

UNIVERSIDAD NACIONAL TECNOLÓGICA DE LIMA SUR

FACULTAD DE INGENIERÍA Y GESTIÓN

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



**“DISEÑO DE UN DATAMART PARA LA MEJORA DE LA GESTIÓN
DE LOS ACTIVOS DE INFORMACIÓN EN UNA ENTIDAD
FINANCIERA”**

TRABAJO DE SUFICIENCIA PROFESIONAL

Para optar el Título Profesional de

INGENIERO DE SISTEMAS

PRESENTADO POR EL BACHILLER

MEJÍA VÁSQUEZ, YSBÉL MIRIAN

Villa el Salvador

2019

DEDICATORIA

A mis padres, por brindarme su apoyo incondicional y permitirme llegar hasta este punto de mi carrera.

A mis hermanos Saray e Isaac, por ser mi soporte y cómplices de largas noches de estudio.

A Chicky y Tara, por sus desvariados gestos de aliento y alegría que me impulsaron a continuar hasta el final.

AGRADECIMIENTOS

En primer lugar a Dios, por ser la luz que me guía en el camino de la vida.

Asimismo, a mis profesores por brindarme los conocimientos necesarios que me permitieron culminar satisfactoriamente la carrera y desarrollarme profesionalmente, quienes mediante su orientación y sus altas exigencias me permitieron culminar el presente trabajo.

A todas y cada una de las personas que formaron parte de mi vida universitaria.

ÍNDICE

INTRODUCCIÓN	8
CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA	8
1.1. Descripción de la Realidad Problemática	9
1.2. Justificación del Problema	9
1.3. Delimitación del Proyecto	10
1.3.1. Teórico.....	10
1.3.2. Temporal	10
1.3.3. Espacial.....	11
1.4. Formulación del Problema.....	11
1.4.1. Problema General	11
1.4.2. Problemas Específicos	11
1.5. Objetivos	11
1.5.1. Objetivo General.....	11
1.5.2. Objetivos Específicos	11
CAPÍTULO II: MARCO TEÓRICO.....	12
2.1. Antecedentes	12
2.1.1. Internacionales	12
2.1.2. Nacionales.....	14
2.2. Bases Teóricas.....	17
2.2.1. Toma de Decisiones.....	17
2.2.2. Business Intelligence.....	18
2.2.3. Data Warehouse.....	22
2.2.4. Datamart.....	22
2.2.5. Modelo Dimensional	23
2.2.6. Procesos ETL	27
2.2.7. Calidad de Datos	29

2.2.8. Metodología de Ralph Kimball.....	30
2.2.9. Herramientas de Extracción y Explotación de Datos.....	36
2.3. Definición de Términos Básicos	37
CAPÍTULO III: DESARROLLO DEL TRABAJO DE SUFICIENCIA PROFESIONAL	
.....	41
3.1. Modelo de Solución Propuesto.....	41
3.1.1. Planificación	41
3.1.2. Identificación de Requerimientos.....	41
3.1.3. Definición de las herramientas de software a utilizar.....	43
3.1.4. Análisis Dimensional.....	44
3.1.5. Arquitectura de la solución.	47
3.1.6. Proceso de Extracción, Transformación y Carga.	48
3.1.7. Proceso de Explotación.....	61
3.2. Resultados	61
CONCLUSIONES.....	64
RECOMENDACIONES	65
BIBLIOGRAFÍA	66
ANEXOS	69
Anexo A: Guía de entrevista	69
Anexo B: Cronograma del Proyecto.....	70
Anexo C: Scrip para la creación del Datamart	71
Anexo D: Proceso ETL realizado en SSIS	72

LISTADO DE TABLAS

Tabla 1. <i>Requerimientos funcionales</i>	42
Tabla 2. <i>Requerimientos no funcionales</i>	43
Tabla 3. <i>Granularidad</i>	46
Tabla 4. <i>Medidas</i>	47
Tabla 5. <i>Descripción de las tablas fuente de dimEmpresa</i>	49
Tabla 6. <i>Estandarización y limpieza de datos para dimEmpresa</i>	49
Tabla 7. <i>Fuentes de datos para dimEmpresa</i>	50
Tabla 8. <i>Tabla destino dimEmpresa</i>	50
Tabla 9. <i>Descripción de las tablas fuente de dimTiempo</i>	51
Tabla 10. <i>Estandarización y limpieza de datos para dimTiempo</i>	52
Tabla 11. <i>Fuentes de datos para dimTiempo</i>	52
Tabla 12. <i>Tabla destino dimTiempo</i>	52
Tabla 13. <i>Descripción de las tablas fuente de dimFuente</i>	53
Tabla 14. <i>Estandarización y limpieza de datos para dimFuente</i>	53
Tabla 15. <i>Fuentes de datos para dimFuente</i>	54
Tabla 16. <i>Tabla destino dimFuente</i>	54
Tabla 17. <i>Descripción de las tablas fuente de dimUbicacion</i>	55
Tabla 18. <i>Estandarización y limpieza de datos para dimUbicacion</i>	55
Tabla 19. <i>Fuente de datos1 para dimUbicacion</i>	56
Tabla 20. <i>Fuentes de datos2 para dimUbicacion</i>	56
Tabla 21. <i>Tabla destino dimUbicacion</i>	57
Tabla 22. <i>Descripción de las tablas fuente de factTransacciones</i>	58
Tabla 23. <i>Estandarización y limpieza de datos para factTransacciones</i>	58
Tabla 24. <i>Fuentes de datos1 para factTransacciones</i>	59
Tabla 25. <i>Fuente de datos2 para factTransacciones</i>	59
Tabla 26. <i>Fuentes de datos3 para factTransacciones</i>	60
Tabla 27. <i>Fuentes de datos4 para factTransacciones</i>	60
Tabla 28. <i>Tabla destino factTransacciones</i>	60

LISTADO DE FIGURAS

Figura 1. Niveles jerárquicos de una organización.	18
Figura 2. Ciclo de la Inteligencia de Negocios.	20
Figura 3. Principales problemas de diseño de un Data Warehouse.	25
Figura 4. Modelo Estrella.	26
Figura 5. Modelo de Copo de Nieve.	27
Figura 6. Proceso ETL.	28
Figura 7. Elementos básicos del Data Warehouse según Kimball.	32
Figura 8. Ciclo de vida de Kimball.	33
Figura 9. Diagrama de flujo del proceso dimensional de Kimball.	35
Figura 10. Cronograma del Proyecto.	41
Figura 11. Modelo Estrella para el Datamart.	44
Figura 12. Arquitectura de la solución.	48
Figura 13. Dashboard general de las transacciones.	62

INTRODUCCIÓN

Desde tiempos memorables, las empresas han generado diversos datos provenientes de sus diferentes procesos de negocio. Sin embargo, es desde las últimas décadas que este crecimiento se ha incrementado de manera sustancial dentro de las organizaciones, produciendo así grandes volúmenes de datos en cortos periodos de tiempo.

Los datos, dentro del actual mundo empresarial, constituyen uno de los primordiales activos de información debido al alto valor que estos representan dentro de las estrategias competitivas para la toma de decisiones. Sin embargo, los datos por sí solos, sin ningún tipo de procesamiento o transformación, son irrelevantes para las Instituciones. El punto clave para las mismas es transformar sus datos y obtener como resultado conocimiento, y no una excesiva carga de información que solo entorpecería el proceso de crear estrategias exitosas.

En los últimos años, obtener aquel éxito se ha convertido en un sueño no tan lejano debido a la innovación y desarrollo de nuevas tecnologías como la Inteligencia de Negocios o Business Intelligence (BI), la cual se ha incorporado para brindar soporte a las diversas necesidades estratégicas de las organizaciones.

Actualmente, a Business Intelligence se le atribuye el valor de sintetizar y suministrar información vital, clave para obtener ventajas competitivas y brindar soporte a la toma de decisiones (Méndez, 2006).

Por lo antes mencionado, el presente proyecto busca diseñar un Datamart para mejorar la gestión de los activos de información de una entidad financiera. Además de, explotar los cubos de información mediante dashboards en Power BI.

La estructura de este trabajo consta de 3 capítulos; en el primero, se plantea la problemática del proyecto al detalle, así como los objetivos; en el segundo, se establecen las bases teóricas relacionadas al proyecto, además de otros proyectos realizados sobre la misma línea de trabajo; y finalmente en el tercero, se detalla la aplicación del proyecto basado en la metodología de Kimball.

CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA

1.1. Descripción de la Realidad Problemática

Tener información de calidad y que esta se vea reflejada en los reportes y tomas de decisiones es indispensable dentro de cualquier empresa y, así lo es dentro de la entidad financiera seleccionada como objeto de estudio del presente trabajo. En el departamento de una de las gerencias más significativas de esta Institución, se vienen reportando problemas como incongruencia entre datos, reportes con mala o escasa información útil y demoras en la salida de reportes.

El principal problema de los eventos que vienen ocurriendo radica en que, la fuente de los datos utilizados para la elaboración de los diferentes reportes, que la gerencia requiere, se encuentra en diferentes repositorios y cada uno con diferente tecnología; por consiguiente, se torna difícil el proceso de extracción e integración de datos para la realización de los reportes; lo cual además, se viene realizando manualmente.

Asimismo, existen incongruencias de estructuras dentro de algunas fuentes de datos y por ese motivo un mal manejo de la data y errores en las revisiones; lo que hace más engorrosa la tarea de generar reportes con información real y fidedigna, útil para el correcto análisis que deben realizar los analistas.

1.2. Justificación del Problema

Según Caralt y Díaz (2010), la aplicación de Business Intelligence en una Organización no solo brinda una visión única e integrada de la misma, sino también genera información actualizada y de buena calidad; lo cual se ve reflejado en la mejora competitiva de las empresas.

El presente trabajo muestra el desarrollo de un Datamart como una solución ideal para mejorar la gestión de los activos de información y apoyar

a la toma de decisiones en lo referente a las Encuestas y la información relacionada, necesaria para la publicación de los Reportes Estadísticos.

El desarrollo de un Datamart permitirá integrar de manera óptima la data relevante de los diferentes repositorios de información, además de brindar un amplio y adecuado análisis de la información, que permitirá la obtención de reportes con data veraz y fidedigna. El uso de un Datamart también permitirá la utilización de dimensiones, a través de las cuales se podrá visualizar de manera más fácil y rápida, información a detalle sobre las utilidades o transacciones realizadas por una empresa en un determinado espacio de tiempo, por ejemplo, o las variaciones de utilidades de una empresa entre un trimestre y otro.

Los reportes o dashboards, productos de la explotación de los cubos de información, no solo brindarán información detallada de las transacciones que realizan las empresas, sino también permitirán realizar un análisis completo y actualizado de la información contenida en los cubos. Todo ello consentirá a los analistas realizar una adecuada toma de decisiones, evitando así los retrasos y demoras que vienen ocurriendo al realizar la verificación de la data, además de los errores en las revisiones.

1.3. Delimitación del Proyecto

1.3.1. Teórico

El presente proyecto abarca los conceptos teóricos referentes a Data Warehouse, Datamart, Business Intelligence, Análisis Dimensional, toma de decisiones, Metodología de Kimball, activos de información del banco, entre otros.

1.3.2. Temporal

El presente proyecto abarca una duración de 2 meses; empezando en octubre del 2019, y culminando en noviembre del 2019.

1.3.3. Espacial

Se realizará en la Sede Central de la entidad financiera, en la ciudad de Lima, Perú.

1.4. Formulación del Problema

1.4.1. Problema General

¿Será posible diseñar un Datamart para la mejora de la gestión de los activos de información de la entidad financiera?

1.4.2. Problemas Específicos

- ¿De qué manera se podrá identificar la data relevante en los diferentes repositorios de información?
- ¿Cuáles serán los criterios a utilizar para la configuración de la data relevante en cubos de información?
- ¿De qué manera explotar los cubos de información a través de la herramienta Power BI?

1.5. Objetivos

1.5.1. Objetivo General

Diseñar un Datamart para la mejora de la gestión de los activos de información de una entidad financiera.

1.5.2. Objetivos Específicos

- Identificar la data relevante en los diferentes repositorios de información.
- Configurar la data relevante encontrada en cubos de información.
- Explotar los cubos de información mediante la herramienta Power BI.

