



Universidad
Inca Garcilaso de la Vega
Nuevos Tiempos. Nuevas Ideas

Facultad de Ingeniería de Sistemas, Cómputo y Telecomunicaciones

**Implementación de un Datamart para la mejora en la toma de
decisiones en el control de la demanda eléctrica del Comité de
Operaciones Económicas del Sistema Interconectado Nacional**

Tesis para optar el Título de Ingeniero de Sistemas y Cómputo

Bach. Juan Eduardo Zamora Saldarriaga

1964

Asesor

MSc. Christian Almóguer Martínez

Lima – Perú

Noviembre de 2017

Dedicatoria

Este trabajo está dedicado a Dios por la vida y sus bendiciones, en especial por mis padres por todo el amor y apoyo incondicional que me dan, a mis abuelos por el cariño y por compartirme la experiencia que ganaron a través de su vida, a toda mi familia y amigos quienes siempre están a mi lado muchas gracias por todo.



ÍNDICE

ÍNDICE DE FIGURAS.....	6
ÍNDICE DE TABLAS.....	11
RESUMEN.....	13
ABSTRACT	14
INTRODUCCIÓN.....	15
CAPÍTULO 1: PLANTEAMIENTO DEL PROBLEMA.....	17
1.1 Situación Problemática	17
1.2 Problema de la investigación.....	18
1.3 Objetivos.....	19
1.4 Justificación.....	19
1.5 Alcance.....	21
CAPÍTULO 2: MARCO TEÓRICO.....	22
2.1 Antecedentes de la investigación.....	22
2.2 Bases teóricas.....	24
2.2.1 On Line Analytical Processing (OLAP).....	24
2.2.2 Business Intelligence.....	27
2.2.3 Datamart.....	30
2.2.4 Metodología de Ralph Kimball.....	40
2.2.5 Demanda Eléctrica	48
2.2.6 Modelo de calidad de producto de datos: ISO/IEC 25012	49
2.3 Glosario de términos	53
CAPÍTULO 3: VARIABLES E HIPÓTESIS	55

3.1	Variables e Indicadores	55
3.2	Hipótesis	55
3.3	Matriz de coherencia interna	56
CAPÍTULO 4: METODOLOGÍA DE DESARROLLO		57
4.1	Comparación de metodologías para desarrollo de un Datamart	57
4.2	Metodología de Ralph Kimball	61
4.2.1	Etapas de la metodología.....	61
4.2.2	Artefactos	64
CAPÍTULO 5: SOLUCIÓN TECNOLÓGICA		66
5.1	Planificación del proyecto	66
5.2	Definición de requerimientos del negocio	70
5.3	Diseño de la arquitectura técnica	70
5.4	Diseño físico	74
5.5	Modelado dimensional	74
5.6	Diseño e implementación del subsistema de ETL	80
5.7	Especificación de aplicaciones de BI	106
5.8	Desarrollo de aplicaciones de BI	107
5.8.1	Desarrollo mediante la herramienta Power Pivot.....	107
5.8.2	Desarrollo mediante la herramienta Power BI.....	107
5.9	Implementación	110
5.9.1	Implementación con la herramienta Power Pivot	110
5.9.2	Explotación de datos con Power BI	112
CAPÍTULO 6: RESULTADOS		114
CONCLUSIONES		123
RECOMENDACIONES		124

REFERENCIAS BIBLIOGRÁFICAS.....125

ANEXO129



ÍNDICE DE FIGURAS

<i>Figura 1: Demanda ejecutada y programada en el día de la máxima demanda del 2016</i>	17
<i>Figura 2: Aplicación de herramientas analíticas por área de negocio</i>	20
<i>Figura 3: Interacción de usuario con cubo multidimensional</i>	25
<i>Figura 4: Aplicaciones de Business Intelligence</i>	30
<i>Figura 5: Relación entre Datamart y Datawarehouse</i>	31
<i>Figura 6: Formación de los modelos relacional y multidimensional</i>	33
<i>Figura 7: Representación de dimensiones en los ejes de un cubo OLAP</i>	34
<i>Figura 8: Representación de una medida dentro de un cubo OLAP</i>	34
<i>Figura 9: Esquema en estrella</i>	35
<i>Figura 10: Esquema en copo de nieve</i>	35
<i>Figura 11: Concepto de jerarquía por Ubicación</i>	36
<i>Figura 12: Concepto de jerarquía por Tiempo</i>	37
<i>Figura 13: Escenarios de uso del lenguaje MDX</i>	38
<i>Figura 14: Query en lenguaje MDX para consultar un cubo multidimensional</i>	40
<i>Figura 15: Estrategia de procesamiento de datos según metodología Ralph Kimball</i>	41
<i>Figura 16: Flujo del ETL (Extract, Transform and Load)</i>	46
<i>Figura 17: Diagrama del ciclo de vida de la metodología Kimball</i>	62
<i>Figura 18: Flujo de las etapas para el desarrollo del proyecto</i>	69
<i>Figura 19: Diagrama de arquitectura para el desarrollo del proyecto</i>	71
<i>Figura 20: Diagrama de arquitectura de las tecnologías implementadas</i>	73
<i>Figura 21: Herramientas tecnológicas utilizadas para el desarrollo del datamart</i>	74
<i>Figura 22: Nivel de granularidad de la dimensión Horario</i>	75
<i>Figura 23: Nivel de granularidad de la dimensión Tiempo</i>	75
<i>Figura 24: Nivel de granularidad de la dimensión Ubigeo</i>	75
<i>Figura 25: Nivel de granularidad de la dimensión Sector</i>	75
<i>Figura 26: Nivel de granularidad de la dimensión Actividad Económica</i>	76

<i>Figura 27: Nivel de granularidad de la dimensión Usuario Libre OS</i>	76
<i>Figura 28: Nivel de granularidad de la dimensión Usuarios Libres COES</i>	76
<i>Figura 29: Nivel de granularidad de la dimensión generador</i>	77
<i>Figura 30: Modelo físico de la Base de Datos COES – Replica y Provisional</i>	78
<i>Figura 31: Modelo físico de Datamart</i>	79
<i>Figura 32: Proceso de preparación de la información para la carga en el modelo dimensional</i>	82
<i>Figura 33: Proceso ETL con la herramienta SQL Server Data Tools - Inicio</i>	83
<i>Figura 34: Script SQL para limpiar datamart en el proceso de ETL</i>	84
<i>Figura 35: Script SQL para crear datamart en el proceso de ETL (Fact_Demanda_Extranet y Dim_Usuario_Libre_OS)</i>	84
<i>Figura 36: Data flow para cargar dimensión Actividad Económica</i>	85
<i>Figura 37: Mapeo de origen a destino en la dimensión actividad económica</i>	85
<i>Figura 38: Data flow para cargar dimensión Tiempo</i>	85
<i>Figura 39: Mapeo de origen a destino en la dimensión Tiempo</i>	86
<i>Figura 40: Data flow para cargar dimensión Horario</i>	86
<i>Figura 41: Mapeo de origen a destino en la dimensión Horario</i>	86
<i>Figura 42: Data flow para cargar dimensión Sector</i>	87
<i>Figura 43: Mapeo de origen a destino en la dimensión Sector</i>	87
<i>Figura 44: Data flow para cargar dimensión Ubigeo</i>	87
<i>Figura 45: Mapeo de origen a destino en la dimensión Ubigeo</i>	87
<i>Figura 46: Data flow para cargar dimensión Usuarios Libres OS</i>	88
<i>Figura 47: Mapeo de origen a destino en la dimensión Usuarios Libres OS</i>	88
<i>Figura 48: Data flow para cargar la dimensión Usuarios Libres COES</i>	88
<i>Figura 49: Sub proceso de conversión para el campo razón social en la dimensión Usuarios Libres COES</i>	89
<i>Figura 50: Mapeo de origen a destino en la dimensión Usuarios Libres COES</i>	89
<i>Figura 51: Data flow para cargar la dimensión Generador</i>	90
<i>Figura 52: Mapeo de origen a destino en la dimensión Generador</i>	90
<i>Figura 53: Data flow para cargar la tabla de hechos Máxima Demanda</i>	90
<i>Figura 54: Mapeo de origen a destino en la tabla de hechos Máxima Demanda</i>	90
<i>Figura 55: Data flow para cargar la tabla de hechos MAPE</i>	91

<i>Figura 56: Mapeo de origen a destino en la tabla de hechos MAPE</i>	91
<i>Figura 57: Data flow para cargar la tabla de hechos PBI</i>	91
<i>Figura 58: Mapeo de origen a destino en la tabla de hechos PBI</i>	91
<i>Figura 59: Data flow para cargar la tabla de hechos Demanda SCADA</i>	92
<i>Figura 60: Mapeo de origen a destino en la tabla de hechos Demanda SCADA</i>	92
<i>Figura 61: Data flow para cargar la tabla de hechos Demanda OSINERGMIN</i>	92
<i>Figura 62: Mapeo de origen a destino en la tabla de hechos Demanda OSINERGMIN</i>	93
<i>Figura 63: Data flow para cargar la tabla de hechos Demanda Extranet</i>	93
<i>Figura 64: Mapeo de origen a destino en la tabla de hechos Demanda Extranet</i>	93
<i>Figura 65: Data flow para cargar la tabla de hechos Medidores</i>	94
<i>Figura 66: Mapeo de origen a destino en la tabla de hechos Medidores</i>	94
<i>Figura 67: Proceso ETL con la herramienta SQL Server Data Tools al finalizar</i>	95
<i>Figura 68: Tabla Dim_Actividad_Economica del Datamart</i>	95
<i>Figura 69: Tabla Dim_Tiempo del Datamart</i>	96
<i>Figura 70: Tabla Dim_Horario del Datamart</i>	96
<i>Figura 71: Tabla Dim_Sector del Datamart</i>	96
<i>Figura 72: Tabla Dim_Ubigeo del Datamart</i>	97
<i>Figura 73: Tabla Dim_Usuario_Libre_COES del Datamart</i>	97
<i>Figura 74: Tabla Dim_Usuario_Libre_OS del Datamart</i>	97
<i>Figura 75: Tabla Dim_Generador del Datamart</i>	98
<i>Figura 76: Tabla Fact_Demanda_Extranet del Datamart</i>	98
<i>Figura 77: Tabla Fact_Demanda_OSINERGMIN del Datamart</i>	98
<i>Figura 78: Tabla Fact_Demanda_SCADA del Datamart</i>	99
<i>Figura 79: Tabla Fact_MAPE del Datamart</i>	99
<i>Figura 80: Tabla Fact_Maxima_Demanda del Datamart</i>	99
<i>Figura 81: Tabla Fact_Medidores del Datamart</i>	100
<i>Figura 82: Jerarquía de la dimensión Tiempo</i>	100
<i>Figura 83: Jerarquía de la dimensión Horario</i>	100
<i>Figura 84: Jerarquía de la dimensión Sector</i>	101

<i>Figura 85: Jerarquía de la dimensión Ubigeo</i>	101
<i>Figura 86: Jerarquía de la dimensión Actividad Económica</i>	101
<i>Figura 87: Jerarquía de la dimensión Usuario Libre OS</i>	101
<i>Figura 88: Jerarquía de la dimensión Usuario Libre COES</i>	102
<i>Figura 89: Jerarquía de la dimensión Generador</i>	102
<i>Figura 90: Cubo PBI en la herramienta SQL Server Data Tools</i>	102
<i>Figura 91: Cubo Demanda SCADA en la herramienta SQL Server Data Tools</i>	103
<i>Figura 92: Cubo MAPE en la herramienta SQL Server Data Tools</i>	103
<i>Figura 93: Cubo Demanda OSINERGMIN en la herramienta SQL Server Data Tools</i>	104
<i>Figura 94: Cubo Máxima Demanda en la herramienta SQL Server Data Tools</i>	104
<i>Figura 95: Cubo Demanda Extranet en la herramienta SQL Server Data Tools</i>	105
<i>Figura 96: Cubo Medidores en la herramienta SQL Server Data Tools</i>	106
<i>Figura 97: Selección de fuente de datos en Power BI Desktop</i>	108
<i>Figura 98: Conexión al datamart desde Power BI Desktop</i>	108
<i>Figura 99: Selección de tablas para obtener datos en Power BI Desktop</i>	108
<i>Figura 100: Tablas cargadas en Power BI Desktop</i>	109
<i>Figura 101: Reporte gráfico en relación a los registros de máxima demanda en el SEIN</i>	109
<i>Figura 102: Datos del modelo dimensional (datamart) desde Power Pivot</i>	110
<i>Figura 103: Modelo de datos de las tablas obtenidas desde Power Pivot</i>	111
<i>Figura 104: Reporte en tablas dinámicas del datamart desde Power Pivot</i>	111
<i>Figura 105: Reporte gráfico de Power BI embebido en HTML</i>	112
<i>Figura 106: Reporte gráficos de Power BI en la web</i>	113
<i>Figura 107: Reporte de registros de máxima demanda desde Power BI Mobile</i>	113
<i>Figura 108: Gráfico estadístico de la relación entre los encuestados y la información utilizada</i>	114
<i>Figura 109: Gráfico estadístico de respuestas de la segunda pregunta de la encuesta</i>	115
<i>Figura 110: Gráfico estadístico de respuestas de la tercera pregunta de la encuesta</i>	115
<i>Figura 111: Gráfico estadístico de respuestas de la cuarta pregunta de la encuesta</i>	116
<i>Figura 112: Gráfico estadístico de respuestas de la quinta pregunta de la encuesta</i>	117
<i>Figura 113: Gráfico estadístico de respuestas de la sexta pregunta de la encuesta</i>	118

