

**UNIVERSIDAD NACIONAL TECNOLÓGICA DE LIMA SUR**

**FACULTAD DE INGENIERÍA Y GESTIÓN  
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**



**IMPLEMENTACION DE UN DATAMART PARA MEJORAR LA TOMA DE  
DECISIONES EN LA DIRECCIÓN EJECUTIVA DE CENSOS Y  
ENCUESTAS DE EMPRESAS Y ESTABLECIMIENTOS EN EL  
INSTITUTO NACIONAL DE ESTADÍSTICA E INFORMÁTICA**

**TRABAJO DE SUFICIENCIA PROFESIONAL**

Para optar el Título Profesional de

**INGENIERO DE SISTEMAS**

**PRESENTADO POR EL BACHILLER**

VELASQUE RAMIREZ, JOSÉ

**Villa El Salvador  
2017**

## DEDICATORIA

La concepción de este proyecto está dedicada a mis padres, Julio Velasque Cilvera y María Magdalena Ramirez Rivera, pilares fundamentales en mi vida. Sin ellos, jamás hubiese podido conseguir lo que hasta ahora he logrado. Su tenacidad y lucha insaciable han hecho de ellos el gran ejemplo a seguir y destacar, no solo para mí, sino para mi familia en general.

## **AGRADECIMIENTO**

Agradezco a Dios por darme la vida, sabiduría, y fortaleza para seguir luchando y culminar una etapa más en mi vida como profesional.

A la Facultad de Ingeniería y Gestión, por darme la oportunidad de formar parte de ella y formarme como profesional.

A los Docentes que impartieron sus saberes durante toda mi carrera.

Al Ing. Teodoro Nery Díaz Leyva, quien estuvo incondicionalmente ayudándome como tutor en el transcurso del desarrollo del proyecto.

A todos quienes estuvieron durante mi carrera de una u otra forma apoyándome incondicionalmente.

## INDICE

INTRODUCCIÓN.....	X
CAPITULO I.....	1
PLANTEAMIENTO DEL PROBLEMA .....	1
1.1. DESCRIPCIÓN DE LA REALIDAD PROBLEMÁTICA .....	1
1.2. JUSTIFICACION DEL PROBLEMA.....	4
1.3. DELIMITACIÓN DE LA INVESTIGACIÓN .....	5
1.4. FORMULACION DEL PROBLEMA.....	7
1.4.1. PROBLEMA GENERAL .....	7
1.4.2. PROBLEMAS ESPECÍFICOS.....	7
1.5. OBJETIVOS. ....	8
1.5.1. OBJETIVO GENERAL.....	8
1.5.2. OBJETIVOS ESPECÍFICOS. ....	8
CAPITULO II.....	9
MARCO TEORICO .....	9
2.1. ANTECEDENTES.....	9
2.2. BASES TEÓRICAS.....	13
2.2.1 Inteligencia de Negocios.....	13
2.2.2 Sistemas OLTP (On-Line Transaction Processing).....	13
2.2.3 Sistemas OLAP (On-Line Analytical Processing).....	13
2.2.4 Sistemas MOLAP .....	14
2.2.5 Sistemas ROLAP .....	15
2.2.6 Sistemas HOLAP .....	16
2.2.7 DataWarehouse.....	16

2.2.8 Datamart.....	16
2.2.9. Metodología de Ralph Kimball.....	17
2.2.10 Definición de toma de decisiones .....	17
2.3. MARCO CONCEPTUAL.....	21
2.3.1. Inteligencia de Negocios (BI) .....	21
2.3.2. Metodología de Ralph Kimball.....	23
2.2.2. Metodología de Bill Inmon.....	31
2.3.4. Proceso ETL .....	34
2.3.6. Analysis Services SQL Server (SSAS).....	37
2.3.7. Sistema de soporte de decisiones (DSS).....	37
2.3.8. Data Warehouse .....	38
2.3.9. Datamart.....	39
2.3.10. Dimensiones.....	39
2.3.11. Métricas .....	41
2.3.12. Indicadores.....	42
2.3.13. Fact Table .....	42
2.3.14. Esquema estrella .....	44
2.3.16. Enfoque “Bottom-up” .....	46
2.3.17. Enfoque “Top-down”.....	47
2.3.18. Power BI .....	47
CAPITULO III.....	51
DESARROLLO DE LA METODOLOGIA.....	51
3.1. Analisis de la Metodología del proyecto .....	51
3.1.1 Metodología de Bill Inmon.....	51
3.1.2 Metodología de Ralph Kimball.....	52
3.1.3 Elección de metodología.....	52

3.2 Construcción del modelo Datamart .....	56
3.2.1 Planificación del Proyecto .....	56
3.2.2 Definición de los Requerimientos del Negocio .....	59
3.3. Revisión y Consolidación de Resultados .....	96
CONCLUSIONES .....	103
RECOMENDACIONES .....	104
BIOBLOGRAFIA .....	105
ANEXOS .....	108
ANEXO 1: FLUJO DE COSTOS .....	108
ANEXO 2: ENCUESTA DE VALORACION DE DATAMART .....	109

## INDICE DE FIGURAS

Figura 2-1: Los sistemas de información en las organizaciones .....	19
Figura 2-2: Pirámide Jerárquica de la Toma de Decisiones. ....	20
Figura 2-3: Los componentes de inteligencia de negocios .....	22
Figura 2-4: Fases de la metodología Ralph Kimball .....	24
Figura 2-5: Enfoque de la metodología de Bill Inmon .....	31
Figura 2-6: Enfoque del proceso ETL .....	34
Figura 2-7: Esquema estrella .....	44
Figura 2-8: Esquema copo de nieve.....	45
Figura 2-9: Aplicación Power BI.....	48
Figura 2-10: Fuentes de Power BI .....	49
Figura 2-11: Power BI .....	50
Figura 3-1: EDT del Proyecto.....	56
Figura 3-2: EDT del Proyecto.....	57
Figura 3-3: Modelo Físico del DATAMART .....	72
Figura 3-4: Modelo Dimensional del DATAMART .....	73
Figura 3-5: Entorno Power BI .....	91
Figura 3-6: Conexión con el origen de datos .....	92
Figura 3-7: Conexión con la base de datos .....	92
Figura 3-8: Visualización de las dimensiones y la fact table.....	93
Figura 3-9: Dimensiones y la tabla de hechos en Power BI .....	93
Figura 3-10: Visor de Tablas .....	94
Figura 3-11: Visualización de dimensiones y la tabla de hechos .....	94
Figura 3-12: Visualización de reportes solicitados de la EEA en la DECEEE .....	95
Figura 3-13: cantidad de empresas por sector económico a través de los años .....	101
Figura 3-14: Cantidad de Empresas por Departamento.....	101

## INDICE DE TABLAS

Tabla 3-1: Comparación entre metodologías de Inmon Y Kimball.....	52
Tabla 3-2: Recursos y Costos de la EDT.....	57
Tabla 3-3: Costo Total del Proyecto.....	58
Tabla 3-4: Identificador de Cantidad de empresas – Tiempo – Departamento – Participación – Empresa.....	59
Tabla 3-5: Identificador de Cantidad de Personal Ocupado – Tiempo – Sector Economico – Participación – Empresa.....	60
Tabla 3-6: Identificador de Cantidad de Empresas Omisas – Tiempo – Departamento – Participación – Exoneracion – Empresa.....	60
Tabla 3-7: Identificador de Total de Ventas Acumuladas por Sector Económico – Tiempo – Sector Economico – Empresa.....	61
Tabla 3-8: Identificador de Total de Ventas Acumuladas por Provincias – Tiempo – Departamento – Provincia – Empresa.....	61
Tabla 3-9: Identificador de Cantidad de Empresas por Sector economico – Tiempo – Sector Economico – Participacion – Empresa.....	62
Tabla 3-10: Identificador de Cantidad de Empresas por Departamento – Tiempo – Departamento – Participacion – Empresa.....	62
Tabla 3-11: Identificador de Cantidad de Empresas por Organización juridica – Tiempo – Organización Juridica – Participacion – Empresa.....	63
Tabla 3-12: Identificador de Cantidad de Empresas por Estrato Economico – Tiempo – Estrato Economico – Participacion – Empresa.....	63
Tabla 3-13: Participación de empresas.....	64
Tabla 3-14: Personal ocupado.....	64



Tabla 3-15: Empresas omisas .....	64
Tabla 3-16: Total de ventas .....	64
Tabla 3-17: Identificador de Cantidad de Empresas por .....	65
Tabla 3-18: Empresas por Sector Economico.....	65
Tabla 3-19: Empresas por departamento .....	65
Tabla 3-20: Empresas por organización jurídica .....	65
Tabla 3-21: Empresas por estrato económico.....	65
Tabla 3-22: Objetivos vs Resultados .....	97
Tabla 3-23: Evaluación de tiempo, proceso manual.....	98
Tabla 3-24: Muestra del Excel, cantidad de empresas por sector económico .....	99
Tabla 3-25: Evaluación de tiempo, usando Datamart.....	100

## INTRODUCCIÓN

Actualmente, el instituto nacional de estadística e informática (INEI), tiene como objetivo producir y difundir información estadística oficial que el país necesita con calidad, oportunidad y cobertura requerida, con el propósito de contribuir al diseño, monitoreo y evaluación de políticas públicas y al proceso de toma de decisiones de los agentes socioeconómicos, el sector público y la comunidad en general. Debido a ello la Dirección Nacional de Censos y Encuestas (DNCE) a través de la Dirección Ejecutiva de Censos y Encuestas de Empresas Y Establecimientos (DECEEE), realiza el análisis de la información para orientar su toma de decisiones respecto a los censos o encuestas que están a cargo de la dirección, como es el caso de la Encuesta Económica Anual (EEA).

Sin embargo, en el proceso de análisis de la información se presentan diferentes problemas al momento de realizar los cruces o validación de datos de forma manual, debido a que estos se encuentran en diversos gestores de base de datos, los cuales contienen datos históricos, actuales y los suministrados por las diversas entidades del estado (Sunat, ministerios, etc.). Ante esto, es necesario implementar una solución de inteligencia de negocios, que nos permita crear procesos cuantitativos para mejorar la toma de decisiones.

Finalmente, nace la propuesta de implementar un datamart para mejorar la toma de decisiones en la DECEEE en el INEI, con lo que se busca combinar la información de las empresas desde de una perspectiva histórica y actual, además pudiendo predecir eventos futuros y así llegar a tomar las mejores de decisiones para el beneficio de la institución.

## **CAPITULO I**

### **PLANTEAMIENTO DEL PROBLEMA**

#### **1.1. DESCRIPCIÓN DE LA REALIDAD PROBLEMÁTICA**

Con el paso de los años la tecnología evoluciona constantemente, haciendo posible cada vez más los desarrollos de soluciones integrales en informática, bases de datos y comunicaciones.

En la actualidad, las organizaciones diariamente generan grandes volúmenes de datos, las cuales muchas veces son imposibles de analizar o la mayor parte de estos datos generados no aportan información para la toma de decisiones organizacionales. Para poder utilizarlos es necesario que se transformen en información útil para los usuarios.

Casi todas las organizaciones en la actualidad disponen de bases de datos que almacenan datos sobre actividades, tales como: clientes, colaboradores, personal, distribuidores, proveedores, etc. Todo esto por medio de distintos programas informáticos como son: facturación, planillas, contabilidad, etc.

Por lo tanto, se puede afirmar que las empresas disponen por lo general, de una gran multitud de datos históricos, confiables y voluminosos de todas las actividades realizadas. Ante esto es lógico pensar que dichos datos pueden ser filtrados, agrupados, tratados y analizados para intentar extraer información adecuada, para la mejor toma de decisiones dentro de las organizaciones.

Las organizaciones en el Perú no son ajenas a esta realidad ya que existen una gran cantidad de ellas que hacen uso de bases de datos, y que necesitan tratar mejor los datos para lograr la obtención de información que sea de utilidad para la empresa. El limitante para cubrir esta necesidad es que las herramientas necesarias para este tipo de procesos son poco difundidos o en su defecto son de nula difusión y por ello no son aplicados.

El problema que aqueja actualmente el INEI, en la DECEEE, es el inadecuado manejo de la información, debido a que la información requerida en el área, se extrae en diferentes gestores de base de datos, donde 30% de los datos son extraídos del servidor SIGED (datos proporcionados por otras instituciones ej. SUNAT), que es administrado con el sistema Microsoft Access, otro 35% de los datos son extraídos del servidor VRMSV-QUIPU (datos históricos) que es administrado con el sistema Microsoft SQL Server, otro 30% de los datos son extraídos del servidor RAWA (datos actuales) que es administrado por el sistema Microsoft SQL Server y los otros restantes, 5% de los datos son extraídos de la herramienta Microsoft Excel (datos ubicación y dirección); esta dirección tiene como función principal determinar las metodologías y estrategias para la ejecución de los censos y encuestas de empresas y establecimientos, siendo su principal fuente de recopilación de información la EEA, la cual tiene por finalidad proporcionar información clasificada por actividad económica de las empresas del país sobre ventas,

personal ocupado, remuneraciones, egresos, inversión, valor agregado, estado de situación financiera, estado de resultados por naturaleza, movimiento de activo fijo, depreciación, impuestos, existencias, volúmenes producidos, costos de producción, entre otros.

Dentro de este problema, se observa el inapropiado análisis de información, que fue recopilada mediante la EEA a través de los años, debido a que el proceso de elaboración de reportes se efectúa manual y sistemáticamente, la cual genera retrasos en los tiempos de respuesta a las solicitudes realizadas, esta generación representa aproximadamente un 37% de horas de laborables por día de un personal informático en la dirección, siendo esto equivalente a un promedio de 3 horas diarias (ver tabla 3-23), todo esto provocando retrasos en las atenciones de las solicitudes, haciendo que en muchos casos la entrega de información sea inoportuna y tardía.

Finalmente como consecuencia de los problemas que posee la DECEEE en cuanto a el inadecuado análisis de la información de las empresas recopilado a través de los años en la EEA, datos almacenados en distintos gestores de base de datos y el elevado tiempo requerido por un informático para la elaboración de los reportes, se propone realizar un datamart, el cual ayudara al análisis de la información de las empresas en un menor tiempo y así mejorar la toma de decisiones para determinar las metodologías y estrategias, para la ejecución de los Censos y Encuestas de Empresas y Establecimientos llevados a cargo por esta dirección.

## **1.2. JUSTIFICACION DEL PROBLEMA**

El presente proyecto desarrollado en la Instituto Nacional de Estadística e Informática busca implementar un Datamart para mejorar la toma de decisiones en la DECEEE, debido a que se requiere analizar la información de las empresas que se encuentran almacenados en diferentes gestores de base de datos tales como: Microsoft Access, Microsoft Excel y SQL server que fueron recopiladas a través de los años por la EEA, con esto se busca generar reportes en el menor tiempo posible, medir el porcentaje de empresas que fueron tomadas continuamente en la encuesta por sectores económicos, niveles de crecimiento económico a través de los años de las empresas representativas por división económica o sector económico, porcentajes de empresas omisas por sectores a través de los años, comparación de volúmenes de ventas por sector, número de trabajadores promedio por empresa de acuerdo los sectores, empresas omisas que son representativas por sector y división, niveles de crecimiento respecto a la cantidad de empresas por sector económico, sectores más estables y volátiles a través del tiempo respecto a las altas y bajas de empresas por sector, porcentajes de crecimiento de número de empresas respecto al tamaño de empresa a través de los años, empresas por departamento que entregan información constantemente. A fin de determinar las metodologías y estrategias, para la ejecución de los Censos y Encuestas de Empresas y Establecimientos.

Los resultados obtenidos por la implementación del Datamart serán de gran utilidad, ya que beneficiará a los proyectos a realizarse en el INEI, específicamente a la DECEEE, la misma que podrá generar reportes, combinar la información de las empresas desde una perspectiva histórica y actual y además puedan predecir eventos futuros, que les

permitirá llegar a tomar las mejores decisiones. Finalmente el presente proyecto de investigación causará un alto impacto, debido a que realizará una profunda transformación en los procedimientos de manejo de datos e información en la DECEEE del INEI. A su vez puede servir para otras Instituciones que lleven las mismas reglas de negocio.

### **1.3. DELIMITACIÓN DE LA INVESTIGACIÓN**

**Espacial:** El presente proyecto se desarrollara en la Av. Gral. Garzón 654 - 658, Jesús María, Sede central del Instituto Nacional de Estadística e Informática, Lima-Perú.

**Temporal:** Comprende el período de: abril del 2017 a julio del 2017.

**Teórica:** El presente proyecto de investigación hace uso de conceptos teóricos como:

#### **Indicadores.**

Permite cuantificar alguna dimensión conceptual y su aplicación, genera un valor numérico que por lo general suele ser empleado para comparar desempeños entre períodos.

#### **Gestores de base de datos**

Permiten el almacenamiento, modificación y extracción de la información en una base de datos, además de proporcionar herramientas para añadir, borrar, modificar y analizar los datos

#### **Reportes.**

Se entiende por reportes a los informes que organizan y exhiben la información contenida en una base de datos de modo que tenga un diseño atractivo y que sea fácil de interpretar por los usuarios.



**Censo**

Consiste básicamente en obtener mediciones del número total de individuos mediante diversas técnicas de recuento; además este se realiza cada determinado período.

**Encuesta.**

Permite recopilar datos por medio de un cuestionario previamente diseñado, sin modificar el entorno ni el fenómeno en donde se recoge la información.

**Muestra.**

Se entiende por muestra a un subconjunto de casos o individuos de una población. En diversas aplicaciones interesa que una muestra sea una muestra representativa y para ello debe escogerse una técnica de muestra adecuada que produzca una muestra aleatoria adecuada.

**Metodología.**

Se hace referencia al estudio o elección de un método pertinente o adecuadamente aplicable a un determinado objeto.

**Estrategia**

Describe los procesos de desarrollo e implementación de planes para alcanzar propósitos y objetivos

**Encuesta Económica Anual (EEA)**

Permite medir la actividad económica – financiera de los sectores económicos del país, así como para obtener información indispensable para la elaboración de las variables macroeconómicas e indicadores, que contribuyan a un mejor conocimiento de la realidad económica sectorial y nacional.

## **1.4. FORMULACION DEL PROBLEMA**

### **1.4.1. PROBLEMA GENERAL**

¿Cómo implementar un datamart para mejorar la toma de decisiones en la dirección ejecutiva de censos y encuestas de empresas y establecimientos en el instituto nacional de estadística e informática?

### **1.4.2. PROBLEMAS ESPECÍFICOS.**

- ¿Cómo el análisis de los requerimientos de información de la encuesta económica anual mejorará toma de decisiones en la dirección ejecutiva de censos y encuestas de empresas y establecimientos en el instituto nacional de estadística e informática?
- ¿Cómo el análisis y la construcción del Datamart de la encuesta económica anual mejorará toma de decisiones en la dirección ejecutiva de censos y encuestas de empresas y establecimientos en el instituto nacional de estadística e informática?
- ¿Cómo la generación de reportes de la encuesta económica anual mejorará la toma de decisiones en la dirección ejecutiva de censos y encuestas de empresas y establecimientos en el instituto nacional de estadística e informática?

## **1.5. OBJETIVOS**

### **1.5.1. OBJETIVO GENERAL**

Implementar un Datamart para mejorar la toma de decisiones en la dirección ejecutiva de censos y encuestas de empresas y establecimientos en el instituto nacional de estadística e informática.

### **1.5.2. OBJETIVOS ESPECÍFICOS.**

- Analizar los requerimientos de información de la encuesta económica anual para mejorar la toma de decisiones en la dirección ejecutiva de censos y encuestas de empresas y establecimientos en el instituto nacional de estadística e informática.
- Analizar y construir el Datamart de la encuesta económica anual para mejorar toma de decisiones en la dirección ejecutiva de censos y encuestas de empresas y establecimientos en el instituto nacional de estadística e informática.
- Generar reportes de la encuesta económica anual para mejorar la toma de decisiones en la dirección ejecutiva de censos y encuestas de empresas y establecimientos en el instituto nacional de estadística e informática.

## **CAPITULO II**

### **MARCO TEORICO**

#### **2.1. ANTECEDENTES**

**Análisis, diseño e implementación de un Data Mart de clientes para el área de marketing de una entidad aseguradora (Fernandez, 2010).**

La empresa en donde se implementó la presente solución maneja en su sistema transaccional información de clientes, productos, siniestros, pólizas, certificados y planes. El Área de Marketing necesita explotar la misma para poder analizar la información y tomar mejores decisiones.

La tesis consiste en atender las necesidades de explotación de información del Área de Marketing mediante reportes e indicadores de una entidad aseguradora. La metodología de la elaboración del producto es Business Intelligence Roadmap, que consiste en una guía de buenas prácticas que sigue 16 pasos donde se recogen las actividades a realizar en un proyecto de inteligencia de negocios.

**Tesis: Análisis, diseño e implementación de un Datamart para la Dirección Financiera y Recursos Humanos de la Escuela Politécnica del Ejército para una toma de decisiones efectiva (Lopez, 2007).**

Se orienta a brindar una herramienta de apoyo en la tecnología Business Intelligence de Oracle que sea capaz de solventar la información solicitada en la Dirección Financiera y Recursos Humanos de la Escuela Politécnica del Ejército, dando la posibilidad de formar cubos dimensionales de decisión.

Para el trabajo de la tesis, se utilizó la metodología Rapid Warehousing, este método es iterativo, y está basado en el desarrollo incremental del proyecto de DataWarehouse, que se basa en la definición de un Datamart como la combinación de estructuras de datos especializados para resolver requerimientos de análisis de información en distintas áreas de una institución.

La tesis se enfoca en el control de recursos financieros y humanos y guarda relación con el tema de investigación, de ahí que, ambos se enfocan en la ejecución de un datamart y así conseguir una efectiva toma de decisiones.

**Análisis, diseño e Implementación de una solución de inteligencia de negocios para el Área de Finanzas de la Municipalidad Metropolitana de Lima (Nuñez, 2010).**

Esta investigación tiene como objetivo analizar, diseñar e implementar una solución de inteligencia de negocios para el Área de Finanzas, de manera que, los usuarios cuenten con información sólida, confiable y de alta disponibilidad.

La realidad problemática de esta tesis consiste en que cada uno de los usuarios al realizar los procedimientos de reportes de forma manual, invierten mucho tiempo y los resultados obtenidos no siempre son exactos. Esto debido a que, al manipular una gran cantidad de datos, están propensos al error.

La metodología que usó el proyecto fue una adaptación de la guía del PMBOK (Project Management Body of Knowledge– guía de los fundamentos de la Dirección de Proyectos), desarrollada por el PMI (Project Management Institute- Instituto de Administración de Proyectos). La guía del PMBOK es un estándar reconocido internacionalmente que agrupa un conjunto de buenas prácticas generalmente aceptadas en lo que se refiere a la gestión de proyectos. Está estructurado en cinco grupos de procesos básicos y nueve áreas de conocimiento comunes a casi todos los tipos de proyectos.

La relación que guarda con esta tesis es el tema de investigación, porque ambos se enfocan en el desarrollo de un datamart para una entidad pública. Otro aspecto relacional, se encuentra en la realidad problemática, por lo mismo, que la información requerida no se proporciona en el momento oportuno para la toma de decisiones.

**Tesis: "Análisis, diseño e implementación de un sistema gerencial basado en una suite integrada de Datamarts para las áreas de finanzas, contabilidad, recursos humanos y comercial"(Paitan, 2011).**

Menciona que los modelos multidimensionales de cada uno de los datamarts debe ser lo más completa posible y permitir escalabilidad, debido a que los usuarios siempre podrán

tener nuevos requerimientos en cuanto a dimensiones o variables a analizar y la solución debe permitir estos cambios sin tener que realizar demasiado mantenimiento.

**Tesis: “Análisis, diseño e implementación de una DataWarehouse de soporte de decisiones para un hospital del sistema de salud público” (Villanueva, 2008).**

La realidad problemática de esta tesis precisa que al generar un reporte en el sistema transaccional de salud pública demora días e incluso semanas, ya que se tiene que cumplir obligatoriamente una serie de pasos administrativos para elaborarlo. En cambio, con herramientas tecnológicas, estos reportes son generados, en forma personalizada, en minutos o en segundos.

Este proyecto de tesis, tiene como objetivo implementar una DataWarehouse que ayude al proceso de la toma de decisiones para un hospital del sistema de salud pública.

## **2.2. BASES TEÓRICAS**

### **2.2.1 Inteligencia de Negocios**

Es un enfoque para la gestión empresarial que le permite a una organización definir qué información es útil y relevante para la toma de decisiones corporativas. Inteligencia de negocios es un esquema polifacético que fortalece a las organizaciones para tomar mejores decisiones rápidamente, convertir los datos en información y usar una estrategia inteligente para la gestión empresarial (Cabanillas, 2011).

### **2.2.2 Sistemas OLTP (On-Line Transaction Processing)**

Los Sistemas de tipo OLTP lo conforman todos los sistemas de información que ejecutan datos operacionales del día a día y maximizan la capacidad de procesamiento de transacciones. Estos sistemas realizan principalmente 3 tipos de acciones sobre una Base de Datos, las cuales son INSERT, DELETE y UPDATE que se manejan a nivel de transacciones (Cabanillas, 2011).

