



UNIVERSIDAD CATÓLICA DE CUENCA

Comunidad Educativa al Servicio del Pueblo

UNIDAD ACADÉMICA DE TECNOLOGÍAS DE LA INFORMACIÓN Y COMUNICACIÓN(TIC)

CARRERA DE INGENIERIA DE SISTEMAS

**INTELIGENCIA DE NEGOCIOS APLICADA AL ÁREA DE COMERCIALIZACIÓN
DE LA EMPRESA PÚBLICA MUNICIPAL DE AGUA POTABLE,
ALCANTARILLADO Y SANEAMIENTO AMBIENTAL (EMAPAL-EP) DE
AZOGUES**

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL
TÍTULO DE INGENIERO DE SISTEMAS**

AUTOR: DARWIN ARMANDO RIVERA SANCHEZ

DIRECTOR: BLANCA LUCIA AVILA CORREA

AZOGUES - ECUADOR

2020

*Yo me gradué en los
50 años de La Cato!*

DECLARACIÓN

Yo, **DARWIN ARMANDO RIVERA SÁNCHEZ**, declaro bajo juramento que el trabajo descrito es de mi autoría, ya que no ha sido previamente presentada para ningún grado o calificación profesional; y que he consultado las referencias bibliográficas que se incluyen en este documento

La Universidad Católica puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y la normativa institucional vigente

DARWIN ARMANDO RIVERA SÁNCHEZ

CERTIFICACIÓN

Certifico que le presente trabajo fue desarrollado por Darwin Armando Rivera Sánchez, bajo mi supervisión.

A handwritten signature in blue ink, appearing to read 'Blanca Lucía Avila Correa', with a large, stylized flourish extending to the right.

BLANCA LUCIA AVILA CORREA
DIRECTORA

AGRADECIMIENTOS

Este proyecto de investigación quiero agradecer principalmente a Dios, por prestarme la vida y llenarme de fuerzas para seguir día a día el camino correcto, la lucha contra momentos de dificultad y que pensé haberme caído.

Quiero agradecer a mis padres: Jaime e Inés, por ser quienes me impulsaron a seguir este sueño, implantando valores, consejos y buenos principios.

Así mismo quiero expresar mi reconocimiento a la Empresa Pública Municipal de Agua Potable, Alcantarillado y Saneamiento Ambiental EMAPAL-EP, de la ciudad de Azogues, por su disposición en información brindada para realizar este trabajo.

Finalmente quiero plasmar mi más sincero agradecimiento a mi tutora: Ing. Blanca Ávila, que con su gran conocimiento me oriento al desarrollo y culminación exitosa de este trabajo de titulación.

DEDICATORIA

Dedico esta tesis a mis padres quienes me han forjado como persona; este logro se lo debo a ustedes, gracias por sembrar en mí principios, esfuerzo y las ganas de cumplir mis sueños sin temor a las adversidades.

ÍNDICE DE CONTENIDO

DECLARACIÓN.....	- 1 -
CERTIFICACIÓN.....	- 3 -
AGRADECIMIENTOS.....	- 4 -
DEDICATORIA.....	- 5 -
ÍNDICE DE CONTENIDO.....	- 6 -
LISTA DE FIGURAS.....	- 9 -
LISTA DE TABLAS.....	- 10 -
LISTA DE ILUSTRACIONES.....	- 11 -
LISTA DE ANEXOS.....	- 12 -
RESUMEN.....	- 13 -
ABSTRACT.....	- 14 -
CAPITULO I. FUNDAMETACION TEORICA.....	- 15 -
1.1. INTRODUCCIÓN.....	- 15 -
1.2. FUNCIONAMIENTO DE BI EN LAS ORGANIZACIONES.....	- 17 -
1.3. DATA WAREHOUSE -DWH.....	- 18 -
1.3.1. Objetivos del Data Warehouse.....	- 19 -
1.3.2. Arquitectura de dwh.....	- 19 -
1.3.2.1. Data Sources o Fuentes de Datos.....	- 20 -
1.3.2.2. Arquitectura del Sistema ETL.....	- 20 -
1.3.2.2.1. Extracción.....	- 21 -
1.3.2.2.2. Limpiar y Conformar.....	- 22 -
1.3.2.2.3. Integración y Entrega.....	- 22 -
1.3.2.2.4. Servicios de Gestión ETL.....	- 23 -
1.3.2.2.5. Almacenes de Datos ETL.....	- 23 -
1.3.2.2.6. Metadatos ETL.....	- 24 -
1.3.2.2.7. <i>Data Mart</i>	- 24 -
1.4. ARQUITECTURA DE BI.....	- 25 -
1.4.1. Tipos de aplicaciones de BI.....	- 26 -
1.4.1.1. Gestión de Servicios de BI. [13].....	- 28 -
1.4.1.2. Almacenes de datos de BI. [13].....	- 28 -
1.5. MINERÍA DE DATOS (DATA MINING).....	- 28 -

1.6. OLAP (PROCESAMIENTO ANALÍTICO EN LÍNEA)	- 29 -
1.6.1. Modelos de Datos	- 29 -
1.6.1.1. Esquema Estrella.....	- 30 -
1.6.1.2. Esquema Copo de Nieve	- 31 -
1.6.1.3. Normalizado – 3FN.....	- 31 -
1.6.2. ENFOQUES PARA DESARROLLO DE APLICACIONES BI	- 32 -
1.6.2.1. Enfoques de diseño del almacén de datos	- 32 -
1.6.2.2. Evaluación de la metodología de diseño del almacén de datos	- 34 -
1.6.3. BENEFICIOS DE BI	- 35 -
1.6.3.1. Manejo del crecimiento de la empresa.....	- 36 -
1.6.3.2. Control de costos.....	- 36 -
1.6.3.3. Clientes.....	- 36 -
1.6.3.4. Indicadores de gestión.....	- 36 -
1.6.4. METODOLOGÍA DE IMPLEMENTACIÓN DE BI	- 36 -
1.6.4.1. Definición.....	- 38 -
1.6.4.2. Levantamiento de información.	- 38 -
1.6.4.3. Diseño.	- 39 -
1.6.4.4. Desarrollo.....	- 39 -
1.6.4.5. Pruebas	- 40 -
1.6.4.6. Puesta en producción.....	- 40 -
1.6.5. HERRAMIENTAS PARA EL DESARROLLO DEL PROYECTO	- 41 -
1.6.5.1. Power Desginer.	- 41 -
1.6.5.2. Exasol.....	- 41 -
1.6.5.3. Exaplus.....	- 42 -
1.6.5.4. PostregSQL.	- 42 -
1.6.5.5. Knime.....	- 42 -
1.6.5.6. Tableau.....	- 43 -
1.6.6. Costo Total de propiedad de las herramietas	- 44 -
CAPITULO 2. DIAGNOSITCO SITUACIONAL	- 47 -
2.1. ANTECEDENTES.....	- 47 -
2.2. DEFINICIÓN DEL PROBLEMA	49
2.3. JUSTIFICACIÓN	50
2.4. OBJETIVOS	50

2.4.1. objetivo General	50
2.4.2. Objetivos Específicos	50
2.5. ALCANCE.....	51
2.6. TRABAJOS RELACIONADOS	51
CAPITULO 3. PROPUESTA.....	53
3.1. Definición:.....	53
3.2. Levantamiento de la Información:	56
3.3. Diseño.....	62
3.3.1. Diseño del Modelo Dimensional	63
3.4. desarrollo.....	70
3.4.1. Creación del Data Mart.....	70
3.5. Proceso ETL	73
3.6. Pruebas.	76
3.6.1. Instalación del ambiente de pruebas.	77
3.6.2. Verificación del Data Mart	77
3.6.3. Pruebas ETL	77
3.6.4. Pruebas de Interfaces de Visualización	77
3.7. Puesta en marcha.....	77
3.8. RESULTADOS	79
CONCLUSIONES.....	84
RECOMENDACIONES	84
BIBLIOGRAFÍA	86

LISTA DE FIGURAS

Fig. 1. Datos, Información y Conocimiento. [1]	- 15 -
Fig. 2. BI en las Organizaciones.. [6]	- 18 -
Fig. 3. Arquitectura de DWH. [11]	- 20 -
Fig. 4. Arquitectura ETL. [13]	- 21 -
Fig. 5. Data Marts. [17]	- 24 -
Fig. 6 Front End. [20].....	- 26 -
Fig. 7. Tipos de Aplicaciones BI. [13]	- 27 -
Fig. 8. Esquema en Estrella.	- 30 -
Fig. 9. Esquema Copo de Nieve. [17]	- 31 -
Fig. 10. Esquema. [17]	- 31 -
Fig. 11. Esquema en FN. [17]	- 32 -
Fig. 12. Arquitectura de almacenamiento de Kimball. [13].....	- 33 -
Fig. 13. Arquitectura de almacenamiento de Inmon. [22].....	- 34 -
Fig. 14. Metodnologia Scrum. [28]	- 37 -
Fig. 15. Cuadrante Mágico de Garter de las Herramientas de Calidad de los Datos. [31]..-	41 -
Fig. 16. Cuadrante mágico para plataformas de ciencia de datos y aprendizaje automático. [37]	- 43 -
Fig. 17 Cuadrante Mágico de las Plataformas de Analítica e Inteligencia de Negocios. [40]	- 44 -
Fig. 18. Organigrama de EMAPAL. [42].....	48
Fig. 19. Modelo Lógico del Negocio. Autor (2020)	67
Fig. 20. Modelo Físico. Autor (2020)	69
Fig. 21. Conexión a la Base de Datos Analítica. Autor (2020).....	71
Fig. 22. Esquema del Data Mart - Exaplus. Autor (2020).....	73
Fig. 23. DWH. Autor(2020)	75
Fig. 24 Tarifas Recaudadas. Autor (2020)	79
Fig. 25 Abonados Por Mes. Autor (2020)	80
Fig. 26. M3 Consumidos. Autor (2020)	81
Fig. 27 M3 de Cliente. Autor (2020).....	82
Fig. 28 Dashboard Emapal. Autor (2020)	83

LISTA DE TABLAS

TABLA I	- 35 -
TABLA II	- 44 -
TABLA III	- 45 -
TABLA IV	- 45 -
TABLA V	- 46 -
TABLA VI	- 46 -
TABLA VII	54
Tabla VIII	55
TABLA IX	56
TABLA X	57
TABLA XI	57
TABLA XII	58
TABLA XIII	59
TABLA XIV	60
TABLA XV	61
TABLA XVI	62
TABLA XVII	63
TABLA XVIII	63
TABLA XIX	63
TABLA XX	64
TABLA XXI	64
TABLA XXII	65
Tabla XXIII	70
Tabla XXIV	70
Tabla XXV	78

LISTA DE ILUSTRACIONES

Ilustración 1. Generar Modelo Físico del Datawarehouse. Autor (2020)	68
---	----

LISTA DE ANEXOS

ANEXO. A ANTIPLAGIO	90
ANEXO. B PERMISO DE SUBIR LA TESIS AL REPOSITORIO	91
ANEXO. C. CERTIFICADO DE NO ADEUDAR LIBROS A LA BIBLIOTECA ...	92
ANEXO. D ACTA DE CONSTITUCION DEL PROYECTO	93
ANEXO. E CERTIFICADO DEL PROYECTO REALIZADO EN LA EMPRESA	95
ANEXO. F ESPECIFICACIONES DEL SERVIDOR	96

RESUMEN

En la actualidad, la información es uno de los activos más importantes de las organizaciones. Su análisis es la base fundamental que les permite mejorar sus procesos y la entrega de bienes y servicios de mejor calidad. A pesar de ello, no todas las organizaciones hacen una gestión eficiente y eficaz de la información; muy pocas en nuestro medio apuestan por ello y aprecian los beneficios que proporciona un adecuado análisis de información a través de las herramientas de Inteligencia de Negocios. En efecto, la Inteligencia de Negocios permite que las organizaciones se conozcan a sí mismas, a sus clientes, y a sus competidores, lo que contribuye a lograr ventaja competitiva.

El presente trabajo consiste en dotar de una solución de inteligencia de negocios, que apoye de manera vertiginosa la disponibilidad de la información requerida para la toma de decisiones informada a nivel gerencial, táctico y operativo en el área de comercialización de la Empresa Pública Municipal de Agua Potable, Alcantarillado y Saneamiento Ambiental EMAPAL-EP de la ciudad de Azogues.

Este documento se encuentra desarrollado en tres amplios capítulos. En el CAPÍTULO I se aborda una breve introducción y el marco teórico que fundamenta los aspectos de la Inteligencia de Negocios y el *Data Warehouse*, que van desde los conceptos, arquitectura hasta sus metodologías y herramientas. En el CAPÍTULO II se abordan los antecedentes, la problemática, el objetivo, justificación y alcance que tendrá este trabajo. Finalmente, el en CAPÍTULO III se hace una descripción detallada de la solución a la problemática, en base a los requerimientos planteados. De esta manera pretende apoyar con información oportuna, relevante, precisa para la toma de decisiones.

Palabras clave: Inteligencia de Negocios, *Data Warehouse*, Toma de decisiones informada.

ABSTRACT



CENTRO DE IDIOMAS

ABSTRACT

Darwin Armando Rivera Sánchez

At present, information is one of the most important assets organizations possess. Its analysis is the fundamental basis that allows them to improve their processes and the delivery of better-quality goods and services. Despite this, not all organizations make efficient and effective management of information; a few in our environment opt for it and appreciate the benefits provided by an adequate analysis of information through Business Intelligence tools. Indeed, Business Intelligence allows organizations to know themselves, their customers, and their competitors, which contributes to achieving competitive advantage.

The present work consists of providing a business intelligence solution, which vertiginously supports the availability of information required for informed decision making at the managerial, tactical and operational level in the marketing area of the Municipal Public Company of Drinking Water, Sewerage and Environmental Sanitation EMAPAL-EP in the city of Azogues.

This document is developed in three chapters. In CHAPTER I, a brief introduction, and the theoretical framework that bases the aspects of Business Intelligence and Data Warehouse are discussed, ranging from the concepts, architecture to its methodologies and tools. In CHAPTER II it is approached the antecedents, the problem, the objective, justification, and reach that this work will have. Finally, in CHAPTER III there is a detailed description of the solution to the problem, based on the requirements. In this way, it intends to support with timely, relevant, and precise information for decision making.

Keywords: business intelligence, data warehouse, informed decision-making process.

CAPITULO I. FUNDAMETACION TEORICA

1.1. INTRODUCCIÓN

Con el pasar del tiempo la informática ha sido considerada una herramienta con la única funcionalidad de dar soporte a las funciones operativas, esta idea debe ser modificada, debido a que las herramientas informáticas no son sólo instrumentos para la reducción de costos, sino que además son útiles para el mejor manejo de la información que dispone la empresa, generar ventajas competitivas y obtener nuevos beneficios. Actualmente gran parte de las organizaciones dedican gran parte del tiempo y de recursos económicos y humanos a la obtención, proceso, aplicación y proyección de información, es por esto por lo que la información se convierte en el activo más importante dentro del ámbito empresarial.

Frecuentemente, quienes conforman la Alta Dirección de las empresas encuentran pocos beneficios en la información en bruto que producen las transacciones que a diario se registran en los sistemas de información empresariales. Su necesidad trasciende a un nivel superior en la que la información se convierte en conocimiento. El análisis de los datos operativos proporciona conocimiento en forma de tablas, gráficos de pastel, líneas de tendencia, gráficos de barras, y otras herramientas estadísticas. Este análisis requiere acceder a una variedad de fuentes de datos que son transformados y visibilizados a través de interfaces amigables y fáciles de usar e interpretar [1].

La diferencia entre dato, información y conocimiento es claramente definida por Springer [1], donde manifiesta que: los Datos son “recopilados diariamente en forma de bits, números, símbolos u objetos”; Información son “datos organizados que se procesan, limpian, organizan en estructuras y carece de redundancia” y finalmente el conocimiento se refiere a la información integrada, que incluyen sucesos o hechos y relaciones percibidos. Ver figura 1.

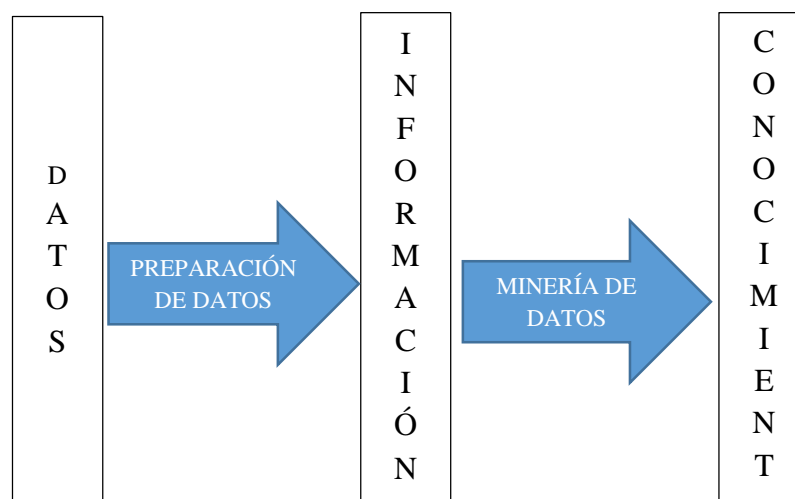


Fig. 1. Datos, Información y Conocimiento. [1]

Esta perspectiva del conocimiento fundamenta la toma de decisiones en las organizaciones; razón por la cual es necesario que las empresas implementen soluciones de Inteligencia de

Negocios. La Inteligencia de Negocios permite una toma de decisiones informada, que responda a las necesidades gerenciales y lograr ventaja competitiva. A través de la Inteligencia de Negocios las empresas pueden contar con información clave de sus clientes, proveedores e incluso de sus competidores; para ello deben intensificar sus estrategias para integrar grandes cantidades de datos que se encuentran dispersos en diversas fuentes [2].

Soejarto como se citó en [3] hace referencia que a pesar de que la Inteligencia de negocios sea un término de uso no muy común hoy en día, sigue creciendo incluso cuando la demanda de más productos de TI es baja. Diariamente va incrementando la necesidad de saber que ocurre en las empresas y organizaciones, pero la disposición de tiempo de las personas encargadas de este proceso es limitada, es por ello que se ha visto la necesidad de obtener toda esa información en el menor tiempo posible.

La Inteligencia de Negocios o en inglés llamada *Business Intelligence*, es conocida como un medio que nos facilita a la toma de decisiones en técnicas de negocios, comúnmente llamada como “BI”, nos ayuda en el proceso interactivo que busca y compara información organizada sobre un sitio para descubrir procesos, de los cuales obtendremos ideas y mediante la misma la obtención de una conclusión y conocimientos [4].

Muchos autores definen a la BI de manera diferente:

Según Calzada [5], el termino Inteligencia de Negocios procura caracterizar una amplia variedad de tecnologías, plataformas de software, especificaciones de aplicaciones y procesos. El objetivo principal que propone todo autor en la Inteligencia de Negocios es contribuir a tomar mejores decisiones que mejoren el desempeño de la empresa y promover su ventaja competitiva en el mercado. Este concepto se requiere analizar desde tres perspectivas: Hacer mejores decisiones en menos tiempo, convertir datos en información, y usar una aplicación relacional para la administración

Hatch [2] menciona que: “es la combinación de prácticas, capacidades y tecnologías usadas por las compañías para recopilar e integrar información, aplicar reglas de negocio y asegurar la visibilidad de la información en función de una mejor comprensión del mismo y, en última instancia, para mejorar el desempeño”

Para Médes [2] es; “El conjunto de herramientas y aplicaciones para la ayuda a la toma de decisiones que posibilitan acceso interactivo, análisis y multiplicación de la información corporativa de misión crítica. Estas aplicaciones aportan un conocimiento valioso sobre la información operativa identificando problemas y oportunidades de negocio. Con ellas los usuarios son capaces de acceder a grandes cantidades de información para establecer y analizar relaciones y comprender tendencias que, a la postre, soportarán decisiones de negocio.”

Al analizar las definiciones, se puede observar que existe coincidencia en algunos aspectos que se deben resaltar:

1. La Inteligencia de Negocios engloba; herramientas, tecnologías y metodologías que apoyan la toma de decisiones.

2. La Inteligencia de Negocios, permite visibilizar con mayor comprensión el conocimiento producto de procesar grandes cantidades de información.
3. No se debe confundir la gestión de grandes cantidades de información con su almacenamiento [2].

Entonces; la Inteligencia de Negocios es el nexo para que las organizaciones utilicen sus datos mediante herramientas puestas al servicio de los usuarios para facilitar la de tomar decisiones informada, incluyendo el funcionamiento actual de la empresa o el anticipo de futuros eventos.

El problema más común de la mayoría de las empresas es la carencia de integración, la cual es posible realizar mediante herramientas, técnicas y conceptos que se han crecido exponencialmente, como son: El *Data Warehouse* (DWH). El *Data Mining*, (DM) y el *On-line Analytical Processing*, (OLAP).

Para cubrir los problemas de integración, las herramientas de Inteligencia de Negocios permiten esta integración, proporcionando mayor velocidad en la obtención de datos y ahorro de tiempo.

Para que esto se consolide es necesario agrupar grandes cantidades de información de los sistemas transaccionales, sin importar la fuente, en una bodega de datos denominada; *Data Warehouse*, donde a través de una transformación se cargan para su posterior visualización. El proceso de extraer de las fuentes, transformar y cargar se conoce como ETL y tiene por objeto depurar y preparar los datos que se obtienen de las fuentes de datos antes de cargarlos en el almacén o bodega de datos (DWH).

El objetivo principal de las soluciones de BI es proporcionar la mayor cantidad de información procesable en el debido momento, lugar y de forma adecuada con los insumos adecuados para dar soporte a las personas encargadas de las tomas de decisiones.

1.2. FUNCIONAMIENTO DE BI EN LAS ORGANIZACIONES

Cada área organizativa genera datos pertenecientes a su propia actividad dentro de la empresa; y sus interrelaciones entre estos. Las soluciones tecnológicas como: CRM (*Customer Relationship Management*), SCM (*Supply Chain Management*), ERP (*Enterprise Resource Planning*), EIS (*Executive Information System*), entre otras, son los principales sistemas de información que gestionan los procesos de esas áreas de la empresa. BI debe actuar de manera estructurada y organizada para que se obtengan resultados efectivos. En este contexto, se deben definir tres niveles organizacionales a los que apoya BI: estratégico, táctico y operativo: (ver figura 2). [2]

- A nivel Estratégico, BI apoya en lograr ventaja competitiva. Requiere de una aplicación innovadora de tecnologías de información enfocadas por ejemplo en fidelizar clientes, ampliar el nicho de mercado, proveer productos o servicios con valor agregado.

- A nivel táctico, BI apoya en la toma de decisiones mediante sistemas de información que ayuden a los gerentes y profesionales de los negocios a decidir sobre aspectos relacionados a sus áreas. Por ejemplo: lanzar o retirar portafolio de servicios o productos.
- Finalmente, a nivel operativo las BI apoyan los procesos de negocio definiendo aspectos del día a día del personal de operaciones; Por ejemplo: registrar compras, mantener el inventario al día, evaluar nuevas tendencias, etc.

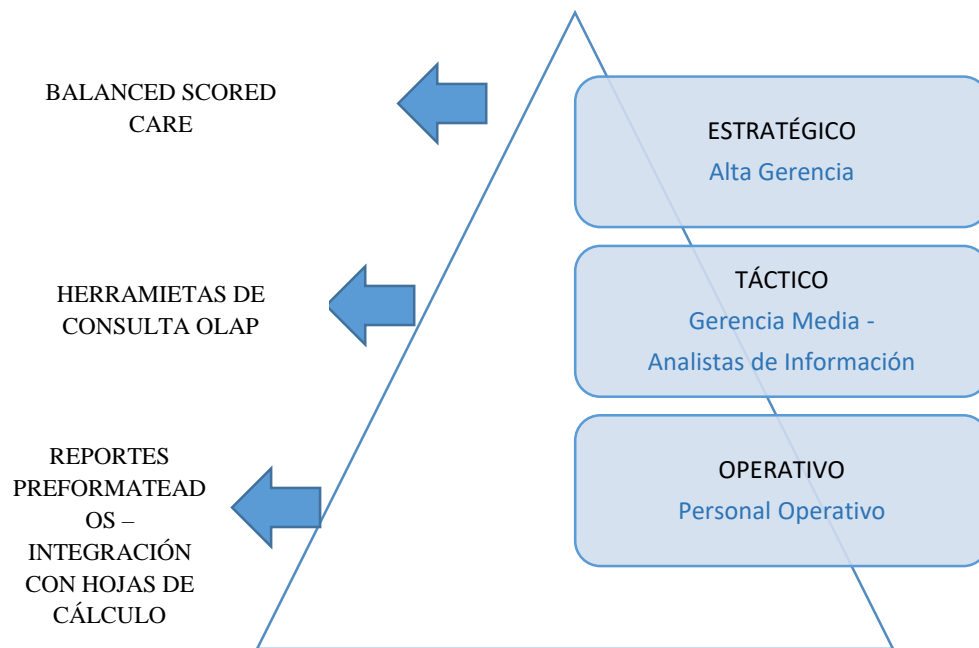


Fig. 2. BI en las Organizaciones.. [6]

1.3. DATA WAREHOUSE -DWH.

Sinnexus [7] afirma que *Data Warehouse* es un término que hace referencia a un almacén de datos. Es una base de datos corporativa, su principal característica es integrar y depurar un gran volumen de datos en el cual está integrada la información obtenida de una o diversas fuentes (CRM, *Google Analytics*, *Twitter*, sistemas contables, administrativos y fiscales, datos de ficheros de texto, ficheros Excel, comentarios de blogs, etc.) que previamente pasaron por un proceso ETL de esta manera se integra de forma homogénea para posteriormente procesarla permitiendo el análisis desde una infinidad de perspectivas que incluye grandes velocidades de respuesta.

