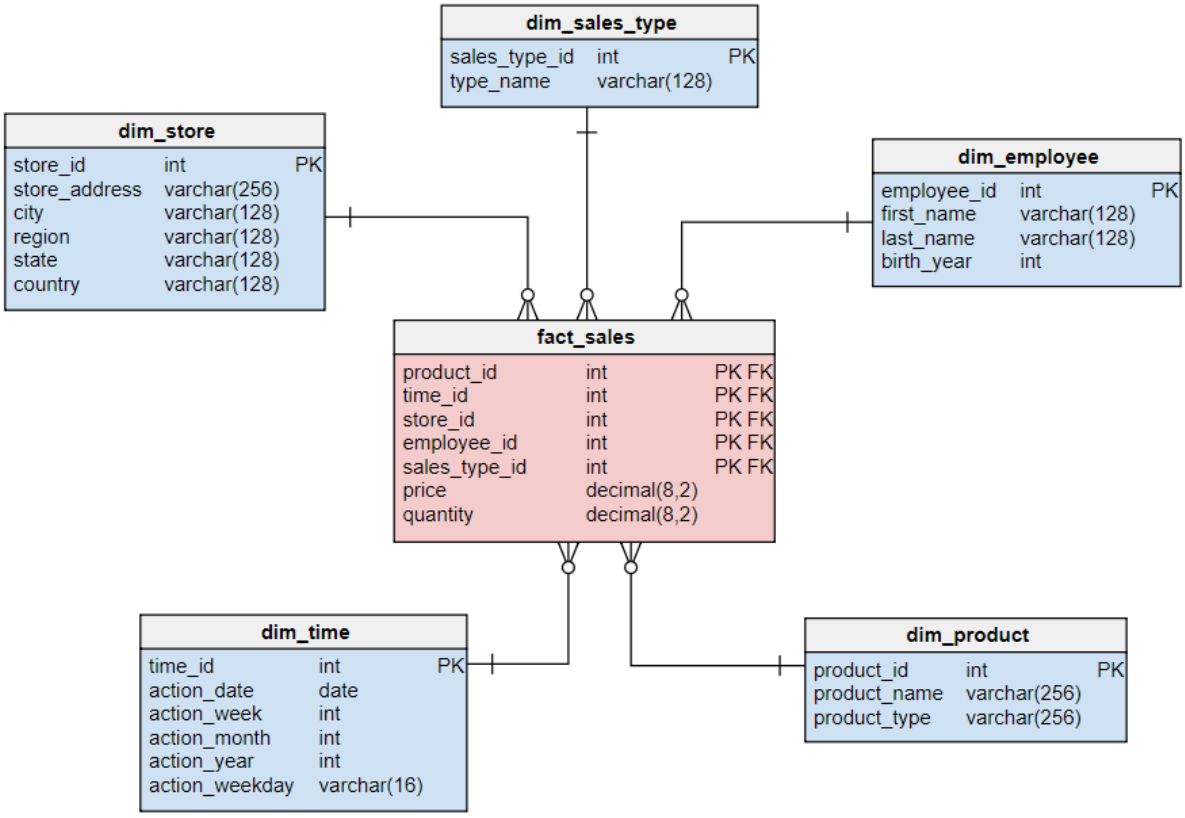


Figura 6. Modelo estrella definido por Ralph Kimball. Fuente: (ALVI, 2019)

El Data Warehouse es una integración de los datos en el nivel empresarial, pero según el área y la necesidad de información pueden existir subconjuntos de datos para ayudar a áreas específicas del negocio en la toma de decisiones, esto se conoce como data mart. Puede decirse que el data mart es “Una poderosa

extensión natural del Data Warehouse para un uso funcional o especifico de un departamento” (PRABHU, 2008).

Dentro del diseño del Data Warehouse siempre existen los procesos ETL (extracción, transformación y carga), estos procesos se utilizan para obtener la información desde los sistemas transaccionales hacia el Data Warehouse, usualmente pasan anteriormente por un área denominada Staging Area, la cual

sirve como antecesor para aplicar ciertas reglas de calidad y transformación antes de pasar al Data Warehouse. En esta fase es donde se aplican los procesos de calidad de datos a fin de que la información que se almacena en el Data Warehouse sea lo más consistente y limpia posible. Este proceso abarca un análisis inicial llamado Data Profiler, que es utilizado para conocer a fondo los datos que se van a tratar.

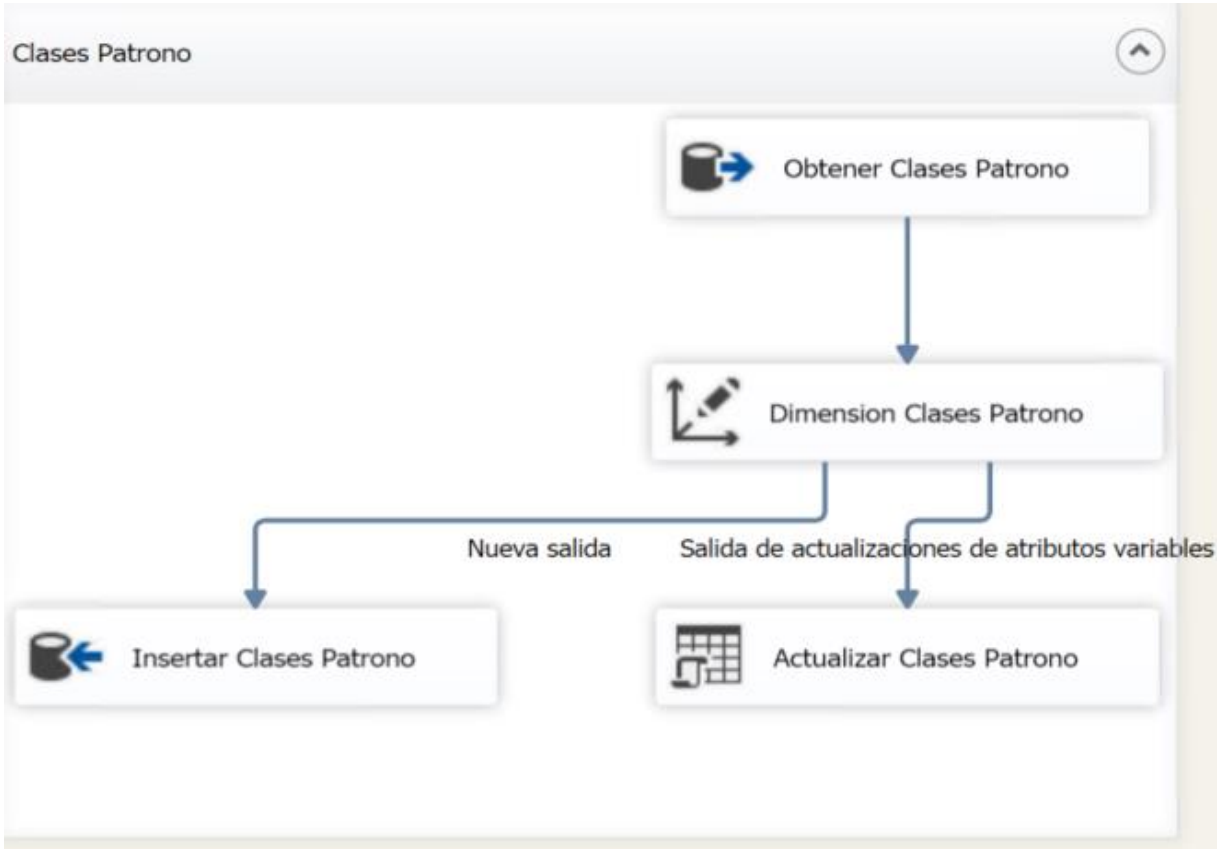


Figura 7. Vista de componentes Integration Services. Fuente: Propia.

Cuando los datos están cargados dentro del Data Warehouse, se les debe dar un procesamiento a fin de construir un cubo analítico OLAP, este modelo puede ser multidimensional o bien tabular. Las bases de datos OLAP poseen características importantes sobre base de datos relacionales, principalmente en la estructura interna para extraer los datos. Se puede destacar:

- Están diseñadas para un análisis general de lo que sucedió.
- El modelado de datos es diferente.
- Son distintos a la tercera forma normal.
- Mayor desempeño en el procesamiento de consultas.
- Permite hacer cálculos más complicados que OLTP y en menor tiempo.
- Manejo de información histórica (Volúmenes grandes).

Para lograr construir un cubo analítico OLAP se utilizan herramientas especializadas que ayuden con la tarea. En este caso particular se utiliza el servicio de Anlisys Services que viene dentro del paquete de SQL Server Data Tools, esta herramienta permite la construcción de un cubo analítico y proporcionar un modelado de datos virtual para el procesamiento del cubo que se requiere realizar. Adicionalmente se pueden construir “perspectivas” que son diferentes vistas que permiten ver cierto grupo de información para las personas autorizadas, así como hacer conversiones del idioma y que este sea entendible por cualquier persona en el nivel empresarial ubicado en diferentes países, dentro de muchas otras características que brinda la herramienta.

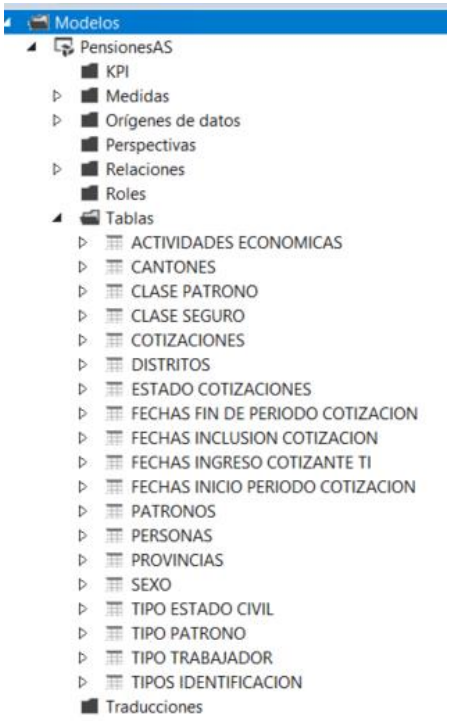


Figura 8. Herramienta SSAS modelo tabular. Fuente: Propia.

Una vez procesada la información en la base de datos OLAP mediante la herramienta de Analisys Services, se puede consultar la información, pero está información necesita de una herramienta que sea fácil de manejar y entendible. En este caso se optó por utilizar la herramienta llamada Power BI de Microsoft debido a su robustez y compatibilidad con las licencias que ya tiene la Gerencia de Pensiones para su uso. Esta es una herramienta orientada al usuario experto para la ejecución del análisis empresarial de la información, de modo que se conecta a la base de datos OLAP para generar información con la combinación de las múltiples variables existentes dentro de su arquitectura.

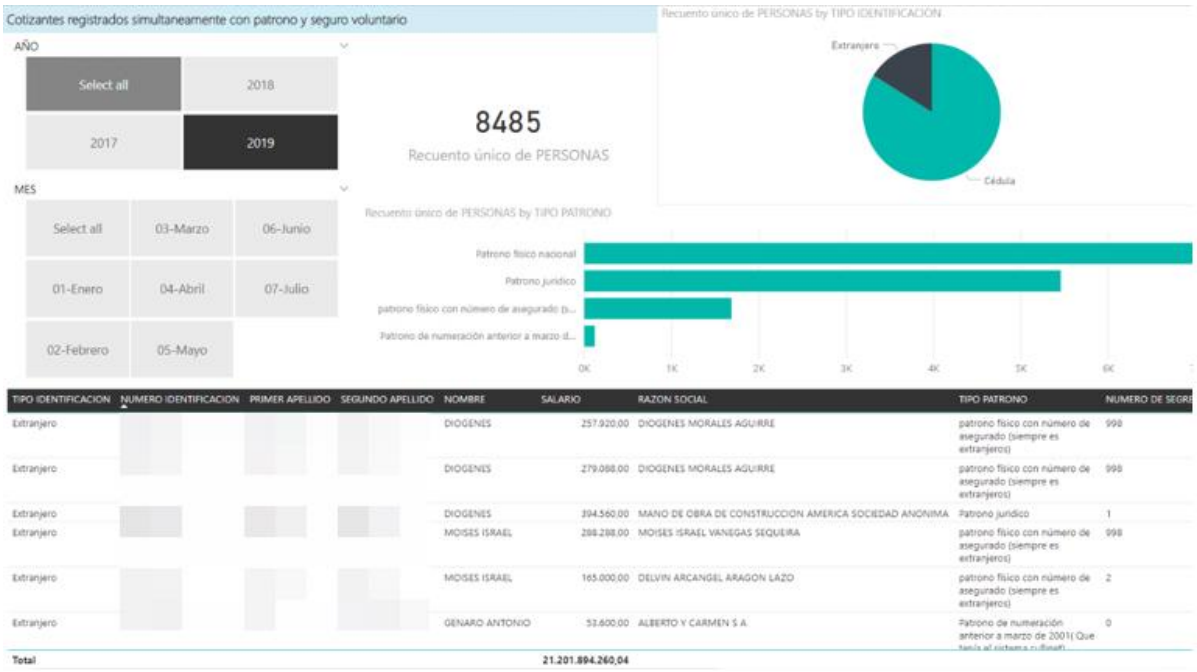


Figura 9. Vista de Dashboard en Power BI. Fuente: Propia.

Capítulo 3. Marco Metodológico

3.1 Tipo de Investigación

El tipo de investigación es aplicada, dado a que se busca desarrollar conocimientos existentes para atender la necesidad del Área de Cuenta Individual y Control de Pagos en cuanto a la propuesta de un modelo de inteligencia de negocios que le permita mejorar el uso del tiempo y reducir la cantidad de tareas necesarias para desarrollar análisis de la información.

3.2 Alcance Investigativo

El estudio tiene alcance descriptivo, dado que demuestra basado en los estudios primarios las siguientes situaciones o contextos:

- Las mejoras en el uso del tiempo.
- La facilidad para elaborar análisis sobre la información.
- El valor de la información sobre el costo-beneficio de aplicar modelos de inteligencia de negocios.
- Procesos de optimización en el diseño del Data Warehouse.
- Comprensión del negocio sobre el impacto de este tipo de soluciones.
- Requisitos del negocio que ayudan a generar estructuras más adecuadas, como índices y vistas materializadas para optimizar las consultas SQL.
- Reconocimiento de los altos mandos del negocio en cuanto a la necesidad de integración de datos y reportería analítica para la toma de decisiones.

Para medir el impacto de la propuesta del modelo de inteligencia de negocios se utilizó como punto de referencia la cantidad de tareas y tiempo que se requiere actualmente para generar reportes con ciertas cantidades de variables a través del

tiempo por parte de los funcionarios del Área de Cuenta Individual y Control de Pagos asignados para esta labor, por ejemplo, en días, trimestres, semestres o anualmente.

3.3 Enfoque

El enfoque que se utilizará será alternativo, al partir del hecho de que el enfoque alternativo lo cuantitativo y lo cualitativo nunca han estado separados (Chavarría, 2011).

De este modo es que en este trabajo se explican las dimensiones epistemológica, ontológica y axiológica de la investigación, sin hacer alusión explícita a un enfoque en particular.

La dimensión epistemológica, al tratarse del estudio de una propuesta de un modelo de inteligencia de negocios enfocado en la resolución de una problemática existente en el Área de Cuenta Individual y Control de Pagos, implica que se espera tener resultados positivos en cuanto a la complejidad en el desarrollo de las tareas y mejoras en el uso del tiempo, por cuanto habrá que involucrarse con el desarrollo del fenómeno. El investigador prestará atención a las métricas de eficiencia, precisión y eficacia en la propuesta del modelo de inteligencia de negocios.

Para describir la dimensión ontológica se utiliza el concepto expresado por (Gruber, 1993). "Lo que existe es exactamente aquello que puede ser representado". La propuesta del diseño ontológico de un modelo de inteligencia de negocios se puede visualizar como se muestra en la figura 10:

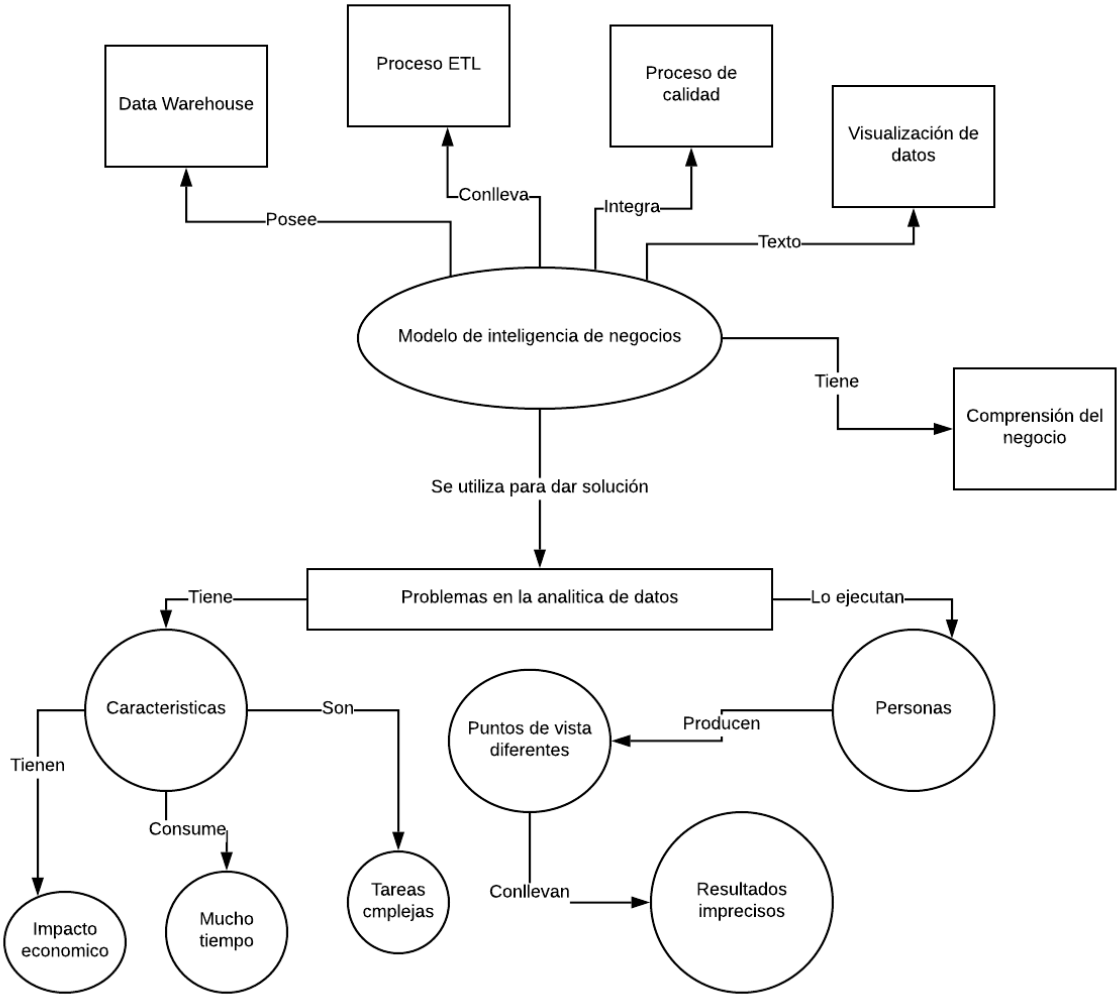


Figura 10. Ontología de modelo de inteligencia de negocios aplicado al problema de la analítica de datos ejecutados de forma manual. Fuente: Propia.

La dimensión axiológica está determinada por las mejoras en cuanto a las cantidades, tiempos, costos e impacto en la obtención de la información y su análisis.

Dentro de lo que destaca:

- Cantidad de tareas requeridas para desarrollar informes analíticos.
- Tiempo necesario para desarrollar informes analíticos.

- Costo económico requerido para desarrollar informes analíticos (humano y tecnológico).
- Impacto de los informes en la toma de decisiones, con las variables actuales versus con variables adicionales y segmentadas a través del tiempo.