CAPÍTULO II: MARCO TEÓRICO

2.1. Antecedentes

2.1.1. Internacionales

Trabajo, titulado **Business Intelligence: ejemplo de una implantación**, presentado por Carlos Ruiz Pérez, España.

El proyecto de Ruiz (2018), expone las diferencias y similitudes existentes entre Business Analytics y Data Science, además de algunas herramientas SAP que son posibles utilizarse para aplicar esta técnica. Así mismo, se ponen en práctica conocimientos para la implantación de una solución de BI, el cual consiste en un nuevo cuadro de mando que se quiere incorporar en el cuadro de mando integral del cliente.

Al finalizar el autor concluye que, mediante la utilización de otras herramientas software de BI también es posible alcanzar los mismos resultados o unos parecidos. Sin embargo se utilizó SAP puesto que la empresa donde se aplicó utiliza este software.

Trabajo, titulado **La utilización de herramientas Business Intelligence para la gestión empresarial. Un estudio aplicado a la empresa Desarrollo y Servicios Infovec S.L.**, presentado por Patricia Jordá, España.

El objetivo de este trabajo es profundizar en cuanto al conocimiento de BI se refiere y además, implementar una herramienta de BI que permita el desarrollo de un cuadro de indicadores para mejorar la toma de decisiones en la empresa Desarrollo y Servicios Infovec, S.L. Para definir la herramienta de BI, se utilizó como base un estudio publicado cada año por la consultoría Gartner, escogiendo

finalmente a Tableau. Los datos a utilizar fueron extraídos del ERP Sage Eurowin que utiliza la empresa, siendo estos extraídos en formato Excel para que puedan ser posteriormente modificados conforme a los cambios que se requieran (Jordá, 2016).

La autora concluye con un efectivo análisis de las herramientas de BI, además que sugiere conocer a detalle cada aspecto de la empresa relacionada a la problemática, ya que sin información exacta no se puede realizar ningún estudio. Mediante la realización de este proyecto la empresa logró tener un mejor control del negocio, así como también una imagen global de su situación al finalizar el año.

Tesis, titulada **Análisis, diseño e implementación de un Datamart utilizando herramientas Open Source para la unidad administrativa y financiera de la ESPE**, presentada por Cristian Dario Álvarez Calvopiña y Diego Armando Esparza Montes, Ecuador.

Los autores, Esparza y Alvarez (2014), proponen la implementación de datamarts para la mejora de procesos y consolidación de la información que administra la Dirección Financiera y Talento Humano de la ESPE; para ello utilizan la metodología de Kimball y la herramienta open source llamada Pentaho.

Como parte de sus resultados se logró reducir el tiempo de creación de cada reporte, teniendo así un mejor análisis dinámico e intuitivo de los datos para la toma de decisiones; además de disminuir el tráfico de consultas a las bases de datos.

Caso de éxito, **Coca-Cola Bottling Company maximizó su eficiencia operativa**, presentado por Tableau.

El problema era el acceso restringido que se tenía a los datos de ventas y operaciones en tiempo real debido a la generación manual de los reportes. Mediante su plataforma de BI, el equipo de Coca-Cola automatizó el proceso de generación de reportes, permitiéndole así al equipo de ventas tener información procesable y oportuna. De esta manera los analistas se pueden enfocar en mejorar sus estrategias competitivas y nuevas formas de innovar, en vez de realizar manualmente un reporte (Tableau, s.f.).

2.1.2. Nacionales

Tesis, titulada **Análisis, diseño e implementación de un datamart para el soporte de toma de decisiones y evaluación de las estrategias sanitarias en las Direcciones de Salud**, presentado por Carmen Pamela Rosales Sedano, Lima.

El objetivo de esta tesis es apoyar a la toma de decisiones de cada dirección de salud del país, a través de la implementación de un datamart, en lo referente a los objetivos específicos de cada estrategia sanitaria. Para ello, muestra la implementación del datamart siguiendo la metodología de Kimball y el uso de software libre para la extracción y explotación de los datos (Rosales, 2009).

Como parte de sus conclusiones la autora menciona que, la implementación del datamart no solo permitió ahorrar tiempo en la generación de reportes, sino también que estos sean más dinámicos para el usuario. Asimismo, el uso de herramientas libres disminuyó los costos, para posteriormente realizar mejores inversiones dentro de las direcciones de salud.

Tesis, titulada **Modelo de Business Intelligence y Analytics soportado por la tecnología Cloud Computing para Pymes del sector retail**, presentado por Milton Elviz López Inga y Ricardo Martín Guerrero Huaranga, Lima.

Para Lopez y Guerrero (2017), el principal objetivo del proyecto es implementar un modelo innovador y tecnológico que combine tecnologías como Business Intelligence, Analytics y Cloud Computing, para que las PYMES retail peruanas puedan tomar decisiones de manera oportuna e informada. Para tales fines, se diseñó un modelo tecnológico que luego fue validado a través de su implementación; y finalmente se propuso un plan de continuidad escalable del modelo.

Las conclusiones mostradas por el autor, muestran que el modelo tuvo efectividad con relación al tiempo e integración. Además, el modelo brinda información oportuna a las personas encargadas de la toma de decisiones sin necesidad que estas estén dentro de la empresa. Se comprobó que los servicios Cloud disminuyen los costos de mantenimiento y administración dentro de las PYMES.

Tesis, titulada **Desarrollo de un modelo de calidad de datos aplicado a una solución de Inteligencia de Negocios en una Institución Educativa: Caso Lambda**, presentado por Marshall André Fernández Sáenz, Lima.

En el proyecto de tesis el autor, Fernández (2018), propone el desarrollo de un Modelo de Calidad de Datos para una adecuada evaluación de los datos almacenados dentro de su Data Warehouse, así como también las métricas relacionadas. El modelo de calidad se desarrolló a través de la técnica de grupo nominal y el método Brosseau, teniendo como base las características encontradas en la norma ISO/IEC 25012.

Como parte de sus conclusiones, el autor menciona que se logró el desarrollo del modelo de calidad identificando los repositorios con data relevante y sus métricas. La evaluación de calidad de datos se desarrolló acorde a la norma ISO/IEC 25040, y se analizaron los resultados obtenidos para luego obtener reportes de ello.

Tesis, titulada **Análisis de métodos y técnicas de limpieza de datos existentes y aplicación en un sistema CRM para una Institución Educativa limeña**, presentado por Ángel Gabriel Sandoval Linares, Lima.

El autor enfoca este proyecto a la realización de una nueva técnica de limpieza de datos, para una posterior implementación de procesos ETL, lo cual evita tener inconsistencias en cada fuente de datos procesada del sistema CRM de una Institución Educativa limeña. Para su realización de la nueva técnica de limpieza de datos se realizó un análisis detallado de los diferentes algoritmos, técnicas y métodos; para luego realizar una combinación de ellos, que permitan optimizar y mejorar la calidad de los datos de la Institución Educativa; y de esta manera ayudar a que la información se mantenga limpia en los sistemas de la empresa (Sandoval, 2018).

Al finalizar el proyecto, el autor concluye que, se logró integrar y normalizar las entidades de las diferentes fuentes y sistemas de información, y además se propuso un nuevo modelo de base datos que pueda integrar toda la información necesaria. Asimismo, se diseñó e implementó un algoritmo para la corrección de los datos y se propusieron algoritmos y procesos ETL para realizar la limpieza de los registros y la posterior integración de los datos en una única base de datos. Para ello se implementó un nuevo módulo de software, el cual integraría la información de la base de datos en un Data Warehouse.

2.2. Bases Teóricas

2.2.1. Toma de Decisiones

Se define como la elección entre las diferentes alternativas mediante la cual se soluciona una situación de la vida (Peña, 2006).

Para Fincowsky (2011) la toma de decisiones se refiere a las elecciones que se realizan frente a una combinación de factores que salen un poco de lo habitual.

La toma de decisiones es importante dentro del concepto de Business Intelligence ya que, está principalmente orientado a obtener mejores resultados de las decisiones que se lleven a cabo y de esta manera obtener ventajas competitivas para la empresa, dejando nunca de lado los objetivos estratégicos de la misma.

Existen varios criterios para tomar una decisión eficaz, según Méndez (2006), son seis y son las siguientes:

- Concentrarse en lo que es realmente importante para la empresa.
- Realizar el proceso de forma lógica y coherente.
- Considerar tanto los elementos objetivos como los subjetivos y utilizar una estructura de pensamiento analítica e intuitiva.
- Recopilar la información necesaria para optar o elegir.
- Recopilar las informaciones, opiniones, que se han formado en torno a la elección.
- Ser directos y flexibles antes, durante y después del proceso.

2.2.1.1. Estructura jerárquica de la toma de decisiones

Para Peña (2006), existen tres niveles jerárquicos para clasificar los roles existentes que ocupan el personal dentro de una organización, dentro de cada uno existen diferentes responsabilidades que cada persona debe de cumplir.

Tradicionalmente estos niveles jerárquicos se muestran en una pirámide dividida en tres partes, esto se considera para destacar los niveles de autoridad dentro de una organización, el impacto de las decisiones y las competencias de las funciones que se realiza en cada nivel. Estos tres niveles son el estratégico, táctico y operativo; los cuales se muestran en la Figura 1.

Figura 1.
Niveles jerárquicos de una organización.



Fuente: (Peña, 2006).

2.2.2. Business Intelligence

Una de las ventajas competitivas en las organizaciones es contar con información oportuna y relevante dentro de una plataforma de Inteligencia de Negocio, que les permita tomar las mejores decisiones.

La información procedente de las operaciones de una organización en una plataforma de Business Intelligence interrelacionada con su cadena de valor se está convirtiendo en un factor crítico de éxito, siendo esencial para la competitividad y aún más, para la supervivencia de la empresa (Medina, 2013).

Según Méndez (2006), BI es mucho más que una simple tecnología, es un nuevo escenario en el cual las empresas buscan la excelencia mediante la definición de sus objetivos, el análisis de sus procedimientos, la gestión de su información, la obtención de conocimientos y la toma de decisiones, seguida de una constante evaluación.

La Inteligencia de Negocios también se puede definir como el conjunto de metodologías, aplicaciones, prácticas y capacidades orientadas a permitir una mejor toma de decisiones a los usuarios de una organización a través de una adecuada creación y administración de la información (Caralt & Díaz, 2010).

Existen diversos conceptos para BI, desde décadas anteriores diferentes autores han definido esta tecnología desde varios aspectos, pero todos coinciden en que ayuda a mejorar la toma de decisiones en una organización.

2.2.2.1. Ciclo de la Inteligencia de Negocios

Para Vitt, Luckevich, y Mismar (2003), el ciclo de BI se compone de cuatro etapas o fases:

- **Análisis.** Se determina los datos a recopilar.
- **Reflexión.** Se basa principalmente en un entendimiento básico de cómo opera la organización.
- **Acción.** Implica relacionar la información obtenida de las etapas anteriores, análisis y reflexión.
- **Medición.** Comprende la valoración y comparación de la información, apoyados del benchmarking y siempre fijando la vista en los objetivos y estrategias a corto y largo plazo de la organización.

Figura 2.
Ciclo de la Inteligencia de Negocios.



Fuente: (Vitt, Luckevich, & Mismar, 2003)

2.2.2.2. Factores que impulsan el BI

La globalización ha sido sin duda una de las principales fuentes impulsadoras del BI. Las empresas generan cada minuto, e incluso cada segundo, una indeterminada cantidad de datos que van creciendo exponencialmente.

Debido a ello, las herramientas de BI necesitan brindar soporte fiable ante estos grandes volúmenes de datos, siendo capaces de administrar, procesar y gestionar la información proveniente de las diversas fuentes de datos que una organización pueda tener.

Para Méndez (2006), existen tres indiscutibles factores de mercado que están obligando, hoy más que nunca, a las organizaciones a adoptar soluciones o estrategias de BI:

- **Incrementar los ingresos y disminuir los costos, mejorando la competitividad.** El BI puede lograr ello a través de una visión integral y oportuna de los datos

convertidos en información, útil para una efectiva toma de decisiones.