Mendez [8] redacta que la creación de *Data Warehouse* constituye la primera fase para implementar una solución segura y confiable de BI, debido a que ofrece una visión integrada de la información, particularmente orientada hacia la toma de decisiones por parte de la administración de la organización.

El *Data Warehouse* presenta diversas ventajas, sin embargo, el principal beneficio radica en las estructuras donde se reúne la información, de esta manera permite la consulta y el tratamiento jerarquizado de la información [9].

Este término es debatido por dos grandes autores:

Ralph Kimball [10] Indica que “el *Data Warehouse* es una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis; es la unión de todos los Data Marts de una entidad”

Bill Inmon [10] determina que “el *Data Warehouse* es una colección de datos orientados al tema, integrados, no volátiles e historizados, organizados para el apoyo de un proceso de ayuda a la decisión”.

1.3.1. OBJETIVOS DEL DATA WAREHOUSE.

La aplicación de un *Data Warehouse* tiene como propósito varios estándares, entre los más destacados: conciben que toda la información de la organización sea accesible, incluyendo que los contenidos abordados sean entendibles y navegables, además el acceso es caracterizado por el rápido desempeño, es decir un cero tiempo de espera. Por otro lado, proporciona que la información sea consistente, adaptable y elástica. Otro de los objetivos principales es que protege los valores de la información, por último, es la fundación de la toma de decisiones, debido a que contiene los datos adecuados para soportar las aplicaciones de toma de decisiones.

1.3.2. ARQUITECTURA DE DWH

Para que se concrete el apoyo de BI en el logro de ventajas competitivas, toma de decisiones o procesos de negocio, es necesario que la información de los almacenes de datos esté disponible y actualizada periódicamente con información de los Sistemas de Datos Operativos, por sus siglas en inglés ODS. No es necesario información en tiempo real, sin embargo, la información debe estar disponible en el momento oportuno. Antes de cargarse en el DWH los datos deben limpiarse y estandarizarse – proceso ETL para finalmente aplicar procesos de visualización de datos mediante herramientas BI. Este proceso define la arquitectura típica de las soluciones de BI que se indican en la figura 3. [11]

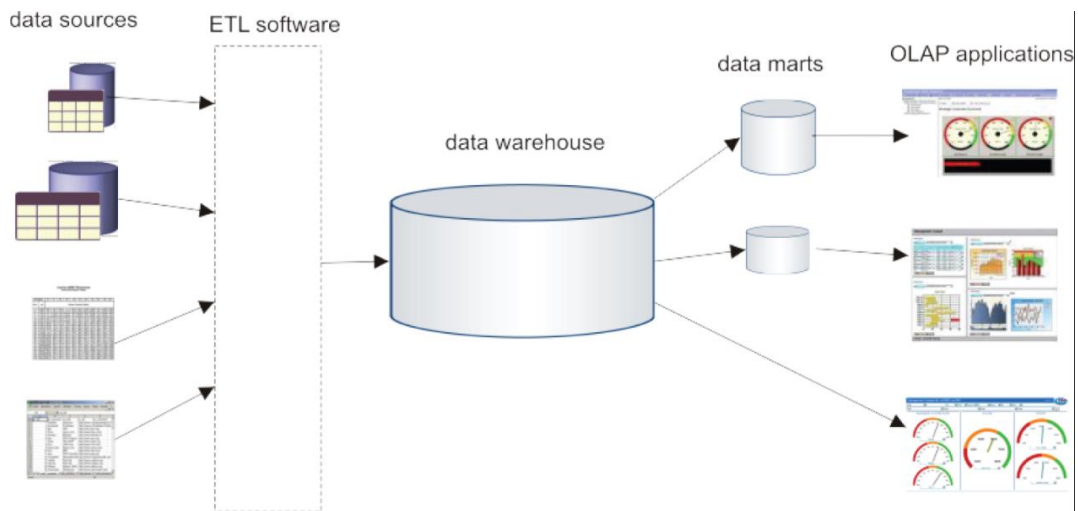


Fig. 3. Arquitectura de DWH. [11]

1.3.2.1. Data Sources o Fuentes de Datos

Munch [3] indica que las fuentes de datos es el componente principal de BI ya que hace referencia a toda la información que se va a obtener, de la cual dependerá la calidad de los datos. El origen de estos datos puede ser interno o externo. T. Balagueró [4] señala que el origen interno de los datos incluye esencialmente fuentes como: Sistemas; CRM, ERP, SRC, Bases de Datos, etc; así como también software de inventario, hojas de cálculo, entre otros. En tanto el origen externo, engloba los datos de redes sociales, como también las fuentes de datos abiertos gubernamentales. Una vez identificado el origen de las fuentes de datos, procedemos a realizar el proceso ETL.

1.3.2.2. Arquitectura del Sistema ETL

El ETL, gestiona los procesos ETL anteriormente mencionados. Es un componente esencial en la estructura de BI, ya que con su combinación permite el traslado de datos de una fuente a otra [12]. Esto se da gracias a la ejecución de tres acciones que comprende este proceso: Extracción, Transformación y Carga (ETL). La Fig. 4 muestra la arquitectura de un sistema ETL [13].

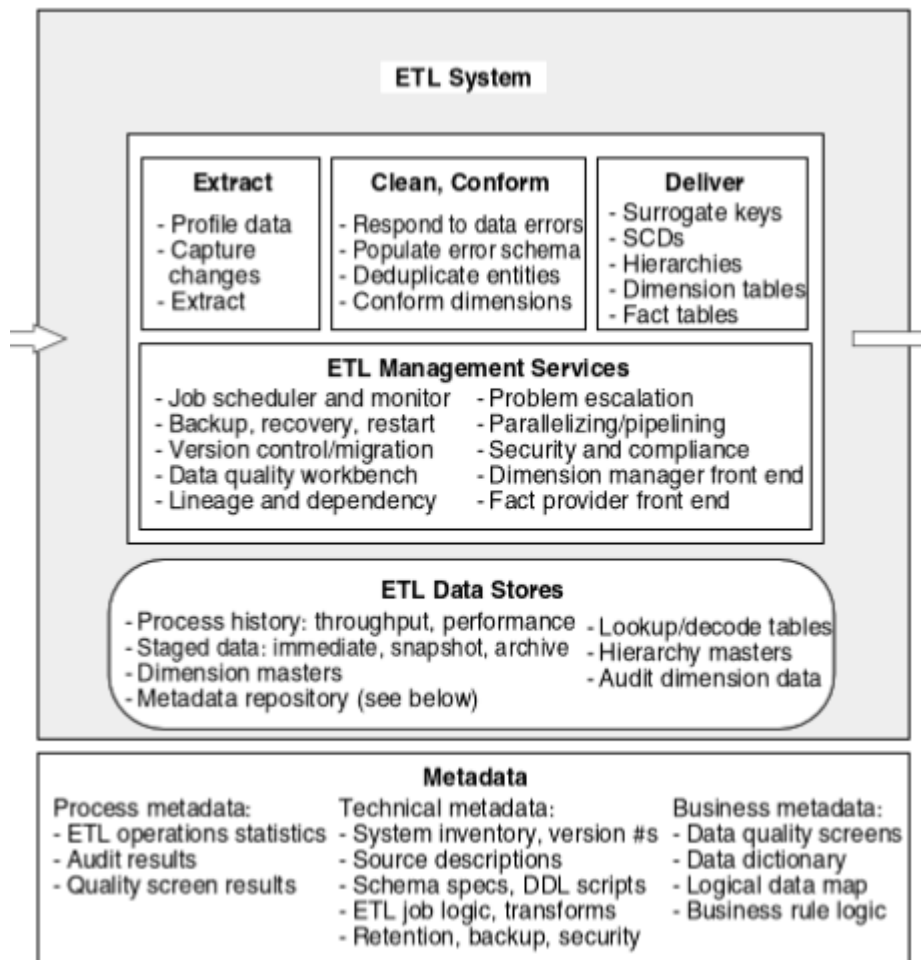


Fig. 4. Arquitectura ETL. [13]

1.3.2.2.1. Extracción.

Tejada [14] explica que la extracción interna o externa de los datos puede llevarse a cabo de dos maneras; ya sea manual o por medio de herramientas de ETL. La extracción manual, requiere programar rutinas a través de lenguajes de programación con el fin de extraer los datos de las fuentes de datos origen. Sin embargo, la opción más productiva es la que proporcionan las herramientas especializadas de ETL, debido a que son prediseñadas para elaboración esta función, y se hace más factible monitorear el proceso.

El desafío en el proceso de extracción de datos es determinar; qué datos extraer y qué clase de filtros aplicar. Campos que pertenecen a varios usuarios, valores que probablemente no existan, transacciones hechas de cuentas que no pueden ser creadas y otros errores. En este punto es necesario entender los requerimientos de los procesos de extracción que pueden determinar cuáles de los siguientes servicios se pueden necesitar [13].

- Perfil de los datos
- Cambiar captura de datos
- Sistema de Extracción.

1.3.2.2.2. Limpieza y Conformación.

La limpieza y conformación de datos son el núcleo de la calidad de los datos en el proceso ETL. En este punto se realizan una serie de transformaciones para convertir los datos en algo valioso y presentable para la organización. Los cinco servicios más importantes, para la limpieza y conformación de datos son: Sistema de limpieza de datos, rastreo de eventos de error, creación de dimensión de auditoria, eliminación de duplicidad y conformación [13].

Prudenciano [15] manifiesta que, al extraer los datos de los sistemas transaccionales, existe la probabilidad de que estos datos no fueran depurados en su totalidad, por lo cual posteriormente deben ser limpiados.

ETL, cuenta con herramientas propias para llevar a cabo esta acción. El proceso de limpieza de los datos previene errores que pueden alterar los datos, debido a varios motivos, entre los más relevantes: valores por defecto, carencia de valor, campos con diversas utilidades, valores crípticos, uso inadecuado de los campos, identificadores que no son exclusivos, entre otros.

Prudenciano [15] añade que el proceso de limpieza cuenta con cinco subprocesos:

- **Depuración:** Se refiere al reconocimiento de manera individual de los componentes de la información a fin de aislarlos en los ficheros destino.
- **Corrección:** Se basa en una corrección en cuanto a los valores individuales de los atributos empleando algoritmos de corrección además de fuentes externas.
- **Estandarización:** En esta etapa de adaptan rutinas de conversión con el objetivo de transformar los valores en formatos definidos además de consistentes, esto se da a través de la aplicación de procedimientos de estandarización definidos previamente por las reglas del negocio.
- **Relación:** Constituye la identificación y relación de valores de los registros, para posteriormente corregirlos y estandarizarlos en base a las reglas de negocio, con la finalidad de suprimir los duplicados.
- **Consolidación:** Examina y reconoce las relaciones, entre los registros relacionados, para luego juntarlos en una sola representación para que pueda continuar al proceso de la transformación.

Muñoz [16], establece que una vez que los datos culminan el proceso de limpieza, empiezan por la transformación, esta etapa realiza tareas como: filtración de datos, cambios de formato, conversión de códigos, y finalmente los valores derivados y/o agregados. En el caso de los valores agregados, como en la suma de las ventas, por lo general se pre calculan y se alojan para de esta manera obtener mayores rendimientos al momento de lanzar las consultas que necesiten el cálculo de totales al DWH. Además, en esta fase, se ajusta el nivel de detalle; es decir, se puede obtener detalle a nivel de líneas de factura en los datos previamente extraídos, sin embargo, en el DWH lo que se almacena son las ventas ya sean semanales o mensuales.

1.3.2.2.3. Integración y Entrega

La data limpia y alineada debe prepararse para para su consumo y entrega en la capa de presentación. Los subsistemas en el *Back End* de un sistema ETL están conformados por:

Administrador de cambio lento de dimensión (SCD : *Slowly Changing Dimension Manager*), Generador de claves subrogadas, Administrador de Jerarquía, Administrador de Dimensiones especiales, Constructor de tablas de hechos, Constructor de cubos OLAP, Administrador de propagación de datos, constructor de agregaciones, sistema proveedor de tablas de hechos, entre otros [13].

En esta etapa, se culmina el proceso, y se refiere a la integración en el DWH; es decir, es el instante en el cual cargamos los datos, posteriormente debe verificarse si coinciden con la información que se encontraba previamente en el sistema transaccional. Además, se debe corroborar que los valores que presentan los registros cargados se corresponden a los definidos en el DWH. Es esencial cerciorarse que el proceso se haya desarrollado de manera correcta, en su defecto existe la posibilidad de llevar a decisiones de tipo erróneas a los usuarios [4].

El proceso de carga de los datos en un DWH, se hace con herramientas especializadas para llevar a cabo acciones, entre ellas: [4]

- **Diseño Gráfico:** Herramienta que brinda acceso a los desarrolladores para que se pueda establecer la relación entre las fuentes de datos, transformaciones, procesos y tareas, con el propósito de desarrollar la carga. Una vez realizado el diseño, se debe alojar en un repositorio *Meta data*.
- **Gestión de Meta data:** Proporciona un repositorio en el cual se define, documenta y se gestiona la información desarrollada en el proceso ETL y su ejecución. Además, esta herramienta debería ser alcanzable desde distintas aplicaciones.
- **Extracción:** Extrae la información a través de conectores como: ODBC65, SQL, los cuales deben acceder al *Meta data* para de esta manera establecer que información deberá ser extraída y que forma.
- **Transformación:** Proporcionan librerías de transformación, así pues, otorgan acceso a los desarrolladores para poder transformar los datos origen en los destinos, implementando los nuevos componentes creando tablas de agregación con el propósito de mejorar la calidad del rendimiento.
- **Carga:** Permite el uso de adaptadores que servirán para agregar o modificar los datos en el DWH.
- **Administración y operación:** Brinda acceso a los administradores a: programar, ejecutar y monitorear las tareas de ETL, resultados, además gestionar errores, recuperar fallos y finalmente agrupar los resultados con los sistemas originales.

1.3.2.2.4. Servicios de Gestión ETL

Estos servicios incluyen: Calendario de tareas, Sistema de respaldos, Recuperación y reinicio, control de versión, control de migración, monitor de flujo de trabajo, Ordenamiento, Ascendencia y Dependencia, Problemas de escalabilidad, paralelismo y canalización, la gestión de la conformidad, seguridad y el repositorio de metadatos [13].

1.3.2.2.5. Almacenes de Datos ETL

Los almacenes de datos son lugares de asentamientos ser temporales o permanentes de los sistemas DWH o BI. El tipo almacén de datos tiene implicaciones en la arquitectura de los

sistemas DWH o BI. Pueden ser: Almacenes de Datos del Sistema ETL, tablas de búsqueda y decodificación, almacenes de datos de calidad.

1.3.2.2.6. Metadatos ETL

Los objetos de la *meta data* del ETL se clasifican en tres categorías: de Procesos tales como las operaciones estáticas, resultados de auditoría o pantallas de resultados de calidad; Técnicos como sistemas de inventario que incluye el número de la versión, descripción de las fuentes, métodos de acceso, etc. y metadatos del negocio, tales como especificaciones de pantallas de calidad de datos, diccionario de datos, mapa lógico de datos, reglas lógicas del negocio [13].

1.3.2.2.7. Data Mart.

Un *data mart*, es un tipo especial de almacén de datos de un área funcional específica del negocio. Corresponde en sí a un subconjunto o pequeños DWH centrados en un tema. Su objetivo es ayudar en la toma de decisiones a las áreas específicas de una organización, Ver Fig. 5.

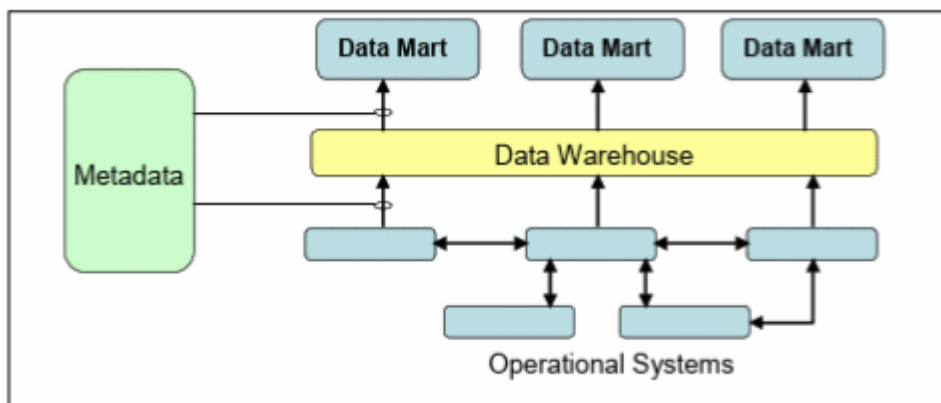


Fig. 5. Data Marts. [17]

Pueden ser explotados por los usuarios de las áreas funcionales de forma diversa de acuerdo a sus necesidades. Las consultas se hacen a través de herramientas OLAP (*On Line Analytical Processing* – Procesamiento Analítico en Línea) el cual ofrece una visión multidimensional de la información.

Para diferenciar un *Data Mart* de un DWH [18] propone el análisis comparativo breve que se indica en la Tabla I.

CARACTERÍSTICA	DWH	Data Mart
Tamaño	Más de 100 Gb.	Menos de 100 GB.
Asunto	Trata Varios temas	Trata un solo tema en específico.
Fuentes	Fuentes Internas y Externas de datos empresariales.	Herramientas específicas para el negocio.
Alcance	Trabaja con múltiples líneas de la empresa.	Un solo lineamiento en la empresa o un departamento multifuncional.

Decisión	Respalda decisiones estratégicas que afectan a toda la empresa.	Apoyo en las decisiones según los objetivos de un solo departamento.
Integración	Todas las integraciones comerciales.	Las que sean necesarias para un solo departamento.

Nota: Análisis de la diferencia entre DWH y Data Mart de acuerdo a sus descripciones. [18]

Un *Data Warehouse* es un repositorio donde se almacena grandes volúmenes de información provenientes de diferentes fuentes, tanto internas como externas, esta información debe ser depurada, mientras que un *Data Mart* es una herramienta de alcance limitado, el cual está orientado a satisfacer las necesidades de un área en específico, o solo a un grupo de usuarios en específico.

Debido a que la implementación de DWH corresponde a una tarea significativa que se desarrolla en períodos largos de tiempo, los *data marts* permiten hacer implementaciones más cortas de áreas específicas de la organización. Estas implementaciones pueden ir evolucionando de forma natural. [19]

El propósito de los *data marts* según Ballard [19] se resume en:

- Brindar acceso rápido a la información para necesidades analíticas específicas.
- Controlar el acceso del usuario final a la información
- Representa la visión del usuario final y la interfaz de datos en el DWH.
- Crear una vista multidimensional de los datos para mejorar el análisis
- Almacenar información cargada previamente para mejorar los tiempos de respuesta d en las consultas.

1.4. ARQUITECTURA DE BI

En el *Front End* de la Arquitectura están los sistemas BI. La arquitectura BI comprende lo que el usuario del negocio puede ver y con lo que puede trabajar el día a día. Podría comprenderse como la interfaz de usuario del DWH. En esta capa de la arquitectura los usuarios desconocen el tiempo, energía y recursos que están por debajo (*Back End*) y el acceso a los datos es complejo. Sin embargo, el Modelo Dimensional ayuda a reducir esa complejidad, pero los negocios abundan en reglas y excepciones que deben ser incluidas en las bodegas de datos de manera que los analistas de negocios puedan entender su impacto. [13]

La coyuntura de los sistemas DWH/BI permiten que el acceso a la información sea posible de manera clara y con formatos entendibles; por lo tanto, es necesario construir aplicaciones de BI que ayuden a buscarla y encontrarla. [13]

La arquitectura de BI (figura 6) muestra el *Front End* de la Arquitectura técnica de los sistemas DWH/BI. En ésta se agrupan todas las diferentes aplicaciones y los servicios BI para desplegar la información para los usuarios y gestionar el medioambiente.

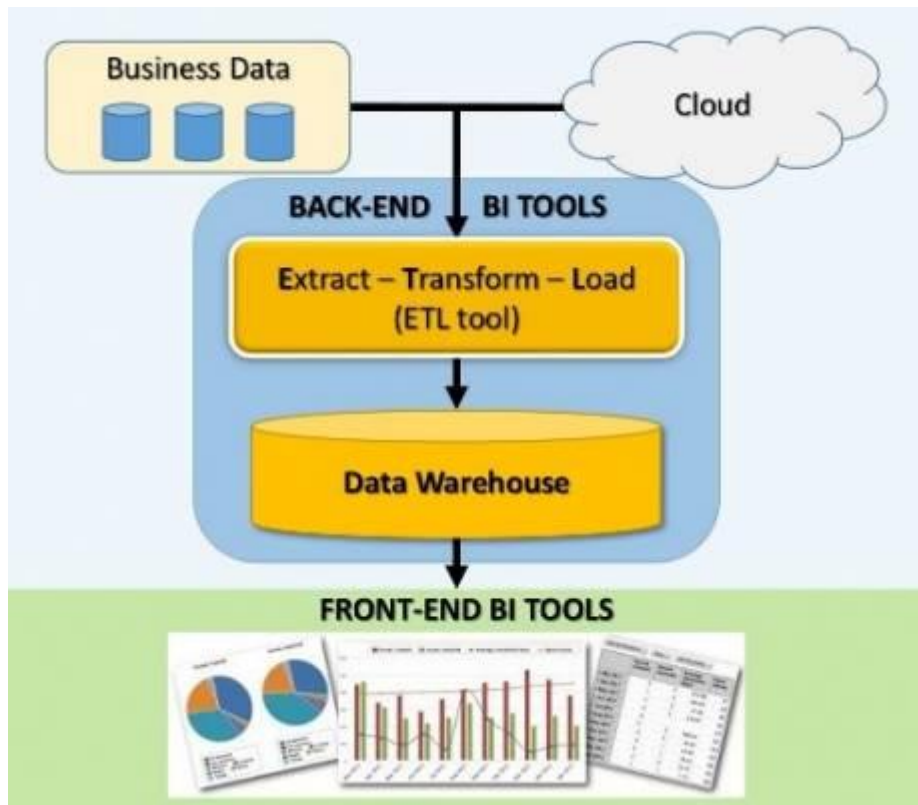


Fig. 6 Front End. [20]

1.4.1. TIPOS DE APLICACIONES DE BI

Existe una amplia gama de aplicaciones que demandan datos desde el servidor de presentación que deben ser diseñadas de forma simétrica, flexible y escalable que soporten diferentes clases de solicitudes simultáneamente. Estas aplicaciones son:

- Herramientas de consulta de escritorio, que permiten dar respuesta a las solicitudes de información de los usuarios del negocio: [13]
- Reportes estándares que proporciona reportes periódicos desde un portal de BI o hojas de cálculo o PDFs de librerías en línea;
- Aplicaciones analíticas que contienen poderosos algoritmos de análisis para consultas de BD relacionales. Incluyen presupuestos, previsión y seguimiento de la actividad empresarial.

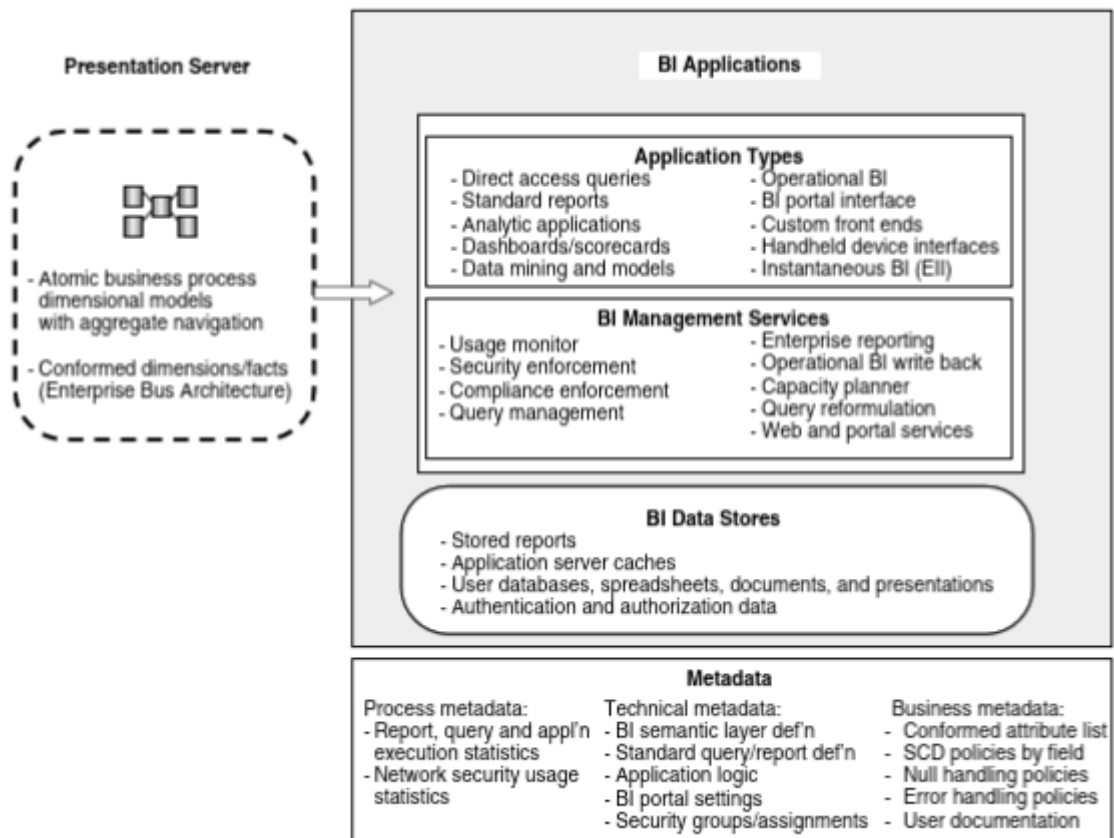


Fig. 7. Tipos de Aplicaciones BI. [13]

- *Dashboards y Scorecards* son interfaces de usuario multiusuario que gestionan indicadores clave de rendimiento (KPIs) a través de gráficos y texto.
- Minería de Datos y modelos, proporcionan análisis exploratorio de grandes conjuntos de observación, generalmente descargados del DWH al software de minería de datos. La minería de datos usada también para crear modelos usados por aplicaciones analíticas y operativas.
- BI Operacional, permite consultas en tiempo real de estado operativo, frecuentemente acompañadas de interface de transacciones de escritura.