Dado que en la revisión de la literatura de los estudios primarios y la investigación que se pudo realizar, no fue posible localizar una definición concreta o estándar que permita encasillar los elementos anteriores en pesos o valores porcentuales que permitan determinar niveles de aceptación del diseño de inteligencia de negocios. A su vez se utilizará para definir estos valores o porcentajes un criterio basado en la mejoría de los procesos en tiempo y cantidades esperados con el diseño que se plantea, al tomar como referencia la rúbrica de un modelo académico y al considerar las sugerencias del negocio. La tabla 6 ilustra los porcentajes y mecanismos utilizados para lograr tales efectos:

Variable	**Valor porcentual**	**Criterios**
Cantidad de tareas requeridas para desarrollar informes analíticos.	25%	Reducción de la cantidad de tareas. Escala:25%>=5,20%=4,15%=3,10%=2,5%=1
Tiempo necesario para desarrollar informes analíticos.	25%	Reducción del tiempo en segundos. Escala: 25%>=25200,20%>=21600,15%>=18000,10%>=14400,5%>=10800

Costo económico requerido para desarrollar informes analíticos (humano y tecnológico).	25%	Reducción del costo económico anual. Escala: 25%>=¢262.719,20%=131.359, 15%>=65.679,10%>=¢32.839,5%=¢16.419
Impacto de los informes en la toma de decisiones, con las variables actuales versus con variables adicionales y segmentadas a través del tiempo.	25%	Determinado por el jefe de CI.
Total	**100%**	
Niveles de aceptación		
Excelente	**Mayor o igual a 90%**	
Muy bueno	**Mayor o igual a 80%**	
Bueno	**Mayor o igual a 70%**	

Tabla 6. Porcentajes de evaluación del enfoque. Fuente: Propia.

La definición para cada nivel de aceptación se define de la siguiente forma:

- **Excelente:** El modelo es muy prometedor, existe una fuerte disposición a implementarlo.
- **Muy bueno:** El modelo responde a las necesidades, se deben hacer algunos ajustes mínimos para mejorar ciertas variables.
- **Bueno:** El modelo debe someterse a un análisis más profundo para determinar factores que podrían brindar mayor eficiencia en las variables a fin de que estas respondan a las necesidades del negocio.

Para cada una de las variables se utiliza un valor equitativo, por cuanto se definen 4 variables para lograr el 100%. La tabla se basa en los niveles de aceptación definidos usualmente en las encuestas.

3.4 Diseño

El diseño que se utilizó para desarrollar el presente trabajo es de tipo estudio de caso. Dado que se realizó un examen detallado, comprensivo, sistemático y profundo sobre la problemática actual, particularmente en la complejidad y tiempo que toma la generación de reportes analíticos avanzados para la toma de decisiones en el Área de Cuenta Individual y Control de Pagos, de modo que se obtenga como resultado un producto descriptivo del fenómeno estudiado.

La propuesta busco dar sentido y valor a las categorías de análisis respecto de los modelos de inteligencia de negocios que el tema de investigación abarca.

En este sentido se tienen 4 áreas importantes de análisis:

- Tiempos de duración.
- Cantidad de tareas.
- Costo económico.
- Impacto en la toma de decisiones.

Todo lo anterior en lo que respecta a la generación de informes con fines analíticos para el área y dirección involucrada.

3.5 Población y Muestreo

Para obtener una muestra válida y que se relevantes para las conclusiones de la investigación se tomó en cuenta para la población al jefe del Área de Cuenta Individual y Control de Pagos y 3 analistas de datos (funcionarios) del Área de

Cuenta Individual y Control de Pagos encargados de generar está información, por ende, el tipo de muestreo será no probabilístico.

3.6 Instrumentos de Recolección de Datos

Se aplicaron los siguientes instrumentos en procura de obtener la información más relevante para el tema que se desarrolla:

- Entrevistas estructuradas a los expertos del negocio y encargados de Tecnologías de Información.
- Recolección de documentos como informes de los indicadores de gestión utilizados para la toma de decisiones.
- Observación de campo.
- Reuniones y correos electrónicos.

3.7 Técnicas de Análisis de Información

Para este propósito, se planteó utilizar las siguientes técnicas:

- Diagramas de flujo de datos
- Espinas de Ishikawa
- Diagramas UML
- Mapas conceptuales

3.8 Estrategia de Desarrollo de la Propuesta

El modelo de inteligencia de negocios abarcó las mejores prácticas en cuanto a las técnicas para desarrollar las bases de datos OLAP, al seguir la estructura de un modelo tabular y al preservar la consistencia en la calidad de datos mediante criterios definidos por el negocio y propuestos por la investigación realizada. Asimismo, las visualizaciones de datos mediante la herramienta Power BI abarcó las recomendaciones de diseño e impacto sobre la percepción visual y la manera de

comunicar claramente la información, para los paneles y Dashboard que indicó el negocio, los estudios primarios y el desarrollo del trabajo.

Capítulo 4. Análisis del Diagnóstico

Recolección de información y resultados

Una vez aplicados los instrumentos diseñados del apartado 3.6, se obtuvieron los siguientes resultados:

Entrevistas al personal de TI.

Pregunta	*Respuesta*
¿Se cuenta con un modelo de inteligencia de negocios actualmente?	No.
¿Existen propuestas concretas para desarrollar inteligencia de negocios en el área de Cuenta Individual y Control de Pagos?	No. Existe una contratación, pero para desarrollarse en otro sistema de información que es inversiones.
¿Cuentan con las herramientas tecnológicas necesarias para desarrollar un modelo de inteligencia de negocios?	Si. Tienen las licencias de SQL Server, Analisys Service, Power BI con subscripción de office 365.
¿Está de acuerdo en que se desarrolle un modelo de inteligencia de negocios para el Área de Cuenta Individual y Control de Pagos? Dirigida al jefe de informática.	Sí. Estoy de acuerdo.

Tabla 7. Resultado entrevista personal de TI. Fuente: Propia.

Entrevista al personal del Área de Cuenta individual y Control de Pagos

Pregunta	*Respuesta*
¿Cuáles son los indicadores que desean obtener producto de la propuesta del modelo de inteligencia de negocios?	En total son 10 indicadores que son importantes para el área.
¿Actualmente como ejecutan el proceso para obtener está información?	Actualmente no se posee una fuente directa para generar estos informes. Cuando se requieren se realizan una solicitud al encargado de base de datos, quien ejecuta vía script la consulta y la devuelve adjunta en un Excel, dependiendo del tamaño un xml. Posteriormente se toma está información y se consolida con información local para generar el informe.
¿Cuánto tiempo en promedio para lograr generar un informe con alguno de estos indicadores?	Depende de cuanto tarde el encargado de base de datos en brindarla, y adicionalmente el cotejo local con los archivos de Excel. En promedio 1 hora o más.
¿Qué tanto esfuerzo requieren para poder generar está información en	Es complicado, requiere especificar en las solicitudes estas temporalidades,

diferentes temporalidades, como mensual, trimestral o anual?	pero según comentan los compañeros de base de datos a mayor rango más se tarda en obtener la información mediante consulta por la cantidad de registros.
¿Qué les parece la idea de generar ustedes mismos este tipo de información de forma eficiente y con más características?	Estaría muy de acuerdo, vendría agilizar mucho el proceso y brindaría mucho apoyo en la gestión para obtener la información requerida.

Tabla 8. Resultado entrevista al personal del Área de Cuenta Individual y Control de Pagos. Fuente: Propia.

A continuación, se muestra el listado de indicadores importantes que el Área de Cuenta de Individual y Control de Pagos definió a fin de desarrollar dentro de la propuesta de inteligencia de negocios.

Indicadores Cuenta Individual

N°: 1

Nombre del reporte: Cotizantes asalariados reportados con montos inferiores a la Base Mínima Contributiva.

Periodicidad: Mensual, trimestral, anual.

Objetivo: Verificar el cumplimiento del artículo 34 del Régimen IVM.

Producto esperado: Informar a las autoridades superiores con el propósito de facilitar la toma de decisiones.

N°: 2

Nombre del reporte: Cotizantes con edad menor a 18 años y mayores a 65.

Periodicidad: Mensual, trimestral, anual.

Objetivo: Comprobar que la aportación al Régimen IVM provenga de registros válidos y no sea producto de un error.

Producto esperado: Depurar la base de datos de la cuenta individual como parte del mantenimiento que se le debe brindar para contar con datos fidedignos. Adicionalmente se verificarán los derechos a los beneficiarios del Seguro de Pensiones IVM de los afiliados mayores de 65 años.

N°: 3

Nombre del reporte: Trabajadores independientes que no aparecen cotizando para el Seguro de Pensiones IVM.

Periodicidad: Mensual, trimestral, anual.

Objetivo: Verificar el cumplimiento del artículo 2° del reglamento Trabajador Independiente.

Producto esperado: Una vez aplicadas las variables, verificar si efectivamente existen afiliados que no se encuentran cotizando para el Seguro de Pensiones IVM, con el propósito de informar a las autoridades competentes, a afectos de

que actúen de acuerdo con la normativa vigente.	
N°: 4 **Nombre del reporte:** Cotizantes reportados con salarios mayores a 15 millones de colones excluyendo a los que corresponden a planillas adicionales **Periodicidad:** Mensual, trimestral, anual.	**Objetivo:** Verificar la veracidad de los montos reportados por el patrono conforme al historial salarial de los afiliados. **Producto esperado:** Informar a la Dirección de Inspección o al Área Técnica de la Sucursal correspondiente para que se realice la investigación pertinente.
N°: 5 **Nombre del reporte:** Cotizantes que registran cuotas posteriores a una fecha de defunción. **Periodicidad:** Mensual, trimestral, anual.	**Objetivo:** Identificar la existencia de cotizantes en esta condición, para evitar el otorgamiento de beneficios o que los usuarios del Sistema de Pensiones brinden información incorrecta de las cotizaciones aportadas al Seguro de Pensiones IVM. **Producto esperado:** Control y seguimiento de la información de la

base de datos de la cuenta individual del Seguro de Pensiones IVM.
Depurar la base de datos de la cuenta individual como parte del mantenimiento que se le debe brindar para contar con datos fidedignos.

información del Seguro de Pensiones IVM de forma improcedente.

N°: 6

Nombre del reporte: Afiliados que registran cotizaciones anteriores a la fecha de nacimiento.
Periodicidad: Mensual, trimestral, anual.
Objetivo: Identificar la existencia de cotizantes en esta condición, para evitar el otorgamiento de beneficios o

Producto esperado: Control y seguimiento de la información de la base de datos de la cuenta individual del Seguro de Pensiones IVM.
Depurar la base de datos de la cuenta individual como parte del mantenimiento que se le debe brindar para contar con datos fidedignos.

N°: 7

Nombre del reporte: Cotizantes registrados simultáneamente con patrono y seguro voluntario.
Periodicidad: Mensual, trimestral, anual.

Objetivo: Determinar si hubo incumplimiento del Articulo 5 de la Ley Constitutiva CCSS y evitar el otorgamiento de beneficios o que los usuarios del SIP brinden información

incorrecta de las cotizaciones aportadas al Seguro de Pensiones IVM a los clientes.

Producto esperado: Una vez aplicadas las variables determinar si efectivamente existen asegurados que presenten registros simultáneos con el propósito de realizar las modificaciones pertinentes e informar a las unidades correspondientes.

N°: 8

Nombre del reporte: Cotizantes reportados sin salario, pero que registran la cotización del Seguro de Pensiones IVM.

Periodicidad: Mensual, trimestral, anual.

Objetivo: Verificar la existencia de casos en la condición indicada.

Producto esperado: Informar a la Dirección de Inspección o al Área Técnica de la Sucursal correspondiente para que se realice la investigación pertinente.

Periodicidad: Mensual, trimestral, anual.

Objetivo: Velar por el cumplimiento de lo que se establece en el Artículo 2° del Reglamento de Pensiones de IVM.

N°: 9

Nombre del reporte: Afiliados desglosados por sector laboral que aparecen cotizando para el SEM pero no para el Seguro de Pensiones IVM.

Producto esperado: Una vez aplicadas las variables determinar la existencia de asegurados que cotizan únicamente para el SEM y deben

cotizar también para el Seguro de Pensiones IVM, con el propósito de informar a las autoridades competentes para que se realice la investigación pertinente.

N°: 10

Nombre del reporte: Anulaciones automáticas aplicadas a las cuentas individual de aquellos casos en los que los asegurados son pensionados por el Seguro de Pensiones IVM.

Periodicidad: Mensual, trimestral, anual.

Objetivo: Identificar pensiones otorgadas de forma improcedente.

Producto esperado: Control y seguimiento de la información de la base de datos de la cuenta individual del Seguro de IVM. Informar al lugar donde se otorgó el beneficio para que se realice la investigación y trámites pertinentes.

Observación de campo y reuniones

La observación de campo se logra mediante el acceso al lugar de trabajo donde las personas titulares encargadas del análisis de datos generan información a partir de los recursos que poseen. Asimismo, mucha de la información se obtiene mediante reuniones con los encargados del Área de Cuenta Individual y Control de Pagos.

Dentro de la información importante obtenida mediante estos instrumentos se obtiene lo siguiente:

Para representar el proceso actualmente que sigue la generación de alguno de estos indicadores, se construyó el siguiente caso de uso:

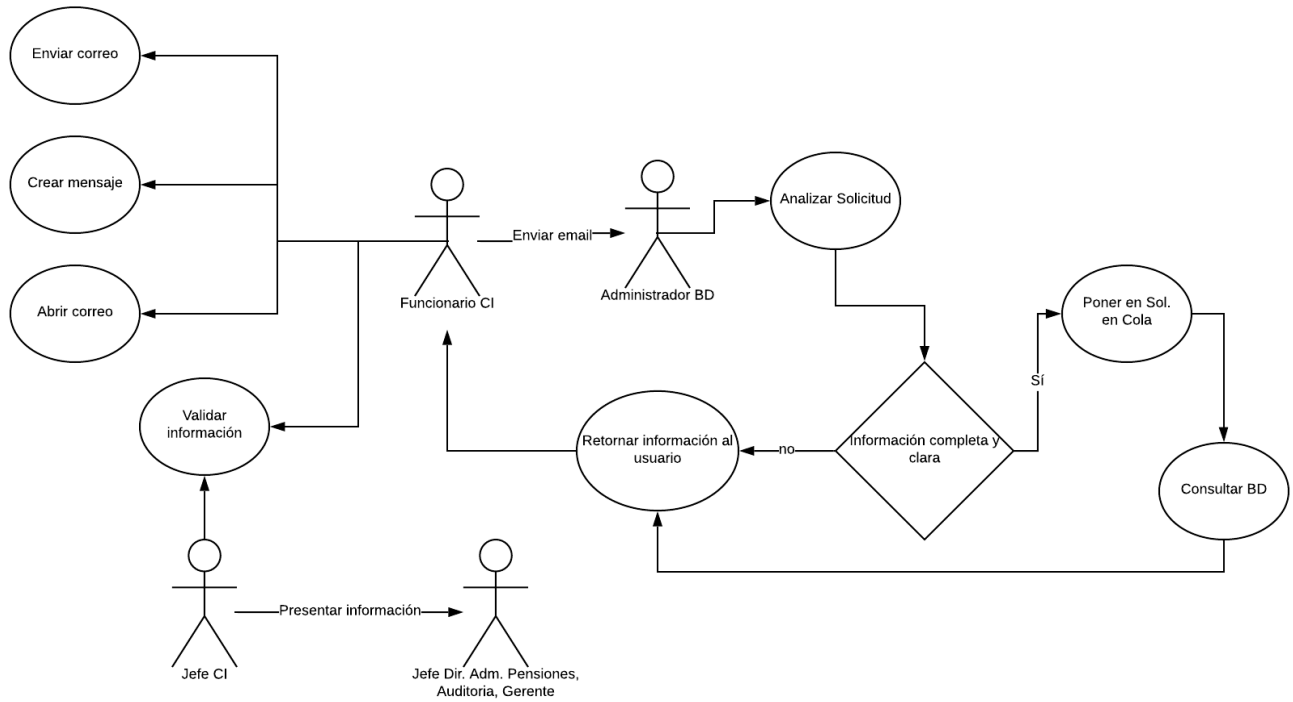


Figura 11. Caso uso en la generación de indicadores. Fuente: Propia.

- Como se puede ver en el proceso existen 4 actores relacionados directamente con la generación de esta información: Funcionario de Cuenta Individual, Administrador de Base de Datos, Jefe de Cuenta Individual y las

autoridades de las altas direcciones como el jefe de la Dirección de Administración de Pensiones.

- Se puede denotar que el medio principal para obtener está información es el correo electrónico.
- Hay que tener presente que los tiempos para generar está información puede variar dependiendo de la carga de trabajo de los administradores de base de datos. En promedio se puede decir que para generar está información es necesario al menos 3 horas hasta un día, se acuerdo con las consultas realizadas a los funcionarios del Área de Cuenta Individual.
- Si la información se requiere en otra temporalidad entonces será necesario pasar por el mismo ciclo.

Producto de este análisis se puede obtener la siguiente tabla que ilustra la cantidad y tiempos aproximados en cada una de las tareas.

N° Tarea	Nombre Tarea	Tiempo Promedio (horas)	Actor	Observaciones
1	Abrir correo y crear mensaje.	0,25	Funcionario Cuenta Individual	Tarea que inicia la solicitud de alguno de los indicadores, a fin de aplicar un informe con las variables requeridas.
2	Recepción y Análisis de la solicitud	5,5	Administrador Base de datos	Funcionarios de base de datos reciben petición, por lo que realizan el análisis respectivo para determinar si la información está clara. Para obtener el

				promedio se sumó el tiempo mínimo de 3 horas más el tiempo máximo de 8 horas laborales dividido entre dos.
3	Solicitud en cola	3	Administrador de base de datos	Tiempo promedio de las solicitudes en cola.
4	Desarrollar consulta SQL	1	Administrador de base de datos	Desarrollo propio de la consulta en SQL, se valida información y se exporta algún formato.
5	Preparar respuesta y enviar al funcionario de CI	0,25	Administrador de base de datos	Crear correo electrónico y enviar la información.
6	Recibir respuesta y validar con jefatura	2	Funcionario Cuenta Individual y jefatura	Revisar la información obtenida contra otros reportes e información manual.
7	Preparar y presentar información a las autoridades superiores	3	Jefatura Cuenta Individual	Preparar gráficos en Excel o generar PDF para presentar a las autoridades superiores de forma clara.
Total de horas		15		

Tabla 9. Tareas y tiempos en la generación de indicadores. Fuente: Propia.

De la tabla anterior se puede deducir que:

- El mayor conglomerado de tiempo lo tienen en la parte del proceso que corresponde al DBA, en lo que respecta al análisis seguido de la cola en el que la solicitud debe procesarse.
- En total son 7 tareas principales que se requieren para generar información de alto impacto y calidad, con una inversión en tiempo de 15 horas aproximadas.
- No se están contemplando muchas variables que tienen un alto impacto en los tiempos como por ejemplo que al momento de solicitar los indicadores la información no está clara y se tenga que pedir aclaraciones.
- Si bien es cierto, este tipo de tareas se puede automatizar de alguna forma, siempre existirá alguna dependencia hacia el encargado de base de datos, por ejemplo, si se necesita obtener una temporalidad diferente e inclusive en plazo de años o bien otras incluir otras cualidades a la información que no se tienen actualmente.

Con la información anterior es posible obtener el costo aproximado que consume cada actor en este proceso, basado en el salario mínimo de un profesional en sus áreas respectivas, de acuerdo con el portal de recursos humanos de la Caja Costarricense del Seguro Social.

Actor	Perfil del profesional	Salario base total (base+dedicación)	Horas Aprox. Mens	Costo Hor. Profesional	Costo mensual	Costo anual
Funcionario Cuenta Individual	Profesional 1	¢949.607	2,25	¢3.956	¢8.902	¢106.830
Funcionario Base de Datos	Analista de Sistemas 4	¢1.240.232	6,75	¢5.168	¢34.884	¢418.608
Totales		¢2.189.839	9	¢9.124	¢43.786	¢525.438

Tabla 10. Costo aproximado en la generación de indicadores actualmente. Fuente: Propia.

Para la tabla anterior es importante anotar:

- La información solo contempla los tiempos en función de un reporte o indicador que un funcionario del Área de Cuenta Individual necesita, pero eventualmente esto puede ser más de un indicador por tanto los tiempos tienden a sumar.
- Las horas aproximadas totales para los actores corresponden a la tabla 9, en los cuales existe un tiempo efectivo consumido para tal tarea y es la calculada.
- No se contempla el costo en el que incurre el actor de la jefatura de la Cuenta Individual ni el actor de los jefes superiores, como gerente o jefe de las direcciones.
- El costo por hora se obtiene con base en 30 días entre 8 horas laborales.

Una de las herramientas que permite representar gráficamente el problema que presenta el Área de Cuenta Individual y Control y Pago, en cuanto a la

utilización de una plataforma de inteligencia de negocios que le apoye en la gestión analítica para la toma de decisiones, es el diagrama de Ishikawa. Este permite visualizar las causas del problema de una forma muy explícita y al destacar lo importante.