Figura 114: Gráfico estadístico de respuestas de la séptima pregunta de la encuesta.....118
Figura 115: Gráfico estadístico de respuestas de la octava pregunta de la encuesta119
Figura 116: Gráfico estadístico de respuestas de la novena pregunta de la encuesta.....120
Figura 117: Gráfico estadístico de respuestas de la décima pregunta de la encuesta.....121
Figura 118: Gráfico estadístico de respuestas de la undécima pregunta de la encuesta.....121



ÍNDICE DE TABLAS

<i>Tabla 1: Diferencias entre los tipos de OLAP</i>	26
<i>Tabla 2: Elementos de OLAP</i>	27
<i>Tabla 3: Tipos de datamarts</i>	31
<i>Tabla 4: Funciones de lenguaje MDX</i>	39
<i>Tabla 5: Características inherentes del sistema según ISO/IEC 25012</i>	50
<i>Tabla 6: Características dependientes del sistema según ISO/IEC 25012</i>	50
<i>Tabla 7: Características inherentes y dependientes del sistema según ISO/IEC 25012</i>	51
<i>Tabla 8: Correspondencia entre las características del ISO 9126 e ISO/IEC 25012</i>	52
<i>Tabla 9: Comparación de metodologías para el desarrollo de estructuras dimensionales</i>	57
<i>Tabla 10: Niveles de evaluación para el análisis comparativo de metodologías</i>	58
<i>Tabla 11: Evaluación general de metodologías propuestas</i>	58
<i>Tabla 12: Evaluación de metodologías propuestas según características</i>	59
<i>Tabla 13: Niveles de evaluación de repuesta para el análisis comparativo de metodologías</i>	59
<i>Tabla 14: Evaluación de metodología propuesta según criterio en la relación con el proyecto</i>	60
<i>Tabla 15: Evaluación de metodología propuesta según criterios generales del proyecto</i>	60
<i>Tabla 16: Cronograma de tareas identificadas para el desarrollo del Datamart</i>	68
<i>Tabla 17: Descripción del diagama de arquitectura para el desarrollo del Datamart</i>	73
<i>Tabla 18: Relación entre la información y los sistemas transaccionales</i>	82
<i>Tabla 19: Herramientas para el acceso a la información del datamart</i>	107
<i>Tabla 20: Respuestas de la encuesta para el indicador de accesibilidad – Pregunta 2</i>	114
<i>Tabla 21: Respuestas de la encuesta para el indicador de accesibilidad – Pregunta 3</i>	115
<i>Tabla 22: Respuestas de la encuesta para el indicador de fiabilidad – Pregunta 4</i>	116
<i>Tabla 23: Respuestas de la encuesta para el indicador de fiabilidad – Pregunta 5</i>	117
<i>Tabla 24: Respuestas de la encuesta para el indicador de eficiencia – Pregunta 6</i>	117
<i>Tabla 25: Respuestas de la encuesta para el indicador de eficiencia – Pregunta 7</i>	118

Tabla 26: Respuestas de la encuesta para el indicador de eficiencia – Pregunta 8119

Tabla 27: Respuestas de la encuesta para el indicador de disponibilidad – Pregunta 9120

Tabla 28: Respuestas de la encuesta para el indicador de disponibilidad – Pregunta 10120

Tabla 29: Respuestas de la encuesta para el indicador de disponibilidad – Pregunta 11121



RESUMEN

El Comité de Operaciones Económicas del Sistema Interconectado Nacional (COES) trabaja diariamente con información operativa histórica que es registrada en una base de datos transaccional, para poder analizar los datos obtenidos del Sistema Eléctrico Interconectado Nacional (SEIN) con el fin de sobrellevar un nivel óptimo de la demanda eléctrica. El proceso de obtención de información para el análisis no llegaba a ser el más óptimo, debido a que el almacenamiento de los datos no se encontraba bajo un enfoque de análisis que permita el aprovechamiento del gran volumen de información histórica que el COES maneja en relación a la demanda eléctrica, lo que originaba un sobre esfuerzo generado por un tiempo de espera y varios procesos de descarga para obtener la información histórica que uno deseaba analizar, así como también el tiempo que tomaba estructurar la información para generar los reportes gráficos; por estas razones, se optaba por manejar la información histórica descargada en hojas de cálculo, lo cual no permitía mantener una trazabilidad frente a las actualizaciones que se generaban, ya que era necesario volver a descargar la información para acoplar los cambios. La finalidad del presente trabajo es la implementación de un datamart para la mejora en la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional; para ello se empleó la metodología de Ralph Kimball, y las herramientas tecnológicas Power BI y Power Pivot para entregar al usuario el acceso a la información del datamart. Según los resultados obtenidos mediante la encuesta elaborada en el presente trabajo de investigación se demostró que el datamart cumple con los indicadores de accesibilidad, fiabilidad, eficiencia y disponibilidad. El datamart influyó satisfactoriamente para la mejora en la toma de decisiones del control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional, asimismo la accesibilidad, fiabilidad, usabilidad y eficiencia del datamart propuesto, influyeron positivamente para el cumplimiento de los procesos.

Palabras clave: Datamart, Demanda eléctrica, Metodología Ralph Kimball, Toma de decisiones.

ABSTRACT

The Committee of Economic Operations of the National Interconnected System (COES) works daily with historical operative information that is registered in a transactional database, in order to analyze the data obtained from the National Interconnected Electric System (SEIN) in order to overcome an optimal level of electricity demand. The process of obtaining information for the analysis was not the most optimal, because the storage of the data was not under an analysis approach that allows the use of the large volume of historical information that the COES manages in relation to the electrical demand, which caused an overexertion generated by a waiting time and several download processes to obtain the historical information that one wanted to analyze, as well as the time it took to structure the information to generate the graphic reports; for these reasons, it was decided to handle the historical information downloaded in spreadsheets, which did not allow to maintain a traceability against the updates that were generated, since it was necessary to download the information again to match the changes. The purpose of the present work is the implementation of a datamart for the improvement in the decision making in the control of the electrical demand of the Committee of Economic Operations of the National Interconnected System; For this purpose, Ralph Kimball's methodology and the Power BI and Power Pivot technology tools were used to provide the user with access to the datamart information. According to the results obtained through the survey developed in the present research work, it was demonstrated that the datamart complies with the indicators of accessibility, reliability, efficiency and availability. The datamart successfully influenced the decision making of the electric demand control of the Economic Operations Committee of the National Interconnected System, as well as the accessibility, reliability, usability and efficiency of the proposed datamart, positively influencing the compliance of the processes.

Keywords: Datamart, Electric demand, Ralph Kimball Methodology, Decision making.

INTRODUCCIÓN

Las organizaciones han utilizado diferentes tipos de recursos con lo que buscan un incremento de productividad y mejora continua; debido a los nuevos paradigmas tecnológicos existentes, se ven obligados a utilizar uno de los elementos que con el tiempo ha ido en crecimiento, no solo en cantidad, sino también en relevancia para la toma de decisiones de las empresas, los datos. A este fenómeno informático, que consta en el aprovechamiento de grandes volúmenes de información para la toma de decisiones, se le ha denominado Business Intelligence (BI), el cual no solo es un término, sino que muchos revistas e informes reafirman la importancia de dicho paradigma tecnológico para las empresas, debido a que permite obtener grandes resultados basados en distintos casos de éxito.

El sector eléctrico, el cual forma parte de nuestra vida cotidiana, conforma una de las grandes redes operativas a nivel nacional; existen diversos organismos que de manera conjunta se encargan de lograr que el SEIN, se encuentre estable y en un buen nivel operativo, teniendo en cuenta que éste no es un sistema estático, al contrario, éste se encuentra en un crecimiento continuo. El COES, “es un organismo técnico cuya finalidad es coordinar la operación de corto, mediano y largo plazo del SEIN al mínimo coste, preservando la seguridad del sistema y el mejor aprovechamiento de los recursos energéticos” (Comité de Operaciones Económicas del Sistema Interconectado Nacional, 2016). Diariamente, la información operativa es registrada en la base de datos que el COES emplea como repositorio de almacenamiento, tanto de información en tiempo real, como de data periódica que es cargada por los especialistas. Así mismo, esta información es empleada para generar reportes (gráficos o tablas) que se muestran en el portal web de la organización, por otro lado, hay informes que son entregados a la alta dirección del COES y distribuidos a las jefaturas que también necesitan tener la información sintetizada; la Sub Dirección de Gestión de Información (SGI), es un área, que trabaja con grandes volúmenes de datos históricos, entre la información que manejan tenemos la producción de energía en el SEIN, la máxima demanda declarada en medidores de bornes de generación, la demanda ejecutada y programada, diagramas unifilares de las instalaciones del SEIN, diagramas topológicos de centrales hidroeléctricas, entre otra información, las cuales provienen de diversas fuentes como pueden ser la declaración de información de los Agentes por medio del Extranet, registros de señales SCADA, archivos enviados vía correo por los agentes del COES y registro de la base de datos del COES. En relación con todo lo mencionado, la Sub Dirección de Gestión de Información (SGI) elabora una estadística anual, boletines mensuales, reportes e informes, los cuales tienen como finalidad mostrar diversos indicadores del estado actual del SEIN, en los cuales se realizan comparaciones entre la data de diferentes años. También, se realiza el seguimiento necesario a diversos indicadores, uno de ellos es el caso de la demanda de Usuarios Libres, para realizar éste proceso, necesitan identificar a los Usuarios Libres que han excedido el nivel de potencia establecido (10 MW), y para ello utilizan la información que el Organismo Supervisor de la Inversión en Energía y Minería elabora y publica en su portal web, para lo cual deben realizar una descarga de los archivos correspondientes y posterior a ello trabajar dichos datos en hojas de cálculo. En

el Anexo I se puede apreciar el flujograma de dicho proceso, así mismo, de forma complementaria, en los Anexos II, III, IV y V se puede observar el flujograma de los procesos realizados por SGI. Con toda esta información histórica que se genera, llevar el control y almacenamiento con la estructura empleada, genera un esfuerzo adicional. Sin embargo, al aprovechar las herramientas que existen, podemos generar un repositorio de los datos históricos que se tienen, y posterior a una correcta adecuación, se logrará generar los reportes visuales necesarios, explotando la data mediante un análisis detallado y entregando al usuario las herramientas necesarias que le permitirán trabajar con la información correctamente actualizada, con ello se logrará evitar el uso de horas/hombres en procesos que pueden ser automatizados; según lo mencionado se determina que es necesario el desarrollo de una estructura dimensional, en éste caso un datamart.

La presente tesis consta de los siguientes capítulos:

- **Capítulo 1: Planteamiento del problema;** tiene por finalidad describir la situación problemática, el problema general, los problemas específicos, objetivo general y objetivos específicos.
- **Capítulo 2: Marco teórico;** consta del desarrollo del marco teórico, el cual comprende los antecedentes, las bases teóricas y el glosario.
- **Capítulo 3: Variables e hipótesis;** se definen las variables, indicadores e hipótesis del trabajo de investigación.
- **Capítulo 4: Metodología de desarrollo;** se describe la metodología de desarrollo a utilizar para el desarrollo de la solución tecnológica.
- **Capítulo 5: Solución tecnológica;** se describe el desarrollo de la solución tecnológica, que comprende la realización de las distintas etapas de la metodología.
- **Capítulo 6: Resultados;** se describen los resultados obtenidos.

Finalmente se presentan la conclusiones, recomendaciones y referencias bibliográficas.

CAPÍTULO 1: PLANTEAMIENTO DEL PROBLEMA

En el primer capítulo se presenta la situación problemática existente sobre la cual se ha desarrollado la tesis, así mismo, se señalan los problemas de la investigación y los objetivos de la solución tecnológica desarrollada; también se define el alcance de la tesis y la justificación del trabajo de investigación en relación con la problemática planteada.