- **Gestionar la complejidad.** Esto ocurre ya que, gracias al BI las empresas pueden detectar patrones de comportamiento y tendencias que suelen ser difíciles de descubrir. Todo esto mediante el análisis y minería de la información.
- **Explotar las inversiones existentes.** La Inteligencia de Negocios añade valor a las diversas inversiones de TI de una compañía, mediante la integración y el análisis de los datos capturados por los diferentes sistemas y aplicaciones de la organización.

2.2.2.3. Beneficios del BI

No hay duda que la adecuada aplicación de BI brinda maravillosos y extraordinarios beneficios para la organización, mejorando incluso la reducción de costos y gastos innecesarios.

Para Caralt y Díaz (2010), entre los beneficios que más destacan son los siguientes:

- Generar un círculo virtuoso de la información, en el cual los datos se transformen en información y, de estos provenga conocimiento útil que permita mejorar las decisiones tomadas, las cuáles se reflejan en la obtención de mejores resultados que luego generarán nuevos y mejores datos.
- Brindar una visión única, histórica, integrada, persistente y de buena calidad de la información.
- Generar, administrar y mantener métricas, indicadores claves de rendimiento (KPI) e indicadores claves de metas (KGI) indispensables para toda organización.

- Brindar información actualizada a nivel agregado, así como también en detalle.
- Mejorar la comunicación y disminuir la brecha de distancia entre el área de TI y la organización.
- Mejorar la comprensión y documentación de los diferentes sistemas de información de la compañía.
- Mejorar la competitividad de la organización.

2.2.3. Data Warehouse

Según Inmon (2005), es definido como una colección de datos orientados a temas integrados, variantes en el tiempo y no volátiles que apoya al proceso de toma de decisiones.

El Data Warehouse es un repositorio de datos basado en estructuras relacionales o multidimensionales, en las cuales se almacena la información, calculando antes todas las combinaciones existentes en todos los niveles de todas las aperturas de análisis. Prácticamente es como un producto cartesiano en el cual se almacenan todas las combinaciones posibles (Kimball & Ross, 2002).

Un Data Warehouse se puede definir entonces como una base de datos empresarial o corporativa, la cual se caracteriza por integrar data de diferentes fuentes de base de datos para su posterior análisis desde diversas perspectivas y obteniendo respuestas a grandes velocidades.

2.2.4. Datamart

Se define como un conjunto de datos estructurados para que facilite su análisis, los cuales provienen de las diferentes aplicaciones operacionales de la organización. Un Datamart es un subconjunto de un Data Warehouse y está orientado en la información de una determinada área de la empresa, por lo que consta de un alcance de contenido limitado (Kimball & Ross, 2002).

Se puede decir que un Datamart es una base de datos departamental, especializada tan sólo en un área en específico del negocio y no en todo como lo hace un Data Warehouse. Un Datamart se caracteriza por tener una óptima estructura de datos que permite un rápido y eficaz análisis de la data recolectada de las diferentes fuentes de información.

2.2.4.1. Clasificación del Datamart

Según Inmon W. H. (2005), se observa la siguiente clasificación:

Datamart dependiente

Los Datamart dependientes son aquellos que tienen una única fuente de datos, siendo esta un Data Warehouse.

Datamart independiente

Los Datamart independientes son aquellos que sus fuentes de datos las toman de los sistemas transaccionales de la organización y no dependen de un Data Warehouse.

Datamart híbrido

Los Datamart híbridos son aquellos que combinan sus fuentes de datos de un Data Warehouse con las de los sistemas transaccionales y/u operacionales de la organización.

2.2.5. Modelo Dimensional

El modelo dimensional es la manera en que modelamos los datos para efectuar el análisis. Brinda a los usuarios finales una forma más rápida de obtener información, logrando un mayor desempeño. Este modelo produce una base de datos la cual es simple de navegar e interactuar, conteniendo una menor cantidad de tablas y relaciones.

Aquí las perspectivas o dimensiones se combinan dentro de una entidad llamada hechos. Las medidas se pueden analizar desde cada una de las perspectivas que se hayan definido (Kimball & Ross, 2002).

2.2.5.1. Tabla de hechos (Fact Table)

Para Kimball y Ross (2002), la tabla de hechos es la principal tabla del modelo dimensional y contiene los valores del negocio que se desea analizar. Cada tabla de hechos consta de sus llaves foráneas, las cuales se relacionan con sus tablas de dimensiones y las columnas con las que serán analizadas. Está constituida por medidas y llaves primarias o en algunos casos por llaves foráneas.

2.2.5.2. Dimensiones

Son objetos calificadores del negocio que dan sentido a la tabla de hechos, los cuales se organizan en función de algún aspecto de interés del usuario para su posterior análisis sobre la tendencia y comportamiento del mismo.

Las variadas definiciones de las dimensiones se basan en las políticas de la empresa, describiendo un aspecto de la empresa, y en la manera en que esta interpreta su información para segmentar el análisis y proporcionando un acceso intuitivo a los datos (Kimball & Ross, 2002).

2.2.5.3. Medidas o métricas

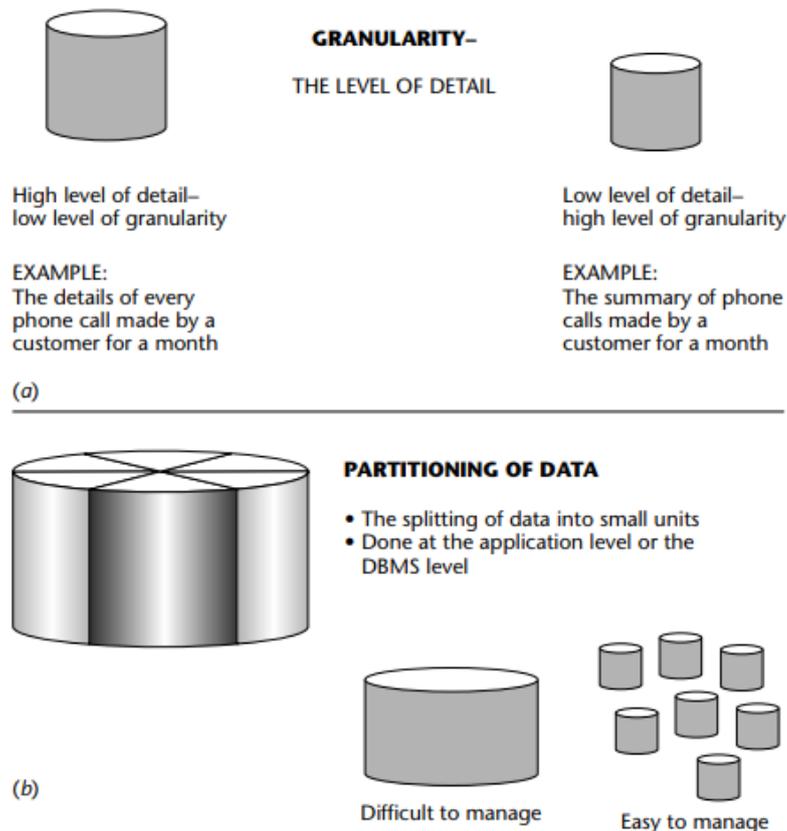
Son características cualitativas o cuantitativas de lo que se desea analizar de la organización. Una medida es una columna cuantitativa dentro de la tabla de hechos. Estas medidas representan los objetos que son analizados y están dadas por valores o cifras porcentuales como por ejemplo, el importe de ventas en soles o dólares y la cantidad de unidades en stock (Kimball & Ross, 2002).

2.2.5.4. Granularidad (Granularity)

Para Inmon (2005), el problema de la granularidad es el aspecto más relevante del diseño de un Data Warehouse. La granularidad hace referencia al nivel de detalle de los datos dentro de un Data Warehouse; mientras más detalles haya, menor será el nivel de granularidad y cuantos menos detalles haya, mayor será el nivel de granularidad.

Hace referencia a los hechos y está definido por el mínimo nivel de detalle de las dimensiones (Kimball & Ross, 2002).

Figura 3.
Principales problemas de diseño de un Data Warehouse.

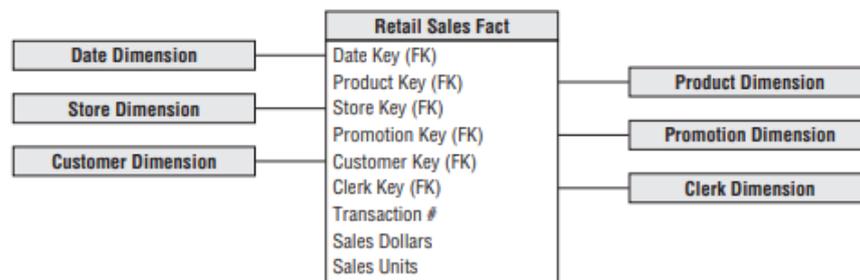


Fuente: (Inmon, 2005)

2.2.5.5. Esquema estrella

Es una estructura que consiste en una tabla central denominada Tabla de Hechos y un conjunto de tablas de dimensiones conectadas a esta y organizadas alrededor de la misma. Cada tabla de dimensión se vincula con la Tabla de Hechos a través de un identificador.

Figura 4.
Modelo Estrella.



Fuente: (Kimball & Ross, 2002)

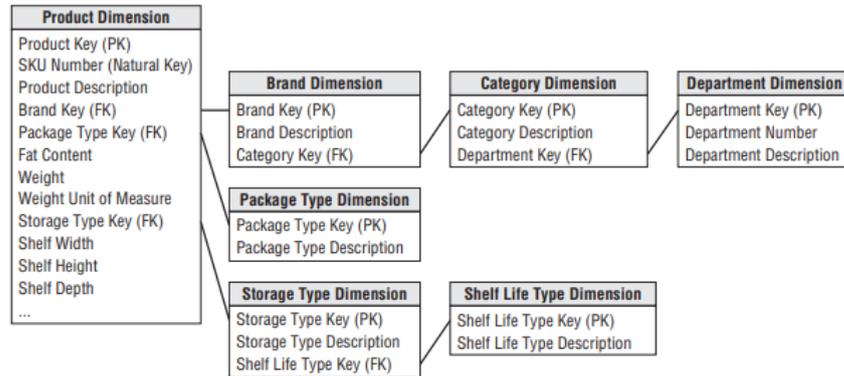
2.2.5.6. Esquema copo de nieve

Es una variante del esquema estrella en donde cada una de las puntas de la estrellas pueden dividirse en más puntas, es decir, las tablas de dimensión se encuentran normalizadas y pueden incluir claves que apunten a otras tablas de dimensión.

Las ventajas de este tipo de esquema es la reducción del tamaño y redundancia en las tablas de dimensión. Sin embargo, dentro de las desventajas esta que el incremento de la cantidad de tablas lleva a que se requiera de una mayor cantidad de operaciones de unión al realizar las consultas, lo que empeora el rendimiento; además del mantenimiento que

requiere cada una de las tablas adicionales (Kimball & Ross, 2002).

Figura 5.
Modelo de Copo de Nieve.



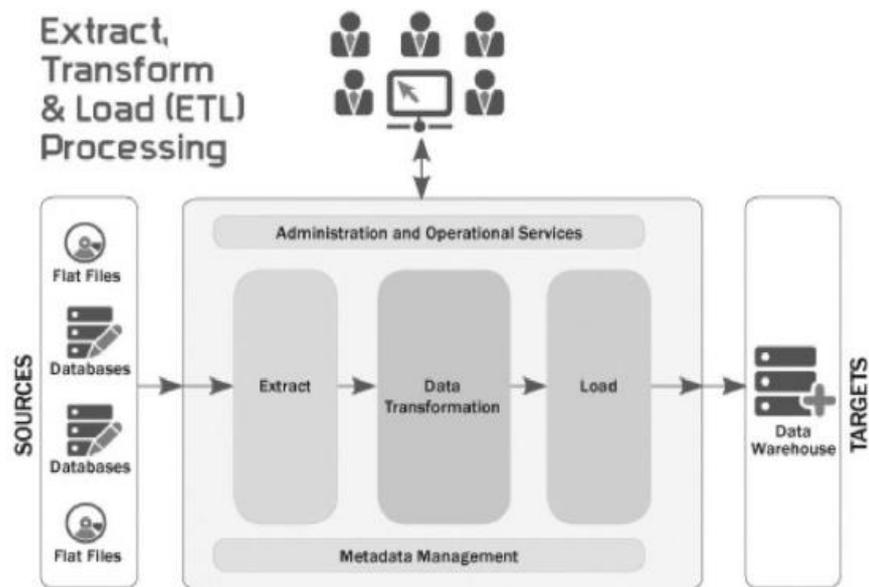
Fuente: (Kimball & Ross, 2002)

2.2.6. Procesos ETL

Los procesos ETL son conocidos por permitir el procesamiento de los datos almacenados en las fuentes de información, teniendo como entrada la data contenida en una base de datos transaccional y almacenándola luego en un Data Warehouse.

