



Universidad Cenfotec

Maestría en Tecnología de Bases de Datos
Documento final del Proyecto de Investigación Aplicada 2

Proyecto

Tema:

Propuesta de un modelo de minería de datos para la Municipalidad de Esparza,
Puntarenas

Estudiantes:

Fallas Romero Huberth F.

Ruiz Fernández Jorge F.

Fecha: Septiembre de 2018

Declaratoria de derechos de autor

Nosotros, Jorge Félix Ruiz Fernández, con cédula de identidad 602310557 y Huberth Francisco Fallas Romero, con cédula de identidad 603020435, conocedores de las sanciones legales con la que la ley penal de la República de Costa Rica castiga el falso testimonio y el Reglamento Disciplinario de la Universidad Cenfotec, en adelante Cenfotec, declaramos, bajo fe de juramento, lo siguiente: Que ambos somos estudiantes de la Carrera de Maestría en Bases de Datos de Cenfotec y como requisito de graduación y culminación de la carrera, debemos realizar el Trabajo Final de Graduación, en este caso con el título: **Propuesta de un modelo de minería de datos para la Municipalidad de Esparza, Puntarenas. Periodo 2018.**

Por lo anterior, manifestamos que la investigación ha sido elaborada siguiendo las disposiciones exigidas por Cenfotec, de igual forma autorizamos a la Universidad en la conservación del documento, para préstamo y consulta en forma física dentro de la biblioteca de la universidad, así mismo la disposición del documento en forma digital, para la inclusión, comunicación y divulgación dentro del repositorio institucional de Cenfotec en caso de que esta lo posea, con la mención expresa en que ambas modalidades física y digital el préstamo es de carácter académico.

Firmamos en la ciudad de San José a las _____ del _____ de septiembre de 2018.

Huberth Francisco Fallas Romero Cédula 603020435 _____

Jorge Félix Ruiz Fernández Cédula 602310557 _____

Dedicatoria y agradecimientos

A la Universidad Cenfotec, por darnos la oportunidad de acrecentar nuestros
conocimientos.

A nuestras familias, porque son la inspiración de nuestras vidas.

A Dios padre, por habernos dado el soplo de la vida.

Hoja de aprobación del proyecto

Contenido

Capítulo 1. Introducción	1
1.1 Generalidades	2
1.2 Antecedentes del Problema	2
1.3 Definición y Descripción del Problema	4
1.4 Justificación	4
1.5 Viabilidad	5
1.5.1 Punto de Vista Técnico	5
1.5.2 Punto de Vista Operativo	6
1.5.3 Punto de Vista Económico	6
1.6 Objetivos	7
1.6.1 Objetivo General	7
1.6.2 Objetivos Específicos	7
1.7 Alcances y Limitaciones	8
1.7.1 Alcances	8
1.7.2 Limitaciones	8
1.8 Marco de Referencia Organizacional y Socioeconómico	9
1.8.1 Historia	9
1.8.2 Tipo de Negocio y Mercado Meta	9
1.8.3 Misión, Visión y Valores	10
1.8.4 Políticas Institucionales	11
1.9 Estado de la cuestión	12
1.9.1 Concepto	12
1.9.2 Enfoque del estudio	12
1.9.3 Entorno contextual del estudio	13
1.9.4 Criterios utilizados para la selección de fuentes	15
Capítulo 2. Marco Teórico	16
2.1 Data Warehouse	17
2.2 Tipos de modelos de Data Warehouse	18
2.2.1 Modelo de Estrella	18
2.2.2 Modelo en Copo de Nieve	22
2.3 Enfoques del Data Warehouse	23

2.3.1 Ralph Kimball	23
2.3.2 Bill Inmon	25
2.4 Minería de datos	27
Capítulo 3. Marco Metodológico	34
3.1 Tipo de Investigación.....	34
3.2 Alcance Investigativo.....	35
3.3 Enfoque.....	36
3.4 Diseño	37
3.5 Población y Muestreo.....	38
3.6 Instrumentos de Recolección de Datos	39
3.6.1 Estudios de correlación.....	39
3.6.2 Estudios de desarrollo.....	40
3.6.3 El Cuestionario.....	40
3.6.4 La Entrevista.....	40
3.7 Técnicas de Análisis de la Información.....	41
3.8 Estrategia de Desarrollo de la Propuesta	41
Capítulo 4. Análisis del Diagnóstico.....	43
Capítulo 5. Propuesta de Solución	47
Capítulo 6. Conclusiones y Recomendaciones	81
6.1 Conclusiones	81
6.2 Recomendaciones	82
Bibliografía.....	84

Lista de figuras, gráficos, cuadros y/o ilustraciones

Figura 1. Modelo en Estrella.....	19
Figura 2. Modelo en Copo de Nieve.	22
Figura 3. Arquitectura de un Data Warehouse con un Staging Área.	24
Figura 4. Arquitectura de un Data Warehouse con un Data Marts.	26
Figura 5. Minería de Datos como un paso en el proceso de descubrimiento de conocimiento.	28
Figura 6. Disciplinas para un efectivo Gobierno de Datos.	31
Figura 7. Modelo de Madurez para Gobierno de Datos.	32
Figura 8. Cuadro resumen encuesta Gobernanza de datos.....	44
Figura 9. Cuadro resumen encuesta Gobernanza de datos.....	45
Figura 10. Ausencia de Gobernanza de Datos.....	46
Figura 11. Migración de datos en archivos.....	52
Figura 12. Análisis de archivos para migración.....	53
Figura 13 .Script de migración para los archivos .dmp.....	59
Figura 14. Segmento de Base de Datos Operacional.....	60
Figura 15 Propuesta de Staging Area del Proyecto.....	61
Figura 16 Vista general E.T.L.	62
Figura 17. Cargando Staging Area de Ingresos.	63
Figura 18 Cargando Staging Area de Solicitudes.	64
Figura 19. Vista General proceso Carga de Data Warehouse.	65
Figura 20 Cargando las Dimensiones.....	66
Figura 21 . Cargando las Dimensiones Propiedad y Tiempo.....	67
Figura 22 Cargando la Dimensión transacción.	68
Figura 23. Cargando tablas de Hechos.	69
Figura 24 . Proceso completo de carga con Limpieza Staging Area.....	70
Figura. 25 Modelo de Data Warehouse.....	71
Figura 26 Visualización de Ingresos de los Servicios por Periodo.	72
Figura 27. Visualización de Solicitudes por Distrito.....	73
Figura 28 Análisis Descriptivo Cantidad de Propiedades por Distrito.....	74
Figura 29 Análisis descriptivo clientes por rango de edad.....	75
Figura 30 Análisis correlación variables zonas homogéneas por distrito.....	76
Figura 31. Gráfico distribución de zonas por distrito.....	77
Figura 32. Gráfico distribución zona específica por distrito.	78
Figura 33. Matriz de distribución de mujeres con base en su rango de edad por distrito.....	79

Resumen Ejecutivo

La manera paulatina en que las tecnológicas orientadas al almacenamiento y administración de bases de datos han venido cambiando, así como la tendencia de las empresas por reconocer la importancia de adoptar una cultura de datos que permita su propia sostenibilidad a lo largo del tiempo, garantizando seguridad, calidad y accesibilidad en todo momento y el hecho de que en la Municipalidad de Esparza, así como en muchas otras del país, no se cuenta con modelos de minería de datos que permitan, de alguna forma, automatizar la elaboración de información oportuna para la toma de decisiones, generan el planteamiento de este proyecto.

El fin primordial es abordar el tema de minería de datos y de los procesos que esta conlleva, que sean adecuados tanto para la implementación de modelos descriptivos como predictivos sobre los datos administrados, considerando, además los siguientes factores:

1. Los datos se encuentran distribuidos tanto en forma estructurada, como no estructurada.
2. Los procesos actuales realizados por las diferentes instancias internas de la municipalidad, para la toma de decisiones.
3. La oportunidad de negocio que estos nuevos procesos podrían arrojar.
4. La posibilidad de establecer las bases de una cultura orientada a los datos, en forma integral, dentro del municipio.

Capítulo 1. Introducción

La forma en la cual las empresas hacen uso de tecnología de punta para el apoyo de sus actividades diarias permite el análisis de diferentes alternativas en el mercado; estas tendencias tecnológicas no solo posibilitan enfocarse en aspectos orientados a actividades muy especiales, sino en aquellos concretos por los cuales se puede sacar provecho a los diferentes recursos de datos tratados y recolectados por dichas tendencias. Resulta de suma importancia identificar los diferentes orígenes de datos, clasificarlos según el tipo y generar, además, una cultura donde la empresa los considere como un activo valioso. Es a partir de este momento cuando la generación de soluciones a través de mecanismos de minería de datos juega un papel primordial, mediante el cual las empresas pueden generar datos fiables que apoyen el proceso en la toma de decisiones. El presente proyecto tiene como objetivo la integración de los diferentes insumos y con ello, proporcionar un modelo de minería de datos que permita satisfacer las diferentes necesidades de la Municipalidad de Esparza.

1.1 Generalidades

En lo referente a las generalidades para este proyecto es importante indicar que los datos de índole sensible se encuentran sujetos a la aplicación de la Ley de protección de datos del Gobierno de Costa Rica, así como cualquier otro tipo de información privada, en casos de atención de violencia doméstica o seguridad ciudadana, que pongan en riesgo el actuar de las instituciones destinadas a la atención de esta naturaleza de incidencias.

1.2 Antecedentes del Problema

La repentina y creciente generación de transferencias de competencias del gobierno central enmarcado en los diferentes ámbitos destinados al régimen municipal, genera no solamente la atención de actividades extraordinarias a las habituales, sino, por consiguiente, propicia, entre otros aspectos, el establecimiento de mecanismos de control interno que aseguren medir el nivel de cumplimiento y alcances concretados. La atención de todas estas obligaciones promueve la integración interinstitucional que requiere de la concentración de datos, tanto estructurados como no estructurados, de manera integral, que sirvan de aporte para el apoyo en la toma de decisiones en aspectos referentes a la administración municipal, salud, seguridad ciudadana, entre otros.

A esta situación se une el hecho de que a mediados del 2017 la Municipalidad de Esparza logró concretar y poner en marcha un proceso de migración de datos, de un

modelo basado en archivos dbf administrados por Visual Fox, a un modelo de bases de datos relacional; dicho proceso de migración arrojó una serie de hallazgos, tales como:

1. Inconsistencia de los datos en términos de integridad referencial.
2. Formato de los datos, por ejemplo, fechas y comas en lugar de punto decimal para números.
3. Campos que deberían haber sido mandatorios o exigidos con valores nulos.
4. Duplicidad de registros.

Así, aunque se encontraron problemas como los anteriores, estos no tienen una representación significativa ante el volumen de datos migrados.

Otro elemento importante de mencionar, radica en que la gerencia de la Municipalidad de Esparza quiere mejorar los tiempos de respuesta a los clientes del municipio en cuanto a trámites se refiere, y de igual forma, incrementar la capacidad de resolver problemas relacionados con los proyectos de desarrollo comunal en el cantón; así, aprovechando las nuevas prestaciones obtenidas a raíz de la migración de datos y que no existe una cultura de datos para la toma de decisiones más precisa y más amplia, surge la necesidad de desarrollar este proyecto.

1.3 Definición y Descripción del Problema

En el sentido de que las municipalidades proporcionan, coordinan y gestionan múltiples servicios conocidos como lo tradicional del municipio, se ven impactadas por la atención de nuevos conceptos que el Gobierno central de la República otorga mediante el traslado de competencias. Estos nuevos servicios, tareas y conceptos generan información que, en muchos de los casos, debe verse en concejos integrados como son los CCCI, los cuales, en este caso corresponden al Concejo Cantonal de Coordinación Interinstitucional del cantón de Esparza. Los datos generados en estos entes, así como los pertenecientes a la administración municipal en sus múltiples quehaceres, no se encuentran presentes en un contenedor principal que permita la integridad de los múltiples contextos que interactúan entre sí; además de la generación de modelos de minería de datos que puedan generar el apoyo en la toma de decisiones, considerando aspectos como seguridad ciudadana, impactos de riesgos naturales, análisis de tendencias contexto social del cantón de Esparza y lo referente a la administración, en general.

1.4 Justificación

Es importante para la Municipalidad de Esparza, como gobierno local, generar estrategias tanto de negocio como para el apoyo interinstitucional, que permitan mejorar las diferentes acciones en las cuales se encuentra involucrado el municipio. Contar con un contenedor principal de datos que no solamente considere los de la administración municipal, sino además, los proporcionados por otras organizaciones

del cantón en diferentes ámbitos de acción, permitirá construir modelos de minería de datos que puedan predecir algunos comportamientos, a futuro, que afecten las condiciones actuales del cantón. Otra condición es encontrar nuevas oportunidades de negocio que puedan generar mejores condiciones para los habitantes; el rol de las municipalidades comienza a tomar fuerza cada vez con una ciudadanía que demanda más y mejores servicios de calidad. El análisis de los datos generado en la confección de una ciudad inteligente, serán igual de valiosos para fortalecer un cantón que no solo busca crecer en tecnología, sino anticipar, conocer y prever todo lo que pueda contribuir con el concepto de una ciudad más integral.

1.5 Viabilidad

1.5.1 Punto de Vista Técnico

Es importante señalar que desde el punto de vista técnico los requerimientos solicitados son factibles de elaborar; al ser un proyecto tan integral se deben establecer los alcances de este, para enfocarse en productos específicos pertenecientes a temas señalados por la administración municipal para enfocar esfuerzos en estos temas de gran relevancia. Por otra parte, los entornos tecnológicos existentes en el mercado son de conocimiento de la Municipalidad de Esparza, a través de su Unidad de Tecnología de Información; por consiguiente, la identificación y estudio de los diferentes recursos de datos estructurados y no estructurados en ámbitos tanto internos como externos, serán accesibles para

elaborar las diferentes propuestas de estructuras que logren satisfacer los posteriores modelos o mecanismos de minería de datos.

1.5.2 Punto de Vista Operativo

El punto de vista operativo se encuentra plenamente identificado en las diferentes áreas involucradas en el proyecto, donde posteriormente se comprometerán para analizar el comportamiento de los análisis proporcionados por los mecanismos de minería de datos y además suministrar los datos necesarios para la continuidad de la herramienta. En el ámbito técnico, la municipalidad contará con el personal profesional capacitado, en caso de requerirse algún cambio dentro de los requerimientos iniciales establecidos para seguir brindando una herramienta acorde con la lógica de negocio plasmada.

1.5.3 Punto de Vista Económico

La factibilidad económica resulta de vital importancia en un proyecto. En este caso, los requerimientos tecnológicos serán suministrados mediante presupuesto municipal, contemplado por la Unidad de Tecnología de Información, para satisfacer los diferentes aspectos, tanto tecnológicos como materiales, dentro del proyecto, donde se ven contemplados ambientes productivos y de desarrollo para la implementación de la propuesta.

Tabla 1

Insumos necesarios para el desarrollo del proyecto

Insumos Proyecto	Cantidad	Costos de desarrollo
Adquirir Licenciamiento SABD Oracle	50	¢30,000,000.00
Renovar Seguro Actualización SABD Anual	1	¢7,000,000.00
Horas consultor para el Análisis, desarrollo, implementación y capacitación del proyecto.	300	¢ 2,550,000.00

Nota. Recuperado de estudio de mercado realizado por la Unidad de Tecnología de Información de la Municipalidad de Esparza.

1.6 Objetivos

1.6.1 Objetivo General

Elaborar un data *warehouse* que permita realizar minería de datos para la toma de decisiones en la Municipalidad de Esparza.

1.6.2 Objetivos Específicos

Identificar el nivel de importancia que poseen las diferentes fuentes de datos externos para cada unidad funcional de la Municipalidad de Esparza.

Aplicar las mejores prácticas en la confección de los diversos *datamart*, de forma que satisfagan los requerimientos de datos de las diferentes unidades institucionales.

Confeccionar cada uno de los procesos E.T.L necesarios para la incorporación de las fuentes de datos al contenedor único de la Institución.

Considerar elementos descriptivos de la minería de datos para lograr la presentación de los datos acorde con las necesidades de negocio de la organización del cantón.

1.7 Alcances y Limitaciones

1.7.1 Alcances

Lograr la implementación de un cubo de datos que permita satisfacer los requerimientos funcionales para cada una de las entidades internas de la municipalidad.

Confeccionar elementos descriptivos de minería de datos que permitan generar insumos de información de carácter descriptivo para el apoyo en la toma de decisiones.

1.7.2 Limitaciones

Considerando las diversas instancias municipales que desarrollan proyectos tanto a lo interno como a lo externo de la Institución, este proyecto de investigación no podrá establecer los *DataMarts* en su totalidad, para cubrir la demanda de cada una de las entidades o departamentos que forman el municipio.

Lograr la conversión, tanto efectiva como automática, de aquellos datos no estructurados que podrían ser requeridos. Así como lograr una cultura real orientada a los datos, tanto en las entidades internas del municipio, como en el CCCI.

1.8 Marco de Referencia Organizacional y Socioeconómico

1.8.1 Historia

El origen del nombre del cantón se debe al Gobernador de la Provincia don Diego de Artieda y Chirino, quien, en 1577, en memoria de la Villa de Esparza, situada en el Valle de Salazar (Navarra), en las cercanías de Pamplona, España, de donde era originario, le agregó a la denominación de la ciudad del Espíritu Santo las palabras “de Esparza”.

La primera municipalidad instalada en el año de 1877, fue integrada por los señores Felipe Herrera, Ignacio Pérez (en cuya memoria se bautizó el parque central de la ciudad) y Marcelino Zúñiga.

1.8.2 Tipo de Negocio y Mercado Meta

Para el caso de negocio, la Municipalidad de Esparza se encuentra como un gobierno local que realiza la prestación de servicios, entre los cuales destacan: obras y ornato, recolección de desechos sólidos, reciclaje, limpieza de vías y sitios públicos.

En cuanto al mercado meta, hace referencia a todos los ciudadanos que habitan el cantón, o bien los que cuentan con un bien inmueble dentro de este; además, al turismo presente en ciertas zonas.

1.8.3 Misión, Visión y Valores

Misión

Ser un gobierno local transparente, moderno, eficaz y eficiente, que promueve la participación ciudadana y el compromiso de todos los sectores de la sociedad para lograr el desarrollo humano de Esparza.

Visión

Ser un gobierno municipal participativo, con capacidad de gestión, económicamente sostenible, con recursos humanos calificados y con credibilidad para liderar, coordinar y concertar el desarrollo humano de Esparza.

Valores en acción

VALORES	PRINCIPIOS
Integridad	Rectitud y ética en el ejercicio de nuestras labores.
Solidaridad	Sensibilidad ante la condición de los sectores más necesitados.
Transparencia	Capacidad de realizar nuestro trabajo y dar cuenta por los resultados.
Compromiso	Involucrarse de lleno en lo que hacemos.
Responsabilidad	Cumplimiento efectivo de nuestros deberes.

1.8.4 Políticas Institucionales

Desarrollo Institucional Municipal	<p>Las aspiraciones de desarrollo del pueblo esparzano serán atendidas por medio del fortalecimiento de las unidades gestoras del ingreso municipal.</p> <p>Cambio de actitud institucional hacia un nuevo rol de gobierno local.</p> <p>Fortalecimiento de las capacidades de los Concejos de Distritos.</p> <p>Gestión inmediata de la edificación de un palacio municipal moderno que permita brindar un servicio de calidad a los esparzanos.</p>
Equipamiento	<p>Los servicios comunales existentes son apoyados por medio del suministro de equipo requerido.</p>
Desarrollo económico local	<p>La Municipalidad de Esparza se da a la tarea de identificar las actividades que produzcan el mayor impacto en el desarrollo económico sostenible del cantón.</p> <p>La Municipalidad de Esparza acompaña aquellos proyectos productivos que responden a actividades identificadas como altamente impactantes en el desarrollo del cantón.</p>
Medio Ambiente	<p>Garantizar la disponibilidad de agua en el largo plazo es un compromiso de la Municipalidad de Esparza.</p>
Política Social Local	<p>La identidad cultural y la promoción de valores son parte del quehacer permanente de la municipalidad.</p> <p>Potenciar el desarrollo humano de la comunidad esparzana.</p>
Infraestructura	<p>La Municipalidad de Esparza priorizará la infraestructura a desarrollar de acuerdo con la expresión de los ciudadanos del cantón.</p>

Servicios Municipales	Los servicios municipales brindados deberán ser de calidad, con autosuficiencia financiera y además, amigables con el ambiente.
Ordenamiento Territorial	El Plan Regulador cantonal y el Plan Regulador costero constituirán el marco de referencia para el desarrollo del cantón.

1.9 Estado de la cuestión

1.9.1 Concepto

También conocido como estado del arte, se refiere meramente al estudio previo realizado por el investigador para determinar el conocimiento actual sobre el tema o campo de estudio por tratar. “En consecuencia, tales estudios muestran el conocimiento relevante y actualizado, las tendencias, los núcleos problemáticos, los vacíos, los principales enfoques o escuelas, las coincidencias y las diferencias entre esas hipótesis y los avances sobre un tema determinado”. (Bernal, 2010, p.112)

1.9.2 Enfoque del estudio

La situación económica que enfrenta Costa Rica, de unos años a la fecha y aunado a situaciones climatológicas que han provocado daños tanto en la infraestructura nacional, como comercio e industria, han provocado la reducción del presupuesto o finanzas públicas para el desarrollo de proyectos o programas gubernamentales, en sus instituciones y por ende, en el país.

Existen, además, otros factores, tales como:

1. Que los municipios como gobiernos locales han venido incrementado su rol en el desarrollo cantonal y aparte de eso, como garantes del uso adecuado de los recursos del Estado.
2. Que existe una política nacional de contención del gasto público.

Por lo anteriormente mencionado, los objetivos plasmados en esta investigación adquieren importancia relevante para la gerencia de la Municipalidad de Esparza, especialmente el de contar con un contenedor y mecanismos de minería de datos adecuados para la toma de decisiones oportunas, que permitan maximizar los recursos económicos del municipio, en pro del desarrollo de la zona administrada; así como el de mejorar la recaudación de fondos a través de sus diferentes impuestos y cánones.

Entender las razones por las cuales la Municipalidad de Esparza debe contar con la tecnología informática que brinde soporte a la toma de decisiones, es el primer paso para la formulación de los alcances de este proyecto; en forma general, establecen la implementación y puesta en marcha de un cubo de datos y estrategias de consulta que suplan los intereses del municipio.

1.9.3 Entorno contextual del estudio

La Municipalidad de Esparza posee una serie de normativas, formularios y plantillas que definen tanto los datos como el proceso de registro de estos, así como

el propietario interno, considerando cada uno de los departamentos de la Institución; no obstante, todavía no existe una orientación institucional hacia los datos, que permita, entre otros aspectos, inventariarlos y determinar, de la colección existente, aquellos que son comunes entre departamentos.