1.1 Situación Problemática

El sector eléctrico, está compuesto por diversas empresas (generadores, transmisores, distribuidores y usuarios libres), así mismo, en cada país hay una organización que cumple el rol de operador eléctrico; desde una perspectiva a nivel internacional, observando los casos de Europa y Latinoamérica, el consumo de energía eléctrica (demanda eléctrica) es un indicador diario que permite observar el comportamiento de los sistemas eléctricos de cada país, y ello influye directamente en la generación de energía, por ésta razón, los organismos encargados de la operación eléctrica a nivel nacional, realizan el cálculo de una demanda prevista para entregar una base de referencia a las instalaciones de generación con el fin que la demanda sea satisfactoriamente cubierta, dicha programación es realizada para distintos periodos (corto plazo, mediano plazo y largo plazo) lo que facilita realizar las coordinaciones respectivas y mantener un óptimo desarrollo sostenible del sistema eléctrico. En la Figura 1, podemos observar el comportamiento de la demanda ejecutada en comparación de la programada en el Sistema Interconectado Nacional del Perú, en el día de la máxima demanda del año 2016.

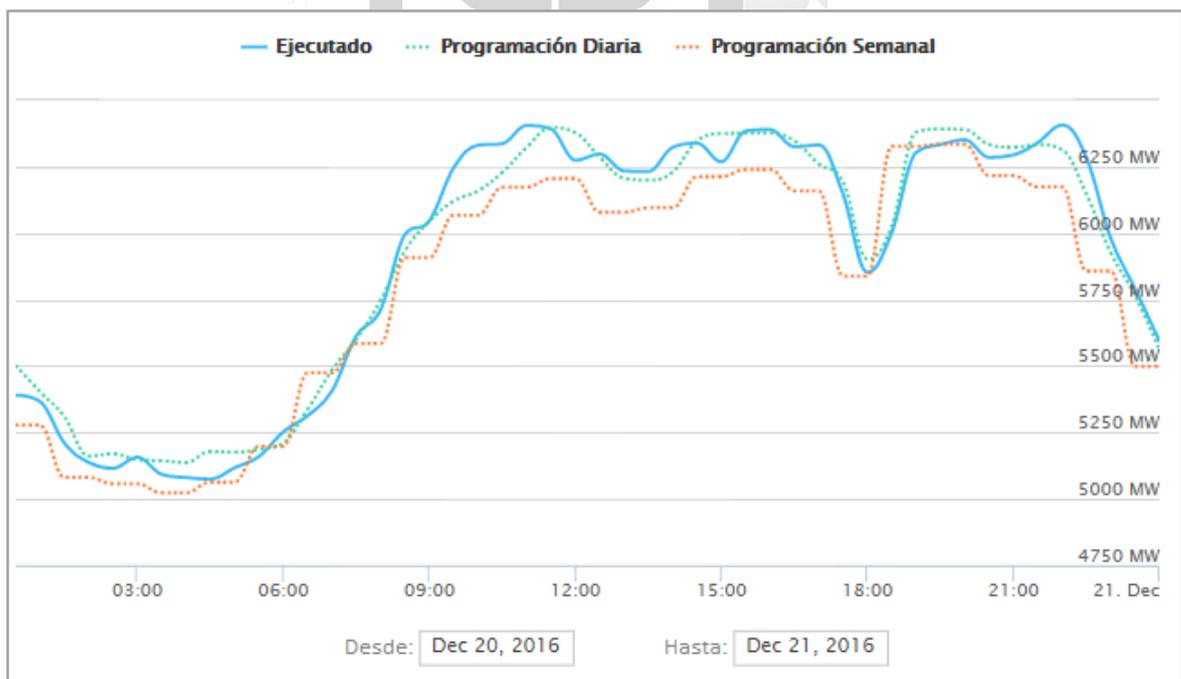


Figura 1: Demanda ejecutada y programada en el día de la máxima demanda del 2016
Fuente: Comité de Operaciones Económicas del Sistema Interconectado Nacional. (2017)

En Perú, la función de operador eléctrico la cumple el Comité de Operaciones Económicas del Sistema Interconectado Nacional (COES), para ello, trabaja diariamente con información operativa, la cual es registrada en la base de datos que el COES emplea como repositorio, el cual cuenta con información histórica, pero ésta no se encuentra bajo un enfoque de análisis, generando que el almacenamiento y el seguimiento de los resultados obtenidos mediante los indicadores obligue a realizar un esfuerzo adicional en la preparación de las fuentes de información; así mismo, se conoce, que para la obtención de dichos indicadores, primero se realiza la descarga de un gran volumen de información, para luego transformarla y recién empezar a trabajarla, lo que va a ir dificultándose en cuanto a la cantidad de periodos de tiempo (años, meses, semanas, trimestres) que uno desee analizar, ya que ello conlleva un mayor número de descargas, lo cual genera un largo tiempo de espera para que toda la información sea comprimida en un solo archivo.

El esfuerzo empleado previo a la elaboración de los diversos reportes debido a la actual tecnología empleada para el manejo de datos, no es el más óptimo, ya que hoy en día existen diversas alternativas tecnológicas (datawarehouse, datalakes, datamarts, etc.), que siendo correctamente desarrolladas, son capaces de resolver el problema y optimizar los procesos, dando la capacidad en tomar decisiones y de realizar el respectivo control que se maneja de manera interna en el Comité de Operaciones Económicas del Sistema Interconectado Nacional teniendo un apoyo o sustento definido y concreto, garantizando una correcta elección.

1.2 Problema de la investigación

Problema General:

¿En qué medida la implementación de un datamart influye en la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional?

Problemas Específicos:

- ¿En qué medida el nivel de accesibilidad de un datamart influye en la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional?
- ¿En qué medida el nivel de fiabilidad de un datamart influye en la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional?
- ¿En qué medida el nivel de eficiencia de un datamart influye en la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional?
- ¿En qué medida el nivel de disponibilidad de un datamart influye en la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional?

1.3 Objetivos

Objetivo General

Determinar la influencia de un datamart para la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional.

Objetivos Específicos

- Determinar el nivel de influencia de la accesibilidad de un Datamart para la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional.
- Determinar el nivel de influencia de la fiabilidad de un Datamart para la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional.
- Determinar el nivel de influencia de la eficiencia de un Datamart para la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional.
- Determinar el nivel de influencia de la disponibilidad de un Datamart para la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional.

1.4 Justificación

El desarrollo del trabajo de la tesis tiene como finalidad mejorar la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional mediante la implementación de un datamart, que almacene adecuadamente la información histórica de la demanda eléctrica de los Usuarios Libres permitiendo observar su comportamiento desde distintas dimensiones, debido a su relevancia para procesos operacionales de corto, mediano y largo plazo. Así mismo, éste trabajo de investigación demuestra la importancia del manejo de grandes volúmenes de datos históricos para mejorar la toma de decisiones en base al control de la demanda eléctrica y obtener mejoras en el desarrollo de distintos procesos.

Con la presente tesis, se brinda una solución tecnológica a los usuarios para acceder a la información histórica relacionada a la demanda con la estructura adecuada, la cual les permita a los especialistas del negocio disponer de dichos datos. El desarrollo del trabajo de investigación permite generar una mayor cantidad de reportes gráficos a comparación de los que se manejaban, los cuáles pueden ser consultados por diferentes dimensiones (tiempo, zona geográfica, empresa, entre otras) que ayudarán a parametrizar la información que se desea obtener y visualizar. Con ello se logró obtener un beneficio con la disminución de tiempos que toma realizar ciertos procesos (control de la demanda eléctrica, generar reportes para el análisis de la información, obtener información histórica de la demanda eléctrica, medidores de generación y demanda en señales

SCADA) y a su vez permite una óptima administración de la información para el control de la demanda eléctrica que el COES realiza.

Por otro lado, el caso de estudio de la presente tesis se relaciona con la operación del sector eléctrico; según Salvador, Fabrizio (2014) en la Figura 2 observamos el nivel de tendencia a utilizar aplicaciones de herramientas analíticas según el área del negocio, lo que nos permite identificar el alto nivel de rendimiento que posee el área de estudio (operaciones y producción), por debajo del área de Gestión financiera y presupuestos, ello nos ayuda a justificar la implementación de la propuesta tecnológica desarrollada.

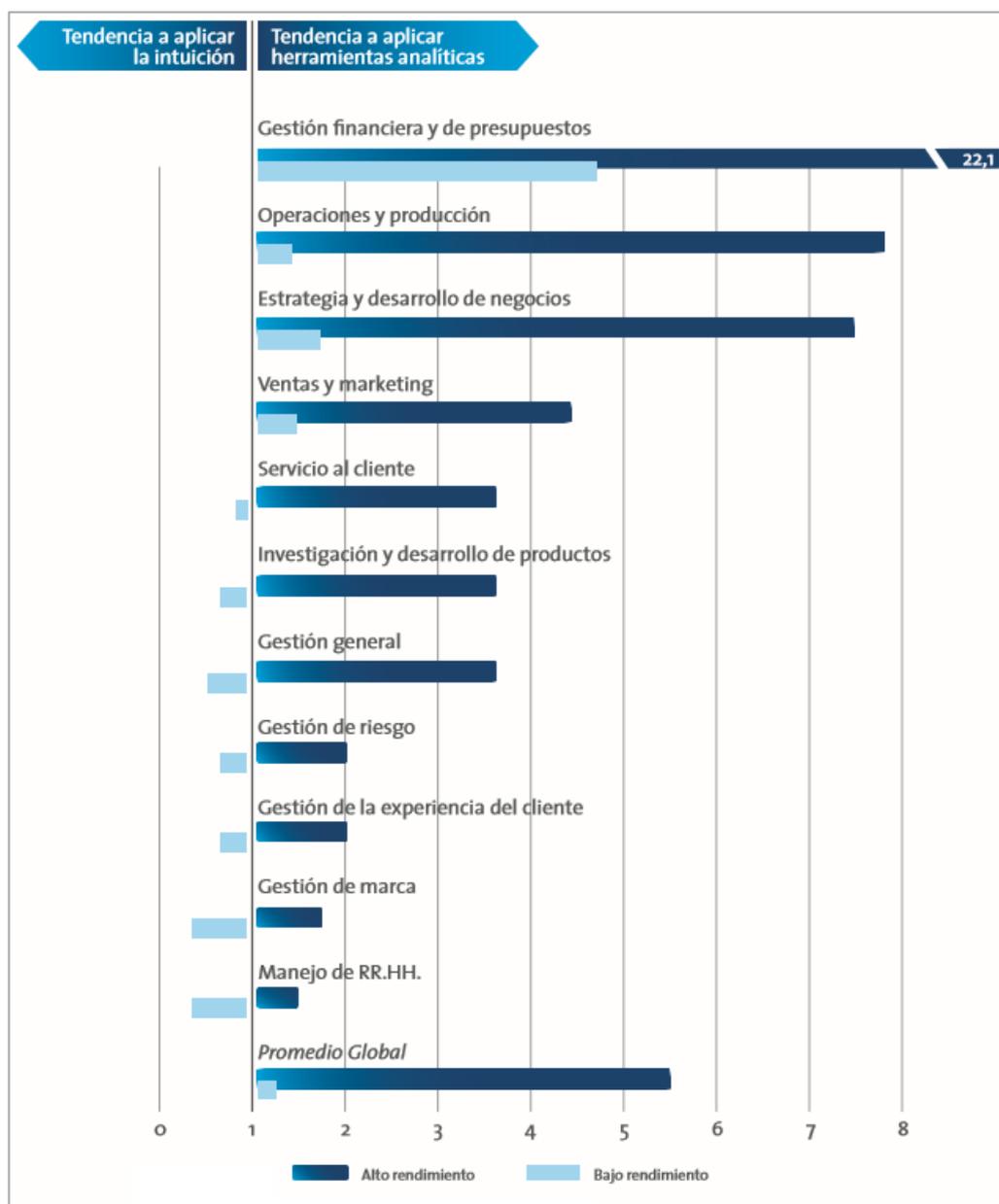


Figura 2: Aplicación de herramientas analíticas por área de negocio
Fuente: Salvador, Fabrizio. (2014)

1.5 Alcance

La presente tesis se ha desarrollado con la finalidad de obtener un repositorio de datos dimensional, el cual resuelva los problemas planteados en el presente trabajo de investigación. Mediante la investigación se realizaron las siguientes actividades:

- Estudio del problema de la investigación.
- Análisis de las metodologías para Business Intelligence.
- Implementación de la solución tecnológica.
- Implementación de herramientas de acceso al repositorio multidimensional, de tal manera que utilicen la información histórica almacenada en el datamart.
- Emisión de reportes gráficos estandarizados.



CAPÍTULO 2: MARCO TEÓRICO

2.1 Antecedentes de la investigación

- González Marroquín, Horacio (2012). *Inteligencia de Negocios en el Desarrollo de Sistemas de Monitoreo de Mercado para el Sector Eléctrico*. Tesis para obtener el grado de Magister. Pontificia Universidad Católica de Chile. Santiago de Chile, Chile.