Gonzáles, Ramón, y Guitart (2016), hacen referencia a los procesos ETL de manera muy singular, refiriéndose a la ETL como la máquina o cocina del sistema de BI puesto que, aunque no se ve, nada funciona si antes no se ha “cocinado bien”. Además, mencionan que para limitar los riesgos durante esta etapa es indispensable la participación de los usuarios clave, propietario de la data y el patrocinador para realizar una adecuada toma de decisiones en caso se den discrepancias que puedan afectar los objetivos estratégicos del proyecto.

Figura 6.
Proceso ETL.



Fuente: (Caralt & Díaz, 2010)

Para Kimball y Ross (2002), el nombre con que cuentan estos procesos es lo que define lo que harán:

2.2.6.1. Extracción (Extract)

Se empieza con la extracción de los datos de los sistemas fuente, almacenándolos luego de tal manera que, estén listos para realizar una transformación.

2.2.6.2. Transformación (Transform)

Después del proceso de extracción, se realiza la transformación de los datos, y esta se hace en dos partes: la primera que es la limpieza, se realiza para eliminar inconsistencias que hayan habido en los datos fuente, o eliminar los casos de datos duplicados; esto se soluciona normalizando la información, de manera que la data obtenida tenga calidad y sea consistente; y la segunda que es la

transformación en sí, en la que se hacen los cambios a los datos que sean requeridos por las reglas de negocio y lo solicitado por los usuarios, así como la unión de los datos de múltiples fuentes.

2.2.6.3. Carga (Load)

Los datos son cargados en el Data Warehouse para su posterior presentación a los usuarios a través de la explotación de la información, lo cual se realiza mediante diversas herramientas que permitan obtener reportes o Dashboards.

2.2.7. Calidad de Datos

Para hablar de calidad de datos, primero se tiene que conocer el concepto de calidad como tal. Calidad, según la International Organization for Standardization (2015), es el grado en que un conjunto de características propias de un objeto cumple con los requisitos establecidos.

La Calidad de Datos también conocida como Data Quality está directamente relacionada con sus dimensiones, dentro de las cuales se puede mencionar la exactitud, completitud, consistencia, actualidad de los datos, entre otros (Bianchi & Valverde, 2009). Por lo general, se puede decir que los datos son correctos si es que están aptos para ser usado en las operaciones, toma de decisiones y planificación del negocio (Redman, 2002).

La calidad de datos es un problema común encontrado en diversas bases de datos, el cual se va acumulando con el pasar de los años y puede llegar a convertirse en un castillo de naipes que puede derrumbarse en cualquier momento (Méndez, 2006). Esto ocurre debido a una mala toma de decisiones y a la construcción mediante un defectuoso análisis y sus correspondientes predicciones.

Según Kimball y Caserta, (2014), para garantizar la precisión de los datos estos tienen que ser:

- Correctos (sus valores tienen que representar fielmente a los objetos que representan).
- No ambiguos (los valores solo pueden tener un significado).
- Consistentes (los valores deben guiarse por un único estándar general).
- Completos (los valores deben estar completos y no deben perderse durante su uso).

2.2.8. Metodología de Ralph Kimball

Para Kimball, un Data Warehouse no es más que la unión de todos los Datamarts de una organización, por ello defiende una metodología Bottom – Up, la cual consta en empezar la construcción de un Data Warehouse a partir de pequeñas partes que vendrían a ser los Datamarts, y seguir evolucionando hacia estructuras más desarrolladas. Esto parte de la premisa de que, un Data Warehouse se construye a partir de un Datamart que está orientado a un proceso de negocio y normalmente está referido a un tema en específico. Luego se desarrollan otros Datamarts haciendo que las dimensiones sean comunes y sean las fuentes de integración (Kimball & Ross, 2002).

Además, mencionan que la metodología Bottom - Up producirá resultados mucho más rápidos en comparación a un enfoque Top – Down, el cual tomará más tiempo dependiendo del tamaño del proceso de negocio específico que se haya modelado en el Datamart.

El modelo de Kimball deja en un segundo plano la necesidad de tener un Data Warehouse debido a que la mayoría de usuarios desea obtener datos detallados, por tal motivo Kimball argumenta que es mejor almacenar los datos en Datamarts independientes y lógicamente conectados usando dimensiones.

La principal ventaja que se puede encontrar en este enfoque es que, al estar conformado por diferentes Datamarts estructurados dimensionalmente; el Data Warehouse puede ser explotado por completo directamente por las herramientas de reporting y análisis de datos sin tener que acudir a estructuras intermedias.

Para Kimball y Ross (2002), el Data Warehouse debe tener cuatro importantes componentes. Estos se pueden visualizar en la Figura 7 y son los siguientes:

a) *Sistemas operacionales fuente (Operational Source Systems)*

Son los que capturan cada transacción del negocio, las consultas se caracterizan porque actualizan o leen un registro a la vez. Sus prioridades son el procesamiento y la disponibilidad.

b) *Área de limpieza de datos (Data Staging Area)*

Es el área de almacenamiento y conjunto de procesos ETL (extracción, transformación y carga); permanece entre las fuentes y el Data Warehouse o Datamarts cuyo objetivo es facilitar la extracción, limpieza y calidad de datos mediante los procesos ETL.

c) *Área de presentación de datos (Data Presentation Area)*

En esta área los datos se encuentran almacenados, organizados y disponibles para ser consultados y analizados por los usuarios a través de las herramientas de acceso.

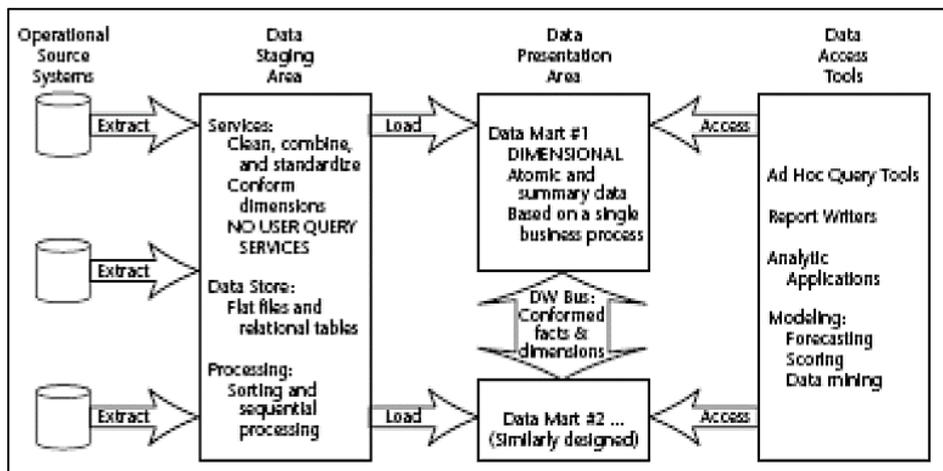
Presenta un conjunto de Datamarts integrados en el que cada uno presenta datos de un solo proceso del negocio. Utiliza el modelado dimensional como técnica para que los usuarios obtengan los datos, ya que lo que se busca es obtener un rápido acceso a la data a través de las consultas y no remover los datos redundantes.

d) *Herramientas de acceso a datos (Data Access Tools)*

La capa de herramientas de acceso es también denominada capa de presentación, la cual puede llegar a ser tan simple como un “ad hoc” query, o al contrario, puede ser un complejo y sofisticado Data Mining. A través del uso de las herramientas de acceso y los resultados que estos puedan mostrar, los usuarios pueden tomar decisiones en beneficio del negocio.

Figura 7.

Elementos básicos del Data Warehouse según Kimball.



Fuente: (Kimball & Ross, 2002)

2.2.8.1. Ciclo de vida

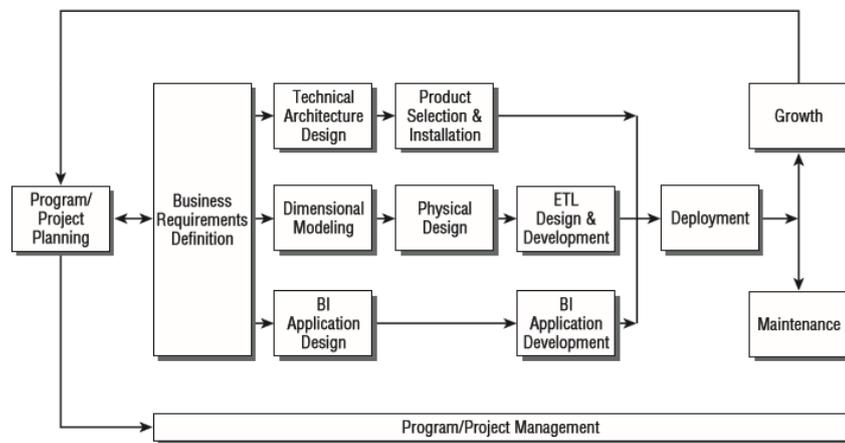
Para Kimball y Ross (2002), la metodología se basa en lo que denominan el Ciclo de Vida Dimensional del Negocio y presentan tres principios claves para el éxito de un Data Warehouse, que son los siguientes:

- Centrarse en las necesidades de la empresa.
- Presentar datos dimensionalmente estructurados a los usuarios.

- Abordar proyectos manejables e iterativos.

En la Figura 8 se puede apreciar el ciclo de vida de Kimball, el cual muestra una secuencia de tareas y dependencias por las que los proyectos de BI deben encaminarse. Este diagrama es una especie de roadmap que permite guiar a los equipos de BI a realizar las cosas en el momento adecuado (Kimball & Ross, 2002).

Figura 8.
Ciclo de vida de Kimball.



Fuente: (Kimball & Ross, 2002)

a) Planificación (Program / Project Planning)

Aquí se determina el propósito del proyecto, sus objetivos y el alcance del mismo, además se definen los principales riesgos a los que está expuesto y las necesidades de información.

b) Definición de requerimientos (Business Requirements Definitions)

La definición de requerimientos tiene como principal objetivo recolectar información de la

situación actual del negocio, así como sus necesidades, todo ello a través de la ejecución de entrevistas al personal involucrado, desde los que se encargan de la toma de decisiones, hasta aquellos que realmente ven y saben qué tipo de problemas ocurren con los datos. Lo que se busca a través de la obtención de información es priorizar los requerimientos y procesos críticos que se encuentren.