Revisando el contexto nacional se ha encontrado un documento o tesis universitaria, que permite establecer patrones de conformación de los cubos con problemas reales de Costa Rica; obsérvese:

1. Informe final de tesis por el grado de Máster en Computación de la postulante Giannina Ortiz Quesada, denominado “ESTABLECIMIENTO DE UN GRUPO DE INDICADORES PARA LA TOMA DE DECISIONES EN EL SECTOR VIVIENDA UTILIZANDO TÉCNICAS DE *MINERIA DE DATOS*”, del Instituto Tecnológico de Costa Rica. (Obtenido del <https://repositoriotec.tec.ac.cr>)

Fuera de este documento no se encontró otro proyecto relacionado con el tema de minería de datos en el entorno de empresas municipales. Lo anterior no significa, necesariamente, que no se estén realizando proyectos en esta área en el entorno nacional.

Actualmente la Municipalidad de Esparza, al igual que otras en el país, se encuentran realizando procesos de migración de datos de sus sistemas actuales a una nueva plataforma; por ello, relacionado con el tema de estudio, no existen otras experiencias previas a nivel municipal, en torno a minería y cubos de datos, así como de inteligencia y analítica de negocios.

1.9.4 Criterios utilizados para la selección de fuentes

Criterios para la selección de fuentes bibliográficas

1. Relacionados con el diseño e implementación de cubos de datos; esto incluye desde buenas prácticas, hasta su ejecución en un motor de datos.
2. Relacionados con los procesos de extracción, carga y transformación de los datos, así como procesos y buenas prácticas relacionadas con la limpieza de datos.
3. Material relacionado con el almacenamiento, procesamiento y consulta de datos que los municipios tengan en su haber, por ejemplo, normativas sobre el registro y aprobación de datos, derechos de uso por usuarios internos.

Criterios para la selección del personal municipal

1. Deben intervenir en los procesos de toma de decisión de las diferentes dependencias de la institución.
2. Deben conocer los procesos de registro, tratamiento, depuración de los datos administrados.
3. Deben tener conocimiento sustancial sobre la afectación de los datos administrados sobre la toma de decisiones de otras dependencias internas.

Capítulo 2. Marco Teórico

Briones, citado por Hugo Cerda (1998) y este a su vez citado por César Bernal (2010), señala:

El marco teórico a niveles más específicos y concretos comprende la ubicación del problema en una determinada situación histórico – social, sus relaciones con otros fenómenos, las relaciones de los resultados por alcanzar con otros ya logrados, como también definiciones de nuevos conceptos, redefiniciones de otros, clasificaciones, tipologías por usar, etc. (p.125)

Los cambios en la forma en cual la Municipalidad de Esparza quiere tratar sus datos internos para una mejor toma de decisiones, de elaborar mecanismos que le ayuden a predecir el comportamiento del desarrollo cantonal, así como incorporar los datos que los CCCI manipulan y ahondado en el hecho de que el municipio se encuentra en un proceso de migración y depuración de datos a un entorno de bases de datos relacionales, evidencian la necesidad de aprovechar el impulso que dicha institución está planteando para convertirse a una cultura orientada a los datos.

En este capítulo se abordarán aquellos conceptos relacionados con la implementación de un *Data Warehouse* y su correspondencia expresa con los objetivos de este proyecto de investigación, así se podrá justificar el porqué de los conceptos mencionados y además, se abordarán aquellos relativos a los procesos del negocio atinentes a esta investigación.

2.1 Data Warehouse

“Un Data Warehouse permite la unificación de información obtenida mediante la recolección de diversas fuentes de datos relacionados con los procesos operacionales y administrativos de la empresa”. (Darmawikarta, 2007, p.1)

Considerando este concepto, la Municipalidad de Esparza, a partir de una iniciativa nacida en el seno de la Unidad de Informática, se plantea la necesidad de establecer un contenedor de datos que permita agrupar los provenientes de los diversos sistemas internos que esta posee, como lo son: el sistema de ingresos propiamente relacionado con las actividades de cobros por concepto de impuesto de bienes e inmuebles, patentes comerciales, derechos de nichos en el cementerio local y de los fondos que gira el Gobierno central; además, otros datos provenientes del sistema de catastro que identifica la asignación de fincas con sus correspondientes dueños y que se contrasta con los datos obtenidos del Registro Nacional de la Propiedad.

A estos datos deben sumarse los relacionados con el sistema de egresos, que propiamente determina cómo son invertidos los fondos municipales, a partir de los diferentes proyectos de inversión en el cantón.

De acorde con lo expresado, un *Data Warehouse*, es una base de datos relacional que permite distribuir los datos recolectados para efectos de consultas, que ayuden a la toma de decisiones.

Los datos de los sistemas externos deben pasar por un proceso conocido por sus siglas en inglés como E.T.L (extracción, transformación y carga), con cierta periodicidad al *Data Warehouse*, para que este se mantenga lo más actualizado posible, acorde con el nivel de operatividad de dichos sistemas externos.

Este requerimiento funcional del *Data Warehouse*, le permite convertirse en un contenedor de datos tanto actuales como históricos, donde los primeros permiten a la gerencia de la empresa reconocer la situación actual de esta y los segundos le permiten generar las estadísticas y proyecciones requeridas.

De esta forma, la Municipalidad de Esparza contaría con un repositorio que le permitiría consultar los movimientos recientes, para contrastarlos con lo esperado, medir el nivel, por ejemplo, de ingresos en relación con periodos anteriores, o en su parte histórica, ver cómo se comporta el crecimiento comercial en los distritos administrados y poder calcular cómo se comportaría este a un futuro inmediato.

2.2 Tipos de modelos de Data Warehouse

2.2.1 Modelo de Estrella

Esta técnica consiste en declarar una tabla de hechos, rodeada por una serie de tablas conocidas como *dimensiones*. Principalmente se busca tener una visión multidimensional de un proceso, para lo cual la tabla de hechos contiene los

elementos que van a ser consultados en primera instancia y las tablas dimensionales los datos complementarios a cada elemento hecho consultado.

En la siguiente figura se puede apreciar cómo se vería un modelo de estrella

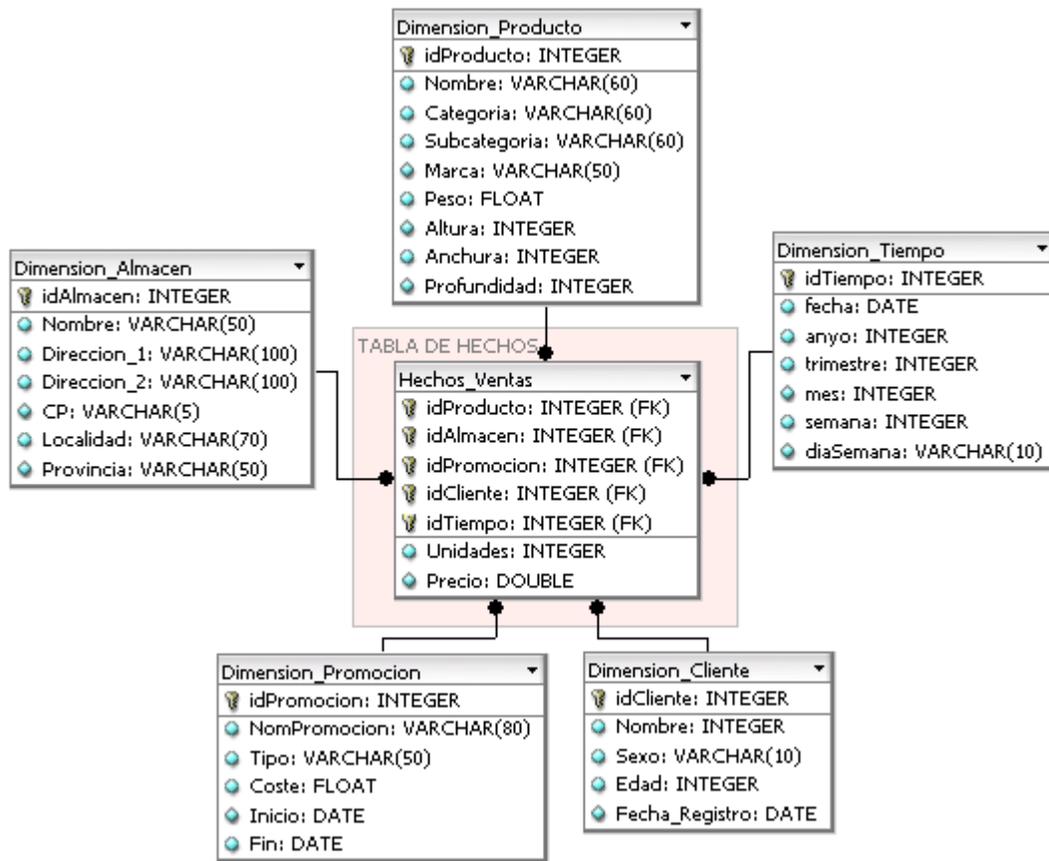


Figura 1. Modelo en Estrella.

Figura 1. Modelo en Estrella.

Fuente: Wikipedia (2018). Modelo en Estrella. [Figura] (Recuperado de

https://es.wikipedia.org/wiki/Esquema_en_estrella)

Aunque una tabla de hechos puede tener la estructura anterior donde todos los campos que corresponden a llaves foráneas son una llave principal compuesta, puede también contener una llave primaria, completamente diferente a la indicada y queda únicamente la propuesta de dichos campos como meras llaves de relación.

Este modelo es una de las formas más básicas de implementar consultas; permite el almacenamiento de datos redundantes con el objetivo de minimizar los tiempos de respuesta, en contraposición con el modelo relacional, de tal forma que la tabla de hechos se encuentra en tercera forma normal y las tablas dimensionales en segunda forma.

Esta situación permite construir consultas más simples, donde media el acceso a una o pocas tablas.

Normalmente las bases de datos usadas para producción cumplen con una serie de reglas de diseño y normalización que exigen la fragmentación de los datos en forma equilibrada y generan relaciones y dependencias entre las tablas que conforman la base de datos como tal. La propuesta del modelo en estrella más bien indica:

1. Que las tablas dimensionales contengan los datos necesarios para satisfacer las consultas del cliente, aunque esto involucre mezclar datos de varias tablas en una sola.

2. Crear una tabla de tiempo, con una serie de campos que permitan asociar cada parte del tiempo en valores más simples; esto con la idea de agilizar las consultas sobre campos nativos, tales como enteros en lugar de campos tipo fecha, los cuales son más complejos de procesar por parte del motor de datos. Esta tabla tiene, como función, dimensionar los periodos en que suceden los hechos por consultar.

3. La tabla de hechos solo debe contener las llaves de todas aquellas tablas dimensionales involucradas en la consulta, incluyendo la dimensión del tiempo.

4. Este diseño permite establecer la funcionalidad de una base de datos multidimensional, utilizando una base de datos relacional.

2.2.2 Modelo en Copo de Nieve

Consiste en declarar una tabla de hechos rodeada por una serie de tablas también conocidas como *dimensiones*; pero he aquí la diferencia con él modelo de estrella, pues las *tablas dimensiones*, pueden tener dependencias sobre otras tablas.

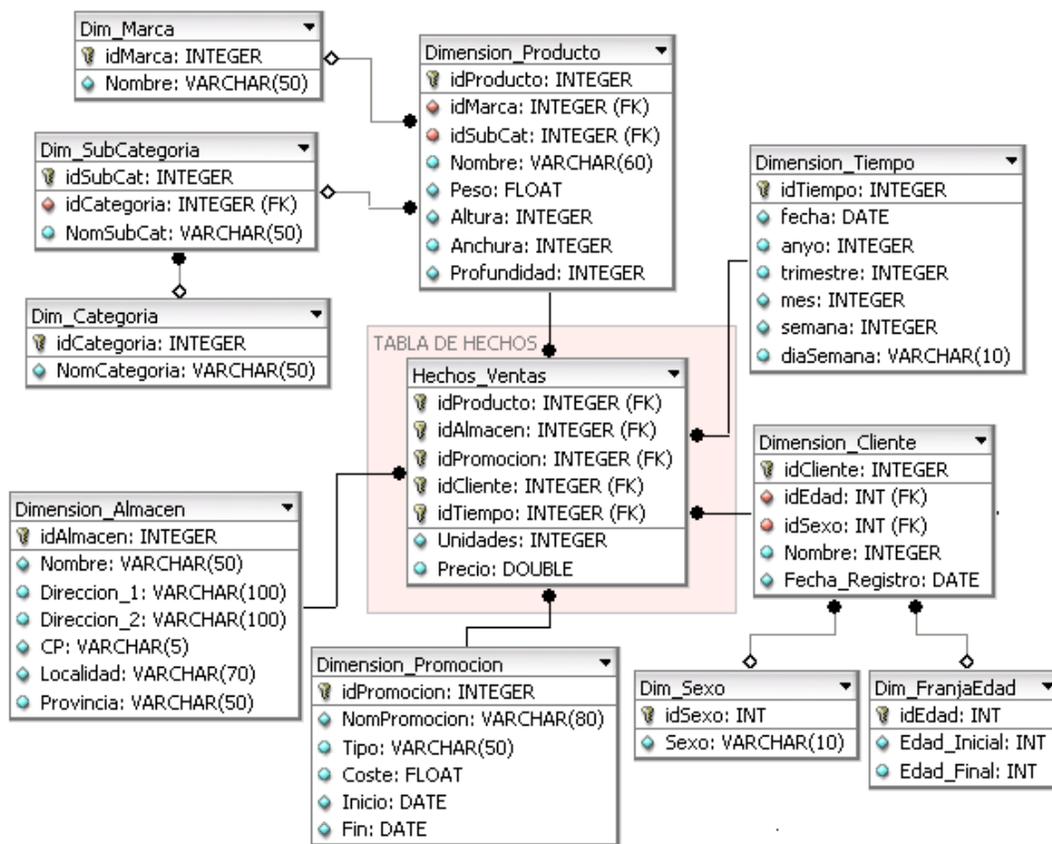


Figura 2. Modelo en Copo de Nieve.

Figura 2. Modelo en Copo de Nieve.

Fuente: Wikipedia (2013). Modelo en Copo de Nieve. [Figura] (Recuperado de https://es.wikipedia.org/wiki/Esquema_en_copo_de_nieve)

El modelo de copo de nieve puede emplearse para dividir la información de una tabla dimensional que se vuelve muy pesada para ser consultada; se extrae de dicha tabla, aquellos campos que podrían tener valores repetidos y colocarlos en una tabla adicional.

Esta alternativa libera peso en la propuesta de la tabla dimensional, pero agrega una dependencia con la nueva tabla y por tanto, la complejidad de armar aquellas consultas que requieran de esos datos extras.

En relación con las necesidades de la Municipalidad de Esparza, aun no se tiene propiamente definido cómo sería la propuesta del *Data Warehouse*.

2.3 Enfoques del Data Warehouse

2.3.1 Ralph Kimball

En forma resumida, Kimball define que el *Data Warehouse* debe construirse a partir de la identificación de los principales procesos del negocio y del principio de cuáles son las principales preguntas que el *Data Warehouse* debe responder. Para elaborar el *Data Warehouse*, se deben realizar los siguientes pasos:

1. Analizar las fuentes de datos que habrán de usarse (estructuradas o no).
2. Definir de estas los datos requeridos.
3. Definir los procesos de E.T.L adecuados para la recolección del dato.
4. Uso del *Staging Área* para los procesos de E.T.L.

Aquí el diseño del *Data Warehouse* asume una estructura desnormalizada, por lo que los procesos de E.T.L deben encargarse de crear los datos en su nuevo formato, para ser almacenados en las tablas dimensionales, legibles para el usuario final y como única relación la existente con la tabla de hechos.

Esta propuesta aplica el modelo o esquema de estrella. Kimball además, indica que el *Data Warehouse* podría contener tantos esquemas de estrella como se requieran, para resolver los principales intereses del negocio.

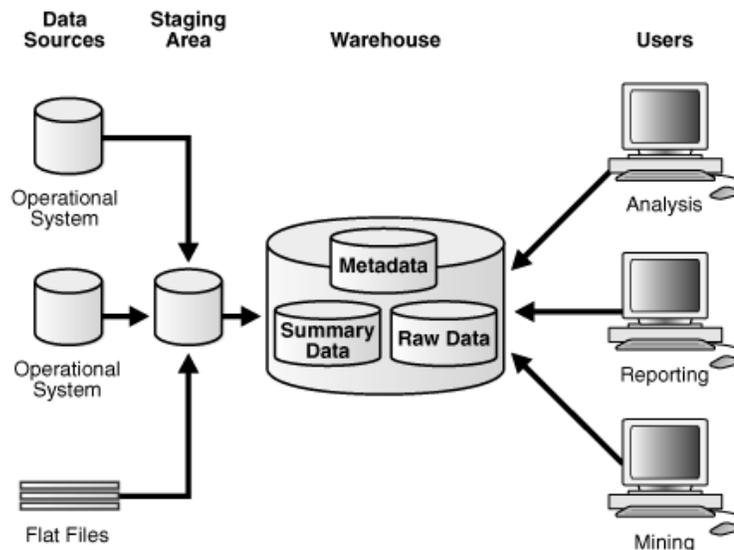


Figura 3. Arquitectura de un Data Warehouse con un Staging Área.

Figura 3. Arquitectura de un Data Warehouse con un Staging Área.

Fuente: Oracle and affiliates, (2017). Arquitectura de un Data Warehouse con un Staging Área. [Figura] (Recuperado de <https://docs.oracle.com/database/121/DWHSG/concept.htm#DWHSG8073>)

Esta propuesta permite que los procesos de consulta no necesariamente sean realizados por un experto en bases de datos, sino, puede ser llevado a cabo por

personal de cada departamento con conocimientos en herramientas de consulta, como tablas pivote en Excel o herramientas de Inteligencia de Negocio depuradas.

2.3.2 Bill Inmon

En forma resumida, Inmon define el *Data Warehouse* a partir del modelo de datos corporativo y no solamente las fuentes de datos operativas, donde se identifican las principales áreas de la empresa, las cuales, al final corresponden a los principales consumidores del *Data Warehouse*. En este enfoque se procede a definir la estructura física que habrá de tener el *Data Warehouse*, muy similar a como lo propone Kimball, en cuanto a:

5. Analizar las fuentes de datos que habrán usarse (estructuradas o no).
6. Definir de estas los datos requeridos.
7. Definir los procesos de E.T.L adecuados para la recolección del dato.
8. Uso del Staging Área para los procesos de E.T.L

No obstante, presentan las siguientes diferencias:

1. Exige que el modelo aplique normalización sobre las tablas propuestas en el *Data Warehouse*, lo cual hace que este sea complejo.
2. El modelo o esquema por utilizar es el de copo de nieve.
3. Otra diferencia es que se desarrolla, en conjunto al *Data Warehouse* principal, una serie de contenedores de datos independientes para cada uno de los

consumidores o líneas del negocio que posee la empresa, conocidos como *Data Marts*.

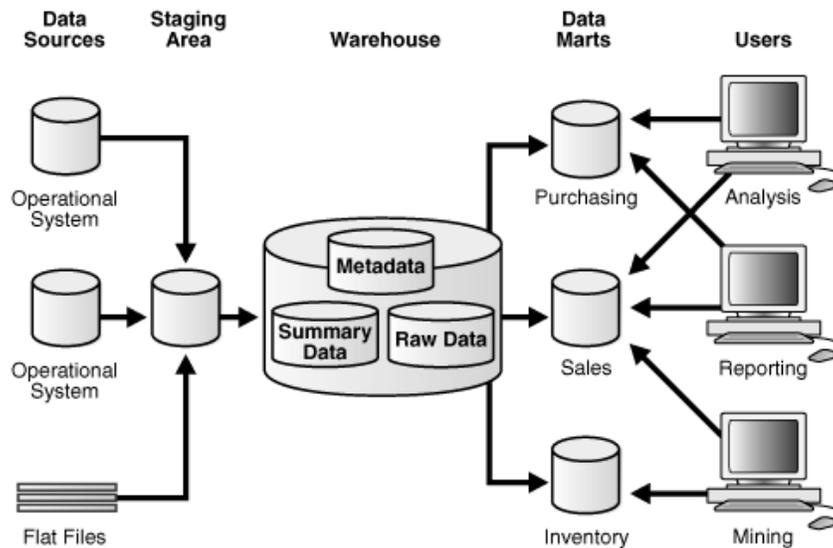


Figura 4. Arquitectura de un Data Warehouse con un Data Marts.

Figura 4. Arquitectura de un Data Warehouse con un Data Marts.

Fuente: Oracle and affiliates, (2017). Arquitectura de un Data Warehouse con Data Marts. [Figura] (Recuperado de <https://docs.oracle.com/database/121/DWHSG/concept.htm#DWHSG8075>)

En la figura anterior, en forma de ejemplo aparecen diferentes *Data Marts* orientados a diferentes líneas de negocio, como lo son *ventas, inventario, compras* y *los procesos básicos de minería de datos; reporte y análisis* se desarrollan sobre los *Data Marts* y no sobre el *Data Warehouse* propiamente.

Este modelo presentado, por Inmon, permite una mejor personalización de los datos para cada uno de los consumidores internos de la empresa, donde cada *Data*

Mart, posee la colección de datos que interesa a cada uno de estos; adicionalmente, las consultas que se realizan sobre cada *Data Mart* no deberían alterar el rendimiento sobre las hechas a los otros *Data Marts* de la propuesta.

En contraposición, el desarrollo de esta técnica supone tanto espacio extra para colocar cada uno de los *Data Marts*, como de procesos ETL para colocar los datos correspondientes del *Data Warehouse* principal a cada uno de los *Data Marts* requeridos y que probablemente requieran de un experto tanto en el modelaje de datos, como del negocio, debido al esquema normalizado que se aplica en este enfoque.

2.4 Minería de datos

“El Data Mining es un conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos que expliquen el comportamiento de estos datos”. (Ester Ribas, 2018)

La minería de datos se convierte en un paso clave para obtener conocimiento por medio de los datos, el descubrir el valor agregado contenido en los datos aporta a la entidad aspectos que permiten apoyar la toma de decisiones; esto mediante el conocimiento adquirido. El proceso de descubrir el conocimiento requiere puntualizar en acciones claves, dentro de las cuales destacan la limpieza de los datos, su integración, selección datos, transformación datos, minería de datos, evaluar patrones y presentar conocimiento.

Para evidenciar cómo se efectúa el proceso se incorpora la siguiente figura, la cual detalla esas acciones en un orden lógico.