### **2.2.3 Sistemas OLAP (On-Line Analytical Processing)**

Los Sistemas de tipo OLAP tiene como principal objetivo realizar un análisis del negocio mediante la identificación de indicadores de gestión que puedan servir de apoyo en la toma de decisiones de una empresa. Este se basa principalmente en realizar consultas de Base de Datos Multidimensional que permita el análisis de variables numéricas. Estos tipos de sistemas analíticos buscan facilitar la creatividad,



centrándose en la estrategia a largo plazo y la ventaja competitiva dentro del negocio de una empresa (Paitan, 2011).

#### **2.2.4 Sistemas MOLAP**

“La arquitectura MOLAP usa unas bases de datos multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente. Un sistema MOLAP usa una base de datos propietaria multidimensional, en la que la información se almacena multidimensionalmente, para ser visualizada en varias dimensiones de análisis.

El sistema MOLAP utiliza una arquitectura de dos niveles: la bases de datos multidimensionales y el motor analítico. La base de datos multidimensional es la encargada del manejo, acceso y obtención del dato.

El nivel de aplicación es el responsable de la ejecución de los requerimientos OLAP. El nivel de presentación se integra con el de aplicación y proporciona una interfaz a través del cual los usuarios finales visualizan los análisis OLAP. Una arquitectura cliente/servidor permite a varios usuarios acceder a la misma base de datos multidimensional” (Sinnexus, 2017).

La arquitectura MOLAP requiere unos cálculos intensivos de compilación. Lectura de datos pre-compilados, y tiene capacidades limitadas de crear agregaciones dinámicamente o de hallar ratios que no se hayan pre-calculados y almacenados previamente.

### **2.2.5 Sistemas ROLAP**

“La arquitectura ROLAP, accede a los datos almacenados en un DataWarehouse para proporcionar los análisis OLAP. La premisa de los sistemas ROLAP es que las capacidades OLAP se soportan mejor contra las bases de datos relacionales.

El sistema ROLAP utiliza una arquitectura de tres niveles. La base de datos relacional maneja los requerimientos de almacenamiento de datos, y el motor ROLAP proporciona la funcionalidad analítica. El nivel de base de datos usa bases de datos relacionales para el manejo, acceso y obtención del dato. El nivel de aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios.

El motor ROLAP se integra con niveles de presentación, a través de los cuáles los usuarios realizan los análisis OLAP. Después de que el modelo de datos para el DataWarehouse se ha definido, los datos se cargan desde el sistema operacional. Se ejecutan rutinas de bases de datos para agregar el dato, si así es requerido por el modelo de datos. Se crean entonces los índices para optimizar los tiempos de acceso a las consultas.

Los usuarios finales ejecutan sus análisis multidimensionales, a través del motor ROLAP, que transforma dinámicamente sus consultas a consultas SQL. Se ejecutan estas consultas SQL en las bases de datos relacionales, y sus resultados se relacionan mediante tablas cruzadas y conjuntos multidimensionales para devolver los resultados a los usuarios.” (Sinnexus, 2017).

### **2.2.6 Sistemas HOLAP**

“Un desarrollo un poco más reciente ha sido la solución OLAP híbrida (HOLAP), la cual combina las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas: desempeño superior y gran escalabilidad. Un tipo de HOLAP mantiene los registros de detalle (los volúmenes más grandes) en la base de datos relacional, mientras que mantiene las agregaciones en un almacén MOLAP separado.

Se puede concluir que ROLAP son sistemas analíticos relacionales que usan Sistemas de Gestores de Bases de Datos (SGBD) para extraer la información, por el contrario la tecnología MOLAP son herramientas de análisis” (Sinnexus, 2017).

### **2.2.7 DataWarehouse**

Es un almacén o repositorio de datos que integra información de diferentes fuentes (base de datos, archivos de texto, hojas de cálculo, etc.) Y permite un análisis para la toma de decisiones. Muchos expertos definen el DataWarehouse como un almacén de datos centralizados que introduce datos en un almacén de datos específico llamado Datamart. Otros aceptan una amplia definición de DataWarehouse, como un conjunto integrado de Datamarts (Cabanillas, 2011).

### **2.2.8 Datamart**

Es un conjunto de datos que son estructurados de una forma que facilite su posterior análisis. Un Datamart contiene la información referente a un área, un tema o

una función en particular, con datos relevantes que provienen de las diferentes aplicaciones operacionales. Los Datamarts pueden ser de diversas bases de datos OLAP dependiendo del tipo de análisis que se quiera desarrollar (Cabanillas, 2011).

### **2.2.9. Metodología de Ralph Kimball**

“La metodología de Kimball se enfoca principalmente en el diseño de la base de datos que almacenará la información para la toma de decisiones. El diseño se basa en la creación de tablas de hechos (FACTS) que son tablas que contienen la información numérica de los indicadores a analizar, es decir la parte cuantitativa de la información” (Leon, 2014).

### **2.2.10 Toma de decisiones**

“La toma de decisiones es el proceso de identificación de un problema u oportunidad y la selección de una alternativa de acción entre varias existentes, es una actividad diligente clave en todo tipo de organización” (Diaz, 2005).

Se puede decir que el proceso de toma de decisiones es la integración de todos los mecanismos que intervienen en el tratamiento de los datos hasta su conversión en conocimiento, por parte de los administrativos, y por último, la ejecución de las medidas.

#### **a. Proceso de toma de decisiones.**

Este proceso debe abarcar las cuatro funciones administrativas fundamentales: planeamiento, organización, conducción y control.

- ✓ **La Planeación:** Selección de misiones y objetivos así como de las acciones para cumplirlas. Esto implica “Toma de decisión”. Se responden preguntas como:
  - ¿Cuáles son los objetivos de la organización, a largo plazo?
  - ¿Qué estrategias son mejores para lograr este objetivo?
  - ¿Cuáles deben ser los objetivos a corto plazo?
  
- ✓ **Organización:** Establecimiento de la estructura que desempeñan los individuos dentro de la organización. Se responden preguntas como:
  - ¿Cuánta centralización debe existir en la organización?
  - ¿Cuándo debe una organización instrumentar una estructura diferente?
  
- ✓ **Dirección:** Esta función requiere que los administradores influyan en los individuos para el cumplimiento de las metas organizacionales y grupales. Se responden preguntas como:
  - ¿Cuál es el estilo de liderazgo más eficaz para una situación dada?
  - ¿Cómo afectará un cambio específico a la productividad del trabajador?
  - ¿Cuándo es adecuado estimular el conflicto?
  
- ✓ **Control:** Es la medición y corrección del desempeño individual y organizacional de manera tal que se puedan lograr los planes. Se responden preguntas como:
  - ¿Qué actividades en la organización necesitan ser controladas?
  - ¿Cómo deben controlarse estas actividades?

- ¿Cuándo la organización está desempeñándose de manera efectiva?

Como se muestra en la Figura 2.1 el proceso de la toma de decisiones abarca 4 funciones importantes que se debe tomar en cuenta a la hora de tomar decisiones en la empresa.



Figura 2-1: Los sistemas de información en las organizaciones

Fuente: <http://sistemasinformaciongerencial1.blogspot.pe/2012/09/tipos-de-sig.html>

#### b. Estructura jerárquica

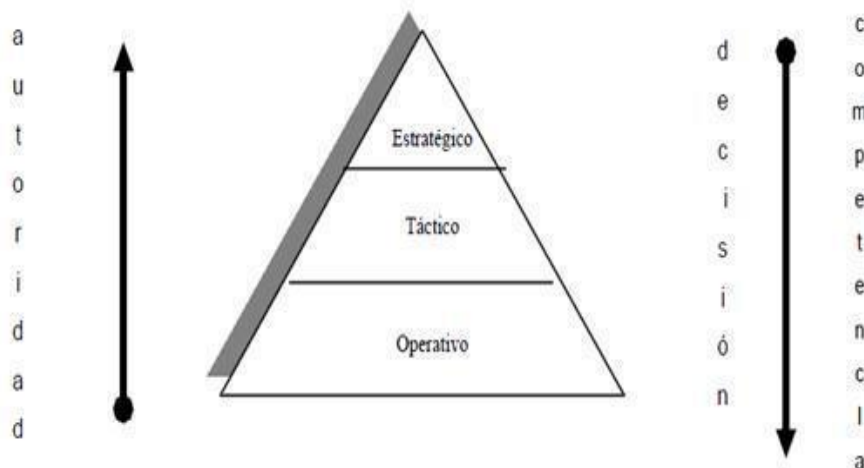
Generalmente el rol que ocupa el personal de las empresas se agrupa en alguno de los tres niveles jerárquicos de autoridad y responsabilidad, particularmente en la toma de decisiones, empleo de recursos, ejecución del trabajo y supervisión de resultados. Estos niveles jerárquicos se conocen como:

- ✓ **Estratégico:** Es el máximo órgano de autoridad y liderazgo, integrado por el consejo administrativo, presidente y directores donde descansa la conducción de la empresa.

- ✓ **Táctico:** Es el nivel intermedio de autoridad, compuesto por los responsables de aplicar los lineamientos dispuestos por los superiores y de dirigir al personal encargado de las tareas.
- ✓ **Operativo:** Formulado por el personal y supervisores especializados en el desempeño de actividades específicas del proceso de producción de bienes o de prestación de servicios representantes del giro comercial de la empresa.

Tradicionalmente los niveles anteriores se ilustran como una pirámide dividida en tres partes, para destacar el orden ascendente de autoridad, el impacto de las decisiones y competencia de las funciones que cada rango realiza.

Como se muestra en la Figura 2.2 se puede observar los niveles o jerarquías de la toma de decisiones.



**Figura 2-2: Pirámide Jerárquica de la Toma de Decisiones.**

**Fuente:** <http://www4.ujaen.es/~cruiz/diplot-5.pdf>

## **2.3. MARCO CONCEPTUAL**

### **2.3.1. Inteligencia de Negocios (BI)**

En el “Boletín de Asesoría Gerencial” proporcionado por Espiñera, Sheldon y asociados, indican que la definición de inteligencia de negocios incluyen una amplia categoría de metodologías, aplicaciones y tecnologías que permiten reunir, acceder, transformar y analizar los datos, transacciones e información no estructurada (interna y externa), con el propósito de ayudar a los usuarios de una compañía a tomar mejores decisiones de negocio (Espiñera, Sheldon y asociados, 2008).

Por otra parte, en el libro Business Intelligence: Josep Lluís Cano, menciona que el objetivo básico de inteligencia de negocios es apoyar de forma sostenible y continuada a las organizaciones para mejorar su competitividad, facilitando la información necesaria para la toma de decisiones (Cano, 2008).

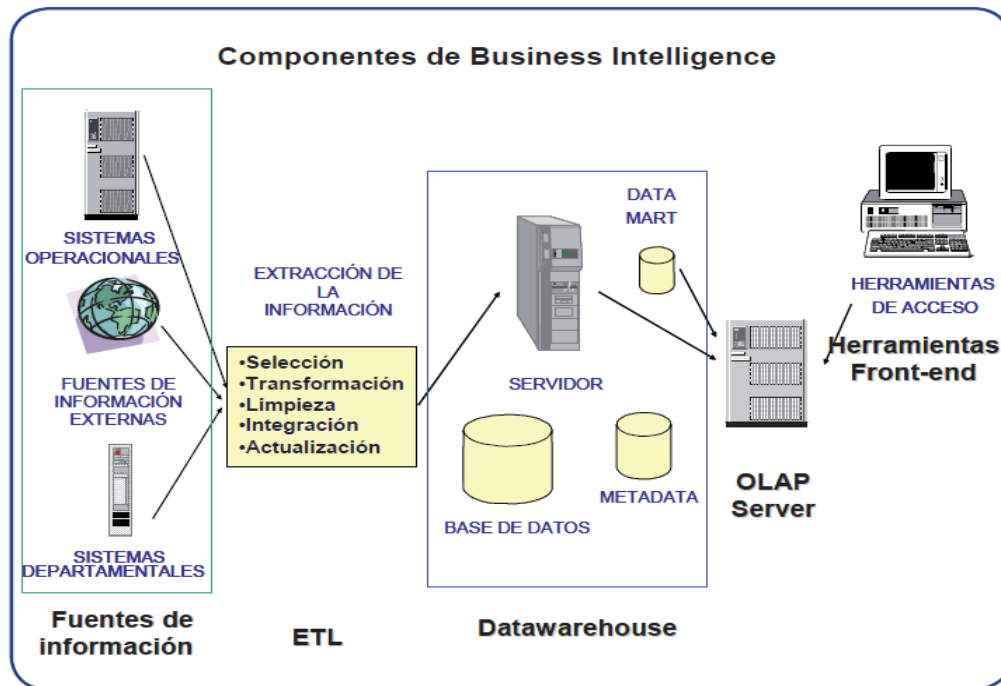
Asimismo, describe tres tipos de beneficios que se pueden obtener a través del uso de inteligencia de negocios, los cuales son:

- ✓ Beneficios tangibles: por ejemplo, reducción de costos, generación de ingresos, reducción de tiempos para las distintas actividades del negocio.
  
- ✓ Beneficios intangibles: el hecho de que tengamos disponible la información para la toma de decisiones hará que más usuarios la utilicen para tomar decisiones y mejorar nuestra posición competitiva.



- ✓ Beneficios estratégicos: la formulación de estrategias, especificarán a qué clientes, mercados o con qué productos dirigimos (Cano, 2008).

A su vez, el mismo autor, nos indica los componentes de inteligencia de negocios:



**Figura 2-3: Los componentes de inteligencia de negocios**

**Fuente:** Business Intelligence: competir con información, Josep Lluís Cano

- ✓ Fuentes de información, de las cuales partiremos para alimentar de información el DataWarehouse.
- ✓ Proceso ETL de extracción, transformación y carga de los datos en la DataWarehouse. Antes de almacenar los datos en una DataWarehouse, éstos deben ser transformados, limpiados, filtrados y redefinidos. Normalmente, la

información que tenemos en los sistemas transaccionales no está preparada para la toma de decisiones.

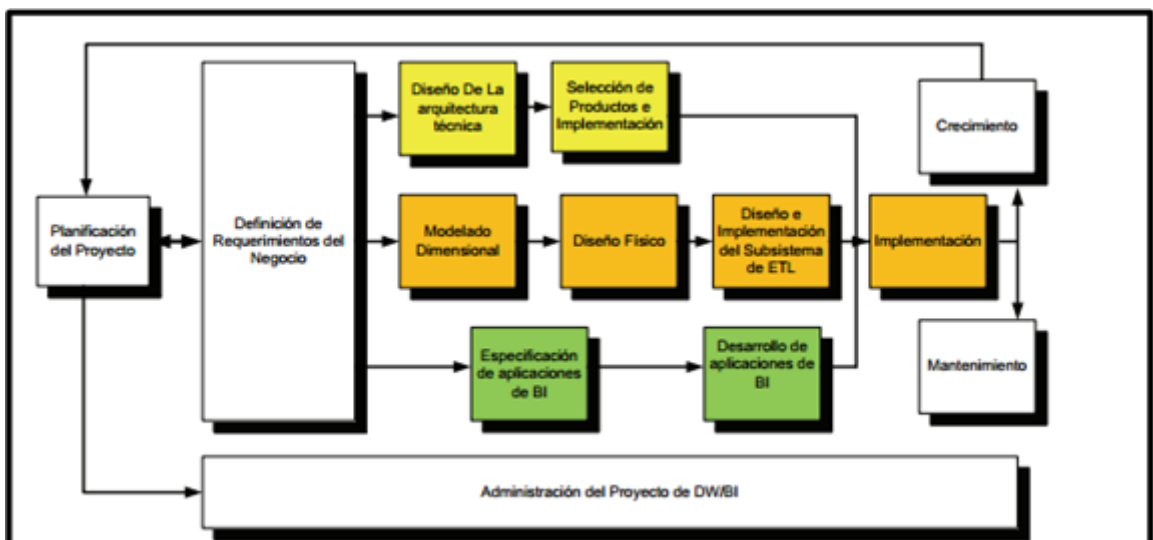
- ✓ La propia DataWarehouse: se busca almacenar los datos de una forma que maximice su flexibilidad, facilidad de acceso y administración.
- ✓ El motor OLAP, nos provee capacidad de cálculo, consultas, funciones de planeamiento, pronóstico y análisis de escenarios en grandes volúmenes de datos. En la actualidad, existen otras alternativas tecnológicas al OLAP.
- ✓ Las herramientas de visualización, nos permiten el análisis y la navegación a través de los mismos (Kimball, 2002).

### **2.3.2. Metodología de Ralph Kimball**

La metodología de Ralph Kimball nos indica que la DataWarehouse es un conglomerado de todos los Datamarts dentro de una empresa, siendo una copia de los datos transaccionales estructurados de una forma especial para el análisis, de acuerdo, al modelo dimensional (no normalizado) que incluyen las dimensiones de análisis y sus atributos, su organización jerárquica, así como los diferentes hechos de negocio que se quieren analizar. Por un lado, tenemos tablas para representar las dimensiones y por otro lado, tablas para los hechos (las facts tables). Los diferentes Datamarts están conectados entre sí, por la llamada bus structure, que contiene los elementos anteriormente citados a través de las dimensiones conformadas (que permiten que los usuarios puedan realizar queries conjuntos sobre los diferentes Datamarts, pues este bus contiene los elementos en común que los comunican). Una dimensión conformada puede ser, por ejemplo, la

dimensión cliente, que contienen todos los atributos o elementos de análisis referentes a los clientes y que puede ser compartida por diferentes Datamarts (ventas, pedidos, gestión de cobros, etc.).

Este enfoque también se referencia como Bottom-up, pues al final el DataWarehouse Corporativo no es más que la unión de los diferentes Datamarts, que están estructurados de una forma común, a través de la bus structure. Esta característica permite que sea flexible y sencillo de implementar, pues podemos construir una Datamart como primer elemento del sistema de análisis, y luego ir añadiendo otros que comparten las dimensiones, ya definidas u otras nuevas. En este sistema, los procesos ETL extraen la información de los sistemas operacionales y los procesan igualmente en el Área Stage, realizando posteriormente, el llenado de cada uno de los Datamart de una forma individual, respetando la estandarización de las dimensiones (dimensiones conformadas). Esta metodología de desarrollo presenta las siguientes fases:



**Figura 2-4: Fases de la metodología Ralph Kimball**

**Fuente:** The Data Warehouse Lifecycle Toolkit, Ralph Kimball

### **a. Planificación del proyecto:**

La planificación busca identificar la definición y el alcance del proyecto de DataWarehouse, también justificaciones del negocio y evaluaciones de factibilidad.

La planificación del proyecto se focaliza sobre recursos, perfiles, tareas, duraciones y secuencialidad. El plan de proyecto resultante identifica todas las tareas y las partes involucradas.

Esta etapa se concentra sobre la definición del proyecto (identificación del escenario del proyecto para saber de dónde surge la necesidad del DataWarehouse). Según sentencia Kimball, “Antes de comenzar un proyecto de DataWarehouse o Datamart, hay que estar seguro si existe la demanda y de dónde proviene. Si no se tiene un sólido usuario sponsor y no hay usuarios entusiasmados, posponga el proyecto”. Factores asociados con estas etapas añaden: identificación de los usuarios sponsors, convincentes motivaciones del negocio, cooperación entre áreas de sistemas y negocios, cultura analítica de la organización y análisis de factibilidad (tanto tecnológica como de disponibilidad de datos).

### **b. Definición de los requerimientos del negocio**

Un factor determinante en el éxito de un proceso de DataWarehousing es la interpretación correcta de los diferentes niveles de requerimientos expresados por los diferentes niveles de usuarios.

La técnica utilizada para relevar los requerimientos de los analistas del negocio difiere de los enfoques tradicionales guiados por los datos. Los diseñadores de los DataWarehouses deben entender los factores claves que guían al negocio para

determinar efectivamente los requerimientos y traducirlos en consideraciones de diseño apropiadas.

Los usuarios finales y sus requerimientos impactan siempre en las implementaciones realizadas de un DataWarehouse. Según la perspectiva de Kimball, los requerimientos del negocio se posicionan en el centro del “universo del DataWarehouse”. Como destaca siempre Kimball, los requerimientos del negocio deben determinar el alcance del DataWarehouse (qué datos debe contener, cómo debe estar organizado, cada cuánto debe actualizarse, quiénes y desde dónde accederán, etc.).

### **c. Modelo dimensional:**

La creación de un modelo dimensional es un proceso dinámico e altamente iterativo.

El proceso de diseño comienza con un modelo dimensional de alto nivel obtenido a partir de los procesos priorizados de la matriz de requerimientos.

El proceso iterativo consiste en cuatro pasos:

- ✓ Elegir el proceso del negocio: el primer paso es elegir el área a modelar. Esta es una decisión de la dirección, y depende fundamentalmente del análisis de requerimientos y de los temas analíticos anotados en la etapa anterior.
  
- ✓ Establecer el nivel de granularidad: es decir, significa especificar el nivel de detalle. La elección de la granularidad depende de los requerimientos del negocio y lo que es posible a partir de los datos actuales. La sugerencia general es comenzar a diseñar el DW al mayor detalle posible, ya que se podría luego realizar agrupamientos al nivel deseado.

- ✓ Elegir las dimensiones: surgen naturalmente de las discusiones del equipo, y facilitadas por la elección del nivel de granularidad y de la matriz de procesos/dimensión. Las tablas de dimensiones tienen un conjunto de atributos (generalmente textuales) que brindan una perspectiva o forma de análisis sobre una medida en una tabla hechos.
  
- ✓ Identificar medidas y las tablas de hechos: el último paso consiste en identificar las medidas que surgen de los procesos de negocio. Una medida es un atributo (campo) de una tabla que desea analizar, agrupando sus datos usando los criterios de corte conocidos como dimensión. Las medidas habitualmente se vinculan con el nivel de granularidad, y se encuentran en tablas que denominamos tablas de hechos. Cada tabla de hechos tiene como atributos una o más medidas de un proceso organizacional de acuerdo a los requerimientos.

#### **d. Diseño físico:**

Se focaliza sobre la selección de las estructuras necesarias para soportar el diseño lógico. Algunos de los elementos principales de este proceso son la definición de convenciones estándares de nombres y seteos específicos del ambiente de la base de datos.

#### **e. Diseño e implementación del subsistema de ETL**

Esta etapa es típicamente la más subestimada de las tareas en un proyecto de DataWarehouse.

Las principales sub-etapas de esta zona del ciclo de vida son: la extracción, la transformación y la carga

Se define como proceso de extracción a aquellos requeridos para obtener los datos que permitirán efectuar la carga del modelo físico acordado.

También se definen como proceso de transformación el convertir o recodificar los datos fuente, a fin poder efectuar la carga efectiva del modelo físico. Por otra parte, los procesos de carga de datos son los requeridos para poblar la DataWarehouse.

Todas estas tareas son altamente críticas, pues tienen que ver con la materia prima de la DataWarehouse: los datos. La desconfianza y pérdida de credibilidad de la DataWarehouse serán resultados inmediatos e inevitables, si el usuario choca con información inconsistente. Es por ello, que la calidad de los datos es un factor determinante en el éxito de un proyecto de DataWarehousing. Es en esta etapa donde deben sanearse todos los inconvenientes relacionados con la calidad de los datos fuente. Como advierte Kimball, el proceso de Data Staging es el iceberg de un proyecto de DataWarehousing. Son muchos los desafíos que deben enfrentarse para lograr datos de alta calidad de los sistemas fuentes. En general, es una de las etapas más subestimadas que siempre termina tomando más tiempo del previsto.

#### **f. Diseño de la arquitectura técnica**

Los ambientes de DataWarehousing requieren la integración de numerosas tecnologías. Se debe tener en cuenta tres factores: los requerimientos del negocio, los actuales ambientes técnicos y las directrices técnicas estratégicas futuras planificadas para de esta forma poder establecer el diseño de la arquitectura técnica del ambiente de DataWarehousing.

Del mismo modo, que en una construcción, los planos sirven para comunicar los deseos entre los clientes y el arquitecto, como así también para medir esfuerzo y materiales necesarios para la obra (comunicación, planificación, flexibilidad y mantenimiento, documentación, productividad y reúso). Finalmente, argumenta Kimball, “Un buen conjunto de planos, como cualquier buena documentación, nos ayudara más tarde cuando sea tiempo de remodelar o hacer incorporaciones”.

#### **g. Selección del producto e implementación**

Utilizando como marco el diseño de arquitectura técnica es necesario evaluar y seleccionar componentes específicos de la arquitectura como ser la plataforma de hardware, el motor de base de datos, la herramienta de ETL o el desarrollo pertinente, herramientas de acceso, etc.

Una vez evaluados y seleccionados los componentes determinados se procede con la instalación y prueba de los mismos en un ambiente integrado de DataWarehousing.