Estas aplicaciones llegan a los usuarios de negocio a través de aplicaciones como: [13]

- Portales de BI e interfaces personalizadas, proporcionan fácil acceso a aplicaciones BI basadas en la web o consultas complejas específicas y resultados de pantalla
- interfaces de dispositivos portátiles; diseñadas para pantallas portátiles y dispositivos de entrada.
- BI instantáneo, con una arquitectura de datos en tiempo real y conexión directa desde el sistema de transacciones de la pantalla de origen del usuario.

1.4.1.1. Gestión de Servicios de BI. [13]

La demanda de clientes ligeros, los costos de los sistemas DWH/BI y el impulso que ha tenido la industria de las tecnologías de información para migrar a las arquitecturas orientadas a servicios SOA, han hecho que el acceso a los datos, pasen de los equipos de escritorio a las aplicaciones. Los sistemas DWH/BI.

La gestión de los servicios de BI abarca desde los servicios compartidos que residen en el servidor de presentación hasta los servicios de escritorio que normalmente se presentan a nivel de usuario u que pertenecen a la definición de informes y visualización de resultados: Servicios compartidos, Servicios de seguridad, servicios de metadatos, monitoreo de uso, gestión de consultas, servicios de reportes empresariales, acceso web, servicios del portal.

1.4.1.2. Almacenes de datos de BI. [13]

Aunque a simple vista no debe haber un almacén de datos en el *Front End* del sistema DWH/BI, los datos que residen en el servidor de presentación, se recuperan cuando sea necesario un análisis o informe. En realidad, cuando la respuesta configurada para una solicitud de datos específica sale del servidor de presentación, generalmente termina en el navegador del usuario y puede guardarse como parte del archivo del informe.

Alternativamente, el conjunto de resultados puede introducirse en un almacén analítico local como un motor OLAP, Acces o Excel puede terminar en una aplicación de BI.

En este grupo se encuentran: Informes Almacenados, Cachés del Servidor de Aplicaciones (cachés locales de los servicios orientados a datos), Bases de Datos de Usuario Local, Almacenes de Datos Analíticos Temporales, Resultados de Aplicaciones Analíticas, Sistemas descendentes, seguridad del almacén de datos.

1.5. MINERÍA DE DATOS (DATA MINING)

Data mining hace referencia a una etapa que consiste en compilar los beneficios de varias áreas dentro de un proceso mayor denominado extracción de conocimiento en bases de datos en un entorno comercial. Dicho de otra manera, *data mining* es considerado una tecnología que se encuentra estructurada por fases que incluye varias áreas. En la actualidad existen diversas aplicaciones o herramientas comerciales de *data mining* de gran potencial las cuales contiene un sinnúmero de utilidades, que tienen la función de que el desarrollo de un proyecto sea de una manera factible. [21]

Aplicaciones de uso.

En general el proceso de *data mining* se estructura de cuatro etapas: [21]

- Determinación de objetivos.
- Pre procesamiento de datos.
- Determinación del modelo.
- Análisis de resultados.

1.6. OLAP (PROCESAMIENTO ANALÍTICO EN LÍNEA)

Procesamiento analítico en línea El procesamiento analítico en línea (OLAP) es una tecnología clave en el almacenamiento de datos. La funcionalidad OLAP se caracteriza por un análisis multidimensional dinámico de datos empresariales consolidados que respaldan las actividades analíticas y de navegación del usuario final. Proporciona una interfaz intuitiva y fácil de usar para los usuarios comerciales y puede procesar los datos de manera muy eficiente. [19]

Algunas de las capacidades funcionales de OLAP se resumen en: [19]

- Cálculos y modelado aplicados en todas las dimensiones, a través de jerarquías y/o entre miembros.
- Análisis de tendencias en períodos de tiempo secuenciales
- Seccionar subconjuntos para visualización en pantalla
- Desglosar a niveles más profundos de consolidación
- Alcance a datos detallados subyacentes
- Rotación a nuevas comparaciones dimensionales en el área de visualización

Algunas variantes de OLAP comunes son: [19]

- MOLAP: se refiere a un OLAP multidimensional. Aquí, la base de datos se almacena en una estructura especial, típicamente propietaria, que está optimizada (a través del cálculo previo) para un tiempo de respuesta de consulta muy rápido y un análisis multidimensional. Tiene limitaciones en lo que respecta a la escalabilidad y es posible que no permita la actualización.
- ROLAP: Significa OLAP relacional. Aquí, el modelo de base de datos también es multidimensional MOLAP, pero se utiliza una base de datos relacional estándar y el modelo de datos puede ser un esquema de estrella o un esquema de copo de nieve. Proporcionando un tiempo de respuesta rápido a la consulta, pero eso se rige en gran medida por la complejidad del SQL utilizado, así como por el número y tamaño de las tablas que deben unirse para satisfacer la consulta. Su ventaja está en la escalabilidad significativa lograda ya que está alojada en una base de datos relacional estándar.
- HOLAP: habilita una versión híbrida de OLAP. Es un híbrido de ROLAP y MOLAP. Se puede pensar en una base de datos HOLAP como una base de datos virtual en la que los niveles superiores de la base de datos se implementan como MOLAP y los niveles inferiores de la base de datos como ROLAP.

1.6.1. MODELOS DE DATOS

Los modelos de datos que típicamente se utilizan para diseñar bases de datos, *data marts* y *data warehouses* se basan en diferentes tecnologías que están destinadas a proporcionar el tipo de soporte de acceso a datos, organización y rendimiento, deseado en una situación

particular. Los más comunes son: Esquema de Estrella (desnormalizado) y Esquema de Copo de Nieve y Esquema Normalizado - 3FN) [19].

La elección del modelo a usar depende de: el rendimiento, la facilidad para comprender y navegar por el modelo de datos.

1.6.1.1. Esquema Estrella

Es muy común en los *Data Marts* y DWH debido a que proporciona mejor rendimiento en las consultas que el modelo relacional. Es de fácil comprensión. Como ya se mencionó en el apartado (TABLA I) consta de una tabla de Hechos y una serie de tablas con datos descriptivos que se relacionan con la tabla central denominadas Tablas de Dimensiones. Ver figura 8.

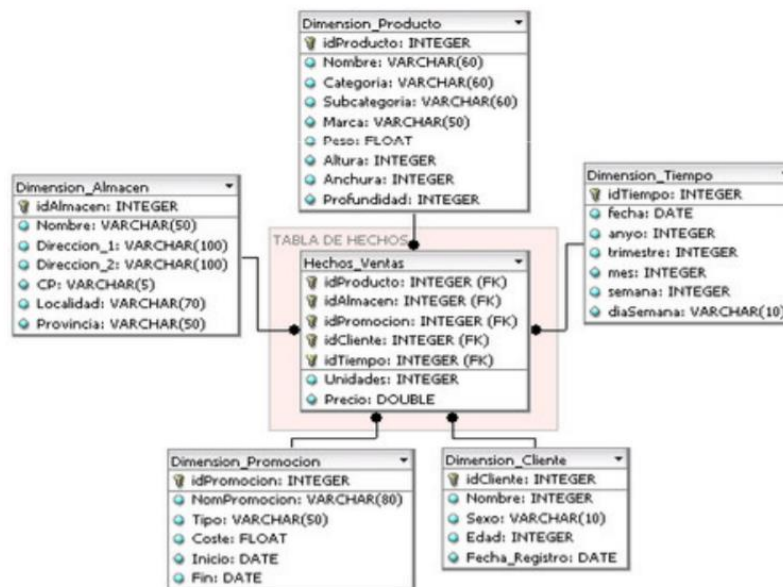


Fig. 8. Esquema en Estrella.

Los elementos del esquema estrella son:

- Tabla de Hechos. Los hechos corresponden a las transacciones o eventos; que se refieren a los datos de contexto y las métricas.
- Tabla de Dimensiones. Colección de miembros o unidades que describen los datos de hechos desde un punto de vista particular. Cada tabla de hechos está ligada a las dimensiones a través de una clave principal, la que puede ser subrogada, natural o combinada.
- Métricas. Es un atributo numérico de un hecho, que representa el desempeño o comportamiento de la empresa en relación con las dimensiones.

1.6.1.2. Esquema Copo de Nieve

Corresponde a un modelo estrella expandido que se forma precisamente cuando hay la necesidad de definir dimensiones de las dimensiones a través de una clave primaria que las relacione. (ver figura 9)

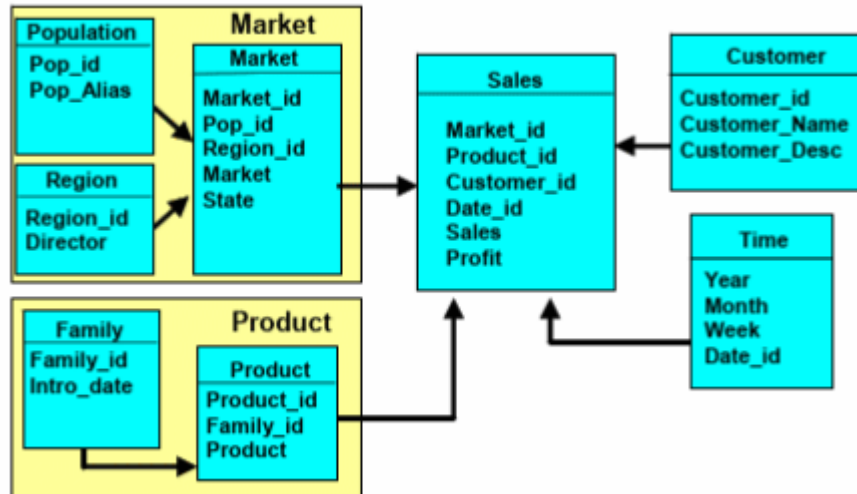


Fig. 9. Esquema Copo de Nieve. [17]

El esquema copo de nieve se puede crear cuando es necesario definir niveles de granularidad de las entidades de las dimensiones y cuando es más probable que las dos entidades sean suministradas por un sistema fuente diferente (Ver Figura 10). [19]

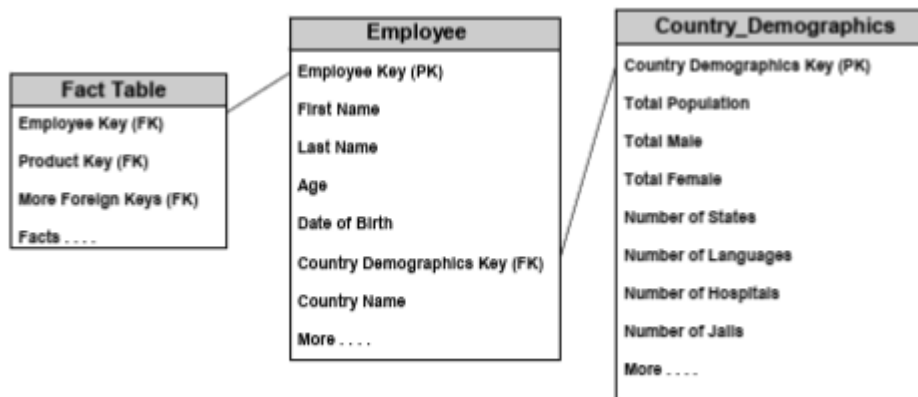


Fig. 10. Esquema. [17]

1.6.1.3. Normalizado – 3FN

Usada en el procesamiento OLTP (*OnLine Transaction Processing* – Procesamiento de Transacciones en línea). Una tabla está en 3FN:

- Si cada columna NO-CLAVE es independiente de las otras columnas NO-CLAVE y depende de una sola clave.
- Otra forma de definirla es “La clave, la clave completa y nada más que la clave”

Un esquema en 3FN puede verse como el de la Figura 11.



Fig. 11. Esquema en FN. [17]

1.6.2. ENFOQUES PARA DESARROLLO DE APLICACIONES BI

Antes de plantar una metodología para aplicarla en el presente trabajo, es necesario conocer los conceptos y métodos establecidos por Ralph Kimball y Bill Inmon ya que son los enfoques más ampliamente discutidos, es decir, cuál de estas dos metodologías es la más eficaz para el uso de las empresas. Hasta la actualidad no existe una respuesta establecida ya que ambos métodos presentan sus beneficios y desventajas, para elegir el uso de una de las dos metodologías varía en función de las necesidades y requerimientos de la empresa

1.6.2.1. Enfoques de diseño del almacén de datos

Antes de establecer el ciclo de vida del proyecto de Inteligencia de Negocios, es necesario comprender las metodologías diseño del almacén de datos o DWH desde los enfoques propuestos por Ralph Kimball y Bill Inmon. La elección del enfoque depende de las necesidades y requerimientos de la organización.

Kimball propone un enfoque ascendente de diseño de arquitectura de datos. En efecto, su propuesta consiste en definir primero los *Data Marts* basados en los requisitos del negocio que luego se irán cargando en un modelo dimensional desnormalizado.

El modelo dimensional desnormalizado por naturaleza está compuesto por una tabla de hechos (tabla central) y varias tablas de dimensiones (modelo estrella o copo de nieve). Cada *data mart* puede integrarse por dimensiones de datos conformadas. Estas dimensiones conformadas son como tablas de dimensiones compartidas entre diferentes tablas de hechos dentro del almacén o como las mismas tablas de dimensiones en varios *data marts* [9], de modo que se garantice la integridad y consistencia de la información.

Kimball también propone una herramienta de diseño de “matriz de bus empresarial de Kimball” que es frecuentemente usada por los equipos de gestión empresarial para priorizar qué dimensiones conformadas deben implementarse primero.

Es un enfoque **Táctico** se deriva de la visión de Kimball, quien prescribe un enfoque dimensional de abajo hacia arriba para BI basado en *data marts* funcionales que finalmente se agregan al DWH corporativo.

La Figura 12 indica la arquitectura básica de almacenamiento propuesta por Kimball

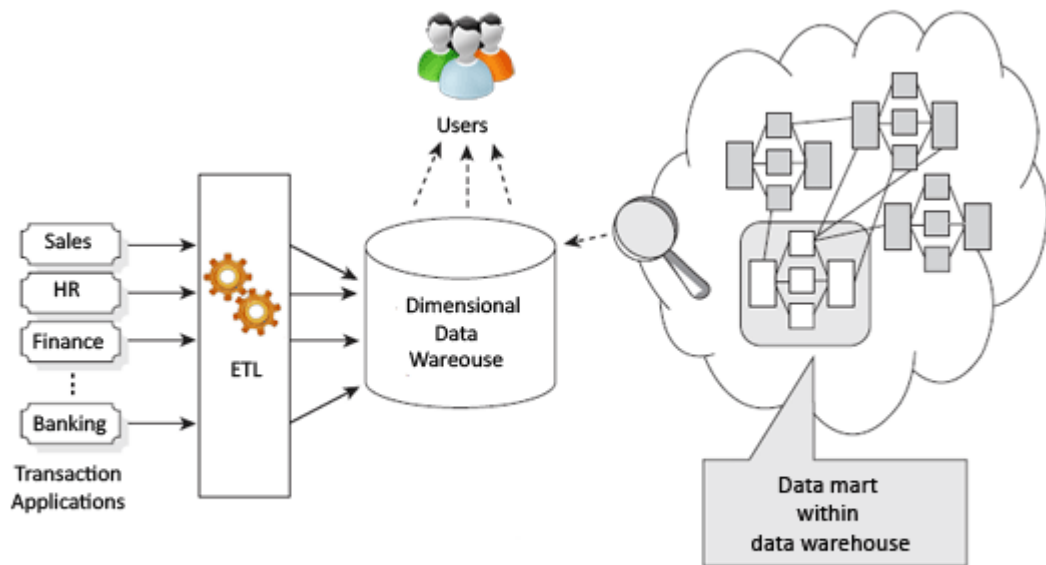


Fig. 12. Arquitectura de almacenamiento de Kimball. [13]

Para Bill Inmon el desarrollo del almacén de datos debe comenzar con el diseño del modelo de datos del almacén corporativo, que incluya las principales funciones y entidades de la organización y a partir de este se crea un modelo lógico para el producto con todos los atributos asociados a esa entidad. [9]

Este enfoque usa la forma normalizada en la construcción de la estructura de cada entidad, de esta forma se evita, en lo posible, la redundancia de los datos. Este enfoque integra los *data marts* de cada área funcional creados por separado garantizando integridad y coherencia, debido a que la única fuente de datos es el almacén corporativo. [9] La Figura 13 ilustra la arquitectura básica de almacenamiento de datos propuesta por Inmon.

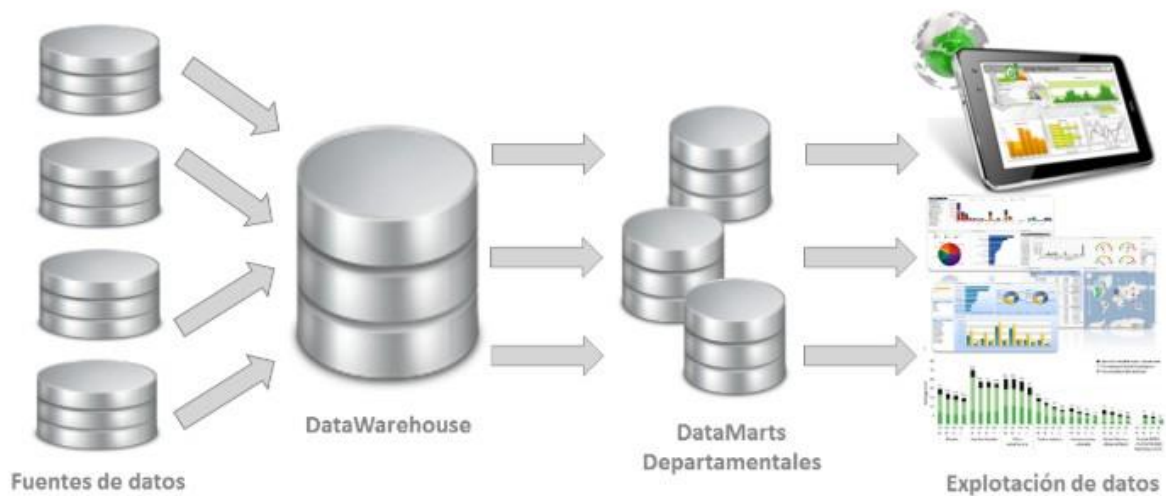


Fig. 13. Arquitectura de almacenamiento de Inmon. [22]

Este enfoque Inmon es consistente con un enfoque estratégico que según Davenport y Harris (2007) y el *Gartner's Business Intelligence and Performance Management Framework* en el que se define “un simple repositorio organizacional de datos de toda la empresa en varias líneas del negocio y sujeto a áreas que contienen masivamente datos

Es **estratégico** de arriba hacia-abajo prescrito por Dave & Harris y el marco de referencia de Gartner como: Un único repositorio organizacional holístico, de muchas líneas de negocios y áreas temáticas que contiene datos masivos e integrados; representa la visión organizacional completa de la información necesaria para administrar y comprender el negocio.

1.6.2.2. Evaluación de la metodología de diseño del almacén de datos

A la hora de elegir la metodología para desplegar soluciones BI, es necesario preguntarse ¿Qué metodología permitirá un enfoque más ágil de la inteligencia de negocios en organizaciones con limitaciones de recursos y experiencia? . [23]

Las organizaciones que buscan implementar ágiles niveles de BI, pero necesitan medios efectivos para evaluar el estado actual y la madurez de sus prácticas de gestión de la información empresarial para identificar y abordar las brechas críticas que podrían inhibir los retornos deseados en inversiones exitosas. [23]

Las organizaciones con limitaciones de recursos y experiencia necesitan una metodología que resuelva la dicotomía entre contar con un BI estratégico de escala empresarial y un perfil más ágil y de menor costo de las BI centradas en áreas funcionales específicas. [23]

Un reporte de Gartner publicado en 2019 enfatiza la necesidad de marcos de referencia estructurados que ayude a los líderes de negocios y de TI a “diseñar un plan más completo para BI, análisis y gestión del rendimiento (PM) que alinee las personas, los procesos y las aplicaciones y, en última instancia, produzca mejores rendimientos “ [24]; es decir que cada organización debe hacer configuraciones apropiadas (hoja de ruta) en función de sus propios objetivos estratégicos y circunstancias comerciales. [23]

La selección entre la necesidad de un enfoque estratégico (Inmon) o Táctico se basa en el análisis de las características que diferencian a uno y de acuerdo a la necesidad de la organización. La Tabla II. sintetiza las características favorables:

TABLA I
KIMBALL VS. INMON

CARACTERÍSTICA	FAVORECE KIMBALL	FAVORECE INMON
Naturaleza de los requisitos de apoyo a la toma de decisiones de la organización	Táctico	Estratégico
Requerimientos de Integración de Datos	Áreas individuales del negocio	Integración a lo largo de toda la organización
Estructura de Datos	Métricas de Negocio, medidas de rendimiento y Scorecards (Registros de puntuación)	datos no métricos y datos que se aplicarán para satisfacer múltiples y variadas necesidades de información.
Escalabilidad	Necesidad de adaptarse a necesidades altamente volátiles dentro de un alcance limitado.	El alcance creciente y los requisitos cambiantes son fundamentales.
Persistencia de los datos	Sistemas fuente son relativamente estables	Alto grado de cambio de las fuentes de datos
Requerimientos de personal y habilidades	equipos pequeños de generalistas	Equipos grandes de especialistas
Tiempo de entrega	Necesidad de la primera aplicación de Data Warehouse es urgente	Los requisitos organizacionales permiten un tiempo de puesta en marcha más largo
Costos de despliegue	Menores costos de puesta en marcha, con cada proyecto subsiguiente costando aproximadamente lo mismo	Mayores costos de puesta en marcha, con menores costos de desarrollo de proyectos posteriores.

Nota: Directrices del apoyo de Kimball y FAVORECE INMON con sus características. [23]

De acuerdo con el análisis comparativo de [25] en la tabla I; la propuesta de Kimball es ágil y permite reducir costos en la puesta en marcha y despliegue más rápido de la solución de BI táctica que está enfocada en un área funcional específica del negocio. Sin embargo, para los críticos, éste enfoque tiende a carecer de integridad debido a la falta de enfoque estratégico.

Por su parte, la propuesta de Inmon está orientada precisamente para lograr integridad empresarial, aunque los costos de puesta en marcha son mayores, los tiempos de despliegue más largos y suele ser rígido y difícil de implementar, especialmente para empresas con habilidades, recursos y tiempo limitados.

1.6.3. BENEFICIOS DE BI.

Los beneficios que nos proporciona unas soluciones de Inteligencia de negocios son diversos, por lo cual se hace factible para una mejor comprensión clasificarlos: [26]

1.6.3.1. Manejo del crecimiento de la empresa.

Toda empresa tiene como principal objetivo evolucionar, es decir crecer a través de los cambios que se vayan implicando. Lo cual la Inteligencia de negocios es una solución fundamental para monitorear que tan eficaces son los procesos para enfrentar dichos cambios y las necesidades puntuales de la empresa.

1.6.3.2. Control de costos.

El control de los costos en un negocio es usualmente el que obliga a la empresa a considerar el uso de una solución de inteligencia de negocios, para que de esta manera se pueda medir los gastos y posteriormente ver esto a un nivel de detalle que identifique la línea de negocio, producto, centro de costo, etc.

1.6.3.3. Clientes.

Todas las empresas reúnen una infinita cantidad de información de suma importancia relacionada a los clientes. El objetivo de las empresas es transformar esta información en conocimiento y posteriormente este conocimiento es dirigido a una gestión comercial que simbolice un tipo de ganancia para la empresa. Las ventas, cruzadas e incrementales junto a la retención de clientes, anticipos de oportunidades, identificación de patrones de compra y el comportamiento de los clientes hacen de BI un elemento esencial para lograr los objetivos de la empresa.

1.6.3.4. Indicadores de gestión

Estos indicadores representan las medidas dirigidas al desempeño organizacional con la capacidad de representar la estrategia organizacional en: objetivos, métricas, iniciativas y tareas enfocadas a un grupo o individuos en la organización.

En las capacidades funcionales de indicadores incluyen:

Monitoreo. Es decir, monitorea procesos cruciales de negocio además de las actividades esto se lleva a cabo con el uso de métricas que dan una alerta acerca de un problema potencial.

Análisis. Hacer un análisis del origen de los problemas, inspeccionando la información desde diversas perspectivas en múltiples niveles de detalles.

Administración. En este punto se administra los recursos y procesos para dirigir la toma de decisiones, además de optimizar el desempeño. Posteriormente podremos obtener una visión global de la empresa con una capacidad de dirigir la organización en la dirección adecuada.