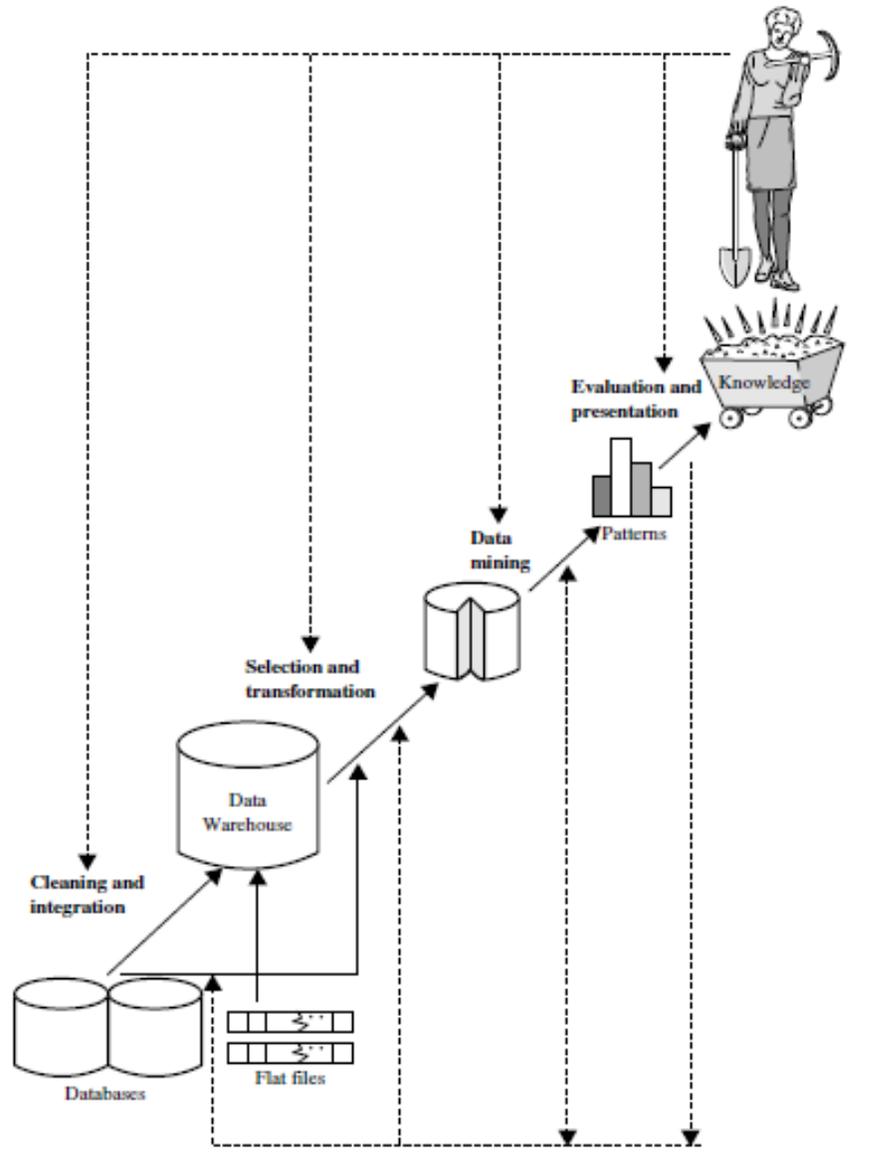


Figura 5. Minería de Datos como un paso en el proceso de descubrimiento de conocimiento.

Figura 5. Minería de Datos como un paso en el proceso de descubrimiento de conocimiento.

Fuente: Han, J. (2012). Minería de Datos como un paso en el proceso de descubrimiento de conocimiento. [Figura]

(Recuperado de Data Mining Concepts and Techniques 3rd Edition)

La persecución de un panorama claro que oriente y obligue en la administración y control de los datos, resalta el hecho del concepto de la gobernanza de los datos, que de forma implícita se manifiesta en las acciones para obtener conocimiento.

Así, “... el gobierno de datos es un conjunto de procesos que garantiza que los activos de datos importantes se gestionan formalmente en toda la empresa. La gobernanza de datos garantiza que los datos pueden ser confiables y que las personas pueden ser responsables de cualquier evento adverso que ocurra debido a la baja calidad de los datos”. (Sarsfield, 2009, p.23)

La ausencia de un gobierno de datos resalta la omisión de una cultura donde los usuarios expertos no asumen su responsabilidad sobre el control de sus datos y enfrentan características inadecuadas que ponen en riesgo la operación de la empresa, dentro de las cuales resaltan: la redundancia, desactualización, incompletitud y la omisión en la revisión de los datos.

La idoneidad de un gobierno de datos, como lo menciona Sarsfield, generará la garantía de los datos, responsabilizando a los individuos de cualquier eventualidad que ponga en duda la calidad de los datos a su cargo.

En busca de establecer una gobernanza de los datos, Jimmy Martínez, en su artículo para la empresa IBM, detalla los siguientes pasos para lograr un gobierno de estos:

1. Establecer metas. Sentencias principales que guían la operación y desarrollo de la cadena de suministro de información.

2. Definir métricas. Conjunto de medidas usadas para evaluar la efectividad del programa y los procesos de gobierno asociados.
3. Tomar decisiones. La estructura organizacional y el modelo de cambio ideológico para analizar y crear políticas de decisión.
4. Comunicar políticas. Herramientas, habilidades y técnicas usadas para comunicar decisiones políticas a la organización.
5. Medir resultados. Comparar resultados de las políticas con las metas, entradas, modelos de decisión y comunicación para proveer constante retroalimentación sobre la efectividad de la política.
6. Auditar. Herramienta usada para comprobar todo.

Martínez, Jimmy. (2012). (Recuperado de <https://www.ibm.com/developerworks/ssa/data/library/techarticle/gobierno-datos/index.html>)

Considerando la propuesta de los pasos mencionados para la aplicación de un gobierno de datos, se considera de gran relevancia potenciar su desarrollo mediante disciplinas que generen un valor agregado en la efectividad de su alcance.

Disciplinas para un efectivo Gobierno de Datos



Figura 6. Disciplinas para un efectivo Gobierno de Datos.

Figura 6. Disciplinas para un efectivo Gobierno de Datos.

Fuente: Martínez, J. (2012). Seis pasos para el Gobierno de Datos. [Figura] (Recuperado de <https://www.ibm.com/developerworks/ssa/data/library/techarticle/gobierno-datos/index.html>)

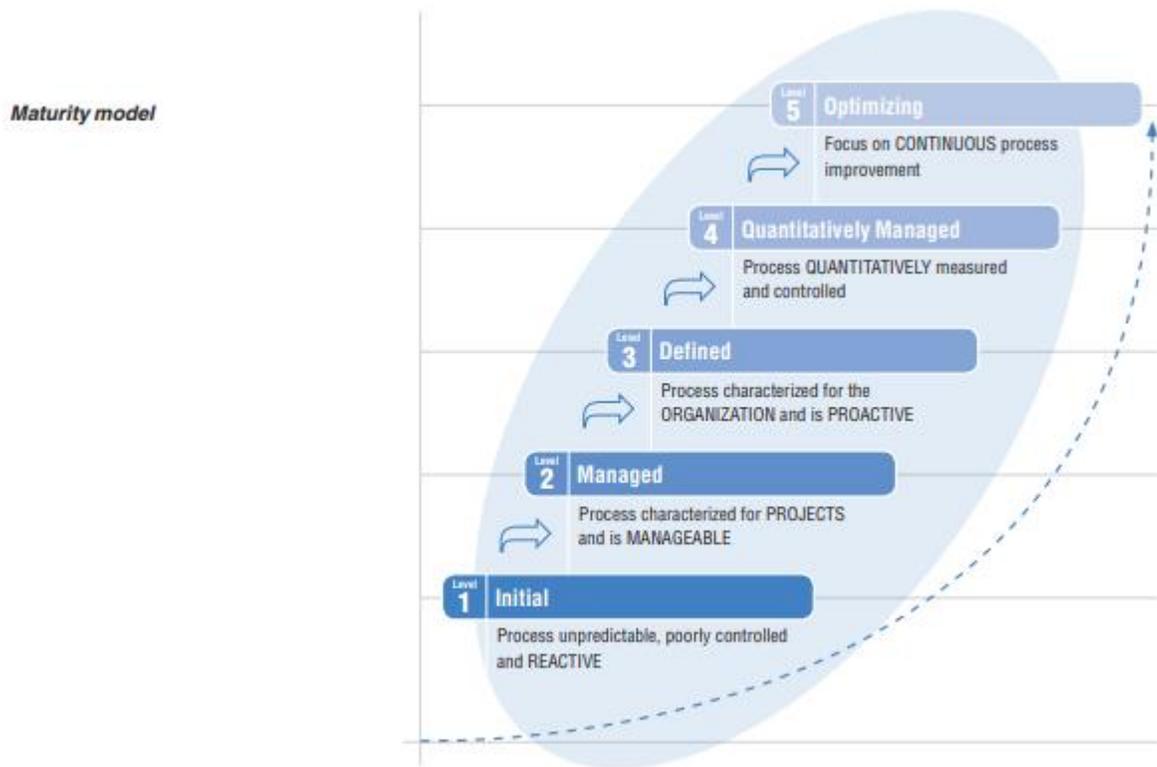


Figura 7. Modelo de Madurez para Gobierno de Datos.

Figura 7. Modelo de Madurez para Gobierno de Datos.

Fuente: IBM. (2007). IBM Data Governance Council Maturity Model. [Figura] (Recuperado de http://www-935.ibm.com/services/uk/cio/pdf/leverage_wp_data_gov_council_maturity_model.pdf)

La importancia de plasmar la madurez de una organización en cuanto al manejo de un gobierno de datos facilitará la implementación de técnicas enfocadas a datos inventariados y con usuarios finales responsables de sus datos. Su propósito no es solo satisfacer los requerimientos de información de la unidad o departamento a cargo, sino, de todos aquellos clientes externos que presentan, en algún grado, la proporción de datos adecuados para incrementar sus resultados y productividad.

Mediante este enfoque se proyecta el uso de técnicas que incrementarán el análisis sobre los datos; se amplían conceptos que permitirán el desarrollo de técnicas como la minería de datos, para un aporte tanto descriptivo como predictivo, con apoyo a la toma de decisiones.

Capítulo 3. Marco Metodológico

3.1 Tipo de Investigación

El tipo de investigación que más se acopla a las necesidades de este proyecto, es el de investigación aplicada, pues, en este caso, el producto final no solo es exponer los conceptos relacionados con *Data Warehouse* y Minería de Datos, sino cómo utilizar el concepto teórico para desarrollar un modelo físico y práctico para utilizar por parte de los mandos gerenciales de la Municipalidad de Esparza.

De igual forma, el producto final no solo contempla un *Data Warehouse* para realizar los procesos de consulta, sino desarrollar, en paralelo, un conjunto de demostraciones que permitan a los mandos gerenciales, contraponer la forma actual y la forma propuesta de desarrollar inteligencia de negocios y toma de decisiones.

Por lo anterior, se propone el proceso de recolección de datos que exponga o resuma la información, de manera cuidadosa, donde puedan analizarse los resultados, a fin de extraer las generalidades significativas que permitan contribuir al conocimiento del caso planteado.

Etapas

1. Examinar las características del problema escogido.
2. Definición y formulación del problema.
3. Selección de temas y fuentes apropiadas.
4. Selección y elaboración de técnicas para la recolección de datos.
5. Clasificación de los datos mediante categorías, que permitan aportar al propósito de la investigación relaciones significativas entre estos.
6. Validar las técnicas administradas para la recolección de datos.
7. Generar las observaciones que permitan demostrar la objetividad y exactitud de la investigación realizada.
8. Analizar los resultados de la investigación a través de los datos obtenidos, mediante un criterio técnico claro y preciso.

3.2 Alcance Investigativo

La Municipalidad cuenta en este momento con diferentes contextos que requieren de análisis desde una perspectiva empresarial, orientada en función de los datos; eso permitirá describir cómo se realizan algunos de sus procesos y los datos generados a raíz de la operatividad. Además, uno de los principales logros es enfocar este proyecto en la identificación, descripción y generación de propuestas de mejora continua; la institución puede obtener especificaciones propias con base en sus características, así como las diferentes mediciones que generan información

relevante para construir modelos de concentradores de datos, donde se evalúa la calidad de los existentes, así como las diferentes estructuras de datos necesarias para los procesos de extracción, transformación y carga de estos.

Resulta sumamente importante señalar que, para obtener cada una de esas estructuras, son necesarias la identificación y recolección de los datos sobre las diferentes unidades operativas de la municipalidad.

3.3 Enfoque

El enfoque mixto de este proyecto permitirá utilizar lo cuantitativo y cualitativo para generar una convergencia tal, que aportará los insumos necesarios para su desarrollo.

La cuantificación resumida como el conteo de los datos, permitirá a la administración municipal determinar los individuos que cumplen o vislumbran alguna condición con base en una característica especial; se logra la aplicación de algún escenario ideal para la organización en materia del mejoramiento en la atención de clientes, bajo una perspectiva de análisis empresarial.

Posteriormente, la capacidad de deducir algunas situaciones en el ámbito municipal, con base en las generalidades del comportamiento de los datos, permite relacionar las diferentes acciones acordes en el diario actuar municipal. El logro de estas probabilidades refuerza, a través de los datos cuantitativos, el apoyo de todos los modelos de minería de datos resultantes de este proyecto.

3.4 Diseño

El proyecto pretende realizar un aporte innovador en la conceptualización de la administración, control y seguridad de los datos. En busca de este fin, se propone evaluar el grado de madurez en la gobernanza de los datos presente en la organización. Este proceso identifica los conjuntos de datos considerados en la integración, para el proceso de generar la simplificación del modelo de base de datos, del cual serán tomados aquellos que permitirán el apoyo en la toma de decisiones.

Para el logro de un proceso automatizado en la carga de datos, se ha confeccionado el *Staging Área*, con los respectivos procesos de extracción, transformación y carga (E.T.L) de los datos provenientes del modelo simplificado, una vez contenidos dichos datos con los formatos idóneos y el tratamiento necesario para lograr su homologación.

Posteriormente, se realiza la construcción de una estructura trascendente en un aspecto primordial para lograr el modelo de un *Data Warehouse* que concentra los datos, lo cual permite convertirse en el contenedor principal de datos universal.

Mediante los datos integrados se satisfacen las necesidades de las diferentes unidades de la organización, al confeccionar los múltiples *Data Marts* que permitirán la satisfacción de cada una de ellas, a las cuales se pretende apoyar en materia de la toma de decisiones.

3.5 Población y Muestreo

La población de los datos se centra en áreas específicas dentro de la administración municipal, principalmente, el personal administrativo correspondiente a la Municipalidad de Esparza, concretamente a las unidades funcionales generadoras de datos como insumos necesarios para este proyecto.

El elemento de integridad proviene de algunos contextos desarrollados en el ámbito de la minería de datos, con la organización, cuando se considere necesario. En resumen, dentro del área de la administración municipal se encuentran las diferentes unidades funcionales de la organización; se detalla su población:

Tabla 2

Detalle de la población para el desarrollo del proyecto

Área Funcional	Cantidad Empleados
Cobros	3
Bienes Inmuebles	3
Catastro	1
Proveeduría	2
Alcaldía	2
Plataforma Municipal	2
Administración Financiera	5
Gestión Desarrollo Humano	2
Gestión Vial Municipal	3

Nota. Recuperado del Manual de Recursos Humanos de la Municipalidad de Esparza.

En cuanto a la muestra, es probabilística, con el objetivo de demostrar algunas situaciones desarrolladas en la labor cotidiana de la municipalidad y el ámbito cantonal.

3.6 Instrumentos de Recolección de Datos

3.6.1 Estudios de correlación

En la construcción de los modelos de minería de datos se establecen los estudios necesarios para identificar la correlación entre las variables que conformarán los modelos de regresión logística, resultantes de la investigación en los diferentes contextos que resulten de interés prioritario para la atención de la parte administrativa de la municipalidad; esto con el objetivo de identificar el grado en que las variaciones sufridas por un factor, corresponden con las que experimenta el otro.

Por lo expuesto con anterioridad, la correlación depende de la medida en que los valores de dos variables aumenten o disminuyan en la misma o diferente dirección.

Cuando los valores de las variables aumentan o disminuyen de la misma manera, se presenta una correlación positiva; si por el contrario, los valores de una variable aumentan en tanto que disminuyen los de la otra, se trata de una correlación negativa; también podría darse muy poca o ninguna correlación.

3.6.2 Estudios de desarrollo

Los estudios de desarrollo permiten determinar no solo las interrelaciones y el estado en que se encuentran algunos eventos, sino los cambios que puedan presentarse en un lapso; describen el desarrollo que experimentan las variables.

3.6.3 El Cuestionario

La utilización de cuestionarios busca la obtención de datos claves, provenientes de los individuos seleccionados en la muestra considerada para el desarrollo del proyecto; la variedad y selección de los diferentes aspectos busca medir el nivel de conocimiento de la administración municipal, en el sentido de temas como gobernanza de los datos, calidad, integridad, seguridad y administración de los datos.

Para tal fin, se aplica el siguiente instrumento considerado como el anexo 1.

3.6.4 La Entrevista

Para el presente proyecto se utilizará la entrevista, con el objetivo de recabar datos de una manera verbal, de cada uno de los individuos considerados en la muestra; esto por medio de preguntas que logren resaltar aspectos de importancia para el proyecto. Dicha técnica será aplicada de manera individual o grupal.

Para tal fin se aplica el instrumento considerado como anexo 2.

3.7 Técnicas de Análisis de la Información

Las técnicas utilizadas para el proyecto se enfocan en aspectos que puedan detallar cómo se realizan algunos procesos; la aplicación de la mejora continua en vista de que los requerimientos de los clientes de la organización son cambiantes en los diferentes contextos generados en el desempeño de las funciones diarias y la afectación de las diferentes tendencias tecnológicas, podrían también incidir o generar un efecto de cambio en la municipalidad.

Para medir y dejar plasmado un análisis de información que permita ser evaluado por la administración municipal, se proponen las siguientes técnicas para este proyecto:

1. Diagramas UML.
2. Ciclo de Deming.
3. Diagrama de Ishikawa (Causa - Efecto).
4. Mapas Conceptuales.
5. Herramientas Estadísticas (Gráficos de Pareto, Gráficos de Dispersión).
6. Técnicas para Analítica Empresarial.
7. Técnicas de Minería de Datos y BPM.

3.8 Estrategia de Desarrollo de la Propuesta

La propuesta cuenta con el interés de generar estrategias para aportar un panorama que contribuya a la toma de decisiones y genere un ambiente propicio para la mejora continua.

Para esto, se plantea utilizar técnicas OLAP para la construcción de contenedores de datos integradores de los múltiples recursos de estos, tratados mediante procesos de ETL que permitan su homologación.

Una vez integrados estos datos, se procederá a generar recursos que permitan apoyar a los usuarios finales en la forma como pueden interpretar y generar sus propios aportes para la toma de decisiones.

Con los datos recolectados se procederá a realizar una evaluación de las diferentes variables identificadas en aquellos prioritarios para la organización y se propiciará la construcción de modelos de técnicas de minería de datos que contribuyan a la analítica empresarial; además, generar propuestas de escenarios para vislumbrar las diferentes acciones que podría efectuar la Municipalidad de Esparza en sus diferentes actividades, financieras, operativas, sociales o culturales.

Capítulo 4. Análisis del Diagnóstico

En lo referente al análisis de diagnóstico, se aplican los diferentes instrumentos considerados para el desarrollo del proyecto planteado; eso, con el objetivo de plasmar la necesidad de contar con un gobierno local que se empodere en el manejo de una gobernanza de los datos más sólida, mediante mecanismos tecnológicos que faciliten el control de estos.

Para conocer el criterio de los diferentes usuarios finales en el sentido de cómo se integran, controlan y protegen los datos de la organización, se aplica el siguiente instrumento de validación.

El resultado de la aplicación del siguiente cuestionario se resume de esta manera:

Registro de los cuestionarios
 Detalle de los valores

Excelente	1
Muy bueno	2
Bueno	3
Regular	4
Malo	5
Sí	6
No	7
No responde	0

Número	Departamento	Pregunta 1	Pregunta 2	Pregunta 3	Pregunta 4	Pregunta 5	Pregunta 5.1
1	Proveeduría	1	2	2	2	7	1
2	Cobros	4	4	4	3	7	1
3	Bienes e Inmuebles	2	2	2	1	7	2
4	Secretaría Consejo	2	2	4	2	7	0
5	Secretaría Consejo	2	2	2	2	7	0
6	Gestión Jurídica	3	3	3	2	6	2
7	Alcaldía	2	1	2	1	6	1
8	Gestión de Ordenamiento Territorial	3	4	4	1	7	1
9	Financiera Tributaria	2	3	3	1	6	0
10	Financiera	3	2	3	2	7	0
11	Financiera	2	2	3	2	7	0
12	Financiera	4	5	4	5	7	0
13	Auditoría Interna	3	4	5	1	7	0
14	Auditoría Interna	3	3	2	2	6	2
15	Gestión Desarrollo Humano	1	4	4	1	7	2
16	Patentes	2	3	2	1	6	1
17	Presupuesto	3	3	3	2	7	1
18	Planificación	4	3	5	1	6	1
19	Gestión de Ordenamiento Territorial	2	4	2	1	7	2
20	Valoración	3	4	4	1	7	1
21	Informática	2	3	3	4	6	2
22	Proveeduría	4	4	4	2	7	2
23	Proyectos	5	5	5	1	6	1
24	Catastro	2	2	2	1	7	1
25	Cobros	2	2	2	1	7	1
26	Cobros	2	2	2	1	7	1
27	Recursos Humanos	2	2	1	1	6	1
28	Gestión Admin.	3	3	4	4	6	1
29	Financiera	3	3	3	2	7	1
30	Valoración	2	3	2	2	7	2

Figura 8. Cuadro resumen encuesta Gobernanza de datos.

Figura 8. Cuadro resumen encuesta Gobernanza de datos.

Fuente: Desarrollo propio (2018)

Registro de los cuestionarios
 Detalle de los valores

Excelente	1
Muy bueno	2
Bueno	3
Regular	4
Malo	5
Sí	6
No	7
No responde	0

Número	Departamento	Pregunta 6	Pregunta 7	Pregunta 7.1	Pregunta 8	Pregunta 9
1	Proveeduría	1	7	1	1	1
2	Cobros	1	7	3	2	2
3	Bienes e Inmuebles	1	7	2	2	2
4	Secretaría Consejo	1	0	0	1	1
5	Secretaría Consejo	1	7	0	1	1
6	Gestión Jurídica	2	7	2	2	2
7	Alcaldía	1	6	1	1	1
8	Gestión de Ordenamiento Territorial	1	6	1	1	2
9	Financiera Tributaria	1	7	1	1	1
10	Financiera	2	7	0	2	2
11	Financiera	2	7	0	3	2
12	Financiera	1	7	0	1	1
13	Auditoría Interna	1	6	1	1	1
14	Auditoría Interna	2	6	2	2	2
15	Gestión Desarrollo Humano	1	7	1	1	1
16	Patentes	1	7	0	1	1
17	Presupuesto	1	7	1	1	1
18	Planificación	1	6	1	1	1
19	Gestión de Ordenamiento Territorial	1	7	2	2	2
20	Valoración	1	7	0	1	2
21	Informática	2	6	2	2	2
22	Proveeduría	2	7	2	2	2
23	Proyectos	1	6	1	1	1
24	Catastro	1	7	1	1	1
25	Cobros	1	7	1	2	1
26	Cobros	1	7	1	1	1
27	Recursos Humanos	1	6	1	1	1
28	Gestión Admin.	1	6	1	1	2
29	Financiera	1	7	1	1	1
30	Valoración	1	7	2	2	1

Figura 9. Cuadro resumen encuesta Gobernanza de datos.

Figura 9. Cuadro resumen encuesta Gobernanza de datos.