#### **h. Especificación de aplicaciones de BI**

No todos los usuarios del Warehouse necesitan el mismo nivel de análisis. Es por ello, que en esta etapa se identifican los diferentes roles o perfiles de usuarios para determinar los diferentes tipos de aplicaciones necesarias sobre la base al alcance de los diferentes perfiles (gerencial, analista del negocio, vendedor, etc.) Kimball se concentra sobre el proceso de creación de aplicaciones “templates”. Comienza definiendo el concepto de la aplicación para usuario final y su rol en el acceso a la información del negocio. Brinda un marco metodológico bastante estándar en lo que ha desarrollo de aplicaciones (como piezas de software) se refiere. Divide el proceso de creación de las aplicaciones para usuarios finales en dos grandes fases: especificación y



desarrollo. Clasifica a los usuarios según su perfil de consulta, así tenemos, usuarios con un perfil más estratégico y menos predecibles (power users) otra de las clases son los usuarios netamente operacionales que consumen una serie de reportes estándares (final users) pasando por los usuarios gerenciales con uso de interfaces push-button.

**i. Desarrollo de aplicaciones de BI:**

Siguiendo a la especificación de las aplicaciones para usuarios finales, el desarrollo de las aplicaciones de los usuarios finales involucra configuraciones de la meta data y construcción de reportes específicos.

Una vez que se ha cumplido con todos los pasos de la especificación y se tiene la posibilidad de trabajar con algunos datos de prueba, comienza el desarrollo de la aplicación.

**j. Implementación:**

La implementación representa la convergencia de la tecnología, los datos y las aplicaciones de usuarios finales accesible desde el escritorio del usuario del negocio. Existen factores extras que aseguran el correcto funcionamiento de todas estas piezas, entre ellos, se encuentran la capacitación, el soporte técnico, la comunicación, las estrategias de feedback. Todas estas tareas deben ser tenidas en cuenta antes de que cualquier usuario pueda tener acceso al DataWarehouse.

**k. Mantenimiento y crecimiento:**

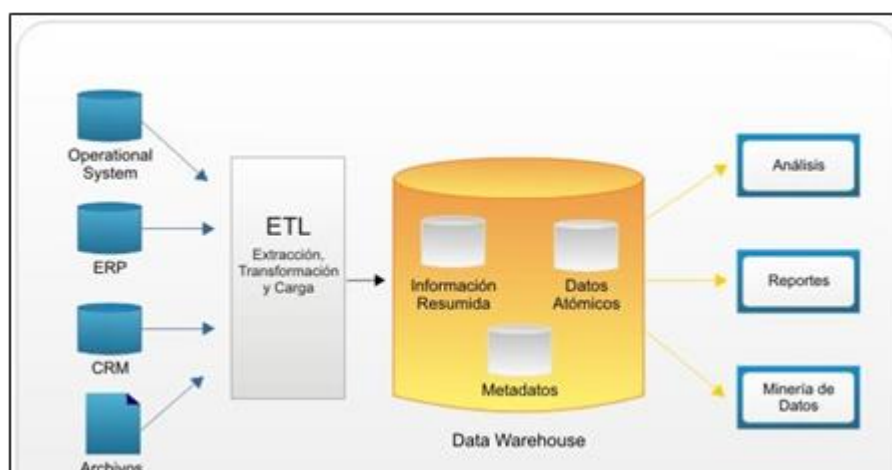
El DataWarehousing es un proceso (etapas bien definidas, con comienzo y fin, pero de naturaleza espiral), pues acompaña a la evolución de la organización durante toda su historia. Se necesita continuar con los relevamientos de forma constante para poder

seguir la evolución de las metas por conseguir. Según, afirma Kimball, “Si se ha utilizado el ciclo de vida dimensional del negocio, la Data Warehouse está preparado para evolucionar y crecer”. Al contrario de los sistemas tradicionales, los cambios en el desarrollo deben ser vistos como signos de éxito y no de falla. Es importante establecer las prioridades para poder manejar los nuevos requerimientos de los usuarios y de esa forma poder evolucionar y crecer.

### 1. Administración del proyecto

El gerenciamiento del proyecto asegura que las actividades del ciclo de vida dimensional del negocio se lleven en sincronizadas y en la mejor forma. Entre las actividades principales se encuentra el monitoreo del estado del proyecto, la comunicación entre los requerimientos del negocio y las restricciones de información para poder manejar correctamente las expectativas en ambos sentidos (Kimball, 2002).

### 2.2.2. Metodología de Bill Inmon



**Figura 2-5: Enfoque de la metodología de Bill Inmon**

**Fuente:** Building the Data Warehouse, William H. Inmon,

Bill Inmon ve la necesidad de transferir la información de los diferentes OLTP (sistemas transaccionales) de las organizaciones a un lugar centralizado donde los datos puedan ser utilizados para el análisis (sería el CIF o Corporate Information Factory). Insiste, además, en que ha de tener las siguientes características:

- ✓ Orientado a temas: los datos sobre la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.
- ✓ Integrado: la base de datos contiene los datos de todos los sistemas operacionales de la organización, y estos deben ser consistentes.
- ✓ No volátil: la información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en información de sólo lectura, y se mantiene para futuras consultas.
- ✓ Variante en el tiempo: los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones.

La información ha de estar a los máximos niveles de detalle. Las Data Warehouse departamentales o Datamarts son tratados como subconjuntos de este Data Warehouse corporativo, que son construidos para cubrir las necesidades individuales de análisis

de cada departamento, y siempre a partir de este Data Warehouse Central (del que también se pueden construir los ODS ( Operational Data Stores ) o similares).

El enfoque Inmon también se referencia normalmente como Top- down. Los datos son extraídos de los sistemas operacionales por los procesos ETL y cargados en las áreas de Stage, donde son validados y consolidados en el DW corporativo, y además existen los llamados metadatos que documentan de una forma clara y precisa el contenido del DW. Una vez realizado estas etapas, los procesos, de refresco de los Datamart departamentales obtienen la información de él, y con las consiguientes transformaciones, organizan los datos en las estructuras particulares requeridas por cada uno de ellos, refrescando su contenido.

Al tener este enfoque global es más difícil de desarrollar en un proyecto sencillo (pues estamos intentando abordar el “todo”, a partir del cual, luego, iremos al “detalle”).

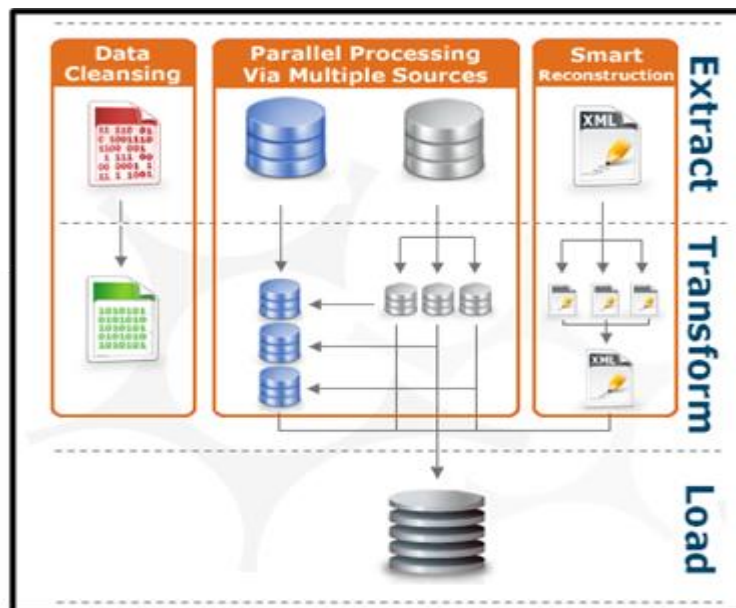
Al establecer una comparación entre las dos metodologías más importantes que son la metodología de Ralph Kimball (y su enfoque dimensional), y la metodología de Bill Inmon (y su enfoque empresarial Warehouse), en el caso particular de este proyecto de tesis, se analizará desde el punto de vista de la construcción de una datamart que es una parte de un dataWarehouse.

La metodología de Inmon es más apropiada para sistemas complejos, donde se quiere asegurar la perdurabilidad y consistencia de la información aunque cambien los procesos de negocio de la organización. Para proyectos pequeños donde se quiere asegurar la usabilidad de los usuarios que permita un desarrollo rápido e incremental de

la solución donde no se tiene claro el panorama global, el enfoque de Kimball es el más apropiado.

Para la construcción del Datamart en la DECEEE del INEI, usaremos la metodología propuesta por Ralph Kimball, dado que el ámbito y los recursos destinados al proyecto son de un área en específico, proporcionando un enfoque de menor a mayor, muy versátil y dejando la puerta abierta a una ampliación posterior dentro del ámbito de la institución. Kimball asegura la usabilidad de los usuarios con un sistema fácil de entender y con un rápido desarrollo.

#### 2.3.4. Proceso ETL



**Figura 2-6: Enfoque del proceso ETL**

**Fuente:** Espinera, Sheldon y asociados. La inteligencia de negocios

Los procesos ETL son procesos que leen los registros de las fuentes de datos, aplican las transformaciones necesarias para prepararlos y los cargan en el destino de datos.

El proceso ETL se divide en 5 subprocesos:

- ✓ Extracción: recupera los datos físicamente de las distintas fuentes de información. En este momento disponemos de los datos en bruto.
- ✓ Limpieza: recupera los datos en bruto y comprueba su calidad, elimina los duplicados y, cuando es posible, corrige los valores erróneos y completa los valores vacíos, es decir se transforman los datos -siempre que sea posible para reducir los errores de carga. En este momento disponemos de datos limpios y de alta calidad.
- ✓ Transformación: recupera los datos limpios, de alta calidad, los estructura y los suma en los distintos modelos de análisis. El resultado de este proceso es la obtención de datos limpios, consistentes, sumados y útiles.
- ✓ Integración: valida que los datos que cargamos en la Data Warehouse sean consistentes con las definiciones y formatos del Data Warehouse; los integra en los distintos modelos de las distintas áreas de negocio que hemos definido en el mismo. Estos procesos pueden ser complejos.
- ✓ Actualización: es el que nos permite añadir los nuevos datos a la DataWarehouse [12].

### **2.3.5. Integration Services SQL Server (SSIS)**

Salvador Ramos, en el documento “Integration Services, pieza fundamental en los proyectos de BI” define a Microsoft Integration Services como una plataforma para la creación de soluciones empresariales de transformaciones de datos e integración de datos. Integration Services sirve para resolver complejos problemas empresariales mediante la copia o descarga de archivos, el envío de mensajes de correo electrónico como respuesta a eventos, la actualización de almacenamientos de datos, la limpieza y minería de datos, y la administración de objetos y datos de SQL Server. Los paquetes pueden funcionar por separado o conjuntamente con otros paquetes para hacer frente a las complejas necesidades de la empresa. Integration Services puede extraer y transformar datos de muchos orígenes distintos, como archivos de datos XML, archivos planos y orígenes de datos relacionales, y, posteriormente, cargarlos en uno o varios destinos.

Integration Services contiene un variado conjunto de tareas y transformaciones integradas, herramientas para la creación de paquetes y el servicio Integration Services para ejecutar y administrar los paquetes. Las herramientas gráficas de Integration Services se pueden usar para crear soluciones sin escribir una sola línea de código. También, se puede programar el amplio modelo de objetos de Integración Services para crear paquetes mediante programación y codificar tareas personalizadas y otros objetos de paquete.

Integration Services no es solo una herramienta para ETL, sino que tiene una serie de tareas orientadas a la administración que serán utilizadas por los DBA’s.

Integration Services es una herramienta muy completa y con muchas funcionalidades, e incluso con ciertos riesgos si no hacemos un uso apropiado de ella (Ramos, 2011).

### **2.3.6. Analysis Services SQL Server (SSAS)**

La empresa consultec en el documento Microsoft SQL Server 2005 funcionalidades de business intelligence define que Microsoft SQL Server Analysis Services es una herramienta que permite a las organizaciones mejorar y optimizar la toma de decisiones empresariales dentro de la organización.

Incorpora funcionalidades OLAP (Online Analytical Processing) y de data Mining para las aplicaciones de inteligencia empresarial. Analysis Services soporta la creación y administración OLAP, que son unas bases de datos analíticas, permitiéndole diseñar, crear y gestionar estructuras multidimensionales que contienen datos agregados a partir de orígenes diversos, como bases de datos relacionales.

### **2.3.7. Sistema de soporte de decisiones (DSS)**

Es un sistema interactivo provisto de programas y herramientas, para ayudar a los responsables de la toma de decisiones a utilizar tecnologías de comunicaciones, datos, documentos, conocimiento y/o modelos para identificar y resolver problemas, para completar tareas del proceso de decisión, y para tomar decisiones.

Es una amplia área de análisis que sirve para que la gente examine datos a fin de tomar decisiones, ya sean grandes o pequeños, sobre los negocios de sus compañías.



### **2.3.8. Data Warehouse**

Es un repositorio de información extraída de otros sistemas corporativos, sean estos sistemas transaccionales, bases de datos departamentales, o Intranet de la compañía, a la que los hombres de negocios de la empresa pueden acceder.

Los sistemas Data Warehouse están orientados a procesos de consultas en contraposición con los procesos transaccionales, sus tablas pueden no estar normalizadas y se admite redundancia en los datos.

Mejor dicho, la Data Warehouse es un sistema, no un producto, en el que se almacenan datos. Es una técnica para consolidar y administrar datos de variadas fuentes con el propósito de responder preguntas de negocios y tomar decisiones, de una forma rápida.

Una Data Warehouse se vale de una base de datos relacional diseñada para el acceso rápido y análisis y no al proceso transaccional. La Data Warehouse separa la carga del análisis y normalmente contiene datos históricos derivados de datos transaccionales.

Sus principales características son:

- ✓ Orientado hacia información relevante de la organización
  
- ✓ Datos integrados
  
- ✓ Variable en el tiempo
  
- ✓ No Volátil

### **2.3.9. Datamart**

Es una DataWareHouse solo que más pequeña; en otras palabras, es una Data Warehouse orientada a algún tema. Los datamart suelen ser usados por un departamento o grupo de usuarios en una compañía, para un conjunto definido de tareas.

Un datamart se considera independiente, ya que recibe datos desde un Data Warehouse. Los Datamart aislados, es decir los que toman sus datos directamente desde sistemas transaccionales y no dependen de otros Data Warehouse, reciben el nombre de “Datamarts Independientes”.

### **2.3.10. Dimensiones**

Son los diferentes puntos de vista por los que queremos analizar la información. Las dimensiones contienen los diversos atributos que queremos analizar, además se estructuran en forma jerárquica, conforme a diferentes niveles de detalle.

Las tablas de dimensiones se construyen con todos los atributos que incluyen de una forma desnormalizada y con una clave que identifica el mínimo nivel de detalle.

Podemos distinguir varios tipos de dimensiones:

- ✓ Dimensiones normales: aquellas que agrupan diferentes atributos que están relacionados por el ámbito al que se refieren (todas las características de un cliente, los diferentes componentes de la dimensión Tiempo, etc).

- ✓ Dimensiones causales: son atributos que pueden causar cambios en los procesos de negocio (por ejemplo, la dimensión promoción en el proceso de negocio de ventas).
  
- ✓ Dimensiones heterogéneas: que agrupan conjuntos heterogéneos de atributos, que no están relacionados entre sí.
  
- ✓ Dimensiones roll-up: es un subconjunto de otra, necesarias para el caso en que tenemos tablas de hechos con diferente granularidad (ver la entrada anterior del blog).
  
- ✓ Dimensiones Junk: agrupa indicadores de baja cardinalidad como pueden ser flags o indicadores.
  
- ✓ Dimensiones role-playing: cuando una misma dimensión interviene en una tabla de hechos varias veces (por ejemplo, la fecha en una tabla de hechos donde se registran varias fechas referidas a conceptos diferentes), es necesario reutilizar la misma dimensión, pues no tiene sentido crear tantas dimensiones como usos se hagan de ella. En consecuencia, se definen las dimensiones role-playing. Podemos crear vistas sobre la tabla de la dimensión completa que nos permiten usarla varias veces o jugar con los alias de tabla. La misma dimensión juega un rol diferente según el sitio donde se utiliza.

- ✓ Dimensiones degeneradas: no tienen ningún atributo y por tanto, no tienen una tabla específica de dimensión. Incluyen para ellas un identificador en la tabla de hechos, que identifica completamente a la dimensión (por ejemplo, un pedido de ventas). Nos interesa tener determinada la transacción (para realizar datamining, por ejemplo), pero los datos interesantes de este elemento los tenemos repartidos en las diferentes dimensiones (cliente, producto, etc).
- ✓ Mini dimensiones o dimensiones Outtrigger: conjunto de atributos de una dimensión que se extraen de la tabla de dimensión principal, pues se suelen analizar de forma diferente. El típico ejemplo son los datos socio demográficos asociados a un cliente (que se utilizan, por ejemplo, para el datamining).

### **2.3.11. Métricas**

Son valores que recogen el proceso de una actividad o los resultados de la misma y son usados por los analistas en sus query's para medir la performance del comportamiento de un proceso o un objeto del negocio.

Las medidas candidatas son los datos numéricos, pero no cada atributo numérico en una medida candidata. Estas medidas proceden del resultado de la actividad de negocio.

- ✓ Métricas de realización de actividad: miden la realización de una actividad. Por ejemplo, la participación de una persona en un evento.
- ✓ Métricas de resultado de una actividad: recogen los resultados de una actividad. Por ejemplo, la cantidad de unidades vendidas.

### **2.3.12. Indicadores**

Es una variable dimensional unitaria, expresada como un cociente, que correlaciona dos variables cualesquiera. A través de este concepto es posible relacionar diversas variables presentes en los procesos de una empresa. Ejemplo: N° Ciclos/Hora, LtsComb, /Maquina.

La importancia de un indicador radica en la particularidad de informar al usuario el estado actual del sistema. Revelar el estado del sistema, nos permitirá tomar decisiones preventivas o correctivas de acuerdo a los resultados de las comparaciones entre el valor esperado y el valor obtenido del sistema.

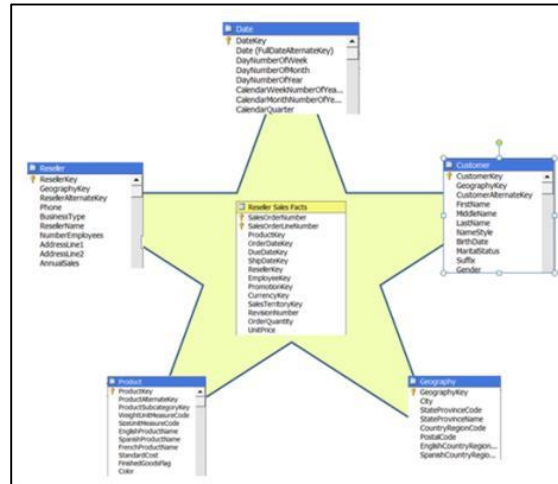
### **2.3.13. Fact Table**

Los hechos son los indicadores de negocio que dan sentido al análisis de las dimensiones. Las tablas de hechos incluyen los indicadores asociados a un proceso de negocio en concreto y las claves de las dimensiones que intervienen en dicho proceso, en el mínimo nivel de granularidad o detalle. Podemos tener varios tipos de tablas de hechos.

- ✓ Transaction fact tables: representan eventos que suceden en un determinado espacio-tiempo. Se caracterizan por analizar los datos con el máximo detalle. Reflejan las transacciones relacionadas con nuestros procesos de negocio (ventas, compras, inventario, contabilidad, etc).

- ✓ Factless fact tables: no tienen medidas y representan la ocurrencia de un evento determinado. Por ejemplo, la asistencia a un curso puede ser una tabla de hechos sin métricas asociadas.
  
- ✓ Periodic snapshot fact tables: son tablas de hecho usadas para recoger información de forma periódica a intervalos de tiempo regulares sobre un hecho. Nos permiten tomar una foto de la situación en un momento establecido (por ejemplo, al final del día, de una semana o de un mes). Un ejemplo puede ser la foto del stock de materiales, al final de cada día.
  
- ✓ Accumulating snapshot fact table: representan el ciclo de vida completo de una actividad o proceso, que tiene un principio y final. Suelen representar valores acumulados.
  
- ✓ Consolidated fact tables: tablas de hechos construidas como la acumulación, en un nivel de granularidad o detalle diferente, de las tablas de hechos de transacciones.

### 2.3.14. Esquema estrella



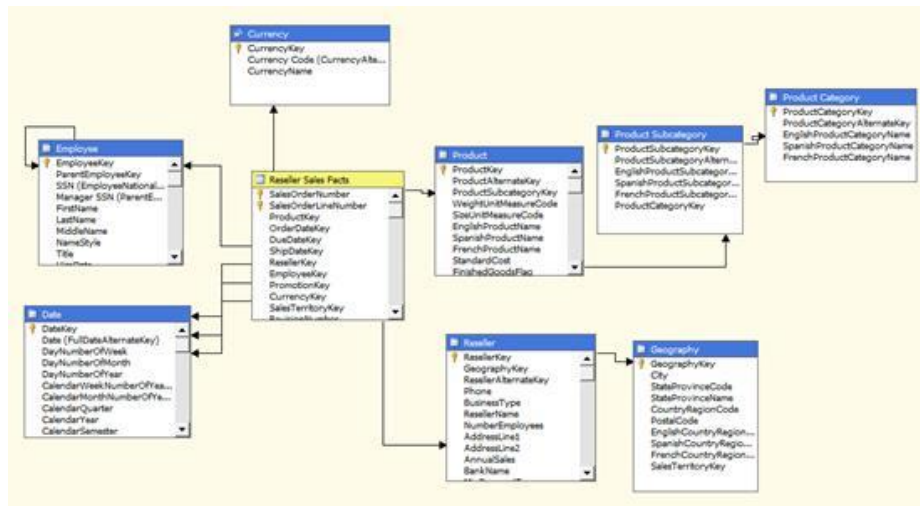
**Figura 2-7: Esquema estrella**

**Fuente:** Curso MS Business Intelligence, Salvador Ramos

Un modelo estrella es un modo de representar datos multidimensionales en una base de datos relacional. Las tablas dimensiones guardan información descriptiva acerca de sus miembros y sus relaciones. Mientras, que las tablas hechos almacenan datos de negocio. Los query's multidimensionales son hechos haciendo joins entre las tablas de hechos y de dimensiones

Los sistemas OLAP que utilizan este tipo de modelo de base de datos son llamados sistemas ROLAP

### 2.3.15. Esquema copo de nieve



**Figura 2-8: Esquema copo de nieve**

**Fuente:** Curso MS Business Intelligence, Salvador Ramos

Las bases de datos relacionales, a menudo emplean esquemas de copo de nieve para proporcionar los mejores tiempos de respuesta posibles, a las consultas complejas. Los esquemas de copos de nieve contienen una tabla de hechos centrales para el tema y numerosas tablas de dimensiones para la información descriptiva sobre las dimensiones del tema. La tabla hechos puede contener varios millones de filas. La información a la que se tiene acceso con más frecuencia se agrega previamente y se resume para mejorar aún más, el rendimiento.

Si bien el esquema copo de nieve se considera fundamentalmente una herramienta con la que el administrador de datos puede aumentar el rendimiento y simplificar el diseño del almacén de datos, también se utiliza para representar la información del almacén de datos, de modo tal, que tenga más sentido para los usuarios finales.



Suele aplicarse cuando diversos atributos caracterizan a los niveles más altos de jerarquía. Se puede elegir el normalizar solo algunas dimensiones y otras no, aumentando así, la complejidad del diseño y metadatos.

Ventajas:

- ✓ No se necesita el atributo de nivel.
  
- ✓ Fácil para definir jerarquías.
  
- ✓ Mejora considerablemente, el rendimiento cuando un gran número de requisitos solicita datos agregados.
  
- ✓ Los requerimientos escanean un reducido número de filas

Desventajas:

- ✓ Aumenta la complejidad de mantener la meta información debido al aumento del número de tablas.

### **2.3.16. Enfoque “Bottom-up”**

El enfoque Bottom-up, establece que el desarrollo de una solución de inteligencia de negocios debe partir por el desarrollo de los Datamarts y que el conjunto de estos se constituyen en la DataWarehouse

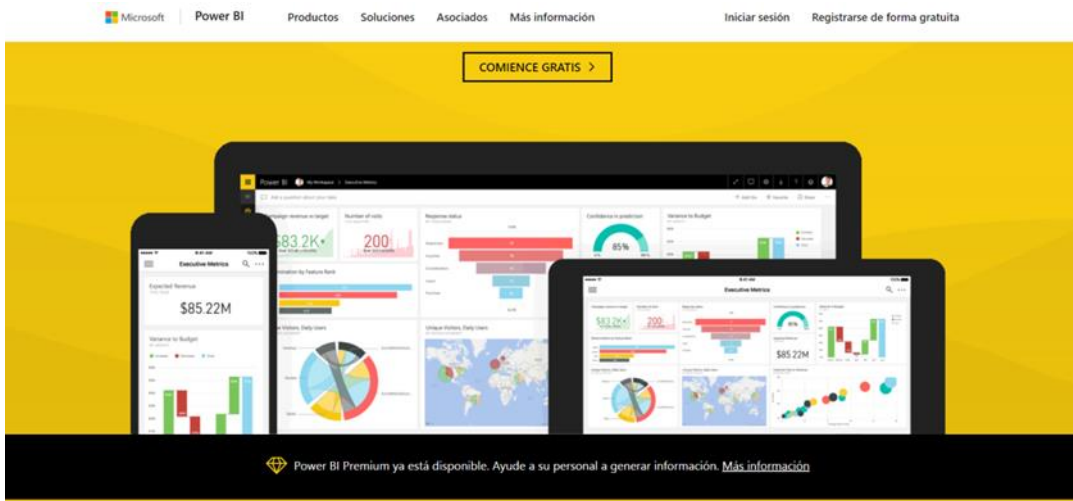
### **2.3.17. Enfoque “Top-down”**

El enfoque Top-down, establece que el desarrollo de una solución de inteligencia de negocios debe partir por el desarrollo del DataWarehouse que deriva en la Datamarts.