1.6.4. METODOLOGÍA DE IMPLEMENTACIÓN DE BI

En función de la Evaluación de la metodología de diseño del almacén de datos realizada en el apartado 1.6.2.2. y en base a la necesidad de la EMAPAL que ha llevado a la realización del presente trabajo, la cual es analizada en el capítulo II; es apropiado basar la metodología de trabajo en la propuesta de Kimball.

En este sentido, una metodología ágil, como SCRUM, para el desarrollo de soluciones de software, bien puede ser adaptada en proyectos de DWH y consecuentemente para soluciones de BI. Esto se debe a que las metodologías ágiles tienden a ser usadas por organizaciones pequeñas y departamentos, permiten entregas rápidas a necesidades urgentes en aplicaciones analíticas y de *reporting*.

“En la práctica, los métodos ágiles aplicados correctamente a grandes proyectos de integración de datos y visualización de información han reducido las horas de desarrollo necesarias y los defectos de codificación a cero” [27].

Los métodos ágiles para proyectos de BI contribuyen en gran medida a resolver las necesidades de información de los usuarios. Sin embargo, puesto que las soluciones BI están ligadas a los proyectos de DWH, que requieren frecuentemente una integración significativa de datos, estos métodos pueden no estar preparados para cumplir con este desafío. [27]

Para evitar que el fracaso en las primeras iteraciones será necesario hacer una adaptación del método para respaldar de mejor manera el trabajo de integración. [27]

No es parte de este trabajo explicar sobre SRUM; sin embargo, la Figura 14 ilustra cómo trabaja este método y los roles que deben cumplir.

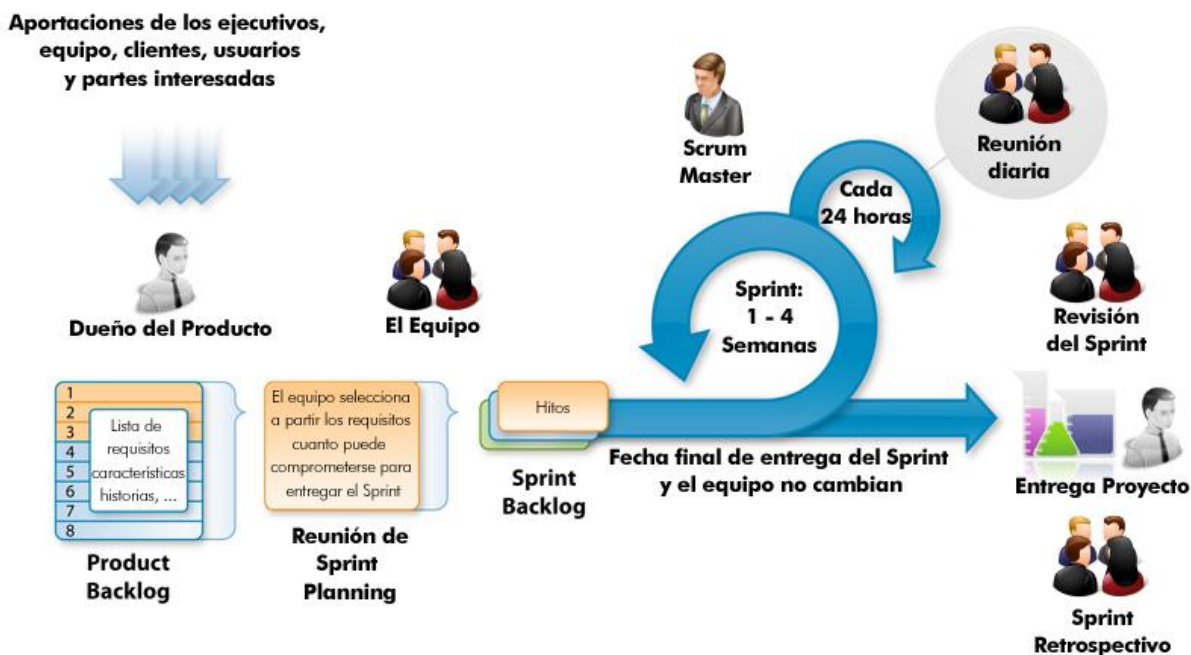


Fig. 14. Metodología Scrum. [28]

El presente trabajo toma a SCRUM como el marco de trabajo para el desarrollo del primer sprint de la solución de BI que se plantea. La estructura del trabajo ha sido adecuada en base al trabajo de titulación denominado: Metodología de desarrollo de proyecto de Inteligencia de Negocios desarrollado por Elka Buitrago de la Universidad Andrés Bello. [29]

La metodología que plantea Buitrago [29], es una adaptación de las fases de ejecución de un proyecto como: Levantamiento de información, Diseño, Desarrollo, Pruebas, Puesta en

Marcha, Entrenamiento y Cierre. Su enfoque adaptado considera la construcción de la solución de BI en varias etapas o iteraciones, precisamente tal como lo plantea el SCRUM (ver Figura 14). Este enfoque asegura que los requerimientos a corto plazo se visibilicen en versiones del producto (solución de BI) y que al mismo tiempo se vayan incorporando nuevas funcionalidades que podrían ir complementando las necesidades de áreas del negocio relacionadas [29].

La idea de Elka Buitrago [29] es tener una visión global de las necesidades del negocio, las cuales se vayan organizando en etapas que precisen las actividades requeridas. Estas fases son: *Definición, Levantamiento de información, Diseño, Desarrollo, Pruebas, Puesta en Producción, Entrenamiento y Cierre*, y se describen en los siguientes apartados:

1.6.4.1. Definición.

En esta etapa se debe concretar el inicio del proyecto. Se desarrolla en base a una reunión, en la cual se incluye a todos los involucrados y tiene por objeto dar a conocer sobre el inicio de proyecto (ver Anexo D). Es importante que en esta etapa se defina el alcance, las funcionalidades, áreas, procesos y por último objetos del negocio que se van a desarrollar. Además, es necesario conformar el equipo de trabajo, identificando los principales involucrados, tanto el personal de negocio, como el personal técnico. Por último, en esta etapa se debe definir el plan de trabajo, lo cual hace referencia a la elaboración de una guía la cual se ampliará al momento del levantamiento de información [29].

Es importante documentar esta etapa con la Aceptación formal del proyecto, documentar el alcance, las especificaciones del equipo técnico y el plan de trabajo [29].

1.6.4.2. Levantamiento de información.

En esta etapa se debe realizar diversas actividades con el fin de conocer el modelo de negocio y los requerimientos de información en las áreas específicas. Se debe llevar a cabo la identificación de áreas y procesos del negocio, para conocer cómo funciona el negocio. Se involucra el levantamiento de la infraestructura tecnológica, el cual consiste en identificar las facilidades tecnológicas que se tiene para la implementación de la plataforma de BI. También se debe determinar las aplicaciones operacionales, las cuales suministran información a las bases de datos de la organización. Asimismo, se deberá desarrollar la identificación de los requerimientos de la información, en la organización cada área requiere de información por lo cual en esta etapa se identificarán cuáles son esos requerimientos [29].

La identificación de requerimientos de la aplicación, requiere la definición de *cómo* los usuarios desean acceder a dicha información y *qué* funcionalidad debe tener la aplicación. También se definirán los requerimientos no funcionales como: el tiempo, seguridad, rendimientos, capacidad, escalabilidad y actualizaciones. Por último, se debe realizar un análisis de las fuentes de datos y los requerimientos para así tener en cuenta la disponibilidad de la información [29].

Los entregables en esta etapa se resumen en: documentación de las áreas del negocio y sus responsables, documentación de las facilidades de la infraestructura tecnológica y

recomendaciones de cambio si es necesario, documentación de riesgos y limitaciones, documentación de las aplicaciones, bases de datos y otras fuentes de datos [29].

La identificación de requerimientos funcionales y no funcionales se documentarán incluyendo información de los usuarios, la especificación de los requerimientos de reportes o consultas, especificación de los almacenes de datos históricos, casos de uso, documentación de fuentes de datos, el documento de los requerimientos no funcionales y la identificación de la disponibilidad de la información, entre otros [28].

1.6.4.3. Diseño.

En esta etapa se agrupan las sub-tareas de diseño de los componentes que estructuran BI, en función del resultado del levantamiento de información. Los entregables que se requieren en esta fase son [28]:

- Diseño del modelo del negocio, que incluya las áreas del negocio, los objetos de información que se manejan y su interrelación.
- Diseño del *Data warehouse* el cual se hace a partir de las necesidades de información que requieren los usuarios y en base a la disponibilidad de los datos.
- Diseño del ETL esto con el fin de poblar el *Data warehouse* con los datos que se encuentran en las distintas fuentes de datos de la organización.
- Diseño de reportes y otras interfaces de análisis. Se puede incluir el diseño de los reportes tabulares y gráficos, como también las interfaces para el análisis de datos y esquemas gráficos del negocio.
- Diseño de indicadores, se debe diseñar cuales son los indicadores de gestión que se quieren evidenciar para la toma de decisiones. Deben orientarse a medir el cumplimiento de las metas de negocio de acuerdo a los objetivos estratégicos.
- Diseño del modelo de análisis o diseño de OLAP, el cual incluye el diseño de los esquemas que se usan en la visualización de la información mediante reportes dinámicos y análisis.
- Diseño de la aplicación, en la que se integran los reportes para el análisis y visualización de los indicadores previamente definidos. Debe incluir la arquitectura de la información, así como también la navegación y el diseño de los componentes que se requieren para su implementación de las funcionalidades necesarias.
- Por último, el *Capacity Planning*, en donde se determina los recursos que posee el hardware y software teniendo en cuenta los requerimientos de almacenamiento de datos, el número de usuarios, demanda de recursos de la aplicación.

1.6.4.4. Desarrollo.

Corresponde propiamente a la elaboración de la solución, incluyendo los diferentes componentes de la misma, basada en el diseño previamente establecido. Esta etapa se abarca [28]:

Instalación de la infraestructura de desarrollo, la cual hace referencia a la instalación y configuración de software, herramientas y los equipos a ser utilizados en el desarrollo. La

construcción del *Data warehouse*, que consiste en crear la base de datos asociada al *Data warehouse*. La construcción de procedimientos ETL; es decir, se desarrollan los procesos que van a permitir la extracción de datos desde las bases de datos fuente, su transformación y posterior carga al *Data warehouse*. La construcción de la aplicación, contempla el desarrollo de los elementos de software previamente diseñados para la implementación de las funcionalidades. La construcción de esquemas OLAP, consiste en elaborar los esquemas OLAP, que van a ser utilizados en el análisis de los datos mediante las interfaces graficas establecidas. La construcción de reportes y otras interfaces de visualización, se elaboran en función del diseño previamente realizado.

Se deberá además desarrollar documentación técnica, se debe abarcar las especificaciones para la posterior instalación y configuración del ambiente en el cual se implementará la solución de BI.

1.6.4.5. Pruebas

En esta etapa se ejecuta tareas las cuales posteriormente permiten verificar el correcto funcionamiento de todos los componentes de la solución de BI. Este proceso incluye [28]:

- La definición del plan de pruebas, en el cual se debe realizar dos tipos de pruebas: integradas y formales dirigidas a cada uno de los elementos de BI.
- La instalación del ambiente de pruebas, aquí se valida la instalación y configuración requerida de los equipos, así como también del software previa a su puesta en producción, con lo cual se asegura la solución de conflictos.
- La verificación del *Data warehouse*, para corroborar que el *Data warehouse* elaborado cumpla con todos los requerimientos que fueron establecidos anteriormente.
- Pruebas de procedimientos ETL, aquí se debe verificar la ejecución de los programas de extracción de las fuentes de datos y su posterior carga al *Data warehouse*.
- Pruebas de reportes e interfaces de visualización, para verificar que todos los reportes e interfaces cumplan con los requerimientos establecidos por los usuarios.

1.6.4.6. Puesta en producción.

Cuando las pruebas fueron exitosas en su totalidad se procede a la puesta en producción de la plataforma, para lo cual se requiere la previa preparación e instalación del software requerido, así como también los productos propiamente desarrollados para la plataforma [28].

En este proceso se abarca dos subprocesos: Preparación del ambiente de producción, se incluye la instalación y configuración de los equipos en el ambiente de producción para de esta manera empezar la puesta en marcha de la solución de BI. También en esta etapa se abarca la verificación del ambiente de producción, luego de que fue realizada a instalación se debe realizar una verificación, para de esta manera asegurarnos que el ambiente como la solución de BI, se encuentre en su correcto funcionamiento [28].

1.6.5. HERRAMIENTAS PARA EL DESARROLLO DEL PROYECTO

Para el desarrollo de este proyecto se hace un análisis de todas las herramientas que serán utilizadas, respaldadas por diferentes consultoras y de investigación de tecnologías de la información en el cual perfilan el ranking de las mejores herramientas, en este apartado se detalla cada una de estas herramientas.

1.6.5.1. Power Designer.

Tiene la capacidad para modelar Datos y la gestión de metadatos nivel físico y conceptual, esta herramienta puede alinearse a los requisitos, el lenguaje y los modelos de datos de las empresas, SAP es una de las empresas líderes en el cuadrante mágico de Gartner para la solución de metadatos en el 2019, ya que permite visualizar y manejar de manera ordenada los datos [30].



Fig. 15. Cuadrante Mágico de Garter de las Herramientas de Calidad de los Datos. [31]

1.6.5.2. Exasol

Es una herramienta que está orientada a trabajar con grandes cantidades de datos, la cual tiene un gran rendimiento en memoria permitiendo hacer un seguimiento de los cambios, además se integra con conectores JDBC con otros sistemas, con herramientas de análisis y BI tales como *Tableau*, cuenta con buen rendimiento y una gran estabilidad [32].

1.6.5.3. Exaplus.

Exaplus es una herramienta para la ejecución de scripts SQL mediante la interfaz de líneas de comando, está implementado en Java y se puede utilizar en varios Sistemas Operativos, en la actualidad dispone de tres interfaces estandarizadas para las que *exasol* dispone de controladores: ODBC, JDBC, ADO.NET [33].

Andreas Wesselmann [34], vicepresidente de Big Data de Productos e Innovaciones SAP, dijo “SAP, es la clave para la construcción de una base de datos confiable para ejecutar sistemas operativos y analíticos inteligentes, generando mejores resultados Este reconocimiento de Gartner afirma que *SAP Power Designer* engloba la gestión de metadatos empresariales y ofrece una orquestación de datos ágil.”

1.6.5.4. PostgreSQL.

Es un sistema de base de datos que relaciona los objetos, es *Open Source* que trabaja con lenguaje SQL, con una escalabilidad segura, *PostgreSQL* posee una arquitectura probada, confiable y tiene integridad de datos, ofrece recursos de manera consistente. [35]

1.6.5.5. Knime.

Knime Analytics Platform (ver Figura 16), es un software *Open Source* orientado a la construcción de ciencias de datos, donde se puede crear flujos de trabajo con una interfaz gráfica intuitiva, donde solo es necesario arrastrar y soltar, sin verse en la necesidad de utilizar cierto tipo de codificación, se puede combinar datos de varias fuentes como por ejemplo en formatos de texto simples como: CSV, PDF, XLS, JSON, entre otros, también se puede conectar a diferentes bases de datos como: Oracle, Microsoft SQL, PostgreSQL, etc., además con esta plataforma se puede agregar, clasificar, filtrar y combinar datos, extraer y limpiar datos [36].

Knime dentro del cuadrante mágico de Gartner se ubica en el cuadrante de Visionarios en el año 2019. Esta firma lo considera como una herramienta de innovación, el enfoque visionario de *Knime* para la analítica brinda flexibilidad según las necesidades de los usuarios [37].

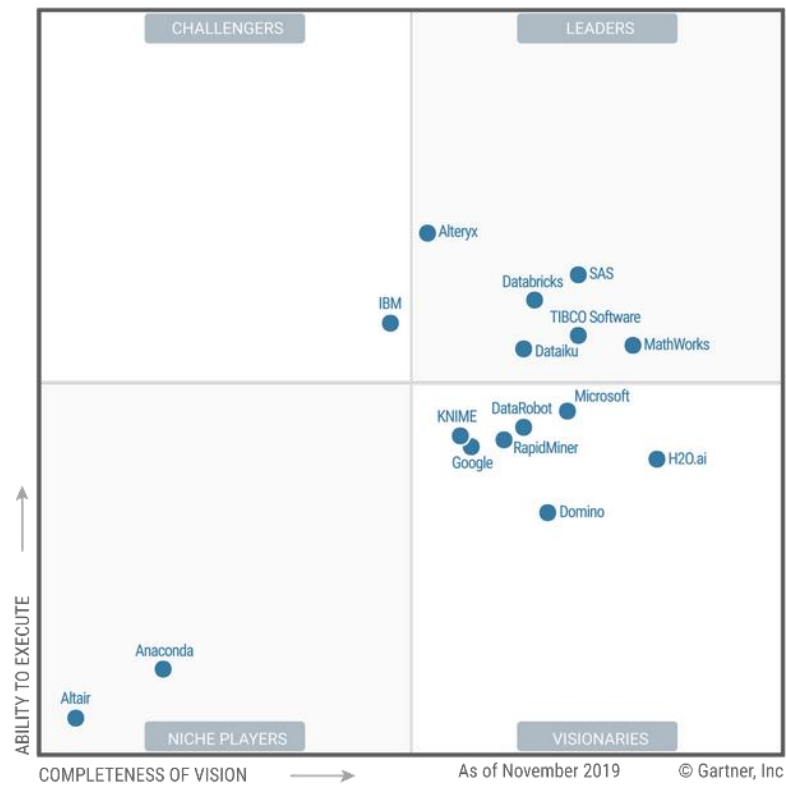


Fig. 16. Cuadrante mágico para plataformas de ciencia de datos y aprendizaje automático. [37]

1.6.5.6. Tableau.

Es uno de los líderes de la inteligencia empresarial moderna. Apenas superado por Power BI en 2020 (ver figura 17.) según Gartner, *Tableau*, es una plataforma intuitiva, fácil de manejar, explotar y administrar los datos, es un sistema innovador de inteligencia de negocios, el cual se puede utilizar en investigaciones simples como en investigaciones complejas, una de las mayores ventajas de estos sistema es que reconoce los datos de cualquier formato, reduce de gran forma el tiempo de análisis de los datos, a través de visualizaciones [38].

- **Intuitiva:** Los Productos de *tableau* están fabricados para el fácil manejo de los usuarios finales, ya que el análisis de los datos debe comprender en hacer preguntas y no en aprender software. *Tableau* permite una visualización sin límite de datos [39].
- **Analítica:** Al poseer licencias basadas en roles, *Tableau* pone el acceso a la información por todos los usuarios, desde científicos hasta usuarios comerciales [39].

Figure 1. Magic Quadrant for Analytics and Business Intelligence Platforms



Fig. 17 Cuadrante Mágico de las Plataformas de Analítica e Inteligencia de Negocios. [40]

1.6.6. COSTO TOTAL DE PROPIEDAD DE LAS HERRAMIENTAS

El costo total de la propiedad de las plataformas y herramientas de software que se han citado en el apartado 1.6.5., las cuales han sido seleccionadas principalmente reportes de Gartner – reconocida firma de consultoría e investigación de las tecnologías de información – para trabajar en el presente proyecto; ha sido determinado en función de las políticas de licenciamiento, soporte y mantenimiento que caracterizan a cada una de ellas.

Para empezar, es necesario tener un conocimiento del escenario en el que se va a implementar. En la Tabla III se proponen tres tipos de escenarios, los cuales fundamentalmente corresponden al número de usuarios de las licencias.

TABLA II.

TABLA DE ESCENARIOS Y USUARIOS

Tamaño	Total Usuarios	Usuarios Creadores	Usuarios Exploradores	Usuarios Lectores
Pequeña	25	1	3	21
Mediana	100	3	12	85
Grande	500	15	53	432

Nota: Tipo de usuario para la determinación del costo total. [29]

Para los diferentes escenarios se empleó la regla de Pareto, 80-20 (18-2) con la cual se establece que; el 80% de los usuarios que interactúan de forma sencilla con las herramientas, mientras que un 20% lo hace de forma intensificada. Dicho de otra manera; el 80% de los usuarios necesitan hacer revisión y lectura de reportes y *dashboards*, del 20% restante un 18% modifican objetos y producen reportes de información y apenas el 2 % restante se refiere a los usuarios analistas que generan los diferentes reportes y *dashboards* para cada departamento de la empresa. [29]

De acuerdo a la información oficial del proveedor de la Plataforma de *Tableau* de dependiendo del escenario, los costos licencia y de soporte a tres años se reflejan en la TABLA IV.

TABLA III
COSTOS DE PROPIEDAD DE TABLEAU – 3 AÑOS

Proveedor	Tamaño	Costo de Licencia /Mes	Costo Licencia /Anual	Costo de soporte 36 meses	COSTO DE PROPIEDAD 3- AÑOS
	Pequeño	\$427,00	\$5.124,00	\$427,00	15.799,00
Tableau	Mediano	\$1.650,00	\$19.800,00	\$1.650,00	61.050,00
	Grande	\$8.089,00	\$97.068,00	\$8.089,00	299.293,00

Nota: Costos de Tableau. [41]

Tableau hace a los usuarios que sea más fácil y rápido administrar y explorar los datos. Es una herramienta de analítica de un extremo a otro muy potente, segura y flexible [41].

En el caso de la herramienta de diseño y modelado de datos *SAP Power Designer*, las políticas de licenciamiento del proveedor, se establecen por costo anual empresarial y el costo de renovación de la licencia. Dependiendo del país este costo difiere. En este contexto, la TABLA V muestra el Costo Total de Propiedad de esta herramienta a tres años.

TABLA IV.
COSTO DE PROPIEDAD DE POWER DESIGNER – 3 AÑOS

Proveedor	Costo de Licencia /Anual	Renovación Anual	COSTO DE PROPIEDAD 3- AÑOS
SAP Power Designer	\$7.500	\$1.650	\$10.800,00

Nota: Costos de Licencia de Power Desinger. [34]

El caso de *Knime Analytics Platform*, los costos de licenciamiento se definen de acuerdo al tipo de licencia que se requieran: paquete por servidor o servicio en la nube. La Tabla VI detalla los costos de esta plataforma de analítica a 3 años en función del tamaño del servidor.

TABLA V.

COSTO DE PROPIEDAD DE KNIME ANALYTICS PLATFORM – 3 AÑOS

Proveedor	Categoría	Núcleos	RAM	Cada Hora	Mensual	Anual	COSTO DE PROPIEDAD 3- AÑOS
Knime Analytics Platform	Básico	1	1.7	\$2.095,00	\$1.558,68	\$18.704,16	\$56.112,48
	Medio	2	3,5	\$2.155,00	\$1.603,32	\$19.239,84	\$57.719,52
	Grande	4	7	\$2.258,00	\$1.679,95	\$20.159,42	\$60.478,26

Nota: Costos de Licencia de Knime. [36]

Exaplus

TABLA VI.

COSTO DE PROPIEDAD DE EXAPLUS– 3 AÑOS

Proveedor	Costo Licencia/ Mensual	Costo de Licencia/ Anual	COSTO DE PROPIEDAD 3- AÑOS
Exaplus	\$2.611,17	\$3.5605,20	\$10.680,00

Nota: Costos de Licencia de Exaplus. [33]

CAPITULO 2. DIAGNOSTICO SITUACIONAL

2.1. ANTECEDENTES

La EMAPAL-EP fue creada el 30 de marzo de 1974 como una decisión del Ilustre Concejo Municipal mediante la Ordenanza Nro. 13. Inicia sus actividades independientes en agosto de ese mismo año, cuando ejercía la Primera Dignidad de la ciudad el Dr. Marco Romero Heredia. En esa época ya se operaba la planta de tratamiento de Zhindilig con una capacidad aproximada de 40 lts./seg. CITA

Para octubre de 1986 se inició con la construcción de la planta de tratamiento de Uchupucún en la administración del Sr. Hipólito Izquierdo y la Gerencia de la empresa el ejercía el Ing. Francisco Zea. Ésta estuvo diseñada para tratar un caudal de 70 lts./seg., de las captaciones provenientes de Nudpud y Llaucay; sin embargo en la actualidad el agua proviene de la nueva captación del llamado sistema Tabacay. (“Reseña Histórica,” n.d.).

Complete el párrafo anterior con información importante sobre la capacidad de tratamiento actual.

Cuenta con 11689 abonados, distribuidos en: Mahuarca, Uchupucun y Zhindilig.

Es una empresa del sector público que ofrece servicios de agua potable, alcantarillado y saneamiento ambiental a la ciudadanía del Cantón Azogues

En la actualidad, la EMAPAL-EP, es una empresa que bajo la dirección del Presidente: Dr. Romel Sarmiento (En su calidad de Alcalde o quien lo subrogue legalmente), Miembros del Directorio: Ing. Francisco Molina Cárdenas(En su calidad de Concejal, representante de la Ilustre corporación, o su alterno), Arq. Gonzalo Ormaza Vintimilla(Representante de la Ciudadanía, miembros del Consejo de Participación Ciudadana), Arq. Eugenio Calle Yumbla (Representante del Colegio de Arquitectos del Cañar), Ing. Jorge Cárdenas Rojas(Representante del colegio de Ingenieros Civiles del Cañar), Secretario: Ing. Santiago Luna Romero(En su calidad de Gerente General, tiene voz informativa en las Sesiones del Directorio), está gestionada por la gerencia, bajo cuyo mando están los departamentos de: Asesoría Jurídica, Dirección de Planificación, Dirección de Comercial, Dirección Financiera y Dirección Administrativa dentro de la cual está el Departamento de TICs (ver figura 18.)