Los autores sostienen que el uso de sistemas de monitoreo de mercado en el sector eléctrico se ha convertido en un tema con mayor influencia para el control del ejercicio de poder de mercado, la eficiencia y la seguridad en la operación. Desafortunadamente, el desarrollo de dichos sistemas normalmente ha sido basado en extensiones de plataformas informáticas propias del sector, que no han logrado abordar efectivamente las dificultades de consolidación y estructuración de grandes volúmenes de datos, provenientes de fuentes dispersas, así como su efectiva visualización y análisis. En este sentido, las metodologías y herramientas de inteligencia de negocios muestran un gran potencial como solución, por su gran madurez técnica y especialización en el manejo de información. El objetivo de esta investigación es mostrar la aplicabilidad del uso de herramientas y metodologías de BI en el desarrollo de los sistemas de monitoreo en mención. Se propone una metodología estructurada de construcción de los respectivos sistemas, utilizando BI como plataforma de desarrollo. Los autores concluyen en que la utilización de BI facilita la obtención y consolidación de los datos del mercado eléctrico, permitiendo destinar menos tiempo en su recolección y extracción, para enfocarse más en el análisis.

- Mencías Imbaquingo, James y Mencías Imbaquingo, Amy (2012). *Desarrollo de un aplicativo Business Intelligence para el área administrativo financiera de la Empresa Eléctrico “Quito” S.A.* Tesis para obtener el título de Ingeniero. Universidad de las Fuerzas Armadas - Escuela Politécnica del Ejército. Sangolquí, Ecuador.

Los autores proponen solventar los problemas directivos y operativos detectados, mediante el desarrollo de un aplicativo Business Intelligence, donde se plantea fusionar los datos con el negocio para obtener información relevante sobre la situación de la empresa; para ello se utilizará un sistema de inteligencia de negocio formado por los datos que se extraigan de las bases de datos de producción existentes, de esta forma se busca facilitar la accesibilidad a la información, apoyo en la toma de decisiones y la orientación al usuario final; a su vez, para conseguir una amplia reducción de costos se ha optado por utilizar un datamart de tipo OLAP, que será construido según los requisitos de cada área o departamento, permitiendo de ésta manera facilitar la identificación de los usuarios potenciales y el acceso a los datos que se

necesitan frecuentemente; se busca facilitar la administración de la información y permitir la personalización en la obtención de la misma en base a indicadores específicos por parte de los usuarios, de esta forma se evitará la existencia de varios reportes estáticos que proporcionan la misma información. Para finalizar, los autores afirman que, mediante el desarrollo de una aplicación BI, se generó información analítica y estratégica basada en los requerimientos obtenidos de los usuarios directivos y el equipo de desarrollo que posee conocimiento en cada uno de los requerimientos involucrados.

- Matallana Caffo, Carlos y Vivanco Valdez, Edwin (2014). *Diseño de un Datamart para el proceso de inscripción en el registro de hidrocarburos de Osinergmin utilizando la metodología Kim Ball*. Tesis para obtener el título de Ingeniero. Universidad Privada Antenor Orrego. Trujillo, Perú.

Los autores sostienen que el Organismo Supervisor de la Inversión en Energía y Minería - OSINERGMIN es una institución pública encargada de regular y supervisar que las empresas del sector eléctrico, hidrocarburos y minero cumplan las disposiciones legales de las actividades que desarrollan, que consta de 4 áreas, tales como: electricidad, hidrocarburos, gas natural y minería. Los datos registrados mediante la producción diaria, correspondientes al área de hidrocarburos, son almacenados en una base de datos transaccional, el almacenamiento y la organización de los datos históricos son migrados a tablas dinámicas elaboradas en Microsoft Excel. La institución presenta el problema con la elaboración de reportes multidimensionales a partir de la organización y almacenamiento de los datos históricos. OSINERGMIN, tomó la decisión de elaborar el diseño de un datamart para mejorar el almacenamiento y organización de sus datos con la finalidad de que sus reportes multidimensionales resulten eficientes para el área de hidrocarburos. Para el desarrollo de esta solución, se utilizó software libre destinados a gestionar datos e información y para el diseño del datamart se empleó MySQL Workbench. Obteniéndose un mejor almacenamiento y una buena organización de los datos históricos y la creación y elaboración de los reportes multidimensionales basados en los datos históricos del datamart.

- Peña Gutiérrez, Gonzalo y Pincheira Henríquez, Iván (2014). *Implementación de Business Intelligence para una Pyme local del rubro eléctrico*. Tesis para obtener el título de Ingeniero. Universidad del Bío-Bío. Concepción, Chile.

Los autores presentan la implementación de Business Intelligence en el área de ventas, para VALFI, una pyme comercializadora del rubro eléctrico. A partir de una serie de entrevistas con los dueños y empleados de VALFI se consiguió recolectar los requisitos para la construcción del diseño del datamart. En una etapa previa a los procesos ETL y luego de

realizado un análisis de las herramientas, se generó un proceso de modelado dimensional, usando como referencia la metodología Kimball, para la definición de las dimensiones y hechos que van a componer el datamart; posterior a ello, se realizó el proceso de extracción, transformación y carga de los datos a partir de la base de datos transaccional de VALFI hacia un datamart, todo ello se realizó con el objetivo de crear un análisis OLAP y un cuadro de mando integral o dashboard. Culminada la implementación, concluyen que dicha aplicación representa una herramienta que va a permitir ver el comportamiento de las ventas durante los doce meses del año, ver la procedencia de los clientes con más ventas o ver cuáles son los productos más vendidos durante un año, generando un impacto positivo en VALFI.

- Galarza Torres, Wilfredo y Valdivieso Zavala, Diego (2015). *Implementación de Inteligencia de Negocios para la mejora de calidad de suministro de Edelnor*. Tesis para obtener el título de Ingeniero. Universidad San Martín de Porres. Lima, Perú.

Los autores explican que la implementación de inteligencia de negocios para la mejora de calidad de suministro de Edelnor, tiene como finalidad obtener los indicadores de las interrupciones del servicio eléctrico y reducir efectivamente el pago de compensaciones por fallas en la continuidad del servicio que brinda la empresa Edelnor S.A.A. Para ello, evaluaron los procesos de compensación por calidad de suministro, registro de interrupciones y reportes legales a Osinergmin. Se basaron en la metodología de Ralph Kimball para la implementación de inteligencia de negocios, desde la construcción de una ETL, el cual permitió obtener los datos más importantes para la construcción del datawarehouse y con ello construir los reportes que serán de gran utilidad para la toma de decisiones. Así también se utilizó la metodología Procede para las fases de desarrollo del producto, mediante la cual se identificó las necesidades del cliente, a su vez se planificó el desarrollo del producto para pasar al diseño y construcción del software, finalizando con la evaluación del grado de satisfacción del cliente. Como resultado, se logró mejorar los tiempos de solución y prevención de las fallas de los suministros eléctricos, disminuyendo los pagos por compensación de suministros establecidos por Osinergmin hacia los usuarios del servicio eléctrico; así como obtener las estadísticas e indicadores de fallas para una adecuada toma de decisiones.

2.2 Bases teóricas

2.2.1 On Line Analytical Processing (OLAP)

Según De los Angeles Ibarra, María (2006), define los sistemas OLAP como sistemas utilizados por los ejecutivos para proporcionar a nivel estratégico la información útil mediante la cual se pueda tomar una decisión. Asimismo, mediante OLAP los datos son clasificados en diferentes

dimensiones, que pueden ser visualizadas desde diferentes perspectivas e incluso, realizar combinaciones para obtener distintas formas de analizar los datos. Cuando se emplea un modelo OLAP, la información es vista como cubos, los cuales consisten en categorías descriptivas (dimensiones) y valores cuantitativos (medidas). El modelo de datos multidimensional simplifica a los usuarios formular consultas complejas, generar reportes, obtener datos detallados y filtrar los datos en subconjuntos significativos. OLAP comprende varias operaciones, analíticas básicas, incluidas la consolidación, “drill-down” y “slicing and dicing”, las cuales son definidas por la autora de la siguiente manera:

- Drill-down y Roll-up: OLAP puede moverse en la dirección contraria y presentar automáticamente datos detallados que abarcan datos consolidados. Esto recibe el nombre de drill-down, caso contrario se desee un menor nivel de detalle se denomina como roll-up.
- Slicing y dicing: se refiere a la capacidad de visualizar a las bases de datos desde diferentes puntos de vista. Una parte de la base de datos de ventas podría mostrar todas las ventas del tipo de producto dentro de las regiones. Otra parte podría mostrar las ventas por canal de ventas dentro de cada tipo de producto. El slicing and dicing con frecuencia se realiza a lo largo de un eje de tiempo con el fin de analizar tendencias y encontrar patrones.

Cabe resaltar que otros autores hacen mención de la operación “pivoting”, la cual consiste en el cambio de ejes al realizar las consultas a los cubos según la dimensión y las métricas sobre las cuales se desea obtener información; la Figura 3 nos muestra como interactúa el usuario con un cubo dimensional utilizando las operaciones analíticas mencionadas.

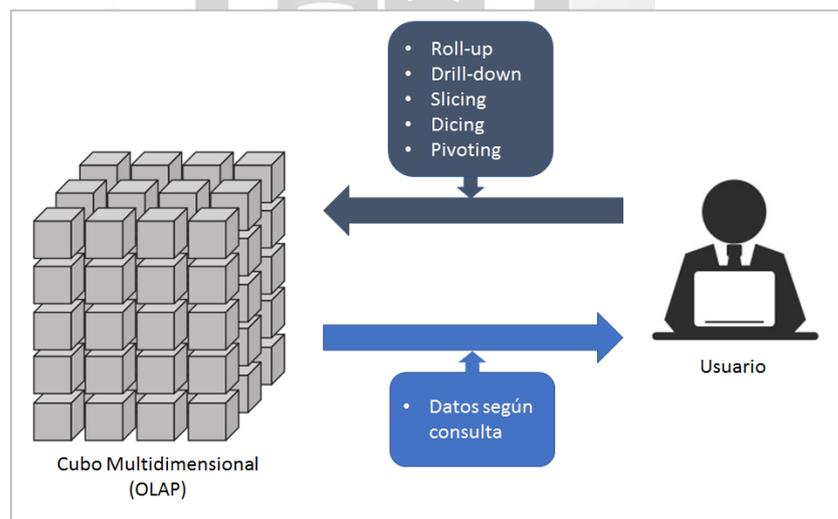


Figura 3: Interacción de usuario con cubo multidimensional
Fuente: Elaboración propia.

Así mismo, existen diferentes tipos de OLAP basados en el almacenamiento de datos que manejan; Conesa Caralt, Jordi y Curto Díaz, Josep (2010) definen las siguientes variaciones:

- MOLAP: es la forma “clásica” de OLAP, consiste en utilizar las estructuras de bases de datos generalmente optimizadas para la recuperación de los mismos y también para mejorar la

velocidad del cálculo. También se optimizan a menudo para la recuperación a lo largo de patrones jerárquicos de acceso. Las dimensiones de cada cubo son típicamente atributos tales como período, localización, producto o código de la cuenta. La manera que cada dimensión será agregada es definida por adelantado por unas o más jerarquías.

- **ROLAP:** trabaja directamente con las bases de datos relacionales, se almacenan los datos base y las tablas dimensionales como tablas relacionales mientras que tablas nuevas se crean para guardar la información agregada.
- **HOLAP:** No hay acuerdo claro en la industria en cuanto a qué constituye el "híbrido OLAP", a menos que es una base de datos en la que los datos se dividen en almacenaje relacional y multidimensional. Por ejemplo, para algunos vendedores, una base de datos de HOLAP utilizará las tablas relacionales para guardar las cantidades más grandes de datos detallados, y utiliza el almacenaje multidimensional para algunos aspectos de cantidades más pequeñas de datos menos detallados o agregados.
- **DOLAP:** es un caso particular de OLAP ya que está orientado a equipos de escritorio. Consiste en obtener la información necesaria desde la base de datos relacional y guardarla en el escritorio. Las consultas y análisis son realizados contra los datos guardados en el escritorio.
- **In-memory OLAP:** es un enfoque por el que muchos nuevos fabricantes están optando. Consiste en que la estructura dimensional se genera sólo a nivel de memoria y se guarda el dato original en algún formato que potencia su despliegue de esta forma (por ejemplo, comprimido o mediante una base de datos lógica asociativa). En este último punto es donde cada fabricante pone su énfasis.

Cada tipo de OLAP tiene ciertas ventajas, aunque hay desacuerdos específicos sobre las ventajas entre los diferentes proveedores, las cuales se presentan en la Tabla 1:

Tipo de OLAP	Diferencias
MOLAP	Es mejor en sistemas más pequeños de datos, es más rápido para calcular agregaciones y retornar respuestas y necesita menos espacio de almacenaje.
ROLAP	Se considera más escalable. Sin embargo, el pre-proceso de grandes volúmenes es difícil de implementar eficientemente así que se desecha con frecuencia. De otro modo, el funcionamiento de las queries puede ser no óptimo.
HOLAP	Está entre los dos en todas las áreas, pero puede preprocesar rápidamente y escalar bien.