### **2.3.18. Power BI**

Power BI es un conjunto de aplicaciones de análisis de negocios que permite analizar datos y compartir información. Los paneles de Power BI ofrecen a los usuarios una vista de 360 grados con sus métricas más importantes en un mismo lugar. La información se actualiza en tiempo real y está disponible en todos sus dispositivos. Con un solo clic, los usuarios pueden explorar los datos subyacentes del panel mediante herramientas intuitivas que permiten obtener respuestas fácilmente.

La creación de un panel es una sencilla operación gracias a los miles de conexiones a conocidas aplicaciones empresariales, que se completan con paneles pregenerados para ayudarle a ponerse en marcha rápidamente. Asimismo, puede acceder a sus datos e informes desde cualquier lugar con las aplicaciones móviles de Power BI Mobile, que se actualizan automáticamente con los cambios que se realizan en los datos.

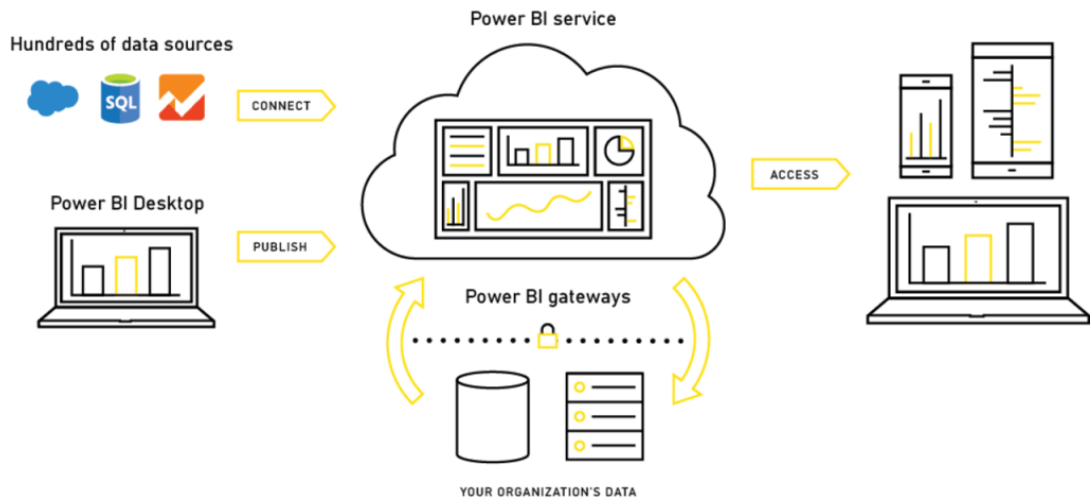


**Figura 2-9: Aplicación Power BI**

**Fuente:** <https://powerbi.microsoft.com/es-es/>

Si es un analista de datos que proporciona informes y análisis a su organización, Power BI le permite liberar todo su potencial creativo y alcanzar la máxima productividad en sus tareas. Power BI Desktop es una herramienta de mashup de datos y creación de informes que incluye numerosas características. Combine datos de bases de datos, archivos y servicios web diversos con herramientas visuales que ayudan a comprender y corregir problemas de formato y calidad de los datos automáticamente. Con más de 20 objetos visuales integrados y una dinámica comunidad de visualizaciones personalizadas, podrá crear informes espectaculares que comuniquen su mensaje con la máxima eficacia.

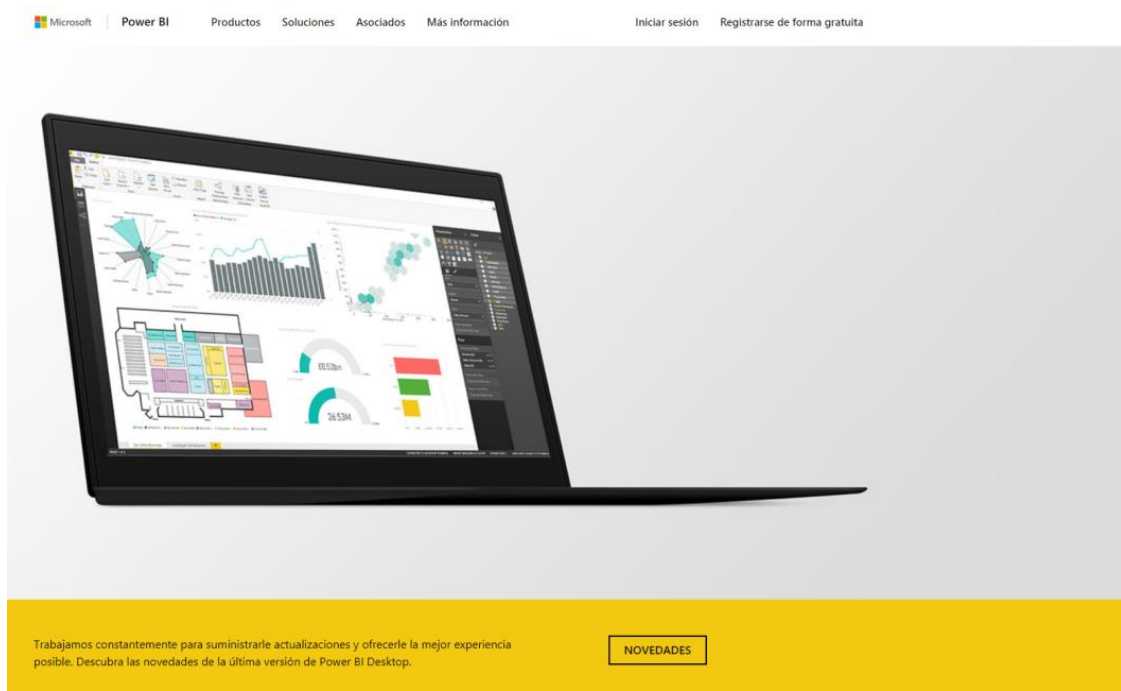
Con el servicio Power BI, publique informes con seguridad en su organización y configure la actualización de datos automática para que todo el mundo disponga de la información más reciente.



**Figura 2-10: Fuentes de Power BI**

**Fuente:** <https://powerbi.microsoft.com/es-es/what-is-power-bi/>

Power BI puede unificar todos los datos de su organización, ya sea en la nube o localmente. Con Power BI Gateways, puede conectar bases de datos SQL Server, modelos de Analysis Services y muchos otros orígenes de datos a los mismos paneles en Power BI. Si ya cuenta con portales o aplicaciones de creación de informes, inserte los informes y paneles de Power BI para disfrutar de una experiencia unificada.



**Figura 2-11: Power BI**

**Fuente:** <https://powerbi.microsoft.com/es-es/desktop>

## **CAPITULO III**

### **DESARROLLO DE LA METODOLOGIA**

#### **3.1. Analisis de la Metodología del proyecto**

El presente proyecto de investigación tiene como objetivo, implementar un datamart para mejorar la toma de decisiones en la dirección ejecutiva de censos y encuestas de empresas y establecimientos en el instituto nacional de estadística e informática, para ello se elegirá una metodología apropiada para el desarrollo y la construcción del datamart. Esto nos lleva a definir la metodología de implementación de la solución de Business Intelligence entre dos metodologías especializadas existentes. La de Bill Inmon y la de Ralph Kimball.

##### **3.1.1 Metodología de Bill Inmon**

La metodología de Bill Inmon recomienda construir un data warehouse que siga un enfoque de arriba hacia abajo. Esto consiste en que se empieza con la construcción de un gran data warehouse empresarial centralizada donde todos los datos disponibles de los sistemas de transacciones son consolidados en una colección de datos integrada, variante en el tiempo y no volátil colección de datos que soporta la toma de decisiones. Entonces los Data Marts son construidos de acuerdo a las necesidades analíticas de cada área de la empresa.

### 3.1.2 Metodología de Ralph Kimball

La metodología de Ralph Kimball recomienda que para construir un Data Warehouse que sigue el enfoque de abajo hacia arriba. Esto empieza con implementar Data Marts críticos que sirven a las necesidades analíticas de los departamentos. Luego se integran estos Data Marts para lograr la consistencia de datos a través de un bus de información. Kimball hace uso de modelo dimensional para hacer frente a las necesidades de los departamentos en diversas áreas de la empresa.

### 3.1.3 Elección de metodología

**Tabla 3-1: Comparación entre metodologías de Inmon Y Kimball**

	<b>INMON</b>	<b>KIMBA</b>
	Arriba hacia abajo	Abajo hacia arriba
<b>Estructura de la Arquitectura</b>	Un Data Warehouse que abarque toda la empresa y que se alimente de las bases de datos de los departamentos.	Modelamiento de un Data Mart por proceso de negocio. El Data Warehouse se obtiene con la unión de Data Marts de la empresa se logra

<b>Orientación de datos</b>	Asunto o manejo de datos	Orientado a procesos
<b>Herramientas</b>	Tradicionales (Diagramas Entidad Relación y Sistemas de interpretación de Data)	Modelado Dimensional; se aparta del tradicional modelado relacional.
<b>Accesibilidad al usuario final</b>	Baja	Alta
<b>Plazos (Dimensiones)</b>	Continuos y Discretos	Cambia lentamente
<b>Métodos (Dimensiones)</b>	Hora y Fecha	Llaves dimensionales



<b>Entrevistas</b>	Área de T.I.	Usuarios finales
<b>Lugar en la Organización</b>	Integral, parte de la Fábrica de información corporativa	Transformador y reten de data operacional
<b>Objetivo</b>	Entregar una solución basada en métodos probados.	Entregar una solución que facilite la labor de los usuarios finales, para consultar directamente los datos con una tasa de respuesta razonable

Fuente: <http://www.bi-bestpractices.com/view-articles/4768>

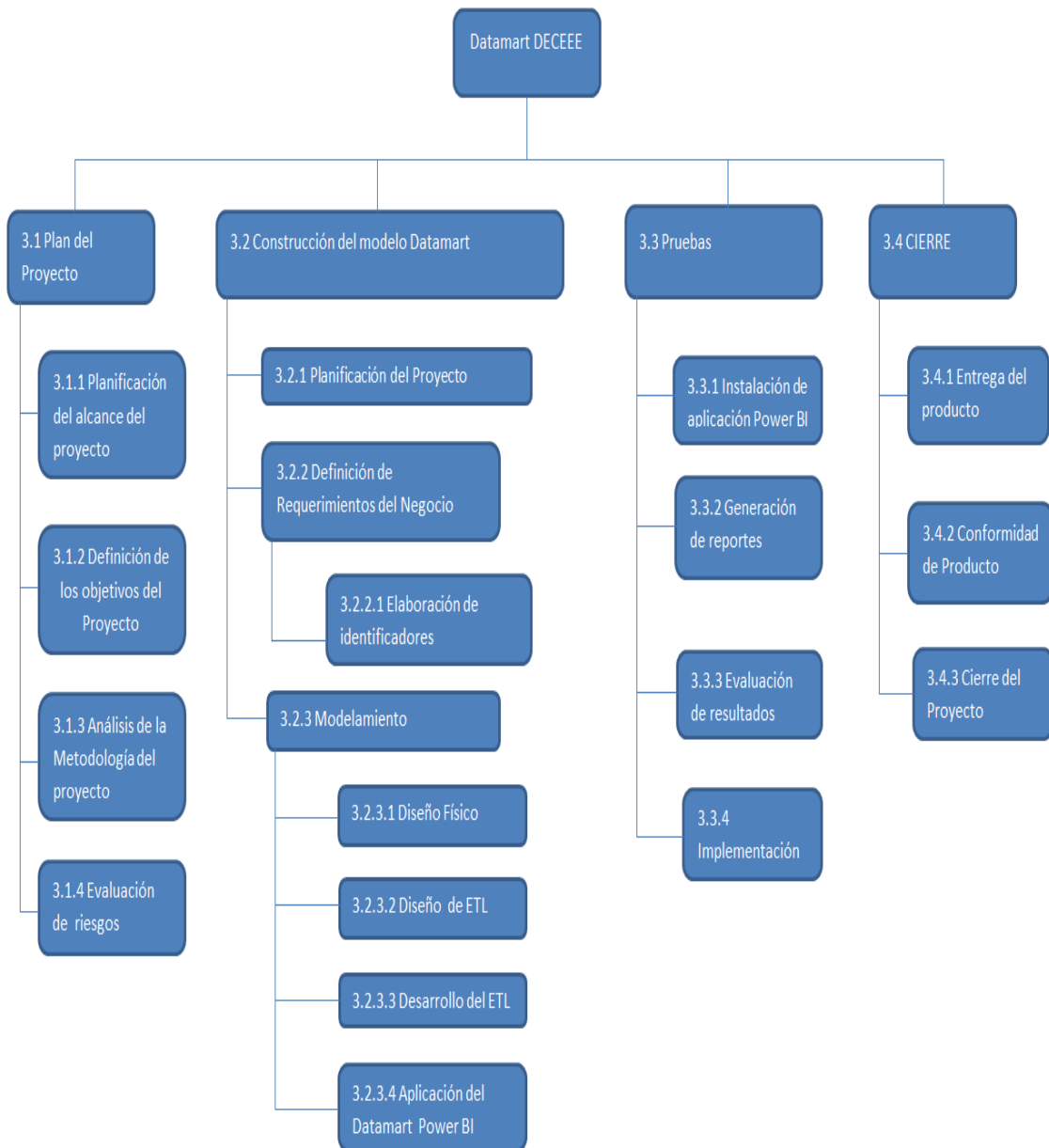
Al establecer una comparación entre las dos metodologías más importantes, que son la metodología de Ralph Kimball (y su enfoque dimensional), y la metodología de Bill Inmon (y su enfoque empresarial Warehouse), se llega a determinar que la metodología de Inmon es más apropiada para sistemas complejos, donde se quiere asegurar la perdurabilidad y consistencia de la información aunque cambien los procesos de

negocio de la organización. Mientras que para proyectos pequeños donde se quiere asegurar la usabilidad de los usuarios, que permita un desarrollo rápido e incremental de la solución donde no se tiene claro el panorama global, el enfoque de Kimball es el más apropiado.

Ante esto en el caso particular de este proyecto de tesis la metodología propuesta por Ralph Kimball es la más adecuada para la construcción del datamart en la DECEEE del INEI, dado que el ámbito y los recursos destinados al proyecto son de un área en específico, proporcionando un enfoque de menor a mayor, muy versátil y dejando la puerta abierta a una ampliación posterior dentro del ámbito de la Instituto Nacional de Estadística e Informática (INEI).

## 3.2 Construcción del modelo Datamart

### 3.2.1 Planificación del Proyecto



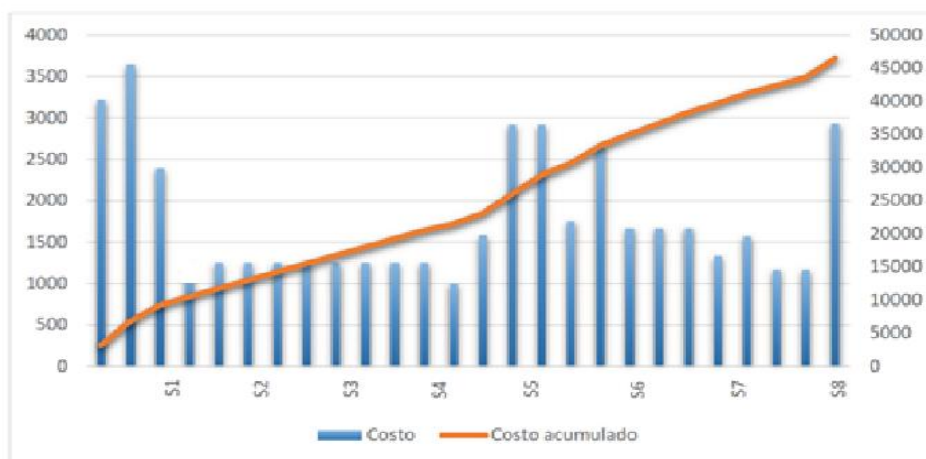
**Figura 3-1: EDT del Proyecto**

Fuente: Elaboración propia. Basado en PMBOK.

**Tabla 3-2: Recursos y Costos de la EDT**

N° Entr.	ENTREGABLES	RECURSOS	TIEMPO	COSTO	
<b>3.1</b>	<b>Plan del Proyecto</b>		<b>7</b>	<b>S/.</b>	<b>4.360,00</b>
3.1.1	Planificación del alcance del Proyecto	Gestor de Proyecto	2	S/.	1.360,00
3.1.2	Definición de los objetivos del Proyecto	Gestor de Proyecto	2	S/.	1.360,00
3.1.3	Análisis de la metodología del Proyecto	Analista Funcional	2	S/.	1.200,00
3.1.4	Evaluación de Riesgos	Analista Técnico	1	S/.	440,00
<b>3.2</b>	<b>Construcción del modelo Datamart</b>		<b>46</b>	<b>S/.</b>	<b>31.440,00</b>
3.2.1	Desarrollo del modelo físico	Diseñador del Modelo	2	S/.	1.440,00
3.2.2	Desarrollo del modelo lógico	Diseñador del Modelo	2	S/.	1.200,00
3.2.2.1	Elaboración de Identificadores	Analista Funcional	2	S/.	1.200,00
3.2.3	Modelamiento	Diseñador del Modelo	42	S/.	28.800,00
3.2.3.1	Diseño Físico	Analista Funcional	10	S/.	6.000,00
3.2.3.2	Diseño de ETL	Diseñador del Modelo	10	S/.	7.200,00
3.2.3.3	desarrollo del ETL	Diseñador del Modelo	20	S/.	14.400,00
3.2.3.4	Aplicación del Datamart Power BI	Analista Funcional	2	S/.	1.200,00
<b>3.3</b>	<b>Pruebas</b>		<b>14</b>	<b>S/.</b>	<b>8.640,00</b>
3.3.1	Instalación de Aplicación Power BI	Analista Funcional	3	S/.	1.800,00
3.3.2	Generación de Reportes	Analista Funcional	5	S/.	3.000,00
3.3.3	Evaluación de Resultados	Gestor de Proyecto	3	S/.	2.040,00
3.3.4	Implementación	Analista Funcional	3	S/.	1.800,00
<b>3.4</b>	<b>Cierre</b>		<b>3</b>	<b>S/.</b>	<b>2.040,00</b>
3.4.1	Entrega del Producto	Gestor de Proyecto	1	S/.	680,00
3.4.2	Conformidad del Producto	Gestor de Proyecto	1	S/.	680,00
3.4.3	Cierre del proyecto	Gestor de Proyecto	1	S/.	680,00
	Total de Días		70	S/.	46.480,00

Fuente: Elaboración propia según la EDT propuesto. Leyenda: Entregables (Entr.),  
Tiempo en días y Costo en soles.



**Figura 3-2: EDT del Proyecto**

**Fuente:** Elaboración propia, basado en flujo de costos. (Véase Anexo 1)

**Tabla 3-3: Costo Total del Proyecto**

Costo del Proyecto	S/.	46.480,00
Gastos Administrativos	S/.	1.000,00
IGV	S/.	8.546,40
Total	S/.	56.026,40

Fuente: Elaboración propia.

### **3.2.1.1. Planificación del proyecto y su alcance**

El alcance de este proyecto se basa en la construcción del Datamart que permitirá analizar los datos que se obtiene de las encuesta económicas anuales que se realizaron durante años anteriores en base a los sectores económicos, empresas, departamentos, donde se distribuyen las empresas económicas para que el director y jefes de la DECEEE interpreten la información y pueda tomar un decisiones oportunas.

La toma de decisiones en la DECEEE se basa en los reportes de las encuestas que emite el sistema transaccional por lo que el análisis es muy tedioso, demorado y existe una pérdida de tiempo que eso significa una gran pérdida para la institución, por lo que se toma decisiones erróneas sobre las metodologías y estrategias para las encuestas, la institución se ve en la necesidad de implementar una solución de inteligencia de negocios, en el área de la DECEEE, esta aplicación controlará los procesos para la toma de decisiones a nivel de dirección para elaborar metodologías y estrategias en la aplicación de los censos y encuestas llevadas a cabo por esta área y así tener mayor réditos para la institución.

### 3.2.2 Definición de los Requerimientos del Negocio

Para definir los requerimientos de la dirección, se tomara en cuenta los indicadores de la participación de la empresas en las encuestas realizadas tanto por los distritos, provincias, departamentos, sectores económicos, estratos económicos, logrando así definir las falencias que tienen las encuestas y ayudar a mejorar las la aplicación de metodologías y estrategias en los censos y encuestas llevadas a cabo.

A continuación se detallan los indicadores que se necesitan para la toma de decisiones en la DECEEE del INEI.

**Tabla 3-4: Identificador de Cantidad de empresas – Tiempo – Departamento – Participación – Empresa**

<b>Identificador:</b>	F01	<b>Nombre:</b>	Cantidad de empresas que participaron
<b>Tipo:</b>	Funcional	<b>Fecha:</b>	20/06/2017
<b>Prioridad:</b>	Alta	<b>Necesidad:</b>	Si
<b>Estabilidad:</b>	Normal	<b>Verificable:</b>	Si
<b>Descripción:</b>	La solución deberá proporcionar el departamento con la mayor cantidad de empresas que participaron y en que año se realizo la EEA, se verifica por la dimensión tiempo, departamento, participacion y empresa.		

**Fuente:** Elaboración propia

**Tabla 3-5: Identificador de Cantidad de Personal Ocupado – Tiempo – Sector  
Economico – Participación – Empresa**

<b>Identificador:</b>	F02	<b>Nombre:</b>	Cantidad de personal ocupado por sector economico
<b>Tipo:</b>	Funcional	<b>Fecha:</b>	20/06/2017
<b>Prioridad:</b>	Alta	<b>Necesidad:</b>	Si
<b>Estabilidad:</b>	Normal	<b>Verificable:</b>	Si
<b>Descripción:</b>	La solución deberá proporcionar el sector economico con la mayor cantidad de personal ocupado y en que año se realizo, se verifica por la dimensión tiempo, sector economico, participacion y empresa.		

**Fuente:** Elaboración propia

**Tabla 3-6: Identificador de Cantidad de Empresas Omisas – Tiempo –  
Departamento – Participación – Exoneracion – Empresa**

<b>Identificador:</b>	F03	<b>Nombre:</b>	Cantidad de empresas omisas por departamento
<b>Tipo:</b>	Funcional	<b>Fecha:</b>	20/06/2017
<b>Prioridad:</b>	Alta	<b>Necesidad:</b>	Si
<b>Estabilidad:</b>	Normal	<b>Verificable:</b>	Si
<b>Descripción:</b>	La solución deberá proporcionar el o los departamento con la mayor y menor cantidad de empresas omisas en la EEA y en que año, se verifica por la dimensión tiempo, departamento, participacion, exoneracion y empresa.		

**Fuente:** Elaboración propia

**Tabla 3-7: Identificador de Total de Ventas Acumuladas por Sector Económico – Tiempo – Sector Economico – Empresa**

<b>Identificador:</b>	F04	<b>Nombre:</b>	Total de ventas acumuladas por sector economico
<b>Tipo:</b>	Funcional	<b>Fecha:</b>	20/06/2017
<b>Prioridad:</b>	Alta	<b>Necesidad:</b>	Si
<b>Estabilidad:</b>	Normal	<b>Verificable:</b>	Si
<b>Descripción:</b>	La solución deberá proporcionar el o los sectores economicos con la mayor y menor cantidad de de eventas en la EEA y en que año, se verifica por la dimensión tiempo, sector economico, y empresa.		

**Fuente:** Elaboración propia

**Tabla 3-8: Identificador de Total de Ventas Acumuladas por Provincias – Tiempo – Departamento – Provincia – Empresa**

<b>Identificador:</b>	F05	<b>Nombre:</b>	Total de ventas acumulados por provincias
<b>Tipo:</b>	Funcional	<b>Fecha:</b>	20/06/2017
<b>Prioridad:</b>	Alta	<b>Necesidad:</b>	Si
<b>Estabilidad:</b>	Normal	<b>Verificable:</b>	Si
<b>Descripción:</b>	La solución deberá proporcionar la provincia con la mayor cantidad de ventas acumuladas en la EEA y en que año, se verifica por la dimensión tiempo, departamento,provincia, y empresa.		

**Fuente:** Elaboración propia



**Tabla 3-9: Identificador de Cantidad de Empresas por Sector economico – Tiempo – Sector Economico – Participacion – Empresa**

<b>Identificador:</b>	F06	<b>Nombre:</b>	Cantidad de empresas por sector economico
<b>Tipo:</b>	Funcional	<b>Fecha:</b>	20/06/2017
<b>Prioridad:</b>	Alta	<b>Necesidad:</b>	Si
<b>Estabilidad:</b>	Normal	<b>Verificable:</b>	Si
<b>Descripción:</b>	La solución deberá proporcionar el sector economico con las mayor y menor cantidad de empresas en la EEA y en que año, se verifica por la dimensión tiempo, sector economico, participacion, y empresa.		