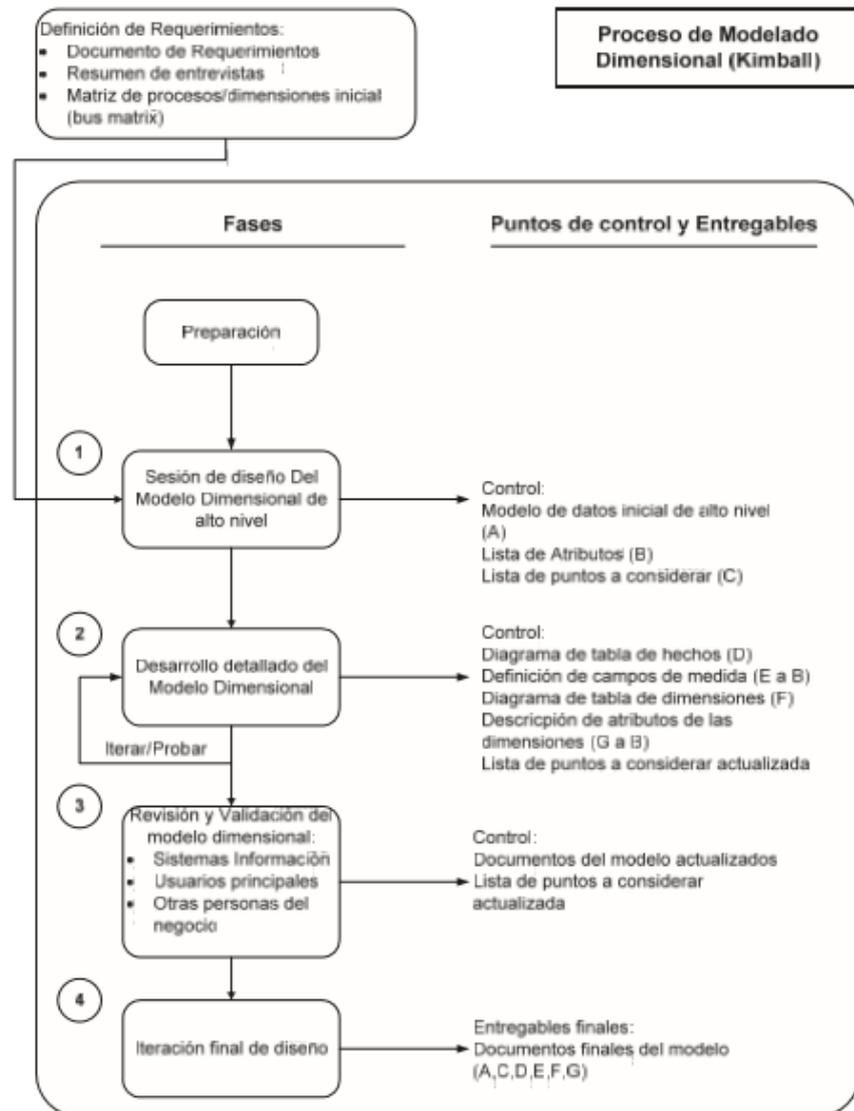
c) Modelado Dimensional (Dimensional Modelling)

Según Mundy y Thornthwaite (2006), el proceso iterativo de modelado consiste en cuatro pasos:

- Elegir el proceso de negocio crítico para la organización.
- Establecer el nivel de granularidad adecuado.
- Escoger las dimensiones.
- Identificar las medidas y tablas de hechos.

Figura 9.

Diagrama de flujo del proceso dimensional de Kimball.



Fuente: (Mundy & Thornthwaite, 2006)

d) Diseño Físico (Physical Design)

Aquí se determina la amplitud del Data Warehouse y los requerimientos que necesite el modelo físico como por ejemplo memoria y servidores, instalación del software en los servidores de desarrollo, prueba y producción; entre otros.

e) Diseño y desarrollo del ETL (ETL Design & Development)

Aquí se determinan los orígenes de datos para el modelo dimensional propuesto, ya que el ETL es la base sobre la cual se alimenta el Data Warehouse. También se pueden definir reglas para definir la calidad y consistencia de los datos.

f) Diseño y desarrollo de aplicaciones BI (BI Application Design / BI Application Development)

Las aplicaciones de BI son la cara de la solución de inteligencia de negocios, por tanto aquí se muestra el diseño y desarrollo de los reportes o dashboards que brindará a los usuarios información útil y real. Existen diferentes tipos de reportes y herramientas para la obtención y análisis de los mismos, desde los más simples, de formato predefinido y/o estático, hasta los más complejos, que pueden incluir algoritmos y modelos de minería de datos; Kimball los llama informes estándar y aplicaciones analíticas respectivamente.

2.2.9. Herramientas de Extracción y Explotación de Datos

2.2.9.1. SQL Server Integration Services - SSIS

SSIS es un componente de Microsoft SQL Server, siendo su principal característica el ser una herramienta flexible para el proceso de ETL. Provee una plataforma escalable y extensible que capacita al equipo desarrollador a construir, mantener, y desplegar soluciones para diseñar, generar y alcanzar soluciones de integración de datos únicas

de acuerdo a las necesidades, además de aplicaciones workflows. También permite la integración de los datos de cualquier fuente.

2.2.9.2. Power BI

Es una herramienta de Microsoft con servicio en la nube y de fácil implementación; que permite compartir y acceder a reportes desde cualquier tipo de dispositivo. La integración de Power BI con Microsoft Office 365 también permite el acceso a datos locales, bases de datos y servicios en la nube.

Las novedades en cuanto a esta herramienta es la conectividad híbrida a fuentes de datos de correo locales; lo que significa que no todos los datos se cargan en la nube de Microsoft Azure.

2.3. Definición de Términos Básicos

- a. Toma de Decisiones:** Se refiere a las elecciones que se realizan frente a una combinación de factores que salen un poco de lo habitual (Fincowsky, 2011).
Se concluye que, el buen manejo decisivo es indispensable para la mejora competitiva dentro de cualquier organización.
- b. Activos de Información:** Para efectos del presente proyecto, los activos de información hacen referencia a las Encuestas que el departamento del banco envía, cada cierto tiempo, a las empresas para obtener información de las transacciones que estas realizan con el exterior. La información recopilada se puede ver reflejada en los “Reportes Estadísticos” y publicaciones que el banco realiza, quienes cuentan con información relevante para la economía.
- c. Business Intelligence (BI):** Según Méndez (2006), se refiere al conjunto de herramientas y aplicaciones orientados a ayudar en la toma de decisiones que permitan acceder interactivamente, analizar y manipular la información de misión crítica de la organización.

De acuerdo con lo anterior, se puede concluir que Business Intelligence es importante para una correcta toma de decisiones dentro de una organización.

- d. **Data Warehouse (DW):** Para Inmon (2005), un Data Warehouse es una colección de datos organizados orientados a temas para soportar las necesidades de las empresas. Estos son integrados, no volátiles y variantes en el tiempo.

Por lo tanto, se puede decir que un Data Warehouse es como una gran base de datos empresarial que integra datos de diversas fuentes de la empresa para su posterior análisis.

- e. **Datamart:** Se refiere al conjunto de datos estructurados que provienen de las diferentes aplicaciones operacionales, de manera que faciliten su posterior análisis. Es una pequeña parte de un Data Warehouse con un alcance de contenido limitado, el cual es utilizado sólo por un área específica o un problema particular de análisis dentro de la organización (Kimball & Ross, 2002).

Se entiende que, un Datamart está orientado al análisis de solo una pequeña parte de una empresa como por ejemplo, un departamento o área de la misma. Se especializa solo en un área en específico del negocio y no en toda la organización.

- f. **Sistemas Transaccionales – OLTP:** Los sistemas transaccionales OLTP (Online Transaction Process), son todos los sistemas de información que soportan las transacciones realizadas día a día en una organización y maximizan la capacidad de procesamiento de transacciones, por ello deben dar respuesta inmediata a cada transacción realizada; además se encargan de recolectar los datos asociados a cada transacción y almacenarlos en bases de datos a través de sistemas de información. Si un sistema OLTP deja de funcionar puede ocasionar grandes pérdidas en una compañía por el hecho de no poder realizar transacciones (Kimball & Ross, 2002).

Son bases de datos operacionales que no han sido diseñadas para realizar el análisis del negocio, sino solo funcionalidades básicas.

- g. **Sistemas analíticos – OLAP:** Para Kimball y Ross (2002), los Sistemas de tipo OLAP (On-Line Analytical Processing) tienen como principal

objetivo obtener indicadores de gestión que puedan servir de apoyo a la toma de decisiones de la organización. Son una tecnología de análisis de datos que presenta una visión multidimensional lógica de los datos en el Data Warehouse.

Este tipo de sistemas organiza los datos de manera multidimensional, lo que permite a los analistas absolver interrogantes en menor tiempo.

h. Cubos OLAP: Son medios de almacenamiento lógico, los cuales contienen datos resumidos de las bases de datos o los sistemas transaccionales y se clasifican en tres: MOLAP, ROLAP y HOLAP (Kimball & Ross, 2002).

i. Multidimensional OLAP (MOLAP): Para Kimball y Ross (2002), el tipo de almacenamiento MOLAP requiere de un preprocesamiento de la información contenida en el cubo OLAP. Todos los datos y sus agregaciones son almacenados dentro de una estructura multidimensional.

Los datos obtenidos del Data Warehouse son pre-agregados en Cubos OLAP y tienen un tiempo de respuesta muy rápido para las consultas.

j. Relational OLAP (ROLAP): Toda la información del cubo, sus datos y agregaciones, son almacenados dentro de una base de datos relacional. Contrario a MOLAP, este accede a las tablas de las bases de datos cada vez que requiere dar respuesta una consulta (Kimball & Ross, 2002).

Generalmente son mucho más lentas en tiempo de respuesta, respecto a los otros tipos; por ello se suelen utilizar cuando se trabaja con gran cantidad de datos que raramente son consultados o cuando se necesita un mayor nivel de detalle.

k. Hybrid OLAP (HOLAP): Combina atributos tanto de MOLAP, como de ROLAP; permitiendo así almacenar las agregaciones en estructuras multidimensionales mientras que los datos a nivel de detalle son almacenados en bases de datos relacionales (Kimball & Ross, 2002).

Es un híbrido entre MOLAP y ROLAP, que usualmente es utilizado cuando se requiere una respuesta rápida o para reducir el espacio ocupado sin perjudicar el rendimiento de las consultas.

l. Procesos ETL: Hace referencia a las herramientas de extracción, transformación y carga de datos. Es la “máquina” o la “cocina” de BI, ya

que aunque no se ve, nada funciona si antes no se ha “cocinado bien” (González, Ramón, & Guitart, 2016).

Se refiere principalmente a todo el procesamiento por el que la data debe pasar antes de la construcción y configuración de los cubos de información, desde la extracción de las diversas fuentes de datos hasta la transformación y/o limpieza de datos en caso se requiera.

m. ETL rules: Son reglas de conversión, transformación y carga que nos indican de qué manera transformar los datos de las tablas de origen a las del DW.

CAPÍTULO III: DESARROLLO DEL TRABAJO DE SUFICIENCIA PROFESIONAL

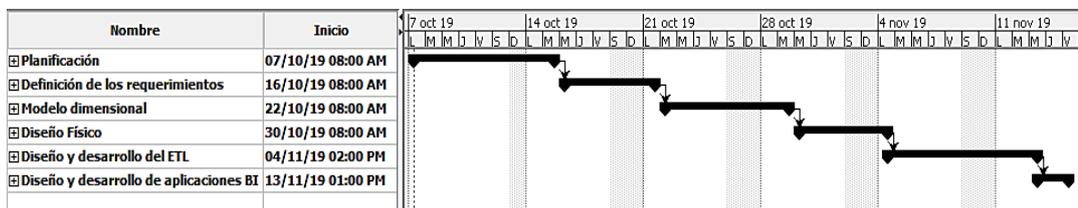
3.1. Modelo de Solución Propuesto

En este apartado se desarrolla la construcción del Datamart basado en la metodología de Kimball, explicada anteriormente. Además, se definen los requerimientos, el análisis dimensional correspondiente y la arquitectura planteada como solución de Inteligencia de Negocios, para los procesos de extracción y explotación.

3.1.1. Planificación

Se muestra el cronograma del proyecto en cada una de sus etapas, respecto al desarrollo del presente trabajo. El detalle de cada una de las etapas se puede visualizar en el Anexo B.

**Figura 10.
Cronograma del Proyecto.**



Fuente: Propia.

3.1.2. Identificación de Requerimientos

Se detallan los requerimientos funcionales y no funcionales necesarios para la elaboración del presente proyecto, los mismos que están asociados a las necesidades de los usuarios de la empresa.

3.1.2.1. *Requerimientos Funcionales*

La Tabla 1 muestra los requerimientos funcionales para el Datamart a construir.

Tabla 1.
Requerimientos funcionales.

Código	Requerimiento	Nivel de Prioridad
RF01	Integrar la data relevante de los diferentes repositorios de información.	1
RF02	Realizar un buen proceso de análisis, previo a la construcción y diseño de la herramienta.	1
RF03	Disminuir el tiempo de elaboración de los reportes.	1
RF04	Generar reportes que permitan visualizar la cantidad total de utilidades de las empresas por trimestre y por año.	1
RF05	Generar reportes que permitan visualizar la cantidad de utilidades por sector.	1
RF06	Generar reportes que permitan diferenciar las utilidades por el lugar geográfico perteneciente a las empresas.	2
RF07	Generar reportes que permitan visualizar la cantidad de utilidades por concepto (reinversión, utilidad pagada y utilidad generada).	1

3.1.2.2. *Requerimientos no Funcionales*

Son aquellos relacionados directamente con el rendimiento y performance del Datamart; así como también, con las herramientas elegidas para la elaboración de estos. La Tabla 2 muestra los requerimientos no funcionales:

Tabla 2.
Requerimientos no funcionales.