Tabla 1: Diferencias entre los tipos de OLAP
Fuente: Adaptado de Conesa Caralt, Jordi y Curto Díaz, Josep (2010)

También se tiene presente, que en el contexto OLAP existen distintos elementos, entre los cuales Conesa Caralt, Jordi y Curto Díaz, Josep (2010) describen los siguientes en la Tabla 2:

Elemento de OLAP	Descripción
Esquema	Es una colección de cubos, dimensiones, tablas de hecho y roles.
Cubo	Es una colección de dimensiones asociadas a una tabla de hecho. Un cubo virtual permite cruzar la información entre tablas de hecho a partir de sus dimensiones comunes.
Jerarquía	Es un conjunto de miembros organizados en niveles. En cuanto a bases de datos, se puede entender como una ordenación de los atributos a una dimensión.
Nivel	Es un grupo de miembros en una jerarquía que tienen los mismos atributos y nivel de profundidad en una jerarquía.
Miembro	Es un punto de la dimensión de un cubo que pertenece a un determinado nivel de una jerarquía. Las métricas (medidas) en OLAP se consideran un tipo especial de miembro que pertenece a su propio tipo de dimensión. Un miembro puede tener propiedades asociadas.
Roles	Permisos asociados a un grupo de usuarios.
MDX	Es un acrónimo de Multidimensional eXpressions (aunque también es como Multidimensional Query eXpression). Es el lenguaje de consulta de estructuras OLAP, fue creado en 1997 por Microsoft, si bien no es un lenguaje estándar, la gran mayoría de fabricantes de herramientas OLAP lo han adoptado como estándar de hecho.
Tabla de hecho, dimensión y métrica	Son los componentes de la estructura dimensional.

Tabla 2: Elementos de OLAP
Fuente: Adaptado de Conesa Caralt, Jordi y Curto Díaz, Josep (2010)

2.2.2 Business Intelligence

De acuerdo con lo planteado por Conesa Caralt, Jordi y Curto Díaz, Josep (2010), existen situaciones en las que la implantación de un sistema de Business Intelligence resulta adecuada. Destacamos, entre todas las que existen:

- La toma de decisiones se realiza de forma intuitiva en la organización.
- Identificación de problemas de calidad de información.

- Uso de Excel como repositorios de información corporativos o de usuario. Lo que se conoce como Excel caos.
- Necesidad de cruzar información de forma ágil entre departamentos.
- Evitar silos de información.
- Las campañas de marketing no son efectivas por la información base usada.
- Existe demasiada información en la organización para ser analizada de la forma habitual. Se ha alcanzado la masa crítica de datos.
- Es necesario automatizar los procesos de extracción y distribución de información.

Lluís Cano, Josep (2007), define los siguientes componentes como componentes de Business Intelligence:

- Fuentes de información, de las cuales partiremos para alimentar de información el datawarehouse.
- Proceso ETL de extracción, transformación y carga de los datos en el datawarehouse. Antes de almacenar los datos en un 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.
- El propio datawarehouse o almacén de datos, con la metadata o diccionario de datos. Se busca almacenar los datos de una forma que maximice su flexibilidad, facilidad de acceso y administración.
- El motor OLAP, que nos debe proveer 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, que también desarrollaremos en el presente capítulo.
- Las herramientas de visualización, que nos permitirán el análisis y la navegación a través de los mismos.

Así mismo, Lluís Cano (2007) enfatiza que el análisis de los actuales usos de Business Intelligence debe tomar en cuenta la evolución de las herramientas, así como también la evolución y las tendencias en el uso por parte de las organizaciones, conociendo a su vez los movimientos que se han producido en el mercado. Históricamente, los distintos usos que soportan las herramientas han sido:

- Reporting: Elaboración de informes.
- Análisis: Herramientas de consultas ad hoc y OLAP.
- Planificación y modelización.
- Monitoring: Dashboards y Scorecards.
- Análisis avanzado: Datamining, Text Mining y Visualización avanzada.

Cada vez más, las distintas herramientas soportan un mayor número de funcionalidades, son más maduras, escalables y más fáciles de usar, desde la carga de datos (procesos ETL) hasta el análisis. También se han producido cambios tecnológicos significativos referentes a acceso vía web, tan sólo con un navegador o con un cliente ligero o corriendo sobre procesadores de 32 bits a 64 bits, herramientas de Text Mining, localización, alertas, compatibilidad XML, etc. Un factor determinante en la evolución de las herramientas Business Intelligence será la adopción de arquitecturas SOA (Service Oriented Architecture), las mejoras en la visualización de la información y sin duda el aumento del uso de las herramientas de Open Source o de código abierto. (Lluís Cano, 2007)

Business Intelligence es un proceso interactivo para explorar y analizar información estructurada sobre un área (normalmente almacenada en un datawarehouse), para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conclusiones. Abarca los procesos, las personas, las herramientas y las tecnologías para convertir datos en información, información en conocimiento y planes para conducir de forma eficaz las actividades de los negocios. Mediante la solución planteada, vamos a obtener KPI (Indicadores Clave de Negocio), los cuales sirven a las organizaciones para evaluar si están alcanzando sus objetivos. Son aquellos factores que en cada empresa resultan ser claves para progresar hacia el éxito, y dependen de la idiosincrasia y características de cada organización. Para lograr la implementación de una herramienta Business Intelligence, se debe realizar un proceso de extracción, transformación y carga de datos desde las fuentes de información disponibles, y las herramientas que nos facilitan este proceso y que nos permitirán alimentar un datawarehouse. (Lluís Cano, 2007)

Así mismo, Han, Jiawei; Kamber, Micheline & Pei, Jian (2012) explican lo fundamental que es para las empresas adquirir una mejor comprensión del contexto comercial de su organización, como sus clientes, el mercado, el suministro y los recursos, y competidores. Las tecnologías de Business Intelligence (BI) proporcionan información histórica, actual y visión predictiva de las operaciones comerciales. Algunos ejemplos son la presentación de informes, procesamiento, gestión del desempeño empresarial, inteligencia competitiva, benchmarking, y análisis predictivo. En la Figura 4 podemos observar las aplicaciones tecnológicas de Business Intelligence. “¿Cuán importante es la inteligencia de negocios?” Sin la minería de datos, muchas empresas pueden no pueda realizar análisis de mercado eficaces, comparar los comentarios de los clientes productos, descubrir las fortalezas y debilidades de sus competidores, conservar valiosos clientes y tomar decisiones de negocios inteligentes. Claramente, la minería de datos es el núcleo de la inteligencia de negocios. Las herramientas de procesamiento analítico en línea en la inteligencia empresarial se basan en el datawarehousing y el datamining multidimensional. Las técnicas de clasificación y predicción son el núcleo de la analítica predictiva en la inteligencia de negocios, para la cual hay muchas aplicaciones en el análisis de mercados, suministros y ventas. Por otra parte, la

agrupación desempeña un papel central en la gestión de la relación con los clientes, que agrupa a los clientes basándose en sus similitudes. Utilizando técnicas de minería de caracterización, podemos entender mejor las características de cada grupo de clientes y desarrollar programas personalizados de recompensa de clientes.



Figura 4: Aplicaciones de Business Intelligence
Fuente: Modak Analytics (2014)

2.2.3 Datamart

Talus Company (1998) menciona que el datamart es una idea cuyo tiempo ha llegado. Su enfoque y sencillez prometen el rápido despliegue de la capacidad de apoyo a las decisiones con el rápido retorno de la inversión exigida por el ritmo de los negocios modernos. El datamart ha evolucionado desde el concepto de datawarehouse. El alcance se ha vuelto altamente centrado - concentrándose en un único tema en lugar de toda la empresa. Esto es importante. De hecho, cambia todo. Controlando el alcance de esta manera, la inversión total en tiempo y dinero se reduce dramáticamente. El riesgo se mitiga de forma correspondiente ya que el datamart entrega el valor ganado en una fracción del tiempo requerido por el almacén de datos a escala empresarial. La siguiente tabla resume las principales diferencias entre los almacenes de datos y los datamarts. Claramente, una estrategia exitosa de datamart puede mitigar el riesgo, limitar el gasto y reducir el tiempo requerido para entregar la funcionalidad de datawarehouse. Debido a que es escalable (cuando correctamente implementado por el profesional experimentado), el mercado de datos puede funcionar bien para las organizaciones de cualquier tamaño y nivel de complejidad. No es difícil entender por qué el mercado de datos se considera ahora como el mecanismo más eficaz para proporcionar una capacidad de soporte de decisiones rápida y confiable. Si bien es cierto que el mercado de datos reduce en gran medida el riesgo asociado con la construcción de un sistema de apoyo a la decisión, sin embargo, requiere considerable habilidad y conocimientos para implementar adecuadamente uno. El oficio de construir el

datamart todavía es joven y los practicantes experimentados siguen siendo pocos en número. La manera más rápida para que un aspirante a constructor de datamart para adquirir las habilidades necesarias es estudiar las lecciones aprendidas por profesionales con experiencia. En la Figura 5, podemos observar cual es la relación entre los datamarts y un datawarehouse:

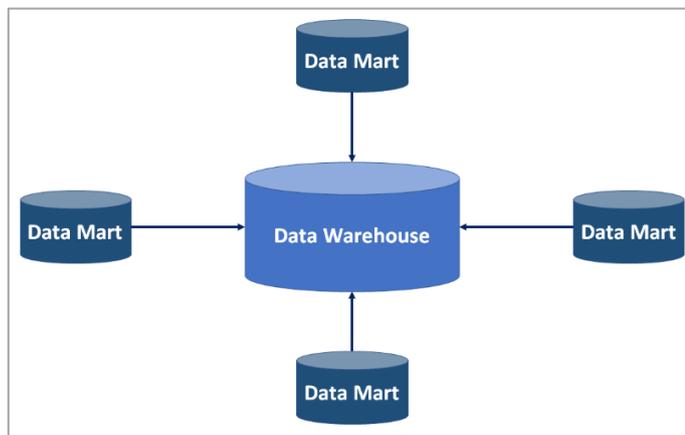


Figura 5: Relación entre Datamart y Datawarehouse
Fuente: Elaboración propia

Para Salvador Ramos (2011), la diferencia de un datamart con respecto a un datawarehouse es solamente en cuanto al alcance. Mientras que un datawarehouse es un sistema centralizado con datos globales de la empresa y de todos sus procesos operacionales, un datamart es un subconjunto temático de datos, orientado a un proceso o un área de negocio específica. Debe tener una estructura óptima desde todas las perspectivas que afecten a los procesos de dicha área. Es más, según Ralph Kimball, cada datamart debe estar orientado a un proceso determinado dentro de la organización, por ejemplo, a pedidos de clientes, a compras, a inventario de almacén, a envío de materiales, etc. Para Ralph Kimball el conjunto de datamarts forma el datawarehouse.

Así mismo, Ballard, Chuck et al. (2005) define 2 tipos de datamart: dependientes e independientes, los cuáles podemos observar en la Tabla 3:

Tipo de Datamart	Descripción
Dependiente	Contienen data que ha sido extraída directamente desde un datawarehouse, por ello, la data es integrada y es consistente con la data que se encuentra en el datawarehouse.
Independiente	Estos datamarts son autónomos, la data no se encuentra integrada y no es consistente con el datawarehouse. En su mayoría, la data es extraída de otras aplicaciones, bases de datos transaccionales o de operational data store (ODS).

Tabla 3: Tipos de datamarts
Fuente: Adaptación de Ballard, Chuck et. al. (2005)

Inmon, William (2005), afirma que las diferencias entre la unión en estrella y la estructura relacional como base para el diseño del almacén de datos son muchas. La diferencia más importante es en términos de flexibilidad y rendimiento. El modelo relacional es muy flexible, pero no está optimizado para el rendimiento de ningún usuario. El modelo multidimensional es altamente eficiente en el servicio de las necesidades de una comunidad de usuarios, pero no es bueno en flexibilidad. Otra diferencia importante en los dos enfoques para el diseño de bases de datos está en el alcance del diseño. Por necesidad, el diseño multidimensional tiene un alcance limitado. Dado que los requisitos de procesamiento se utilizan para moldear el modelo, el diseño comienza a descomponerse cuando se recogen muchos requisitos de procesamiento. En otras palabras, el diseño de la base de datos se puede optimizar para uno y sólo un conjunto de requisitos de procesamiento. Cuando se agrega un conjunto completamente diferente de requisitos de procesamiento a un diseño, la optimización se convierte en una cuestión discutible. Un diseño de base de datos se puede optimizar para el rendimiento de una sola manera. Cuando se utiliza el modelo relacional, no hay una optimización particular para el rendimiento de una manera u otra. Dado que el modelo relacional requiere que los datos se almacenen en el nivel más bajo de granularidad, se pueden añadir nuevos elementos de datos ad infinitum. Simplemente no hay fin a los datos que se pueden agregar al modelo relacional. Por esta razón, el modelo relacional es apropiado para un amplio rango de datos (como un modelo de empresa), mientras que el modelo multidimensional es apropiado para un pequeño rango de datos (como un departamento o incluso un subdepartamento). Las raíces de las diferencias entre el modelo multidimensional y el modelo relacional se remontan a la configuración original de los modelos mismos. La Figura 6 muestra cómo se forman los modelos, también muestra que el entorno relacional está conformado por el modelo de datos corporativo o empresarial. El modelo de unión en estrella o multidimensional está conformado por los requisitos del usuario final. En otras palabras, un modelo - el modelo relacional - está formado por un modelo de datos puros y el otro modelo multidimensional - está modelado por los requisitos de procesamiento. Esta diferencia en la conformación de los modelos conduce a varias consecuencias sutiles, pero muy importantes. La primera consecuencia es en términos de servicio. El modelo relacional se configura alrededor de una abstracción de datos. Debido a esto, el modelo en sí es muy flexible. Pero para toda la flexibilidad del modelo relacional, el modelo relacional no es terriblemente óptimo para el desempeño del acceso directo de datos. Si desea obtener un buen rendimiento del modelo relacional, es óptimo extraer los datos del modelo y volverlos a crear en una forma que sea adecuada para un acceso rápido. A pesar de las limitaciones de rendimiento, ya que el modelo relacional apoya la remodelación de datos, el modelo relacional es bueno para el acceso indirecto de datos. El modelo multidimensional, por otra parte, es bueno para una rápida y eficiente acceso directo a los datos. El modelo multidimensional apoya el acceso directo a los datos, en oposición al acceso indirecto a los datos soportados por el modelo de datos relacional.