**Fuente:** Elaboración propia

**Tabla 3-10: Identificador de Cantidad de Empresas por Departamento – Tiempo – Departamento – Participacion – Empresa**

<b>Identificador:</b>	F07	<b>Nombre:</b>	Cantidad de empresas por departamento
<b>Tipo:</b>	Funcional	<b>Fecha:</b>	20/06/2017
<b>Prioridad:</b>	Alta	<b>Necesidad:</b>	Si
<b>Estabilidad:</b>	Normal	<b>Verificable:</b>	Si
<b>Descripción:</b>	La solución deberá proporcionar el departamento con las mayor y menor cantidad de empresas en la EEA y en que año, se verifica por la dimensión tiempo, departamento, participacion, y empresa.		

**Fuente:** Elaboración propia

**Tabla 3-11: Identificador de Cantidad de Empresas por Organización jurídica –  
Tiempo – Organización Jurídica – Participacion – Empresa**

<b>Identificador:</b>	F08	<b>Nombre:</b>	Cantidad total de empresas por organización jurídica
<b>Tipo:</b>	Funcional	<b>Fecha:</b>	20/06/2017
<b>Prioridad:</b>	Alta	<b>Necesidad:</b>	Si
<b>Estabilidad:</b>	Normal	<b>Verificable:</b>	Si
<b>Descripción:</b>	La solución deberá proporcionar el tipo de organización jurídica con las mayor y menor cantidad de empresas en la EEA y en que año, se verifica por la dimensión tiempo, organización jurídica, participacion, y empresa.		

**Fuente:** Elaboración propia

**Tabla 3-12: Identificador de Cantidad de Empresas por Estrato Economico –  
Tiempo – Estrato Economico – Participacion – Empresa**

<b>Identificador:</b>	F09	<b>Nombre:</b>	Cantidad total de empresas por estrato economico
<b>Tipo:</b>	Funcional	<b>Fecha:</b>	20/06/2017
<b>Prioridad:</b>	Alta	<b>Necesidad:</b>	Si
<b>Estabilidad:</b>	Normal	<b>Verificable:</b>	Si
<b>Descripción:</b>	La solución deberá proporcionar los estratos economicos con las mayor y menor cantidad de empresas en la EEA y en que año, se verifica por la dimensión tiempo, estrato economico, participacion, y empresa.		

**Fuente:** Elaboración propia

### 3.2.3 Modelado Dimensional

Se comienza con una matriz donde se determina la dimensionalidad de cada indicador para luego especificar los diferentes grados de detalle dentro de cada concepto de la encuesta, así tenemos:

**Tabla 3-13: Participación de empresas**

<b>Empresa</b>	<b>Tiempo</b>	<b>Departamento</b>	<b>Participacion</b>
----------------	---------------	---------------------	----------------------

**Fuente:** Elaboración propia

**Tabla 3-14: Personal ocupado**

<b>Empresa</b>	<b>Tiempo</b>	<b>Sector Economico</b>	<b>Participacion</b>	<b>Cantidad de Personal Ocupado</b>
----------------	---------------	-------------------------	----------------------	-------------------------------------

**Fuente:** Elaboración propia

**Tabla 3-15: Empresas omisas**

<b>Empresa</b>	<b>Tiempo</b>	<b>Departamento</b>	<b>Participacion</b>	<b>Exoneracion</b>
----------------	---------------	---------------------	----------------------	--------------------

**Elaborado por:** Investigador

**Tabla 3-16: Total de ventas**

<b>Empresa</b>	<b>Tiempo</b>	<b>Sector Economico</b>	<b>Participacion</b>	<b>Total de Ventas por Sector Economico</b>
----------------	---------------	-------------------------	----------------------	---

**Fuente:** Elaboración propia

**Tabla 3-17: Identificador de Cantidad de Empresas por**

<b>Empresa</b>	<b>Tiempo</b>	<b>Departamento</b>	<b>Provincia</b>	<b>Total de Ventas por Provincia</b>
----------------	---------------	---------------------	------------------	--

**Fuente:** Elaboración propia

**Tabla 3-18: Empresas por Sector Economico**

<b>Empresa</b>	<b>Tiempo</b>	<b>Sector Economico</b>	<b>Participacion</b>	<b>Cantidad de empresas por Sector Economico</b>
----------------	---------------	-------------------------	----------------------	--

**Fuente:** Elaboración propia

**Tabla 3-19: Empresas por departamento**

<b>Empresa</b>	<b>Tiempo</b>	<b>Departamento</b>	<b>Participacion</b>	<b>Cantidad de empresas por Departamento</b>
----------------	---------------	---------------------	----------------------	--

**Fuente:** Elaboración propia

**Tabla 3-20: Empresas por organización jurídica**

<b>Empresa</b>	<b>Tiempo</b>	<b>Organización Juridica</b>	<b>Participacion</b>	<b>Cantidad de empresas por Organización Juridica</b>
----------------	---------------	----------------------------------	----------------------	---

**Fuente:** Elaboración propia

**Tabla 3-21: Empresas por estrato económico**

<b>Empresa</b>	<b>Tiempo</b>	<b>estrato Economico</b>	<b>Participacion</b>	<b>Cantidad de empresas por Estrato Economico</b>
----------------	---------------	------------------------------	----------------------	---

**Fuente:** Elaboración propia

**a. Definición del proceso del Negocio.**

En este primer paso seleccionamos los procesos de negocio del modelo de la Dirección. En el caso de estudio se quiere analizar las cantidades de empresas en la encuesta, cantidad de personal ocupado así como el monto de ventas por sectores a través de los años, cantidad de empresas omisas, cantidad de empresa por organización jurídica, cantidad de empresas por estrato económico, las mismas que deberán ser analizadas por las dimensiones de tiempo, departamento, sector económico, participación, exoneración, organización jurídica, estrato económico y empresa.

**b. Definición del grano o nivel de detalle para el modelo de Datamart.**

Una vez que se ha definido el proceso de negocio, la siguiente tarea será la definición de la granularidad, o lo que es lo mismo, hasta qué nivel de detalle se quiere alcanzar en el modelo del Datamart y más concretamente en la tabla de hechos.

El grano más apropiado para el propósito de este proyecto consistiría en asignar una fila por cada empresa, fecha, departamento, provincia, distrito, sector económico, CIU, organización jurídica, estrato económico que agrupa el sistema. De manera que se pueda estructurar la información en torno a los datos que se puedan obtener de las encuestas, entre los que pueden encontrarse los siguientes:

- ✓ Fecha de Encuesta
- ✓ Código de Empresa
- ✓ Código de Departamento
- ✓ Código de la Provincia
- ✓ Código del Distrito

- ✓ Código del Sector Económico
- ✓ Código de División económica
- ✓ Código de Organización Jurídica
  - ✓ Código de Estrato Económico
  - ✓ Total de empresas
  - ✓ Total de ventas
  - ✓ Total de Personal Ocupado

### **c. Elección de las Dimensiones**

Una vez que se ha escogido el grano, analizamos qué dimensiones tiene asociadas. En este proyecto las dimensiones más apropiadas son: el tiempo, el sector económico, estrato económico, exoneración, CIIU, distrito, provincia, departamento, organización jurídica, participación y la empresa.

Las tablas de dimensiones son catálogos de información complementaria necesaria para la presentación de los datos a los usuarios. Es decir, la información general complementaria a cada uno de los registros de la tabla de hechos.

### **d. Identificación de los hechos que poblarán cada fila de la tabla de hechos**

Las medidas más útiles para incluir en una tabla de hechos son las aditivas, es decir, aquellas medidas que se pueden sumar, como por ejemplo la cantidad de empresas que participaron, la cantidad de ventas o la cantidad de personal ocupado en un sector económico. Todas son medidas numéricas que pueden calcularse con la suma de varias cantidades de la tabla de hechos.

A menudo se describen los hechos como un valor permanente, principalmente como una guía para el diseñador para ayudar a descifrar qué es un hecho frente a un atributo de dimensión. Por lo general, las tablas de hechos constituyen el 90 por ciento o más del espacio total consumido por una base de datos dimensional.

En este proyecto de tesis la tabla de hechos será la FACT\_RESUTADOS\_EEA

**e. Detalle de las tablas de dimensión.**

A continuación vamos a definir de manera clara y concisa cada una de las dimensiones que se han identificado para el proyecto, describiendo todos los atributos que componen las mismas.

**f. Dimensión de Tiempo:**

Dado que el tiempo es una dimensión presente en prácticamente cualquier Datamart, merece una atención especial. Al diseñar la dimensión, hay que prestar especial cuidado, ya que puede hacerse de varias maneras y no todas son igualmente eficientes.

Se define como clave primaria un campo de tipo date, La dimensión de tiempo va a ser relacionada con la tabla de hechos conservando su atributo del tipo date además de ello en esta dimensión se definirán los campos como es el año, el semestre, trimestre, semana, mes, día, etc. para que al momento de realizar los reportes el gerente pueda analizar de acuerdo a su criterio e interpretar para tomar una decisión oportuna al momento de visualizar los datos de las empresas obtenidas en todos los años analizadas por sector económico, departamento, estrato económico, etc.

**g. Dimensión de Empresa:**

La dimensión de la empresa guardará los campos como ruc de la empresa, ID\_ENCUESTA siendo estos dos campos las claves primarias de la dimensión, de modo que se identifique a cada tupla almacenada. Esta dimensión en general guardará algunas descripciones propias de la empresa

**h. Dimensión de Sector Económico:**

La dimensión Sector económico guardará los campos como código del sector, nombre del sector, siendo su código de sector su clave primaria de modo que identifique a cada tupla ID\_SECTOR del sector. una empresa solo puede pertenecer a un sector económico.

**i. Dimensión de Departamento:**

La dimensión Departamento guardará los campos como código del departamento, nombre del departamento, siendo su código de departamento su clave primaria de modo que identifique a cada tupla ID\_DEPARTAMENTO del departamento. Una empresa solo puede pertenecer a un departamento con la indicación de que por una empresa independientemente de cuantos establecimientos tenga, solo se considerara su establecimiento principal.

**j. Dimensión de Provincia:**

La dimensión Provincia guardará los campos como código del Provincia, nombre de la provincia, siendo su código de provincia su clave primaria de modo que identifique a cada tupla ID\_PROVINCIA de la provincia. Esta dimensión también llevara una clave foránea que procede de la dimensión departamento.



**k. Dimensión de Distrito:**

La dimensión Distrito guardará los campos como código del Distrito, nombre del distrito, siendo su código de distrito su clave primaria de modo que identifique a cada tupla ID\_DISTRITO del distrito. Esta dimensión también llevara una clave foránea que procede de la dimensión provincia.

**l. Dimensión de Participación:**

La dimensión Participación guardará los campos como código de participación, descripción, siendo su código de participación su clave primaria de modo que identifique a cada tupla ID\_PARTICIPACION de participación.

**m. Dimensión de Exoneración:**

La dimensión Exoneración guardará los campos como código de Exoneración, descripción, siendo su código de exoneración su clave primaria de modo que identifique a cada tupla ID\_EXONERACION de exoneración.

**n. Dimensión de Organización Jurídica:**

La dimensión Organización Jurídica guardará los campos como código del Organización Jurídica, nombre de la organización jurídica, siendo su código de organización jurídica su clave primaria de modo que identifique a cada tupla ID\_ORG\_JURIDICA de la organización jurídica.

**o. Dimensión de Clasificación Internacional Industrial Uniforme (CIIU):**

La dimensión CIIU guardará los campos como código de CIIU, código de división económica, código de grupo económico, código de clase económica, nombre de división económica, nombre del distrito, siendo su código de CIIU su clave primaria de modo que identifique a cada tupla ID\_CIIU del CIIU.

**p. Dimensión de Estrato Económico:**

La dimensión Estrato Económico guardará los campos como código del estrato económico, ID\_ANIO\_EJERCICIO\_FISCAL siendo estos dos campos las claves primarias de la dimensión, de modo que se identifique a cada tupla almacenada.

**q. Diseño Físico:**

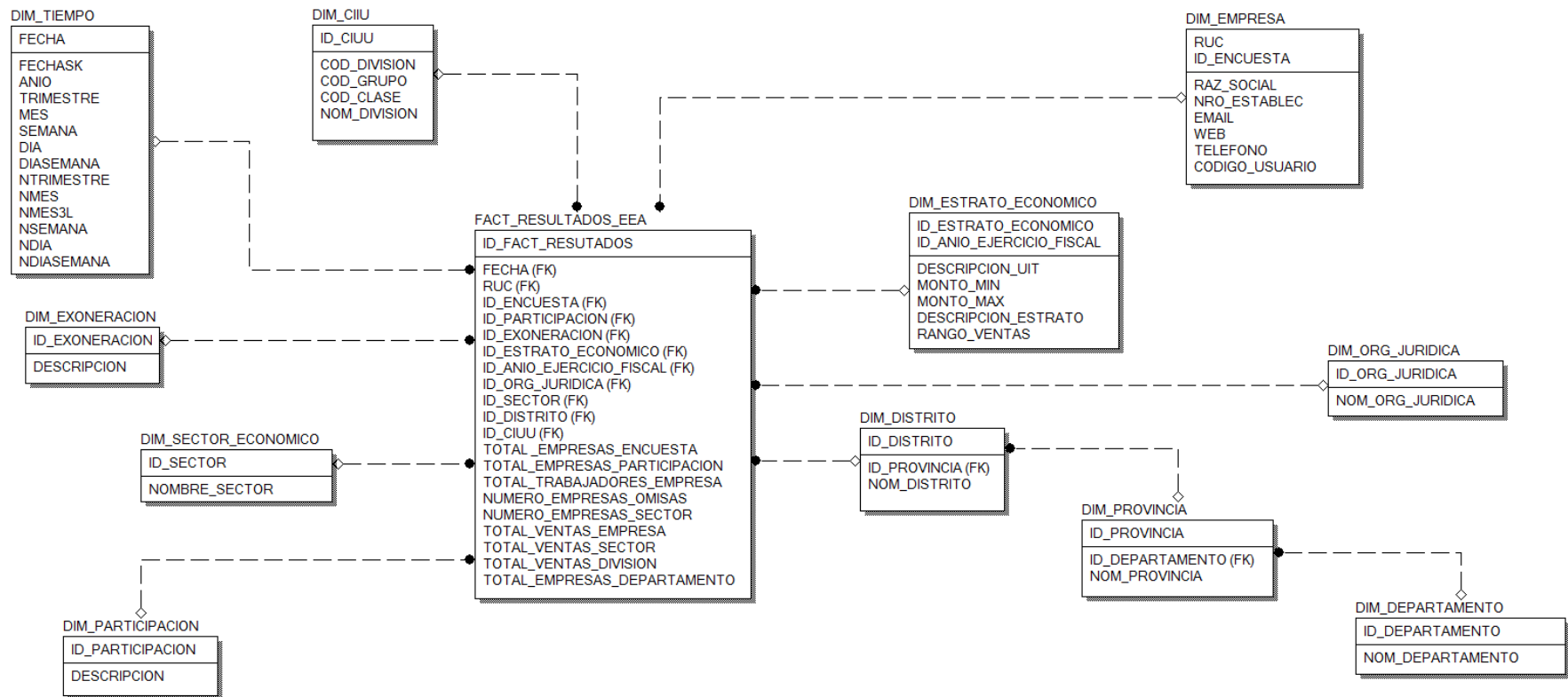
Se centra en la selección de las estructuras necesarias para soportar el diseño lógico. Un elemento principal de este proceso es la definición de estándares del entorno de la base de datos. La indexación y las estrategias de aprisionamiento que se determinan en esta etapa. El modo lógico se presenta en la figura 3-1.

**r. Modelo Entidad Relación:**

En esta etapa se realiza un modelado de las dimensiones y los hechos que conforman el Datamart para analizar y explotar los datos de las empresas recopilados por la EEA.

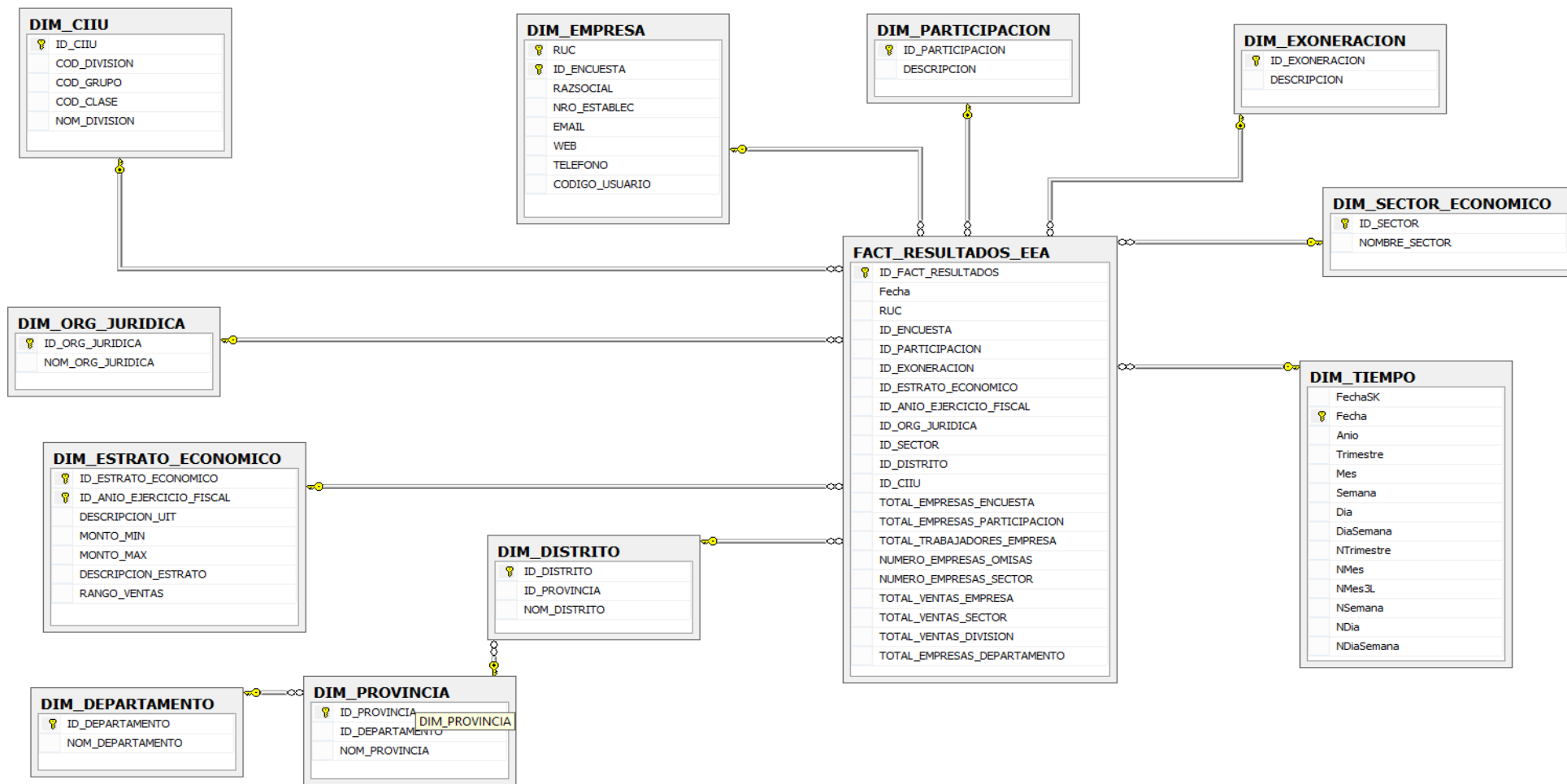
**s. Diseño y Desarrollo de la presentación de datos:**

Tiene como principales actividades la extracción, transformación y carga (ETL). Estas actividades son altamente críticas ya que tienen que ver con la materia prima del DataWarehouse o Datamart que son los datos del BI.



**Figura 3-3: Modelo Físico del DATAMART**

Fuente: Elaboración Propia.



**Figura 3-4: Modelo Dimensional del DATAMART**

Fuente: Elaboración propio

#### **t. Extracción**

Es el primer paso de obtener información hacia el ambiente del Datamart para explotarlos de la mejor manera. Consiste en extraer los datos desde los sistemas de origen hacia el sistema en que se realizara el Datamart.

La extracción convierte los datos a un formato preparado para iniciar el proceso de transformación. Se extrae la información del sistema transaccional, la base de datos de la empresa está en Sql Anywhere y de esta pasa al Datamart que es una parte del datawarehouse para explotarlos de la mejor manera y los datos que se visualicen sean reales y concisos.

#### **u. Transformación**

Luego de que la información fue extraída de la base de datos original de la empresa mediante el ODBC, se pueden realizar diferentes pasos de transformación, como ser: limpieza de la información, botar a la basura lo que consideramos innecesario o sea los datos deben ser lo más precisos posibles para que al momento de la toma de decisiones se realice de manera eficiente, seleccionar campos específicos que consideremos necesarios para el análisis del BI, realizar combinaciones fuentes de datos, etc. Se limpia y depura datos nulos, en blanco o inconsistencias del sistema transaccional y de esta pasa al datamart.

#### **v. Carga**

Al final del proceso de transformación, los datos están en forma para ser cargados. Dependiendo de los requerimientos de la organización, este proceso puede abarcar una amplia variedad de acciones diferentes.

En primer lugar se realiza la conexión de la base de datos del modelo transaccional de la empresa mediante el ODBC creando un origen de datos que se enlace con el Datamart y los datos se puedan extraer de manera rápida.

Para la extracción transformación y carga de los datos de la base de datos original de la empresa se realiza el siguiente código para la creación de las tablas de hechos y dimensiones y definir de manera óptima el modelo del Datamart para que los datos que se analicen sean en tiempo real y eficientes a la hora de tomar una decisión.

A continuación se procede a crear las diferentes dimensiones y hechos que poblaran el Datamart de la empresa para luego ser explotados.

#### **w. Tiempo de Refresco (Políticas de Actualización de datos del Datamart)**

La política de actualización de los datos del Datamart se realiza de acuerdo a las especificaciones de la institución, porque cada cierto periodo ingresan gran cantidad de información de la EEA que se ejecuta y el director y altos jefes de la dirección, necesitan analizar la información de acuerdo a las encuestas realizadas, para luego tomar las mejores decisiones, además con este proceso el director podrá verificar si cumplieron las metas propuestas por la DECEEE.

Con la lógica asociativa se logra que el Datamart se refresque automáticamente sin tener que detener los procesos internos que realizan en la DECEEE, es una ventaja muy importante porque nos ayuda a analizar los datos de manera rápida y eficiente y así las decisiones que tome el director y los altos jefes de la dirección den un giro importante.