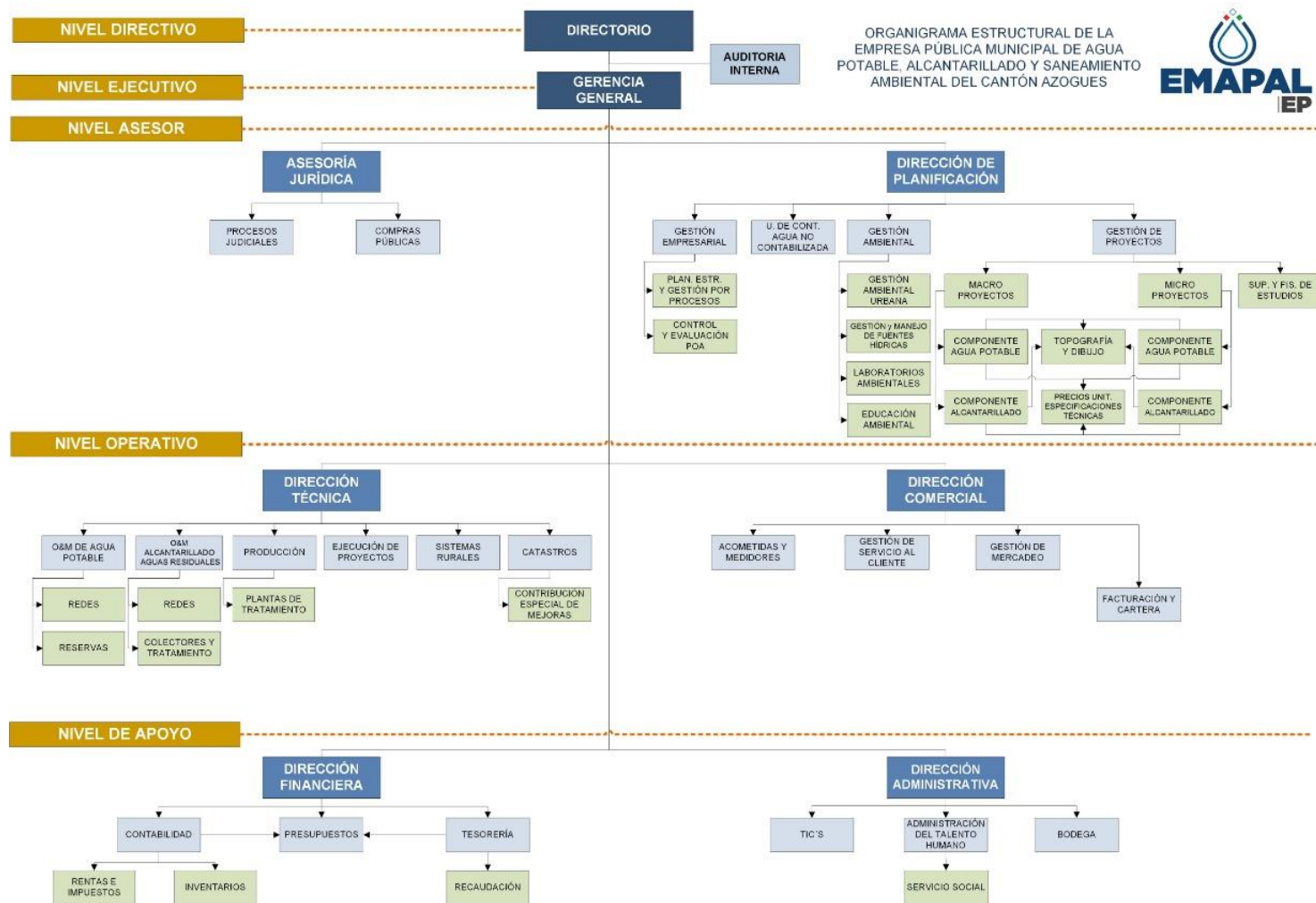


Fig. 18. Organigrama de EMAPAL. [42]

Debido al exponencial crecimiento de esta empresa, tanto en servicios, como en abonados, se ha incrementado de la misma forma la información, a través de una herramienta ERP con la que cuenta actualmente la empresa se manejan grandes volúmenes de información la cual no cuentan con un sistema o herramientas que ayuden al procesamiento y análisis de la información y que permita llevar a cabo una correcta administración basadas en la toma decisiones precisas, por medio del acceso veloz y detalla de la información requerida sobre algún tema de necesidad.

2.2. DEFINICIÓN DEL PROBLEMA

El encargado de TI manifiesta que los tiempos de respuesta para cumplir los requerimientos de información de la gerencia y direcciones departamentales son altos y requiere de mucho esfuerzo realizar este trabajo. Para hacerlo usan la herramienta SIIM. Adicionalmente a ello, en ocasiones las necesidades de información no se satisfacen completamente.

Varias son las necesidades de información de la EMAPAL, sin embargo, según manifiesta responsable de TI, Ing. Giovanni Morocho, una de las áreas críticas es la de comercialización. En efecto en esta área algunos de los problemas son:

- Falta de sistemas informáticos que permitan el análisis y el entendimiento de grandes cantidades de información que se manejan en la empresa
- Ausencia de una solución de inteligencia de negocios en EMAPAL que impide la mejora y ganancia de valor agregado estructural para la empresa.
- Los reportes o informes que se presentan a la parte gerencial muchas de las veces carecen de importancia para la toma de decisiones.
- La información no se obtiene de forma rápida y precisa
- Al momento que el gerente solicita un informe, en el tiempo que transcurre entre la petición del informe y la entrega del informe, se han producido varias actualizaciones en la información, lo que conlleva a que al momento que el gerente tenga el informe, los datos y la información cambiaron.
- Se solicita reportes, y varios no son entregados por la falta de tiempo para su elaboración.
- Al momento de generar un informe, se lo hace de forma manual con la herramienta Excel, lo que hace que el tiempo de entrega sea extenso, ya que se debe depurar la información de acuerdo a la necesidad del reporte.

En este contexto es necesario solventar este problema, la implementación de una solución de inteligencia de negocios que visibilice por medio de dashboards o reportes específicos, información a los niveles operativo y tácticos del departamento de comercialización de la EMAPAL-EP, los cuales serán creados a partir de los datos históricamente almacenados en la base de datos del sistema ERP, con que cuenta.

2.3. JUSTIFICACIÓN

La empresa EMAPAL EP no cuenta con un modelo de gestión que apoye la toma de decisiones informada. Las decisiones se toman de manera intuitiva y dependen de la experiencia y criterio de quienes gestionan esta organización. La tasa de rotación de la gerencia es alta, esto se debe a que, por ser una institución pública, el cargo de gerente es de libre remoción y generalmente éstos se promueven cada cuatro años.

A esta realidad se suma el hecho de que esta empresa aún no cuenta con una gestión basada en procesos, su organización funcional es puramente vertical y no todos los procesos de negocio han sido implementados aún en el sistema ERP con el que cuenta. Sin embargo, los procesos críticos, tal es el caso de la gestión de la recaudación. La información transaccional que genera este sistema es muy importante para la empresa y como tal debe ser tratada de manera especial para que cumpla con las necesidades de información a nivel táctico y estratégico.

Como se manifestó en la definición del problema, la necesidad de información táctica y estratégica requiere el uso de herramientas de Inteligencia de Negocios, que agilicen las consultas de complejos y grandes cantidades de datos y los visualice de manera más simple y relevante. Las gráficas de líneas, barras, en pastel, de mapas, etc., el uso de colores, etiquetas y otros; permitirán destacar la información y concentrarlos en dashboards ejecutivos que mejoren su comprensión y visibilización.

El conocimiento que se genera responderá de manera eficaz a las necesidades de la alta dirección y tomar mejores decisiones, gracias a que gestiona de datos actuales e históricos, que posibilitarán identificar patrones y tendencias

Este proyecto tiene como finalidad dar soporte a los directivos, administrativos del departamento de comercialización de la EMAPAL EP en la toma de decisiones de manera precisa.

2.4. OBJETIVOS

2.4.1. OBJETIVO GENERAL

Implementar una solución de BI para apoyar la extracción de conocimiento del área de comercialización en la EMAPAL

2.4.2. OBJETIVOS ESPECÍFICOS

1. Determinar requerimientos de la información.
2. Diseñar la solución de BI en base a las funcionalidades requeridas (diseño de BI).
3. Construcción de la solución de BI.
4. Probar e implementar la Solución de BI.

2.5. ALCANCE

El presente trabajo de titulación tiene como fin realizar una propuesta de solución de Inteligencia de negocios para el área de comercialización de la EMAPAL-EP la cual forma parte de la cadena de valor de la empresa. En este contexto el trabajo se realizará con un enfoque táctico que inicia con la creación de un *data mart* del área seleccionada para que posteriores trabajos de BI, se pueda escalar hacia otras áreas.

El trabajo se limitará a gestionar los requerimientos iniciales de información y no es parte de este trabajo elaborar consultas que gestionen indicadores que relacionen información de otras áreas. Asimismo, las consultas tratan las necesidades de información requeridas al inicio de este trabajo y de ninguna manera representan la totalidad de lo requerido.

Respecto a la metodología usada, es importante indicar que la propuesta ha sido adaptada con los elementos metodológicos que reflejan la realidad de la EMAPAL y las herramientas usadas, por lo que no fue necesario desarrollar algunos entregables.

Por otra parte, la solución será gestionada por los responsables de TI, razón por la cual, se instalará en un solo equipo del departamento de TI. Son ellos quienes se encargarán de actualizar la información del *data mart* y gestionarán consultas que no están previstas en este trabajo, siempre y cuando se haga uso del mismo modelo de datos dimensional que será diseñado.

La puesta en marcha de la solución dependerá de la decisión de EMAPAL de invertir en la adquisición de las licencias de las plataformas y herramientas necesarias. El trabajo asegura el cumplimiento de los requerimientos a nivel de pruebas, las cuales serán ejecutadas conjuntamente con los responsables de TI y retroalimentados hasta que se consiga satisfacer las necesidades iniciales.

2.6. TRABAJOS RELACIONADOS

- ✓ En [43] se lleva a cabo el análisis de una propuesta de implementación de una herramienta de Bussines Intelligence orientada a la empresa Soldeneg Soluciones de Negocios Cia.Ltda., en este estudio se realiza una comparativa de las distintas herramientas destinadas a la generación de reportes, en este proyecto se propone utilizar Qlik Sense Cloud Business debido a que esta herramienta está diseñada para pequeñas empresas, gracias a la propuesta de Bussines Intelligence realizada en este estudio se pudo analizar la información relevante de la empresa, con el fin de dar soporte a tomar mejores decisiones organizacionales con una administración estratégica, dando paso a la organización a tener mayor competitividad en el mercado.
- ✓ Un estudio realizado en la ciudad de Quito-Ecuador acerca de la construcción de un aplicativo de Inteligencia de Negocios por [44], desarrollado para la empresa Importadora Tomebamba S.A, la misma que tiene como objetivo principal apoyar y facilitar las actividades de dirección y toma de decisiones del área de Ventas de dicha

empresa, para la implementación de la solución de BI, se ha elegido es Pentaho BI *Community Edition* por ser una herramienta Open Source, a través de esta herramienta se logró integrar varias fuentes de información como: archivos planos, archivos de Excel y bases de datos transaccionales. Al unificar toda esta información mediante este aplicativo se provee al gerente las herramientas necesarias para tomar decisiones de acuerdo a las necesidades.

Actualmente dentro de la ciudad de Azogues no se conoce sobre la implementación de una solución de Inteligencia de Negocios en instituciones públicas; de esta manera este caso de estudio sería la primera desarrollada localmente, la cual ha sido desarrollada para apoyar la toma de decisiones de la EMAPAL-EP.

CAPITULO 3. PROPUESTA

La solución de Inteligencia de Negocios para el departamento de Comercialización de la EMAPAL-EP, se basa en la metodología propuesta en el 1.6.4. Sin embargo, por las particularidades de la EMAPAL y del alcance del presente trabajo, algunos entregables no pudieron ser adaptados y, como se menciona en el alcance, se cumplió hasta la etapa de pruebas.

Los siguientes apartados explican de forma detallada el trabajo desplegado en este trabajo:

3.1. DEFINICIÓN DE LA PROPUESTA

Esta etapa se llevó a cabo mediante reuniones previamente establecidas con los responsables de TIC de la EMAPAL, con el propósito de poner en conocimiento acerca de los beneficios que brindan las soluciones de Inteligencia de Negocios y solicitar la aprobación formal para la realización del trabajo. Se estableció como área prioritaria la de Comercialización, se definieron las funcionalidades generales y el alcance. (Ver Anexo D).

Con los requerimientos iniciales se procedió a formular un plan de trabajo. Ver Tablas VIII y IX.

CRONOGRAMA DE ACTIVIDADES.

AÑO 2019

TABLA VII.

CRONOGRAMA DE ACTIVIDADES PROPUESTO

ACTIVIDADES	Abril				Mayo				Junio				Julio				Agosto				Septiembre				Octubre				Noviembre				Diciembre			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Acta de constitución del proyecto.																																				
Definición del alcance.																																				
Definición del plan de trabajo.																																				
Levantamiento de la información.																																				
Levantamiento de la infraestructura tecnológica.																																				
Identificación de los requerimientos de la información.																																				
Análisis de fuentes de datos y requerimientos.																																				
Diseño del modelo de negocio																																				
. Diseño del DW.																																				

AÑO 2020

Tabla VIII.

CRONOGRAMA DE ACTIVIDADES EJECUTADO

ACTIVIDADES	Enero				Febrero				Marzo				Abril				Mayo				Junio				Julio				Agosto				Septiembre				Octubre				Noviembre			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4				
Diseño de procesos ETL	■	■																																										
Diseño de reportes			■																																									
Diseño de la solución de BI.			■	■	■	■																																						
Construcción del DW.					■	■	■	■																																				
Construcción de procesos ETL.									■	■																																		
Construcción de reportes																																												
Construcción de la solución de BI.																																												
Definición del plan de pruebas.																																												
Consultas en Tableau																																												
Aprobación de la solución de BI																																												
Presentación del proyecto de titulación.																																												

NOTA: En el periodo marzo - agosto no se llevaron a cabo las actividades pre establecidas, debido a la emergencia sanitaria

3.2. LEVANTAMIENTO DE LA INFORMACIÓN:

Esta etapa se llevó a cabo mediante entrevistas a los encargados del área de TICs, con el objetivo conocer a la EMAPAL como; su organización y gobierno, los procesos y las necesidades de información más relevantes del área de comercialización. En este contexto se pudo determinar que el área de TI es un área relegada de la empresa, no tiene autonomía y por tanto no tiene asignado un presupuesto para invertir en soluciones tecnológicas como las herramientas analíticas para dar soporte a la gestión de la información. Como consecuencia de esto, los encargados de TI cumplen con los requerimientos de información para soporte a la toma de decisiones de los niveles táctico y estratégico, invirtiendo gran cantidad de tiempo en y esfuerzo intentando, cumplir los requerimientos de información con herramientas inadecuadas y obteniendo resultados tardíos e imprecisos.

De otra parte, se levantó información general sobre los elementos de infraestructura tecnológica, los riesgos y limitaciones los que se muestran en la Tabla X y XI.

TABLA IX
INFRAESTRUCTURA TECNOLÓGICA PCs.

TIPO	NOMBRE	CARACTERÍSTICAS				USO/SERVICIO
		Modelo	Micro P.	RAM	DIS.DURO	
PC	PC	CLONE Q-ONE	CORE i7	8GB	1 TB	Uso de reporting y actividades del personal de TIC
	PC	CLONE Q-ONE	CORE i7	8GB	1 TB	Uso de reportes y explotación de información.
	PC	CLONE Q-ONE	CORE i7	16GB	1 TB	Uso de actividades de TIC

Nota: Redacción de Autoría de los recursos tecnológicos.

TABLA X.
INFRAESTRUCTURA TECNOLÓGICA SERVIDOR.

TIPO	MODELO	CARACTERÍSTICAS				RIESGOS
		Procesador	Memoria RAM	Disco Solido	Fuente de poder	
SERVIDOR	HP ProLiant DL380 Gen9	DL380 Gen9 Intel Xeon E5-2640v4 (2.4 GHz/10-core/25MB/90 W) processor kit- 817937-B21 *S/N: J720C376	HP 16GB (1x16GB) single rank x4 DDR4-2400 CAS-17-17-17 registered memory kit – 805349-B21 *S/N: SH69XM6 *S/N: SHSHVM6	HPE 960GB SATA 6G MU SFF (2.5 IN) SC 3 yr warranty DS firmware SSD – P07926-B21 *S/N: SG00020K9B *S/N: SG00020KB6	HPE 500 W Flex slot platinum hot plug power supply kit- 720478-B21 *S/N: 1721009659	*No se hace respaldos de la información periódicamente. *No disponen la estructura física adecuada para los servidores

Nota: Redacción de Autoría de los recursos tecnológicos (ver Anexo F).

Respecto a las aplicaciones, bases de datos; las XI y XII sintetizan las aplicaciones, bases de datos y otras fuentes de datos que posee la EMAPAL que están involucradas de forma directa o indirectas con el área de comercialización.

TABLA XI.
APLICACIONES RELACIONADAS CON COMERCIALIZACION

Aplicación	Descripción	Relación Comercialización
SIIM	Plataforma que aloja varios módulos que dan soporte a la empresa	Directa
SITAC	Procesos de retenciones	Directa
Modulo Financiero	Recepta información de reportes del área de comercialización	Directa
Recursos Humanos	Destinado a la gestión del personal de la empresa	Indirecta

Nota: Redacción de Autoría de las aplicaciones relacionadas con el área de comercialización.

TABLA XII.

BASES DE DATOS RELACIONADAS CON COMERCIALIZACION

Base de Datos / Fuente de datos	Descripción	Relación Comercialización
Base de Datos SIMM del ERP - PostgreSQL	Destinada al área de comercialización	Directa
SQL	Destinado para GIS	Indirecta
Oracle	Destinada a recursos Humanos	Indirecta
Access	Destinada al área financiera	Directa

Nota: Redacción de Autoría de las aplicaciones relacionadas con el área de comercialización.

Para la identificación de requerimientos funcionales y no funcionales se ha hecho uso de Historias de Usuario, las cuales han recopilado información fundamentalmente de los requerimientos de información del departamento de comercialización y los requerimientos de software e infraestructura tecnológica, que requerirá la propuesta para su implementación. Las historias de usuario más representativas han sido recopiladas en este trabajo y se muestran a continuación; sin embargo, como se mencionó en el alcance, existen algunos niveles de consulta a los que se podrá llegar en futuros *Sprints*, puesto que no son parte de este trabajo.

TABLA XIII
HISTORIA USUARIO 1

Historia de Usuario	
N.1	Historia: Situación Financiera
Nivel: Estratégico	
<p>Como encargado del área de TICS</p> <p>Puedo visualizar en una gráfica lineal el total de las tarifas de la empresa.</p> <p>Para no tener que esperar la entrega de un reporte de dicha información, el cual toma mucho tiempo en ser elaborado.</p> <p>Criterio de Aceptación:</p> <ul style="list-style-type: none"> • Dado que quiero visualizar la situación financiera de la empresa. • Cuando el gerente requiera saber el estado del total de las tarifas, durante un periodo determinado • Entonces se procederá a filtrar la información. <p>Detalles:</p> <ul style="list-style-type: none"> • La grafica deberá presentarse de forma lineal • Se permitirá filtrar la información por año. • Se permitirá agrupar la información de acuerdo al Tipo de Categoría. • En cada punto de la gráfica, con respecto a cada año se presentará, la cantidad (en números) del Total de Tarifa recaudada durante cada año. 	

Nota: Redacción de Autoría de la Historia de usuario

La Historia de Usuario N°1, determina que el usuario el cual utilice el sistema, pueda visualizar de forma gráfica, clara y global de la situación financiera de la empresa, además que se pueda utilizar filtros de datos, de acuerdo a las Categorías como son: Comercial, Industrial, Oficial y Residencial

TABLA XIV.
HISTORIA USUARIO 2

Historia de Usuario	
N.2	Historia: Clientes Actuales
Nivel: Táctico	
<p>Como encargado del área de TICS</p> <p>Puedo filtrar información de los clientes.</p> <p>Para de esta forma valorar la calidad de servicio que presta la empresa e identificar patrones de comportamiento.</p> <p>Criterio de Aceptación:</p> <ul style="list-style-type: none"> • Dado que quiero obtener datos estadísticos de los clientes. • Cuando se opte por tomar decisiones en base a esta información. • Entonces se filtrará la información del incremento o decremento de los clientes. <p>Detalles:</p> <ul style="list-style-type: none"> • Los resultados se presentarán en una gráfica de barras. • El número de clientes se deberá presentar de cada mes, de acuerdo a la fecha que se podrá seleccionar. • Al dar clic sobre la barra que se desee, se deberá presentar el número de clientes para esa barra. 	

Nota: Redacción de Autoría de la Historia de usuario

La Historia de Usuario N°2, indica la necesidad de conocer el índice de crecimiento de los abonados para determinar la calidad de servicio que se brinda, con el fin de proyectar el crecimiento de la organización.

TABLA XV.
HISTORIA USUARIO 3

Historia de Usuario	
N.3	Usuario: Encargado de TICS
Nombre de Historia: Filtrar consumo por Categoría y Tiempo.	
Nivel: Táctico	
<p>Como encargado del área de TICS</p> <p>Puedo visualizar el porcentaje de M3 de Agua Consumido.</p> <p>Para la toma decisiones en función del consumo y de la ubicación.</p> <p>Criterio de Aceptación:</p> <ul style="list-style-type: none"> • Dado que quiero visualizar la información del consumo en una gráfica de pastel • Cuando se requiera obtener un reporte de consumos de agua. • Entonces se presentará un <i>Dashboard</i> con el porcentaje total de consumo. <p>Detalles:</p> <ul style="list-style-type: none"> • Se permitirá escoger la fecha y tipo de categoría. • Los datos serán presentados en una gráfica de pastel. • En la gráfica se presentará el porcentaje de M3 de agua consumida de acuerdo al tipo de Categoría. • Se podrá filtrar los datos de acuerdo a una fecha en específico. 	

Nota: Redacción de Autoría de la Historia de usuario

La Historia de Usuario N°3, explica que se requiere obtener la información de los M3 consumida de agua por categoría al instante debido a que es un reporte que generalmente es solicitado por el gerente de la organización.

TABLA XVI.
HISTORIA USUARIO 4

Historia de Usuario	
N.4	Historia: Buscar un Usuario por Id.
Nivel: Operativo	
<p>Como encargado del área de TICS</p> <p>Puedo buscar un usuario en específico</p> <p>Para identificar de forma rápida a cierto usuario registrado.</p> <p>Criterio de Aceptación:</p> <ul style="list-style-type: none"> • Dado que quiero buscar de forma rápida a un usuario en específico según su identificador. • Cuando se requiera saber, información de los metros cúbicos de agua consumido durante un mes en específico. • Entonces se mostrará de manera instantánea la información del cliente. <p>Detalles:</p> <ul style="list-style-type: none"> • Se permitirá escoger la fecha que se requiera visualizar • Los datos serán presentados en una gráfica lineal. • En la gráfica se presentará los metros cúbicos de agua que ha consumido el cliente. 	

Nota: Redacción de Autoría de la Historia de usuario

La Historia de Usuario N°4, para la búsqueda de un histórico de un usuario en específico en casos de procesos de coactivas.

3.3. DISEÑO

Puesto que el presente trabajo se lo realizó en el área de comercialización, el diseño se orientó a la creación de un *data mart* de esta área funcional. No fue necesario diseñar el modelo de negocio ni el *Data warehouse*. Los elementos de diseño que se mencionan a continuación han sido adaptados a la realidad del área de comercialización de la EMAPAL-EP basados en el enfoque Kimball.

3.3.1. DISEÑO DEL MODELO DIMENSIONAL

En su defecto, el modelo multidimensional propuesto, fue creado con SAP Power Designer. Entonces, usando el gestor de Bases de Datos PostgreSQL y PgModeler, se gestionó la compleja estructura de datos de Base de datos SIMM declarada en la tabla XII, que corresponde a la principal fuente de datos para este ejercicio de BI. Se identificaron los atributos relacionados con la información requerida y sobre esta base se procedió a identificar la tabla de Hechos y las tablas de Dimensiones del modelo, las cuales se detallan a continuación:

- **TABLA DE DIMENSIONES FACTURA:** Esta dimensión contiene todas las facturas que se han generado.

TABLA XVII.

TABLA DE DIMENSIONES FACTURA

Factura		
Columna	Tipo	Descripción
ID_FACTURA	Long integer	Clave principal de la Dimensión Factura.
NROFACTURA	Variable Character(20)	Numero De Factura
TOTALTARIFA	Long float	Total de la Tarifa
VALORBASE	Long float	Valor Base
ABONO	Long float	Abono
PAGO	Long integer	Si esta pagado o no

Nota: Autoría de redacción de la tabla factura en el sistema.

- **TABLA DE DIMENSIONES CATEGORÍA:** Esta dimensión contiene las categorías donde están seccionadas los diferentes clientes como por ejemplo “RESIDENCIAL”.

TABLA XVIII.

TABLA DE DIMENSIONES CATEGORÍA

Categoría		
Columna	Tipo	Descripción
ID_CATEGORIA	Long integer	Clave principal de la Dimensión Categoría.
DESCRIPCION	Variable Character(255)	Tipo de Categoría.

Nota: Autoría de redacción de la tabla categoría en el sistema.

- **TABLA DE DIMENSIONES EMISIÓN:** Esta dimensión contiene las fechas de inicio y fecha de fin de la emisión.