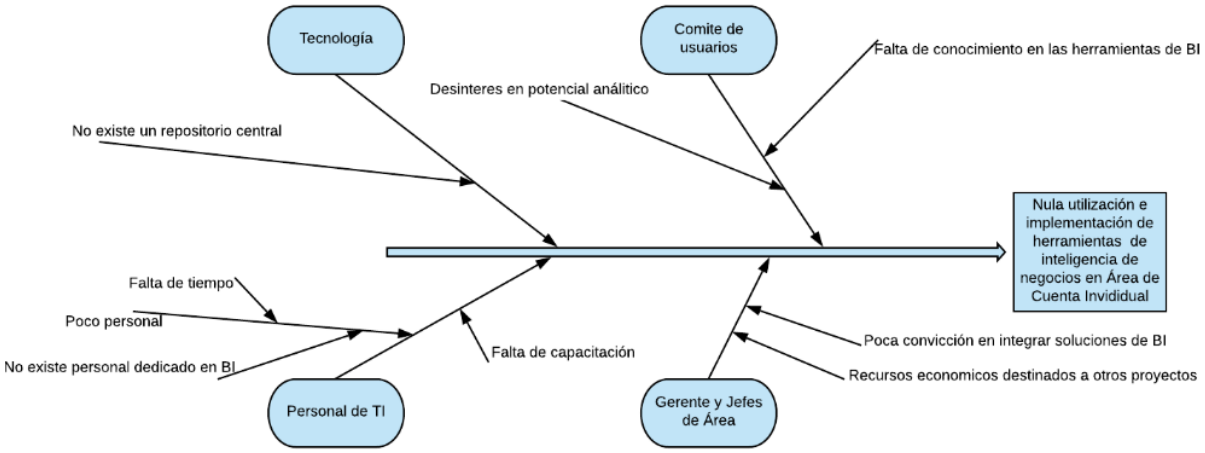


Figura 12. Diagrama Ishikawa. Fuente: Propia.

Como se puede observar en la figura 12, se puede concluir que ha sido una falta de disposición, personal y un tema de conocimiento en el potencial que ofrecen las soluciones de inteligencia de negocios para potenciar el negocio.

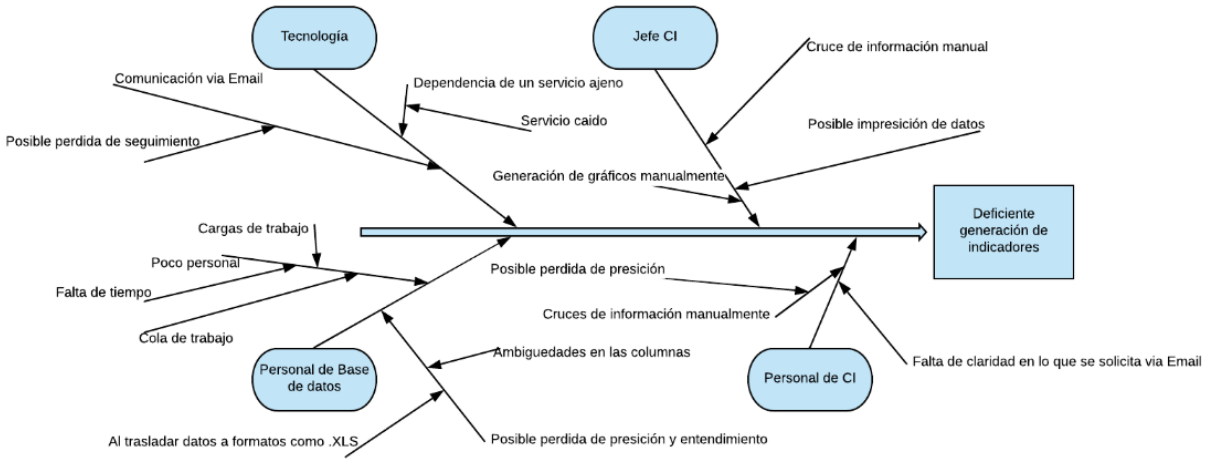


Figura 13. Diagrama Ishikawa deficiencias generando indicadores. Fuente: Propia.

La figura anterior permite observar y concluir que al existir de por medio una comunicación principal como el correo electrónico para generar estos indicadores, hay una alta probabilidad de imprecisión en los datos, dado a que se debe trasladar de una fuente a otra y depende del medio y los conocimientos de la persona pueden impactar en menor o mayor medida los datos que se generan, así como los tiempos. Los cruces de información también impactan en los tiempos dependiendo de las habilidades de las personas, por tanto, como resultado final se obtendrá una deficiente generación de indicadores que a la postre se traduce en costo de horas hombre y en el peor de los casos información errónea.

Es importante conocer que los indicadores solicitados por parte del Área de Cuenta Individual tienen relación directa con las cotizaciones que se reportan al Sistema Centralizado de Recaudación (SICERE). El cual es realizado de forma automática mes a mes. Este proceso incluye las nuevas cotizaciones reportadas a los trabajadores con un mes de retraso. Los patronos son actualizados de la misma forma mensual.

Plataforma tecnológica

Actualmente existen servidores robustos en los cuales se podría llevar a cabo una implementación de la solución.

Los indicadores que solicita el Área de Cuenta Individual y Control de Pagos se encuentran sobre la base de datos de pensiones, esta base de datos es de tipo relacional y su arquitectura opera sobre Microsoft SQL Server. Actualmente está base de datos opera sobre la versión del SQL Server 2016.

A continuación, se da la lista de herramientas que están disponibles para construir la propuesta de inteligencia de negocios para el Área de Cuenta Individual:

Herramienta	Versión	Comentarios
Microsoft SQL Server	2016 (SP1) (KB3182545) - 13.0.4001.0 (X64) Oct 28 2016 18:17:30 Copyright (c) Microsoft Corporation Developer Edition (64-bit)	
Microsoft SQL Server Analysis Services	2016	
Microsoft SQL Server Integration Services	2016	
Power BI	365	Se está gestionando la licencia de versión PRO.
Sistema operativo	Windows Server 2016 Standard 6.3 <X64> (Build 14393:)	

Tabla 11. Plataforma tecnológica actual Gerencia Pensiones. Fuente: Administradores de Base de Datos.

Como parte de la propuesta se solicitó a los encargados de base de datos un espacio en un servidor de pruebas para efectos de construir la solución de inteligencia de negocios y demostrar las capacidades de las herramientas, así como modo de trabajo.

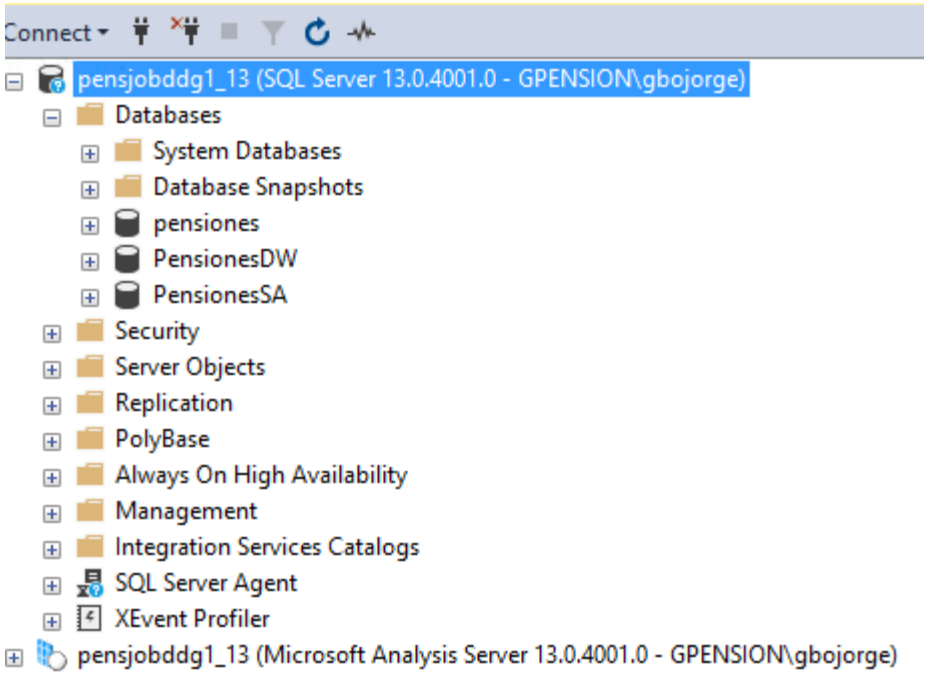


Figura 14. Imagen de servidor de prueba OLTP y Analisys Services. Fuente: Propia.

En la imagen anterior, se puede apreciar en donde estará alojado el servidor OLTP con las bases de datos de pensiones, StagingArea y Data Warehouse, así como el servidor donde estará alojada la base de datos analítica o cubos OLAP.

Hallazgos relevantes

Dentro de los hallazgos importantes de mencionar como parte del trabajo desarrollado en la presente propuesta de inteligencia de negocios, se encuentra que el proceso de calidad de los datos y diseños presentan fuertes deficiencias. Producto del análisis se encontró lo siguiente:

- Para la tabla de cotizaciones, existen campos que no tienen catálogo, estos a su vez están “quemados” dentro de la aplicación.
- Existe un catálogo de actividades económicas de los patronos, este catálogo contiene información repetida y no da mayor valor a la información.

- Al momento de hacer las cargas de las cotizaciones con SICERE, existe un campo que se carga como un comentario dentro de otro campo. Este campo es esencial para diferenciar las planillas ordinarias de las adicionales.
- No existen “Check Constraint” en el nivel de columna para evitar la introducción de estados o tipos incorrectos.
- Existen valores en los catálogos que el encargado de este proceso desconoce.
- La carga de las dimensiones de los cantones y distritos se realizó con el catálogo del sistema de SICERE, dado a que estos presentan incompatibilidad con los códigos del Sistema Integrado de Pensiones (SIP).

Capítulo 5. Propuesta de Solución

El desarrollo de la propuesta se estructura de acuerdo con los indicadores que se brindaron en el capítulo anterior. De modo que está información es verificada con el usuario primeramente a fin de evacuar dudas y plasmar el diseño requerido para el Data Warehouse.

El desarrollo de la propuesta seguirá los siguientes apartados en función de guardar un orden con el desarrollo de la solución.

Apartado	Aspectos contemplados
Desarrollo de consultas OLAP	• Verificación y comprensión de los indicadores. • Reunión con los usuarios encargados de la información a fin de aclarar dudas y cálculos. • Creación de consultas Transact-SQL.
Diseño de StagingArea	• Basado en las consultas Transact-SQL determinar las tablas o funciones requeridas para crear la estructura del StagingArea. • Transformaciones requeridas.
Diseño de Data Warehouse	• Una vez identificado los hechos y dimensiones, se procede a diseñar el modelo idóneo para el Data Warehouse. • Construcción de la estrella. • Validación de campos contra StagingArea.

Proyecto Integration Services	Conformación del proyecto de Integration Services. ETL de OLTP a StagingArea. ETL de StagingArea a Data Warehouse.
Proyecto Analisys Services	Creación del proyecto de Analisys Services, se trasladan los hechos y dimensiones requeridas para la creación del cubo OLAP. Definir perspectivas. Ajuste de nombres de columnas. Agregar columnas calculadas.
Proyecto visualización de datos	Mediante Power BI en su versión de escritorio generar los indicadores en la forma que el usuario los necesita. (Dashboard). Pruebas de Power BI 365 con el usuario.

Tabla 12. Etapas del desarrollo del modelo inteligencia de negocios. Fuente: Propia.

Es importante recordar que la solución será desarrollada en un modelo tabular, asimismo se trabajará mediante proyectos de *Integration Services*, *Analisys Services*, y *Power BI*. El diseño por el cual se opta es el de Ralph Kimball, para conformar la estrella del Data Warehouse. Así como la inclusión del componente básico del Staging Area, en donde se estandarizan, limpian y combinan datos.

Desarrollo de consultas OLAP

Para el desarrollo de las consultas Transact-SQL, se utiliza la herramienta SQL Server Manageament Studio versión 17.9.1. Las consultas se elaboran con base en la optimización con ayuda de la herramienta integrada en Manageament Studio llamada *"Execution Plan"* y estructuración modular, de modo que muchos de los datos puedan ser obtenidos mediante segmentos de consulta específicos a fin de flexibilizar los mantenimientos a futuro, el enfoque sea más especializado y que estos fragmentos de consulta puedan ser reutilizados. Para tal fin se optó por utilizar funciones escalares ("Scalar-valued funtions") y funciones de tabla en línea (Inline Table-valued functions").

A continuación, se muestran del listado de los indicadores proporcionados por el Área de Cuenta Individual y Control de Pagos, las consultas más relevantes a fin de proporcionar documentación del trabajo realizado.

```
2-Cotizantes con edad menor a 18 años - mayores a 65 años*/
Select
pe.tip_identificacion,
pe.num_identificacion,
pe.nom_apellido1,
pe.nom_apellido2,
pe.nom_nombre,
pe.fec_nacimiento,
pensiones.fn_edad_anos(pe.fec_nacimiento,GETDATE()) 'Años',
case when isnull(pa.tip_patrono,-1)=0 then 'Patrono fisico nacional'
when isnull(pa.tip_patrono,-1)=2 then 'Patrono juridico'
when isnull(pa.tip_patrono,-1)=2 then 'Patrono juridico'
when isnull(pa.tip_patrono,-1)=2 then 'Patrono juridico'
else 'Desconocido' end 'Tipo de patrono',
isnull(pa.num_patrono,-1) 'Número de patrono',
isnull(pa.num_segregado,-1) 'Número de segregado',
isnull(pa.num_sector,-1) 'Número de sector',
isnull(pa.cod_razon_social,'Desconocido') 'Nombre Patrono',
co.can_salario_reportado Salario,
pensiones.fn_pensionado_IVM_RNC_OTROS_REGIMENES(pe.tip_identificacion,pe.num_identificacion) 'Pensionado otro regimen',
pensiones.fn_Cant_cuotas_IVM_Persona(pe.tip_identificacion,pe.num_identificacion) 'Cantidad de cuotas validas'
from #TmpCotizaciones co
join pensiones.pens_personas pe on
 co.num_identificacion=pe.num_identificacion and
co.tip_identificacion=pe.tip_identificacion
left join pensiones.pens_patronos pa on
co.num_patrono=pa.num_patrono and co.tip_patrono=pa.tip_patrono
and co.num_segregado=co.num_segregado and co.num_sector=co.num_sector
left join pensiones.pens_unidades_programaticas up on
pa.cod_unidad_programatica=up.cod_unidad_programatica
where --convert(char(06),fec_inicio_periodo_cotizacion,112) = '201904'
 isnull(co.cod_estado,'A') not in ('L','DT','ST','DT')
 and co.tip_cuota <> 'M' and
--and pa.num_segregado not in (998,999)
--998 Asegurados voluntarios
--999 Trabajadores independientes
 (convert(int,pensiones.fn_edad_anos(pe.fec_nacimiento,GETDATE()))<=18 and
    (convert(int,pensiones.fn_edad_anos(pe.fec_nacimiento,GETDATE()))>=65))
order by tip_identificacion,num_identificacion
```

Figura 15. Consulta Transact-SQL de Cotizantes con edad menor a 18 y mayores a 65 años. Fuente: Propia.

Como se puede notar en la figura 11, la consulta realizada utiliza dos funciones escalares a fin de obtener si el cotizantes posee pensiones sobre otros regímenes, además de la cantidad de cuotas validas que tiene está persona en estas condiciones.

Asimismo, se crearon funciones de apoyo para la obtención de fechas, basado en los parámetros de consulta que utiliza las instrucciones Transact-SQL, para las cualidades de datos del indicador de los trabajadores independientes.

```
3-Trabajadores independientes que no aparecen cotizando para el seguro de pensiones de IVM */
Select
pe.tip_identificacion,
pe.num_identificacion,
pe.nom_apellido1,
pe.nom_apellido2,
pe.nom_nombre,
pe.fec_nacimiento,
pensiones.fn_edad_anos(pe.fec_nacimiento,GETDATE()) edad,
case when isnull(pa.tip_patrono,-1)=0 then 'Patrono fisico nacional'
when isnull(pa.tip_patrono,-1)=2 then 'Patrono juridico'
when isnull(pa.tip_patrono,-1)=2 then 'Patrono juridico'
when isnull(pa.tip_patrono,-1)=2 then 'Patrono juridico'
else 'Desconocido' end 'Tipo de patrono',
isnull(pa.num_patrono,-1) num_patrono,
isnull(pa.num_segregado,-1) num_segregado,
isnull(pa.num_sector,-1) num_sector,
isnull(pa.cod_razon_social,'Desconocida') 'Nombre Patrono',
co.can_salario_reportado,
cls.cod_seguro+'-'+cls.dsc_seguro clase_seguro,
pensiones.fn_fecha_ingreso_cotizante(co.num_segregado,co.tip_identificacion,co.num_identificacion) 'Fecha de ingreso',
pensiones.fn_Cant_cuotas_IVM_persona_fecha
(co.tip_identificacion,co.num_identificacion,pensiones.fn_fecha_ingreso_cotizante(co.num_segregado,co.tip_identificacion,co.num_identificacion)) 'Cuotas antes de ser TI',
pensiones.fn_edad_anos(pe.fec_nacimiento,pensiones.fn_fecha_ingreso_cotizante(co.num_segregado,co.tip_identificacion,co.num_identificacion)) 'Edad al ingresar como TI',
pensiones.fn_cant_cuotas_ivm_Persona_TI(pe.tip_identificacion,pe.num_identificacion) 'Cuotas reales siendo TI',
 (co.can_salario_reportado) 'Monto de referencia siendo TI',
 (co.can_cuota_ivm) cuota_ivm,
(co.can_cuota_sem) cuota_sem
from #TmpCotizaciones co
join pensiones.pens_personas pe on
 co.num_identificacion=pe.num_identificacion and
co.tip_identificacion=pe.tip_identificacion
left join pensiones.pens_patronos pa on
co.num_patrono=pa.num_patrono and co.tip_patrono=pa.tip_patrono
and co.num_segregado=co.num_segregado and co.num_sector=co.num_sector
left join pensiones.pens_unidades_programaticas up on
pa.cod_unidad_programatica=up.cod_unidad_programatica
left join pensiones.pens_clases_seguro_cotizaciones cls
on cls.cod_seguro=co.tip_cuota
where --convert(char(06),fec_inicio_periodo_cotizacion,112) = '201904'
isnull(cod_estado,'A') not in ('L','DT','ST','DT')
and co.tip_cuota not in ('M','E','A','H','N','P','R','W')
and pa.num_segregado in (999)
--998 Asegurados voluntarios
--999 Trabajadores independientes
and  co.can_cuota_ivm<=0
order by pe.tip_identificacion,pe.num_identificacion
```

Figura 16. Consulta SQL Trabajadores independientes que no aparecen cotizando para el seguro de pensiones IVM. Fuente: Propia.

Para el indicador de los “Cotizantes registrados simultáneamente con patrono y seguro voluntario”, los datos se obtuvieron con el apoyo de tablas temporales y la utilización del comando Transact-SQL llamado *Intersect*. Este comando permite obtener los datos con información de iguales de una primera consulta contra otra que contenga los mismos metadatos y estos coincidan.

```
select *  into #tmpSimultaneos
from
(
 select pe.tip_identificacion tip_identificacion,
pe.num_identificacion num_identificacion,
@nom_nombre nom_nombre,@nom_apellido1 nom_apellido1,@nom_apellido2 nom_apellido2,@tip_patrono tip_patrono, @num_patrono num_patrono,
@num_segregado num_segregado,@num_sector num_sector,@nom_patrono nom_patrono,@cant_salario_reportado_patrono cant_salario_reportado_patrono,
@cant_salario_reportado_seguro_voluntario cant_salario_reportado_seguro_voluntario,@pensionado_otro_regimen pensionado_otro_regimen
 from #TmpCotizaciones co
 join pensiones.pens_personas pe on
 co.num_identificacion=pe.num_identificacion and
co.tip_identificacion=pe.tip_identificacion
left join pensiones.pens_patronos pa on
co.num_patrono=pa.num_patrono and co.tip_patrono=pa.tip_patrono
and co.num_segregado=pa.num_segregado and co.num_sector=pa.num_sector
 where co.num_segregado in (998)
 and isnull(cod_estado,'A') not in ('L','DT','ST','DT')
 and co.tip_cuota<>'M'
 -- que no sea pensionado
 -- que sea una cuota valida
 group by  pe.tip_identificacion,
pe.num_identificacion


 intersect

 select pe.tip_identificacion tip_identificacion,
pe.num_identificacion num_identificacion,
@nom_nombre nom_nombre,@nom_apellido1 nom_apellido1,@nom_apellido2 nom_apellido2,@tip_patrono tip_patrono,@num_patrono num_patrono,
@num_segregado num_segregado,@num_sector num_sector,@nom_patrono nom_patrono,@cant_salario_reportado_patrono cant_salario_reportado_patrono,
@cant_salario_reportado_seguro_voluntario cant_salario_reportado_seguro_voluntario,@pensionado_otro_regimen pensionado_otro_regimen
 from #TmpCotizaciones co
 join pensiones.pens_personas pe on
 co.num_identificacion=pe.num_identificacion and
co.tip_identificacion=pe.tip_identificacion
left join pensiones.pens_patronos pa on
co.num_patrono=pa.num_patrono and co.tip_patrono=pa.tip_patrono
and co.num_segregado=pa.num_segregado and co.num_sector=pa.num_sector
 where co.num_segregado not in (998)
  and isnull(cod_estado,'A') not in ('L','DT','ST','DT')
 and co.tip_cuota<>'M'
 -- que no sea pensionado
 -- que sea una cuota valida
 group by  pe.tip_identificacion,
pe.num_identificacion
 ) as t
```

Consulta SQL Indicador de cotizantes registrados simultáneamente con patrono y seguro voluntario. Fuente: Propia.