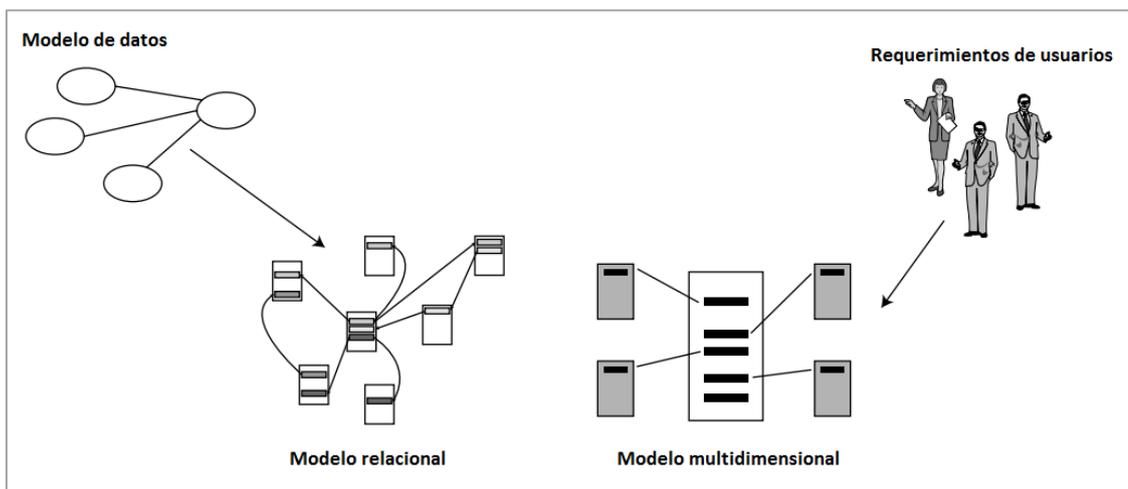


Figura 6: Formación de los modelos relacional y multidimensional
Fuente: Inmon, William (2005)

2.2.3.1 Cubos Multidimensionales

Según Martínez Orol, Jorge (2007), los cubos son elementos claves en OLAP (online analytic processing), una tecnología que provee rápido acceso a datos en un almacén de datos (datawarehouse). Los cubos proveen un mecanismo para buscar datos con rapidez y tiempo de respuesta uniforme independientemente de la cantidad de datos en el cubo o la complejidad del procedimiento de búsqueda. Los cubos son subconjuntos de datos de un almacén de datos, organizado y sumariado dentro de una estructura multidimensional. Los datos se sumarizan de acuerdo con factores de negocio seleccionados, proveyendo el mecanismo para el rápido y uniforme tiempo de respuesta de las complejas consultas. La definición del cubo es el primero de tres pasos en la creación de un cubo. Los otros pasos son, el especificar la estrategia de sumarización diseñando las agregaciones (elementos precalculados de datos), y la carga del cubo para procesarlo. Para definir un cubo, seleccione una tabla objetivo y seleccione las medidas (columnas numéricas de interés a los usuarios del cubo) dentro de esta tabla. Entonces seleccione las dimensiones, cada compuesta de una o más columnas de otra tabla. Las dimensiones proveen la descripción categórica por el cual las medidas son separadas para su análisis por los usuarios del cubo. El autor también menciona 2 elementos principales que conforman un cubo multidimensional: dimensiones y medidas.

Las dimensiones son categorías descriptivas por los cuales los datos numéricos en un cubo son separados para su análisis, en la Figura 7 podemos apreciar su representación de manera gráfica. Por ejemplo, si una medición de un cubo es el conteo de la producción, y las dimensiones son Tiempo, localización de la fábrica y el producto, los usuarios del cubo podrán separar el conteo de la producción, dentro de varias categorías de tiempo, localización de la fábrica y productos. Una dimensión puede ser creada para usarse en un cubo individual o en múltiples cubos. Una dimensión creada para un cubo individual es llamada dimensión privada. Por el contrario, si esta

puede ser usada por múltiples cubos, se le llama dimensión compartida. Estas podrán ser usadas dentro de todo cubo, en la base de datos, así se optimiza el tiempo y se evita el andar duplicando dimensiones privadas. Las dimensiones compartidas, también habilitan la estandarización de las métricas de negocios entre cubos. Por ejemplo, el estandarizar las dimensiones compartidas para el tiempo y localización geográfica, aseguran que los datos analizados, desde diferentes cubos, estén organizados similarmente. (Martínez Orol, Jorge, 2007)

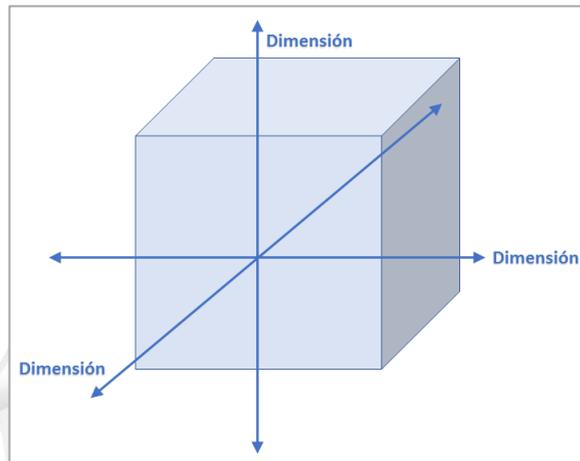


Figura 7: Representación de dimensiones en los ejes de un cubo OLAP
Fuente: Elaboración propia

Así mismo, Martínez Orol, Jorge (2007) afirma que las medidas, son datos numéricos de interés primario para los usuarios del cubo; en la Figura 8 se observa la representación de las medidas en el cubo. Algunas medidas comunes son ventas en unidades, costo de ventas, gastos, conteo de la producción, presupuesto, etc. Estas son usadas por el procedimiento de agregación de los servicios de OLAP y almacenadas para su rápida respuesta a las peticiones de los usuarios. Se puede crear una medida calculada y calcular miembros de dimensiones, combinando expresiones multidimensionales (MDX), fórmulas matemáticas y funciones definidas por el usuario (UDFs). Esta facilidad, habilita a usted a definir nuevas medidas y miembros de dimensión, basados sobre una sintaxis de fórmulas sencillas. Se pueden registrar adicionales bibliotecas de UDFs, para utilizarse en la definición de miembros calculados.

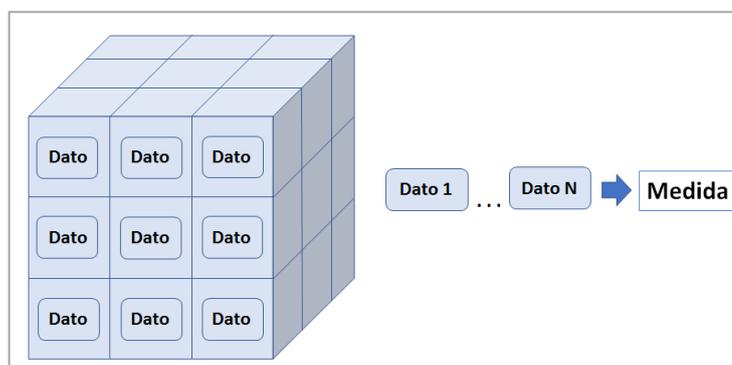


Figura 8: Representación de una medida dentro de un cubo OLAP
Fuente: Elaboración propia

Así mismo, para desarrollar una estructura dimensional, existen distintos esquemas, entre los más conocidos y utilizados tenemos los presentados por Salvador Ramos (2016), que son:

a) Esquema en estrella

A la hora de modelar el datamart o data warehouse, hay que decidir cuál es el esquema más apropiado para obtener los resultados que queremos conseguir. Habitualmente, y salvo excepciones, se suele modelar la base de datos utilizando el esquema en estrella, que se observa en la Figura 9, en el que hay una única tabla central, la tabla de hechos, que contiene todas las medidas y una tabla adicional por cada una de las perspectivas desde las que queremos analizar dicha información, es decir por cada una de las dimensiones:

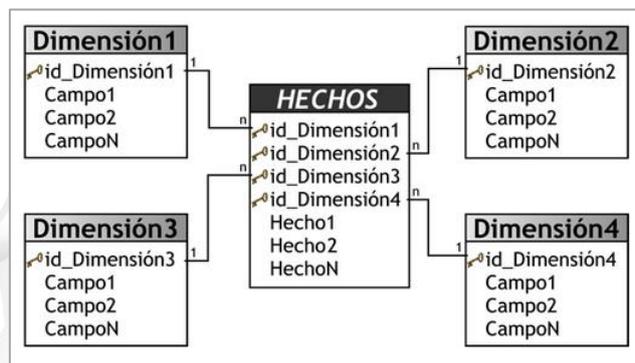


Figura 9: Esquema en estrella
Fuente: R. Darío, Bernabeu (2009)

b) Copo de nieve

La otra alternativa de modelado es la utilización del esquema en copo de nieve. Esta es una estructura más compleja que el esquema en estrella, cuya distribución gráfica la apreciamos en la Figura 10. La diferencia es que algunas de las dimensiones no están relacionadas directamente con la tabla de hechos, sino que se relacionan con ella a través de otras dimensiones. En este caso también tenemos una tabla de hechos, situada en el centro, que contiene todas las medidas y una o varias tablas adicionales, con un mayor nivel de normalización:

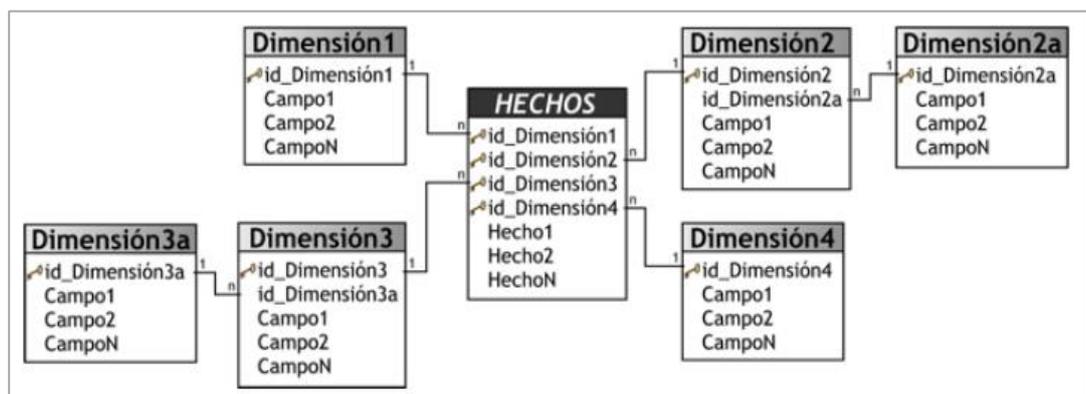


Figura 10: Esquema en copo de nieve
Fuente: R. Darío, Bernabeu (2009)

Cuando trabajamos con cubos multidimensionales, también manejamos un concepto que nos va a permitir observar el grado de granularidad que van a tener las dimensiones con las que vamos a trabajar, permitiéndonos observar los valores de que cada medida tiene según el tipo de consulta que deseamos realizar. Dicha característica que tienen las dimensiones se denominan jerarquías; Han, Jiawei; Kamber, Micheline & Pei, Jian (2012) mencionan que una jerarquía conceptual define una secuencia de asignaciones de un conjunto de conceptos de bajo nivel a conceptos más generales y de nivel superior. Considere una jerarquía de conceptos para la ubicación de la dimensión. Los valores de la ciudad para la ubicación incluyen Vancouver, Toronto, Nueva York y Chicago. Cada ciudad, sin embargo, puede ser mapeada a la provincia o estado al que pertenece. Por ejemplo, Vancouver puede ser mapeado a Columbia Británica, y Chicago a Illinois. Las provincias y los estados pueden ser asignados al país (por ejemplo, Canadá o los Estados Unidos) al que pertenecen. Estas asignaciones forman una jerarquía conceptual para la ubicación de la dimensión, trazando un conjunto de conceptos de bajo nivel (es decir, ciudades) a conceptos más generales y de nivel superior (es decir, países). Esta jerarquía conceptual se ilustra en la Figura 11. Muchas jerarquías de concepto están implícitas en el esquema de la base de datos. Por ejemplo, supongamos que la ubicación de la dimensión se describe por el número de atributos, calle, ciudad, provincia o estado, código postal y país. Alternativamente, los atributos de una dimensión pueden organizarse en un orden parcial, formando una red. Un ejemplo de una orden parcial para la dimensión temporal basada en los atributos día, semana, mes, trimestre y año, esta estructura de la red se muestra en la Figura 12.