Fuente: Desarrollo propio (2018)

En cuanto al resultado, se pueden determinar algunos aspectos, los cuales dejan en claro que la administración y control de los datos no consideran todos los aspectos recomendados en una gobernanza activa de los datos; posteriormente, que la gran mayoría de las unidades pertenecientes a la municipalidad, consideran que existe ausencia en el tratamiento, administración y control de los datos.

Adicionalmente, se confecciona el diagrama de espina de pescado para resaltar problemas encontrados durante el análisis de la condición de la municipalidad, en materia de gobernanza de los datos.

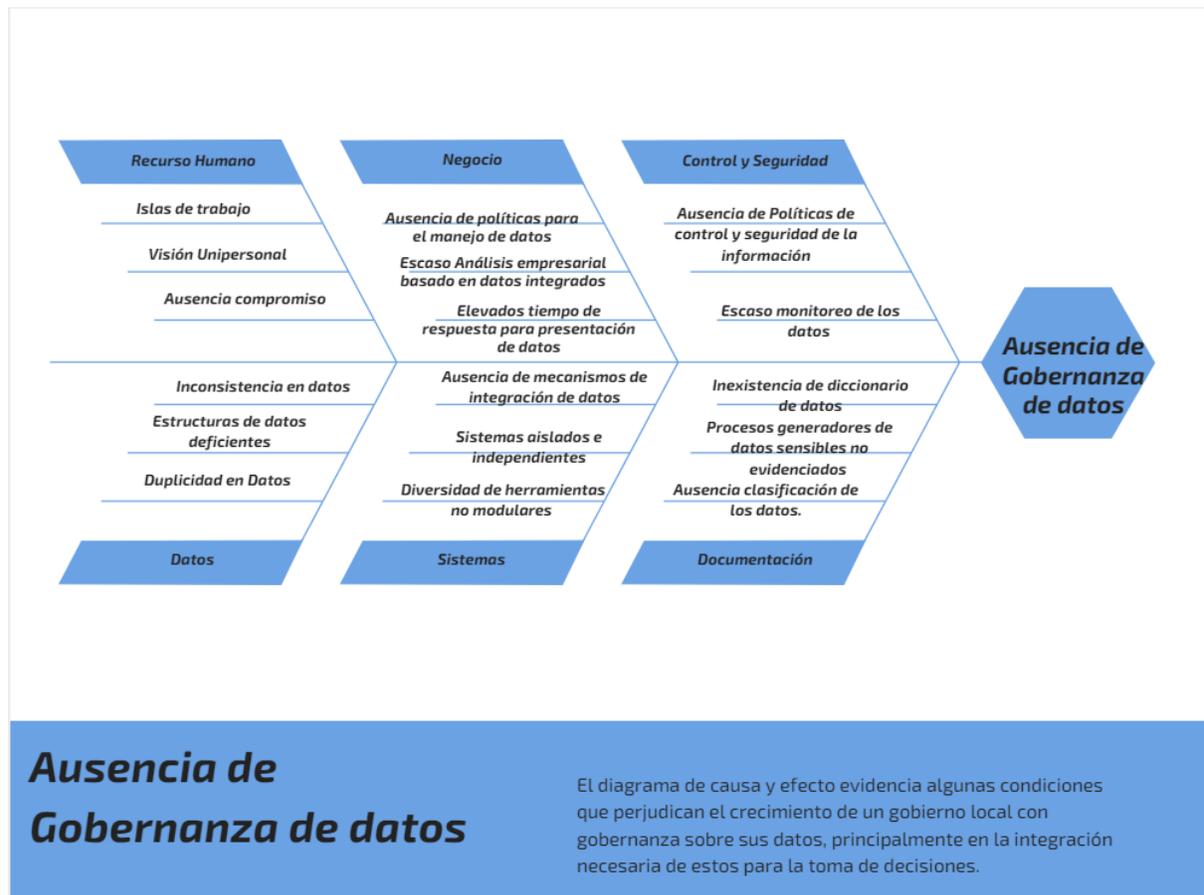


Figura 10. Ausencia de Gobernanza de Datos

Figura 10. Ausencia de Gobernanza de Datos.

Fuente: Desarrollo propio, (2018)

Capítulo 5. Propuesta de Solución

La propuesta planteada proporciona una solución oportuna para la Municipalidad de Esparza, fundamentada en aspectos claves que combinan, mediante mecanismos tecnológicos, la administración, control e integración de los datos.

Las bases de datos digitales de la Municipalidad de Esparza presentan una longevidad de alrededor de treinta años; el primer reto que asume esta propuesta consiste en proporcionar los insumos necesarios para la administración de los datos, con base en la regla de negocio.

El régimen municipal con el pasar de los años, ha recibido el apoyo de diferentes instituciones; dentro de estas, el Instituto de Fomento y Asesoría Municipal. Lastimosamente, muchas de las propuestas a nivel del desarrollo de un sistema integral para el manejo de ingresos y egresos, han fracasado y generaron un retraso en la evolución de los sistemas transaccionales, así como en la tecnología utilizada para la administración de los datos. Durante el año 2017, la Municipalidad de Esparza se propone destinar los recursos necesarios para la adquisición de un sistema integral para el manejo de toda la regla de negocio del ámbito municipal.

Generación de Insumos

Dentro de los retos que encuentra la propuesta está generar los insumos necesarios para la migración de los datos provenientes de estructuras no relacionales, las cuales pertenecían a los sistemas actuales que la institución tenía para su uso diario.

En este sentido, se crean los scripts necesarios para la carga en plantillas de los datos que servirían como recurso para el traslado al nuevo sistema administrador de base de datos; el proceso de implementación considera *Oracle Express* como primer repositorio de datos, debido a que fue el motor ofrecido por la empresa adjudicada en la compra del sistema, argumentando que esto facilitaría el proceso de implementación, pero el requerimiento municipal y de esta propuesta propone el uso de software open source; por lo tanto, se selecciona PostgreSQL en su versión 11.

Calidad, aseguramiento e integridad en los datos

El hecho de contar con recursos que tienen procedencia de un sistema de datos, con un esquema donde no existían elementos tan importantes como llaves primarias, llaves foráneas y relación entre los datos, obligó a la construcción de herramientas de migración que evaluaran el patrón de dichos datos con base en los requerimientos del negocio. La manipulación de los datos mediante una interfaz gráfica que no consideraba elementos de validación, generaba un tratamiento inadecuado de estos; por lo tanto, para asegurar la calidad y completitud de los

datos, se establecieron reglas o modelos heurísticos que permitieran su integridad y calidad.

Definición del ámbito del proyecto

El sistema integral actual adquirido por la municipalidad cuenta con un modelo relacional que presenta una estructura de alrededor de mil seiscientas cincuenta y ocho tablas, enfocadas en satisfacer las necesidades de almacenamiento y administración de los datos perteneciente a múltiples contextos. En el caso de la propuesta, considera dos enfoques importantes para el apoyo en la toma de decisiones de la institución; en este sentido, se enfoca en los ingresos y su comportamiento, además del crecimiento de trámites con un impacto significativo para la generación de esos ingresos.

Para satisfacer esta necesidad, se consideran 296 tablas que logran proporcionar los datos necesarios para la generación de una solución óptima para el desarrollo de este proyecto. La nomenclatura integrada en las tablas corresponde al auxiliar CUF (Cobro Unificado de Finca) posibilita la obtención de los recursos necesarios a nivel de datos, con un enfoque de ingresos y trazabilidad de documentos, para medir aspectos como la generación de nuevos ingresos y el crecimiento de diversos sectores, dentro de los que destacan urbanismo, negocio, arrendamientos, entre otros.

Implementación del entorno de pruebas

Buscando simular el entorno final que habrá de tener el *Data Warehouse*, conforme con la infraestructura de la Municipalidad de Esparza, se implementa una máquina virtual con las siguientes características.

1. Debian 9.5 (stretch) de 64 bits, en modo consola.
2. 16 gigabytes de memoria principal.
3. 38 gigabytes de almacenamiento primario.
4. 5 gigabytes de swap partition y 4 gigabytes de swap file.
5. 2 hilos de acceso al procesador.

Dicha máquina contiene los motores de bases de datos de:

1. Oracle 11g Express Edition en 64 bits.
2. PostgreSQL 10 64bits.

Ambos, configurados para ser accedidos por clientes remotos, pues tanto los clientes de SQL developer para Oracle, como pgAdmin y el pgModeler para PostgreSQL, se encuentran instalados en el sistema anfitrión de Windows 10.

Es importante identificar que el entorno de desarrollo requerido por la municipalidad conlleva un enfoque que apunta a tendencias libres, con aspectos que eviten encarecer el desarrollo del proyecto y generan un equilibrio para su sostenibilidad. Ha sido fundamental para el éxito, establecer la pauta de los alcances, retransmitiendo seguridad a la administración municipal y apoyando el

intercambiando de criterios técnicos con la empresa desarrolladora del sistema integral adoptado. Resultan vitales los procesos de depuración de los datos que aseguran su calidad e integridad; el propósito establece contar con estos en un futuro, para enfocarse en el apoyo a la toma de decisiones, en elementos claves para la administración.

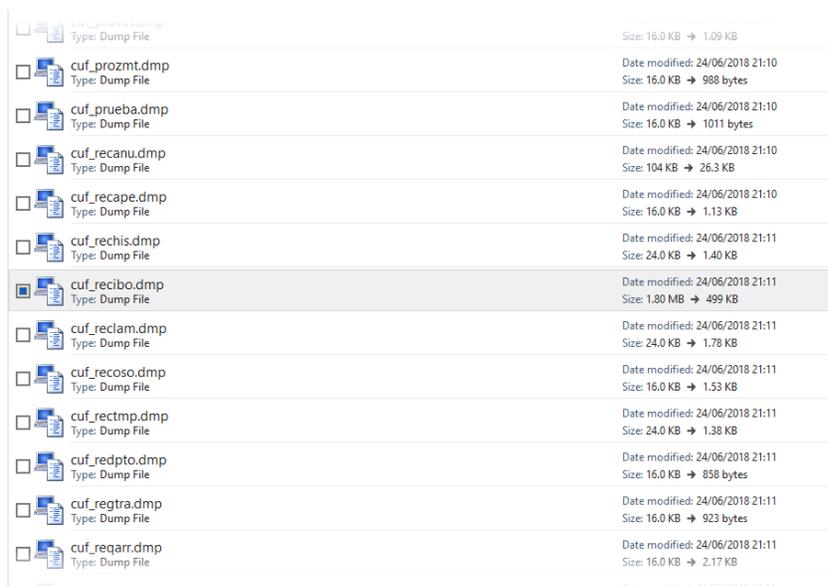
Los datos han seguido una trazabilidad a través de diversos motores de datos como parte de los requerimientos iniciales y finales, que giran en torno del desarrollo del proyecto. En una fase inicial para poner en marcha el proyecto de implementación del sistema se realizó la migración de los datos al sistema administrador de Base de Datos Oracle 11g Express, esto por la infraestructura tecnológica del sistema, contrario a los requerimientos establecidos por la Unidad de Tecnología de Información de la municipalidad, motivo por el cual se acepta utilizar inicialmente el motor de datos antes mencionado, pero posteriormente se realiza la migración de los datos a *PostgreSQL 10*, repositorio final para la elaboración de este proyecto.

Posterior a la implementación de un entorno operativo similar al requerido para la Municipalidad de Esparza, se realizan los siguientes procesos:

Implementar la base de datos Oracle

Como se mencionó anteriormente, el entorno actual del sistema adquirido por la Municipalidad de Esparza se encuentra en estado de implementación y depuración, y los datos están temporalmente alojados en el motor de *Oracle*, mientras dicho proceso termina.

Se desarrolla un proceso de migración de datos del motor local del municipio, al motor local de pruebas, mediante el respaldo de los datos y estructuras para cada una de las 1658 tablas que posee la propuesta adquirida; se crean, para ello, archivos individuales, uno para cada tabla. Este proceso es definido y realizado por la empresa externa. En la siguiente imagen se puede apreciar un extracto del paquete comprimido que contiene los archivos .dmp.



<input type="checkbox"/>	 <i>cu_...</i> Type: Dump File	Size: 16.0 KB → 1.09 KB
<input type="checkbox"/>	 <i>cu_f_prozmt.dmp</i> Type: Dump File	Date modified: 24/06/2018 21:10 Size: 16.0 KB → 988 bytes
<input type="checkbox"/>	 <i>cu_f_prueba.dmp</i> Type: Dump File	Date modified: 24/06/2018 21:10 Size: 16.0 KB → 1011 bytes
<input type="checkbox"/>	 <i>cu_f_recanu.dmp</i> Type: Dump File	Date modified: 24/06/2018 21:10 Size: 104 KB → 26.3 KB
<input type="checkbox"/>	 <i>cu_f_recape.dmp</i> Type: Dump File	Date modified: 24/06/2018 21:10 Size: 16.0 KB → 1.13 KB
<input type="checkbox"/>	 <i>cu_f_rechis.dmp</i> Type: Dump File	Date modified: 24/06/2018 21:11 Size: 24.0 KB → 1.40 KB
<input checked="" type="checkbox"/>	 <i>cu_f_recibo.dmp</i> Type: Dump File	Date modified: 24/06/2018 21:11 Size: 1.80 MB → 499 KB
<input type="checkbox"/>	 <i>cu_f_reclam.dmp</i> Type: Dump File	Date modified: 24/06/2018 21:11 Size: 24.0 KB → 1.78 KB
<input type="checkbox"/>	 <i>cu_f_recoso.dmp</i> Type: Dump File	Date modified: 24/06/2018 21:11 Size: 16.0 KB → 1.53 KB
<input type="checkbox"/>	 <i>cu_f_rectmp.dmp</i> Type: Dump File	Date modified: 24/06/2018 21:11 Size: 24.0 KB → 1.38 KB
<input type="checkbox"/>	 <i>cu_f_redpto.dmp</i> Type: Dump File	Date modified: 24/06/2018 21:11 Size: 16.0 KB → 858 bytes
<input type="checkbox"/>	 <i>cu_f_regtra.dmp</i> Type: Dump File	Date modified: 24/06/2018 21:11 Size: 16.0 KB → 923 bytes
<input type="checkbox"/>	 <i>cu_f_reqarr.dmp</i> Type: Dump File	Date modified: 24/06/2018 21:11 Size: 16.0 KB → 2.17 KB

Figura 11. Migración de datos en archivos

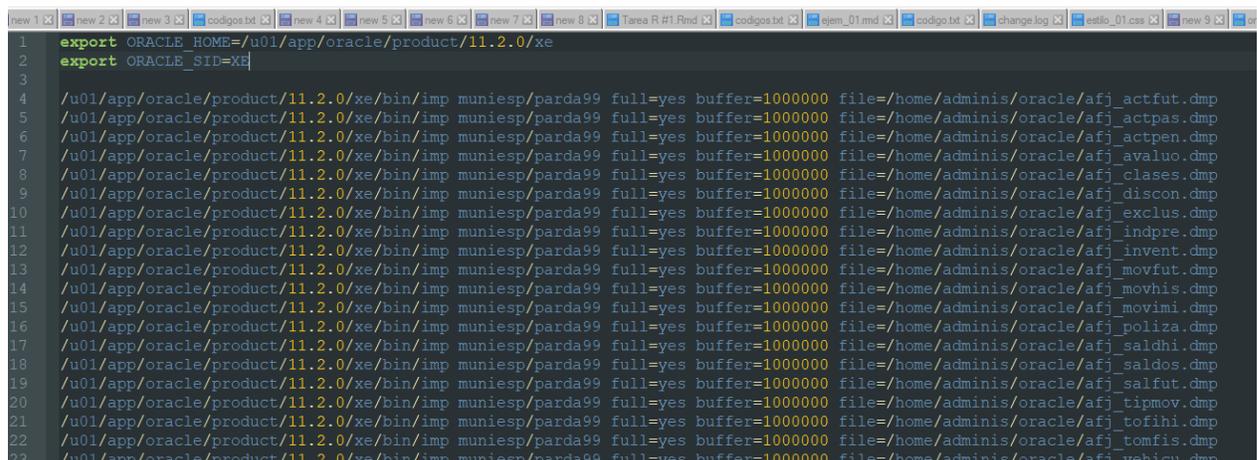
Figura 11. Migración de datos en archivos .dmp.

Fuente: Desarrollo propio (2018)

Posteriormente se debe realizar para la carga de datos, para cada tabla de destino, una instrucción de consola como la siguiente:

```
/u01/app/oracle/product/11.2.0/xe/bin/imp muniesp/parda99 full=yes  
buffer=1000000 file=/home/adminis/oracle/cuf_recibo.dmp
```

Como este proceso se debe efectúa 658 veces, se implementa una solución C#, que lee el directorio con los archivos .dmp y crea un archivo .sh (ejecución de comandos por lotes para el entorno de Linux), el cual luego será transferido al servidor de pruebas en conjunto con los archivos .dmp. La siguiente imagen muestra cómo se observa el contenido de dicho archivo.



```
new 1 x new 2 x new 3 x codigos.txt new 4 x new 5 x new 6 x new 7 x new 8 x Tarea R #1 Pmd x codigos.txt ejem_01.mmd x codigo.txt change.log x estfo_01.css x new 9 x  
1 export ORACLE_HOME=/u01/app/oracle/product/11.2.0/xe  
2 export ORACLE_SID=XE  
3  
4 /u01/app/oracle/product/11.2.0/xe/bin/imp muniesp/parda99 full=yes buffer=1000000 file=/home/adminis/oracle/afj_actfut.dmp  
5 /u01/app/oracle/product/11.2.0/xe/bin/imp muniesp/parda99 full=yes buffer=1000000 file=/home/adminis/oracle/afj_actpas.dmp  
6 /u01/app/oracle/product/11.2.0/xe/bin/imp muniesp/parda99 full=yes buffer=1000000 file=/home/adminis/oracle/afj_actpen.dmp  
7 /u01/app/oracle/product/11.2.0/xe/bin/imp muniesp/parda99 full=yes buffer=1000000 file=/home/adminis/oracle/afj_avaluo.dmp  
8 /u01/app/oracle/product/11.2.0/xe/bin/imp muniesp/parda99 full=yes buffer=1000000 file=/home/adminis/oracle/afj_clases.dmp  
9 /u01/app/oracle/product/11.2.0/xe/bin/imp muniesp/parda99 full=yes buffer=1000000 file=/home/adminis/oracle/afj_discon.dmp  
10 /u01/app/oracle/product/11.2.0/xe/bin/imp muniesp/parda99 full=yes buffer=1000000 file=/home/adminis/oracle/afj_exclus.dmp  
11 /u01/app/oracle/product/11.2.0/xe/bin/imp muniesp/parda99 full=yes buffer=1000000 file=/home/adminis/oracle/afj_indpre.dmp  
12 /u01/app/oracle/product/11.2.0/xe/bin/imp muniesp/parda99 full=yes buffer=1000000 file=/home/adminis/oracle/afj_invent.dmp  
13 /u01/app/oracle/product/11.2.0/xe/bin/imp muniesp/parda99 full=yes buffer=1000000 file=/home/adminis/oracle/afj_movfut.dmp  
14 /u01/app/oracle/product/11.2.0/xe/bin/imp muniesp/parda99 full=yes buffer=1000000 file=/home/adminis/oracle/afj_movhis.dmp  
15 /u01/app/oracle/product/11.2.0/xe/bin/imp muniesp/parda99 full=yes buffer=1000000 file=/home/adminis/oracle/afj_movimi.dmp  
16 /u01/app/oracle/product/11.2.0/xe/bin/imp muniesp/parda99 full=yes buffer=1000000 file=/home/adminis/oracle/afj_poliza.dmp  
17 /u01/app/oracle/product/11.2.0/xe/bin/imp muniesp/parda99 full=yes buffer=1000000 file=/home/adminis/oracle/afj_saldhi.dmp  
18 /u01/app/oracle/product/11.2.0/xe/bin/imp muniesp/parda99 full=yes buffer=1000000 file=/home/adminis/oracle/afj_saldos.dmp  
19 /u01/app/oracle/product/11.2.0/xe/bin/imp muniesp/parda99 full=yes buffer=1000000 file=/home/adminis/oracle/afj_salfut.dmp  
20 /u01/app/oracle/product/11.2.0/xe/bin/imp muniesp/parda99 full=yes buffer=1000000 file=/home/adminis/oracle/afj_tipmov.dmp  
21 /u01/app/oracle/product/11.2.0/xe/bin/imp muniesp/parda99 full=yes buffer=1000000 file=/home/adminis/oracle/afj_tofihi.dmp  
22 /u01/app/oracle/product/11.2.0/xe/bin/imp muniesp/parda99 full=yes buffer=1000000 file=/home/adminis/oracle/afj_tomfis.dmp  
23 /u01/app/oracle/product/11.2.0/xe/bin/imp muniesp/parda99 full=yes buffer=1000000 file=/home/adminis/oracle/afj_vehicu.dmp
```

Figura 12. Análisis de archivos para migración

Figura 12. Análisis de archivos .dmp para migración

Fuente: Desarrollo propio (2018)

Básicamente, este proceso se realiza no solo con el objetivo de tener una copia de los datos a lo externo del proceso de depuración e implementación, sino para, a posteriori, realizar análisis de patrones, para validar que los datos se han migrado adecuadamente y que inconsistencias halladas en dicho proceso de depuración han sido corregidas, sin necesidad de interferir en el rendimiento del entorno de implementación del nuevo sistema.

Implementar la base de datos en *PostgreSQL*

En forma similar al paso anterior, se desarrolla un utilitario en C# que permite leer de la base de datos en *Oracle*, tanto los datos como la estructura, de forma que:

1. Adecua la estructura obtenida de *Oracle* al formato correspondiente en *PostgreSQL*.
2. Formatea los datos originales tratando de eliminar:
 1. Problemas en la conversión de caracteres con acentuación (tildes, eñes).
 2. Formatos de fecha.
 3. Equiparación con valores numéricos.

El utilitario realiza varios pasos:

1. Genera el script para la creación de tablas.
2. Genera un segundo script para la implementación de las llaves primarias.
3. Genera un tercer script para la implementación de las llaves foráneas.
4. Realiza la migración de las tablas requeridas, registrando en los archivos *logs*, tanto información general del proceso de migración, como los errores encontrados, con indicación de cuales datos no pudieron ser procesados y el error identificado por el motor, por el cual fracasa la migración específica de ese dato.
5. Actualiza los datos de fecha de nacimiento de los propietarios, a partir del padrón nacional del Tribunal Supremo de Elecciones.