Cada una de estas consultas de desarrollo en un servidor de pruebas y se validó la información obtenida con el usuario a efectos de consolidar las bases para el diseño del StagingArea y Data Warehouse.

Diseño de StagingArea

Con la base de las consultas para cada uno de los indicadores y con los datos validados por parte de los usuarios de la cuenta individual, se procede a desarrollar el diseño del StagingArea como un componente recomendado de incluir en las soluciones de inteligencia de negocios a fin de proporcionar un estado de los datos de forma pura, para poder realizar las posteriores extracciones, transformaciones y cargas requeridas por el diseño del Data Warehouse.

Una vez analizada la información se procede a conformar las tablas que participan en este diseño. Las cuales son:

- Actividades económicas patronos*
- Cantones
- Provincias
- Distritos
- Clases de seguro*
- Cotizaciones
- Clase de patronos
- Estado de cotizaciones*
- Patronos
- Personas
- Sexo
- Tipos de estado civil*
- Tipos de identificación
- Tipos de patrono*
- Tipos de trabajadores*

Para cada una de las tablas marcadas con *, estas no existen en el nivel del sistema transaccional, sino que fueron construidas a partir del conocimiento y los datos obtenidos requeridos para conformar las dimensiones que el usuario podría utilizar a fin de realizar una analítica más profunda sobre sus datos.

Al momento de construir cada una de estas tablas, se desmarcan de cualquier llave foránea o primaria que tuvieran, solo conservando la anulabilidad del dato que caracteriza tal y como funciona actualmente.

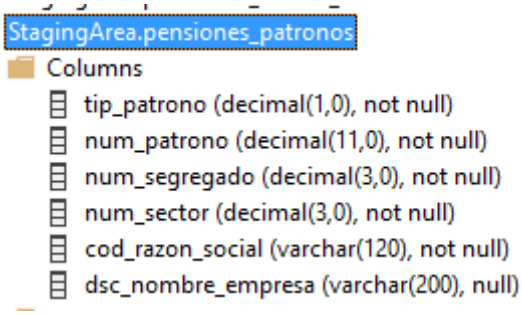


Figura 17. Tabla Patronos en StagingArea. Fuente: Propia.

Como se puede observar, en este caso para la tabla patronos, se conserva los valores nulos para el campo “dsc_nombre_empresa”, de modo que se preserva consistencia de datos desde esta entrada, hasta llegar a la carga de datos del Data Warehouse.

El diseño del StagingArea también se contempló la longitud de los campos y tipos de datos del sistema transaccional de modo que se evitara truncamiento de datos o errores en las cargas de las tablas.

Asimismo, se presenta la tabla dentro del StagingArea que fungirá como la de hechos, en este caso será la de “pens_cotizaciones”. Esta tabla es donde se almacenan mensualmente las aportaciones de los trabajadores desde SICERE, además se cargan las aportaciones diarias de otras fuentes al sistema transaccional.

El diseño detallado quedó conformado de la siguiente forma:

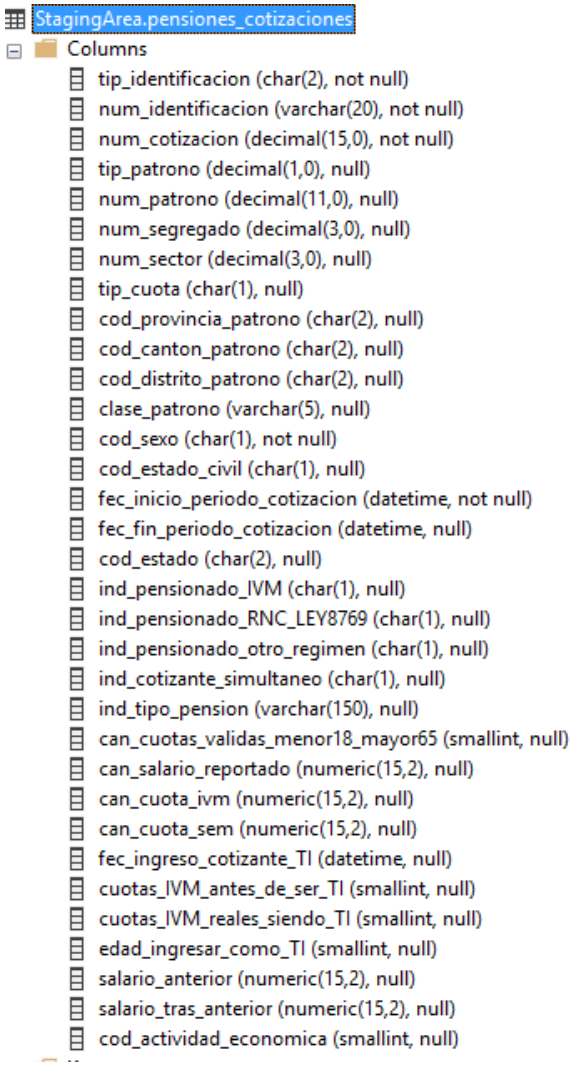


Figura 18. Diseño tabla hechos StagingArea. Fuente: Propia.

Para este diseño fue necesario establecer “banderas” o “indicadores” que proporcionaran características propias de un conjunto de datos, según lo establecieron en los indicadores, por lo cual estos se crearon con un nombre de columna explícito a fin de que sea fácilmente interpretables. Como por ejemplo “ind_pensionado_IVM” el cual indica que si la persona está pensionada por el seguro de pensiones IVM o bien el indicador llamado “ind_tipo_pension” el cual contiene el detalle del riesgo por la cual la persona está pensionada.

Asimismo, estos indicadores se cargan solamente si se cumplen las condiciones, si no es así entonces el dato no se carga. Por ejemplo, el campo

denominado “can_cuotas_validas_menor18_mayr65”, proporciona información sobre la cantidad de cuotas validas que tiene una persona menor de 18 años o mayor a 65.

Los datos de los indicadores que corresponden a los trabajadores independientes se denotan mediante la abreviación de “TI “(Trabajador Independiente). Como por ejemplo los campos “fec_ingreso_cotizante_TI” o bien el campo “cuotas_IVM_antes_de_ser_TI”.

Diseño DataWarehouse

Con el diseño del StagingArea finalizado, se procede con la elaboración del diseño del Data Warehouse. Para lo cual se conforman las dimensiones identificadas y los hechos que corresponden a los indicadores proporcionados.

La estrella se conforma siguiendo el principio de Ralph Kimball en su diseño, debido principalmente a su eficiencia en las consultas y disminución de dimensiones en el modelo final. Su desarrollo de planteo la utilización de nombres estándares y un esquema alusivo al entorno donde se encuentra. Dentro de lo que se destaca:

- El esquema que se utilizó se denominó “DataWarehouse”.
- El nombre de las dimensiones se identifica con tres caracteres “dim” y un guion bajo “_”.
- Todas las estructuras conservan un nombre en plural.
- El nombre de los indicadores se denota con las iniciales “ind_”.
- La tabla de hechos se identifica mediante las iniciales “he_”.

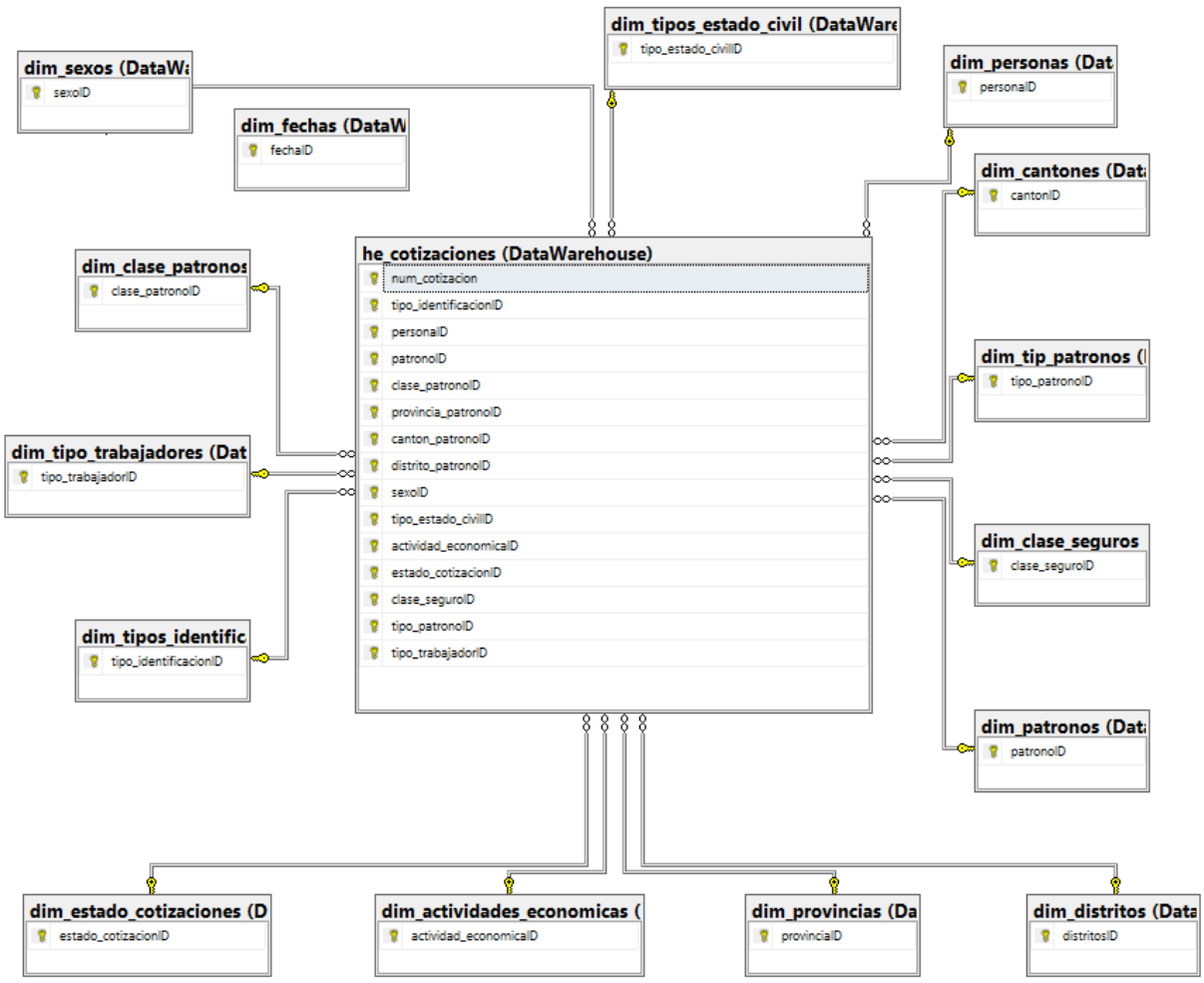


Figura 19. Modelo estrella DataWarehouse. Fuente: Propia.

Como se puede apreciar en la figura anterior, el modelo estrella se conforma por 15 dimensiones y una tabla de hechos. Para la tabla fechas no se contempla referencia integral dado a que, al existir varias fechas en la tabla de hechos, se opta por conformar varias dimensiones de fechas, pero en el nivel de modelo del proyecto de Analisys Services.

La conformación del detalle de los hechos del modelo estrella puede verse en la siguiente figura:

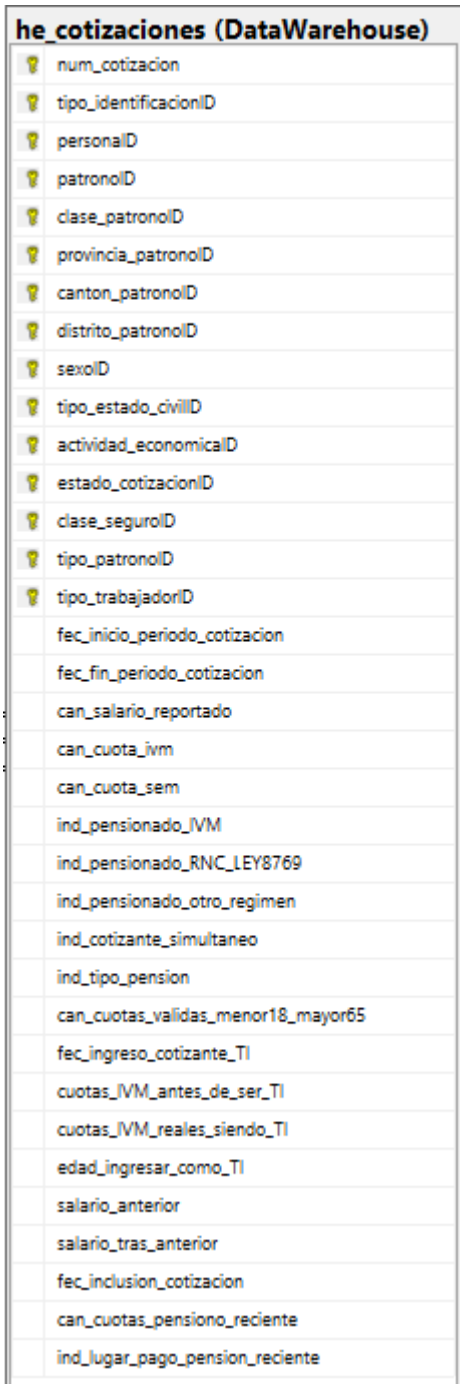


Figura 20. Vista detalle tabla hechos. Fuente: Propia.

La tabla se conforma por 35 columnas, de las cuales 15 corresponden a dimensiones y las otras 20 corresponden a las métricas y características cualitativas del hecho.

Para cada una de estas cualidades del hecho se proporciona información referente sobre el significado de cada uno de los campos:

Nombre columna	Descripción
fec_inicio_periodo_cotizacion	Fecha en la que la persona se le reporta salario.
fec_fin_periodo_cotizacion	Fecha final que corresponde a ese periodo de cotización.
can_salario_reportado	Monto salarial reportado
can_cuota_ivm	Monto que corresponde al seguro IVM
can_cuota_sem	Monto reportado al Seguro de enfermedad y maternidad
ind_pensionado_ivm	Indica si la persona es pensionada por IVM
ind_pensionado_rnc_ley8769	Indica si la persona es pensionada por el Régimen No Contributivo o por la Ley 8769 (PCP).
ind_pensionado_otro_regimen	Indica si la persona esta pensionada por otro régimen como Hacienda o el Magisterio Nacional.
ind_cotizante_simultaneo	Indica si la persona fue reportada como asalariada y con seguro voluntario.
ind_tipo_pension	Indica el detalle de las pensiones que posee esa persona. Una concatenación de los riesgos.
can_cuotas_validas_menor18_mayor65	Obtiene las cuotas validas de las personas menores a 18 años y mayores a 65.
fec_ingreso_cotizante_ti	Fecha de ingreso como trabajador independiente.
cuotas_ivm_antes_de_ser_ti	Cuotas IVM validas antes de ser trabajador independiente.
edad_al_ingresar_como_ti	Edad de la persona al ingresar como trabajador independiente.
salario_anterior	Salario anterior valido reportado a la persona.
salario_tras_anterior	Salario tras anterior valido reportado a la persona.
fec_inclusion_cotizacion	Fecha en la que se incluyó la cotización.
can_cuotas_pensiono_reciente	Cantidad de cuotas con las que la persona se pensiono
Ind_lugar_pago_pension_reciente	Lugar de pago de la pensión más reciente activa.

Tabla 13. Detalle de cualidades en tabla hechos. Fuente: Propia.

Asimismo, se proporciona información referente a cada una de las dimensiones creadas:

Dimensión	Descripción
Dim_actividades_economicas	Dimensión que proporciona información referente a las actividades económicas asociadas a los patronos. Ejemplo: "CULTIVO DE CEREALES Y OTROS CULTIVOS N.C.P."
Dim_cantones	Dimensión de cantones donde está registrado el patrono.
Dim_provincias	Dimensión de provincia donde fue registrado el patrono.
Dim_distritos	Dimensión de distritos donde fue registrado el patrono.
Dim_clase_patronos	Información asociada del patrono que indica su clase. Ejemplo: "ASEGURADO VOLUNTARIO" o "SERVICIO DOMESTICO"
Dim_clase_seguros	Información sobre la clase de seguro que se está reportando. Ejemplo: "COTIZA SEM, IVM Y BPO"
Dim_estado_cotizaciones	Información sobre el estado de las cotizaciones. Ejemplo: COTIZACIÓN ACTIVA- COTIZACIÓN INGRESADA POR LA CARGA MENSUAL
Dim_fechas	Dimensión sobre las fechas
Dim_patronos	Información sobre el patrono, como su número de patrono y razón social.
Dim_sexos	Información sobre el sexo de las personas.
Dim_tip_patronos	Dimensión sobre los tipos de patronos. Ejemplo: "Patrono físico nacional"
Dim_tipo_trabajadores	Dimensión sobre los tipos de trabajadores, se relaciona con el número de segregado. Ejemplo: ASALARIADO.
Dim_tipos_estado_civil	Dimensión sobre los tipos de estado civil de la persona.
Dim_tipo_trabajadores	Dimensión que indica el tipo de trabajado cotizante: "ASALARIADO", "ASEGURADO VOLUNTARIO"
Dim_tipos_identificacion	Dimensión sobre los tipos de identificación de la persona.

Tabla 14. Descripción sobre las dimensiones. Fuente: Propia.

Proyecto Integration Services

El proyecto de Integration Services consistió en la captura y extracción de los datos para cada una de las dimensiones identificadas y hechos. Para tal fin se conformó una solución que contuviera las características de forma en la cual se construyera la base para dar mantenimiento y construir futuras mejora al proyecto.

El diseño se desarrolló de forma modular, previendo la integración de más información que se requiera incluir como parte de esta solución u otras que los funcionarios de la plataforma tecnológica de la Gerencia de Pensiones considere.

El proyecto se conformó de la siguiente forma:



Figura 21. Proyecto Integration Services. Fuente: Propia.

De tal forma que se construyen paquetes separados según el contexto del funcionamiento. En este caso Carga_Hechos_DataWarehouse y Carga_Hechos_StagingArea, son paquetes encargados de cargas los hechos respectivamente en sus tablas. Así mismo los paquetes llamados Carga_Dimensiones_DataWarehouse y Carga_Dimensiones_StagingArea se encargan de proporcionar las cargas para las dimensiones identificadas.

ETL StagingArea

El proceso de carga para las dimensiones del StagingArea se conforma de la siguiente forma:

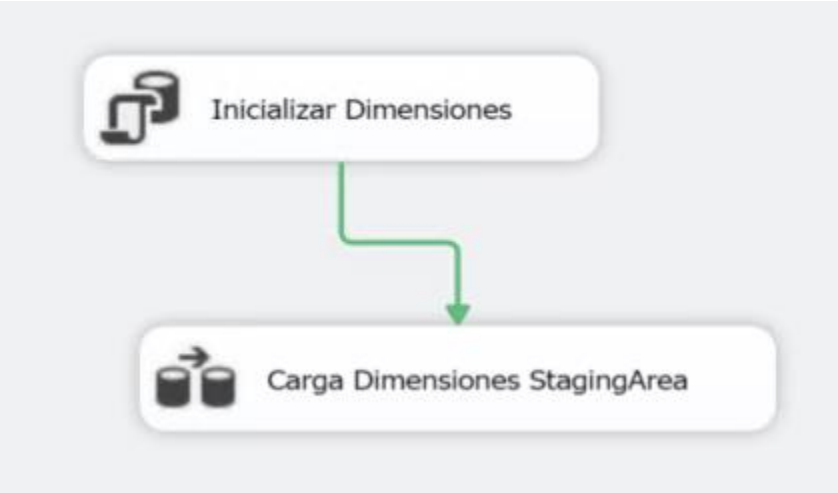


Figura 22. Flujo de control carga de dimensiones StagingArea. Fuente Propia.