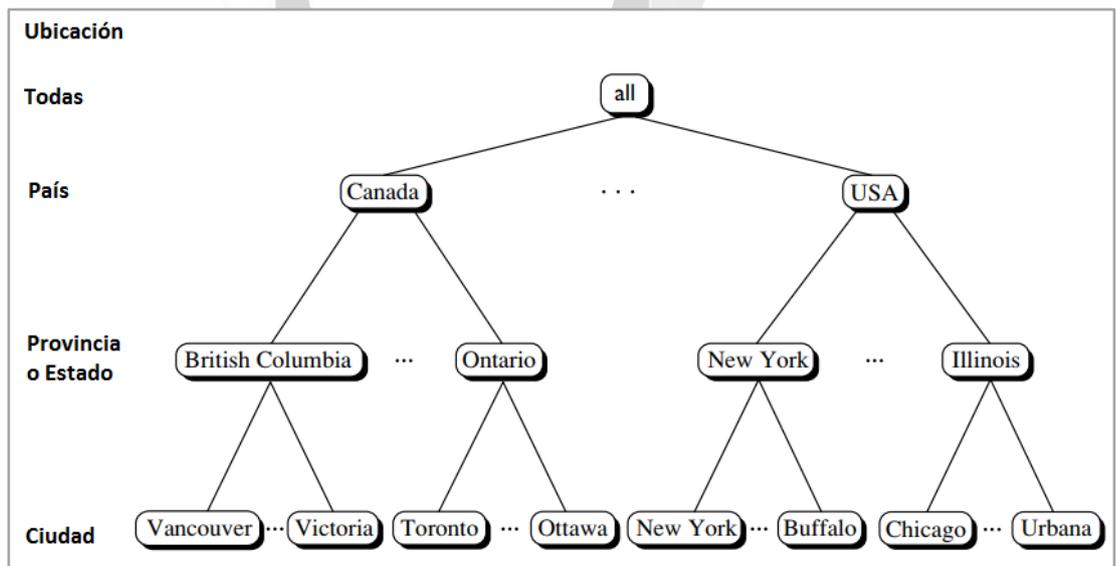


Figura 11: Concepto de jerarquía por Ubicación
Fuente: Han, Jiawei; Kamber, Micheline & Pei, Jian (2012)

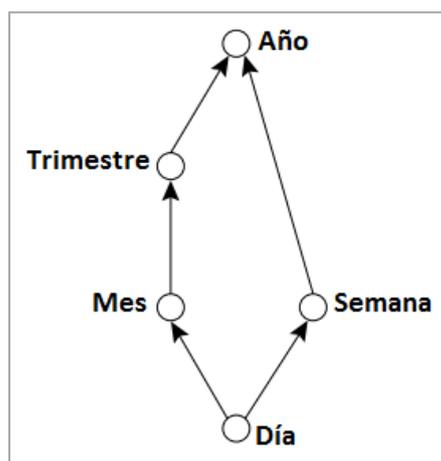


Figura 12: Concepto de jerarquía por Tiempo
Fuente: Han, Jiawei; Kamber, Micheline & Pei, Jian (2012)

Una jerarquía de conceptos que es un orden total o parcial entre atributos en un esquema de base de datos se denomina jerarquía de esquema. Las jerarquías de concepto que son comunes a muchas aplicaciones (por ejemplo, para el tiempo) pueden predefinirse en el sistema de minería de datos. Los sistemas de minería de datos deben proporcionar a los usuarios la flexibilidad necesaria para adaptar jerarquías predefinidas de acuerdo con sus necesidades particulares. Por ejemplo, los usuarios pueden querer definir un año fiscal a partir del 1 de abril o un año académico que comience el 1 de septiembre. Las jerarquías de concepto también pueden definirse agrupando valores para una dimensión o atributo dado, resultando en una jerarquía de agrupación de conjuntos. Un orden total o parcial puede ser definido entre grupos de valores. Puede haber más que una jerarquía de conceptos para un atributo o dimensión determinado, basada en diferentes puntos de vista del usuario. Por ejemplo, un usuario puede preferir organizar el precio mediante la definición de rangos de bajo costo, precio moderado y costoso. Las jerarquías de concepto pueden ser proporcionados manualmente por los usuarios del sistema, expertos o ingenieros de conocimiento, o pueden ser generados automáticamente sobre la base del análisis estadístico de la distribución de datos. La generación automática de jerarquías de concepto se discute como un paso de pre procesamiento en la preparación de la minería de datos; las jerarquías de concepto permiten manejar los datos en diferentes niveles de abstracción. (Han, Jiawei; Kamber, Micheline & Pei, Jian, 2012)

La herramienta que vamos a utilizar para la presente tesis es Microsoft SQL Server, por ello nos vamos a apoyar del Analysis Services para lograr generar el cubo, pero para poder realizar consultas (queries) a un cubo dimensional desde el administrador de base de datos de SQL, no basta con una consulta simple basada en T-SQL, debido a que la estructura dimensional posee características muy diferentes a una base de datos relacional, por ello se utiliza el lenguaje Multi Dimensional eXpressions (MDX); según Clay, C. y Smith, B.

(2009), explican que éste lenguaje es usado para interactuar con los datos de un cubo en Microsoft SQL Server Analysis Services. Es versátil y un lenguaje poderoso, pero que requiere que abordes los datos de una manera totalmente nueva, éste es el principal reto para el uso exitoso del lenguaje MDX. Las dimensiones del cubo se denominarán simplemente dimensiones. Aunque la distinción entre las dimensiones de cubo y las dimensiones del Analysis Services es crítica en el contexto del diseño de cubos, la distinción no suele ser un problema en el desarrollo MDX. Con todo el acceso de datos a través del cubo, los desarrolladores de MDX se interceptan casi exclusivamente con las dimensiones del cubo.

Así mismo en una conferencia de Koo Labrín, Alan (2014), se realizó una introducción al lenguaje MDX, con la que buscaba explotar al máximo todo el potencial del Analysis Services Multidimensional. En dicha conferencia, definía a MDX como un lenguaje de consultas para bases de datos OLAP, análogo a SQL como lenguaje de consulta para base de datos relacionales, con el cuál se obtiene un dataset (conjunto de resultados) multidimensional que consiste de data en celdas y ejes. Así mismo, dicho lenguaje fue originalmente desarrollado por Microsoft a finales de los 90's y adoptado por muchos otros fabricantes de bases de datos multidimensionales. Entre los principales escenarios de uso que menciona el expositor, presentaba las mostradas en la Figura 13:



Figura 13: Escenarios de uso del lenguaje MDX
Fuente: Han, Jiawei; Kamber, Micheline & Pei, Jian (2012)

Así mismo, el lenguaje MDX tiene distintas funciones que van a permitir realizar distintas operaciones, entre las que Microsoft define las funciones presentadas en la Tabla 4:

Categoría de Función	Descripción
Funciones de matriz	Proporcionan matrices para su uso en procedimientos almacenados
Funciones de dimensión	Devuelven una referencia a una dimensión de una jerarquía, un nivel o un miembro.
Funciones de jerarquía	Devuelven una referencia a una jerarquía de un nivel o un miembro.
Funciones de nivel	Devuelven una referencia a un nivel de un miembro, una dimensión, una jerarquía o una expresión de cadena.
Funciones lógicas	Realizan operaciones y comparaciones lógicas en objetos y expresiones.
Funciones de miembro	Devuelven una referencia a un miembro de otros objetos o de una expresión de cadena.
Funciones numéricas	Realizan funciones matemáticas y estadísticas en objetos y expresiones.
Funciones de conjuntos	Devuelven una referencia a un conjunto de otros objetos o de una expresión de cadena.
Funciones de cadena	Devuelven valores de cadena de otros objetos o del servidor.
Funciones de tupla	Devuelven una referencia a una tupla de un conjunto o de una expresión de cadena.

Tabla 4: Funciones de lenguaje MDX
Fuente: Microsoft Technet (2017)

Así mismo, al utilizar el lenguaje MDX, también debemos mencionar la existencia de los valores calculados, los cuales tienen una relación directa con las jerarquías que definimos en nuestras dimensiones. Como mencionan Clay, C. y Smith, B. (2009), se puede pensar en un miembro calculado como otro miembro de un atributo existente o en una jerarquía de usuarios. Sin embargo, a diferencia de los miembros tradicionales, los miembros calculados no tienen datos almacenados realmente en los puntos en el espacio de cubo asociado con ellos. En su lugar, un miembro calculado se asigna una expresión MDX.

Los miembros calculados, como cualquier miembro en un cubo, deben estar asociados a una jerarquía. Finalmente, en la Figura 14, se muestra un ejemplo de una consulta MDX:

```
WITH
MEMBER [Measures].[Total Sales Amount] AS
    ([Measures].[Internet Sales Amount]) + ([Measures].[Reseller
Sales Amount])
SET [Top 10 Products of 2003] AS
    TOPCOUNT(
        {[Product].[Product]. [Product].Members},
        10,
        ([Measures].[Total Sales Amount], [Date].[Calendar Year].[CY
2003])
    )
SELECT
    {[Measures].[Total Sales Amount]} ON COLUMNS,
    {[Top 10 Products of 2003]} ON ROWS
FROM [Step-by-Step]
WHERE ([Date].[Calendar Year].[CY 2004])
```

Figura 14: Query en lenguaje MDX para consultar un cubo multidimensional
Fuente: Clay, C. y Smith, B. (2009)

2.2.4 Metodología de Ralph Kimball

De acuerdo con Rivadera, Gustavo (2010), la metodología se basa en lo que Kimball denomina ciclo de vida dimensional del negocio. Este ciclo de vida del proyecto de datawarehouse, está basado en cuatro principios básicos:

- Centrarse en el negocio: Hay que concentrarse en la identificación de los requerimientos del negocio y su valor asociado, y usar estos esfuerzos para desarrollar relaciones sólidas con el negocio, agudizando el análisis del mismo y la competencia consultiva de los implementadores.
- Construir una infraestructura de información adecuada: Diseñar una base de información única, integrada, fácil de usar, de alto rendimiento donde se reflejará la amplia gama de requerimientos de negocio identificados en la empresa.
- Realizar entregas en incrementos significativos: crear el almacén de datos en incrementos entregables en plazos de 6 a 12 meses. Hay que usar el valor de negocio de cada elemento

identificado para determinar el orden de aplicación de los incrementos. En esto la metodología se parece a las metodologías ágiles de construcción de software.

- Ofrecer la solución completa: proporcionar todos los elementos necesarios para entregar valor a los usuarios de negocios. Para comenzar, esto significa tener un almacén de datos sólido, bien diseñado, con calidad probada, y accesible. También se deberá entregar herramientas de consulta ad hoc, aplicaciones para informes y análisis avanzado, capacitación, soporte, sitio web y documentación.

Rivadera, Gustavo (2010) afirma que la construcción de una solución de Datawarehouse/Business Intelligence es sumamente compleja, y Kimball nos propone una metodología que nos ayuda a simplificar esa complejidad (...). Los requerimientos del negocio son el soporte inicial de las tareas subsiguientes. También tiene influencia en el plan de proyecto. En segundo lugar, podemos ver tres rutas o caminos que se enfocan en tres diferentes áreas:

- Tecnología (Camino Superior). Implica tareas relacionadas con software específico, por ejemplo, Microsoft SQL Analysis Services.
- Datos (Camino del medio). En la misma diseñaremos e implementaremos el modelo dimensional, y desarrollaremos el subsistema de extracción, transformación y carga (ETL) para cargar el datawarehouse.
- Aplicaciones de Inteligencia de Negocios (Camino Inferior). En esta ruta se encuentran tareas en las que diseñamos y desarrollamos las aplicaciones de negocios para los usuarios finales.