A continuación, se muestra el código fuente utilizado para realizar la migración de datos entre los motores previamente descritos.

```
void migradatos(){
    StreamWriter sw = new
StreamWriter(@"D:\Documentos\Capacitaciones\Cenfotec\Maestría\TFG\postgre\migracion_8.log");

    StreamWriter re = new
StreamWriter(@"D:\Documentos\Capacitaciones\Cenfotec\Maestría\TFG\postgre\resumen_8.log");

    //Auto instancia
    Program prg = new Program();

    //Objetos para la conexión con Postgre
    NpgsqlConnection conexP;
    NpgsqlCommand comandoP;

    //Cadena de conexión con Postgre
    String auxConexionP = "Host=10.236.2.100; " +
```

```

        "Port=5432; " +
        "Database=muniesp; " +
        "User Id=?????"; " +
        "Password = ??????";

//Objetos para la conexión con Oracle
OracleConnection conex;
OracleCommand comando;
OracleDataReader lector;

//Cadena de conexión con Oracle
String auxConexion = "Data Source = 10.236.2.100:1521; User Id = ?????? ; Password = ?????? ";

//Colección con el nombre de las tablas
ArrayList lista = new ArrayList();

//Declaración de una expresiones regulares o patrones de validación
Regex entero = new Regex(@"^[0-9]+$");
Regex flotante = new Regex(@"^(\d|-)?(\d|,)*\.\d*$");
try{
    //intenta conectar con el servidor Postgre
    conexP = new NpgsqlConnection(auxConexionP);
    conexP.Open();
    Console.WriteLine("Conexión satisfactoria con PostgreSQL");

    //intancia el objeto de comandos para Postgre
    comandoP = new NpgsqlCommand();
    comandoP.Connection = conexP;

    //intenta conectar con el servidor oracle
    conex = new OracleConnection(auxConexion);
    conex.Open();
    Console.WriteLine("Conexión satisfactoria con Oracle");

    //intenta ejecutar instrucción contra el servidor
    comando = new OracleCommand();
    comando.Connection = conex;

```

```

comando.CommandType = CommandType.Text;
comando.CommandText = "select table_name from user_tables where substr(table_name,1,3) = 'CUF' order
by table_name";

//Obtiene la lista de tablas de interés desde Oracle
lector = comando.ExecuteReader();

if (lector.HasRows)
{
    lista.Add("COM_PERSON");

    //Carga los nombres de las tablas
    while (lector.Read())
    {
        lista.Add(lector["table_name"]);
    }
}

//Se cierra la consulta anterior
lector.Close();

//Se verifica que la consulta regresara valores
if (lista.Count > 0)
{
    string salida = "";
    int conta = 0, tuplas = 0;
    for (int i = 0; i < lista.Count; i++)
    {
        comando.CommandText = "select * from " + lista[i].ToString() + "";

        //Obtiene la lista de tablas de interés desde Oracle
        lector = comando.ExecuteReader();
        if (lector.HasRows)
        {
            tuplas = 0;
            conta = 0;
            while (lector.Read())
            {

```

```

tuplas++;
salida = "insert into " + lista[i].ToString().ToLower() + " values(";
for (int j = 0; j < lector.FieldCount; j++)
{
    if (string.IsNullOrEmpty(lector[j].ToString()))
    {
        salida = salida + "null";
    }
    else if (flotan.IsMatch(lector[j].ToString()))
    {
        salida = salida + lector[j].ToString();
    }
    else if (entero.IsMatch(lector[j].ToString()))
    {
        salida = salida + lector[j].ToString();
    }
    else
    {
        salida = salida + "'" + prg.fechita(lector[j].ToString()) + "'";
        //salida = salida + "'" + lector[j].ToString() + "'";
    }

    if ((j + 1) != lector.FieldCount)
        salida = salida + ",";
}
comandoP.CommandType = CommandType.Text;
comandoP.CommandText = salida + ")";
try
{
    comandoP.ExecuteNonQuery();
    conta++;
}
catch(Exception ex)
{
    sw.WriteLine(comandoP.CommandText + " -- ERROR... " + ex.Message.ToString());
}
}

```

```

        Console.WriteLine("Procesando tabla [" + lista[i].ToString() + "]);
        Console.WriteLine("Registros leídos....: " + tuplas);
        Console.WriteLine("Registros migrados...: " + conta + "\n");

        re.WriteLine("Procesando tabla [" + lista[i].ToString() + "]);
        re.WriteLine("Registros leídos....: " + tuplas);
        re.WriteLine("Registros migrados...: " + conta);
        re.WriteLine("-----");

        // Console.ReadKey();
    }
    //Se cierra la consulta anterior
    lector.Close();
}

}
sw.Close();
re.Close();
}
catch (OracleException ex)
{
    Console.WriteLine(ex.Message);
}
Console.WriteLine("Script para migración de datos ejecutado satisfactoriamente");
Console.ReadKey();
}

```

Figura 13 .Script de migración para los archivos .dmp

Figura 13. Script de migración para los archivos .dmp.

Fuente: Desarrollo propio (2018)

Aunque este utilitario puede migrar toda la base de datos, fue ajustado solo para migrar las tablas correspondientes al cobro unificado de fincas, correspondiente a 296 tablas.

Base de datos operacional

La base de datos operacional utilizada para el desarrollo de este proyecto contiene alrededor de unas mil seiscientas tablas en total; como se ha señalado en algunos segmentos de este proyecto, al ser un sistema integral, dicho modelo satisface las necesidades de almacenamiento de los datos provenientes de dos conceptos claros, como son los ingresos y egresos municipales. La estructura de datos tiene un enfoque multicompañía, por lo cual se considera dicho aspecto; la estructura utilizada se resume en todo lo relacionado con el cobro unificado de finca por sus siglas CUF, que proporcionan los datos referentes a la regla de negocio encargada de la generación, control y administración de ingresos con base en las propiedades. Para este proyecto se consideran solamente 296 tablas para brindar una propuesta con enfoque en este sentido.

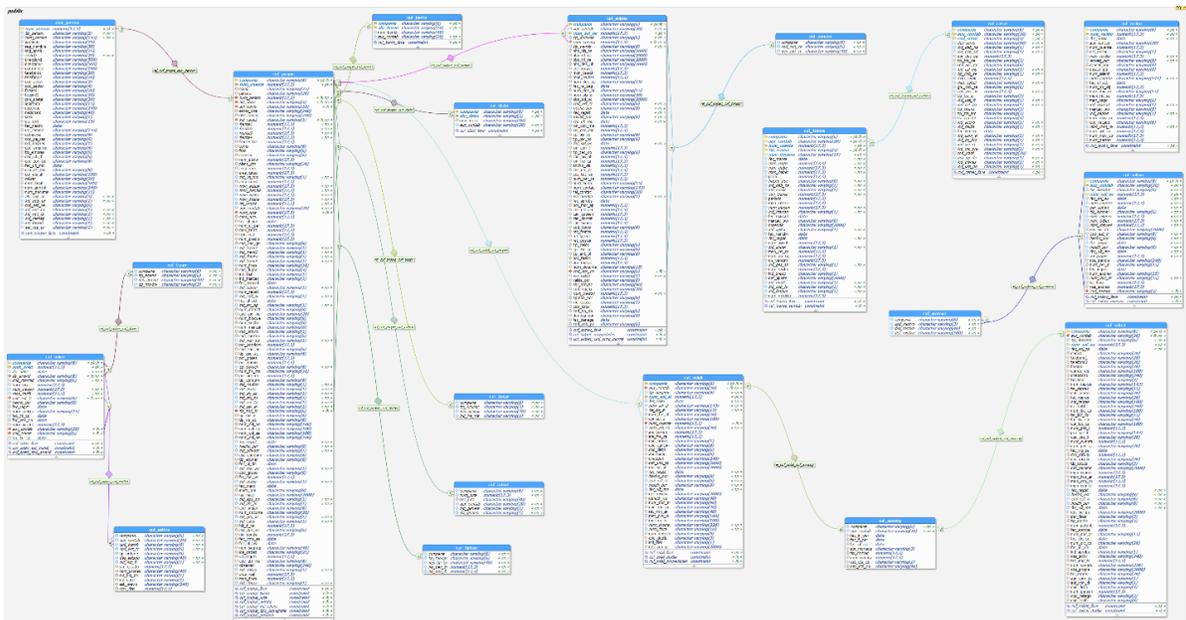


Figura 14. Segmento de Base de Datos Operacional

Figura 14. Segmento de Base de Datos Operacional

Fuente: Decsa. (2017). Modelo Base Datos Operacional Sistema Decsis. [Figura] Recuperado de Diccionario de datos Sistema Decsis.pdf

Staging Area

La propuesta planteada sobre la base datos operacional plasma la propuesta del Staging Area, que posibilita la carga de los datos importantes para negocio en el contexto que permite satisfacer el proyecto. Estas tablas presentan los insumos suficientes que detallan la relación, dependencia y características ideales para obtener las asociaciones necesarias para construir posteriormente el Data Warehouse que podrá cargarse.

public			
<ul style="list-style-type: none"> num_cuenta numeric(17,3) num_person numeric(17,3) aux_contab character varying(20) abr_distri character varying(3) abr_barrio character varying(10) senas_lote character varying(250) area_regis numeric(17,3) area_catas numeric(17,3) num_derech character varying(8) por_posesi numeric(17,3) por_domini numeric(17,3) tip_derech character varying(6) num_fin_ma character varying(10) tip_condom character varying(6) cod_zon_ho character varying(20) num_bloque character varying(8) ind_cordon character varying(1) ind_acera character varying(1) ind_alk_sa character varying(1) ind_alk_pu character varying(1) ind_ser_te character varying(1) ind_ser_el character varying(1) uso_condom character varying(6) area_const numeric(17,3) num_finca character varying(10) 	<ul style="list-style-type: none"> aux_contab character varying(20) cod_servic character varying(3) des_servic character varying(50) ind_deb_hab character varying(1) 	<ul style="list-style-type: none"> abr_distri character varying(3) nom_distri character varying(40) 	<ul style="list-style-type: none"> abr_barrio character varying(10) nom_barrio character varying(40)
<ul style="list-style-type: none"> aux_contab character varying(20) num_cuenta numeric(17,3) num_docume character varying(15) mon_origin numeric(17,3) tip_transa character varying(3) periodo numeric(17,3) num_person numeric(17,3) num_recibo numeric(17,3) 	<ul style="list-style-type: none"> num_person numeric(17,3) cedula character varying(20) nom_person character varying(143) apellidos character varying(25) seg_nombre character varying(30) seg_apelli character varying(25) tip_person character varying(2) sexo character varying(1) est_civil character varying(15) fec_nacim date 	<ul style="list-style-type: none"> num_recibo numeric(17,3) fec_envio date num_cuenta numeric(17,3) est_recibo character varying(1) mon_recibo numeric(17,3) enviad_por character varying(8) ind_deposi character varying(1) num_asient numeric(17,3) num_docume character varying(15) tip_transa character varying(3) periodo numeric(17,3) num_person numeric(17,3) 	
<ul style="list-style-type: none"> num_sol_al numeric(17,3) aux_contab character varying(20) fec_solici date num_cuenta numeric(17,3) cod_distri character varying(6) abr_barrio character varying(6) 	<ul style="list-style-type: none"> num_solici numeric(17,3) aux_contab character varying(20) fec_solici date num_person numeric(17,3) 	<ul style="list-style-type: none"> num_sol_ex numeric(17,3) aux_contab character varying(20) fec_sol_ex date num_cuenta numeric(17,3) 	
<ul style="list-style-type: none"> num_sol_pa numeric(17,3) aux_contab character varying(20) fec_sol_pa date num_cuenta numeric(17,3) num_person numeric(17,3) cod_distri character varying(3) 	<ul style="list-style-type: none"> num_solicitud numeric(17,3) tipo_documento character varying(15) fec_solicitud date num_cuenta numeric(17,3) num_persona numeric(17,3) abr_distrito character varying(6) abr_barrio character varying(10) 	<ul style="list-style-type: none"> num_sol_pe numeric(17,3) aux_contab character varying(20) fec_sol_pe date num_cuenta numeric(17,3) num_person numeric(17,3) 	

Figura 15 Propuesta de Staging Area del Proyecto

Figura 15. Propuesta de Staging Area del Proyecto

Fuente: Desarrollo propio (2018)

Una vez completada las diferentes propuestas del diseño correspondiente tanto a la base de datos propuesta como el Staging Area, se hace el desarrollo de los procesos E.T.L factores determinantes en la extracción, transformación y carga de los datos; sobre este proceso se logra proporcionar el formato adecuado a los datos que se procesan, así como su carga. La siguiente figura permite visualizar, de una forma integral, todos los procesos considerados para la carga de datos.

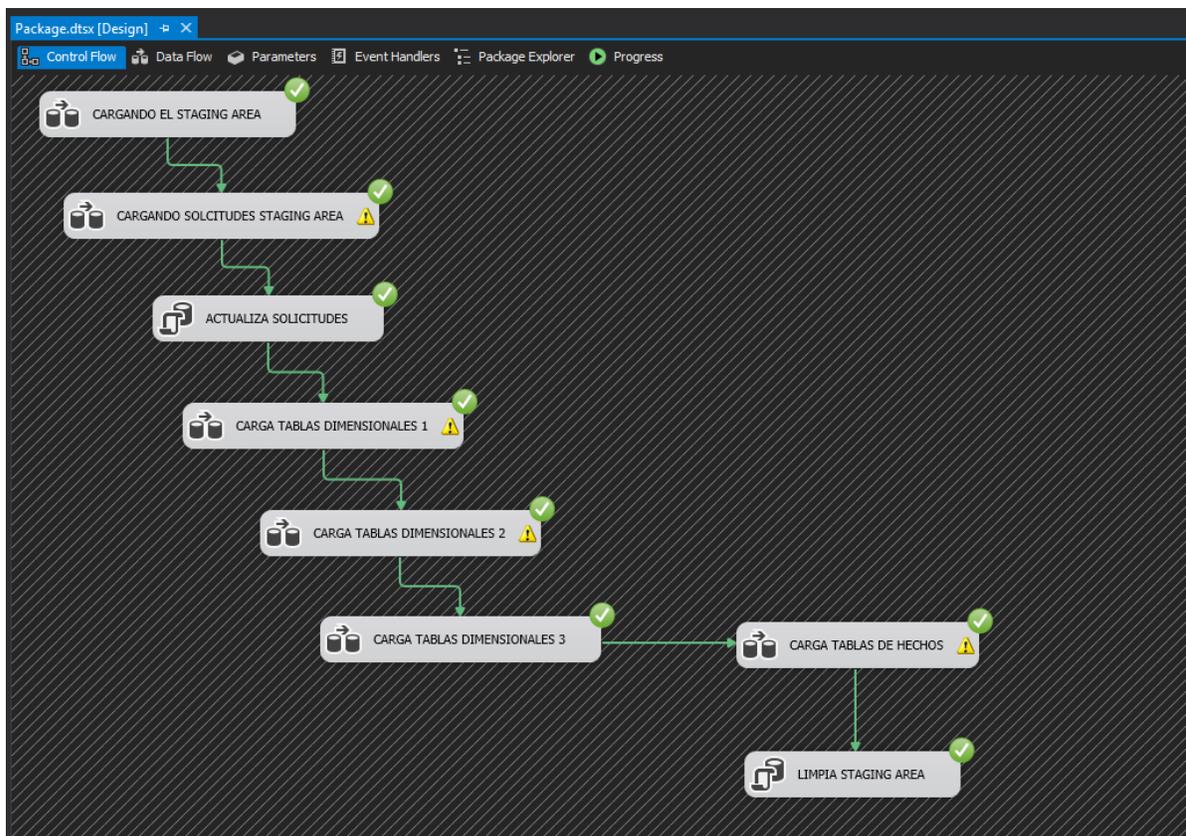


Figura 16 Vista general E.T.L.

Figura 16. Vista general E.T.L.

Fuente: Desarrollo propio (2018)

La primera carga corresponde al conjunto de datos correspondiente al sector de ingresos de la Municipalidad de Esparza, donde radican todas las transacciones de los diferentes servicios brindados por la institución; posteriormente, se realiza la carga de los datos correspondiente a todas las solicitudes de trámites municipales, donde destacan nuevas patentes, permisos de construcción, declaración de bienes inmuebles, entre otros.

Nótese que en la figura aparecen indicadores de advertencia, los cuales consisten en la interpretación realizada por el SQL Data Tools, con respecto a los datos con un tipo varchar que se concatenan en las consultas; estos son interpretados por esta aplicación como campos de tipo texto, que, por su naturaleza, son de mayor capacidad. Se realiza la advertencia de que el dato original o sea, el concatenado, podría registrarse en forma truncada.

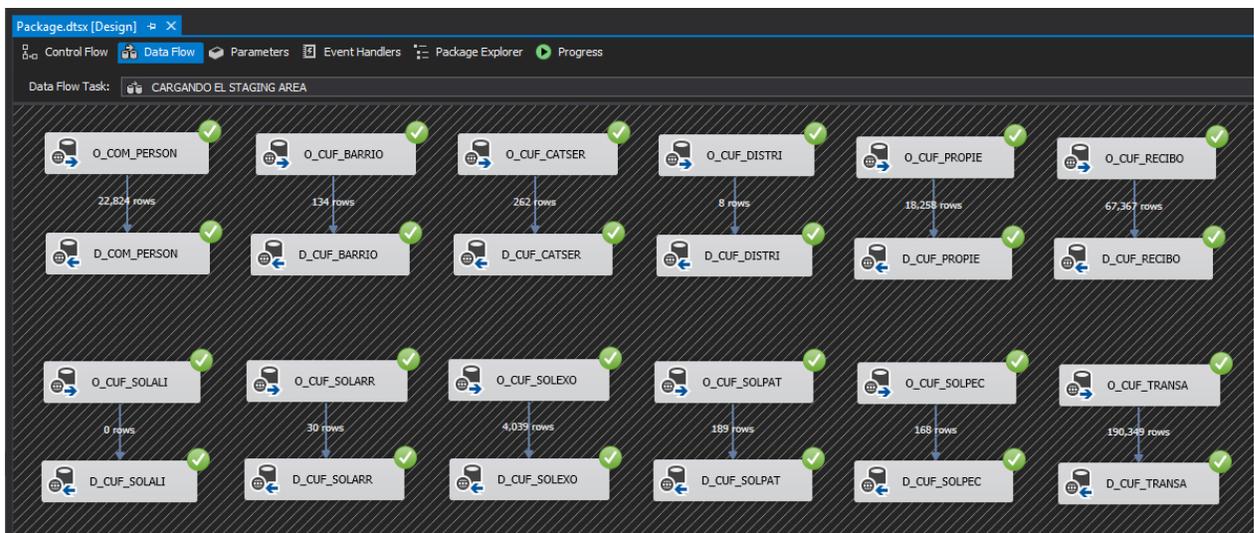


Figura 17. Cargando Staging Area de Ingresos.

Figura 17. Cargando Staging Area de Ingresos.

Fuente: Desarrollo propio (2018)

Representando los procesos, se realiza la descripción de la carga de los insumos de datos necesarios en el Staging Area, correspondiente al sector de ingresos; se consideran las entidades necesarias identificadas para satisfacer la carga de elementos, como las personas que interactúan en las diferentes reglas de negocio de la organización, las propiedades que se incorporan según el movimiento catastral respectivo al cantón de Esparza, la distribución de los ingresos acorde con el catálogo de servicios que detalla cada uno de los rubros considerados en las proyecciones presupuestarias y las transacciones que dan origen a los diferentes recibos cancelados.

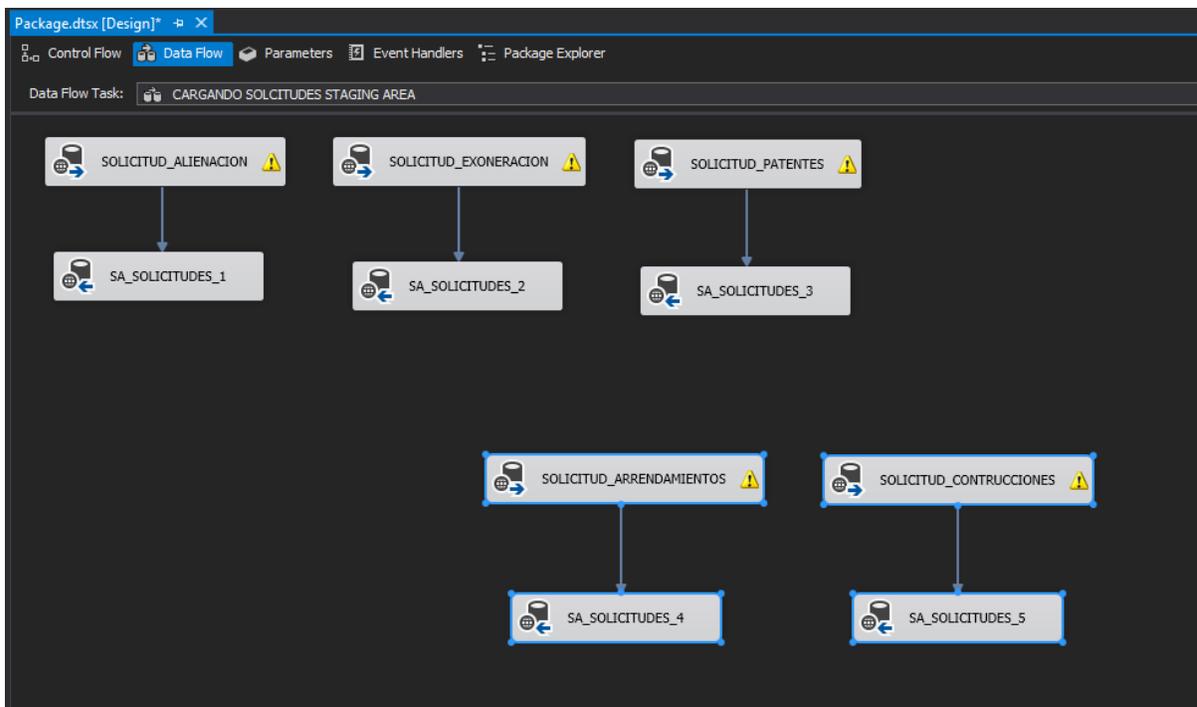


Figura 18 Cargando Staging Area de Solicitudes.

Figura 18. Cargando Staging Area de Solicitudes.

Fuente: Desarrollo propio (2018)

La carga de las diferentes solicitudes dentro del Staging Area permitirá contar con los datos necesarios para evaluar el crecimiento en los diferentes servicios generadores de ingresos, en los diferentes segmentos de ingresos de la gestión municipal.

En la figura anterior aparecen nuevamente los mensajes de advertencia, los cuales corresponden a los descritos con anterioridad para la figura 15 de este apartado.