Para la inicialización de las dimensiones en esta fase se utiliza un procedimiento almacenado que trunca todas las tablas a fin de cargar la información reciente de las dimensiones. El procedimiento almacenado posee la siguiente forma:

```
1   use PensionesSA
2   go
3   CREATE PROCEDURE StagingArea.spInicializaDimensionesSA
4   AS
5   BEGIN
6   truncate table StagingArea.[pensiones_tipos_identificacion]
7   truncate table StagingArea.[pensiones_personas]
8   truncate table StagingArea.[pensiones_patronos]
9   truncate table StagingArea.[pensiones_clase_patronos]
10  truncate table StagingArea.[pensiones_provincias]
11  truncate table StagingArea.[pensiones_cantones]
12  truncate table StagingArea.[pensiones_distritos]
13
14  END
```

Figura 23. Estructura procedimiento almacenado inicializa dimensiones StagingArea. Fuente: Propia.

Para el flujo de datos que corresponde a la “Carga Dimensioens StagingArea” se conforman las cargas mediante el origen de datos OLEDB, por sus características en cuanto a rendimiento y opciones de carga en el destino.

Para lo cual se extraen los datos en limpio del origen transaccional hacia el destino de StagingArea.

Una vista parcial del trabajo realizado se puede visualizar desde está imagen:



Figura 24. Vista Parcial de la carga StagingArea. Fuente: Propia.

Posteriormente se procede a obtener los hechos que conforman el núcleo del DataWarehouse, por lo cual se construye la consulta Transact-SQL que proporcionará estos datos.

La consulta se desarrolló dentro de la estructura de un procedimiento almacenado, con el fin de poder tener los datos mejor organizados y además evitar hacer cambios en el proyecto de análisis ante futuras modificaciones a variables existentes.

Dentro del proceso para capturar las variables requeridas por los hechos del Data Warehouse, se conforma una tabla temporal como marca de inicio a cada una de las variables, de forma que estas se vayan trabajando conforme se realizan las capturas y actualizaciones de la información.

```
create table #tmp_cotizaciones
(
[tip_identificacion] [char](2)  not NULL,
[num_identificacion] [varchar](20) not  NULL,
[num_cotizacion] [decimal](15, 0) not NULL,
[tip_patrono] [decimal](1, 0) NULL,
[num_patrono] [decimal](11, 0) NULL,
[num_segregado] [decimal](3, 0) NULL,
[num_sector] [decimal](3, 0) NULL,
[tip_cuota] [char](1) NULL,
[cod_provincia_patrono] [char](1) NULL,
[cod_canton_patrono] [char](2) NULL,
[cod_distrito_patrono] char(03) null,
[clase_patrono] [varchar](5) NULL,
[cod_sexo] [char](1) NULL,
[cod_estado_civil] [char](1) NULL,
[fec_inicio_periodo_cotizacion] [datetime] NOT NULL,
[fec_fin_periodo_cotizacion] [datetime] NULL,
[cod_estado] [char](2) NULL,
[ind_pensionado_IVM] [char](1) NULL,
[ind_pensionado_RNC_LEY8769] [char](1) NULL,
[ind_pensionado_otro_regimen] [char](1) NULL,
[ind_cotizante_simultaneo] [char](1) NULL,
[ind_tipo_pension] [varchar](150) NULL,
[can_cuotas_validas_menor18_mayor65] [smallint] NULL,
[can_salario_reportado] [numeric](15, 2) NULL,
[can_cuota_ivm] [numeric](15, 2) NULL,
[can_cuota_sem] [numeric](15, 2) NULL,
[fec_ingreso_cotizante_TI] [datetime] NULL,
[cuotas_IVM_antes_de_ser_TI] [smallint] NULL,
[cuotas_IVM_reales_siendo_TI] [smallint] NULL,
[edad_ingresar_como_TI] [smallint] NULL,
[salario_anterior] [numeric](15, 2) NULL,
[salario_tras_anterior] [numeric](15, 2) NULL,
[cod_actividad_economica] smallint NULL,
[fec_inclusion_cotizacion] datetime,
[tip_trabajador] char(2),
[can_cuotas_pensiono_reciente] int NULL,
[ind_lugar_pago_pension_reciente] VARCHAR(100) NULL
CONSTRAINT [pk_cotizaciones] PRIMARY KEY CLUSTERED
(
    [tip_identificacion] ASC,
    [num_identificacion] ASC,
    num_cotizacion asc
)
)
```

Figura 25. Vista parcial del procedimiento carga hechos StagingArea. Fuente: Propia.

La conformación de los campos se da basado en las columnas existentes de la tabla de hechos del StagingArea. Por efecto de rendimiento se colocan llaves

primarias sobre esta, a fin de que las actualizaciones o búsquedas sean más rápidas.

Para efectos de prueba de la solución se cargaron los datos de las cotizaciones ingresadas desde enero hasta julio del 2019. Se obtienen los datos relevantes y los demás se dejan en valor nulo para posteriormente cargarlos. En promedio se tienen por carga mensual un aproximado de 2 millones de registros. En total se cargaron para los meses del 2019 un total de 14.192.215 registros.

```
insert into #tmp_cotizaciones
select
[tip_identificacion]=pe.tip_identificacion,
[num_identificacion]=pe.num_identificacion,
[num_cotizacion]=co.num_cotizacion,
[tip_patrono]=co.tip_patrono,
[num_patrono]=co.num_patrono,
[num_segregado]=co.num_segregado,
[num_sector]=co.num_sector,
[tip_cuota]=co.tip_cuota,
[cod_provincia_patrono]=NULL,
[cod_canton_patrono]=NULL,
cod_distrito_patrono=null,
[clase_patrono]=pa.cls_patrono,
[cod_sexo]=pe.cod_sexo,
[cod_estado_civil]=pe.cod_estado_civil,
[fec_inicio_periodo_cotizacion]=co.fec_inicio_periodo_cotizacion,
[fec_fin_periodo_cotizacion]=co.fec_fin_periodo_cotizacion,
[cod_estado]=co.cod_estado,
[ind_pensionado_IVM]=null,
[ind_pensionado_RNC_LEY8769]=NULL,
[ind_pensionado_otro_regimen]=NULL,
[ind_cotizante_simultaneo]=NULL,
[ind_tipo_pension]=NULL,
[can_cuotas_validas_menor18_mayor65]=NULL,
[can_salario_reportado]=co.can_salario_reportado,
[can_cuota_ivm]=co.can_cuota_ivm,
[can_cuota_sem]=co.can_cuota_sem,
[fec_ingreso_cotizante_TI]=null,
[cuotas_IVM_antes_de_ser_TI] = NULL,
[cuotas_IVM_reales_siendo_TI] = NULL,
[edad_ingresar_como_TI] = NULL,
[salario_anterior] = null,
[salario_tras_anterior] = null,
cod_actividad_economica=null,
[fec_inclusion_cotizacion] =fec_inclusion_cotizacion
 from pensiones.pens_cotizaciones co
 join pensiones.pens_personas pe on
 co.num_identificacion=pe.num_identificacion and co.tip_identificacion=pe.tip_identificacion
 join pensiones.Pensiones.pens_patronos pa on
 pa.num_patrono=co.num_patrono and pa.tip_patrono=co.tip_patrono and
 pa.num_sector=co.num_sector and pa.num_segregado=co.num_segregado
 where
convert(char(06),fec_inclusion_cotizacion,112) in ('201905')
--convert(date,co.fec_inclusion_cotizacion)='20190505'
and tip_cuota<>'M'
```

Figura 26. SQL de carga hechos OLTP a StagingArea. Fuente: Propia.

Como parte de este proceso es importante mencionar que se están excluyendo las cotizaciones que dentro de su clase de seguro indica “pensionado

con su propio sistema", dado a que es información que no es de interés en este momento. Asimismo, se están utilizando funciones de tabla en línea. Esto son funciones que retornan tablas basadas en parámetros, o visto de otra forma son vistas que permiten parámetros. La ventaja que tienen estas tablas es que pueden utilizarse de forma derivada tal y como se podría hacer con una tabla normal. Para esto es necesario utilizar la función "cross apply" de tal forma que permite utilizar los sub-columnas dentro de la función y determinar condiciones sobre los datos.

Ejemplo de ello son los indicadores de pensionado por RNC o LEY 8769, tal y como muestra la siguiente imagen:

```
--indicador de pensionado por RNC y LEY8769
update a
set a.ind_pensionado_RNC_LEY8769='S'
from #tmp_cotizaciones a
cross apply pensiones.StagingArea_pensionado_IVM_RNC
(a.tip_identificacion,a.num_identificacion,'RNC') pensionado_rnc
where
pensionado_rnc.tip_identificacion=a.tip_identificacion and
pensionado_rnc.num_identificacion=a.num_identificacion and
pensionado_rnc.ind_riesgo='RNC'
```

Figura 27. Carga de indicador pensionado por RNC o Ley8769. Fuente: Propia.

ETL Data Warehouse

Para la carga de las dimensiones del Data Warehouse, se utiliza el mismo componente de origen OLEDB y el componente dimensión de variación lenta (*Slow Changing Dimension*) para efectos de cargar nueva información o actualizar la existente. La siguiente imagen ilustra el flujo de control que posee la carga de dimensiones del Data Warehouse.



Figura 28. Flujo de control carga dimensiones Data Warehouse. Fuente Propia.

Como se pude notar en la imagen anterior, primeramente, se carga la dimensión de las fechas, posteriormente se cargan las dimensiones de los hechos. La siguiente imagen muestra el detalle de la carga para las dimensiones del estado de las cotizaciones y las clases de seguro.

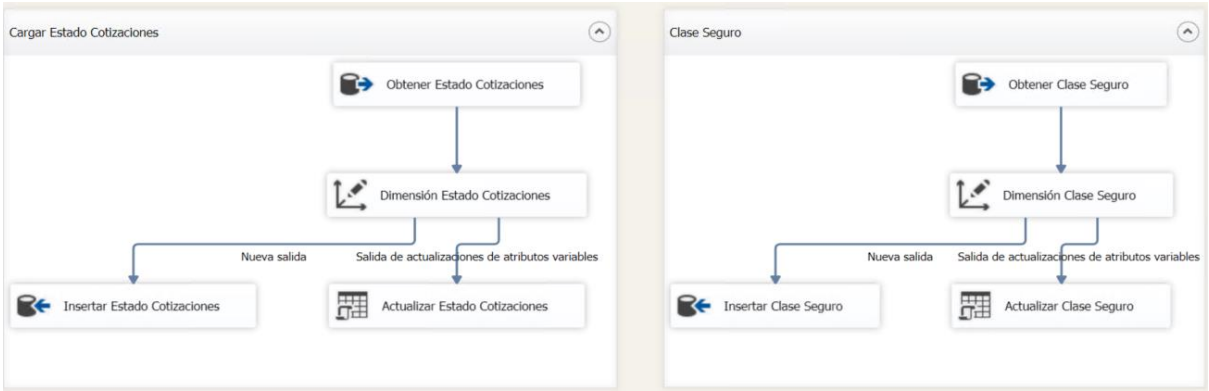


Figura 29. Vista parcial carga dimensión estado de cotizaciones y clase de seguro. Fuente: Propia.

Cuando la información de las dimensiones es muy grande como lo es el caso de la dimensión de las personas (6 millones) o los patronos (2 millones), se debe optar por utilizar otra alternativa de componente para actualizar o ingresar nuevos valores de esas dimensiones, ya que con el componente de dimensión de variación

lenta es sumamente “pesado” para procesar tanta cantidad de información. Es útil cuando la cantidad de datos es pequeña de al menos 10 mil registros hacia abajo.

Para las cargas de las personas y patronos se utilizaron varios componentes a fin de agilizar la carga de estos datos. Estos son: el ordenamiento, la combinación de mezcla, la división condicional, el destino de inserción OLE DB y el comando OLE DB para la actualización.

Visto gráficamente las cargas de estas dimensiones se miran de la siguiente forma:



Figura 30. Carga dimensión patronos Data Warehouse. Fuente: Propia.

En el caso de los patronos, se tuvo que someter a una transformación de los tipos de datos a fin de obtener un mejor rendimiento y optimización de espacio,

dado a que sus llaves principales eran decimales, se optó por cambiarlas a tipo char de longitud fija.

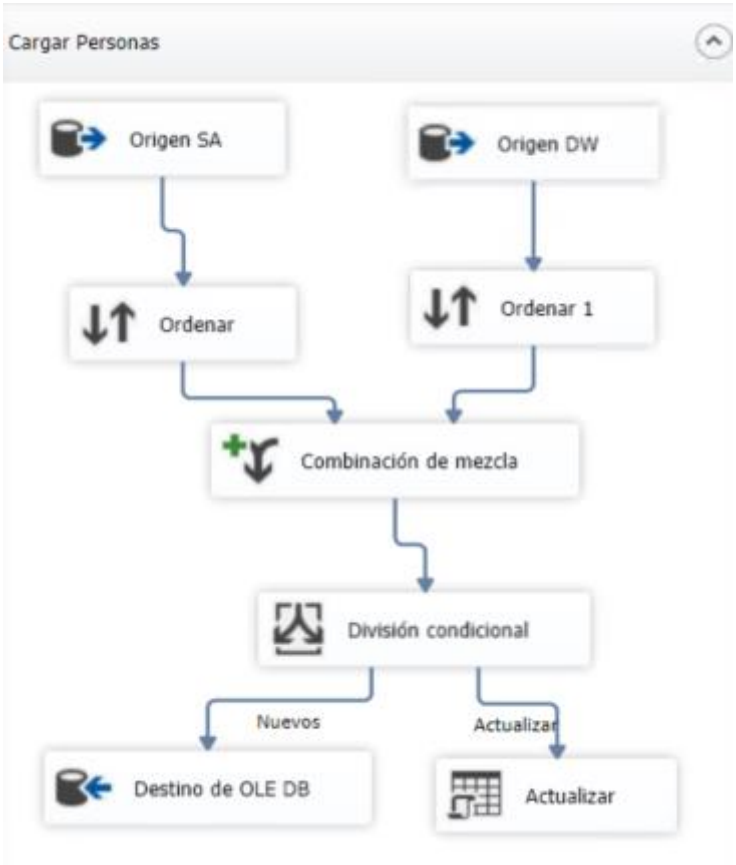


Figura 31. Carga dimensión personas Data Warehouse. Fuente: Propia.

Para lograr las cargas de los hechos al Data Warehouse, se utilizó la siguiente consulta:

```
select
co.num_cotizacion,
--Dimensiones
--Se maneja -1 para las dimensiones que no existen
[tipo_identificacionID]=isnull(DtiposID.tipo_identificacionID,-1),
[personaID]=isnull(Dpersona.personaID,-1),
[patronoID]=isnull(Dpatrono.patronoID,-1),
[clase_patronoID]=isnull(Dclasep.clase_patronoID,-1),
[provincia_patronoID]=isnull(Dprovincia.provinciaID,-1),
[canton_patronoID]=isnull(Dcanton.cantonID,-1),
[distrito_patronoID]=isnull(Ddistrito.distritosID,-1),
[sexoID]=isnull(Dsexo.sexoID,-1),
[tipo_estado_civilID]=ISNULL(DtipEstadoCivil.tipo_estado_civilID,-1),
[actividad_economicaID]=ISNULL(Dactividad_economica.actividad_economicaID,-1),
[estado_cotizacionID]=ISNULL(Destado_cotizacion.estado_cotizacionID,-1),
[clase_seguroID]=ISNULL(Dclase_seguro.clase_seguroID,-1),
[tipo_patronoID]=ISNULL(Dtippatrono.tipo_patronoID,-1),

--Indicadores
[ind_pensionado_RNC_LEY8769]=ISNULL(co.ind_pensionado_RNC_LEY8769,'N'),
[ind_pensionado_otro_regimen]=ISNULL(co.ind_pensionado_otro_regimen,'N'),
[ind_pensionado_IVM]=ISNULL(co.ind_pensionado_IVM,'N'),
[ind_cotizante_simultaneo]=ISNULL(co.ind_cotizante_simultaneo,'N'),
[ind_tipo_pension]=ISNULL(co.ind_tipo_pension,'No tiene'),
[can_cuotas_validas_menor18_mayor65]=ISNULL(co.can_cuotas_validas_menor18_mayor65,0),
[cuotas_IVM_antes_de_ser_TI]=ISNULL(co.cuotas_IVM_antes_de_ser_TI,0),
[cuotas_IVM_reales_siendo_TI]=ISNULL(co.cuotas_IVM_reales_siendo_TI,0),
[edad_ingresar_como_TI]=ISNULL(co.edad_ingresar_como_TI,0),
[salario_anterior]=ISNULL(co.salario_anterior,0),
[salario_tras_anterior]=ISNULL(co.salario_tras_anterior,0),

--Metricas
[can_salario_reportado]=isnull(co.can_salario_reportado,0),
[can_cuota_ivm]=isnull(co.can_cuota_ivm,0),
[can_cuota_sem]=ISNULL(co.can_cuota_sem,0),

--fechas
[fec_inicio_periodo_cotizacion]=CAST(co.fec_inicio_periodo_cotizacion AS INT),
[fec_fin_periodo_cotizacion]=CAST(isnull(co.fec_fin_periodo_cotizacion,0) AS INT),
[fec_ingreso_cotizante_TI]=CAST(isnull(co.fec_ingreso_cotizante_TI,0) AS INT)

from
PensionesSA.StagingArea.pensiones_cotizaciones co
left join PensionesDW.DataWarehouse.dim_personas Dpersona
on co.num_identificacion=Dpersona.ref_num_identificacion
and co.tip_identificacion=Dpersona.ref_tip_identificacion
left join pensionesDW.DataWarehouse.dim_tipos_identificacion DtiposID
on co.tip_identificacion =DtiposID.ref_tip_identificacionID
left join pensionesDW.DataWarehouse.dim_clase_patronos Dclasep on
co.clase_patrono=Dclasep.ref_cls_patrono
left join PensionesDW.DataWarehouse.dim_provincias Dprovincia on
co.cod_provincia_patrono =Dprovincia.ref_cod_provincia
left join PensionesDW.DataWarehouse.dim_cantones Dcanton on
co.cod_canton_patrono=Dcanton.ref_cod_canton and
co.cod_provincia_patrono=Dcanton.ref_cod_provincia
left join PensionesDW.DataWarehouse.dim_distritos Ddistrito on
co.cod_provincia_patrono=Ddistrito.ref_cod_provincia and
co.cod_canton_patrono=Ddistrito.ref_cod_canton and
co.cod_distrito_patrono=Ddistrito.ref_cod_distrito
left join PensionesDW.DataWarehouse.dim_tip_patronos Dtippatrono on
co.tip_patrono=Dtippatrono.ref_tip_patrono
left join PensionesDW.DataWarehouse.dim_clase_seguros Dclase_seguro on
co.tip_cuota=Dclase_seguro.ref_cod_seguro
left join PensionesDW.DataWarehouse.dim_sexos Dsexo on
co.cod_sexo=Dsexo.ref_cod_sexo
left join PensionesDW.DataWarehouse.dim_estado_cotizaciones Destado_cotizacion on
co.cod_estado=Destado_cotizacion.ref_cod_estado
left join PensionesDW.DataWarehouse.dim_tipos_estado_civil DtipEstadoCivil on
co.cod_estado_civil=DtipEstadoCivil.ref_cod_estado_civil
left join PensionesDW.DataWarehouse.dim_actividades_economicas Dactividad_economica on
co.cod_actividad_economica=Dactividad_economica.ref_cod_actividad_economica
left join PensionesDW.DataWarehouse.dim_patronos Dpatrono on
co.tip_patrono=Dpatrono.ref_tip_patrono and co.num_patrono=Dpatrono.ref_num_patrono
and co.num_segregado=Dpatrono.ref_num_segregado and co.num_sector=Dpatrono.ref_num_sector
```

Figura 32. Consulta SQL carga hechos Data Warehouse. Fuente: Propia.

Como se puede apreciar en la imagen anterior, se aplica una organización de las variables basado en su contexto, que son: dimensiones, indicadores, métricas y fechas. Esto con el fin de obtener un código más limpio y ordenado. Adicionalmente se maneja el tema de los valores nulos, de acuerdo con los datos presentes en las dimensiones. En este caso se toma como “-1” valores no existentes para las dimensiones, 0 para los valores que son enteros pero que fungen como indicadores de un cierto grupo de información, se determina que se conserve el valor por defecto cuando no exista “1900-01-01” para las fechas de “fec_fin_periodo_cotizacion”, “fec_ingreso_cotizante_TI”, y finalmente se cargan con el valor “No tiene” los campos “ind_tipo_pension” e “ind_lugar_pago_pension_reciente”.