Código	Requerimiento	TIPO
RNF01	El tiempo de respuesta del sistema debe ser de a lo mucho 1 minuto para la obtención de la data para un reporte.	DISPONIBILIDAD
RFN02	La herramienta BI de explotación debe permitir realizar técnicas de consulta multidimensional.	
RFN03	La actualización de la data se realizará trimestralmente, ya que los reportes se suelen elaborar en aquellas fechas.	UTILIDAD
RFN04	La herramienta debe ser escalable y permitir cualquier tipo de consulta.	ESCALABILIDAD

3.1.3. Definición de las herramientas de software a utilizar

3.1.3.1. Herramientas de extracción de datos

Se utilizará la herramienta SSIS para la extracción, transformación y carga de datos; tomando como base las características de esta herramienta revisadas con anterioridad. Asimismo, se utiliza esta herramienta también para el procesamiento de cubos, utilizando las funcionalidades de Analysis Services para el procesamiento de cubos OLAP.

3.1.3.2. Herramientas de explotación de datos

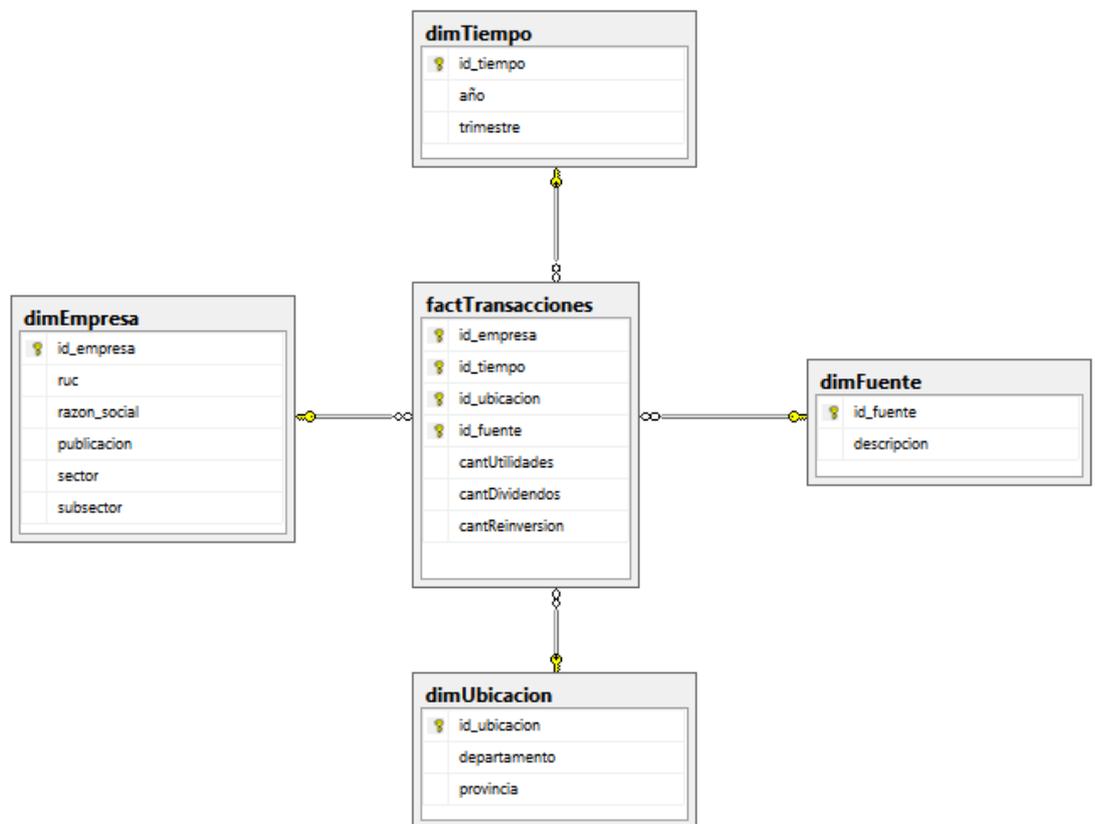
Se elige la herramienta Power BI para la creación de reportes y/ dashboards debido a las características revisadas anteriormente sobre la herramienta como por ejemplo, el análisis en todas las dimensiones y el soporte a múltiples

plataformas. Además, los usuarios cuentan con experiencia previa trabajando con el software ya que, es la herramienta donde la mayoría de reportes del área son actualmente elaborados; lo que permitirá su fácil uso y la rápida adaptación de estos ante la nueva forma de explotar los datos.

3.1.4. Análisis Dimensional

Muestra el análisis de las dimensiones para el Datamart, además de la granularidad y medidas de las fact tables. La siguiente Figura muestra el diseño del Datamart en un Modelo Estrella.

Figura 11.
Modelo Estrella para el Datamart.



Fuente: Propia

3.1.4.1. Dimensiones

Las dimensiones que conforman el Datamart son las siguientes:

- Dimensión Empresa
- Dimensión Fuente
- Dimensión Tiempo
- Dimensión Ubicación

A continuación se pasa a describir cada una de las dimensiones del Datamart:

a) Dimensión Empresa

Es aquella dimensión que posee las características de las empresas con las que trabaja la organización, como lo son la razón social y el ruc.

Referente a las jerarquías, tiene tres niveles los cuales son Publicación, Sector y Subsector.

b) Dimensión Fuente

Esta dimensión contiene datos de la fuente de información de donde se obtuvieron los montos.

c) Dimensión Tiempo

La Dimensión Tiempo posee los niveles en tiempo sobre los cuales se crearán los filtros para las consultas.

Referente a las jerarquías, tiene dos niveles los cuales son Año y Trimestre.

d) Dimensión Ubicación

Contiene los datos del lugar de procedencia de las transacciones de las empresas, específicamente del departamento.

3.1.4.2. Facts

Fact de Transacciones

Esta fact abarca todos los montos reportados por las empresas, los cuales son las utilidades, dividendos y la reinversión. La fact permitirá cumplir con la elaboración de los reportes que se verán en el apartado 3.2.

Tabla 3.

Granularidad.

Nº	Dimensión	Descripción	Llave primaria
1	dimEmpresa	Comprende la información de las empresas que envían sus encuestas.	Si
2	dimTiempo	Comprende la información del periodo en que se reportaron las transacciones.	Si
3	dimUbicación	Comprende la información acerca del lugar de dónde provienen las transacciones de la empresa.	Si
4	dimFuente	Comprende la información de dónde o a través de qué medio se obtuvieron los montos de las transacciones.	Si

Tabla 4.

Medidas.

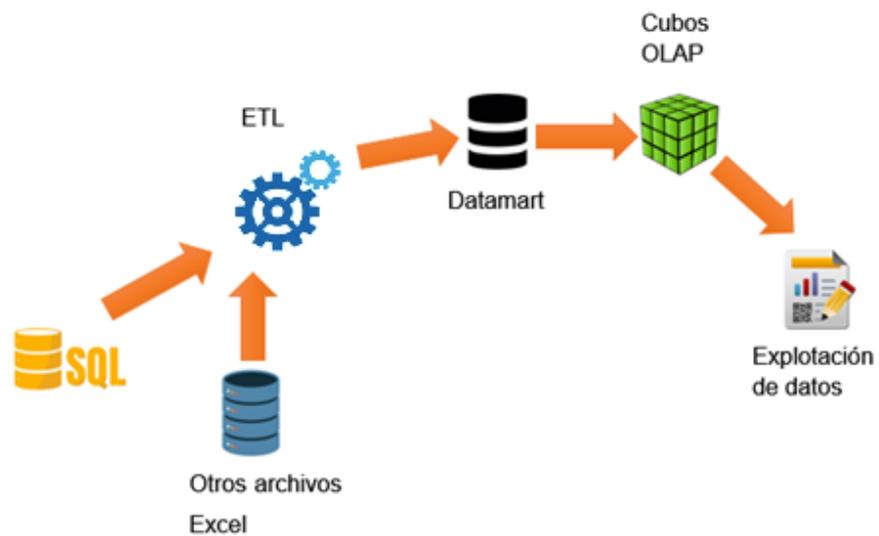
N°	Nombre	Descripción
1	Reinversión	Mide el monto en miles de soles de las reinversiones que una empresa ha realizado.
2	Utilidades	Mide el monto en miles de soles de las utilidades que una empresa ha generado.
3	Dividendos	Mide el monto en miles de soles de las utilidades que una empresa ha pagado a sus subsidiarias o accionistas.

3.1.5. Arquitectura de la solución.

En la figura 12 se muestra la arquitectura usada como solución de Inteligencia de Negocios para el presente proyecto. En ella, las fuentes de datos para el Datamart serán la base de datos transaccional del departamento, además de otros archivos en formato Excel.

Para el proceso ETL, en el cual se realizarán previamente todas las operaciones y transformaciones necesarias antes que los datos pasen a ser almacenados en el Datamart, se utilizará la herramienta SQL Server Integration Services (SSIS), la cual también se empleará como herramienta OLAP para la construcción de los cubos.

Figura 12.
Arquitectura de la solución.



Fuente: Propia.

3.1.6. Proceso de Extracción, Transformación y Carga.

Para la extracción se tomó en cuenta que la información es obtenida a partir de la base de datos que se viene utilizando hasta ahora, la cual está hecha en SQL y además se obtuvieron datos de algunos archivos Excel, que es una de las herramientas que más utilizan. La base multidimensional que se utilizará será una base de datos en SQL.

Para realizar la carga de los datos se seguirá el siguiente proceso:

- 1) Extraer los datos de las fuentes de información
- 2) Aplicar procesos de transformación y limpieza de datos sobre los registros.
- 3) Cargar los registros hacia el destino en el Data Mart.

3.1.6.1. Carga de la Dimensión Empresa

a) Descripción

Representa la carga de la dimensión Empresa desde la tabla Sectorizacion que se encuentran en la base de datos de Utilidades.

b) Descripción de las tablas fuente

Tabla 5.

Descripción de las tablas fuente de dimEmpresa.

Tipo de Fuente	Tabla	Descripción
Base de datos Utilidades (SQL Server)	Sectorizacion	La tabla contiene los datos de las empresas.

c) Estandarización y limpieza de datos

Se especifica el formato que tendrán los campos de la tabla final.

Tabla 6.

Estandarización y limpieza de datos para dimEmpresa.

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
id_empresa	PK	Int	Número correlativo	No debe ser nulo.	No tiene
Ruc		Varchar (11)	Texto		No tiene
razon_social		Varchar (250)	Texto		No tiene
Publicación		Varchar (250)	Texto		No tiene
Sector		Varchar (250)	Texto		No tiene
Subsector		Varchar (250)	Texto		No tiene

d) *Fuentes de datos*

Tabla 7.

Fuentes de datos para dimEmpresa.

Tabla: Sectorizacion				
Nombre	Llave	Tipo	Formato	Consideración importante.
RUC	PK	nvarchar (255)	Texto	No debe ser nulo
[RAZON SOCIAL]		nvarchar (255)	Texto	No tiene
SECTOR		nvarchar (255)	Texto	No tiene
SUBSECTOR		nvarchar (255)	Texto	No tiene
PUBLICACIONES		nvarchar (255)	Texto	No tiene

e) *Tabla destino*

Se indica la descripción general del mapeo.

Tabla 8.

Tabla destino dimEmpresa.

Tabla: dimEmpresa		
Campo	Tipo	Mapeo
id_empresa	Int	Campo incremental.
Ruc	Varchar(11)	Sectorizacion.RUC
razon_social	Varchar(250)	Sectorizacion.[RAZON SOCIAL]
Publicación	Varchar(250)	Sectorizacion.PUBLICACIONES
Sector	Varchar(250)	Sectorizacion.SECTOR
Subsector	Varchar(250)	Sectorizacion.SUBSECTOR

f) *Proceso*

Toma los datos de acuerdo al mapeo realizado, de la tabla Sectorizacion extrae los atributos RUC, [RAZON SOCIAL], PUBLICACIONES, SECTOR y SUBSECTOR. Finalmente, se envía los datos a la tabla dimEmpresa.