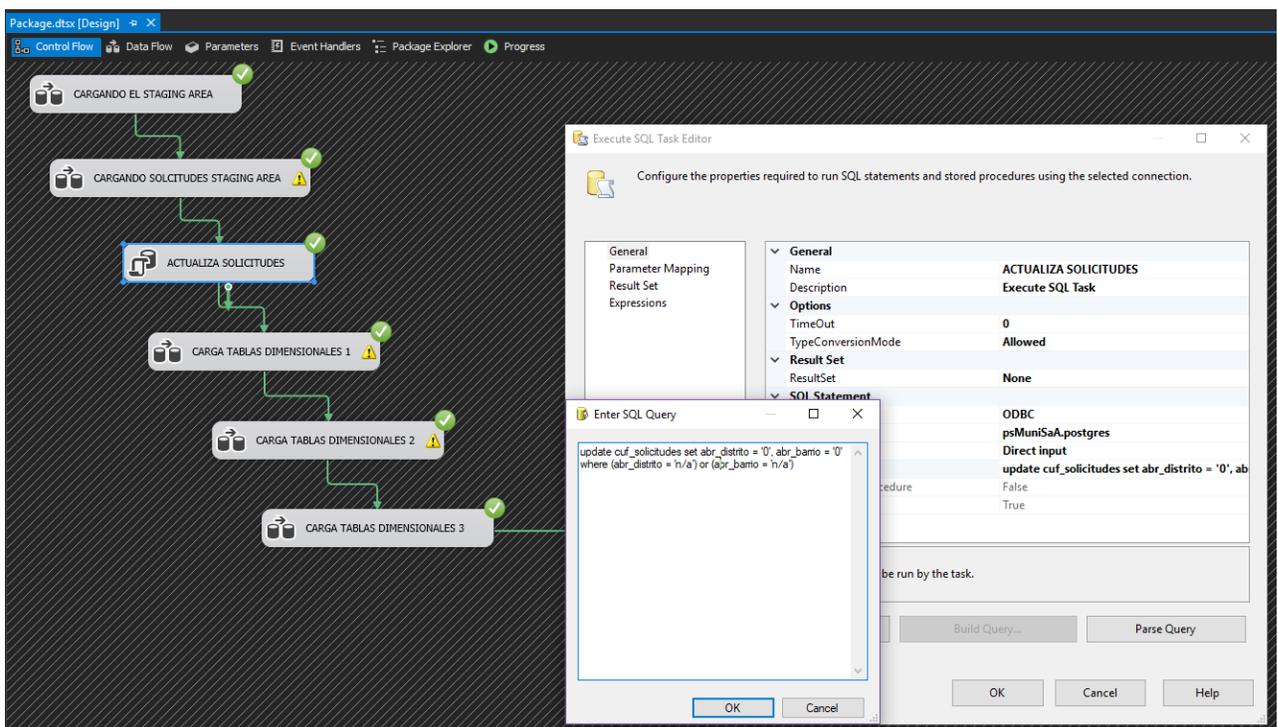


Figura 19. Vista General proceso Carga de Data Warehouse.

Figura 19. Vista General proceso Carga de Data Warehouse.

Fuente: Desarrollo propio (2018)

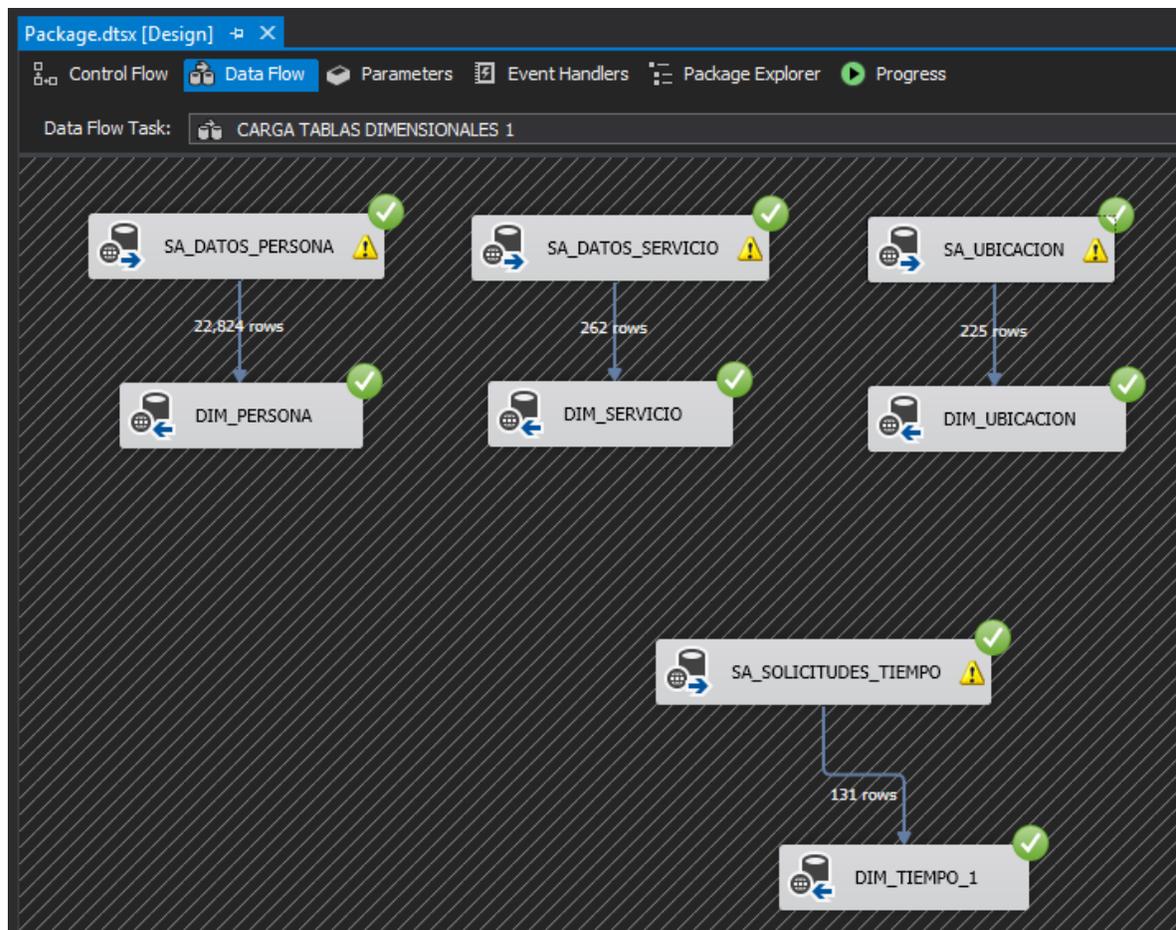


Figura 20 Cargando las Dimensiones.

Figura 20. Cargando las Dimensiones.

Fuente: Desarrollo propio, (2018)

Los datos contenidos en los diferentes Staging Area son trasladados a las dimensiones consideradas en el Data Warehouse, brindando los insumos necesarios para el contenedor universal de los datos; se asegura la integridad de los datos para su eventual uso en la toma de decisiones.

En la figura anterior aparecen nuevamente los mensajes de advertencia, los cuales corresponden a los descritos con anterioridad para la figura 15 de este apartado.

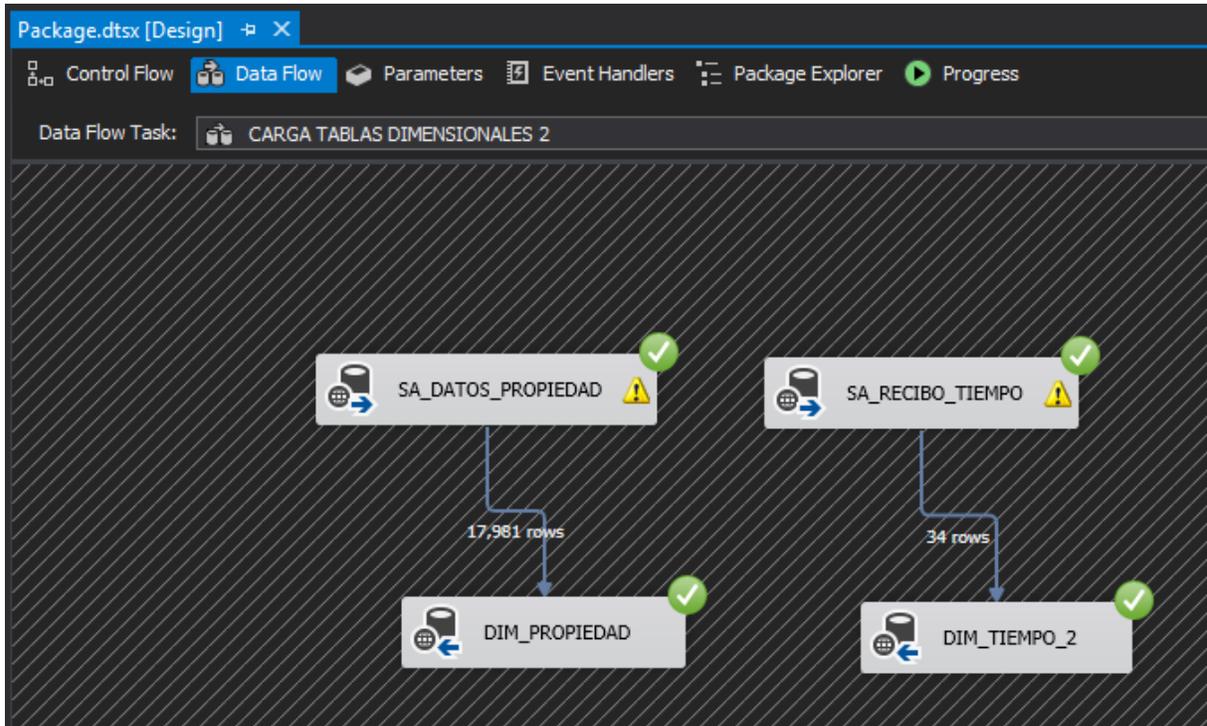


Figura 21 . Cargando las Dimensiones Propiedad y Tiempo.

Figura 21. Cargando las Dimensiones Propiedad y Tiempo.

Fuente: Desarrollo propio (2018)

El traslado de los datos pertenecientes a las propiedades, contenido en su nueva estructura, según los datos simplificados, es migrado a la dimensión *propiedad* y para los recibos existentes en la Staging Area se carga la dimensión *tiempo*.

En la figura anterior aparecen nuevamente los mensajes de advertencia, los cuales corresponden a los descritos con anterioridad para la figura 15 de este apartado.

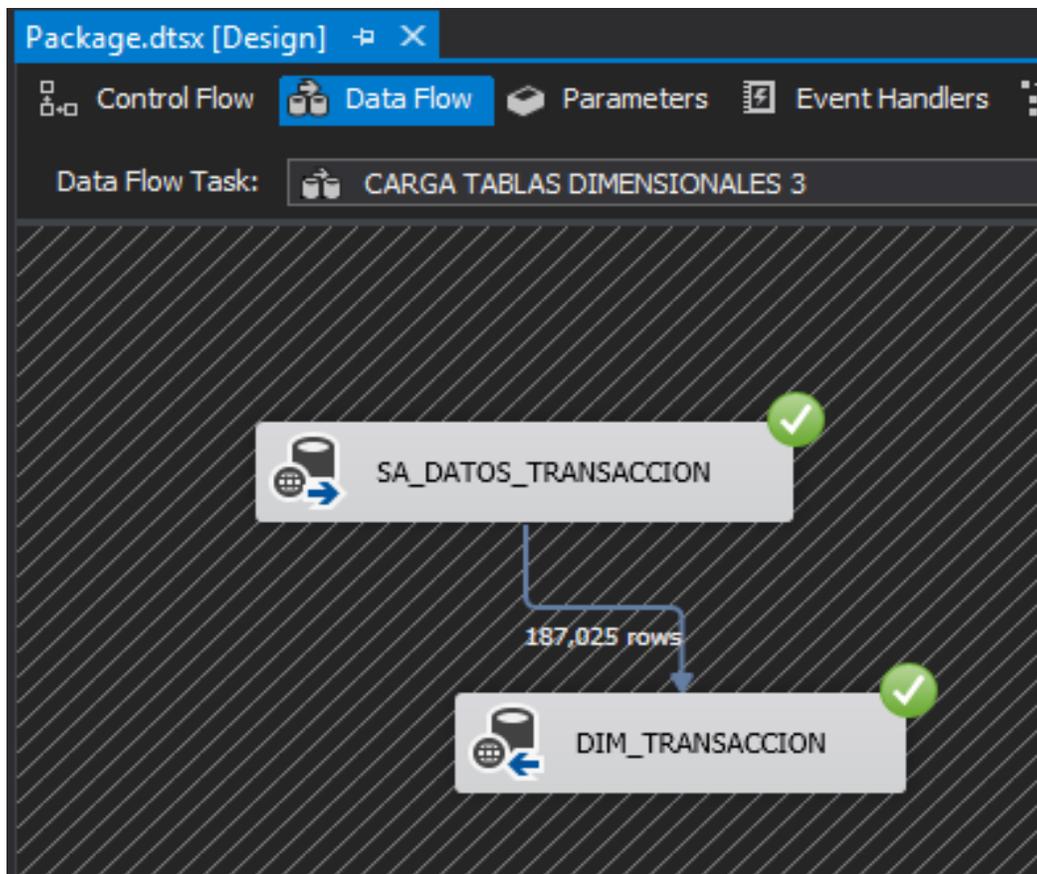


Figura 22 Cargando la Dimensión transacción.

Figura 22. Cargando la Dimensión transacción.

Fuente: Desarrollo propio (2018)

Las transacciones existentes en el Staging Area son migradas a la dimensión *transacción*; es acá donde se encuentran los diferentes pendientes de pago con base en el catálogo de servicios existente y sus tarifas.

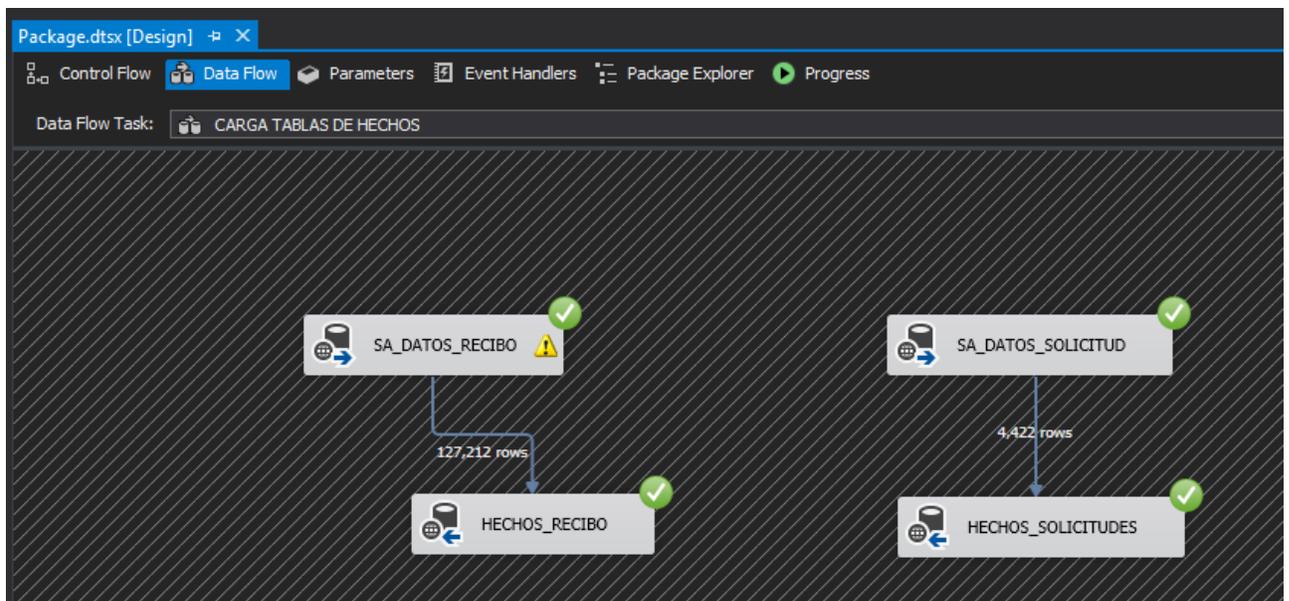


Figura 23. Cargando tablas de Hechos.

Figura 23. Cargando tablas de Hechos.

Fuente: Desarrollo propio (2018)

La elaboración de la tabla de hechos es representada en la figura 22.

En la figura anterior aparecen nuevamente los mensajes de advertencia, los cuales corresponden a los descritos con anterioridad para la figura 15 de este apartado.

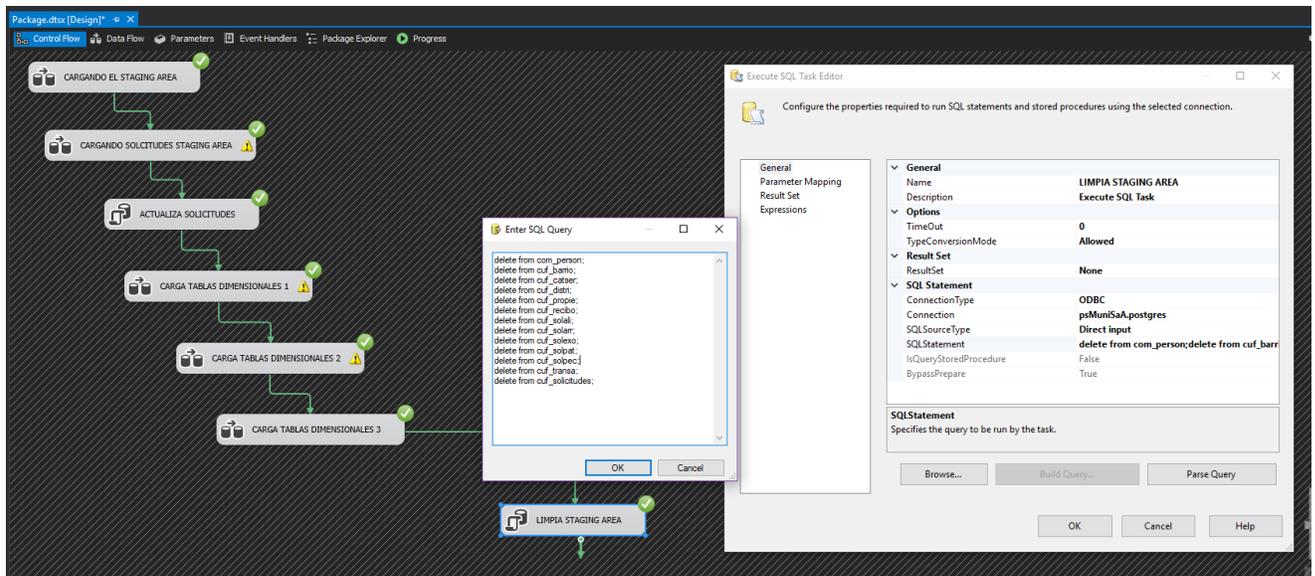


Figura 24 . Proceso completo de carga con Limpieza Staging Area.

Figura 24. Proceso completo de carga con limpieza de Staging Area.

Fuente: Desarrollo propio (2018)

Muestra del proceso que cumple la carga de los datos, además de la limpieza del Staging Area, una vez procesada la carga de datos mediante el flujo requerido.

Ello, para el caso de algunas visualizaciones que sirven como muestra para facilitar el acceso a los datos con una interpretación visual acorde con las necesidades de la organización municipal.

En la figura anterior aparecen nuevamente los mensajes de advertencia, los cuales corresponden a los descritos con anterioridad para la figura 15 de este apartado.

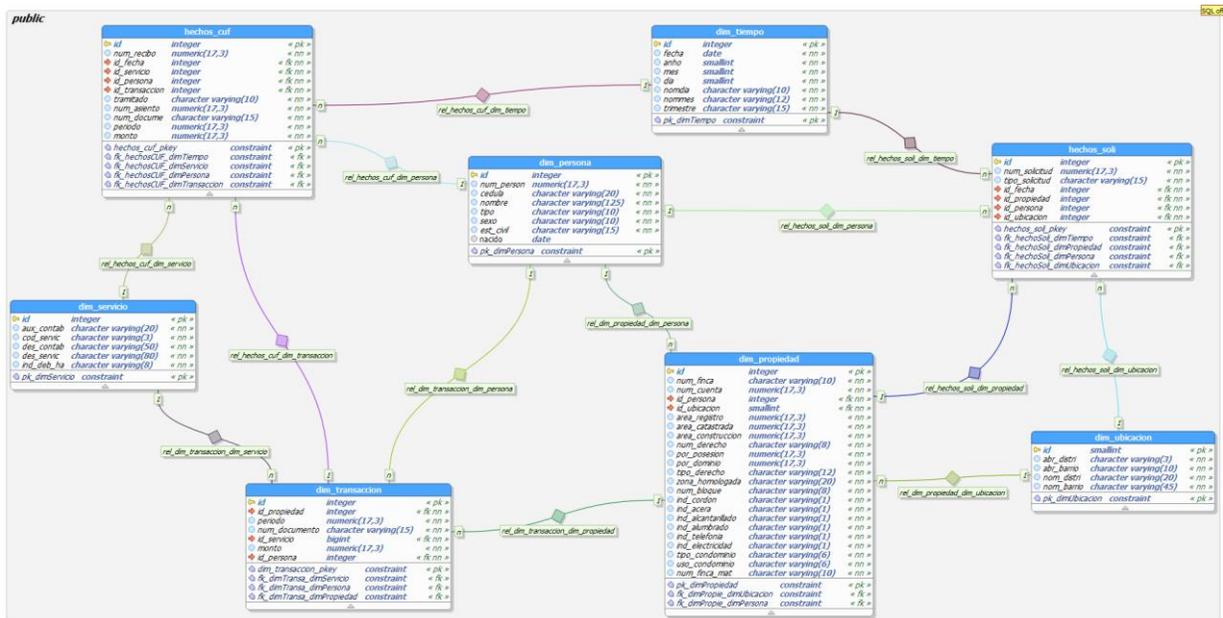


Figura. 25 Modelo de Data Warehouse.

Figura 25. Modelo de Data Warehouse.

Fuente: Desarrollo propio (2018)

El resultado final consiste en un Data Warehouse capaz de contener los datos simplificados que podrán ser sometidos a métricas con base en los hechos. Este contenedor permitirá a los usuarios identificar los datos seleccionados como importantes para la administración municipal.

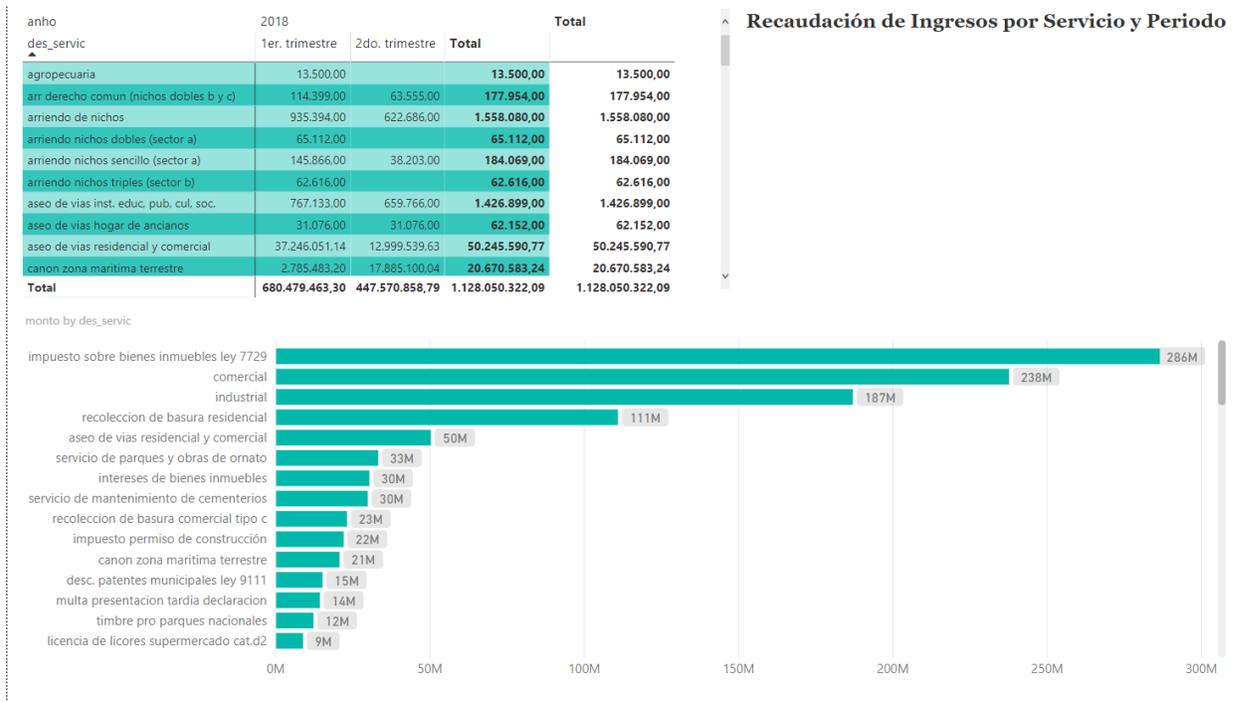


Figura 26 Visualización de Ingresos de los Servicios por Periodo.