Para cada uno de los valores nulos como ya se indicó, se estableció un valor por defecto, a fin de no dejar estos valores en la representación de los hechos del Data Warehouse, por lo que se incluyen de forma manual estos valores. Estos

valores, al ser identidad dentro de las dimensiones, debe deshabilitarse esta condición e insertar el valor para luego volver a habilitar el valor de identidad. Ejemplo de esto es el ingreso del valor desconocido para el estado de las cotizaciones:

```
--estado cotizacion
SET IDENTITY_INSERT PensionesDW.[DataWarehouse].dim_estado_cotizaciones ON

INSERT PensionesDW.[DataWarehouse].dim_estado_cotizaciones
(estado_cotizacionID,ref_cod_estado,ind_cuota_valida_pension,dsc_estado)
VALUES (-1,'-1','NO','Estado desconocido')

SET IDENTITY_INSERT PensionesDW.[DataWarehouse].dim_estado_cotizaciones OFF
```

Figura 33. Inserción de valores desconocidos en dimensión estado de cotizaciones. Fuente: Propia.

Para cada una de las dimensiones se constató que este valor no existiera efectivamente y no hubiera confusión con la información que se obtiene.

Es importante mencionar que las siguientes dimensiones se cargaron de forma manual al no existir en su origen dentro del sistema transaccional:

- dim_actividades_económicas
- dim_clase_seguros
- dim_estado_cotizaciones
- dim_sexos
- tip_patronos
- tipo_trabajadores
- tipos_estado_civil

Una vez realizadas las configuraciones dentro del proyecto de Integration Services, se procede con las cargas:



Figura 34. Procesamiento de carga en tabla de hechos, de enero 2019 a julio 2019.

Fuente: Propia.

Proyecto Analisys Services

El proyecto de Analisys Services, se llevó a cabo en un servidor interno de la organización facilitado para esta tarea, de modo que, una vez procesadas las cargas sobre los diseños creados, se procede a crear el proyecto y definir los ajustes necesarios como nombres identificativos, relaciones del modelo y procesamiento del cubo analítico para que este sea consumido por la herramienta Power BI.

La vista del diagrama desde el modelo creado del proyecto de Analisys Services se muestra a continuación:

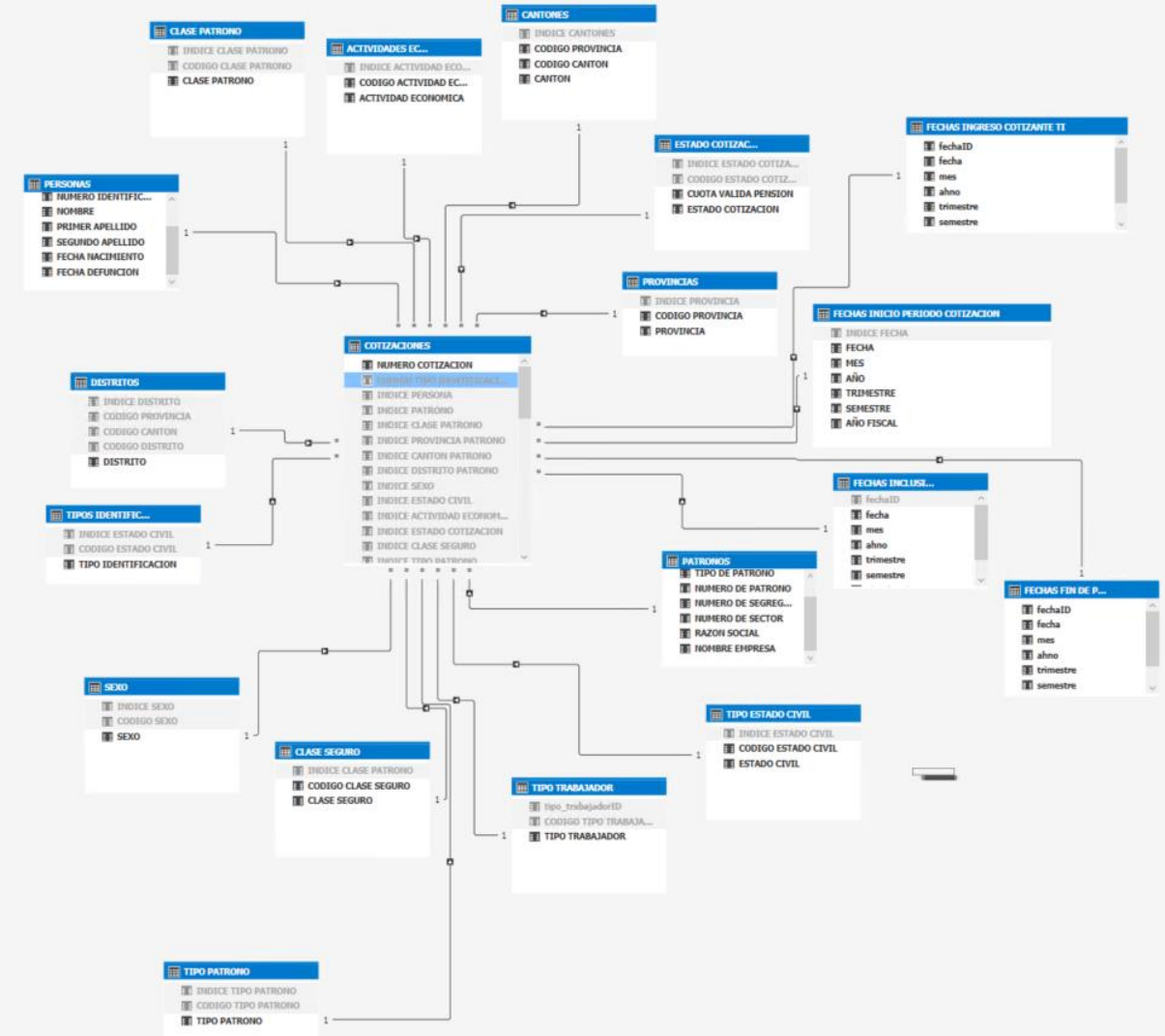


Figura 35. Vista diagrama de modelo en Analisys Services. Fuente: Propia.

La imagen anterior muestra los hechos en conjunto con las dimensiones relacionadas. Se puede denotar la creación de las dimensiones de fechas para cada una de las establecidas en el nivel de los hechos o en este caso las cotizaciones.

Una vez que se validó el modelo en su funcionamiento y ajustes se procedió a ejecutar el procesamiento de todas las tablas involucradas. Con un total de 19 objetos por procesar.

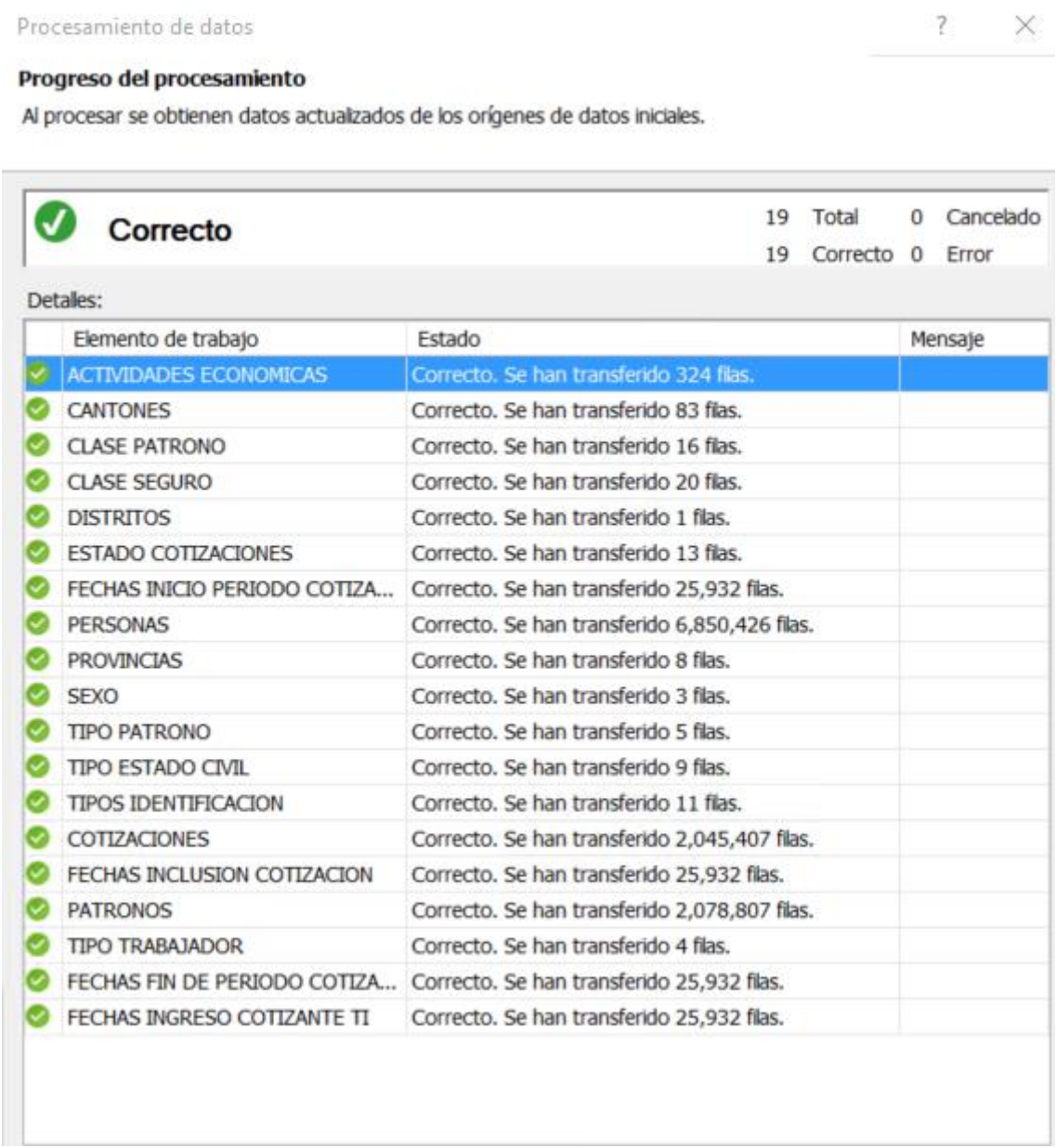


Figura 36. Procesamiento de datos en Analisys Services. Fuente: Propia.

Como resultado del procesamiento se obtuvieron 19 correctos, 0 cancelados y 0 con errores. Adicionalmente es importante mencionar que se integraron algunas columnas calculadas que se realizaron a solicitud de los funcionarios de Cuenta Individual y Control de Pagos, como el rango salarial. Para calcular esta columna se utilizó la función DAX permitida por el motor de Analisys Services para calcular datos: =IF([SALARIO]<=500000 ;"Menor a 500 mil";IF([SALARIO]<=1000000;"500 mil a 1 millón";IF([SALARIO]<=2000000;"De 1 a 2 millones";IF([SALARIO]<=3000000;"De 2 a 3 millones";IF([SALARIO]<=5000000;"De 3 a 5 millones";IF([SALARIO]<=10000000;"De 5 a 10

millones";IF([SALARIO]<=15000000;"De 10 a 15 millones";"Mayor a 15 millones"))))))). Asimismo se realizó la columna calculada para obtener la edad de las personas, mediante la siguiente formula DAX: =INT (Year(today())- Year([FECHA NACIMIENTO])).

El cálculo de esta columna arrojó datos interesantes sobre las edades de las personas que son almacenadas en el sistema transaccional, por ejemplo, indica que el promedio de la edad es de 62 años, asimismo existen edades cuyo valor máximo está en 2019 años y valores mínimos de -2581 años. Esto refleja un problema en la calidad de datos en el nivel de las fechas de nacimiento de las personas.

En el desarrollo de las métricas se construyeron las siguientes:

- Suma de salario tras anterior.
- Suma de salario anterior.
- Promedio de edad al ingresar como trabajador independiente.
- Suma y promedio de cuotas IVM siendo trabajador independiente.
- Suma y promedio de cuotas IVM antes de ser trabajador independiente.
- Suma y promedio de salario.
- Suma y promedio de cuota IVM.
- Suma y promedio de cuota SEM.
- Recuento del número de registros por número de cotización reportada.
- Recuento único de personas.

Una vez que se ha procesado el cubo analítico, implementa en ambiente pruebas y como resultado se construyen las estructuras definidas en el proyecto:

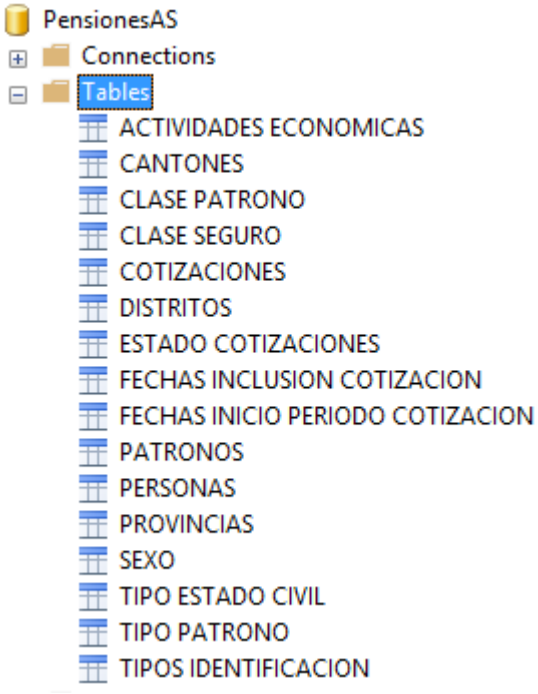


Figura 37. Vista el cubo analítico OLAP. Fuente: Propia.

Proyecto Power BI

Esta etapa conlleva la creación de los Dashboard según determinaron los usuarios a efectos de realizar los análisis de información.

El diseño acordado con el usuario divide el panel del Dashboard en cuatro vistas, las cuales se visualizan de la siguiente forma:

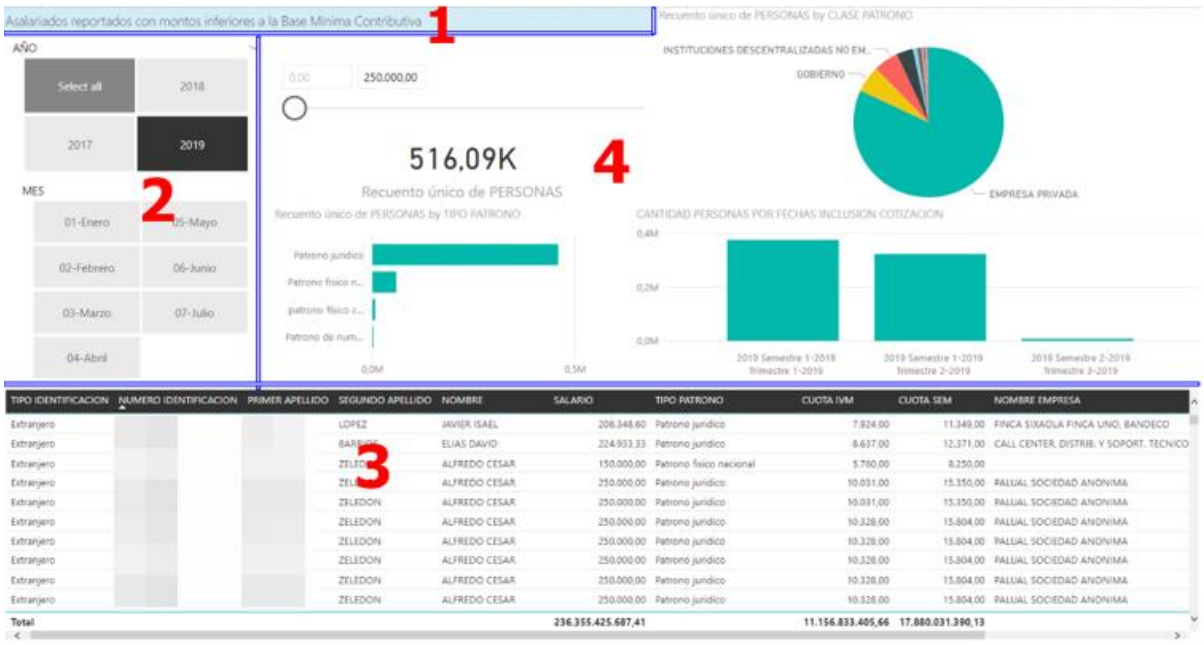


Figura 38. Diseño de Dashboard Power BI. Fuente: Propia.

- Vista 1: Contiene el nombre del indicador.
- Vista 2: Filtros del indicador.
- Vista 3: Detalles del indicador.
- Vista 4: Gráficos de interés para el indicado. Está vista usualmente tendrá recuento total de las personas identificadas mediante los filtros y un gráfico de barras vertical que proporcionará una visión segmentada en una jerarquía de fechas para poder visualizar las cantidades de los registros en esas temporalidades (Año, semestre, trimestre, mes, día).

Como parte de la estrategia para definir los reportes y al buscar flexibilización del usuario contra la herramienta de Power BI para hacer de esta más fácil en su adaptación, los filtros principales para cada uno de los indicadores se colocan en el nivel de página y algunos se colocan en el nivel visual como lo son montos o cantidades. Cada uno de los Dashboard conserva un diseño estándar como se vio en la figura 38, a fin de preservar

una estética consistente y no confundir a las personas que consultan la información.

A continuación, se detallan algunas imágenes de los Dashboard construidos y su contexto.

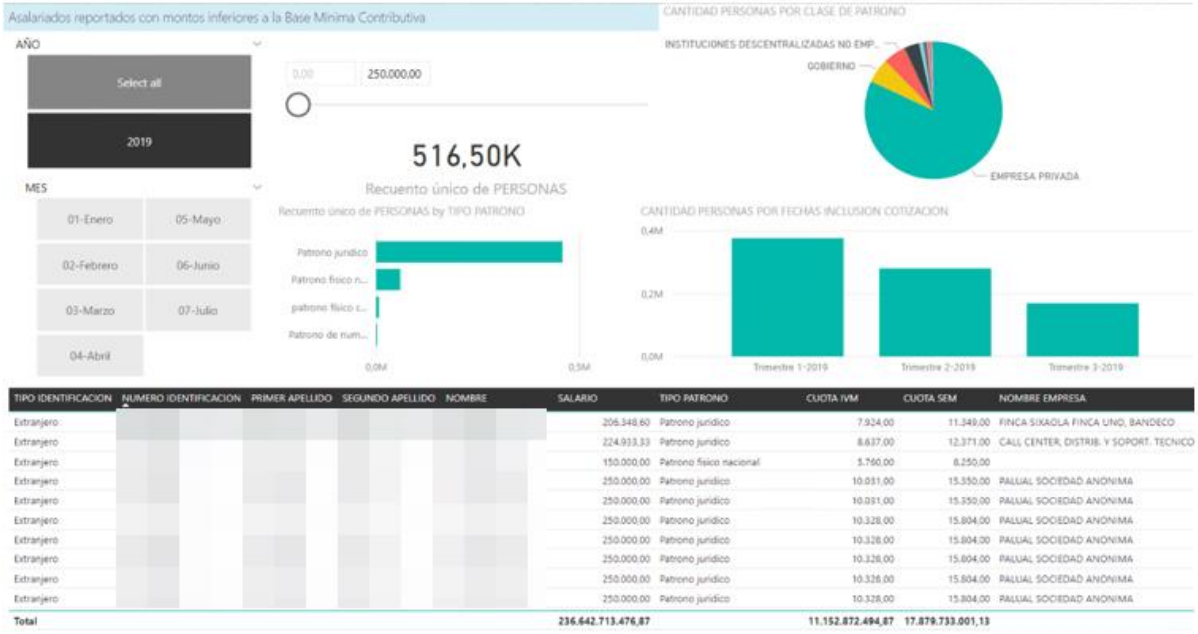


Figura 39. Dashboard de Asalariados reportados con montos inferiores a la Base Mínima Contributiva. Fuente: Propia.