TABLA XIX.

TABLA DE DIMENSIONES EMISIÓN

Emisión

Columna	Tipo	Descripción
ID_EMISION	Long integer	Clave principal de la Dimensión Emisión.
EMISION	Variable Character(20)	Código de la Emisión
FECHAAPERTURA	date	Fecha De Inicio
FECHACIERRE	date	Fecha De Cierre

Nota: Autoría de redacción de la tabla Emisión en el sistema.

- **TABLA DE DIMENSIONES CLIENTE:** Esta dimensión contiene los datos de cada cliente

TABLA XX.

TABLA DE DIMENSIONES CLIENTE

Cliente		
Columna	Tipo	Descripción
ID_ABONADO	Long integer	Clave principal de la Dimensión Cliente.
ID_RUTA	Long Integer	Clave principal de la Dimensión Ruta.
CEDULA	Variable Character (20)	Cedula del Cliente
APELLIDO	Variable Character (20)	Apellido del Cliente
NOMBRE	Variable Character (20)	Nombre del Cliente
DIRECCION	Variable Character (20)	Dirección del Cliente
FECHANACIMIENTO	Date	Fecha De Nacimiento del Cliente
ABONADOCLIENTE	Long Integer	Clave principal de la Dimensión Cliente.
NROMEDIDOR	Variable Character (30)	Número del Medidor
ESTADO	Integer	Indica si está activo o inactivo.

Nota: Autoría de redacción de la tabla cliente en el sistema.

- **TABLA DE DIMENSIONES UBICACIÓN:** Esta dimensión contiene los datos de donde están ubicados los servicios registrados.

TABLA XXI.

TABLA DE DIMENSIONES UBICACIÓN

Ubicación		
Columna	Tipo	Descripción
ID_RUTA	Integer	Clave principal de la Tabla Ruta.

DESCRIPCIONRUTA	Variable Characters(200)	Nombre de la Ruta.
ID_SECTOR	Variable Characters(4)	Clave principal de la Tabla sector.
DESCRIPCIONSECTOR	Variable Characters(200)	Nombre del sector.
ID_ZONA	Variable Characters(4)	Clave principal de la Tabla Cliente.
DESCRIPCIONZONA	Variable Characters(200)	Nombre de la Zona.
ID_PARROQUIA	Variable Characters(4)	Clave principal de la Tabla Parroquia.
DESCRIPCIONPARROQUIA	Variable Characters(200)	Nombre de la Parroquia.
ID_CANTON	Variable Characters(4)	Clave principal de la Tabla Cantón.
DESCRIPCIONCANTON	Variable Characters(200)	Nombre del Cantón.
ID_PROVINCIA	Variable Characters(4)	Clave principal de la Tabla Provincia.
DESCRIPCIONPROVINCIA	Variable Characters(200)	Nombre de la Provincia.

Nota: Autoría de redacción de la tabla ubicación en el sistema.

- **TABLA DE HECHOS AGUA_LIQUIDACION:** De igual manera se identificó la tabla de hechos, que es la tabla principal y central de nuestro modelo dimensional, la cual se relaciona con las dimensiones a través de las claves foráneas, que combinadas forman la clave principal de los hechos. Esta tabla además contiene las métricas que servirán posteriormente para presentar la información cuantitativa. La tabla identificada como tabla de hechos es:

TABLA XXII.

TABLA DE HECHOS DE AGUA_LIQUIDACIÓN

Agua_Liquidacion		
Columna	Tipo	Descripción
ID_AGUA_LIQUIDACION	Long integer	Clave principal de la Tabla de hechos.

ID_FACTURA	Long integer	Clave principal de la Dimensión Factura.
ID_ABONADO	Long integer	Clave principal de la Dimensión Abonado.
ID_EMISION	Long integer	Clave principal de la Dimensión Emisión.
ID_CATEGORIA	Long integer	Clave principal de la Dimensión Categoría.
LECTURAANTERIOR	Long float	Lectura Anterior del medidor
LECTURAACTUAL	Long float	Lectura Actual del medidor

Nota: Autoría de redacción de la tabla Agua_liquidación en el sistema.

Ya identificadas las tablas de Hecho y Dimensiones, se construyó el Modelo Logico del Datos, como se muestra en la Figura 19.

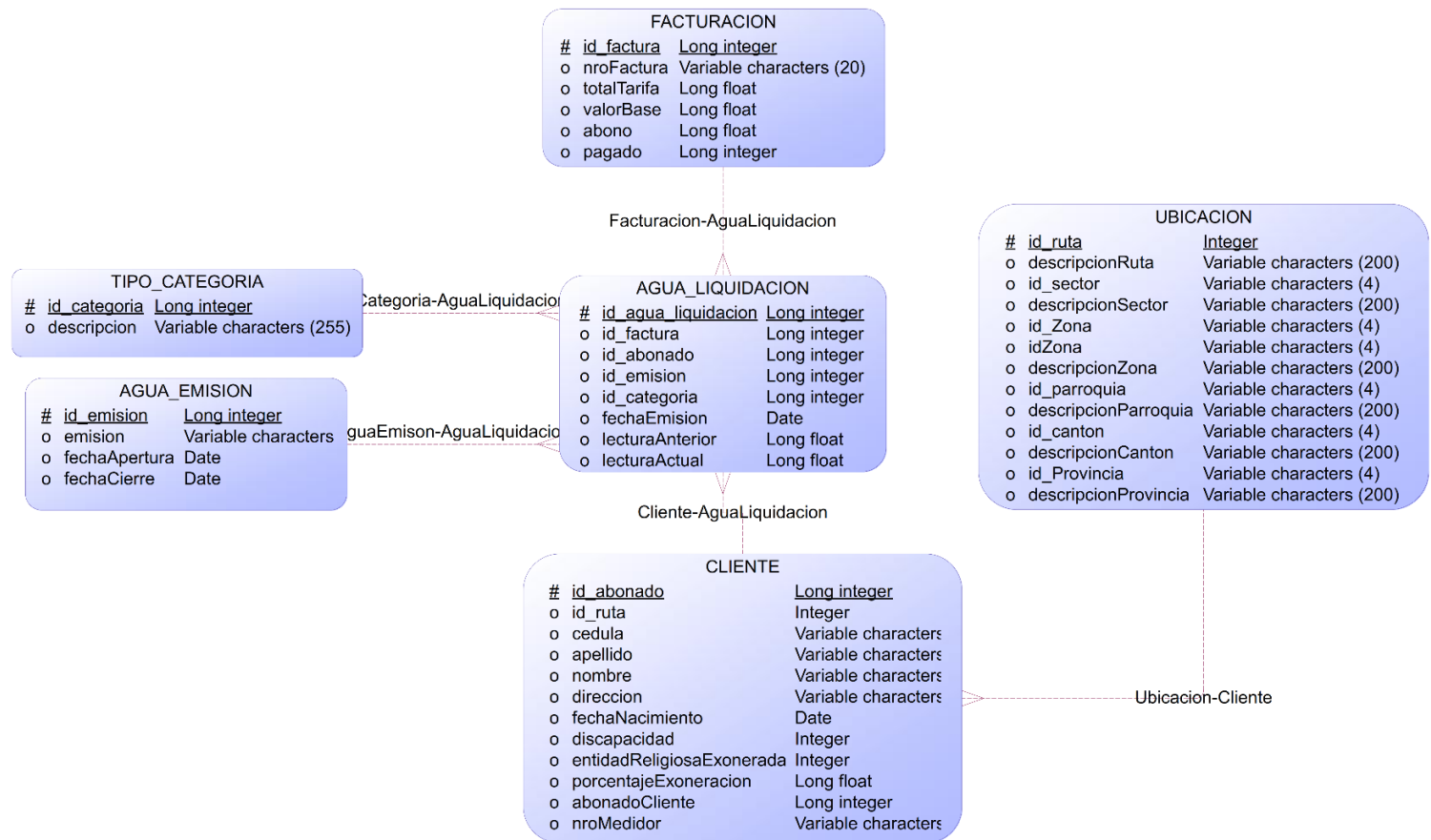


Fig. 19. Modelo Lógico del Negocio. Autor (2020)

A partir del modelo lógico creado anteriormente creamos, el Modelo Físico, el cual se lo realizó con la misma herramienta *SAP Power Designer* tal como se indica en la Ilustración 1:

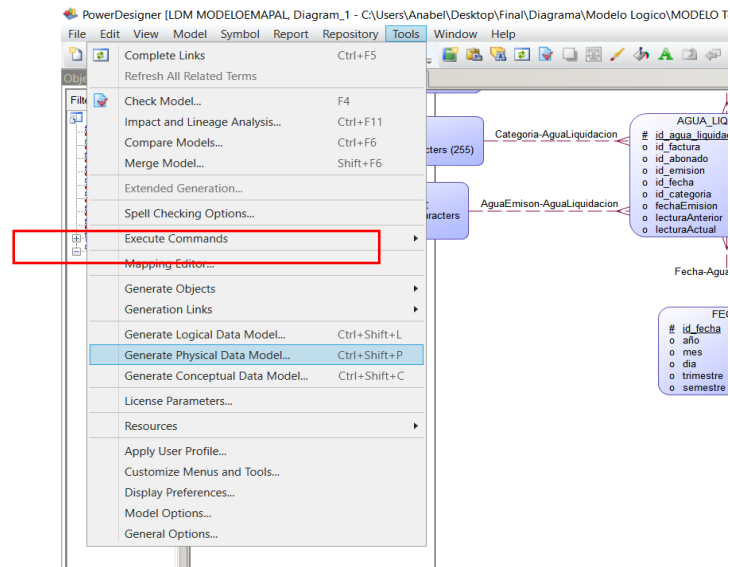


Ilustración 1. Generar Modelo Físico del Datawarehouse. Autor (2020)

De esta forma obtenemos el modelo Físico del *Data Mart* como se ilustra en la Figura 20. que posteriormente será utilizado para generar el Script del modelo.

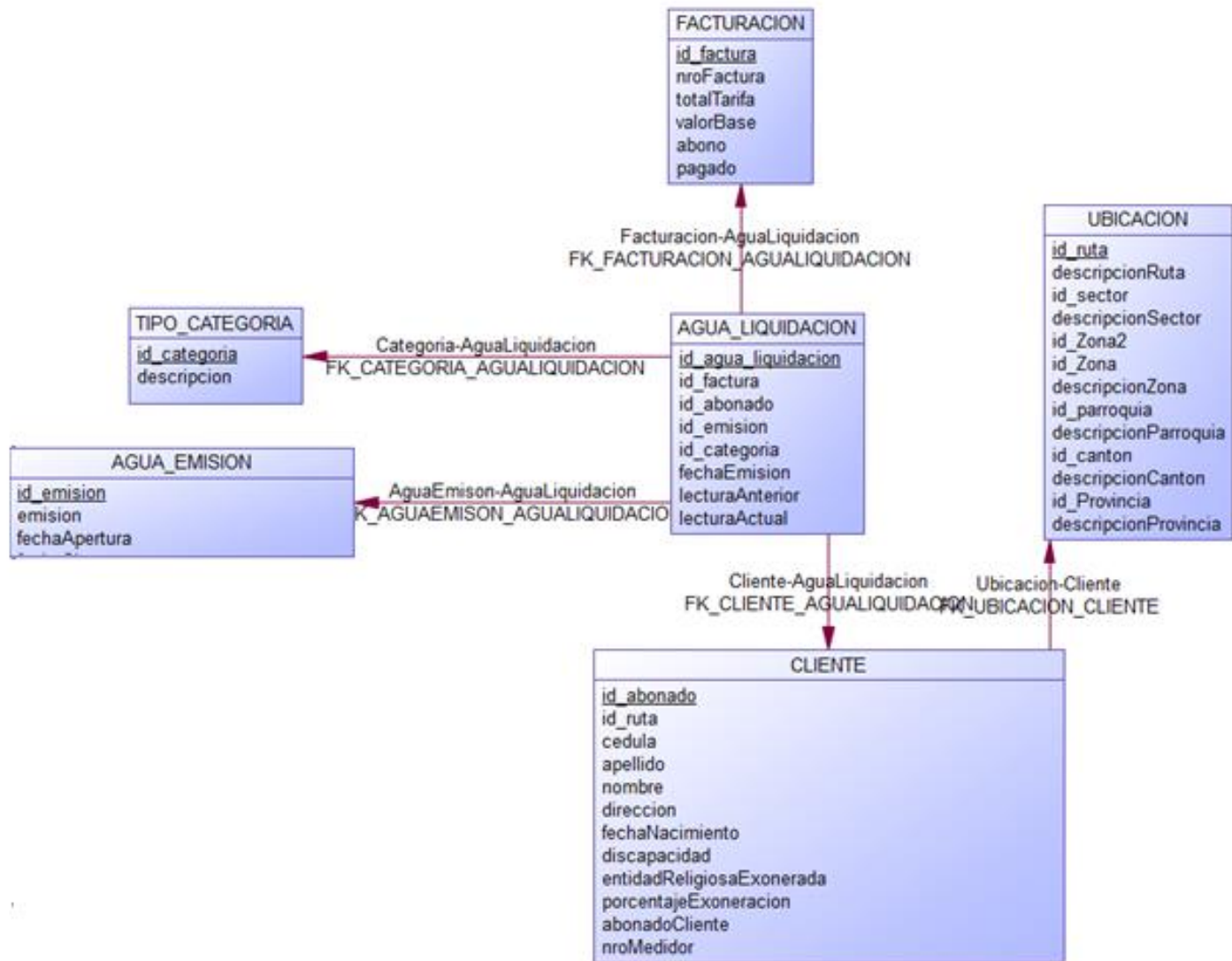


Fig. 20. Modelo Físico. Autor (2020)

Para el diseño y tratamiento de los datos fuente, se hará uso de herramientas de analítica e integración de datos para su posterior gestión de reportes y *Dashboards*. Es importante en este punto, de acuerdo al volumen de las fuentes de datos, se seleccionen los recursos adecuados tanto de hardware y como de software para garantizar la efectividad en el procesamiento del ETL.

La Tabla XXIV especifica los recursos de hardware y la Tabla XXV refiere los recursos de software que se aprovecharon para el avance de este proyecto.

Tabla XXIII

RECURSOS DE HARDWARE USADOS

HARDWARE	
Hardware	Características
Modelo	Laptop
Marca	Dell
Procesador	Intel Core i5
Memoria RAM	12 GB
Memoria ROM	1 TB

Nota: Recursos Tecnológicos para el proyecto. Autor (2020)

Tabla XXIV

RECURSOS DE SOFTWARE USADOS

SOFTWARE	
Software	Versión
Microsoft Windows	10
PostgreSql	4.13
Power Designer	16.5
Exaplus	6.0
Knime Analytics Plataforma	3.5.2
Tableau	2019.3

Nota: Recursos Tecnológicos para el proyecto. Autor (2020)

3.4. DESARROLLO

3.4.1. CREACIÓN DEL DATA MART

A partir del modelo físico se puede generar el *Script* de modelo de *Data Mart* usando la misma herramienta de modelado, *SAP Power Designer*. Para la creación del *Script* del modelo, debemos elegir la opción *Change Current DBMS* en el menú *Database* de la herramienta y a continuación seleccionamos la opción de ODBC 3.0. En necesario agregar del ODBC en Windows, de 64 bits para definir la conexión a la base de datos analítica *Exasol*. (Ver Figura 21.)

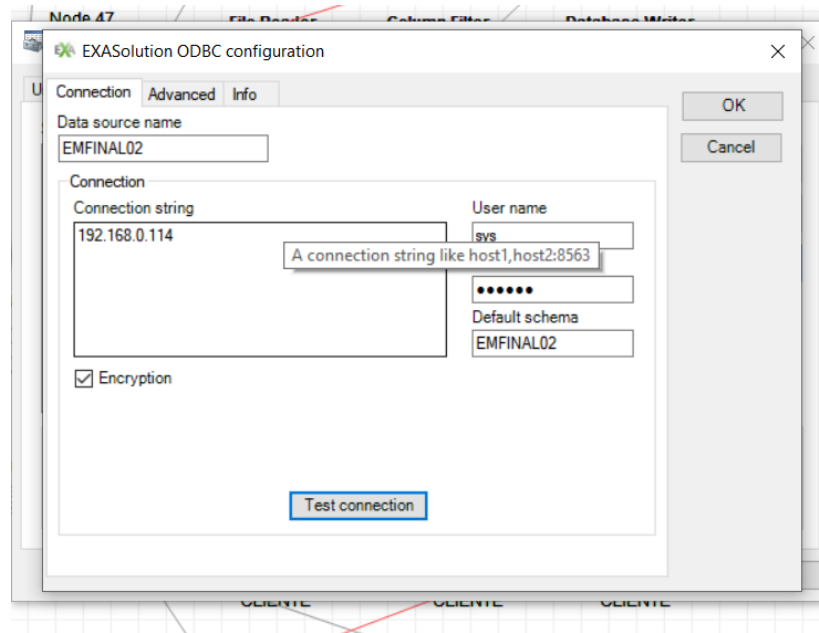


Fig. 21. Conexión a la Base de Datos Analítica. Autor (2020)

Una vez hecha la conexión con la Base de datos analítica *Exaplus*, a continuación; en el menú *Database* de *SAP Power Designer*, se selecciona *Generate DataBase* y luego se elige *Direct Generation*. Se selecciona el ODBC que fue creado anteriormente, luego el directorio donde guardar el *Script* del esquema, se le asigna un nombre que debe ser identificado y localizado fácilmente con EXAPLUS. En esta herramienta, se ejecuta el *Script* con SQL y se generará el Esquema de *Data mart*. El *script* del modelo para la solución propuesta se muestra a continuación.

Script del Modelo del Modelo del Data Warehouse

```

CREATE SCHEMA EMFINAL03;

CREATE TABLE AGUA_EMISION (
ID_EMISION  DECIMAL(36,0) NOT NULL ENABLE,
EMISION     VARCHAR(20) UTF8,
FECHAAPERTURA DATE,
FECHACIERRE  DATE
);

CREATE TABLE UBICACION (

);

CREATE TABLE CLIENTE (
ID_CLIENTE      DECIMAL(36,0) NOT NULL ENABLE,
ID_RUTA         DECIMAL(18,0),
CEDULA         VARCHAR(20) UTF8,
APELLIDO       VARCHAR(100) UTF8,

```

```

NOMBRE          VARCHAR(100) UTF8,
DIRECCION       VARCHAR(100) UTF8,
FECHANACIMIENTO DATE,
DISCAPACIDAD    DECIMAL(18,0),
ENTIDADRELIGIOSAEXONERADA DECIMAL(18,0),

PORCENTAJEEXONERACION DOUBLE PRECISION,
ABONADOCLIENTE DECIMAL(36,0),
NROMEDIDOR      VARCHAR(30) UTF8,
ESTADO          DECIMAL(18,0)
);

```

```

CREATE TABLE FECHA (

```

```

);

```

```

CREATE TABLE TIPO_CATEGORIA (

```

```

);

```

```

CREATE TABLE AGUA_LIQUIDACION (

```

```

);

```

```

CREATE TABLE FACTURA (

```

```

ID_FACTURA DECIMAL(30,10),
NROFACTURA VARCHAR(255) UTF8,
VALORBASE DECIMAL(30,10),
TOTALTARIFA DECIMAL(30,10),

```

```

ABONO DECIMAL(30,10),
PAGADO DECIMAL (18,0)
);

```

```

ALTER TABLE AGUA_EMISION ADD CONSTRAINT
SYS_13238452064968827258806784 PRIMARY KEY(ID_EMISION) ENABLE;

```

```

ALTER TABLE CLIENTE ADD CONSTRAINT SYS_13238452064969153259019776
PRIMARY KEY(ID_CLIENTE) ENABLE;

```

Al ejecutar el *script* en EXAPLUS se crea el esquema del *Data Mart*, quedará como se muestra en la figura 22:

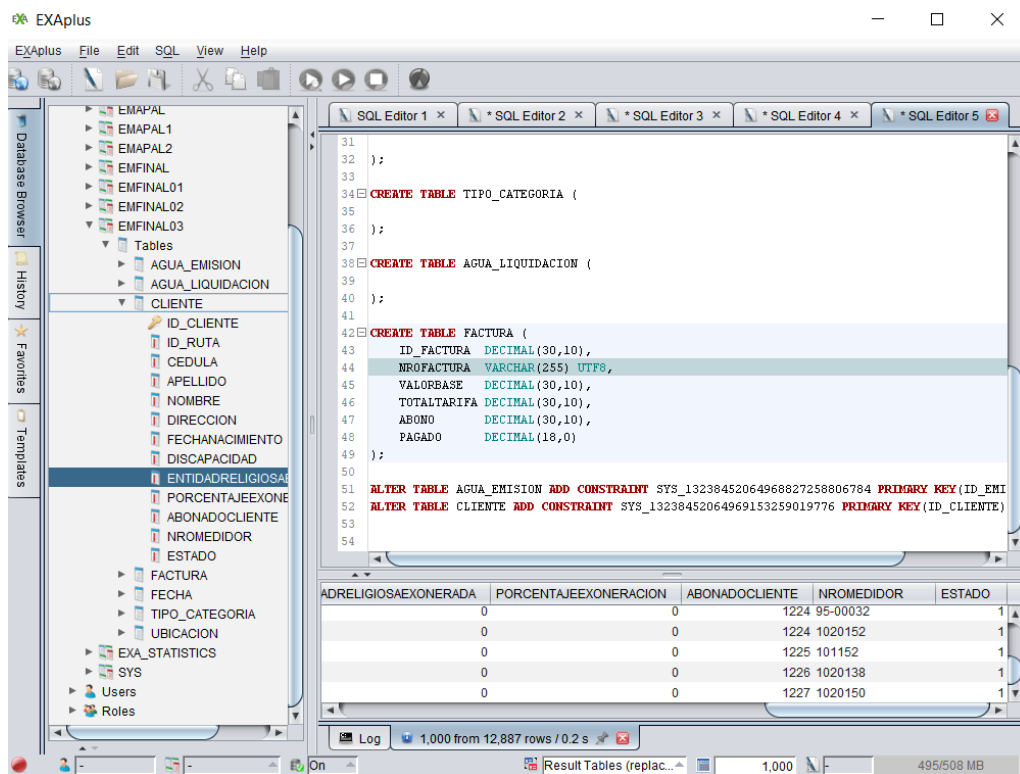


Fig. 22. Esquema del Data Mart - Exaplus. Autor (2020)

3.5. PROCESO ETL

Llegado a este punto es necesario definir el proceso de extracción, transformación y carga de datos en el Almacén de Datos. Es a través del ETL, que se pueden integrar los datos de que provienen de las fuentes de datos en una estructura multidimensional.

- **Extracción:** Este proceso de extracción de datos se lo realizó desde la base de datos SIMM *PostgreSql* de la EMAPAL-EP, en la cual están alojadas todas las tablas y datos necesarios de acuerdo a las necesidades previamente analizadas y que fueron explicadas anteriormente. *Knime Analytics Platform* es la herramienta que permite hacer la extracción de los datos, a través del componente *Database Reader* con una conexión previa a la Base de Datos mediante el nodo *Database Conector*. Es necesario hacer la extracción de cada dimensión y de la tabla de hechos respetando las reglas de integridad referencial. En la presente propuesta la extracción se realizó de las siguientes tablas:
 - AguaEmission.
 - AguaLiquidacion.
 - Cliente.
 - Abonado.
 - Factura.
 - TipoCategoria.
 - Ubicación.
 - Ruta.
 - Sector.

- Zona.
- Parroquia.
- Cantón.
- Provincia.

La extracción de los datos requirió de sentencias SQL en las diferentes tablas para combinarlas, cambiar los tipos de datos y otros. Así también se utilizó el componente *Column Filter*, para filtrae los campos que serán cargados posteriormente.

- **Transformación:** Tras la extracción de datos se procede a la transformación para posteriormente cargar los datos. Se puede decir que la transformación es uno de los procesos más importantes; ya que, una transformación de datos correcta garantizará la limpieza de los datos y su análisis. En este punto los datos que fueron filtrados en la extracción, requieren aplicar operaciones de manipulación de números, cadenas de caracteres, definición de reglas o uso de funciones. Se requirieron transformaciones de los tipos de datos, entre otras.
- **Carga:** Cuando los datos han sido tratados, lo siguiente es cargarlos al esquema que fue creado en EXAPLUS y dejarlos disponibles para que sean posteriormente explotados con la plataforma de Inteligencia de Negocios *Tableau*. Ésta se encargará de la generación de datos útiles.

Para poblar el esquema del *Data Mart*, se utilizó el componente *Database Writer* de *Knime* el cual debe estar conectado directamente al *Database Connector*. Con este se crea una conexión a una Base de datos Exaplust, mediante un controlador y una dirección URL de la base de datos, es en *Exaplust* donde los datos son cargados, para posteriormente mediante *Tableau* establecer una conexión con el esquema en el cual fueron cargados los datos.

Para el procesamiento de los datos con la herramienta *Knime*, se hace un procesamiento secuencial, es decir, se ejecuta cada proceso seguido del otro como se visualiza en la Figura. 23.