Figura 26. Visualización de Ingresos de los Servicios por Periodo.

Fuente: Desarrollo propio (2018)

Un insumo importante para la administración municipal será poder contar con recursos visuales que permitan describir sus datos, para apoyo de la toma de decisiones, considerando las diferentes variables que enriquezcan el control de los datos procesados.

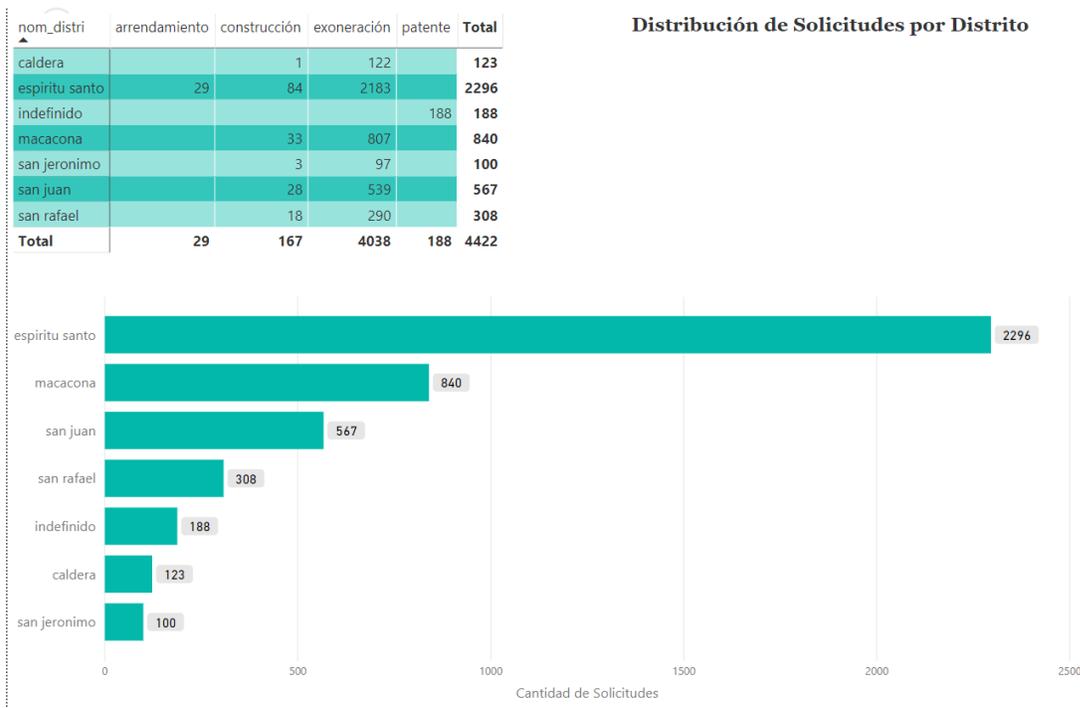


Figura 27. Visualización de Solicitudes por Distrito.

Figura 27. Visualización de Solicitudes por Distrito.

Fuente: Desarrollo propio (2018)

El data Warehouse satisface dos necesidades en concreto: la primera apunta al control sobre los ingresos municipales; la segunda, como se ha mencionado con anterioridad, a las solicitudes tramitadas por parte de los clientes y que de alguna forma tienen un efecto considerable en la generación de ingresos. El contar con un control que identifique el crecimiento del cantón permitirá tener una referencia de los nuevos ingresos proyectados a futuro.



Figura 28 Análisis Descriptivo Cantidad de Propiedades por Distrito.

Figura 28. Análisis Descriptivo Cantidad de Propiedades por Distrito.

Fuente: Desarrollo propio (2018)

Los modelos de minería de datos descriptivos con base en los datos cualitativos existentes en la estructura del Data Warehouse, permite obtener insumos que describen los datos; se logra, de esta forma, evidenciar aspectos que pueden generar insumos, los cuales, posteriormente, generen acciones para atender condiciones existentes en los datos o estrategias, que posibiliten el apoyo en la toma de decisiones.

La figura anterior permite hacer un análisis descriptivo de las propiedades, según su distrito de procedencia; se logra identificar, en promedio, la cantidad de propiedades que pertenecen a un distrito en específico.

```

64
65
66 {r}
67
68 # crea la consulta para determinar la edad promedio de propietario
69 datosPropietarios = set_utf8(dbGetQuery(conex, 'select * from edadPersonas' ));
70
71
72 # Crea el intervalo de años para definir el factor de edad por rangos
73 niveles <- c(-Inf, 17, 28, 38, 48, 58, 68, 78, 88, 98, Inf);
74
75
76 # Crea las etiquetas para el factor de edad por rangos
77 etiquetas <- c("menor de edad", "18 - 28", "29 - 38", "39 - 48", "49 - 58", "59 - 68", "69 - 78", "79 - 88", "89 - 98", "mayor de 98");
78
79
80 # Crea una nueva columna llamada Fac_Edad, basada en el campo Edad, aplicando los
81 # niveles y asignando las etiquetas.
82 datosPropietarios$Fac_Edad = cut(datosPropietarios$Edad, niveles, etiquetas);
83
84
85 # Se crean las agrupaciones por distrito y por rango de edades
86 datosPropietarios$nom_distri <- factor(datosPropietarios$nom_distri);
87 datosPropietarios$Fac_Edad <- factor(datosPropietarios$Fac_Edad);
88
89
90 # Se crea tabla para determinar la frecuencia de ocurrencia entre ambos factores
91 table(datosPropietarios$nom_distri, datosPropietarios$Fac_Edad);
92
93

```

```

91:65  Chunk 4 - R Markdown
Console Terminal
D:/Documentos/Cursos UTN/2108 - 2/BDA/Clases/R/Ejemplo01/
>
> # Se crea tabla para determinar la frecuencia de ocurrencia entre ambos factores
> table(datosPropietarios$nom_distri, datosPropietarios$Fac_Edad);

```

	menor de edad	18 - 28	29 - 38	39 - 48	49 - 58	59 - 68	69 - 78	79 - 88	89 - 98	mayor de 98
caldera	3	11	42	130	241	276	170	109	41	96
espíritu santo	31	188	735	1233	1695	1343	749	368	106	212
macacona	6	94	455	630	656	481	266	106	48	74
otr	0	3	2	2	1	3	2	0	0	1
san jeronimo	2	19	78	114	184	216	98	51	16	39
san juan	5	87	314	407	493	410	206	89	42	70
san rafael	3	31	195	249	318	293	151	82	8	62

Figura 29 Análisis descriptivo clientes por rango de edad.

Figura 29. Análisis descriptivo clientes por rango de edad.

Fuente: Desarrollo propio (2018)

La figura anterior, mediante herramientas de minería de datos, permite la clasificación de los clientes de la organización. Este insumo posibilita dirigir múltiples acciones en los diferentes contextos donde interviene la regla de negocio de la organización. Estos datos claramente podrían tener un grado de granularidad todavía con un detalle más específico, donde se indica la población impactada por algún

servicio o acción social por ejercer; dichos datos podrían ser presentados por una granularidad por barrio.

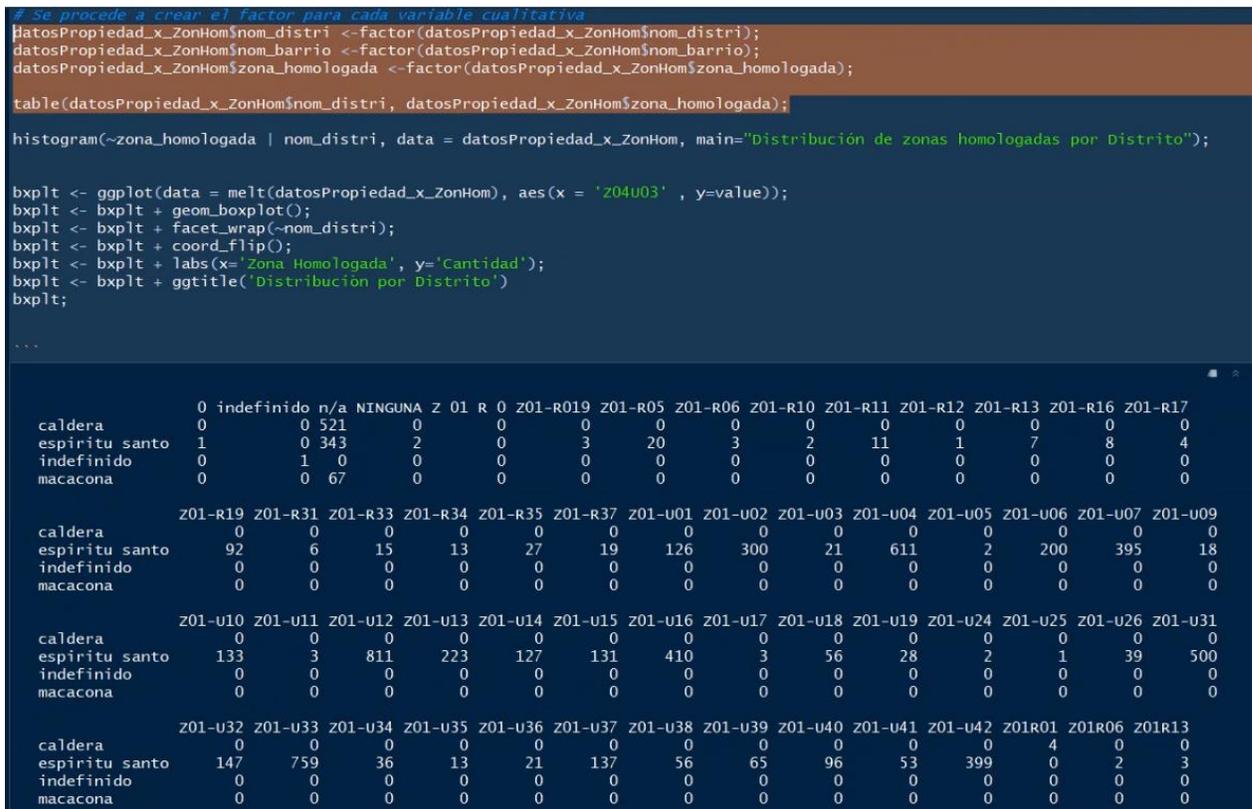


Figura 30 Análisis correlación variables zonas homogéneas por distrito.

Figura 30. Análisis correlación variables zonas homogénea por distrito.

Fuente: Desarrollo propio (2018)

Un aspecto de importancia es el crecimiento urbanístico presente en el cantón de Esparza; la distribución de las propiedades existentes, con base en la clasificación a cada una de las zonas homogéneas, brindará una noción del crecimiento en cuanto a infraestructura constructiva predominante en las diferentes zonas.

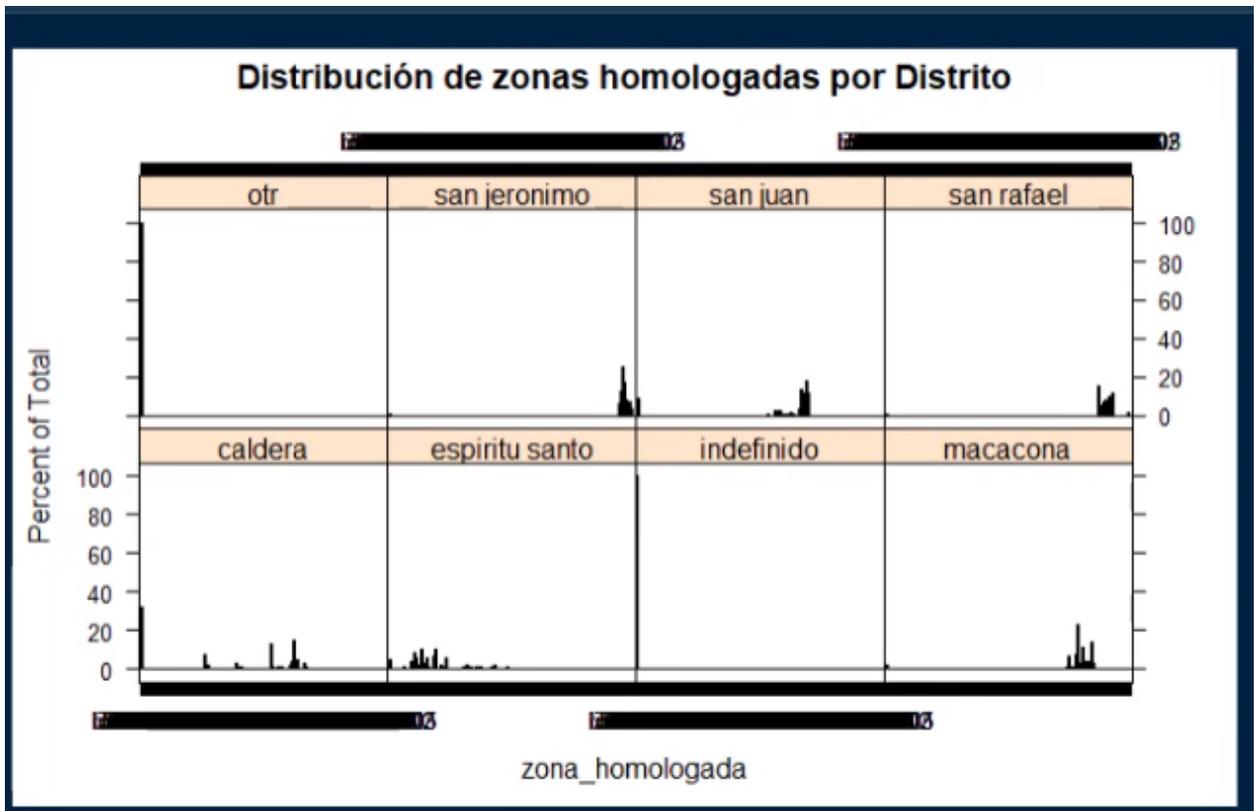


Figura 31. Gráfico distribución de zonas por distrito.

Figura 31. Gráfico distribución de zonas por distrito.

Fuente: Desarrollo propio (2018)

En la figura anterior se muestra la distribución de los distritos con base en las zonas homogéneas existentes. Como se mencionó con anterioridad, se pueden generar estrategias de ordenamiento territorial que aporten un valor adicional en la toma de decisiones.

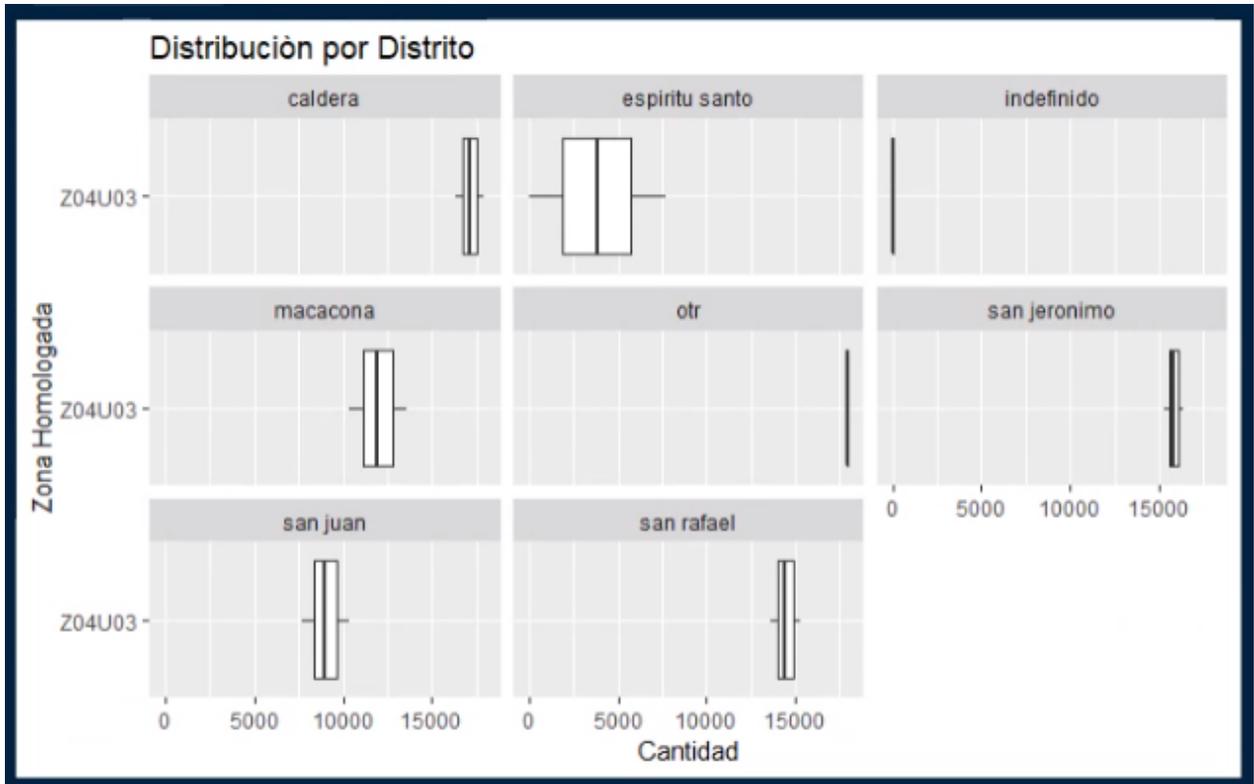


Figura 32. Gráfico distribución zona específica por distrito.

Figura 32. Gráfico distribución zona específica por distrito.

Fuente: Desarrollo propio (2018)

En este sentido, se pueden minar los datos evidenciando de forma específica el comportamiento de los cualitativos, donde, con base en el distrito, se puede observar que la zona seleccionada tiene una representación significativa en Espíritu Santo. Esto permitirá identificar si realmente esas propiedades se encuentran zonificadas de manera correcta y su valor está acorde con la zona.

```

# Crea el factor de agrupamiento por el campo sexo
datosPropietarios$sexo <-factor(datosPropietarios$sexo);

# Crea un filtro para solo obtener el subconjunto de personas mayores a 58 años y que sean mujeres
# propietarias en la zona de Esparza
distFemDist <- datosPropietarios[which(datosPropietarios$sexo=='Femenino' & datosPropietarios$edad>58),];
table(distFemDist$nom_distri, distFemDist$Fac_Edad);
...

<18 18 - 28 29 - 38 39 - 48 49 - 58 59 - 68 69 - 78 79 - 88 89 - 98 >98
caldera      0      0      0      0      0      0      12      11      5      2      0
espíritu santo 0      0      0      0      0      0      88      61      21      2      7
macacona     0      0      0      0      0      0      64      20      3      6      1
otr          0      0      0      0      0      0      0      0      0      0      0
san jeronimo 0      0      0      0      0      0      16      0      7      0      0
san juan     0      0      0      0      0      0      22      13      4      0      4
san rafael   0      0      0      0      0      0      20      8      4      0      4

...{r}
# crea la consulta para determinar la edad promedio de propietario
datosPropiedad_x_ZonHom = set_utf8(dbGetQuery(conex, 'select pr.id, ub.nom_distri, ub.nom_barrio, pr.zona_homologada
from dim_propiedad pr, dim_ubicacion ub
where ub.id = pr.id_ubicacion' ));

# Se procede a crear el factor para cada variable cualitativa
datosPropiedad_x_ZonHom$nom_distri <-factor(datosPropiedad_x_ZonHom$nom_distri);
datosPropiedad_x_ZonHom$nom_barrio <-factor(datosPropiedad_x_ZonHom$nom_barrio);
datosPropiedad_x_ZonHom$zona_homologada <-factor(datosPropiedad_x_ZonHom$zona_homologada);

```

Figura 33. Matriz de distribución de mujeres con base en su rango de edad por distrito.

Figura 33. Matriz de distribución de mujeres con base en su rango de edad por distrito.

Fuente: Desarrollo propio (2018)

Nuevamente la organización puede considerar aspectos demográficos para la toma de decisiones. La figura anterior detalla, para cada distrito, las mujeres dueñas de propiedades y los rangos de edad correspondientes para cada una.

Finalmente, una de la perspectivas de este proyecto es interiorizar en la Municipalidad de Esparza el uso de herramientas y técnicas de avanzada, para la exploración y aprovechamiento de sus datos, en todos sus contextos, ampliar su uso

y enfocarlo en esfuerzos que logren descubrir sus virtudes. Este proyecto posibilita la integración de los datos, mediante los cuales se podrán evaluar modelos de minería, en el ámbito predictivo, de algunos comportamientos que pueden controlarse con el apoyo de estos.

Capítulo 6. Conclusiones y Recomendaciones

6.1 Conclusiones

Para la Municipalidad de Esparza resulta un insumo importante la identificación de los recursos de datos que aportarán las capacidades necesarias tanto para satisfacer el alcance de este proyecto, como para impactar un aspecto de suma relevancia como es la política de datos abiertos. El presente proyecto logró identificar las fuentes de datos necesarias para la construcción del *Data Warehouse* y proporcionar los insumos necesarios, para comenzar con el proceso de apertura de datos de la Municipalidad de Esparza.

Dentro del proyecto se implementan por lo menos dos *Data Marts*, que establecen las mejores prácticas en su confección, considerando los requerimientos idóneos y el proceso de mejora continua para su diseño; además, se definieron las técnicas que permitieron plasmar la visualización de los datos necesarios para los usuarios finales, por medio del perfil de acceso.

Se desarrolla el análisis de las estructuras de datos y fuentes de origen de una forma detallada, para comprender la regla de negocio y la relación existente entre dichos datos; esto permite generar un preámbulo de los aspectos claves de diseño por considerar para los procesos de extracción, transformación y carga para el *Data Warehouse* elaborado.

El proyecto identifica recursos de minería de datos descriptiva para satisfacer las necesidades de negocio; por su parte, estos recursos pueden optimizarse a través del proceso de mejora continua, asumiendo procesos de reingeniería que buscan claramente satisfacer los aspectos cambiantes de negocio y las consideraciones indicadas por los usuarios finales, en la búsqueda del valor agregado en el apoyo de la toma de decisiones facilitando de manera visual dicho aspecto.

6.2 Recomendaciones

Las fuentes de datos deben someterse a una evaluación con base en el ciclo de vida de los datos, para determinar si existen cambios que puedan afectar la calidad, seguridad y rendimiento de estos y considerando aspectos claves en la toma de decisiones. Además, esta evaluación asegura que los conjuntos de datos publicados, con el sentido de satisfacer los requerimientos de los datos abiertos, pueden verse.