La imagen anterior muestra el indicador de los “Asalariados reportados con montos inferiores a la Base Mínima Contributiva”, como se puede notar con los datos cargados desde enero a julio del 2019, existen más de 516 mil personas cuyos patronos han reportado menos que la base mínima contributiva, en este caso el filtro visual del monto mínimo se estableció en ¢250.000. Este filtro visual se estableció con el objeto Slicer. Adicionalmente se estableció un gráfico para visualizar la cantidad de personas bajo está condición distribuidas por la clase de patrono. Asimismo, se estableció el gráfico de barras vertical que permite visualizar las cantidades de personas en esta condición en las jerarquías de fechas disponibles. Para este caso se puede observar la visualización de forma

trimestral, con los datos procesados hasta esas fechas, es posible notar una reducción mínima desde el primer trimestre hasta el tercer trimestre del 2019 en la cantidad de personas cuyo salario está por debajo la base mínima contributiva, aunque para completar el tercer trimestre hacen falta cargar dos meses.

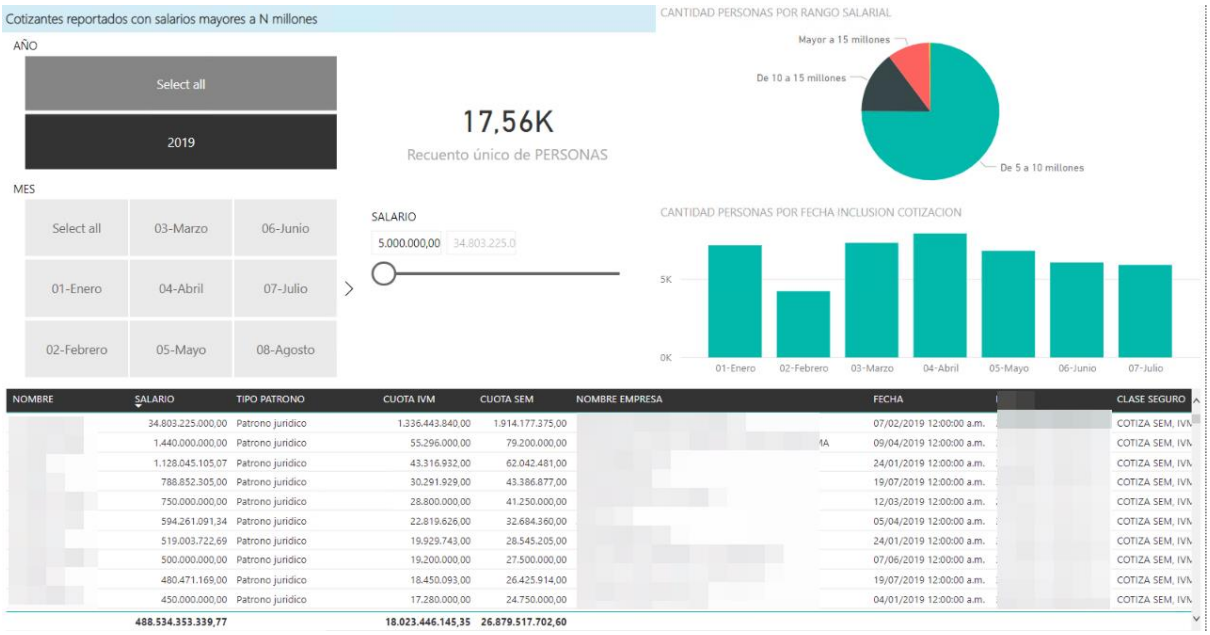


Figura 40. Dashboard Cotizantes reportados con salarios mayores a N millones. Fuente: Propia.

La imagen anterior, brinda el panorama para los salarios cuyo monto es superior a ¢5.000.000 para los salarios reportados de enero a julio de 2019. Con esta información se procede a crear un gráfico que muestra la distribución en los rangos de salarios en colones:

- Menores a 500 mil.
- De 500 mil a 1 millón.
- De 1 a 2 millones.
- De 2 a 3 millones.
- De 3 millones a 5 millones.

- De 5 a 10 millones.
- De 10 a 15 millones.
- Mayores a 15 millones.

Asimismo, se visualiza mediante el gráfico de barras vertical la información de las personas reportadas bajo esta condición en esos meses y ver comportamientos. Se denota que febrero es el mes en el que existieron menos reportes de cotizantes con salarios mayores a 5 millones. Es importante mencionar que parte del objetivo del indicador es verificar la veracidad de los montos reportados por el patrono conforme el historial salarial de los afiliados. Por lo que cualquier monto bajo estas clasificaciones podrá ser capturado fácilmente por los encargados de CI, también es posible que obtengan a la persona con el mayor monto reportado a partir de la ordenación en el detalle de la columna salario. En este caso es interesante observar que para estos meses existe un reporte de un salario de más de ¢34.803.225.000,00 que en letras se lee treinta y cuatro mil ochocientos tres millones doscientos veinticinco mil.

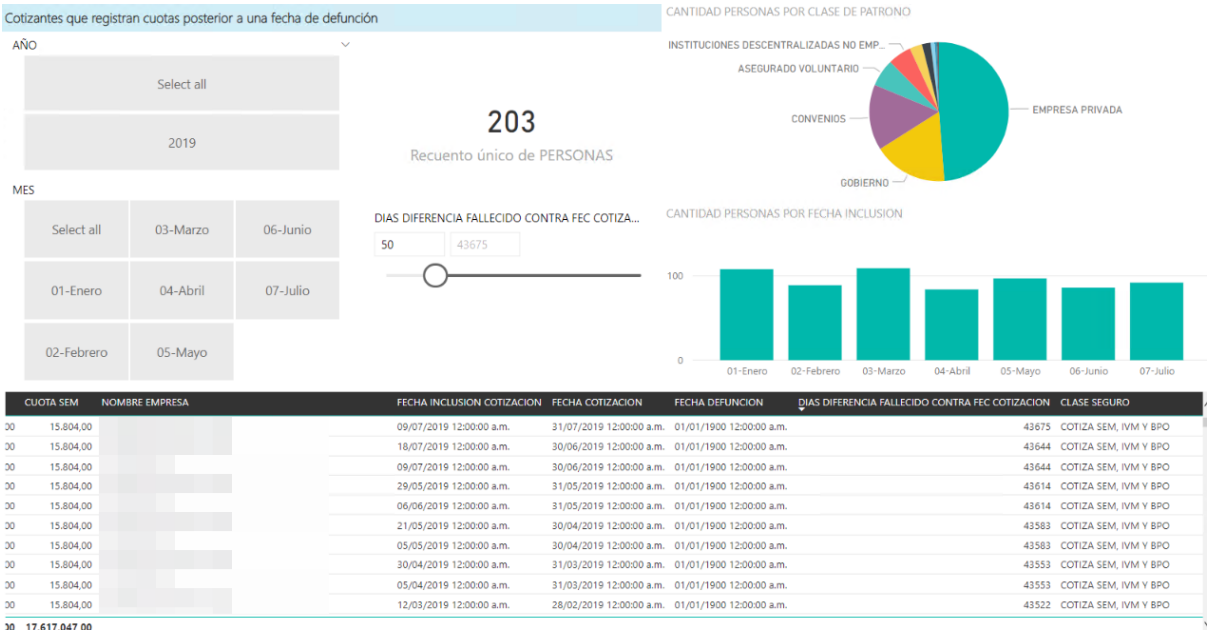
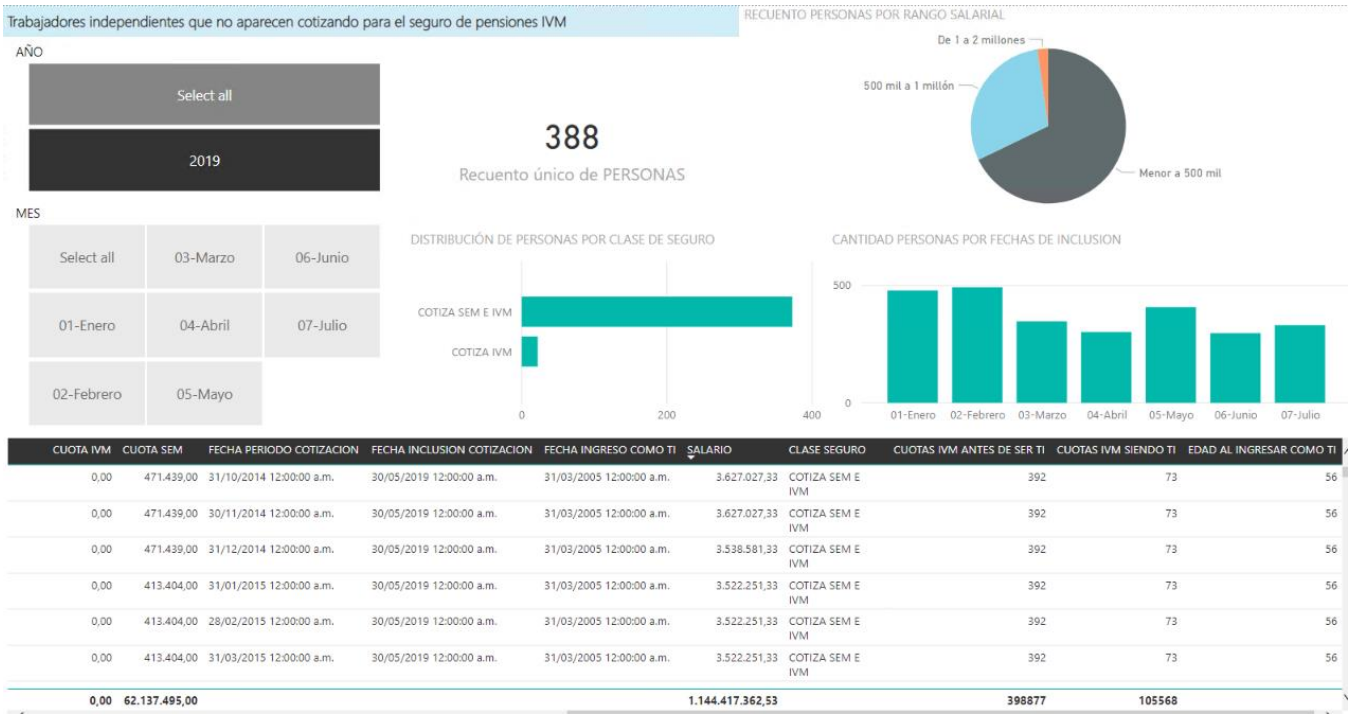


Figura 41. Dashboard Cotizantes que registran cuotas posteriores a una fecha de defunción. Fuente: Propia.

El Dashboard anterior nos muestra personas bajo la condición de reportes de cotizaciones posteriores a su fecha de defunción para los meses cargados de este 2019.

La visualización también contiene un Slicer, el cual proporciona un filtro adicional con respecto a los días de diferencia desde la fecha de defunción de la persona hasta la fecha que del periodo que reportan su cotización. En este caso se filtra a aquellas personas cuyos días sean mayores a 50 días después de su fecha de defunción. Se puede notar en el gráfico pastel, la distribución por la clase de patrono que reporta, el cual muestra que la mayoría de los casos vienen dados por la empresa privada, además el gráfico de barras vertical indica que por mes existe una tendencia en promedio mensual en esta condición de 100 personas. En total para el año 2019, existen 203 personas cuya fecha de período de cotización es mayor a su fecha de defunción.

Es importante mencionar que este indicador en específico viene a fungir como un control y seguimiento de la información de la base de datos de la CI del seguro de pensiones de IVM de acuerdo con el producto esperado por el Área de Cuenta Individual y Control de Pagos. Esto con el fin de identificar a los cotizantes bajo esta condición a fin de evitar el otorgamiento de beneficios improcedentes o que los usuarios del sistema SIP proporcionen información incorrecta de las cotizaciones aportadas al Seguro de Pensiones IVM.



CUOTA IVM	CUOTA SEM	FECHA PERIODO COTIZACION	FECHA INCLUSION COTIZACION	FECHA INGRESO COMO TI	SALARIO	CLASE SEGURO	CUOTAS IVM ANTES DE SER TI	CUOTAS IVM SIENDO TI	EDAD AL INGRESAR COMO TI
0,00	471.439,00	31/10/2014 12:00:00 a.m.	30/05/2019 12:00:00 a.m.	31/03/2005 12:00:00 a.m.	3.627.027,33	COTIZA SEM E IVM	392	73	56
0,00	471.439,00	30/11/2014 12:00:00 a.m.	30/05/2019 12:00:00 a.m.	31/03/2005 12:00:00 a.m.	3.627.027,33	COTIZA SEM E IVM	392	73	56
0,00	471.439,00	31/12/2014 12:00:00 a.m.	30/05/2019 12:00:00 a.m.	31/03/2005 12:00:00 a.m.	3.538.581,33	COTIZA SEM E IVM	392	73	56
0,00	413.404,00	31/01/2015 12:00:00 a.m.	30/05/2019 12:00:00 a.m.	31/03/2005 12:00:00 a.m.	3.522.251,33	COTIZA SEM E IVM	392	73	56
0,00	413.404,00	28/02/2015 12:00:00 a.m.	30/05/2019 12:00:00 a.m.	31/03/2005 12:00:00 a.m.	3.522.251,33	COTIZA SEM E IVM	392	73	56
0,00	413.404,00	31/03/2015 12:00:00 a.m.	30/05/2019 12:00:00 a.m.	31/03/2005 12:00:00 a.m.	3.522.251,33	COTIZA SEM E IVM	392	73	56
0,00	**62.137.495,00**				**1.144.417.362,53**		**398877**	**105568**	

Figura 42. Dashboard Trabajadores independientes que no aparecen cotizando para el seguro de pensiones IVM. Fuente: Propia.

La visualización anterior, permite obtener a los TI que no están cotizando para el seguro de pensiones IVM, nótese que no existe mayor filtro visual o un Slicer que permite seleccionar está información. Esto es debido a que el filtro se aplica en el nivel de página con las restricciones establecidas por los encargados de la CI.

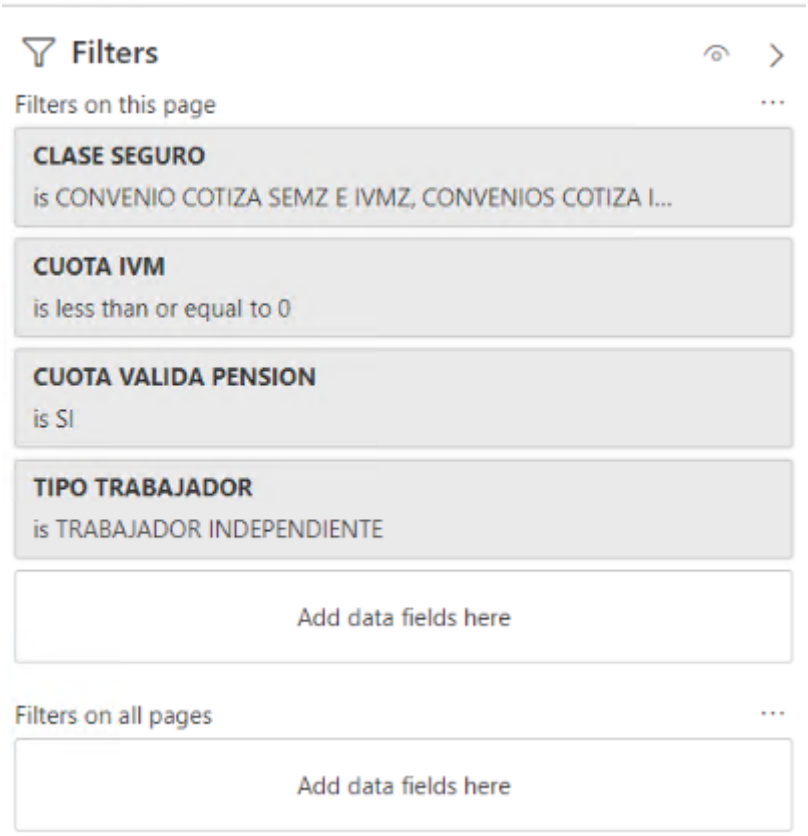


Figura 43. Filtros en el nivel de página para el indicador de los TI que no cotizan IVM. Fuente: Propia.

Dentro de lo que se destaca en este Dashboard, se puede observar que existe un total de 388 personas trabajadoras independientes para los meses cargados del 2019. El gráfico de barras horizontal permite observar en cuales tipos de seguro es en el que se presentan las incidencias. Adicionalmente se tiene un gráfico pastel con el recuento de las personas TI por su rango salarial, en este se puede ver que bajo está condición existen personas con rangos de salario desde los 0 a 500 mil, de 500 mil a 1 millón y de 1 millón a 2 millones que no están reportando cuotas al Seguro de Pensiones IVM para las clases de seguro “Cotizaciones SEM E IVM” y “Cotiza IVM”. Según el objetivo de este indicador el Área de Cuenta Individual y Control de Pagos desea verificar el cumplimiento del Artículo 2° del reglamento del Trabajador Independiente, para proporcionar un seguimiento y

control con el propósito de informar a las autoridades competentes a efectos que actúen conforme a la normativa vigente.

Seguidamente se expone un resumen de las ventajas cualitativas del modelo de inteligencia de negocios propuesto y la clasificación de la excelencia obtenida basado en los puntajes de referencia acordados con el negocio.

Ventajas cualitativas del modelo de inteligencia de negocios.

Analítica actual	**Modelo inteligencia de negocios propuesto**
Reportes estáticos.	Flexibilidad en la creación de gráficos.
Velocidad lenta de consulta	La información se obtiene en segundos.
Manejo sumamente complejo para obtener estadísticas en diferentes temporalidades	Las temporalidades se obtienen simplemente seleccionando el filtro adecuado.
El valor se limita a los datos que se obtienen.	Valor añadido mediante la aplicación de técnicas estadísticas de análisis y modelado para encontrar relaciones ocultas.
Exposición de riesgos en la exportación de datos de una fuente a otra.	Única fuente de datos
Se limita a la cantidad permitida por los XLS.	No hay límite de espacio
La comprensión está sujeta a los nombres de las columnas.	Fácil de comprender
Datos no estandarizados.	Datos estandarizados

Tabla 15. Ventajas del modelo propuesto vs actual. Fuente: Propia.

En el capítulo 4 Análisis del Diagnóstico, se determinaron mediante las herramientas de recolección de datos la información sobre las cantidades de tareas y tiempos requeridos para desarrollar los informes de los indicadores actualmente.

El modelo de inteligencia de negocios por su parte expone otra situación y con este se pueden determinar estos datos a fin de valorar y poder determinar el impacto de la solución versus la actual. Mediante la utilización de las variables de impacto utilizadas en el apartado del enfoque 3.3 del capítulo 2 del Marco Metodológico se realiza la siguiente tabla que muestra el análisis de impacto y el contexto de los valores obtenidos.

Variable	Valor logrado	Explicación
Cantidad de tareas requeridas para desarrollar informes analíticos.	2 tareas	El usuario mediante un acceso directo abre un vínculo hacia el sitio de Power BI donde estarán los indicadores, una vez ahí ubica el informe solo lo selecciona y obtiene la información. 1.Abrir vinculo hacia Power BI. 2. Seleccionar informe de interés.
Tiempo necesario para desarrollar informes analíticos.	60 segundos	El tiempo aproximado en promedio que el usuario podría durar para generar alguno de los indicadores desde el sitio Web de Power BI interno de la organización.
Costo económico requerido para desarrollar informes analíticos (humano y tecnológico).	¢70 mensuales. ¢840 anuales.	Se suma el costo únicamente por minuto del profesional del funcionario de CI, dado a que sería este quien accedería al repositorio de datos para obtener la información. Tabla 10. Aproximadamente el costo representa: ¢70. Se excluye el costo anual de la herramienta Power BI en office 365 que

		posee la institución por cuanto se desconoce este valor, así como otros costos por licencias en la utilización del servicio.

Tabla 16. Análisis de variables impacto modelo propuesto de inteligencia de negocios. Fuente: Propia.

Con la tabla de anterior es posible obtener los porcentajes que van a determinar el nivel de aceptación definido en el apartado de enfoque. Seguidamente se exponen los resultados obtenidos para modelo de inteligencia de negocios propuesto al Área de Cuenta Individual y Control de Pagos.

Variable	Valor porcentual obtenido	Valor absoluto logrado
Cantidad de tareas requeridas para desarrollar informes analíticos.	25%	Valor actual: 7. Valor modelo:2 Reducción=5
Tiempo necesario para desarrollar informes analíticos.	25%	Valor actual: 15 horas. (54.000 segundos) Valor modelo: 60 segundos. Reducción en segundos=53.940
Costo económico requerido para desarrollar informes analíticos (humano y tecnológico).	25%	Valor actual anual: ¢525.438 Valor modelo: ¢840 Reducción: ¢524.598
Impacto de los informes en la toma de decisiones, con las variables actuales versus con variables adicionales y segmentadas a través del tiempo.	20%	Basado en la Tabla 14. Ventajas del modelo propuesto vs actual, así como valorando el impacto de las variables anteriores, el jefe de CI determina

		asignar un 20%, reservando a forma discrecional la evolución del modelo.
Total	95%	
Clasificación del modelo	Excelente	

Tabla 17. Clasificación del modelo según impacto de variables. Fuente: Propia.