3.1.6.2. Carga de la Dimensión Tiempo

a) *Descripción*

Es el proceso que consiste en realizar la carga de la dimensión Tiempo desde la tabla UtilidadesBP que se encuentran en la base de datos Utilidades.

b) *Descripción de las tablas fuente*

Tabla 9.

Descripción de las tablas fuente de dimTiempo.

Tipo de Fuente	Tabla	Descripción
Base de datos Utilidades (SQL Server)	UtilidadesBP	La tabla contiene los datos que se requieren en relación al tiempo, como lo son Año y Trimestre.

c) *Estandarización y limpieza de datos*

Se especifica el formato que tendrán los campos de la tabla final.

Tabla 10.

Estandarización y limpieza de datos para dimTiempo.

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
id_tiempo	PK	Int	Número correlativo	No debe ser nulo.	No tiene
año		Int	Número		No tiene
trimestre		Int	Número		No tiene

d) *Fuentes de datos*

Tabla 11.

Fuentes de datos para dimTiempo.

Tabla: UtilidadesBP				
Nombre	Llave	Tipo	Formato	Consideración importante.
RUC	PK	nvarchar (255)	Texto	No debe ser nulo
TRIM		nvarchar (255)	Texto	Solo tomar el número
AÑO		Float	Número	No tiene

e) *Tabla destino*

Se indica la descripción general del mapeo.

Tabla 12.

Tabla destino dimTiempo.

Tabla: dimTiempo		
Campo	Tipo	Mapeo
id_tiempo	Int	Es un campo incremental.
Año	Int	UtilidadesBP.AÑO
trimestre	Int	UtilidadesBP.TRIM

f) *Proceso*

Toma los datos de acuerdo al mapeo realizado, de la tabla UtilidadesBP extrae los atributos AÑO y TRIM. Finalmente, se envía los datos a la tabla dimTiempo.

3.1.6.3. Carga de la Dimensión Fuente

a) *Descripción*

Representa la carga de la dimensión Fuente desde los archivos de Excel de cada periodo.

b) *Descripción de las tablas fuente*

Tabla 13.

Descripción de las tablas fuente de dimFuente.

Tipo de Fuente	Tabla	Descripción
Archivo Excel	Utilidades [trim][año]	Contiene la procedencia de los montos.

c) *Estandarización y limpieza de datos*

Se especifica el formato que tendrán los campos de la tabla final.

Tabla 14.

Estandarización y limpieza de datos para dimFuente.

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
id_fuente	PK	Int	Número correlativo	No debe ser nulo.	No tiene
descripcion		Varchar (250)	Texto		No tiene

d) *Fuentes de datos*

El archivo tiene una estructura que la tabla UtilidadesBP, siendo la columna de fuente el único diferenciador; por ello solo se describirá esa columna.

Tabla 15.

Fuentes de datos para dimFuente.

Tabla: Utilidades				
Nombre	Llave	Tipo	Formato	Consideración importante.
FUENTE		nvarchar (255)	Texto	No tiene

e) *Tabla destino*

Se indica la descripción general del mapeo.

Tabla 16.

Tabla destino dimFuente.

Tabla: dimFuente		
Campo	Tipo	Mapeo
id_fuente	Int	Es un campo incremental.
Descripción	Varchar(250)	Utilidades.FUENTE

f) *Proceso*

Toma los datos de acuerdo al mapeo realizado, del archivo extrae la data de la columna Fuente y finalmente, se envía los datos a la tabla dimFuente.

3.1.6.4. Carga de la Dimensión Ubicación

a) Descripción

Consiste en realizar la carga de la dimensión Ubicación desde el archivo de Excel Utilidades por Geografía.

b) Descripción de las tablas fuente

Tabla 17.

Descripción de las tablas fuente de dimUbicación.

Tipo de Fuente	Tabla	Descripción
Archivo Excel	Geografía	Contiene el departamento y provincia del lugar de procedencia de los montos.

c) Estandarización y limpieza de datos

Se especifica el formato que tendrán los campos de la tabla final.

Tabla 18.

Estandarización y limpieza de datos para dimUbicación.

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
id_ubicacion	PK	Int	Número correlativo	No debe ser nulo.	No tiene
Departamento		Varchar (250)	Texto		No tiene
Provincia		Varchar (250)	Texto		No tiene

d) *Fuentes de datos*

El archivo contiene principalmente dos tablas de las cuáles se extraerá el departamento y provincia.

Tabla 19.

Fuente de datos1 para dimUbicacion.

Tabla: Departamento				
Nombre	Llave	Tipo	Formato	Consideración importante.
COD_DPTO	PK	Int	Número correlativo	No debe ser nulo
NOM_DPTO		nvarchar (255)	Texto	No tiene

Tabla 20.

Fuentes de datos2 para dimUbicacion.

Tabla: Provincia				
Nombre	Llave	Tipo	Formato	Consideración importante.
COD_PROV	PK	Int	Número correlativo	No debe ser nulo
NOM_PROV		nvarchar (255)	Texto	No tiene

e) *Tabla destino*

Se indica la descripción general del mapeo.

Tabla 21.

Tabla destino dimUbicacion.

Tabla: dimUbicacion

Campo	Tipo	Mapeo
id_ubicacion	Int	Es un campo incremental.
departamento	Varchar(250)	Departamento.NOM_DPTO
Provincia	Varchar(250)	Provincia.NOM_PROV

f) Proceso

Toma los datos de acuerdo al mapeo realizado, del archivo extrae los atributos COD_DPTO y COD_PROV de las hojas Departamento y Provincia respectivamente para obtener la descripción de las columnas NOM_DPTO y NOM_PROV; finalmente se envía los datos a la tabla dimUbicacion.

3.1.6.5. Carga de la Fact Transaccion

a) Descripción

Representa la carga de las tablas UtilidadesBP, Departamento, Provincia, Fuente.

b) Descripción de las tablas fuente

Tabla 22.

Descripción de las tablas fuente de factTransacciones.

Tipo de Fuente	Tabla	Descripción
Base de datos Utilidades (SQL Server)	UtilidadesBP	Contiene los datos de las transacciones realizadas por las empresas en un determinado periodo.
Archivo Excel de Geografía	Departamento	Contiene los datos de ubicación geográfica y procedencia de las transacciones.
Archivo Excel de Geografía	Provincia	Contiene los datos de ubicación geográfica y procedencia de las transacciones.
Archivo Excel de Fuente	Fuente	Representa la forma en que se obtuvieron los montos de las transacciones.

c) Estandarización y limpieza de datos

Se especifica el formato que tendrán los campos de la tabla final.

Tabla 23.

Estandarización y limpieza de datos para factTransacciones.

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
id_empresa	PK	Int	Número	No debe ser nulo.	No tiene
Id_tiempo	PK	Int	Número	No debe ser nulo.	No tiene
Id_ubicacion	PK	Int	Número	No debe ser nulo.	No tiene

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
Id_fuente	PK	Int	Número	No debe ser nulo.	No tiene
cantUtilidades		Float	Número		No tiene
cantDividendos		Float	Número		No tiene
cantReinversion		Float	Número		No tiene

d) Fuentes de datos

Tabla 24.

Fuentes de datos1 para factTransacciones.

Tabla: UtilidadesBP				
Nombre	Llave	Tipo	Formato	Consideración importante.
RUC	PK	nvarchar (255)	Texto	No debe ser nulo
TRIM		nvarchar (255)	Texto	Solo tomar el número
AÑO		Float	Número	No tiene
CONCEPTO		nvarchar (255)	Texto	No tiene
VALOR		Float	Número	No tiene

Tabla 25.

Fuente de datos2 para factTransacciones.

Tabla: Departamento				
Nombre	Llave	Tipo	Formato	Consideración importante.
COD_DPTO	PK	Int	Número correlativo	No debe ser nulo
NOM_DPTO		nvarchar (255)	Texto	No tiene

Tabla 26.

Fuentes de datos3 para factTransacciones.

Tabla: Provincia				
Nombre	Llave	Tipo	Formato	Consideración importante.
COD_PROV	PK	Int	Número correlativo	No debe ser nulo
NOM_PROV		nvarchar (255)	Texto	No tiene

Tabla 27.

Fuentes de datos4 para factTransacciones.

Tabla: Utilidades				
Nombre	Llave	Tipo	Formato	Consideración importante.
FUENTE		nvarchar (255)	Texto	No tiene

e) *Tabla destino*

Se indica la descripción general del mapeo.

Tabla 28.

Tabla destino factTransacciones.

Tabla: factTransacciones		
Campo	Tipo	Mapeo
id_empresa	Int	Se obtiene el atributo RUC y se hace un lookup con la dimensión Empresa.
id_tiempo	Int	Se obtiene los atributos AÑO y TRIM, y se hace un lookup con la dimensión Tiempo.

id_ubicacion	Int	Se obtiene los atributos NOM_DPTO y NOM_PROV, y se hace un lookup con la dimensión Ubicacion.
id_fuente	Int	Se obtiene el atributo FUENTE y se hace un lookup con la dimensión Fuente.
cantUtilidades	Float	Se calcula
cantDividendos	Float	Se calcula
cantReinversion	Float	Se calcula

f) Proceso

Toma los datos de acuerdo al mapeo realizado, se crea una tabla temporal para los datos referentes a la fuente y finalmente a través de lookups se cargan los datos a la tabla factTransacciones.

3.1.7. Proceso de Explotación.

En la explotación de la información se refiere a aquellos reportes que servirán de ayuda al departamento para la toma de decisiones. Este proceso se realizó a través de la herramienta Power BI, el cual se conectó al cubo para la obtención y posterior visualización de la data requerida. El detalle de los reportes se visualiza en el siguiente apartado de Resultados.

3.2. Resultados

Los resultados obtenidos satisfacen los objetivos planteados en el Capítulo I, asimismo se cumple con las necesidades especificadas por los usuarios mediante los requerimientos.

Se visualizó que el departamento ya utilizaba varias de las herramientas de Microsoft y por lo tanto se decidió trabajar con productos de la misma línea para la construcción del Datamart.

Mediante la implementación del cubo se logró obtener información de una manera más fácil y en corto tiempo. Además de la posibilidad de visualizar la información a través de gráficos interactivos.

Figura 13.
Dashboard general de las transacciones.



Fuente: Propia.

El dashboard muestra de manera general el total de utilidades, dividendos y reinversión reportado por las empresas en un determinado periodo, además se visualiza una tabla comparativa entre los dos últimos años y de esta manera ver si existe algún monto que haya variado mucho para realizarle algún seguimiento y/o verificar la información existente; asimismo se puede obtener las empresas que reportan los mayores montos, para ponerle un especial énfasis en ellas puesto que los movimientos que estas realizan son de vital interés para el departamento.

Los montos reportados por sector también son muy importantes para la gerencia, debido a que el tipo de empresa o el rubro al cual se dedican influyen mucho en cuanto a las cantidades de sus movimientos, por ejemplo se puede visualizar en el dashboard que el sector minero es uno de los que mayores movimientos realiza dentro del país, seguido del de servicios; por o

mencionado se recomienda ponerle prioridad a las empresas de que conforman estos sectores.

En cuanto a la ubicación geográfica de las empresas, esto es importante tenerlo en cuenta principalmente antes y durante el procesamiento de la información, ya que obtener la información de las empresas residentes en Lima es de alguna manera más factible que las empresas que residen en provincias debido a que se tiene una menor probabilidad de que reciban los cargos físicos (documentos que anuncian formalmente el envío de las encuestas).

CONCLUSIONES

- Se desarrolló el Datamart siguiendo la metodología de Kimball, la cual fue la más apropiada para el desarrollo del presente proyecto; además se realizó los procesos de extracción, transformación y carga mediante la herramienta SSIS para posteriormente mostrar los datos del cubo en dashboards realizados en Power BI.
- Mediante el análisis dimensional correspondiente se analizó y escogió la data relevante de las diferentes fuentes de información para la construcción del Datamart.
- Se construyó un modelo de datos OLAP correspondiente para el Datamart, identificando las medidas y jerarquías necesarias. Esto permitió la ejecución de consultas a partir de información previamente cargada.
- A través de la herramienta Power BI, se generaron dashboards con la data contenida en el cubo, lo cual permitió la entrega oportuna y relevante de información, además de apoyar al departamento en la toma de decisiones.