En cuanto a las mejores prácticas, se debe evaluar constantemente su aplicación; esto con el propósito de mantener un alcance satisfactorio en los diversos *Datamart*, así como cualquier insumo relacionado con este proyecto. Lo anterior permitirá a cualquier profesional capaz de dar sostenibilidad técnica al proyecto, identificar las prácticas utilizadas y el porqué del producto obtenido en razón de estas.

En cuanto a las fuentes tratadas para la generación de los procesos de carga, claramente sus datos son procesados por los diferentes usuarios finales en sus procesos diarios; es importante identificar ciclos de vida de dichos datos, además de establecer una política de gobernanza de estos, que obligue a estos usuarios a generar los insumos considerados por los procesos E.T.L.

Realizar un análisis periódico de los elementos descriptivos de minería de datos para que se ajusten al cambio continuo presente en los negocios, de esta forma, los recursos visuales de minería de datos responderán a las necesidades iniciales, con ajuste al nuevo modelo de negocio, si este se viera sujeto a alguna variación.

Bibliografía

- Bernal, C. A. (2010). *Metodología de la Investigación* (3ra ed.). Bogotá, Colombia: Pearson Educación de Colombia Ltda.
- Cirillo, A. (2017). *R Data Mining*. Birmingham: Packt Publishing Ltd.
- Darmawikarta, D. (2007). *Dimensional Data Warehousing with MySQL. A tutorial*. (1ra ed.). Montreal, Canada: Brainy Software Corp.
- Hernández Sampieri, R. (2010). *Metodología de la Investigación* (5ta ed.). México, D.F, México: McGraw Hill Interamericana.
- Hernández Sampieri, R. (2017). *Fundamentos de Investigación* (1ra ed.). Ciudad de México, Distrito Federal, México: McGraw Hill Interamericana.
- IBM. (2007). *IBM Data Governance Council Maturity Model*. Somers, NY 10589: IBM Software Group.
- Martínez, J. (23 de Julio de 2012). *Ibm.com*. Obtenido de <https://www.ibm.com/developerworks/ssa/data/library/techarticle/gobierno-datos/index.html>
- Oracle. (2017). *Database Data Warehousing Guide*. Recuperado el 8 de Ene de 2018, de <https://docs.oracle.com/database/121/DWHSG/concept.htm#DWHSG-GUID-452FBA23-6976-4590-AA41-1369647AD14D>
- Rangarajan, S. (1 de Sep de 2016). *Data Warehouse Design – Inmon versus Kimball*. Recuperado el 7 de Ene de 2018, de <http://tdan.com/data-warehouse-design-inmon-versus-kimball/20300>
- Ribas, E. (08 de Enero de 2018). *¿Qué es el Data Mining o minería de datos?* Recuperado el 07 de may de 2018, de <https://www.iebschool.com/blog/data-mining-mineria-datos-big-data/#data>
- Ricardo, C. M. (2009). *Bases de Datos* (1ra ed.). México: McGraw Hill Interamericana.
- Sarsfield, S. (2009). *The Data Governance Imperative. IT Governance Publishing*. Cambridgeshire: IT Governance Publishing.

Anexos

Anexo 1

UNIVERSIDAD CENFOTEC

MAESTRIA EN BASE DE DATOS

PROYECTO FINAL DE GRADUACIÓN

DICENTES

HUBERTH FRANCISCO FALLAS ROMERO

JORGE FÉLIX RUIZ FERNÁNDEZ

El cuestionario tiene como objetivo demostrar la necesidad de la implementación de técnicas y mecanismos que permitan la integración de los datos mediante un almacén de datos único, además de la necesidad de construir modelos que permitan la minería de datos en la Municipalidad de Esparza.

Cuestionario

1. ¿Considera que los datos presentes en la base de datos de la Municipalidad de Esparza satisfacen las necesidades o propósitos de todas sus funciones que realiza en la unidad que representa para el desarrollo de sus funciones?

() Excelente () Muy Bueno () Bueno () Regular () Malo

2. ¿Piensa usted que los datos generados por todas las unidades son recolectados e integrados de manera correcta para su posterior utilización?

() Excelente () Muy Bueno () Bueno () Regular () Malo

3. ¿Cree que los datos generados en la Municipalidad de Esparza son utilizados de una manera adecuada para la aplicación de análisis empresarial?

Excelente Muy Bueno Bueno Regular Malo

4. ¿Considera usted el análisis empresarial como una herramienta importante en la actualidad para el apoyo en la toma de decisiones de la Municipalidad de Esparza?

Excelente Muy Bueno Bueno Regular Malo

5. ¿Conoce el término “gobernanza de los datos”?

Sí No **Pasar a la pregunta N # 5.1**

5.1 Se considera la gobernanza de los datos como un factor clave para mejorar la productividad en las organizaciones, ¿cree usted oportuna su implementación en la Municipalidad de Esparza?

Excelente Muy Bueno Bueno Regular Malo

6. ¿Considera importante que la Municipalidad de Esparza pueda contar con una herramienta que permita la integración de los datos y apoye a la toma de decisiones?

Excelente Muy Bueno Bueno Regular Malo

7. ¿Conoce el término “minería de datos”?

Sí No **Pasar a la pregunta N # 7.1**

7.1 ¿Le resulta importante la construcción de modelos de minería de datos para el análisis predictivo de temas relacionados con el ámbito municipal?

Excelente Muy Bueno Bueno Regular Malo

8. ¿Cree usted que la Municipalidad de Esparza debe considerar la implementación de estas técnicas y mecanismos para generar insumos que puedan brindar el apoyo en la toma de decisiones?

Excelente Muy Bueno Bueno Regular Malo

9. ¿Cómo considera usted que podría ser el aporte que puede traer consigo el desarrollo de este tipo de técnicas dentro de la Municipalidad de Esparza?

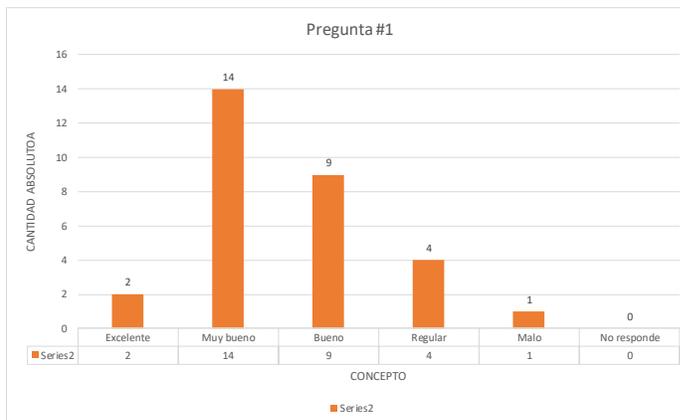
Excelente Muy Bueno Bueno Regular Malo

Anexo 2

Resultado de evaluación pregunta N 1

Universidad Cenfotec Maestría en Base de Datos Trabajo Final de Graduación Cuestionario #1	
Pregunta 1	Considera que los datos presentes en la base de datos de la Municipalidad de Esparza satisfacen las necesidades o propósitos de todas sus funciones que realiza en la unidad que representa para el desarrollo de sus funciones.

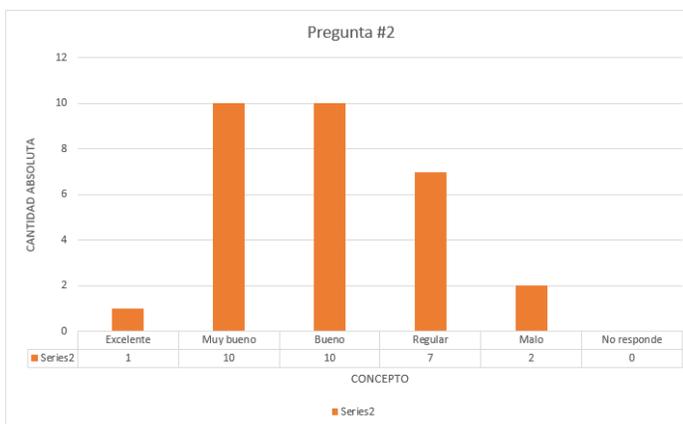
Concepto	Valor	Cantidad
Excelente	1	2
Muy bueno	2	14
Bueno	3	9
Regular	4	4
Malo	5	1
Sí	6	0
No	7	0
No responde	0	0
Total Encuestas		30



Resultado de evaluación pregunta N 2

Universidad Cenfotec Maestría en Base de Datos Trabajo Final de Graduación Cuestionario #1	
Pregunta 2	Piensa usted que los datos generados por todas las unidades son recolectados e integrados de manera correcta para su posterior utilización.

Concepto	Valor	Cantidad
Excelente	1	1
Muy bueno	2	10
Bueno	3	10
Regular	4	7
Malo	5	2
Sí	6	0
No	7	0
No responde	0	0
Total Encuestas		30

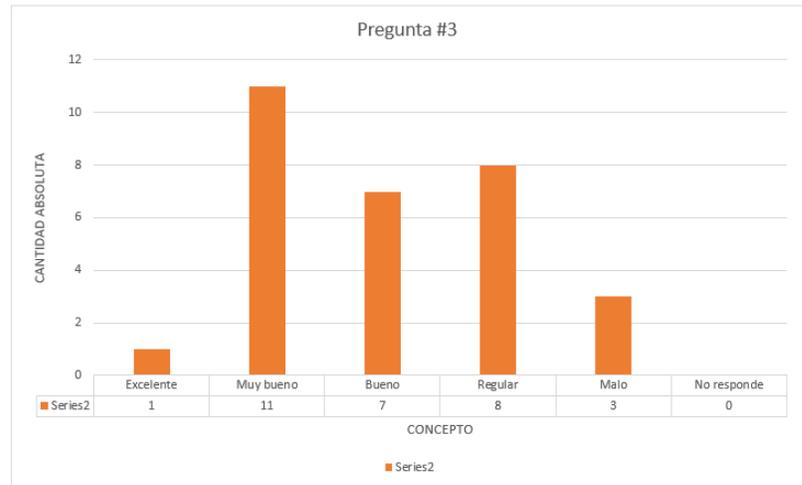


Resultado de evaluación pregunta N 3

Universidad Cenfotec
Maestría en Base de Datos
Trabajo Final de Graduación
Cuestionario #1

Pregunta 3 Cree que los datos generados en la Municipalidad de Esparza son utilizados de una manera adecuada para la aplicación de análisis empresarial.

Concepto	Valor	Cantidad
Excelente	1	1
Muy bueno	2	11
Bueno	3	7
Regular	4	8
Malo	5	3
Sí	6	0
No	7	0
No responde	0	0
Total Encuestas		30

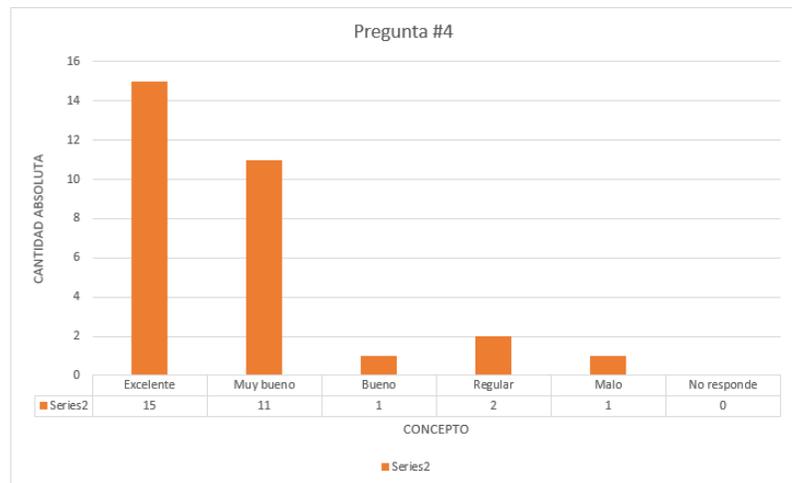


Resultado de evaluación pregunta N 4

Universidad Cenfotec
Maestría en Base de Datos
Trabajo Final de Graduación
Cuestionario #1

Pregunta 4 Considera usted el análisis empresarial como una herramienta importante en la actualidad para el apoyo en la toma de decisiones de la Municipalidad de Esparza.

Concepto	Valor	Cantidad
Excelente	1	15
Muy bueno	2	11
Bueno	3	1
Regular	4	2
Malo	5	1
Sí	6	0
No	7	0
No responde	0	0
Total Encuestas		30

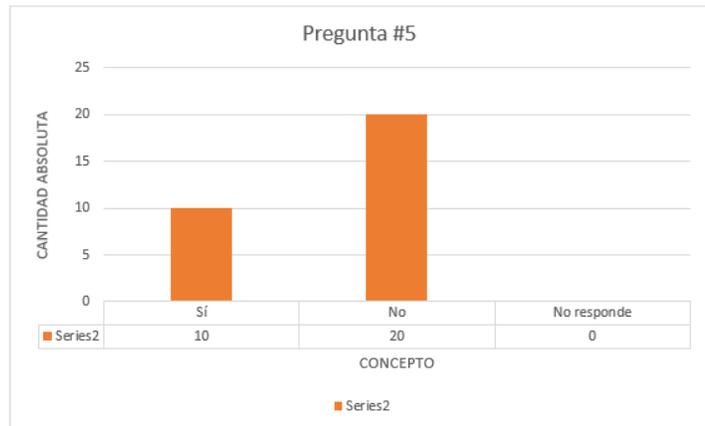


Resultado de evaluación pregunta N 5

Universidad Cenfotec
 Maestría en Base de Datos
 Trabajo Final de Graduación
 Cuestionario #1

Pregunta 5 Conoce el termino de Gobernanza de los datos.

Concepto	Valor	Cantidad
Excelente	1	0
Muy bueno	2	0
Bueno	3	0
Regular	4	0
Malo	5	0
Sí	6	10
No	7	20
No responde	0	0
Total Encuestas		30

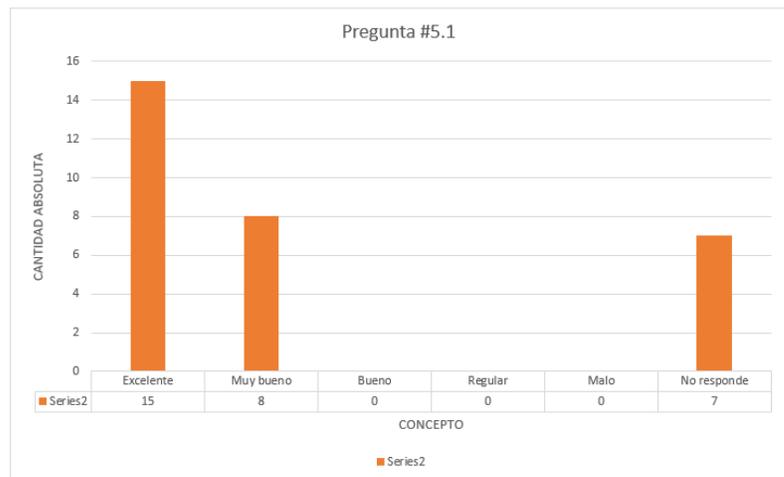


Resultado de evaluación pregunta N 5.1

Universidad Cenfotec
 Maestría en Base de Datos
 Trabajo Final de Graduación
 Cuestionario #1

Pregunta 5.1 Se considera la gobernanza de los datos como un factor clave para mejorar la productividad en las organizaciones, cree usted es oportuna su implementación en la Municipalidad de Esparza.

Concepto	Valor	Cantidad
Excelente	1	15
Muy bueno	2	8
Bueno	3	0
Regular	4	0
Malo	5	0
Sí	6	0
No	7	0
No responde	0	7
Total Encuestas		30

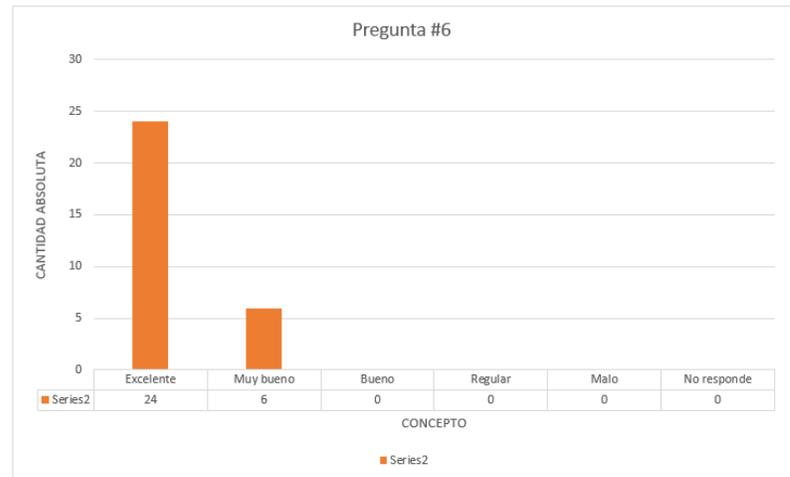


Resultado de evaluación pregunta N 6

Universidad Cenfotec
 Maestría en Base de Datos
 Trabajo Final de Graduación
 Cuestionario #1

Pregunta 6 Considera importante que la Municipalidad de Esparza pueda contar con una herramienta que permita la integración de los datos y apoye a la toma de decisiones.

Concepto	Valor	Cantidad
Excelente	1	24
Muy bueno	2	6
Bueno	3	0
Regular	4	0
Malo	5	0
Sí	6	0
No	7	0
No responde	0	0
Total Encuestas		30

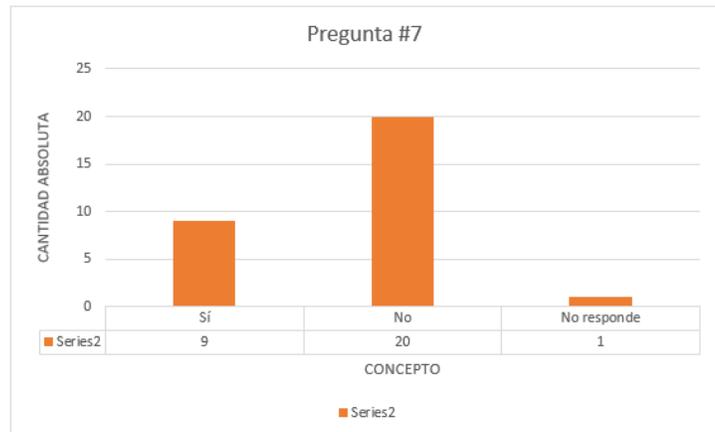


Resultado de evaluación pregunta N 7

Universidad Cenfotec
 Maestría en Base de Datos
 Trabajo Final de Graduación
 Cuestionario #1

Pregunta 7 Conoce el termino minería de datos.

Concepto	Valor	Cantidad
Excelente	1	0
Muy bueno	2	0
Bueno	3	0
Regular	4	0
Malo	5	0
Sí	6	9
No	7	20
No responde	0	1
Total Encuestas		30

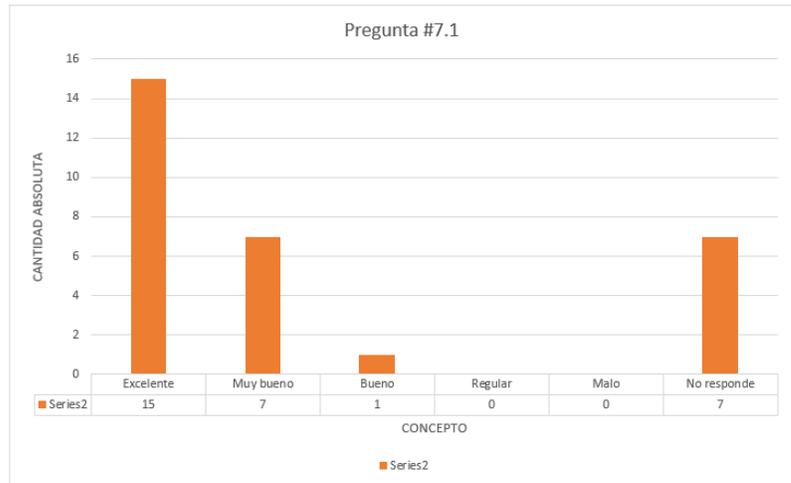


Resultado de evaluación pregunta N 7.1

Universidad Cenfotec
 Maestría en Base de Datos
 Trabajo Final de Graduación
 Cuestionario #1

Pregunta 7.1 Le resulta importante la construcción de modelos de minería de datos para el análisis predictivo de temas relacionados con el ámbito municipal.

Concepto	Valor	Cantidad
Excelente	1	15
Muy bueno	2	7
Bueno	3	1
Regular	4	0
Malo	5	0
Sí	6	0
No	7	0
No responde	0	7
Total Encuestas		30

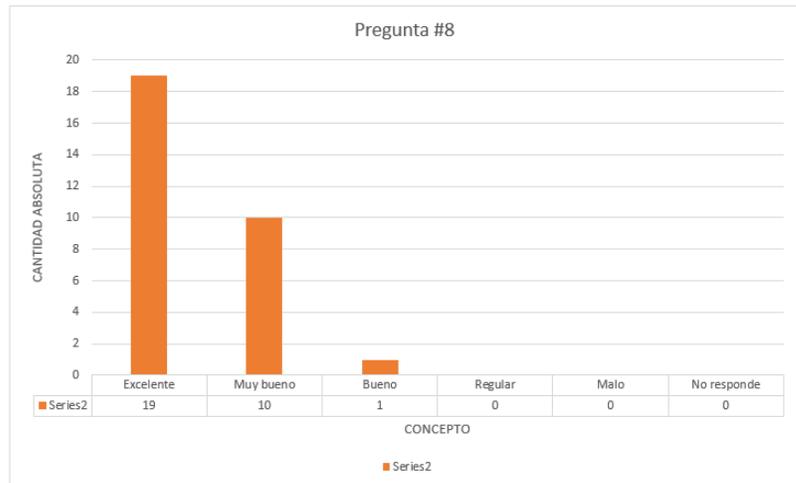


Resultado de evaluación pregunta N 8

Universidad Cenfotec
 Maestría en Base de Datos
 Trabajo Final de Graduación
 Cuestionario #1

Pregunta 8 Cree usted que la Municipalidad de Esparza debe considerar la implementación de estas técnicas y mecanismos para generar insumos que puedan brindar el apoyo en la toma de decisiones.

Concepto	Valor	Cantidad
Excelente	1	19
Muy bueno	2	10
Bueno	3	1
Regular	4	0
Malo	5	0
Sí	6	0
No	7	0
No responde	0	0
Total Encuestas		30



Resultado de evaluación pregunta N 9

Universidad Cenfotec
Maestría en Base de Datos
Trabajo Final de Graduación
Cuestionario #1

Pregunta 9 Cómo considera usted que podría ser el aporte que pueda traer consigo el desarrollo de este tipo de técnicas dentro de la Municipalidad de Esparza.

Concepto	Valor	Cantidad
Excelente	1	18
Muy bueno	2	12
Bueno	3	0
Regular	4	0
Malo	5	0
Sí	6	0
No	7	0
No responde	0	0
Total Encuestas		30