Como se puede observar el modelo cumple con la clasificación de Excelente, de acuerdo con la clasificación desarrollada en el presente trabajo. Cada una de las variables analizadas y su valor fue validada con los encargados del Área de Cuenta Individual y Control de Pagos. Por lo que el modelo es muy recomendado en cada una de las variables de impacto según se pudo determinar, por tanto, su implementación será muy beneficiosa en el corto, mediano y largo plazo.

Capítulo 6. Conclusiones y Recomendaciones

6.1 Conclusiones

- La comprensión de los datos en el Área de Cuenta Individual y Control de Pagos abarcó desde como funcionalmente la información es procesada e interpretada hasta el significado de cada elemento presente en sus datos. Esto debido a que el origen de la información proviene de SICERE y se cruza contra el Sistema de Pensiones.
- La identificación de los indicadores fue un proceso sencillo. La información que requieren está bastante clara, dado que a estos corresponden informes o controles para la toma de decisiones.
- La creación del Data Warehouse, se concluyó de forma satisfactoria en ambiente pruebas. Conllevó un proceso evolutivo que considera las buenas prácticas y la experiencia obtenida en cada una de sus etapas.
- El proceso de ETL es una de las etapas críticas y las que conllevaron más cantidad de tiempo, debido a la gran cantidad de información que debía procesarse, así como de sus características.
- La construcción de los principales Dashboard dentro de la herramienta de Power BI, es de los componentes que más impacto brindó a los usuarios, debido al potencial que tienen para manipular la información y adicionalmente la velocidad y facilidad que tienen para procesar está información en la forma que ellos deseen y necesiten.
- El desarrollo de la propuesta de inteligencia de negocios permitió descubrir variables que, a simple vista son difíciles de identificar, de modo que se crearán oportunidades de análisis para el tratamiento de los datos que proporcionarán conocimiento para lograr estrategias de negocio más efectivas.
- Las etapas de la comprensión de datos y la obtención de información mediante el proyecto de Analisys Services permitió determinar la existencia de un problema en la calidad de datos tanto en las estructuras como en la composición de la información.
- El Área de Cuenta Individual y Control de Pagos deberá crear estrategias preventivas de negocio para los indicadores cuyo objetivo sea depuración para lo cual puede definir metas y dar seguimiento con el proyecto propuesto en los plazos que definan.

- La propuesta de inteligencia de negocios permitirá dar un seguimiento y medir el impacto que se estaría logrando con la aplicación de estas. Adicionalmente fungirá como una herramienta de control para monitorear lo que sucede con sus actividades sustantivas.
- El proyecto permitirá al Área de Cuenta Individual y Control de Pagos optar por transición de una administración reactiva a una administración reactiva.
- La propuesta del modelo de inteligencia de negocios, incentivo el apoyo e influencia en el paradigma de la analítica empresarial tanto en el Área de Cuenta Individual como en el nivel de Gerencia de Pensiones.

6.2 Recomendaciones

- Se recomienda integrar algún servicio como Master Data Services que permita el ingreso de información nueva o actualizar la existente para las dimensiones que no tienen origen transaccional, a fin de que se evite la ejecución de script de forma manual sobre el Data Warehouse.
- Es necesario actualizar la carga de datos de los patronos para la información mensual que proviene de SICERE, a fin de que se carguen mediante campos separados la información adicional de los siguientes campos, dado a que no existen:
    - Cantón, distrito, provincia.
    - Actividades económicas.
- La tabla de cotizaciones debe modificarse a fin de que se pueda capturar y diferenciar el tipo de planilla que se reporta, dado a que actualmente se carga integrado en otro campo a manera descriptiva.

- De acuerdo con las reuniones sostenidas con las autoridades respectivas de la Dirección de Tecnologías de Información, la puesta en producción del presente modelo debe construirse dentro del Data Warehouse institucional de modo que la estructura creada en pruebas para este modelo se debe validar contra lo que ya existe institucionalmente a efectos de conformar la estrella en ese ambiente, así como procesos de carga.
- Si el modelo es construido en el nivel central dentro del Data Warehouse, se ahorrarían costos de mantenimiento, espacios de servidor y recurso humano local para atender incidencias que pudieran afectar el funcionamiento del repositorio.
- Se recomienda realizar un proceso de calidad de datos en general sobre las estructuras de las tablas e información que manejan a los patronos, cotizaciones y personas.
- En el desarrollo de la presente investigación hubo cierta resistencia o poca colaboración para obtener información relacionada con los datos de SICERE. El proceso hubiese abarcado menos complejidad, por tanto, los tiempos de desarrollo se hubieran reducido y como resultado los costos se disminuyen.

Capítulo 7. Reflexiones Finales

El desarrollo del presente trabajo ha permitido desarrollar habilidades en el reconocimiento de patrones de diseño y establecer una visión más clara de las posibles soluciones técnicas que se pueden incluir en la construcción de modelos de repositorio de datos y cubos analíticos. Asimismo, el conocer los alcances y limitaciones de las herramientas que actualmente se utilizan para realizar cada una de las etapas marca un punto de partida referencial en cuanto a los tiempos de desarrollo.

La importancia de conocer el cómo, dónde y forma del origen de la información son puntos vitales en el desarrollo de este tipo de tecnologías, cada origen de datos tiene sus propias características y estas deben ser analizadas de tal forma que se comprenda como se lleva a cabo su procesamiento hasta el destino del Data Warehouse.

La inclusión de esta tecnología y el modo de obtener información conlleva un cambio en la conformación del recurso humano de tecnologías de información de la Gerencia de Pensiones, dado a que este debe ser dedicado para dar mantenimiento a las soluciones existentes y brindar mejora continua a la capacidad de esta tecnología, con el fin de producir resultados más eficientes y notorios en el mediano y largo plazo.

Capítulo 8. Trabajos a Futuro

La propuesta de inteligencia de negocios viene a ser una base fundamental para lograr culturizar este tipo de tecnologías tanto en el Área de Cuenta Individual y Control de Pagos como en la misma Gerencia de Pensiones. En futuras ramificaciones a la solución planteada, se podría pensar en incluir:

- Llaves para medir el rendimiento o KPI.
- Técnicas de Minería de datos descriptivas.
- Técnicas de Minería de datos predictivas.

A fin de disponer de una solución más completa que apoye las actividades sustantivas en la estrategia de negocio que desempeñan.

Glosario

KPI: Key Performance Indicator.

Staging Area: es un área intermedia de almacenamiento de datos utilizada para el procesamiento de estos durante procesos de extracción, transformación y carga.

SIP: Sistema Integrado de Pensiones.

CI: Cuenta Individual.

TI: Trabajador Independiente.

OLAP: OnLine Analytical Processing o procesamiento Analítico en Línea.

BI: Business Intelligence.

DW o DWH: Data Warehouse.

SICERE: Sistema Centralizado de Recaudación.

SSIS: SQL Server Integration Services.

SSAS: SQL Server Analisys Services.

DIM: Dimensión.

HE: Hechos.

MDM: Máster Data Management o método integral que permite a una empresa unificar todos sus datos críticos en un solo repositorio.

Data Integration: Integración de datos. Unificar los datos de modo que sean consistentes entre sí.

Data Governance: Gobierno de datos que establece un conjunto de políticas y reglas para gestionar los datos.

Mobile BI: Visualización de datos analíticos a través de aplicaciones móviles.

Data source: Origen de datos.

OLTP: OnLine Transaction Processing.

Storage: Refiere a almacenamiento en bases de datos.

Metadata: Conjunto de datos que describen el contenido informativo de un recurso, de archivos o de información de estos.

Minería de datos: Proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos.

Tableau: Software para visualizar datos de diferentes fuentes y realizar análisis una opción alternativa a Power BI.

TagCrowd: Software para generar una WordCloud.

WordCloud: Nube de palabras.

Referencias

Biolchini, J. (2005). *Plantilla de Bolchini para Revisión Sistemática.* Rio de Janeiro: Plantilla de Bolchini para Revisión Sistemática.

computerweekly. (05 de 06 de 2019). *computerweekly*. Obtenido de computerweekly: https://www.computerweekly.com/tip/Inmon-or-Kimball-Which-approach-is-suitable-for-your-data-warehouse

Díaz, J. C. (2012). *Introducción al Business Intelligence.* Barcelona, España: Editorial UOC. Recuperado el 05 de 05 de 2019

Kitchenham, B. (2004). *Procedures for Performing.* Australia: Keele University Technical Report TR/SE-0401.

Microsoft. (05 de 06 de 2019). *powerbi.microsoft*. Obtenido de powerbi.microsoft: https://powerbi.microsoft.com/es-es/what-is-power-bi/

Power Data. (05 de 06 de 2019). *blog.powerdata.es*. Obtenido de blog.powerdata.es: https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/data-profiling-el-primer-paso-en-calidad-de-datos

PRABHU, C. (2008). *DATA WAREHOUSING: Concepts, Techniques, Products and Applications* (3 ed.). New Delphi: PHI Learning Pvt. Ltd., 2008.

ALVI, I. A. (02 de 08 de 2019). *datawarehouseinfo.com*. Obtenido de datawarehouseinfo: https://datawarehouseinfo.com/dimensional-data-warehouse-modeling/

Tableau software. (07 de 08 de 2019*). www.tableau.com*. Obtenido de https://www.tableau.com/es-es/learn/articles/business-intelligence

Bibliografía

Área	Sitio Web o Libro
Business Intelligence	Río, L. M. (2006). *Más allá del Business Intelligence.* Barcelona: Ediciones Gestión 2000.
	Kimball, R., & Ross, M. (2011). *The Data Warehouse Toolkit Second Edition.* Estados Unidos: Wiley Computer Publishing.
	Plata, E. M. (2009). *Business intelligence: una guía práctica.* Peru: UPC, Fondo Editorial, 2009.
MDS	Graham, T. (2012). *Microsoft SQL Server 2012 Master Data Services Second Edition.* Estados Unidos: McGraw-Hill.
MDM	Loshin, D. (2010). *Master Data Management.* Estados Unidos: Morgan Kaufmann Publishers.
Power BI	Microsoft. (07 de 08 de 2019). *docs.microsoft.com*. Obtenido de microsoft: https://docs.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-best-practices
SSAS	Russo, M., & Ferrari, A. (2017). *Tabular Modeling in Microsoft SQL Server Analysis Services Second Edition.* Redmond, Washington: Microsfot Press.
	Greenwood, O. (2017). *Bi Solutions Using Ssas Tabular Model for You!* Estados Unidos: CreateSpace Independent Publishing Platform, 2017.

Apéndices



viernes 19/07/2019 9:15 a. m.

Minor Zúñiga Sedó

Archivo Maestro de Patronos

Para María Damaris Jaen Rodríguez

CC Guideón Gabriel Bojorge Bolaños

Mensaje reenviado el 22/07/2019 9:55 a. m..

Provincia_Canton_Distr...
56 KB

Buenos días Doña Damaris.

Según lo conversado el día de ayer, le informo que en el servicio sFTP de la CCSS, en la carpeta de patronos, se encuentra disponible un archivo maestro de patronos con el nombre pat_20190718133947.lzh.

La estructura de dicho archivo es la misma que se ha utilizado anteriormente y que se refiere a continuación:

Campos de Archivo	Descripcion Campo	Longitud Campo	Tipo Campo
NUM_PATRONO_REF	Número Patronal Anterior	11	Numerico
TIP_PATRONO	Tipo Patrono	1	Numerico
NUM_PATRONO	Número Patrono Actual	11	Numerico
NUM_SEGREGADO	Número de Segregado	3	Numerico
NUM_SECTOR	Número de Sector	3	Numerico
NOM_PATRONO	Nombre de Patrono	120	Alfanumerico
NOM_COMERCIAL	Nombre Comercial del Patrono	120	Alfanumerico
COD_PROVINCIA	Código de Provincia	1	Numerico
COD_CANTON	Código de Cantón	2	Numerico
COD_DISTRITO	Código de Distrito	3	Numerico
DSC_DIRECCION	Dirección de Patrono	70	Alfanumerico
NUM_TELEFONO	Número de Teléfono	10	Alfanumerico
V_NUM_FAX	Número de Fax	10	Alfanumerico
IND_ACTIVO	Indicador de Activo/Inactivo	1	Numerico
NUM_PRESUPUESTO	Número de Presupuesto	11	Numerico
NUM_APARTADO	Número de Apartado Postal	15	Alfanumerico

Asimismo, adjunto el catálogo correspondiente a los códigos de Provincia – Cantón – Distrito utilizados.

Cualquier consulta al respecto estamos a sus órdenes.

1. Correo electrónico SICERE para obtener provincias, cantones y distritos de patrono.



viernes 19/07/2019 7:51 a. m.

Guideón Gabriel Bojorge Bolaños

Requerimientos Cuenta Individual 032-2018 y 001-2018

Para María Damaris Jaen Rodríguez

CC Pablo Rodriguez Guzman; Luis Diego Rodriguez Abarca; Adrián Arnoldo Rojas Céspedes

Buenos días Damaris, de acuerdo a la reunión realizada el día de ayer se acordaron los siguientes puntos para los requerimientos de los indicadores 032-2018 e inclusión de campos de cuenta individual 001-2018.

032-2018 Indicadores:

- Area de CI debe enviar nota de solicitud a DTI solicitando los accesos para poder trasladar las estructuras de datos creadas en pensiones hacia el DataWarehouse central.
- Se va a proceder a terminar la solución en pruebas del DataWarehouse y hacer las pruebas correspondientes a efectos de corroborar que la información solicitada cubre sus necesidades.
- Se espera la información del maestro de los patrono de parte de don Minor a fin de actualizar información importante a efectos de obtener una analítica sobre los datos más amplia.
- La puesta en producción de este requerimiento queda sujeta al momento que se brinden los accesos al repositorio central de la DTI para poder integrar las estructuras de datos creadas.

2. Acuerdos de investigador con Jefa de Área de Cuenta Individual y Control de Pagos

ENTREVISTA DE TRABAJO

Tema: Propuesta de un modelo de inteligencia de negocios para el Área de Cuenta Individual y Control y Pagos.

Objetivo: Recabar información relevante respecto a la propuesta de inteligencia de negocios dentro de la plataforma tecnológica de la Gerencia de Pensiones.

Fecha: 02-05-2019

Pregunta	*Respuesta*
¿Se cuenta con un modelo de inteligencia de negocios actualmente?	**No.**
¿Existen propuestas concretas para desarrollar inteligencia de negocios en el área de Cuenta Individual y Control y Pagos?	**No. Existe una contratación, pero para desarrollarse en otro sistema de información que es inversiones.**
¿Cuentan con las herramientas tecnológicas necesarias para desarrollar un modelo de inteligencia de negocios?	**Si. Tenemos licencias de SQL Server, Analisys Service, Power BI con subscripción de office 365.**
¿Está de acuerdo en que se desarrolle un modelo de inteligencia de negocios para el Área de Cuenta Individual y Control y Pagos?	**Si. Estoy de acuerdo.**

Ing. Guidion Bojorges Bolaños
Analista Sistemas
Entrevistador

GERENCIA DE PENSIONES IVM CCSS ÁREA GESTIÓN INFORMÁTICA

Ing. Pablo Rodriguez G.
Coordinador de Desarrollo de Software
Entrevistado

3. Entrevista de trabajo coordinador desarrollo de software

jueves 06/06/2019 2:58 p. m.

MD María Damaris Jaen Rodríguez

RE: Requerimiento SR-SIP-032-2018 Reportes Indicadores Gestión Calidad de Datos

Para Guideón Gabriel Bojorge Bolaños

CC Evelyn María Hernández Pereira; Marielos Chavarría Cordero; Rodrigo Rovira Pantoja; Pablo Rodriguez Guzman

Respondió a este mensaje el 07/06/2019 8:19 a. m..

De: Guideón Gabriel Bojorge Bolaños <gbojorge@ccss.sa.cr>
Enviado el: jueves, 6 de junio de 2019 14:46
Para: María Damaris Jaen Rodríguez <djaenr@ccss.sa.cr>
CC: Evelyn María Hernández Pereira <ehernanp@ccss.sa.cr>; Marielos Chavarría Cordero <mchavarc@ccss.sa.cr>; Rodrigo Rovira Pantoja <rrovira@ccss.sa.cr>; Pablo Rodriguez Guzman <prodriguezg@ccss.sa.cr>
Asunto: RE: Requerimiento SR-SIP-032-2018 Reportes Indicadores Gestión Calidad de Datos

Buenas tardes Damaris, ya tengo listo la estructura de las consultas para obtener la información de los reportes de cuenta individual. Sin embargo es necesario que nos reunamos con Rodrigo Rovira quién es el encargado a fondo del procesamiento de la información con SICERE. Esto con el fin de determinar detalles en cada uno de los reportes y limitantes que tenemos actualmente respecto a la información que solicitas.

Nos indicas que día y hora puedes atendernos para ver el tema.

Quedamos atentos,

4. Comunicaciones de consulta y establecimiento de reuniones respecto a los indicadores

De: Rodrigo Rovira Pantoja
Enviado el: lunes, 29 de julio de 2019 12:20 p. m.
Para: Guideón Gabriel Bojorge Bolaños <gbojorge@ccss.sa.cr>
Asunto: Carga datos de provincia, cantón y distrito, para los patronos

Buenas tardes.

Finalizado el proceso de carga de los datos de provincia, cantón y distrito, para los datos patronales, en el Pensjobddg1_13, procedo a indicarle las siguientes estadísticas:

Total de registros del archivo: 1.786.555

Total direcciones incluidas: 978.871

Registros con código de Provincia erróneo: 418.501

Registros con código de Cantón erróneo: 2.245

Registros con código de Distrito erróneo: 379.846.

Patronos no registrados: 7.092

5. Proceso de carga provincia, cantón y distrito para los patronos

Acuerdo de Confidencialidad.

Entre nosotros, Bojorges Bolaños Guidion, mayor, soltero, informático, con cédula de identidad número 603790307, vecino de San José, en mi condición de Estudiante, en adelante **"EL OBLIGADO"**; y Pablo Rodriguez Guzmán representante , de la Gerencia de Pensiones, mayor, soltero, informático, documento de identidad 111940526, San José, en calidad de Coordinador de Desarrollo y Mantenimiento de Software de la Caja Costarricense de Seguro Social **(CCSS)** en la Gerencia de Pensiones, domiciliada en San José, avenida segunda, calle cinco y siete, cédula jurídica número cuatro - cero cero cero – cero cuarenta y dos mil ciento cuarenta y siete – cero tres, que en adelante se denominará "LA CAJA", procedemos a suscribir el presente acuerdo de confidencialidad, el cual se rige por las siguientes clausulas:

Desarrollo de trabajo final de graduación en Área de Cuenta Individual y Control y Pagos. De ahora en adelante TFG Área Cuenta Individual y Control y Pagos.

CLAUSULA PRIMERA: Objeto. En el marco de TFG Área Cuenta Individual y Control y Pagos,**EL OBLIGADO,** tendrá acceso mediante la red virtual privada de la institución a los sistemas de información en este caso Sistema Integrado de Pensiones en su ambiente de pruebas. Así mismo acceso físico a los lugares involucrados con el desarrollo del TFG, para efectos de realizar análisis y comprender el negocio a fin de lograr los objetivos del TFG.
EL OBLIGADO, será responsable por el manejo y alcances de la información que accese directamente o incluso en los supuestos en que la dinámica de acceso a la información, le exija compartir la información o el acceso con su casa matriz o casa fabricante, cuando el obligado es un representante – sea único, exclusivo o no- de una casa extranjera en el país.

CLAUSULA SEGUNDA: Que según lo dispuesto por la Caja Costarricense de Seguro Social, ***EL OBLIGADO, y CUALQUIER OTRO TERCERO*** al cual el ***OBLIGADO*** le haya valida y legalmente compartido información o acceso a esta, por la naturaleza del requerimiento según lo indicado en la tesis de graduación, deben comprometerse a mantener la mayor reserva, confidencialidad, discreción y secreto, respecto a todos los datos, parámetros de base de datos, diagramas, documentación, procesos, esquemas de cualquier índole (independiente del medio o formato por el que hayan sido obtenidos) y cualquier otro tipo de información o dato sensible, confidencial o restringido propiedad de **LA CAJA**, o los usuarios y pacientes que está posea , respecto a los cuales tuviere conocimiento en virtud de los permisos y/o accesos que se están autorizando.

CONFIDENCIAL

6. Vista parcial acuerdo de confidencialidad