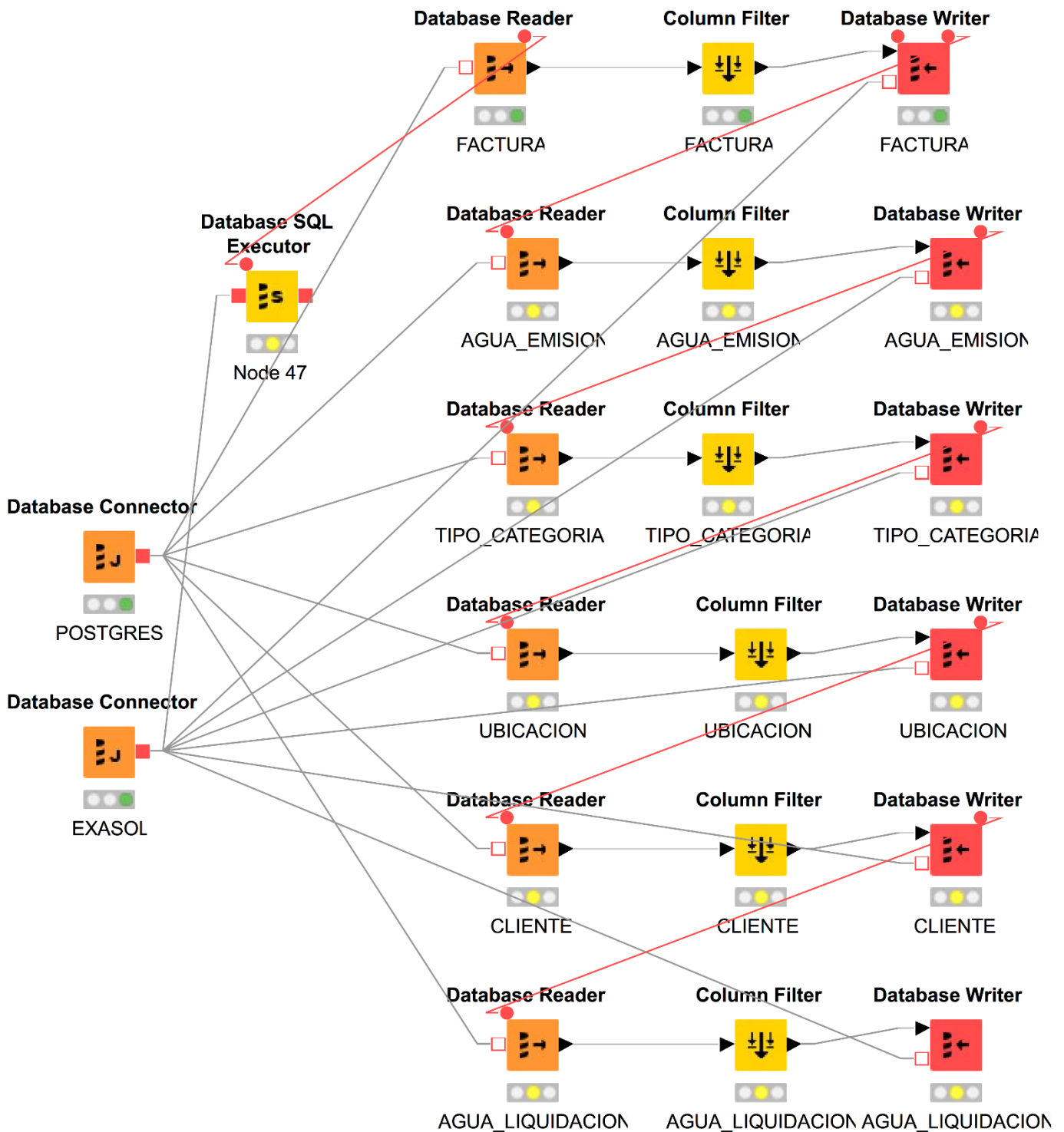


Fig. 23. DWH. Autor(2020)

Para terminar con el desarrollo, luego de finalizado el proceso ETL, se procedió a construir el OLAP mediante la plataforma de Inteligencia de Negocios Tableau. Para iniciar se deberá establecer la conexión del *Data mart* con *Tableau*, y en base a los requerimientos de *reporting* y consultas definidos, se usan las herramientas de visualización de *Tableau* más adecuadas.

Tableau es una plataforma que cuenta con una variedad de graficas de visualización, destinadas a diversas tareas, con la finalidad de convertir la información en conocimiento, expresándolo de manera fácil, rápida y comprensible. Las herramientas que más se adaptaron para el uso de visualización de información en este proyecto destacan: [45]

- **Gráfico de barras:** Este tipo de grafica de visualización es uno de los más utilizados en el medio. Es muy factible para realizar comparaciones rápidas de datos de distintas categorías, además resalta las diferencias de los mismos. Otra de sus funciones es la de mostrar tendencias y valores atípicos, revelar valores mínimos y máximos históricos con solo un vistazo. Estos gráficos son muy eficaces cuando sus datos se pueden dividir en múltiples categorías. Por lo cual fue elegido para filtrar información de los clientes. Para un mayor impacto en el uso de este grafico se recomienda incorporar color a las barras, utilizar barras apiladas o en paralelo con el fin de dividir los datos diversas veces y de esta manera se realiza un análisis profundo, por último, se sugiere la colocación de las barras a ambos lados de un eje.
- **Gráfico de líneas o lineal:** El grafico de líneas o también denominado grafico lineal conecta varios puntos de datos y los presenta como una evolución de tipo continua. Este tipo de gráfico de visualización se utiliza con el fin de ver tendencias en los datos, por lo general durante un periodo. El resultado que presenta es una forma simple de visualizar los cambios de un valor en relación con otro. Los gráficos lineales no se limitan al tiempo. En eje horizontal, se puede usar cualquier dimensión, como tipos de datos, intervalos de tiempo y otros datos ordinales. Por lo cual fue el grafico de elección para representar el total de tarifas de la organización, así como también para la búsqueda de un cliente en específico.
- **Gráfico circular o pastel:** se utiliza la visualización a detalla de la información de manera rápida. Este tipo de grafico fue el electo para visualizar el porcentaje de M3 de agua consumido.

3.6. Pruebas.

La verificación del correcto funcionamiento y cumplimiento de las necesidades de información iniciales, se llevaron a cabo en la etapa de pruebas. En este punto, cada vez que se detectaron algunas inconsistencias y vacíos analizadas con los responsables de TI, se daba inicio a un nuevo *sprint* que retroalmente estos inconvenientes desde el diseño del modelo de datos.

El proceso de pruebas no se basó en un plan de pruebas, sino que se fueron desarrollando de acuerdo a la necesidad al finalizar cada *sprint*. El proceso se llevó a cabo, tal como se describe en los siguientes apartados:

3.6.1. Instalación del ambiente de pruebas.

El proceso de pruebas se llevó a cabo en el área de TICs de la EMAPAL luego de una cita previamente agendada. Se utilizó un *Backup* de la base de datos SIMM la cual contiene toda la información de las recaudaciones. Se instalaron versiones de prueba de las herramientas: *Sap Power Designer*, *PostgreSQL*, *Exasol*, *Exaplus*, *Knime* y *Tableau*.

3.6.2. Verificación del Data Mart

La base de datos de donde fue extraída la información fue desde *PostgreSQL*, fue organizada, depurada e integrada para su correcto funcionamiento y análisis.

3.6.3. Pruebas ETL

Se realizó la extracción de la información de la base de datos SIMM actualizada, se filtró información necesaria y se cargó en la base de datos EXAPLUS.

3.6.4. Pruebas de Interfaces de Visualización

Para constatar el correcto funcionamiento y el cumplimiento de las necesidades de información solicitadas inicialmente, se procedió a realizar el OLAP de la información consolidada, con las herramientas de visualización de *Tableau*.

- Para visualizar la situación financiera de la empresa EMAPAL, se usó la herramienta de líneas de tendencia tal como fuera solicitada en los requerimientos. Permite filtrar la información de las tarifas de manera total o por rangos de fechas.
- Para comprobar la consistencia de la información se hicieron consultas puntuales de información, por ejemplo, búsqueda de usuarios por número de identificación. La consulta muestra información importante del usuario tal como el servicio contratado. Esta información puede ser consultada por fechas.
- La consulta de la cantidad de usuarios del servicio, se verificó que la información obtenida es la correcta.
- La cantidad de metros cúbicos de agua que se han consumido, se obtuvo la información clasificada por categorías, dando cumplimiento a una necesidad urgente de la empresa y las expectativas del usuario como se visualiza en la figura 27.

3.7. Puesta en marcha.

La puesta en producción de la solución, ha quedado condicionada a que, la EMAPAL gestione las licencias de las plataformas y herramientas que fueron usados en el presente trabajo. En este contexto el presente trabajo presenta un análisis del Costo Total de Propiedad del software, el cual fue calculado para un período de tres años. Este análisis fue realizado para todas las herramientas, excepto EXASOL, por la falta de disponibilidad de información sobre licenciamiento de esta plataforma. El costo total de propiedad de las herramientas costeadas en el apartado 1.6.6 se indican en la tabla XXVI.

Este es un costo aproximado que puede ser tomado como referencia para la implementación de la solución de BI en el área de comercialización de la EMAPAL. Este costo incluye el costo

de la inversión inicial, renovaciones, soporte y mantenimiento de algunas herramientas excepto para EXASOL por las razones antes indicadas.

Tabla XXV.

COSTO TOTAL DE PROPIEDAD PARA EMPAL (Estimado) – 3 AÑOS

Proveedor	Precio de Licencia/ Anual
Tableau	\$15.799,00
SAP Power Designer	\$10.800,00
Knime Analytics Platform	\$56.112,48
Exaplus	\$10.680,00
PostgreSQL	\$0.00
TOTAL	93.391,48

3.8. RESULTADOS

Los resultados obtenidos para cumplir la Historia de Usuario N1 se muestran en la Figura 24.

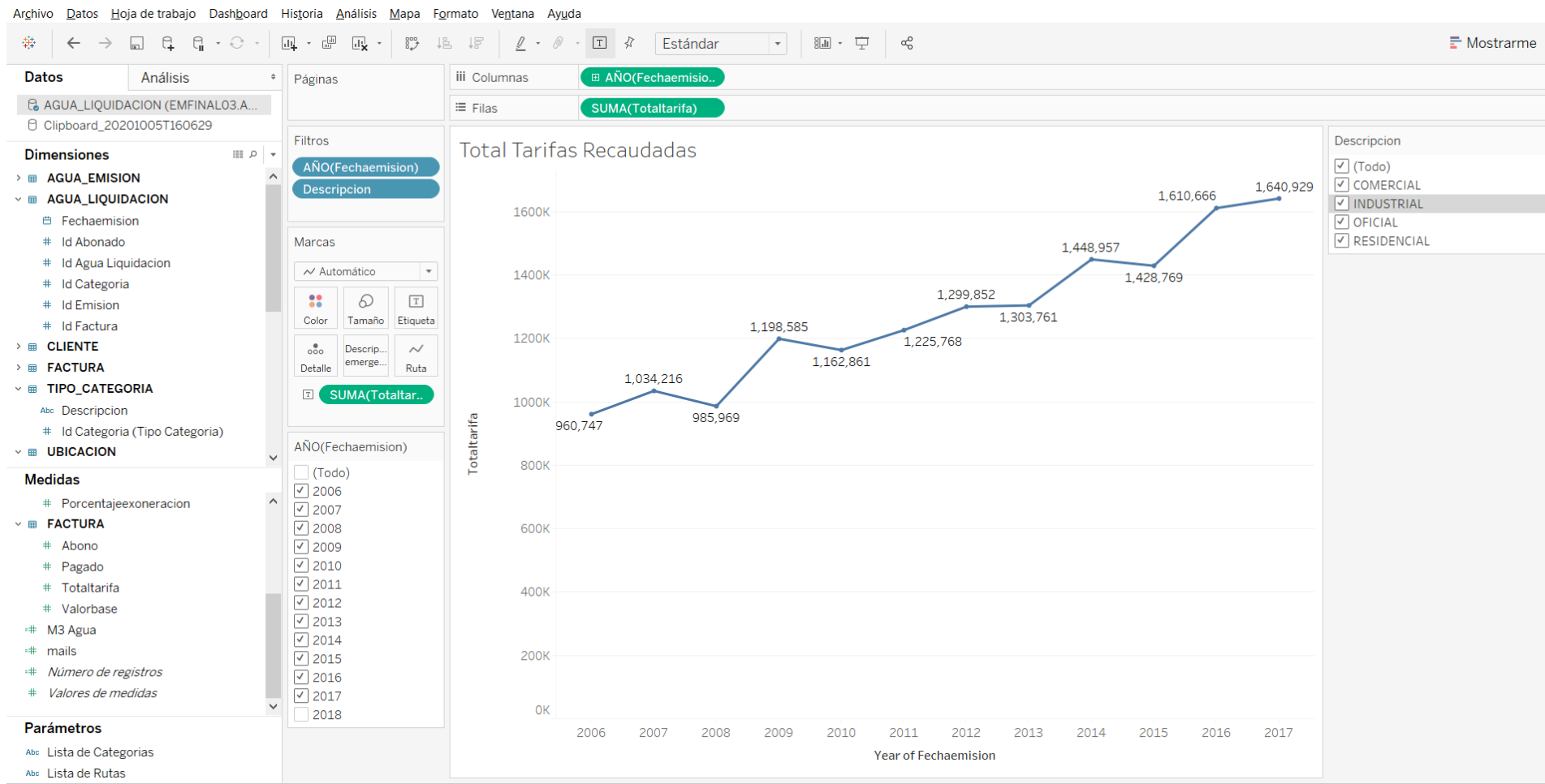


Fig. 24 Tarifas Recaudadas. Autor (2020)

Los resultados obtenidos para cumplir la Historia de Usuario N2 se muestran en la Figura 25.

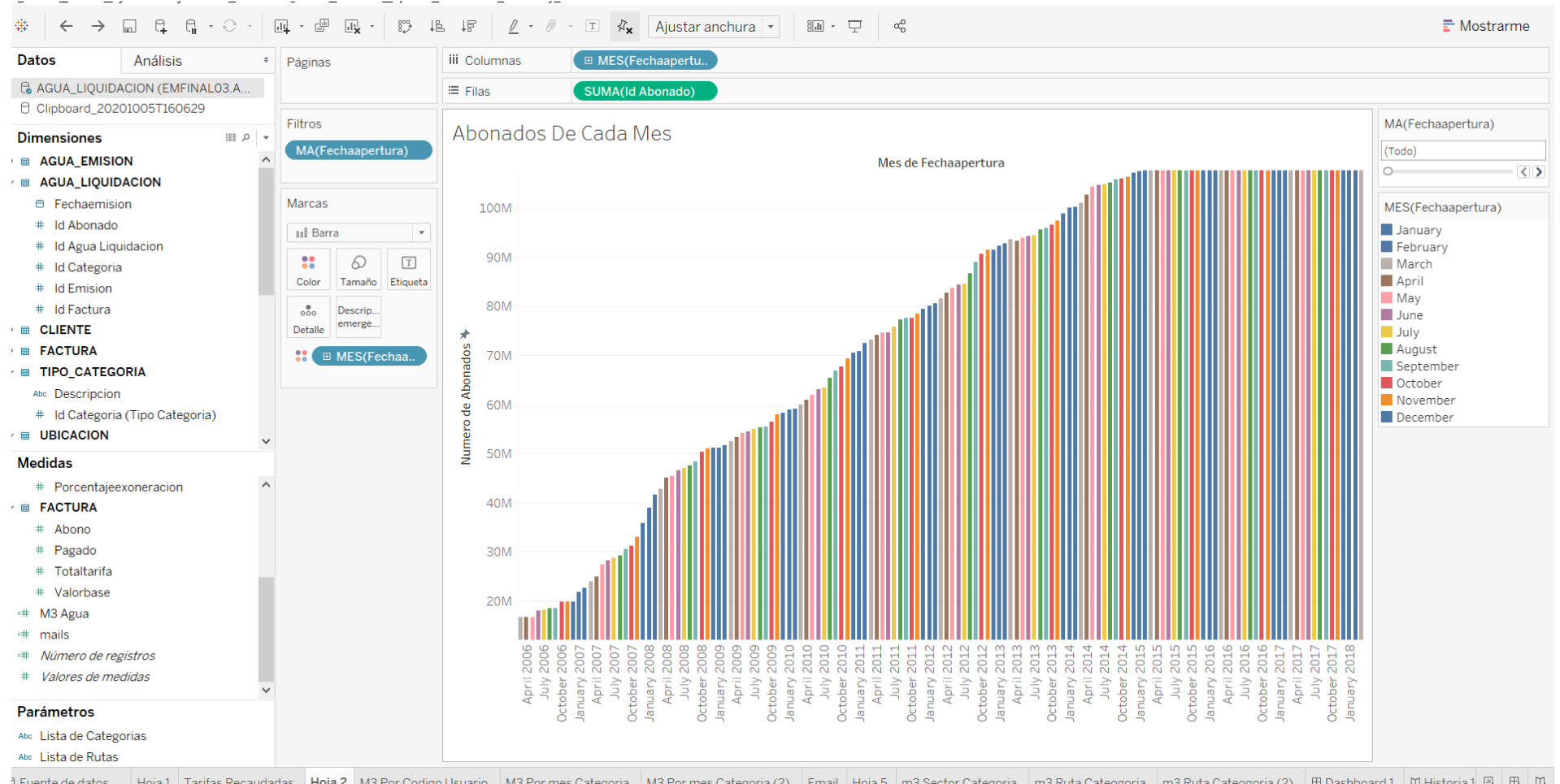


Fig. 25 Abonados Por Mes. Autor (2020)

Los resultados obtenidos para cumplir la Historia de Usuario N3 se muestran en la Figura 26.

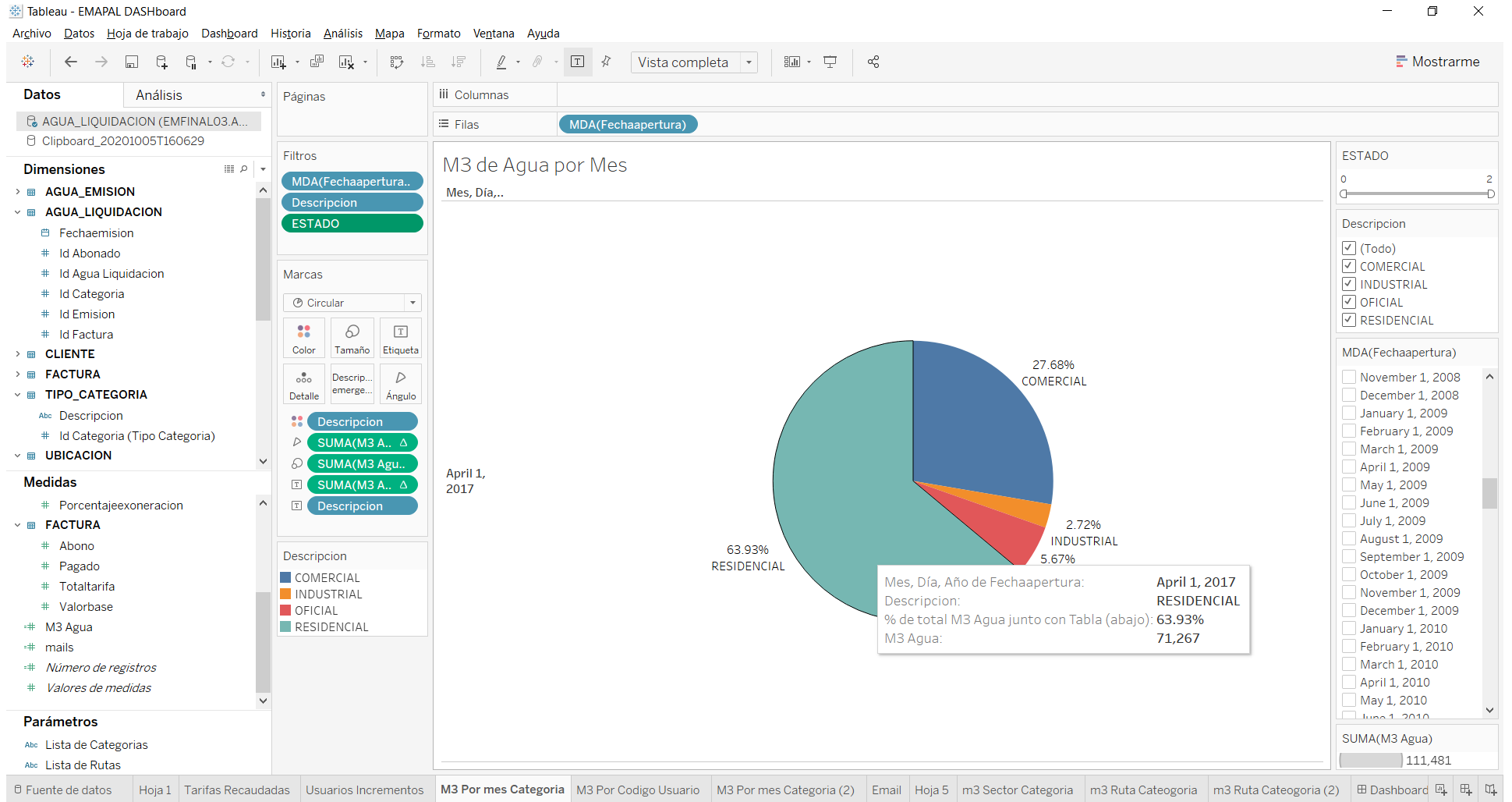


Fig. 26. M3 Consumidos. Autor (2020)

Los resultados obtenidos para cumplir la Historia de Usuario N4 se muestran en la Figura 27.

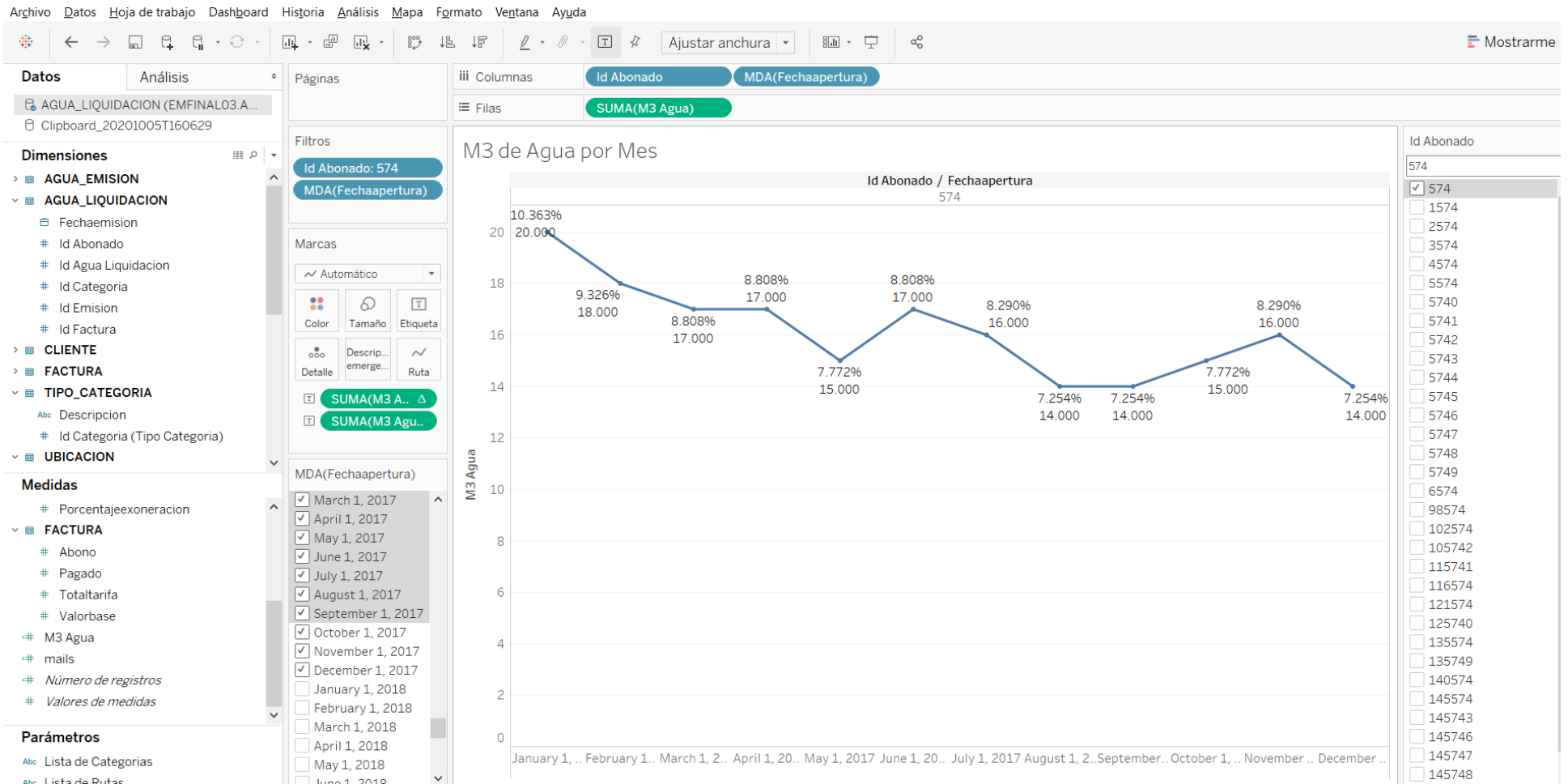


Fig. 27 M3 de Cliente. Autor (2020)

Finalmente se agruparon y organizaron los resultados en un *Dashboard* el cual se ilustra en la Figura 28.