## Creación de las dimensiones en el Datamart

Script para la creacion de las dimensiones

-----CREANDO LAS TABLAS DIMENSION-----

```
CREATE TABLE DIM_DEPARTAMENTO
(
    ID_DEPARTAMENTO    char(2) NOT NULL ,
    NOM_DEPARTAMENTO   varchar(200) NULL ,
)
go
ALTER TABLE DIM_DEPARTAMENTO
    ADD CONSTRAINT XPKDIM_DEPARTAMENTO PRIMARY KEY
    CLUSTERED (ID_DEPARTAMENTO ASC)
Go
```

```
-----
CREATE TABLE DIM_PROVINCIA
(
    ID_PROVINCIA      char(4) NOT NULL ,
    ID_DEPARTAMENTO    char(2) NOT NULL ,
    NOM_PROVINCIA     varchar(200) NULL ,
)
Go

ALTER TABLE DIM_PROVINCIA
    ADD CONSTRAINT XPKDIM_PROVINCIA PRIMARY KEY CLUSTERED
    (ID_PROVINCIA ASC)
Go
```

```
ALTER TABLE DIM_PROVINCIA
    ADD CONSTRAINT R_01 FOREIGN KEY (ID_DEPARTAMENTO)
    REFERENCES DIM_DEPARTAMENTO(ID_DEPARTAMENTO)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION
Go
```

```
-----
CREATE TABLE DIM_DISTRITO
(
    ID_DISTRITO       char(6) NOT NULL ,
    ID_PROVINCIA      char(4) NOT NULL ,
    NOM_DISTRITO      varchar(200) NULL
)
go
```

```
ALTER TABLE DIM_DISTRITO
    ADD CONSTRAINT XPKDIM_DISTRITO PRIMARY KEY CLUSTERED
(ID_DISTRITO ASC)
Go
```

```
ALTER TABLE DIM_DISTRITO
    ADD CONSTRAINT R_02 FOREIGN KEY (ID_PROVINCIA) REFERENCES
DIM_PROVINCIA(ID_PROVINCIA)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION
go
```

---

```
CREATE TABLE DIM_CIIU
(
    ID_CIIU        varchar(4) NOT NULL ,
    COD_DIVISION   varchar(2) NOT NULL ,
    COD_GRUPO      varchar(1) NOT NULL ,
    COD_CLASE      varchar(1) NOT NULL ,
    NOM_DIVISION   varchar(200) NULL
)
go
```

```
ALTER TABLE DIM_CIIU
    ADD CONSTRAINT XPKDIM_CIIU PRIMARY KEY CLUSTERED
(ID_CIIU ASC)
go
```

---

```
CREATE TABLE DIM_EMPRESA
(
    RUC            char(11) NOT NULL ,
    ID_ENCUESTA    char(2) NOT NULL,
    RAZSOCIAL      varchar(200) NULL ,
    NRO_ESTABLEC   char(3) NULL ,
    EMAIL          varchar(255) NULL ,
    WEB            varchar(255) NULL ,
    TELEFONO       varchar(50) NULL ,
    CODIGO_USUARIO varchar(50) NULL
)
go
```

```
ALTER TABLE DIM_EMPRESA
    ADD CONSTRAINT XPKDIM_EMPRESA PRIMARY KEY CLUSTERED
(RUC,ID_ENCUESTA ASC)
go
```

---



```

CREATE TABLE DIM_ORG_JURIDICA
(
    ID_ORG_JURIDICA    char(2) NOT NULL ,
    NOM_ORG_JURIDICA  varchar(50) NULL
)
go
ALTER TABLE DIM_ORG_JURIDICA
    ADD CONSTRAINT XPKDIM_ORG_JURIDICA PRIMARY KEY
    CLUSTERED (ID_ORG_JURIDICA ASC)
go

```

---

```

create table DIM_TIEMPO
(
    FechaSK int not null,
    Fecha date not null,
    Anio smallint not null,
    Trimestre smallint not null,
    Mes smallint not null,
    Semana smallint not null,
    Dia smallint not null,
    DiaSemana smallint not null,
    NTrimestre char(7) not null,
    NMes char(15) not null,
    NMes3L char(3) not null,
    NSemana char(10) not null,
    NDia char(6) not null,
    NDiaSemana char(10) not null
    constraint PK_DIM_TIEMPO PRIMARY KEY CLUSTERED
(
    Fecha asc
)
)
go

```

---

```

CREATE TABLE DIM_SECTOR_ECONOMICO
(
    ID_SECTOR          char(2) NOT NULL ,
    NOMBRE_SECTOR      varchar(200) NULL
)
go
ALTER TABLE DIM_SECTOR_ECONOMICO
    ADD CONSTRAINT XPKSECTOR_ECONOMICO PRIMARY KEY
    CLUSTERED (ID_SECTOR ASC)
go

```

---

```

CREATE TABLE DIM_PARTICIPACION
(
    ID_PARTICIPACION  char(2) NOT NULL ,
    DESCRIPCION      varchar(200) NULL
)
go
ALTER TABLE DIM_PARTICIPACION
    ADD CONSTRAINT XPKPARTICIPACION PRIMARY KEY CLUSTERED
(ID_PARTICIPACION ASC)
go
-----
CREATE TABLE DIM_EXONERACION
(
    ID_EXONERACION  char(2) NOT NULL ,
    DESCRIPCION     varchar(200) NULL
)
go
ALTER TABLE DIM_EXONERACION
    ADD CONSTRAINT XPKEXONERACION PRIMARY KEY CLUSTERED
(ID_EXONERACION ASC)
go
-----
CREATE TABLE DIM ESTRATO_ECONOMICO
(
    ID ESTRATO_ECONOMICO char(2) NOT NULL,
    ID ANIO_EJERCICIO_FISCAL char(2) NOT NULL,
    DESCRIPCION UIT varchar(200),
    MONTO_MIN FLOAT,
    MONTO_MAX FLOAT,
    DESCRIPCION ESTRATO varchar(200),
    RANGO_VENTAS varchar(200)
)
go
ALTER TABLE DIM ESTRATO_ECONOMICO
    ADD CONSTRAINT XPKESTRATO_ECONOMICO PRIMARY KEY
CLUSTERED (ID ESTRATO_ECONOMICO, ID ANIO_EJERCICIO_FISCAL ASC)
go

```

## Creación de la tabla de hechos FACT\_RESULTADOS\_EEA

----- CREANDO LAS FACT-TABLE-----

```
CREATE TABLE FACT_RESULTADOS_EEA
(
    ID_FACT_RESULTADOS integer NOT NULL ,
    Fecha date,
    RUC char(11) NOT NULL,
    ID_ENCUESTA char(2) NOT NULL,
    ID_PARTICIPACION char(2) NOT NULL ,
    ID_EXONERACION char(2) NOT NULL ,
    ID_ESTRATO_ECONOMICO char(2) NOT NULL,
    ID_ANIO_EJERCICIO_FISCAL char(2) NOT NULL,
    ID_ORG_JURIDICA char(2),
    ID_SECTOR char(2),
    ID_DISTRITO char(6) NOT NULL ,
    ID_CIIU varchar(4) NOT NULL ,
    TOTAL_EMPRESAS_ENCUESTA integer NULL ,
    TOTAL_EMPRESAS_PARTICIPACION integer NULL ,
    TOTAL_TRABAJADORES_EMPRESA integer NULL ,
    NUMERO_EMPRESAS_OMISAS integer NULL ,
    NUMERO_EMPRESAS_SECTOR integer NULL ,
    TOTAL_VENTAS_EMPRESA integer NULL ,
    TOTAL_VENTAS_SECTOR integer NULL ,
    TOTAL_VENTAS_DIVISION integer NULL ,
    TOTAL_EMPRESAS_DEPARTAMENTO integer NULL ,
)
go
ALTER TABLE FACT_RESULTADOS_EEA
    ADD CONSTRAINT XPKFACT_RESULTADOS_EEA PRIMARY KEY
    CLUSTERED (ID_FACT_RESULTADOS ASC)
go

ALTER TABLE FACT_RESULTADOS_EEA
    ADD CONSTRAINT R_05 FOREIGN KEY (Fecha) REFERENCES
    DIM_TIEMPO(Fecha)
        ON DELETE NO ACTION
        ON UPDATE NO ACTION
go
ALTER TABLE FACT_RESULTADOS_EEA
    ADD CONSTRAINT R_06 FOREIGN KEY (RUC, ID_ENCUESTA)
    REFERENCES DIM_EMPRESA(RUC, ID_ENCUESTA)
        ON DELETE NO ACTION
        ON UPDATE NO ACTION
go
```

```

ALTER TABLE FACT_RESULTADOS_EEA
    ADD CONSTRAINT R_07 FOREIGN KEY (ID_PARTICIPACION)
REFERENCES DIM_PARTICIPACION(ID_PARTICIPACION )
    ON DELETE NO ACTION
    ON UPDATE NO ACTION

go
ALTER TABLE FACT_RESULTADOS_EEA
    ADD CONSTRAINT R_08 FOREIGN KEY (ID_EXONERACION)
REFERENCES DIM_EXONERACION(ID_EXONERACION )
    ON DELETE NO ACTION
    ON UPDATE NO ACTION

go
ALTER TABLE FACT_RESULTADOS_EEA
    ADD CONSTRAINT R_09 FOREIGN KEY
(ID ESTRATO ECONOMICO, ID ANIO EJERCICIO FISCAL) REFERENCES
DIM ESTRATO ECONOMICO(ID ESTRATO ECONOMICO, ID ANIO EJERCIC
IO FISCAL)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION

go
ALTER TABLE FACT_RESULTADOS_EEA
    ADD CONSTRAINT R_10 FOREIGN KEY (ID_ORG_JURIDICA)
REFERENCES DIM_ORG_JURIDICA(ID_ORG_JURIDICA)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION

go
ALTER TABLE FACT_RESULTADOS_EEA
    ADD CONSTRAINT R_11 FOREIGN KEY (ID_SECTOR) REFERENCES
DIM_SECTOR_ECONOMICO(ID_SECTOR)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION

go
ALTER TABLE FACT_RESULTADOS_EEA
    ADD CONSTRAINT R_12 FOREIGN KEY (ID_DISTRITO) REFERENCES
DIM_DISTRITO(ID_DISTRITO)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION

go
ALTER TABLE FACT_RESULTADOS_EEA
    ADD CONSTRAINT R_13 FOREIGN KEY (ID_CIIU) REFERENCES
DIM_CIIU(ID_CIIU)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION

go

```

## Código de ETL para la carga del Datamart

Script para el llenado de la dimensiones

-----LLENANDO LA DIMENSION TIEMPO-----

```
/*Script de carga*/
DECLARE @FechaDesde as smalldatetime, @FechaHasta as smalldatetime
DECLARE @FechaAAAAMMDD int
DECLARE @Anio as smallint, @Trimestre char(2), @Mes smallint
DECLARE @Semana smallint, @Dia smallint, @DiaSemana smallint
DECLARE @NTrimestre char(7), @NMes char(15)
DECLARE @NMes3l char(3)
DECLARE @NSemana char(10), @NDia char(6), @NDiaSemana char(10)
--Set inicial por si no coincide con los del servidor
SET DATEFORMAT dmy
SET DATEFIRST 1

BEGIN TRANSACTION
    --Borrar datos actuales, si fuese necesario
    --TRUNCATE TABLE FROM DI_TIEMPO

    --Rango de fechas a generar: del 01/01/2006 al 31/12/Año actual+2
    SELECT @FechaDesde = CAST('20060101' AS smalldatetime)
    SELECT @FechaHasta = CAST(CAST(YEAR(GETDATE()))+2 AS CHAR(4)) +
'1231' AS smalldatetime)

    WHILE (@FechaDesde <= @FechaHasta) BEGIN
        SELECT @FechaAAAAMMDD = YEAR(@FechaDesde)*10000+
            MONTH(@FechaDesde)*100+
            DATEPART(dd, @FechaDesde)
        SELECT @Anio = DATEPART(yy, @FechaDesde)
        SELECT @Trimestre = DATEPART(qq, @FechaDesde)
        SELECT @Mes = DATEPART(m, @FechaDesde)
        SELECT @Semana = DATEPART(wk, @FechaDesde)
        SELECT @Dia = RIGHT('0' + DATEPART(dd, @FechaDesde),2)
        SELECT @DiaSemana = DATEPART(DW, @FechaDesde)
        SELECT @NMes = DATENAME(mm, @FechaDesde)
        SELECT @NMes3l = LEFT(@NMes, 3)
        SELECT @NTrimestre = 'T' + CAST(@Trimestre as CHAR(1)) + '/' +
RIGHT(@Anio, 2)
        SELECT @NSemana = 'Sem ' +CAST(@Semana AS CHAR(2)) + '/' +
RIGHT(RTRIM(CAST(@Anio as CHAR(4))),2)
        SELECT @NDia = CAST(@Dia as CHAR(2)) + '' + RTRIM(@NMes)
        SELECT @NDiaSemana = DATENAME(dw, @FechaDesde)
        INSERT INTO Datamart.dbo.DIM_TIEMPO
        (
            FechaSK,
            Fecha,
```

```

Anio,
Trimestre,
Mes,
Semana,
Dia,
DiaSemana,
NTrimestre,
NMes,
NMes3L,
NSemana,
NDia,
NDiaSemana
) VALUES
(
  @FechaAAAAMMDD,
  @FechaDesde,
  @Anio,
  @Trimestre,
  @Mes,
  @Semana,
  @Dia,
  @DiaSemana,
  @NTrimestre,
  @NMes,
  @NMes3l,
  @NSemana,
  @NDia,
  @NDiaSemana
)
--Incremento del bucle
SELECT @FechaDesde = DATEADD(DAY, 1, @FechaDesde)
END
COMMIT TRANSACTION

```

-----LLENANDO LA DIMENSION DEPARTAMENTO-----

```

INSERT INTO DIM_DEPARTAMENTO
(ID_DEPARTAMENTO, NOM_DEPARTAMENTO)
SELECT Ccdd ,Descripcion FROM [EEA_2016].[dbo].[TABLA_Dpto];

```

-----LLENANDO LA DIMENSION PROVINCIA-----

```

INSERT INTO DIM_PROVINCIA
(ID_PROVINCIA, ID_DEPARTAMENTO, NOM_PROVINCIA)
SELECT Ccdd+Ccpp, Ccdd, Descripcion FROM [EEA_2016].[dbo].[TABLA_PROV];

```

-----LLENANDO LA DIMENSION DISTRITO-----

```

INSERT INTO DIM_DISTRITO
(ID_DISTRITO, ID_PROVINCIA, NOM_DISTRITO)
SELECT Ccdd+Ccpp+CCDI, Ccdd+Ccpp, Descripcion FROM
[EEA_2016].[dbo].[TABLA_DIST];

```

```

-----LLENANDO LA DIMENSION CIU-----
INSERT INTO DIM_CIU
(ID_CIU, COD_DIVISION, COD_GRUPO, COD_CLASE, NOM_DIVISION)
SELECT
CodDivision+CodGrupo+CodClase, CodDivision, CodGrupo, CodClase, Descripcion
FROM [EEA_2016].[dbo].[TABLA_Ciu];

```

```

-----LLENANDO LA DIMENSION SECTOR ECONOMICO-----
INSERT INTO DIM_SECTOR_ECONOMICO
(ID_SECTOR, NOMBRE_SECTOR)
SELECT CodSector, Descripcion FROM [EEA_2016].[dbo].[SECTOR];

```

```

-----LLENANDO LA DIMENSION ORGANIZACION JURIDICA-----
-----
INSERT INTO DIM_ORG_JURIDICA
(ID_ORG_JURIDICA, NOM_ORG_JURIDICA)
SELECT Cod_OrgJuridica, Descripcion FROM
[EEA_2016].[dbo].[TABLA_CatOrganizacionJuridica];

```

```

-----LLENANDO LA DIMENSION PARTICIPACION-----
INSERT INTO DIM_PARTICIPACION
(ID_PARTICIPACION, DESCRIPCION) VALUES('00', 'PARTICIPO EN LA
ENCUESTA'), ('01', 'NO PARTICIPO EN LA ENCUESTA');

```

```

-----LLENANDO LA DIMENSION EXONERACION-----
INSERT INTO DIM_EXONERACION
(ID_EXONERACION, DESCRIPCION) VALUES('00', 'EXONERADO'), ('01', 'NO
EXONERADO');

```

```

-----LLENANDO LA DIMENSION EMPRESA-----
INSERT INTO DIM_EMPRESA
(RUC, ID_ENCUESTA, RazSocial, NRO_ESTABLEC, EMAIL, WEB, TELEFONO, COD
IGO_USUARIO)
SELECT RUC, '05', RazSocial, '000', Email, Web, TELEF_1, CODIGO_USUARIO
FROM [EEA_2016].[dbo].[MARCOTOTAL];

```

```

INSERT INTO DIM_EMPRESA
(RUC, ID_ENCUESTA, RazSocial, NRO_ESTABLEC, EMAIL, WEB, TELEFONO, COD
IGO_USUARIO)
SELECT RUC, '04', RazSocial, '000', Email, Web, TELEF_1, CODIGO_USUARIO
FROM [EEA_2015].[dbo].[MARCOTOTAL];

```

```

INSERT INTO DIM_EMPRESA

```

```
(RUC,ID_ENCUESTA,RazSocial,NRO_ESTABLEC,EMAIL,WEB,TELEFONO,CODIGO_USUARIO)
SELECT RUC,'03',RazSocial,'000',Email,Web,TELEF_1,CODIGO_USUARIO
FROM [EEA_2014].[dbo].[MARCOTOTAL];
```

```
INSERT INTO DIM_EMPRESA
(RUC,ID_ENCUESTA,RazSocial,NRO_ESTABLEC,EMAIL,WEB,TELEFONO,CODIGO_USUARIO)
SELECT RUC,'02',RazSocial,'000',Email,Web,TELEF_1,CODIGO_USUARIO
FROM [EEA_2013].[dbo].[MARCOTOTAL];
```

```
INSERT INTO DIM_EMPRESA
(RUC,ID_ENCUESTA,RazSocial,NRO_ESTABLEC,EMAIL,WEB,TELEFONO,CODIGO_USUARIO)
SELECT RUC,'01',RazSocial,'000',Email,Web,TELEF_1,CODIGO_USUARIO
FROM [EEA_2012].[dbo].[MARCOTOTAL];
```

```
DELETE FROM DIM_EMPRESA where ruc=""
```

-----LLENANDO LA DIMENSION ESTRATO ECONOMICO-----

```
INSERT INTO DIM_ESTRATO_ECONOMICO
(ID_ESTRATO_ECONOMICO,ID_ANIO_EJERCICIO_FISCAL,DESCRIPCION_UI
T,MONTO_MIN,MONTO_MAX,DESCRIPCION_ESTRATO,RANGO_VENTAS)
SELECT CODRANVEN,'05',DESCRIPCION,valormin,valormax,estrato,Ven
FROM [EEA_2016].[dbo].[ESTRATO_2016];
```

```
INSERT INTO DIM_ESTRATO_ECONOMICO
(ID_ESTRATO_ECONOMICO,ID_ANIO_EJERCICIO_FISCAL,DESCRIPCION_UI
T,MONTO_MIN,MONTO_MAX,DESCRIPCION_ESTRATO,RANGO_VENTAS)
SELECT CODRANVEN,'04',DESCRIPCION,valormin,valormax,estrato,Ven
FROM [EEA_2016].[dbo].[ESTRATO_2015];
```

```
INSERT INTO DIM_ESTRATO_ECONOMICO
(ID_ESTRATO_ECONOMICO,ID_ANIO_EJERCICIO_FISCAL,DESCRIPCION_UI
T,MONTO_MIN,MONTO_MAX,DESCRIPCION_ESTRATO,RANGO_VENTAS)
SELECT CODRANVEN,'03',DESCRIPCION,valormin,valormax,estrato,Ven
FROM [EEA_2016].[dbo].[ESTRATO_2014];
```

```
INSERT INTO DIM_ESTRATO_ECONOMICO
(ID_ESTRATO_ECONOMICO,ID_ANIO_EJERCICIO_FISCAL,DESCRIPCION_UI
T,MONTO_MIN,MONTO_MAX,DESCRIPCION_ESTRATO,RANGO_VENTAS)
SELECT CODRANVEN,'02',DESCRIPCION,valormin,valormax,estrato,Ven
FROM [EEA_2016].[dbo].[ESTRATO_2013];
```

```
INSERT INTO DIM_ESTRATO_ECONOMICO
(ID_ESTRATO_ECONOMICO,ID_ANIO_EJERCICIO_FISCAL,DESCRIPCION_UI
T,MONTO_MIN,MONTO_MAX,DESCRIPCION_ESTRATO,RANGO_VENTAS)
SELECT CODRANVEN,'01',DESCRIPCION,valormin,valormax,estrato,Ven
```



FROM [EEA\_2016].[dbo].[ESTRATO\_2012];

## Código de ETL para la carga de la tabla de hechos

----- LLENANDO LA TABLA DE HECHOS FACTRESULTADOS\_EEA -----

```
CREATE PROCEDURE SP_CARGA_FACTRESULTADOS_EEA
AS
BEGIN
INSERT INTO FACT_RESULTADOS_EEA
(ID_FACT_RESULTADOS,Fecha,RUC,ID_ENCUESTA,ID_PARTICIPACION,ID_EXONERACION,
ID_ESTRATO_ECONOMICO,ID_ANIO_EJERCICIO_FISCAL,ID_ORG_JURIDICA,
ID_SECTOR,ID_DISTRITO,ID_CIIU
,TOTAL_EMPRESAS_ENCUESTA,TOTAL_EMPRESAS_PARTICIPACION,TOTAL_TRABAJADORES_EMPRESA,
NUMERO_EMPRESAS_OMISAS,NUMERO_EMPRESAS_SECTOR,TOTAL_VENTAS_EMPRESA
,TOTAL_VENTAS_SECTOR,TOTAL_VENTAS_DIVISION,TOTAL_EMPRESAS_DEPARTAMENTO)

SELECT ROW AS ID_FACT_RESULTADOS,ANIO,RUC,ID_ENCUESTA,CASE
WHEN PARTICIPACION = '0' THEN '01' WHEN PARTICIPACION != '0' THEN '00'
END ID_PARTICIPACION
,CASE WHEN EXONERACION = '0' THEN '01' WHEN
EXONERACION != '0' THEN '00' END ID_EXONERACION
,CASE WHEN VENTAS BETWEEN -9999999999 AND 570000
THEN '1'
WHEN VENTAS BETWEEN 570001 AND 6460000
THEN '2'
WHEN VENTAS BETWEEN 6460001 AND 8740000
THEN '3'
WHEN VENTAS BETWEEN 8740001 AND 9999999999
THEN '4' END ID_ESTRATO_ECONOMICO
,ID_ANIO_EJERCICIO_FISCAL,ID_ORG_JURIDICA,ID_SECTOR,ID_DISTRITO,
ID_CIIU
,(SELECT COUNT(RUC) FROM [EEA_2016].[dbo].[MARCOTOTAL]) AS
'TOTAL_EMPRESAS_ENCUESTA'
,(SELECT COUNT(RUC) FROM [EEA_2016].[dbo].[DIRECTORIO] WHERE
NROESTABLEC='000') AS 'TOTAL_EMPRESAS_PARTICIPACION'
,X.PERSONAL_OCUPADO AS 'TOTAL_TRABAJADORES_EMPRESA'
,(SELECT COUNT (MT.RUC)FROM [EEA_2016].[dbo].[MARCOTOTAL] MT
LEFT JOIN [EEA_2016].[dbo].[MARCO_CONTROL] MC ON MT.RUC=MC.RUC
WHERE FLAGEXONERACION!='0')AS 'NUMERO_EMPRESAS_OMISAS'
,CASE WHEN X.ID_SECTOR = '01' THEN (SELECT COUNT(RUC) FROM
[EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CODSECTOR='01')
```

```

    WHEN X.ID_SECTOR = '02' THEN (SELECT COUNT(RUC) FROM
[EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CODSECTOR='02')
    WHEN X.ID_SECTOR = '03' THEN (SELECT COUNT(RUC) FROM
[EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CODSECTOR='03')
    WHEN X.ID_SECTOR = '04' THEN (SELECT COUNT(RUC) FROM
[EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CODSECTOR='04')
    WHEN X.ID_SECTOR = '05' THEN (SELECT COUNT(RUC) FROM
[EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CODSECTOR='05')
    WHEN X.ID_SECTOR = '07' THEN (SELECT COUNT(RUC) FROM
[EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CODSECTOR='07')
    WHEN X.ID_SECTOR = '08' THEN (SELECT COUNT(RUC) FROM
[EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CODSECTOR='08')
    WHEN X.ID_SECTOR = '10' THEN (SELECT COUNT(RUC) FROM
[EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CODSECTOR='10')
    WHEN X.ID_SECTOR = '11' THEN (SELECT COUNT(RUC) FROM
[EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CODSECTOR='11')
    WHEN X.ID_SECTOR = '12' THEN (SELECT COUNT(RUC) FROM
[EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CODSECTOR='12')
    WHEN X.ID_SECTOR = '13' THEN (SELECT COUNT(RUC) FROM
[EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CODSECTOR='13')
    WHEN X.ID_SECTOR = '14' THEN (SELECT COUNT(RUC) FROM
[EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CODSECTOR='14')
    WHEN X.ID_SECTOR = '15' THEN (SELECT COUNT(RUC) FROM
[EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CODSECTOR='15')
    WHEN X.ID_SECTOR = '16' THEN (SELECT COUNT(RUC) FROM
[EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CODSECTOR='16')
END 'NUMERO_EMPRESAS_SECTOR'
,X.VENTAS AS 'TOTAL_VENTAS_EMPRESA'
,CASE WHEN X.ID_SECTOR = '00' THEN '0'
    WHEN X.ID_SECTOR = '01' THEN (SELECT SUM(VENTAS) FROM
[EEA_2016].[dbo].[Directorio_UNIFICADO] WHERE SECTOR='01')
    WHEN X.ID_SECTOR = '02' THEN (SELECT SUM(VENTAS) FROM
[EEA_2016].[dbo].[Directorio_UNIFICADO] WHERE SECTOR='02')
    WHEN X.ID_SECTOR = '03' THEN (SELECT SUM(VENTAS) FROM
[EEA_2016].[dbo].[Directorio_UNIFICADO] WHERE SECTOR='03')

```

```

WHEN X.ID_SECTOR = '04' THEN (SELECT SUM(VENTAS) FROM
[EEA_2016].[dbo].[Directorio_UNIFICADO] WHERE SECTOR='04')
WHEN X.ID_SECTOR = '05' THEN (SELECT SUM(VENTAS) FROM
[EEA_2016].[dbo].[Directorio_UNIFICADO] WHERE SECTOR='05')
WHEN X.ID_SECTOR = '07' THEN (SELECT SUM(VENTAS) FROM
[EEA_2016].[dbo].[Directorio_UNIFICADO] WHERE SECTOR='07')
WHEN X.ID_SECTOR = '08' THEN (SELECT SUM(VENTAS) FROM
[EEA_2016].[dbo].[Directorio_UNIFICADO] WHERE SECTOR='08')
WHEN X.ID_SECTOR = '09' THEN (SELECT SUM(VENTAS) FROM
[EEA_2016].[dbo].[Directorio_UNIFICADO] WHERE SECTOR='09')
WHEN X.ID_SECTOR = '10' THEN (SELECT SUM(VENTAS) FROM
[EEA_2016].[dbo].[Directorio_UNIFICADO] WHERE SECTOR='10')
WHEN X.ID_SECTOR = '11' THEN (SELECT SUM(VENTAS) FROM
[EEA_2016].[dbo].[Directorio_UNIFICADO] WHERE SECTOR='11')
WHEN X.ID_SECTOR = '12' THEN (SELECT SUM(VENTAS) FROM
[EEA_2016].[dbo].[Directorio_UNIFICADO] WHERE SECTOR='12')
WHEN X.ID_SECTOR = '13' THEN (SELECT SUM(VENTAS) FROM
[EEA_2016].[dbo].[Directorio_UNIFICADO] WHERE SECTOR='13')
WHEN X.ID_SECTOR = '14' THEN (SELECT SUM(VENTAS) FROM
[EEA_2016].[dbo].[Directorio_UNIFICADO] WHERE SECTOR='14')
WHEN X.ID_SECTOR = '15' THEN (SELECT SUM(VENTAS) FROM
[EEA_2016].[dbo].[Directorio_UNIFICADO] WHERE SECTOR='15')
WHEN X.ID_SECTOR = '16' THEN (SELECT SUM(VENTAS) FROM
[EEA_2016].[dbo].[Directorio_UNIFICADO] WHERE SECTOR='16')
END 'TOTAL_VENTAS_SECTOR','TOTAL_VENTAS_DIVISION'