RECOMENDACIONES

- Continuar la construcción del modelo multidimensional, abarcando la información de toda la gerencia o subgerencia, lo cual se vio postergado por limitaciones en el tiempo.
- Otra buena para realizar los ETLs es utilizar herramientas Open Source, como lo es Pentaho por ejemplo, para futuras implementaciones, ya que esto permitiría disminuir los costos de licencia.
- Una vez obtenida gran cantidad de datos históricos se recomienda utilizar técnicas de data mining para mejorar los reportes de proyección.
- Para futuras implementaciones es importante tener un proceso de calidad de datos que evite tener inconsistencias, además de posibles errores en el proceso de análisis.

BIBLIOGRAFÍA

- Bianchi, B., & Valverde, M. C. (2009). *Un caso de estudio en Calidad de Datos para Ingeniería de Software Empírica*. Proyecto de Grado, Universidad de la República, Montevideo.
- Caralt, J., & Díaz, J. (2010). *Introducción al Business Intelligence*. Barcelona: Editorial UOC. Recuperado el 21 de Octubre de 2018
- Esparza, D., & Alvarez, C. (2014). *Análisis, diseño e implementación de un Datamart utilizando herramientas Open Source para la unidad administrativa y financiera de la ESPE*. Tesis, Universidad de las Fuerzas Armadas - ESPE, Ecuador.
- Fernández, M. (2018). *Desarrollo de un modelo de calidad de datos aplicado a una solución de inteligencia de negocios en una institución educativa: Caso Lambda*. Pontificia Universidad Católica del Perú, Lima, Lima. Recuperado el 12 de Octubre de 2018
- Fincowsky, E. (2011). Toma de decisiones empresariales. *Contabilidad y Negocios*, 6(11). Recuperado el 15 de Octubre de 2018
- Gonzáles, X., Ramón, J., & Guitart, I. (2016). *¿Cómo planificar un proyecto de inteligencia de negocio?* Barcelona: Editorial UOC.
- Inmon, W. H. (2005). *Building the Data Warehouse*. Estados Unidos de Norteamérica: John Wiley & Sons, Inc.
- International Organization for Standardization. (2015). *ISO 9000*. Recuperado el 21 de Octubre de 2018, de sitio web de International Organization for Standardization: <https://www.iso.org/>
- Jordá, P. (2016). *La utilización de herramientas Business Intelligence para la gestión empresarial. Un estudio aplicado a la empresa Desarrollo y Servicios Infovec S.L.* Trabajo final de Máster, España.

- Kimball, R., & Caserta, J. (2014). *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming and Delivering Data*. John Wiley & Sons.
- Kimball, R., & Ross, M. (2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. Estados Unidos de Norteamérica: John Wiley & Sons, Inc. Recuperado el 30 de Octubre de 2018
- Lopez, M., & Guerrero, R. (2017). *Modelo de Business Intelligence y Analytics soportado por la tecnología Cloud Computing para PYMES del sector retail*. Tesis, Universidad Peruana de Ciencias Aplicadas (UPC)., Lima.
- Medina, E. (2013). *Business Intelligence: la información como arma competitiva*. Universidad Peruana de Ciencias Aplicadas (UPC).
- Méndez, L. (2006). *Más allá del Business Intelligence: 16 experiencias de éxito*. Barcelona, España: Ediciones Gestión 2000.
- Moreno, R. (2014). *Análisis, diseño e implementación de datamarts para las áreas de ventas y recursos humanos de una empresa dedicada a la exportación e importación de productos alimenticios*. Tesis, Lima.
- Mundy, J., & Thornthwaite, W. (2006). *The Microsoft Data Warehouse Toolkit: With SQL Server 2005 and the Microsoft Business Intelligence Toolset*. Indianapolis: Wiley.
- Peña, A. (2006). *Inteligencia de Negocios: una propuesta para su desarrollo en las organizaciones*. Instituto Politécnico Nacional Dirección de Publicaciones. Recuperado el 20 de Octubre de 2018
- Redman, T. (2002). *A Long, strange trip ahead: Process management and Data quality*. .
- Rosales, C. (2009). *Análisis, diseño e implementación de un datamart para el soporte de toma de decisiones y evaluación de las estrategias sanitarias en las Direcciones de Salud*. Tesis, Pontificia Universidad Católica del Perú, Lima, Lima.

Ruiz, C. (2018). *Business Intelligence: Ejemplo de una implantación*. Trabajo de Fin de Grado, Universidad Politécinca de Valencia, España.

Sandoval, A. (2018). *Análisis de métodos y técnicas de Limpieza de Datos existentes y aplicación en un Sistema CRM para una institución educativa limeña*. Tesis, Pontificia Universidad Católica del Perú, Lima, Lima. Recuperado el 12 de Octubre de 2018

Tableau. (s.d.). Acceso em 27 de Octubre de 2019, disponível em Tableau Web site: <https://www.tableau.com/learn/articles/business-intelligence-examples>

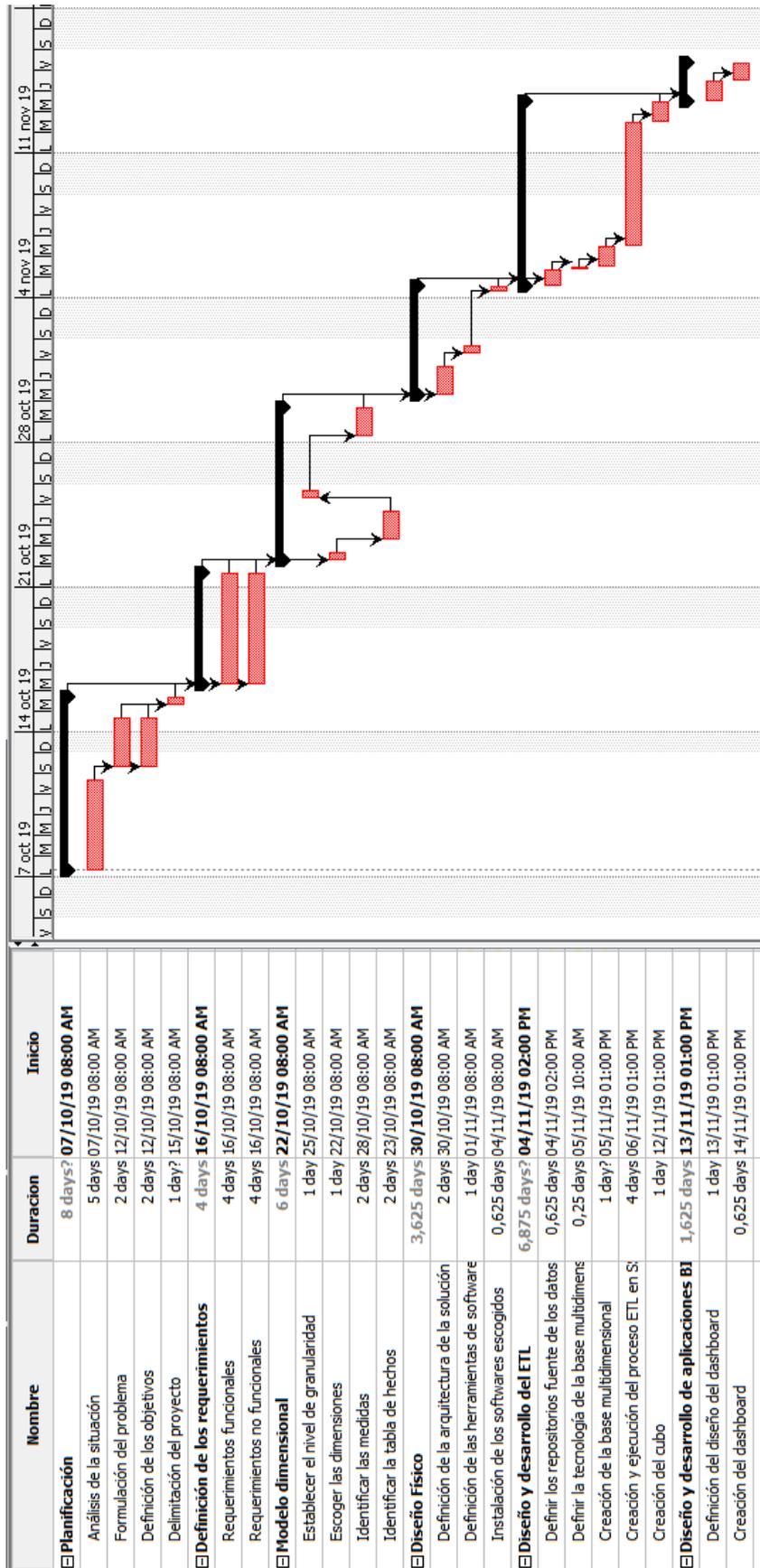
Vitt, E., Luckevich, M., & Mismar, S. (2003). *Business Intelligence: Técnicas De Análisis Para La Toma De Decisiones Estratégicas*. España: McGraw-Hill. Recuperado el 21 de Octubre de 2018

ANEXOS

Anexo A: Guía de entrevista

Fecha: Área: Departamento del banco
Objetivo: Conocer la situación actual del departamento
Preguntas: 1. ¿Cuál es la función del departament dentro del banco? 2. ¿Cómo se realizan los reportes? 3. ¿Cuáles son las principales inconvenientes en cuanto a la generación de los reportes?
Resumen: <p>El área se encarga principalmente de registrar las transacciones económicas, ya sean de exportación o importación, de las diferentes empresas a través de encuestas que se envían cada cierto periodo de tiempo. Esta información es importante, ya que con ella se realizan las publicaciones que se emiten.</p> <p>Dentro de sus principales publicaciones se encuentran los "Informes Semanales", los cuales cuentan con información relevante para la economía, y dentro de ello se encuentran, además de otros reportes, como lo son los "Reportes Estadísticos". La fuente de información de estos reportes provienen de diferentes repositorios de los departamentos de la gerencia, siendo primordialmente dos de los reportes sobre los cuales se ven reflejados el trabajo realizado por el Departamento, presentando así apartados con la información recopilada de las Encuestas.</p> <p>Dentro de los principales problemas que ocurren al generar los reportes son la incongruencia entre datos o datos errados, reportes con mala o escasa información útil y algunas demoras en cuanto a la salida de reportes se refiere.</p>

Anexo B: Cronograma del Proyecto



Anexo C: Scrip para la creación del Datamart

```
CREATE DATABASE DM_Transacciones
GO
```

```
USE DM_Transacciones
GO
```

```
CREATE TABLE dimEmpresa(
    id_empresa INT NOT NULL IDENTITY,
    ruc VARCHAR(11),
    razon_social VARCHAR(250),
    publicacion VARCHAR(250),
    sector VARCHAR(250),
    subsector VARCHAR(250),
    PRIMARY KEY(id_empresa)
)
```

```
CREATE TABLE dimTiempo(
    id_tiempo INT NOT NULL IDENTITY,
    año INT,
    trimestre INT,
    PRIMARY KEY(id_tiempo)
)
```

```
CREATE TABLE dimFuente(
    id_fuente INT NOT NULL IDENTITY,
    descripcion VARCHAR(250),
    PRIMARY KEY(id_fuente)
)
```

```
CREATE TABLE dimUbicacion(
    id_ubicacion INT NOT NULL,
    departamento VARCHAR(250),
    provincia VARCHAR(250),
    PRIMARY KEY(id_ubicacion)
)
```

```
CREATE TABLE factTransacciones(
    id_empresa INT NOT NULL,
    id_tiempo INT NOT NULL,
    id_ubicacion INT NOT NULL,
    id_fuente INT NOT NULL,
    cantUtilidades FLOAT,
    cantDividendos FLOAT,
    cantReinversion FLOAT,
    PRIMARY KEY(id_empresa,id_tiempo,id_ubicacion,id_fuente),
    FOREIGN KEY (id_empresa) REFERENCES dimEmpresa(id_empresa),
    FOREIGN KEY (id_tiempo) REFERENCES dimTiempo(id_tiempo),
    FOREIGN KEY (id_ubicacion) REFERENCES dimUbicacion(id_ubicacion),
    FOREIGN KEY (id_fuente) REFERENCES dimFuente(id_fuente)
)
```

Anexo D: Proceso ETL realizado en SSIS