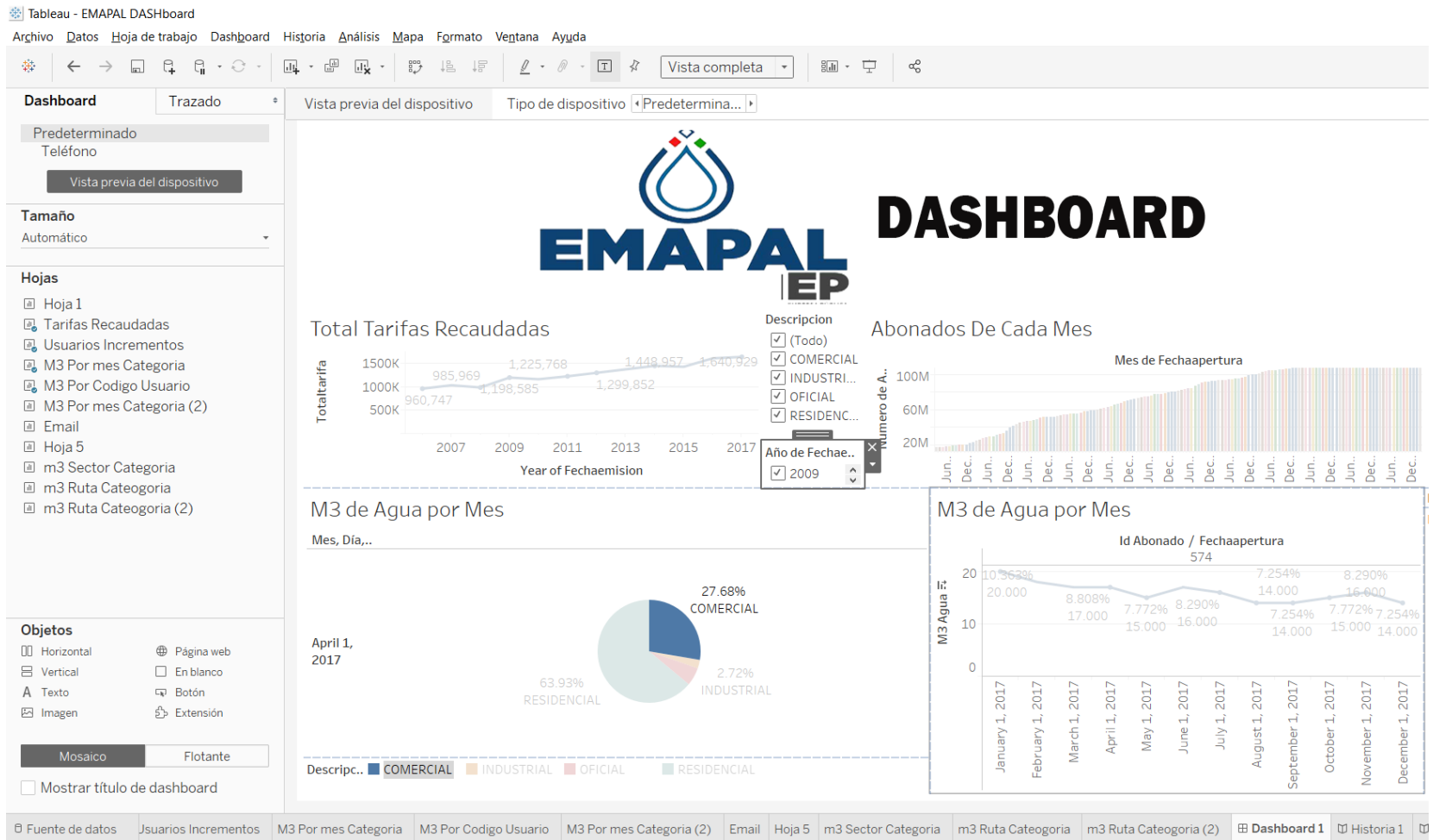


Fig. 28 Dashboard Emopal. Autor (2020)

CONCLUSIONES.

La realización de este trabajo ha dejado algunas lecciones aprendidas y hallazgos más importantes, los cuales se explican a continuación:

- Los beneficios de usar las plataformas BI, para obtener información útil en la toma de decisiones, incrementa el valor de las organizaciones, ya sea por de la optimización de recursos, la disminución de costos, el mejoramiento de la productividad y rendimiento, el incremento de utilidades u otro. Para la EMAPAL-EP contar por ejemplo con información oportuna sobre los volúmenes de distribución del líquido vital; que, al ser comparados con los volúmenes en los tanques, les permitirá determinar los niveles precisos de pérdida de este valioso recurso, evaluar la magnitud y el grado de atención que requiere este importante problema por parte de la Alta Dirección para una atención urgente.
- La facilidad de contar con información oportuna sobre aspectos de la recaudación de EMAPAL a nivel operativo y táctico; ha necesitado la ejecución de varios *sprints* para este ejercicio de BI. El esfuerzo fundamentalmente se debe al trabajo relacionado con el diseño, creación y poblado del *data mart*. Una vez que éste ha sido cargado, las consultas pueden ser fácilmente ejecutadas, lo que posibilitará a los mismos usuarios gestionar otras consultas basadas en el actual modelo dimensional.
- La actualización periódica de la información en el *data mart*, podrá ser ejecutada por los responsables de TI; puesto que, en el ETL que ha se muestra en la figura 21, se ha considerado el proceso de borrado de la data para recargarlo con información actualizada. Este proceso es ejecutado mediante el componente Database SQL Executor conjuntamente con los componentes que permiten repoblar el *data mart*.
- El costo total de propiedad que ha sido definido en el Capítulo III, evidencia que la inversión que EMAPAL deberá hacer a tres años representa un costo mínimo frente a los grandes beneficios que se irán visibilizando a medida que utilice la solución y, cumpla con otras necesidades de información que orienten una toma de decisiones informadas que agreguen valor. Los beneficios crecerán proporcionalmente a medida que se ejecuten más ejercicios de BI en los que se vayan agregando data marts de áreas relacionadas hasta lograr crear un Cuadro de Mando Integral que permita toma de decisiones a nivel estratégico.

RECOMENDACIONES

A modo de sugerencias que garanticen la sostenibilidad de este trabajo, las que deben ser acciones que se deberán tomar mediano y largo plazo por parte de la EMAPAL:

- Se recomienda a la EMAPAL tomar esta solución como el primer paso de una transformación institucional que deberá ser complementada con varios ejercicios de BI que aborden otras áreas críticas de la empresa como son: la Dirección de Planificación, la Dirección Técnica, la Dirección Administrativa y la Dirección Financiera. Para ello se deberá procurar tener procesos definidos que permitan flujos de información entre las diferentes áreas funcionales.

- Al momento que la empresa, opte por implementar la presente solución de BI, con el fin de mejorar la administración de la información de la organización, deberá tener en cuenta la capacidad de los equipos tecnológicos, contar con un equipo técnico y profesional especializado; así como también, considerar el costo total de propiedad aproximado del software que ha sido planteado en este trabajo como un referente para la adquisición de licencias de las herramientas y plataformas.

BIBLIOGRAFÍA

- [1] B. H. Springer, *Adaptive Business Intelligence*, 2007.
- [2] C. Osorio , L. Zúñiga y H. Muñoz, «Inteligencia de los negocios. Clave del Éxito en la era de la información,» *Clío América*, vol. 10, pp. 194-211, 2016.
- [3] S. Negash, «Inteligencia de Negocios,» 15 02 2004. [En línea]. Available: <https://aisel.aisnet.org/cais/vol13/iss1/15/>.
- [4] J. L. Cano, *Business Intelligence: Competir Con Información*, Barcelona: Banesto, Fundacion Cultural, 2007.
- [5] J. L. A. Calzada Leticia, «El impacto de las Herramientas de Inteligencias de Negocios en la toma de decisiones de los ejecutivos,» *International Journal Of Good Conscience* , vol. 4, n° 2, pp. 16-52, 2009.
- [6] H. Muñoz H, R. C. Osorio M y L. M. Zúñiga P, *Inteligencia de los negocios Clave del éxito en la era de la información*, Santa Marta: Clio America, 2016.
- [7] Sinnexus, «Sinergia e Inteligencia de Negocio,» 2007. [En línea]. Available: https://www.sinnexus.com/business_intelligence/datawarehouse.aspx. [Último acceso: 7 Julio 2020].
- [8] A. Mendez, «Fundamentos de Data Warehouse,» *Centro de Actualización Permanente en Ingeniería del Software*, vol. 5, n° 1, pp. 19-26.
- [9] T. Naeem, «Astera Enabling Data-Driven Innovation,» *Astera*, 10 09 2020. [En línea]. Available: <https://www.astera.com/es/type/blog/data-warehouse-concepts/>. [Último acceso: 14 09 2020].
- [10] P. Musso, «monografias.com,» [En línea]. Available: [https://www.monografias.com/trabajos90/datawarehouse-kimball-y-sql-2005/datawarehouse-kimball-y-sql-2005.shtml#:~:text=niveles%20de%20granularidad\),Metodolog%C3%ADa%20de%20Ralph%20Kimball,de%20construir%20un%20Data%20Warehouse..](https://www.monografias.com/trabajos90/datawarehouse-kimball-y-sql-2005/datawarehouse-kimball-y-sql-2005.shtml#:~:text=niveles%20de%20granularidad),Metodolog%C3%ADa%20de%20Ralph%20Kimball,de%20construir%20un%20Data%20Warehouse..) [Último acceso: 16 09 2020].
- [11] F. Wass , R. Wrembel y T. Freudenreich, «On-Demand ELT Architecture for Right-Time BI: Extending the Vision,» *International Journal of Data Warehousing and Mining*, vol. 9, p. 18, 2013.
- [12] T. Balagueró, «DEUSTO FORMACION,» 04 07 2018. [En línea]. Available: <https://www.deustoformacion.com/blog/gestion-empresas/big-data-componentes-business-intelligence>. [Último acceso: 2020 09 03].

- [13] R. Kimball , M. Ross , W. Thornthwite, J. Mundi y B. Becker, The Dara Warehouse Lifecycle Toolkit, 2018.
- [14] Z. Tejada, «Microsoft Docs,» 20 11 2019. [En línea]. Available: <https://docs.microsoft.com/es-es/azure/architecture/data-guide/relational-data/etl>. [Último acceso: 2020 09 04].
- [15] D. Prudenciano, «Tecnomark,» 17 04 2016. [En línea]. Available: <http://tecnomark.es/etl-proceso-herramientas/>. [Último acceso: 09 09 2020].
- [16] L. Muñoz, «Repositorio Institucional de la Universidad de Alicante,» 11 07 2016. [En línea]. Available: https://rua.ua.es/dspace/bitstream/10045/21179/1/9TLA3_18Munoz.pdf. [Último acceso: 10 09 10].
- [17] ITSO, DAta Mart Consolidation Getting Control of Your Enterprise Information, 2005.
- [18] D. Tobin, «Xplenty,» 15 06 2020. [En línea]. Available: <https://www.xplenty.com/blog/data-mart-vs-data-warehouse/>. [Último acceso: 25 10 2020].
- [19] C. Ballard, A. Gupta , V. Krishnan, N. Pessoa y O. Stephan , Data Mart Consolidation : Getting Control of Your Enterprise Information, 2005.
- [20] M. Munch, «dbi services,» 2015 01 29. [En línea]. Available: <https://blog.dbi-services.com/birst-an-attractive-all-in-one-business-intelligence-solution/>. [Último acceso: 2020 14 11].
- [21] E. Martinez , «ITS EL GRULLO Inteligencia de negocios,» [En línea]. Available: <https://sites.google.com/site/itsginteligenciadenegocios/home/1-2-componentes-de-la-inteligencia-de-negocios>. [Último acceso: 7 Julio 2020].
- [22] B. I. Geek, «Arquitectura BI: Los enfoques de William H. Inmon y Ralph Kimball,» medium, 2020. [En línea]. Available: <https://medium.com/@bigEEK/arquitectura-bi-los-enfoques-de-william-h-inmon-y-ralph-kimball-9d4bc57d3ab4>.
- [23] L. Rao, M. L y G. Mansingh, Business Intelligence for small and medium- sized Enterprises an agile roadmap toward business sustainability, 2020.
- [24] N. Rayner , B. Hostmann y G. Herschel, «Gartner Research,» 19 10 2019. [En línea]. Available: <https://www.gartner.com/en/documents/1209327>. [Último acceso: 26 10 2020].
- [25] S. Normiña, «SISTEMA DE BUSINESS INTELLIGENCE PARA LA GESTIÓN DE ATENCIÓN TÉCNICA DE RECLAMOS EN LA EMPRESA ELÉCTRICA RIOBAMBA S.A.,» Ambato , 2016.

- [26] M. Word, «ORACLE,» [En línea]. Available: https://www.oracle.com/ocom/groups/public/@otn/documents/webcontent/317529_esa.pdf. [Último acceso: 30 10 2020].
- [27] R. Hughes, Agile Data Warehousing Project Management Business Intelligence Systems Using Scrum, Usa, 2013.
- [28] Blog soluciones IT, «Scrum,» <http://www.atuservicio.net/>, 2018. [En línea]. Available: <http://www.atuservicio.net/scrum/>. [Último acceso: 2020].
- [29] E. Buitrago, Metodología de desarrollo de proyectos de inteligencia de negocios, Universidad Católica Andrés Bello, 2004.
- [30] S. PowerDesigner. [En línea]. Available: <https://www.sap.com/latinamerica/products/powerdesigner-data-modeling-tools.html>. [Último acceso: 25 09 2020].
- [31] L. Shiff, «bmc blogs,» 2020 09 22. [En línea]. Available: <https://www.bmc.com/blogs/gartner-magic-quadrant-data-quality-solutions/>. [Último acceso: 2020 14 11].
- [32] «Gartner,» Gartner, 24 10 2018. [En línea]. Available: <https://www.gartner.com/reviews/market/data-warehouse-solutions/vendor/exasol/product/exasol-enterprise-cluster/review/view/578232>. [Último acceso: 28 10 2020].
- [33] Exasol, «Exasol Community,» 2020. [En línea]. Available: <https://www.exasol.com/portal/pages/viewpage.action?pageId=4030482>. [Último acceso: 25 09 2020].
- [34] SAP, «Sap Power Designer,» Sap Power Designer , 16 10 2019. [En línea]. Available: <https://www.sap.com/cmp/dg/crm-ya19-pdp-gartner/typ.html>.
- [35] G. d. d. g. d. PostgreSQL, «PostgreSQL: The World's Most Advanced Open Source Relational Database,» [En línea]. Available: <https://www.postgresql.org/>. [Último acceso: 06 10 2020].
- [36] O. f. I. Knime, «Open for Innovation Knime,» [En línea]. Available: <https://www.knime.com/knime-analytics-platform>. [Último acceso: 25 09 2020].
- [37] Gartner, «Cuadrante Mágico Para Plataformas De Diencia De Datos y Aprendizaje Automático,» Gartner, 11 02 2020. [En línea]. Available: <https://www.gartner.com/doc/reprints?id=1-1YDUKTC6&ct=200217&st=sb>.

- [38] I. Bobriakov, «Data Science Central,» 16 01 2019. [En línea]. Available: <https://www.datasciencecentral.com/profiles/blogs/tableau-in-10-minutes-step-by-step-guide>. [Último acceso: 25 09 2020].
- [39] Tableau, «Tableau,» [En línea]. Available: <https://www.tableau.com/why-tableau/what-is-tableau>. [Último acceso: 25 09 2020].
- [40] Microsoft, 2020. [En línea]. Available: <https://info.microsoft.com/ww-landing-2020-gartner-magic-quadrant-for-analytics-and-business-intelligence.html?LCID=EN-US&ls=Website>. [Último acceso: 2020 14 11].
- [41] Tableau, «Tableau,» Tableau , [En línea]. Available: <https://www.tableau.com/why-tableau/what-is-tableau>.
- [42] EMAPAL, «Crecimiento Empresarial,» Emapal, 2020.
- [43] . S. E. Recalde Cuti, Análisis y Propuesta de una herramienta de Bussines Intelligence que permita mejorar la toma de decisiones gerenciales en la empresa Soldeneg Soluciones de Negocios CÍA.Ltda, Universidad Central del Ecuador, Facultad de Ciencias Administrativas, 2018.
- [44] D. E. M. ZAMORA, Desarrollo de un aplicativo Business Intelligence para la empresa importadora Tomebamba S.A., Universidad de las fuerzas armadas ESPE, 2015.
- [45] T. Rodgers, «Tableau,» 2020. [En línea]. Available: <https://www.tableau.com/es-es/learn/whitepapers/which-chart-or-graph-is-right-for-you>. [Último acceso: 15 11 2020].
- [46] J. Salas, «Diseño de una solución de inteligencia de negocios como herramienta de apoyo a la toma de decisiones en el área de ventas de la empresa farmacéutica Dispefarma,» Lima , 2017.
- [47] A. B. Martínez, «Técnicas de modelado de procesos de ETL: una revisión de alternativas y su aplicación en un proyecto de desarrollo de una solución de BI,» *Scientia et Technica*, vol. 18, n° 1, pp. 185-191, 2013.
- [48] P. M. Gutierrez, «DATA WAREHOUSE: MARCO DE CALIDAD.,» [En línea]. Available: <https://core.ac.uk/download/pdf/30046568.pdf>. [Último acceso: 22 09 2020].
- [49] J. M. Marysol Tamayo, *Revista Ingenieria e Invetigacion* , vol. 26, n° 3, pp. 135-142, 2006.
- [50] SAP, «Sap Power designer,» Sap Power designer , [En línea]. Available: <https://www.sap.com/cmp/td/sap-powerdesigner-trial.html>.
- [51] R. Kimball, M. Ross, W. Thornthwite, J. Mundi y B. Becker, The DAta Warehouse Lifecycle Toolkit, 2da ed., John Eiley & Scops Incomeraded, 2008.

ANEXOS

ANEXO. A ANTIPLAGIO

INTELIGENCIA DE NEGOCIOS APLICADA AL ÁREA DE COMERCIALIZACIÓN DE LA EMPRESA PÚBLICA MUINCIPAL DE AGUA POTABLE, ALCANTARILLADO Y SANEAMIENTO AMBIENTAL (EMAPAL-EP) DE AZOGUES

INFORME DE ORIGINALIDAD

4%

INDICE DE SIMILITUD

5%

FUENTES DE
INTERNET

0%

PUBLICACIONES

1%

TRABAJOS DEL
ESTUDIANTE

FUENTES PRIMARIAS

1

www.slideshare.net

Fuente de Internet

2%

2

documents.mx

Fuente de Internet

1%

3

www.plusformacion.com

Fuente de Internet

1%

4

revistas.unimagdalena.edu.co

Fuente de Internet

1%

Excluir citas

Apagado

Excluir coincidencias

< 1%

Excluir bibliografía

Activo



ANEXO. B PERMISO DE SUBIR LA TESIS AL REPOSITORIO



PERMISO DEL AUTOR DE TESIS PARA SUBIR AL REPOSITORIO INSTITUCIONAL

Yo, **RIVERA SANCHEZ DARWIN ARMANDO**, portador (a) de la cédula de ciudadanía Nro. **0302479613**. En calidad de autor/a y titular de los derechos patrimoniales del trabajo de titulación "INTELIGENCIA DE NEGOCIOS APLICADA AL ÁREA DE COMERCIALIZACIÓN DE LA EMPRESA PÚBLICA MUNICIPAL DE AGUA POTABLE, ALCANTARILLADO Y SANEAMIENTO AMBIENTAL (EMAPAL EP) DE AZOGUES" de conformidad a lo establecido en el artículo 114 Código Orgánico de la Economía Social de Los Conocimientos, Creatividad e Innovación, reconozco a favor de la Universidad Católica de Cuenca una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos, Así mismo; autorizo a la Universidad para que realice la publicación de éste trabajo de titulación en Repositorio Institucional de conformidad a lo dispuesto en el artículo 144 de la Ley Orgánica de Educación Superior.

Azogues, 11 de noviembre de 2020

F: 

RIVERA SANCHEZ DARWIN ARMANDO
0302479613

www.ucacue.edu.ec

ANEXO. C. CERTIFICADO DE NO ADEUDAR LIBROS A LA BIBLIOTECA



El Bibliotecario de la Sede Azogues

CERTIFICA:

Que, **RIVERA SÁNCHEZ DARWIN ARMANDO**. Con cédula de ciudadanía **Nro.**

0302479613 de la carrera de **INGENIERÍA EN SISTEMAS**.

No adeuda libros, a esta fecha.

Azogues, 12 de noviembre del 2020.

A handwritten signature in blue ink, appearing to read 'Byron Torres Romo', written over a light blue rectangular background.

Byron Alonso Torres Romo
BIBLIOTECARIO

Biblioteca Universitaria
MONS. FROILAN POZO QUEVEDO

www.ucacue.edu.ec

ANEXO. D ACTA DE CONSTITUCION DEL PROYECTO

ACTA DE CONSTITUCIÓN DEL PROYECTO

Fecha	22-04-2019	Nº de Proyecto	1
Nombre del Proyecto	INTELIGENCIA DE NEGOCIOS APLICADA AL ÁREA DE COMERCIALIZACIÓN DE LA EMPRESA PÚBLICA MUNICIPAL DE AGUA POTABLE, ALCANTARILLADO Y SANEAMIENTO AMBIENTAL (EMAPAL-EP) DE AZOGUES.		

1. Nombre de los Proyectos Revisados

- INTELIGENCIA DE NEGOCIOS APLICADA AL ÁREA DE COMERCIALIZACIÓN DE LA EMPRESA PÚBLICA MUNICIPAL DE AGUA POTABLE, ALCANTARILLADO Y SANEAMIENTO AMBIENTAL (EMAPAL-EP) DE AZOGUES.

2. Presupuesto Asignado

\$0,00

3. Objetivo General.

Implementar una solución de BI para apoyar la extracción de conocimiento del área de comercialización en la EMAPAL.




4. Objetivos Específicos.

1. Determinar requerimientos de la información.
2. Diseñar la solución de BI en base a las funcionalidades requeridas (diseño de BI).
3. Construcción de la solución de BI.
4. Probar e implementar la Solución de BI.

5. Restricciones.

Restricción:
Falta de herramientas open Source, para trabajar liberadamente.
Falta de una metodología específica para la construcción de una solución de BI.

6. Aprobaciones

Nombre	Cargo en el proyecto	Firma	Fecha (DD/MM/AAAA)
A.S. Giovanni Morocho Quinteros	Jefe de la Unidad de Sistemas EMAPAL-EP		22-04-2019
Ing. Blanca Ávila	Tutora del Proyecto		22-04-2019
Darwin Rivera	Estudiantes		22-04-2019

ANEXO. E CERTIFICADO DEL PROYECTO REALIZADO EN LA EMPRESA



CERTIFICADO

Por medio del presente dejo constancia que el Señor **Darwin Armando Rivera Sánchez** portador de la Cedula de Identidad N° 0302479613 estudiante de Facultad de Ingeniería de Sistemas de la Universidad Católica de Cuenca Extensión Azogues, realizo su proyecto de Titulación en la Unidad de Sistemas de la Empresa Pública Municipal de Agua Potable, Alcantarillado y Saneamiento Ambiental del Cantón Azogues, con el desarrollo de un proyecto en el tema de "Inteligencia de Negocios **BI**", aplicada a los procesos del área comercial de la Empresa EMAPAL-EP.

El estudiante **Darwin Armando Rivera Sánchez** realizo su proyecto de Titulación a completa satisfacción, mostrando total responsabilidad, eficiencia y buena formación académica.



Razones por las cuales se le otorga el presente certificado, para los fines que el interesado considere conveniente.

Azogues, 19 de octubre de 2020

A. S. Giovanni Morocho Quinteros
JEFE DE LA UNIDAD DE SISTEMAS EMAPAL-EP



ANEXO. F ESPECIFICACIONES DEL SERVIDOR

		
Ing. Giovanni Morocho	Gerente Departamento de Sistemas	EMAPAL EP
Participantes:	Tec. Andrés Ambrosi	AMBROSS-TI Cía. Ltda.
Ciudad y Fecha:	En la Ciudad de Azogues al 18 de febrero 2020	
Proyecto:	Mejora y optimización de la infraestructura Tecnológica	
Actividad (cronograma):	Entrega de Productos	

ACTA ENTREGA - RECEPCIÓN

PARTES Y PIEZAS

- **Procesador**
 - DL380 Gen9 Intel Xeon E5-2640v4 (2.4 GHz/10-core/25MB/90 W) processor kit - 817937-B21
 - S/N: J720C376
- **Memoria RAM**
 - HP 16GB (1x16GB) single rank x4 DDR4-2400 CAS-17-17-17 registered memory kit - 805349-B21
 - S/N: SH69XM6
 - S/N: SH5HVM6
- **DISCO SOLIDO**
 - HPE 960GB SATA 6G MU SFF (2.5 in) SC 3 yr warranty DS firmware SSD - P07926-B21
 - S/N: SGO0020K9B
 - S/N: SGO0020KB6
- **FUENTE DE PODER**
 - HPE 500 W Flex slot platinum hot plug power supply kit - 720478-B21
 - S/N: 1721009659
- **SWITCH**
 - HPE SWITCH 1920S 48G 4SFP CAPA 3 - JL382A
 - S/N: TW97K3M0QX
 - S/N: CN7AK3MDJW
- **PDU**
 - TRIPP-LITE 1220T - Rack PDU, Basic, 1U, 20A, 120V, (10 tomas) 5-20; 5-20P Horizontal
 - REF: 2830B50IB6434
 - REF: 2849B50IB6434
 - REF: 2849B50IB6434
- **GABINETE**
 - RACK CONNECTION - 42UR 800x800x2055
 - SIN NUMERO DE SERIE

ANTECEDENTES

Francisco Moscoso 6-25 y Av. 10 de Agosto - Telf.: 0969611845 / 0997581734 - email: andres.ambrosi@ambross-ti.com - Cuenca-Ecuador