```

```

,CASE WHEN SUBSTRING(X.ID_DISTRITO,1,2)= '00' THEN '00'
WHEN SUBSTRING(X.ID_DISTRITO,1,2)= '01' THEN (SELECT COUNT(RUC)
FROM [EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CCDD='01')
WHEN SUBSTRING(X.ID_DISTRITO,1,2)= '02' THEN (SELECT COUNT(RUC)
FROM [EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CCDD='02')
WHEN SUBSTRING(X.ID_DISTRITO,1,2)= '03' THEN (SELECT COUNT(RUC)
FROM [EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CCDD='03')
WHEN SUBSTRING(X.ID_DISTRITO,1,2)= '04' THEN (SELECT COUNT(RUC)
FROM [EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CCDD='04')
WHEN SUBSTRING(X.ID_DISTRITO,1,2)= '05' THEN (SELECT COUNT(RUC)
FROM [EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CCDD='05')
WHEN SUBSTRING(X.ID_DISTRITO,1,2)= '06' THEN (SELECT COUNT(RUC)
FROM [EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CCDD='06')
WHEN SUBSTRING(X.ID_DISTRITO,1,2)= '07' THEN (SELECT COUNT(RUC)
FROM [EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CCDD='07')

```

WHEN SUBSTRING(X.ID\_DISTRITO,1,2)= '08' THEN (SELECT COUNT(RUC)  
 FROM [EEA\_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND  
 CCDD='08')

WHEN SUBSTRING(X.ID\_DISTRITO,1,2)= '09' THEN (SELECT COUNT(RUC)  
 FROM [EEA\_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND  
 CCDD='09')

WHEN SUBSTRING(X.ID\_DISTRITO,1,2)= '10' THEN (SELECT COUNT(RUC)  
 FROM [EEA\_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND  
 CCDD='10')

WHEN SUBSTRING(X.ID\_DISTRITO,1,2)= '11' THEN (SELECT COUNT(RUC)  
 FROM [EEA\_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND  
 CCDD='11')

WHEN SUBSTRING(X.ID\_DISTRITO,1,2)= '12' THEN (SELECT COUNT(RUC)  
 FROM [EEA\_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND  
 CCDD='12')

WHEN SUBSTRING(X.ID\_DISTRITO,1,2)= '13' THEN (SELECT COUNT(RUC)  
 FROM [EEA\_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND  
 CCDD='13')

WHEN SUBSTRING(X.ID\_DISTRITO,1,2)= '14' THEN (SELECT COUNT(RUC)  
 FROM [EEA\_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND  
 CCDD='14')

WHEN SUBSTRING(X.ID\_DISTRITO,1,2)= '15' THEN (SELECT COUNT(RUC)  
 FROM [EEA\_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND  
 CCDD='15')

WHEN SUBSTRING(X.ID\_DISTRITO,1,2)= '16' THEN (SELECT COUNT(RUC)  
 FROM [EEA\_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND  
 CCDD='16')

WHEN SUBSTRING(X.ID\_DISTRITO,1,2)= '17' THEN (SELECT COUNT(RUC)  
 FROM [EEA\_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND  
 CCDD='17')

WHEN SUBSTRING(X.ID\_DISTRITO,1,2)= '18' THEN (SELECT COUNT(RUC)  
 FROM [EEA\_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND  
 CCDD='18')

WHEN SUBSTRING(X.ID\_DISTRITO,1,2)= '19' THEN (SELECT COUNT(RUC)  
 FROM [EEA\_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND  
 CCDD='19')

WHEN SUBSTRING(X.ID\_DISTRITO,1,2)= '20' THEN (SELECT COUNT(RUC)  
 FROM [EEA\_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND  
 CCDD='20')

WHEN SUBSTRING(X.ID\_DISTRITO,1,2)= '21' THEN (SELECT COUNT(RUC)  
 FROM [EEA\_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND  
 CCDD='21')

WHEN SUBSTRING(X.ID\_DISTRITO,1,2)= '22' THEN (SELECT COUNT(RUC)  
 FROM [EEA\_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND  
 CCDD='22')

WHEN SUBSTRING(X.ID\_DISTRITO,1,2)= '23' THEN (SELECT COUNT(RUC)  
 FROM [EEA\_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND  
 CCDD='23')

```

    WHEN SUBSTRING(X.ID_DISTRITO,1,2)= '24' THEN (SELECT COUNT(RUC)
FROM [EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CCDD='24')
    WHEN SUBSTRING(X.ID_DISTRITO,1,2)= '25' THEN (SELECT COUNT(RUC)
FROM [EEA_2016].[dbo].[DIRECTORIO] WHERE NROESTABLEC='000' AND
CCDD='25')
END 'NUMERO_EMPRESAS_DEPARTAMENTO'
FROM
(
SELECT ROW_NUMBER() OVER(ORDER BY MT.RUC DESC) AS Row,'2016' AS
ANIO,MT.RUC,D_EMP.ID_ENCUESTA
,ISNULL((SELECT RUC FROM [EEA_2016].[dbo].[DIRECTORIO] D WHERE
NROESTABLEC='000' AND MT.RUC=D.RUC),0) AS PARTICIPACION
,ISNULL((SELECT FLAGEXONERACION FROM
[EEA_2016].[dbo].[MARCO_CONTROL] MC WHERE MT.RUC=MC.RUC),0) AS
EXONERACION
,ISNULL(DU.VENTAS,0)AS VENTAS,ISNULL(DU.PERSONAL,0)AS
PERSONAL_OCUPADO,'05'AS
ID_ANIO_EJERCICIO_FISCAL,ISNULL(D_ORG.ID_ORG_JURIDICA,'00')AS
ID_ORG_JURIDICA
,ISNULL(D_SEC.ID_SECTOR,'00')AS
ID_SECTOR,ISNULL(D_DIST.ID_DISTRITO,'000000')AS
ID_DISTRITO,ISNULL(D_CIU.ID_CIU,'000000')AS ID_CIU
FROM [EEA_2016].[dbo].[MARCOTOTAL] MT
INNER JOIN DIM_EMPRESA D_EMP ON MT.RUC=D_EMP.RUC AND
D_EMP.ID_ENCUESTA='05'
LEFT JOIN [EEA_2016].[dbo].[Directorio_UNIFICADO] DU ON MT.RUC=DU.RUC
AND DU.ANIO='2015'
INNER JOIN DIM ESTRATO_ECONOMICO D_EST ON
D_EMP.ID_ENCUESTA=D_EST.ID_ANIO_EJERCICIO_FISCAL AND
D_EST.ID_ANIO_EJERCICIO_FISCAL ='05'
LEFT JOIN DIM_ORG_JURIDICA D_ORG ON
DU.ORGANIZACION=D_ORG.ID_ORG_JURIDICA
LEFT JOIN dbo.DIM_SECTOR_ECONOMICO D_SEC ON
DU.SECTOR=D_SEC.ID_SECTOR
LEFT JOIN dbo.DIM_DISTRITO D_DIST ON DU.UBIGEO=D_DIST.ID_DISTRITO
LEFT JOIN dbo.DIM_CIU D_CIU ON DU.CIU=D_CIU.ID_CIU
GROUP BY
MT.RUC,D_EMP.ID_ENCUESTA,DU.VENTAS,DU.PERSONAL,D_ORG.ID_ORG_
JURIDICA,D_SEC.ID_SECTOR,D_DIST.ID_DISTRITO,D_CIU.ID_CIU
)AS X
ORDER BY ROW
END
GO

SP_CARGA_FACTRESULTADOS_EEA

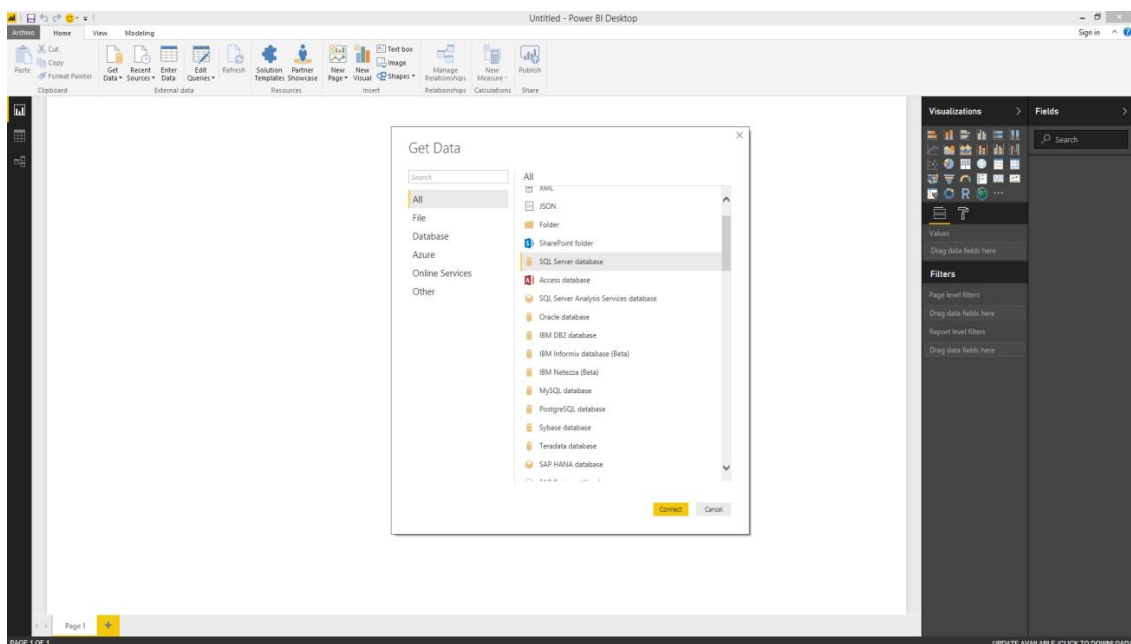
```

- **Aplicación del Datamart con POWER BI**

Luego de crear el script para las dimensiones y la tabla de hechos se procede a crear el proyecto en Power BI para extraer la visualización de los reportes requeridos por la DECEEE de la base de datos transaccional de la dirección.

Para el presente proyecto de tesis se creara una nueva aplicación de Power BI para el análisis de acuerdo a la información que pueda requerir el director o los altos jefes del área, para que puedan tomar la mejor decisión y la institución tenga mejores resultados a nivel institucional. A continuación se detallan los pasos:

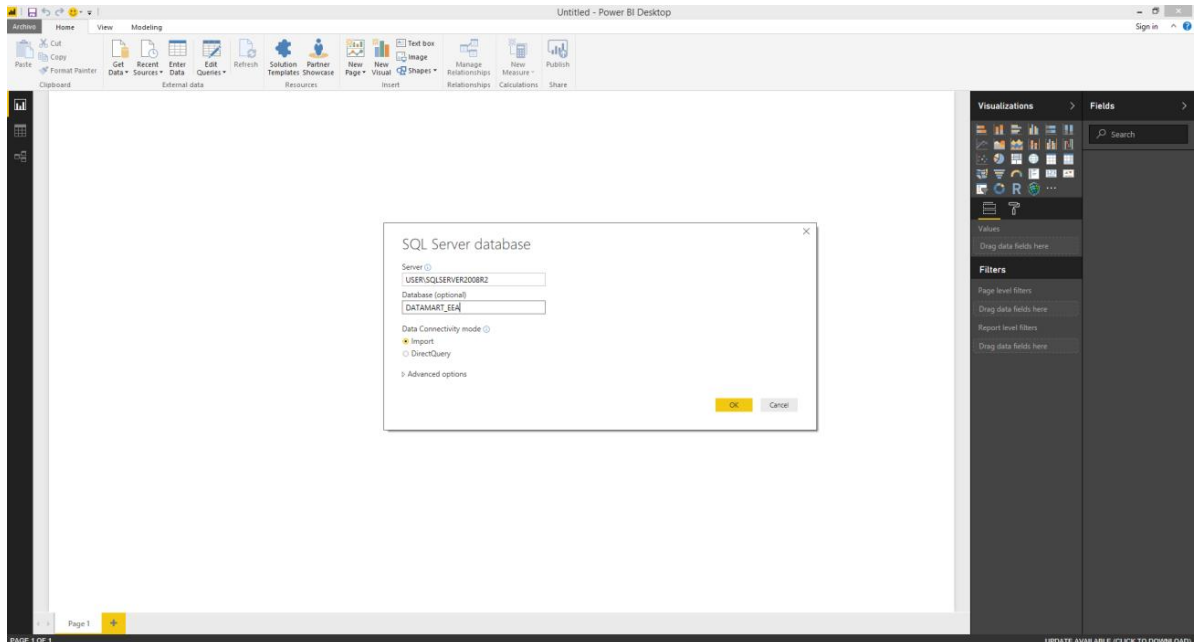
En la figura 3-3, se establece la nueva aplicación y el origen de datos.



**Figura 3-5: Entorno Power BI**

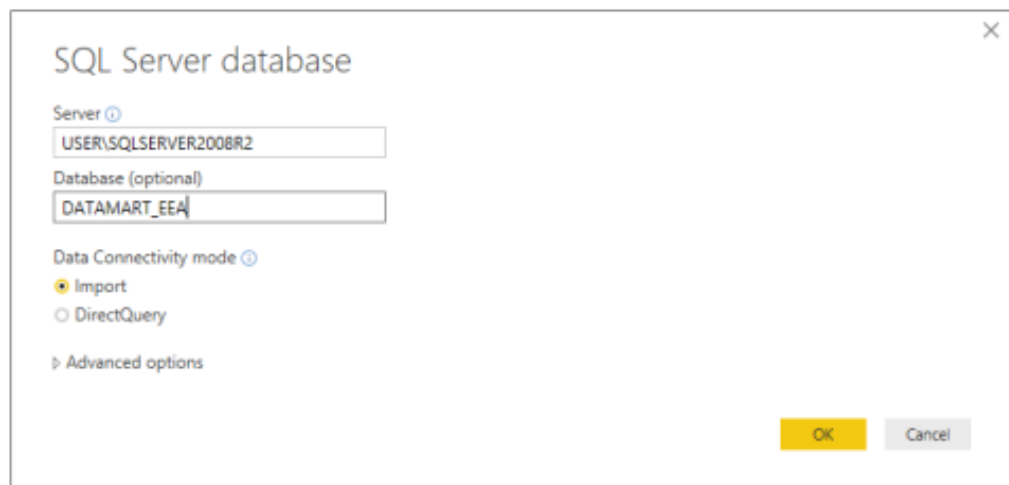
Fuente: Entorno de Desarrollo Propio

Para seleccionar la fuente de datos que a va a utilizar Power BI, se enlaza el origen de datos creado anteriormente, como se muestran en la figura 3-6 y 3-7, para la explotación de la información



**Figura 3-6: Conexión con el origen de datos**

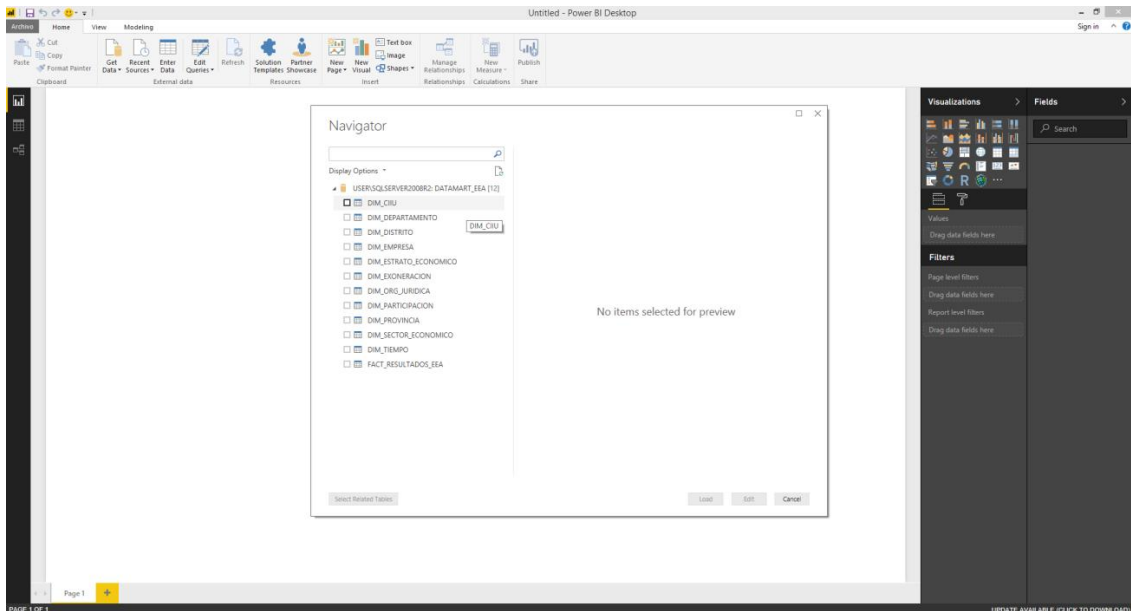
Fuente: Elaboración propia



**Figura 3-7: Conexión con la base de datos**

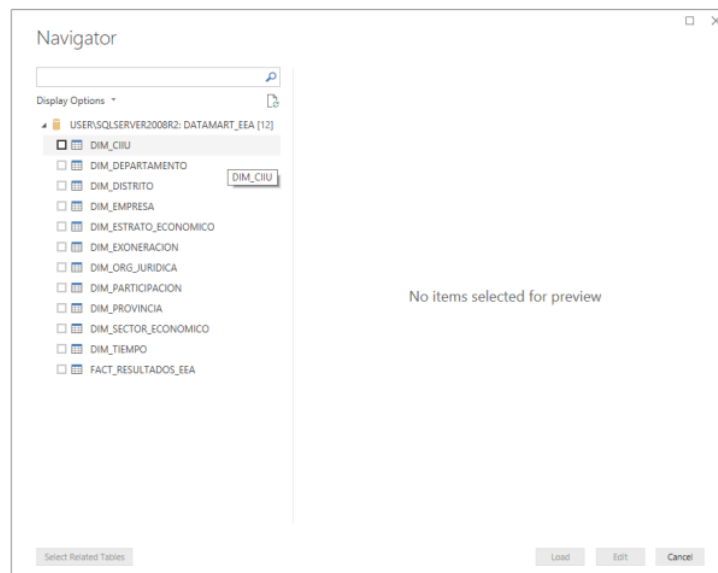
Fuente: Elaboración propia

Una vez cargado el origen de la base de datos, se podrá visualizar las dimensiones y la tabla de hechos, propias de la estructura del Datamart, tal como se muestra en la figura 3-8 y la figura 3-9.



**Figura 3-8: Visualización de las dimensiones y la fact table.**

Fuente: Elaboración propia



**Figura 3-9: Dimensiones y la tabla de hechos en Power BI**

Fuente: Elaboración propia



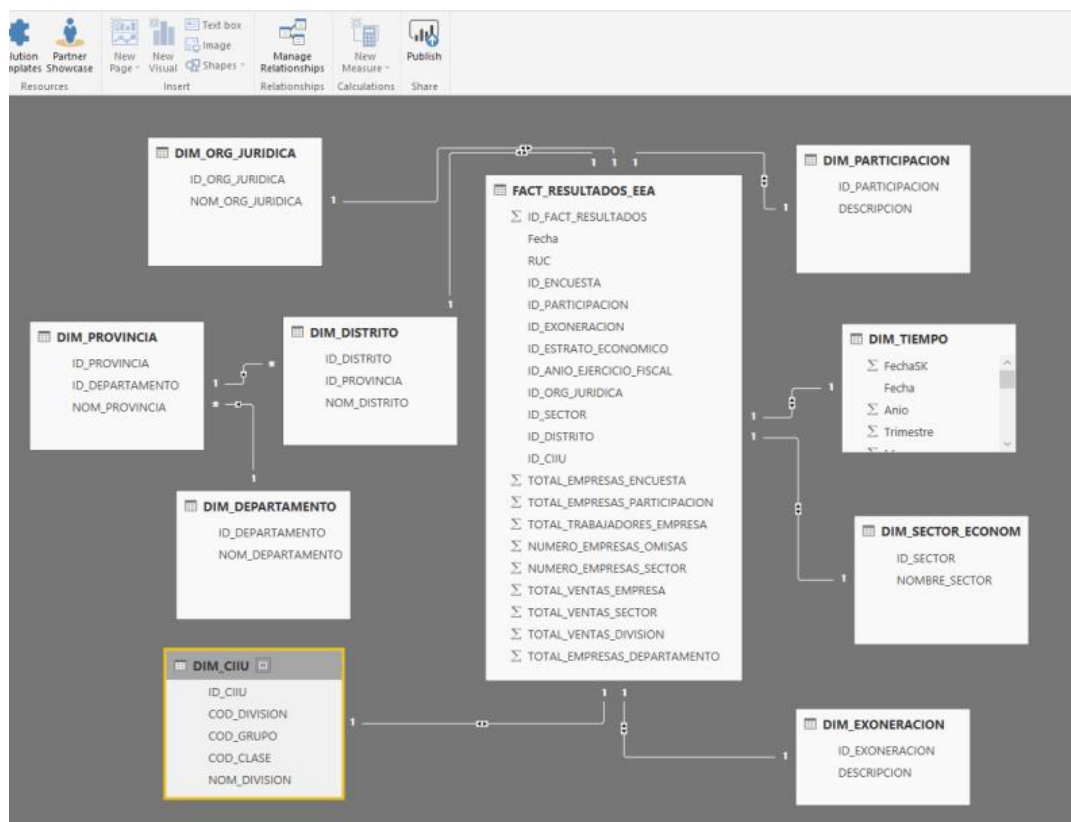
Para poder observar como quedo el esquema del Datamart, se elige la opción administración de relación de tablas y se mostrara las dimensiones y la tabla de hechos, cada uno con sus campos respectivos.



**Figura 3-10: Visor de Tablas**

Fuente: Elaboración propia.

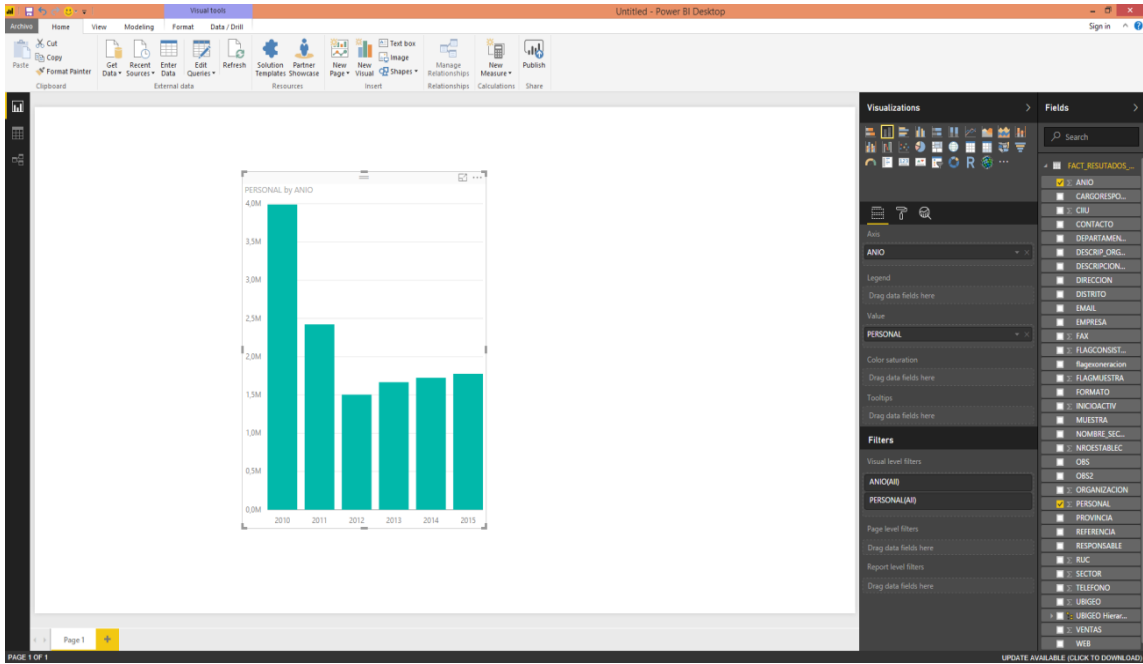
Power BI realiza una relación asociativa donde la información es precisa y confiable para ser explotada de acuerdo a los criterios del director o Jefes de la dirección.



**Figura 3-11: Visualización de dimensiones y la tabla de hechos**

Fuente: Elaboración propia

Y por último se generan los reportes requeridos por el director y los altos jefes de la DECEEE. Esto se logra seleccionando las variables a analizar de las tablas dimensionales y la tabla de hechos. Tal como se muestra en la figura 3-12.



**Figura 3-12: Visualización de reportes solicitados de la EEA en la DECEEE**

Fuente: Elaboración propia

### **3.3. Revisión y Consolidación de Resultados**

Teniendo el marco referencial del uso del Datamart, después de las pruebas realizadas con éxito en la Dirección Ejecutiva de Censos y Encuestas de Empresas y establecimientos (DECEEE) del INEI, se pudo realizar comparaciones del tiempo ahorrado en cada reporte generado, contrastándolo con el tiempo dedicado anteriormente para los reportes realizados de manera manual. Para el análisis se tomó como base entrevistas que se hicieron al personal informático y director de la DECEEE, junto con las respuestas dadas a través de la Encuesta de Satisfacción, ver Anexo 2 Encuesta de Valoración de Datamart.

Para la captura de requerimientos y necesidades se solicitó la planificación de entrevistas, así se obtuvieron datos reales para iniciar el análisis y diseño del Datamart que permita facilitar la generación de reportes que los usuarios finales requieren.

El beneficio obtenido cubrirán los objetivos específicos trazados en el inicio del proyecto de Tesis, se puede ver en la Tabla 3-22.

**Tabla 3-22: Objetivos vs Resultados**

Objetivos del Proyecto (General y Específicos)	Antes de la Implementación del Datamart	Después de la Implementación del Datamart	Satisfacción del Cliente (Encuesta)
Implementar un Datamart para mejorar la toma de decisiones en la dirección ejecutiva de censos y encuestas de empresas y establecimientos en el instituto nacional de estadística e informática.	La DECEEE en el INEI estaba afectada por la falta de una solución tecnológica que apoye la toma de decisiones que permitan dar seguimiento a las a las empresas participes en la EEA.	La DECEEE en el INEI cuenta con un mejor apoyo para la toma de decisiones que permitan dar seguimiento a las empresas que son parte de la EEA.	Muy Satisfecho
Analizar los requerimientos de información de la encuesta económica anual para mejorar la toma de decisiones en la dirección ejecutiva de censos y encuestas de empresas y establecimientos en el instituto nacional de estadística e informática.	Cuenta con un proceso manual para realizar las consultas, reportes y tabulados. La información para la toma de decisiones de la DECEEE del INEI no es oportuna.	Ausencia de la generación manual de reportes. Obtiene la información oportuna Reportes personalizados e intuitivos que apoyan a la toma decisiones para la DECEEE.	Satisfecho
Analizar y construir el Datamart de la encuesta económica anual para mejorar toma de decisiones en la dirección ejecutiva de censos y encuestas de empresas y establecimientos en el instituto nacional de estadística e informática.	Los indicadores con los que cuentan para realizar los reportes no se dan abasto para mostrar la información que desean consultar.	Se cuenta con una matriz de indicadores definidos y establecidos para cada reporte de interés para la DECEEE.	Muy Satisfecho
Generar reportes de la encuesta económica anual para mejorar la toma de decisiones en la dirección ejecutiva de censos y encuestas de empresas y establecimientos en el instituto nacional de estadística e informática.	El proceso de reportes es realizado manualmente.	El proceso de reportes es realizado de forma automatizada.	Satisfecho

Fuente: Elaboración propio.

Para evidenciar con exactitud el éxito de la solución ofrecida fue necesario medir el tiempo que se utiliza para la extracción de la información del sistema transaccional de la EEA, la transformación y la carga de cada reporte, dando uso del Datamart y haciendo una comparación de tiempo invertido antes de ello.

A continuación se muestra el tiempo mínimo y máximo de los reportes más solicitados por el director y los altos jefes de la DECEEE. Véase la tabla 3-23

**Tabla 3-23: Evaluación de tiempo, proceso manual**

N°	Reporte	Tiempo de Realización – Mínimo	Tiempo de Realización- Máximo
1	Reporte de cantidad de empresas por Sector Economico	2.5 hrs.	4 hrs.
2	Reporte de cantidad de empresas por departamento	2.5 hrs.	4 hrs.
3	Cantidad de Población Ocupada por departamento	2.5 hrs.	4 hrs.
4	Reporte de Ventas por sector economico	2.5 hrs.	4 hrs.
5	cantidad de Empresas por estrato Economico	2.5 hrs.	4 hrs.
6	Numero de empresas omisas por CIU	2.5 hrs.	4 hrs.
TOTAL		15 hrs.	24 hrs.

**Fuente:** Elaboración propia

El tiempo de realización de reporte antes de la implementación, es el que abarca todo el proceso manual de generación de reportes, se realiza desde la extracción de la información contenida en la base de datos de la EEA, para ello el especialista informático realiza los scripts y los ejecuta manualmente, luego se realizan tablas en Excel, ver Tabla 3-24 para luego utilizarlas como base para la generación de gráficos.

Todo ello tiene un tiempo promedio ubicado en el rango de 2.5hrs - 4hrs, lo cual multiplicado por seis reportes se obtiene el total de tiempo invertido promedio de 24 horas por la generación de los reportes.

**Tabla 3-24: Muestra del Excel, cantidad de empresas por sector económico**

<b>COD_SECTOR</b>	<b>NOMBRE SECTOR ECONOMICO</b>	<b>CANTIDAD DE EMPRESAS</b>
01	AGENCIAS DE VIAJE	110
02	AGROINDUSTRIA	571
03	CENTROS EDUCATIVOS NO ESTATALES	236
04	COMERCIO	3263
05	CONSTRUCCION	808
07	ESTABLECIMIENTOS DE HOSPEDAJE	391
08	HIDROCARBUROS	81
10	PESCA	436
11	MANUFACTURA	3106
12	SERVICIOS ELECTRICOS	90
13	TRANSPORTES Y COMUNICACIONES	906
14	UNIVERSIDADES NO ESTATALES	85
15	RESTAURANTES	1462
16	SERVICIOS	2375

Fuente: Elaboración propia

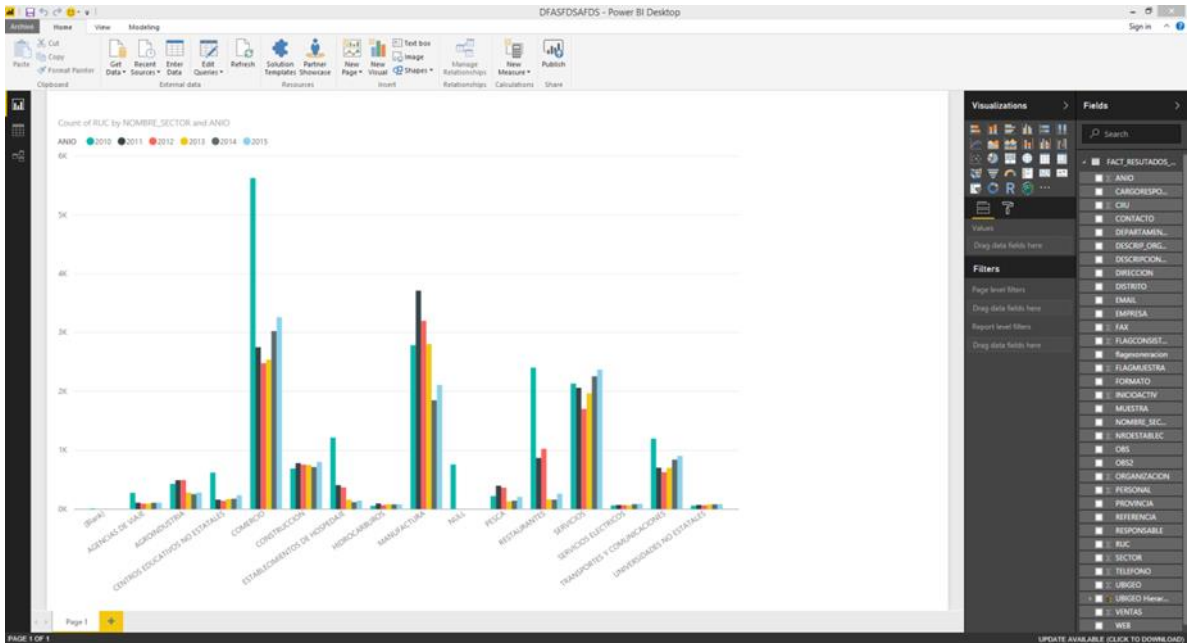
En el Datamart se alojará la información que se extraerá a través de la comunicación con la BD de la EEA, la cual pasará por un proceso ETL (Extracción, Transformación y Carga), el rango de tiempo promedio de su generación es de 0 a 5 minutos como lo podemos observar en la siguiente Tabla 3-25.

**Tabla 3-25:** Evaluación de tiempo, usando Datamart

<b>N°</b>	<b>Reporte</b>	<b>Tiempo de Realización –Mínimo</b>	<b>Tiempo de Realización-Máximo</b>
<b>1</b>	Reporte de cantidad de empresas por Sector Economico	<1 min	5 min
<b>2</b>	Reporte de cantidad de empresas por departamento	<1 min	5 min
<b>3</b>	Cantidad de Poblacion Ocupada por departamento	<1 min	5 min
<b>4</b>	Reporte de Ventas por sector economico	<1 min	5 min
<b>5</b>	cantidad de Empresas por estrato Economico	<1 min	5 min
<b>6</b>	Numero de empresas omisas por CIU	<1 min	5 min
<b>TOTAL</b>		<b>&lt;6 min</b>	<b>30 min.</b>

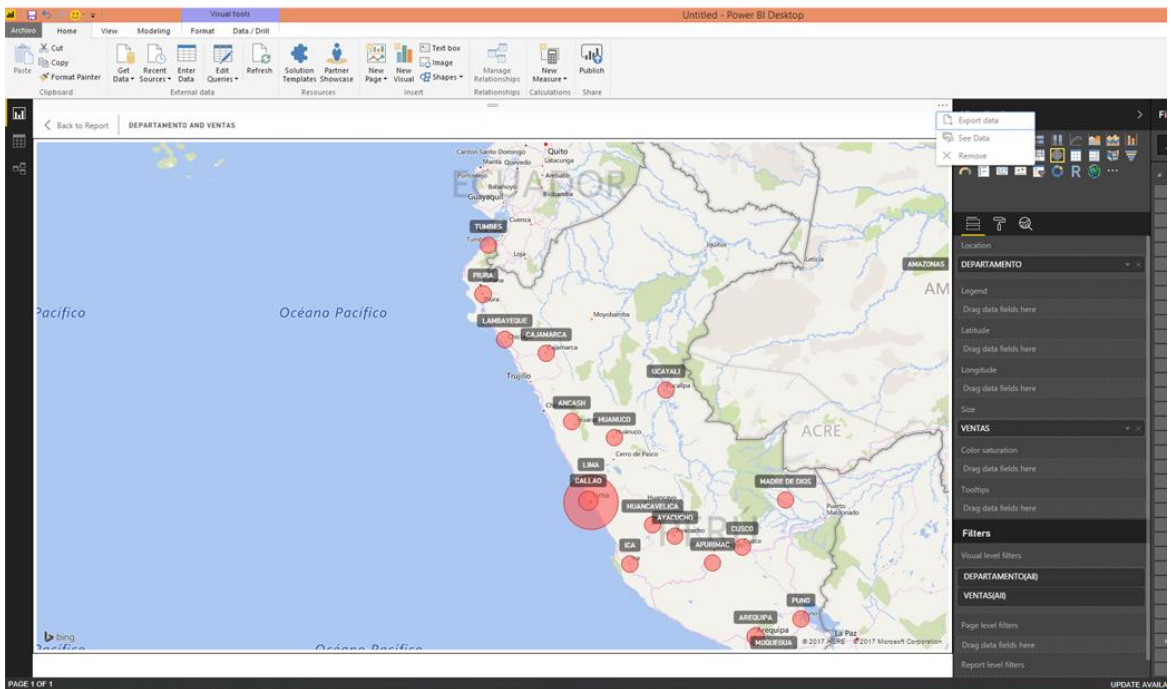
**Fuente:** Elaboración propia

Además de obtener los reportes en menor tiempo, con el uso del aplicativo Power BI los reportes se generan no solo con datos numéricos, sino también permite la visualización de gráficos estadísticos, permitiendo un entorno más amigable y comprensible para los usuarios finales, tal como se pueden observar en las figuras 3-13 y 3-14.



**Figura 3-13: cantidad de empresas por sector económico a través de los años**

Fuente: Elaboración propia



**Figura 3-14: Cantidad de Empresas por Departamento**

Fuente: Elaboración propia



Cálculo de Porcentaje de Reducción de Tiempo:

$$\% \text{ Reduccion de Tiempo} = 100 - \frac{\text{Tiempo de proceso actual} \times 100\%}{\text{Tiempo de Proceso anterior}}$$

$$\% \text{ Reduccion de tiempo} = 100 - (30 \times 100) / 1440$$

$$\% \text{ Reduccion de Tiempo} = 97.92 \%$$

El tiempo invertido ha dado como resultado una reducción de 1410 minutos equivalente a 97.92 % (Intervalo mayor), estos tiempos son de importancia en la toma de decisiones de la DECEEE para realizar y formular las estrategias y metodologías en las nuevas encuestas que la dirección tiene a su cargo, y que a la vez estos impactan que impactan al sector Económico del país.

## CONCLUSIONES

- ✓ La identificación de los principales indicadores son de gran importancia debido a que permite obtener una visión global y detallada del estado de las empresas en un determinado tiempo.
- ✓ La utilización de la metodología Ralph Kimball para este tipo de sistemas es el apropiado porque se acoge a los requerimientos generales definidos por parte de la institución y se puede extraer los datos de las empresas de manera eficaz, obteniendo resultados precisos al momento del análisis de los datos desde perspectivas diferentes a nivel de dirección.
- ✓ La construcción del ETL, nos permite manipular datos desde múltiples fuentes, además reformatearlos, limpiarlos y cargarlos al datamart, con el fin de realizar el análisis de toda la información procesada el cual le da flexibilidad al usuario a realizar diferentes consultas pre elaboradas.
- ✓ La implementación del Datamart nos permite el ahorro de tiempo, en un promedio de casi un 98% respecto al anterior, lo cual implicaba una sobrecarga de trabajo para el área de Informática de la DECEEE en el INEI.

## RECOMENDACIONES

- ✓ Seguir utilizando la metodología Ralph Kimball debido a que se acoge a los requerimientos de la Dirección Ejecutiva de Censos y Encuestas de Empresas y Establecimientos (DECEEE) en el Instituto Nacional de estadística e Informática (INEI).
- ✓ Se recomienda que en una segunda fase se incorporen nuevos procesos, para la generación de nuevos reportes de gestión. En el presente proyecto las bases están construidas para seguir extendiendo el número de indicadores en el datamart.
- ✓ Implementar el Balance Score Card (BSC), con la finalidad de construir indicadores que estén alineados a los requerimientos solicitados en los reportes de la Dirección Ejecutiva de Censos y Encuestas de Empresas y Establecimientos (DECEEE).

## BIOBLOGRAFIA

- ✓ Fernández Eddy. (2010). Análisis, Diseño e Implementación de un Datamart de clientes para el área de Marketing de una Entidad Aseguradora. Tesis de Bachiller publicada. Pontificia Universidad Católica del Perú, Lima, Perú.
- ✓ López, Carlos. (2007). Análisis, Diseño e Implementación de un Datamart para la dirección financiera y recursos humanos de la Escuela Politécnica del Ejército para una toma de decisión Efectiva. Tesis de Bachiller publicada. Departamento de ciencias de la computación de la Escuela Politécnica del Ejército, Sagolqui, Ecuador.
- ✓ Nuñez, Grace. (2010). Análisis, Diseño e Implementación de una Solución de Inteligencia de Negocio para el área de Finanzas de la Municipalidad de Lima. Tesis de Bachiller publicada. Pontificia Universidad Católica del Perú, Lima, Perú.
- ✓ Mendoza Paitan, Sergio Mauricio (2011). Analisis, diseño e implementacion de un sistema gerencial basado en una suite integrada de datamarts para las areas de finanzas, contabilidad, recursos humanos y comerciales. Lima: PUCP.
- ✓ Villanueva Álvaro. (2008). Análisis, Diseño e Implementación de un DataWarehouse de soporte de decisiones para un Hospital del Sistema de Salud Público. Tesis de Bachiller publicada. Pontificia Universidad Católica del Perú, Lima, Perú.
- ✓ Cabanillas, K. G. (2011). Análisis diseño e implementacion de una solución de Inteligencia de Negocios para el área de compras y ventas de una empresa comercializadora de electrodomésticos. Lima: PUCP.

- ✓ Díaz Duarte Daymara. (2005) Toma de decisiones: el imperativo diario de la vida en la organización moderna. ACIMED, Recuperado de :  
[http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1024-94352005000300010&lng=es](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352005000300010&lng=es).
  
- ✓ M. Leon, (2014) «Ciclo de vida de Ralph Kimball,» Recuperado de:  
<http://luisleonin.blogspot.com/2014/02/ciclo-de-vida-de-ralph-kimball.html>.
  
- ✓ Sinnexus, (2016) «Business Intelligence "Persistencia MOLAP, ROLAP, HOLAP",» 05 Mayo 2017. Recuperado de:  
[http://www.sinnexus.com/business\\_intelligence/olap\\_avanzado.aspx](http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx).
  
- ✓ Espiñera, Sheldon y Asociados. (2008) La Inteligencia de Negocios (Business Intelligence). Boletín e Asesoría Gerencial. 2008. Recuperado de:  
<http://www.pwc.com/ve/es/asesoria-gerencial/boletin/assets/boletin-advisory-edicion-10-2008.pdf>
  
- ✓ Lluís, Josep. (2008). BussinesIntelligence: Competir con Información. Barcelona- España: Editorial Dataprix.
  
- ✓ Ralph K. & Ross M. (2002) The data WareHouse toolkit: the complete to dimensional modeling. New York-EEUU: Editorial WILEY.
  
- ✓ Salvador R. (2011) Integration Services: Pieza fundamental en los proyectos de BI (parte 1 de 2). Recuperado de:  
[http://www.solidq.com/sqj/es/JournalDocuments/2011-March-Issue/Integration-Services-pieza-fundamental-en-los-proyectos-de-BI-\(parte-1-de-2\).pdf](http://www.solidq.com/sqj/es/JournalDocuments/2011-March-Issue/Integration-Services-pieza-fundamental-en-los-proyectos-de-BI-(parte-1-de-2).pdf)

- ✓ Espiñera, Sheldon y Asociados. La Inteligencia de Negocios (Business Intelligence). Boletín e Asesoría Gerencial. 2008. [Consultado el 30 de mayo 2017]. Disponible en: <http://www.pwc.com/ve/es/asesoria-gerencial/boletin/assets/boletin-advisory-edicion-10-2008.pdf>
  
- ✓ Lluís, Josep. BussinesIntelligence: Competir con Información. Barcelona-España: Editorial Dataprix. (2008).
  
- ✓ Ralph K. & Ross M. The data Warehouse toolkit: the complete to dimensional modeling. New York-EEUU: Editorial WILEY. (2002).
  
- ✓ Salvador R. (2011) Integration Services: Pieza fundamental en los proyectos de BI (parte 1 de 2). Recuperado de: [http://www.solidq.com/sqj/es/JournalDocuments/2011-March-Issue/Integration-Services-pieza-fundamental-en-los-proyectos-de-BI-\(parte-1-de-2\).pdf](http://www.solidq.com/sqj/es/JournalDocuments/2011-March-Issue/Integration-Services-pieza-fundamental-en-los-proyectos-de-BI-(parte-1-de-2).pdf)
  
- ✓ Fondevila Gascón, J.F. y Olmo Arriaga, J.L. (2013). *El trabajo fin de grado en ciencias sociales y jurídicas: Guía metodológica*. (2ª ed.) Madrid: Ediciones Internacionales Universitarias.
  
- ✓ Cajide Val, J. (Coord.). (2004). *Calidad universitaria y empleo*. Recuperado de <http://app.vlex.com/>.
  
- ✓ Kraig, P. (1990). Experimental archaeology in practice. En H. Cranfield (Ed.), *New trends in archaeology* (pp. 82-105). Birmingham: Birmingham University Publications.
  
- ✓ Porto, A., Mosteiro, M.J. y Castro, M. (2004). Orientación profesional e inserción laboral de los y las estudiantes universitarios. En J. Cajide Val (Coord.). *Calidad universitaria y empleo*. Recuperado de <http://app.vlex.com/>

## ANEXOS

### ANEXO 1: FLUJO DE COSTOS

ITEM	DESCRIPCION	Sem 1	Sem 2	Sem 3	Sem 4	Sem 6	Sem 7	Sem8	TOTAL
1	IMPLEMENTACION UN DATAMART PARA MEJORAR LA TOMA DE DECISIONES EN LA DIRECCIÓN EJECUTIVA DE CENSOS Y ENCUESTAS DE EMPRESAS Y ESTABLECIMIENTOS EN EL INSTITUTO NACIONAL DE ESTADÍSTICA E INFORMÁTICA	S/. 5.800,00	S/. 1.200,00	S/. 6.000,00	S/. 7.200,00	S/. 14.400,00	S/. 9.840,00	S/. 2.040,00	S/. 46.480,00
1.1	Plan del Proyecto	S/. 4.360,00							S/. 4.360,00
1.1.1	Planificación del alcance del Proyecto	S/. 1.360,00							
1.1.2	Definición de los objetivos del Proyecto	S/. 1.360,00							
1.1.3	Análisis de la metodología del Proyecto	S/. 1.200,00							
1.1.4	Evaluación de Riesgos	S/. 440,00							
1.2	Construcción del modelo Datamart	S/. 1.440,00	S/. 1.200,00	S/. 6.000,00	S/. 7.200,00	S/. 14.400,00	S/. 1.200,00		S/. 31.440,00
1.2.1	Desarrollo del modelo físico	S/. 1.440,00							
1.2.2	Desarrollo del modelo lógico		S/. 1.200,00						
1.2.2.1	Elaboración de Identificadores		S/. 1.200,00						
1.2.3	Modelamiento								
1.2.3.1	Diseño Físico			S/. 6.000,00					
1.2.3.2	Diseño de ETL				S/. 7.200,00				
1.2.3.3	desarrollo del ETL					S/. 14.400,00			
1.2.3.4	Aplicación del Datamart Power BI						S/. 1.200,00		
1.3	Pruebas						S/. 8.640,00		S/. 8.640,00
1.3.1	Instalación de Aplicación Power BI						S/. 1.800,00		
1.3.2	Generación de Reportes						S/. 3.000,00		
1.3.3	Evaluación de Resultados						S/. 2.040,00		
1.3.4	Implementación						S/. 1.800,00		
3.4	Cierre							S/. 2.040,00	S/. 2.040,00
1.4.1	Entrega del Producto							S/. 680,00	
1.4.2	Conformidad del Producto							S/. 680,00	
1.4.3	Cierre del proyecto							S/. 680,00	

Fuente: elaboración Propia, elaborado en MS.Project

## ANEXO 2: ENCUESTA DE VALORACION DE DATAMART

Nombre y Apellido:

Cargo:

Área/ Dirección:

Se da a su disposición la siguiente encuesta con la cual puede mostrar su grado de satisfacción con el producto brindado. Su opinión es de máxima importancia para poder definir los resultados obtenidos. Muchas gracias por su tiempo. Coloque dentro del recuadro el valor de su puntuación según su punto de vista en representación de la DECEEE en el INEI.

Se tiene la siguiente escala:

1    2    3    4    5

Dónde: 1=Muy insatisfecho, 2=Insatisfecho, 3=Regular, 4=Satisfecho y 5=Muy satisfecho

Items	Puntuación:
1. ¿Se dispone en forma oportuna los reportes de las Empresas que son o fueron estudiadas en la EEA?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> 1    2    3    4    5
2. ¿Se dispone en forma oportuna los reportes de los niveles de ventas ,población ocupada, cantidad de empresas por sectores económicos ?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> 1    2    3    4    5
3. ¿Esta Ud. Conforme con la implementación del Datamart para la generación de reportes?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> 1    2    3    4    5
4. ¿Los indicadores propuestos van acorde a la información que desea plasmarse en los reportes?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> 1    2    3    4    5
5. ¿Esta Ud. Conforme con que la extracción, transformación y carga de datos se realice con la data contenida en la EEA?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> 1    2    3    4    5
6. ¿Qué tan importante son los reportes generados con la ayuda del Datamart para la DECEEE?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> 1    2    3    4    5
7. ¿El tiempo invertido en la generación de reportes de los procesos consulta se ha reducido considerablemente?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> 1    2    3    4    5

Si Ud. Tiene algún comentario adicional que desee manifestarlo, escríbalo en este espacio.

Fuente: elaboración Propia