A diferencia de la metodología de Bill Inmon, Ralph Kimball propone una estrategia de procesamiento basada en bottom-up, que consiste en generar un data warehouse mediante la creación de los datamarts, como se puede apreciar en la Figura 15:

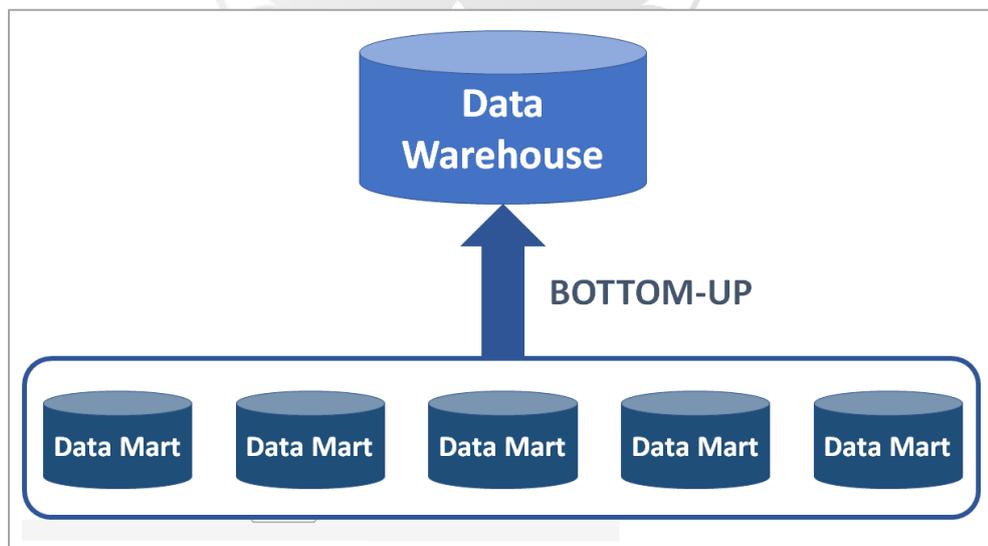


Figura 15: Estrategia de procesamiento de datos según metodología Ralph Kimball
Fuente: Elaboración propia.

Kimball, Ralph y Ross, Margy (2013), propone el desarrollo de la metodología de la siguiente manera:

a) Planificación y gestión de proyectos

- Alcance: Una vez que se sienta cómodo con la preparación de la organización, es hora de poner límites alrededor de un proyecto inicial. El alcance requiere el aporte conjunto tanto de la organización de TI como de la gestión empresarial. El alcance de su proyecto de almacén de datos debe ser significativo en términos de su valor para la organización y manejable. Cuando comienza por primera vez, debe concentrarse en los datos de un solo proceso empresarial. Guarde los proyectos de proceso cruzado más desafiantes para una fase posterior. A veces, los alcances se ven impulsados por una fecha de finalización del objetivo, como el final del año fiscal. Puede administrar el alcance a una fecha de vencimiento de manera efectiva, pero hacerlo puede presentar riesgos adicionales. Incluso con un marco de tiempo establecido, debe mantener su enfoque en el alcance de un proyecto que sea convincente y factible. En ocasiones, los equipos de proyecto sienten que el cronograma de entrega se concreta antes de que se inicie la planificación del proyecto. El proceso de priorización, que describiremos durante la definición de los requisitos del negocio, se puede usar para convencer a la TI y a la administración comercial de que se requieren ajustes. Finalmente, recuerde evitar también la ley al momento del alcance: demasiado firme como para comprometerse con una línea de tiempo demasiado breve que involucre demasiados sistemas de origen y demasiados usuarios en demasiados lugares con requisitos analíticos muy diversos.
- Justificación: Un montón de siglas rodea el proceso de justificación, pero no dejes que te intimiden. La justificación requiere una estimación de los beneficios y costos asociados con un datawarehouse. Con suerte, los beneficios anticipados superan ampliamente los costos. Por lo general, el responsable de derivar los gastos debe determinar los costos aproximados para el hardware y el software necesarios. Los almacenes de datos tienden a expandirse rápidamente, así que asegúrese de que las estimaciones permitan cierto margen de crecimiento a corto plazo
- Desarrollar y mantener el Plan de Proyecto: Los proyectos de depósito de datos son vulnerables a la dispersión del alcance en gran medida debido a nuestro fuerte deseo de satisfacer los requisitos de los usuarios.

b) Definición de requisitos del negocio

- Elegir el foro: Existen dos técnicas principales para reunir los requisitos como las entrevistas o sesiones facilitadas. Ambos tienen sus ventajas y desventajas. Las entrevistas fomentan mucha participación individual. También son más fáciles de

programar. Las sesiones facilitadas pueden reducir el tiempo transcurrido para reunir los requisitos, aunque requieren más tiempo de compromiso de cada participante.

- Identifique y prepare el equipo de requisitos: Independientemente del enfoque, debe identificar y preparar a los miembros del equipo del proyecto que están involucrados. Si está haciendo entrevistas, debe identificar a un entrevistador principal cuya principal responsabilidad es hacer las preguntas geniales abiertas.
- Seleccionar, programar y preparar representantes comerciales: Dentro de la comunidad de usuarios objetivo, debe cubrir la organización verticalmente. Los equipos de proyectos de depósito de datos gravitan naturalmente hacia los analistas de la superpotencia en el negocio. Si bien su visión es valiosa, no puede ignorar a los altos directivos y la gerencia media. De lo contrario, eres vulnerable a estar excesivamente centrado en el aquí y ahora táctico, pero perder de vista la dirección estratégica futura de la organización.
- Recopilación de los requisitos de negocios:
 - Lanzamiento: La responsabilidad de presentar la entrevista debe establecerse antes de reunirse en una sala de conferencias.
 - Flujo de la entrevista: El objetivo de una entrevista es hacer que los usuarios de negocios hablen sobre lo que hacen y por qué lo hacen.

c) **Diseño de arquitectura técnica**

Los equipos de almacenamiento de datos abordan el proceso de diseño de la arquitectura técnica desde extremos opuestos del espectro. Algunos equipos simplemente no entienden los beneficios de una arquitectura y sienten que el tema y las tareas son demasiado nebulosos. Están tan centrados en la entrega del almacén de datos que las arquitecturas se sienten como una distracción e impedimento para el progreso, por lo que optan por evitar el diseño de la arquitectura. En su lugar, reconstruyen los componentes técnicos necesarios para la primera iteración con cuerda flexible y goma de mascar, pero la integración y las interfaces reciben impuestos a medida que agregamos más datos, más usuarios o más funciones. Eventualmente, estos equipos a menudo terminan reconstruyendo porque la estructura no arquitectónica no puede resistir las tensiones. En el otro extremo, algunos equipos quieren invertir dos años diseñando la arquitectura y olvidando que el objetivo principal de un almacén de datos es resolver problemas comerciales, y no abordar ningún desafío técnico plausible (y no tan plausible). Recuerde, cada almacén de datos tiene una arquitectura técnica. La cuestión es si la suya está planificada y explícita o simplemente implícita.

- Establecer una fuerza de trabajo de arquitectura.
- Recopilar requisitos relacionados con la arquitectura.
- Requisitos de arquitectura de documentos.

- Desarrollar un modelo arquitectónico de alto nivel.
- Diseñe y especifique los subsistemas.
- Determinar las fases de implementación de la arquitectura.
- Documente la arquitectura técnica.
- Revisión y finalización de la arquitectura técnica.

d) Selección e implementación de productos

El primer paso antes de seleccionar nuevos productos es comprender los procesos internos de aprobación de hardware y software, nos guste o no. Tal vez los gastos deban ser aprobados por el comité de asignaciones de capital.

- Desarrollar una matriz de evaluación de productos.
- Realizar estudios de mercado
- Limite las opciones a una lista corta y realice evaluaciones detalladas.
- Realice un prototipo, si es necesario.
- Seleccionar producto, instalar en prueba y negociar

e) Modelado Dimensional

En este punto, es hora de hacer un análisis más completo de los datos generados por este proceso. Si bien realizamos una auditoría de alto nivel durante la definición de requisitos comerciales, debemos profundizar en el detalle para evaluar la granularidad, la coherencia histórica, los valores válidos y la disponibilidad de atributos. A menudo, los expertos en temas de negocios o los analistas de poder de la comunidad empresarial pueden arrojar luz rápidamente sobre las inconsistencias o idiosincrasias de los datos en función de los desafíos que han encontrado al intentar analizar los datos. Una vez completada la tarea de análisis de datos, llevamos a cabo talleres de diseño para crear el esquema dimensional. Una vez que el equipo de modelado tiene una confianza razonable sobre su producto de trabajo, comunicamos y validamos el diseño con una audiencia más amplia, primero dentro del equipo de TI y de almacén de datos y luego con otros miembros de la comunidad empresarial. La documentación en el modelo validado debe identificar los nombres de las tablas y columnas, las definiciones y las reglas de cálculo para hechos o las reglas de dimensión que cambian lentamente para los atributos de dimensión. Típicamente capturado en una herramienta de modelado, esta información es parte de la primera entrada (o enlace) a un catálogo de metadatos.

f) Diseño físico

Los modelos dimensionales desarrollados en la sección anterior deben traducirse en un diseño físico. En el modelado dimensional, los diseños lógicos y físicos tienen una gran

semejanza. El modelo físico será diferente del modelo lógico en términos de los detalles especificados para la base de datos física, incluidos los nombres de columnas físicas, tipos de datos, declaraciones clave (si corresponde) y la permisibilidad de nulos. En este punto, el diseño físico también contiene actividades tales como el ajuste del rendimiento, la partición y el diseño del archivo. Cada almacén de datos debe contener tablas de agregación precalculadas y preestablecidas. Dadas nuestras estrictas reglas para evitar la granularidad de tablas de hechos combinados, cada agregación de tabla de hechos debe ocupar su propia tabla de hechos físicos. Cuando agregamos hechos, eliminamos la dimensionalidad o asociamos los hechos con una dimensión acumulada. Estas tablas de dimensiones agregadas y enrolladas deberían ser versiones reducidas de las dimensiones asociadas con la tabla de hechos de base granular. De esta forma, las tablas de dimensiones agregadas se ajustan a las tablas de dimensiones base. Los administradores de bases de datos pueden hiperventilar cuando descubren que las tablas de dimensiones con frecuencia tienen más que un solo índice. Las tablas de dimensiones tendrán un índice único en la clave principal de una sola columna. Además, recomendamos un índice de árbol B en las columnas de atributos de alta cardinalidad utilizadas para las restricciones. Los índices mapeados en bits deberían colocarse en todos los atributos de cardinalidad media y baja. Mientras tanto, las tablas de hechos son los monstruos del almacén de datos, por lo que necesitamos indexarlos con más cuidado. La clave principal de la tabla de hechos es casi siempre un subconjunto de las claves foráneas. Por lo general, colocamos un único índice concatenado en las dimensiones principales de la tabla de hechos. Dado que muchas consultas dimensionales están restringidas en la dimensión de fecha, la clave externa de fecha debe ser el término de índice principal.

g) Diseño e implementación del subsistema de ETL

Uno de los procesos más relevantes para el desarrollo de un Datamart mediante la metodología de Ralph Kimball, es el ETL (Extract, Transform and Load), que consiste en la extracción de la información, la transformación y adecuación de ésta para su posterior carga en los repositorios dimensionales. Según Kimball, Ralph (2004), el sistema Extract-Transform-Load (ETL) es la base del datawarehouse. Un sistema ETL correctamente diseñado extrae datos de los sistemas de origen, aplica estándares de calidad y consistencia de datos, conforma datos para que fuentes separadas puedan ser utilizadas juntas y, finalmente, entrega datos en un formato preparado para la presentación para que los desarrolladores de aplicaciones puedan crear aplicaciones y usuarios finales puede tomar decisiones. El sistema ETL hace o interrumpe el almacén de datos. Aunque la construcción del sistema ETL es una actividad en la parte trasera que no es muy visible para los usuarios finales, consume fácilmente el 70% de los recursos necesarios para la implementación y mantenimiento de un almacén de datos típico. El sistema ETL agrega valor significativo a

los datos. Es mucho más que plomería para obtener datos de los sistemas de origen y en el almacén de datos. Específicamente, el sistema ETL:

- Elimina los errores y corrige los datos que faltan.
- Proporciona medidas documentadas de confianza en los datos.
- Captura el flujo de datos transaccionales para su custodia.
- Ajusta los datos de varias fuentes que se van a utilizar juntos.
- Estructuras de datos que pueden ser utilizadas por las herramientas del usuario final.

Kimball, Ralph y Caserta, Joe (2004), afirma que el ETL es un tema sencillo y complicado. Casi todo el mundo entiende la misión básica del sistema ETL: obtener datos de la fuente y cargarlos en el almacén de datos. Y la mayoría de los observadores están apreciando cada vez más la necesidad de limpiar y transformar datos a lo largo del camino. Tanto para la vista simple. Es un hecho de la vida que el siguiente paso en el diseño del sistema ETL se rompe en miles de subcajas poco, dependiendo de sus propias fuentes de datos extraños, reglas de negocio, software existente, e inusuales aplicaciones de destino para informes. En la Figura 16 observamos el flujo del proceso de ETL.

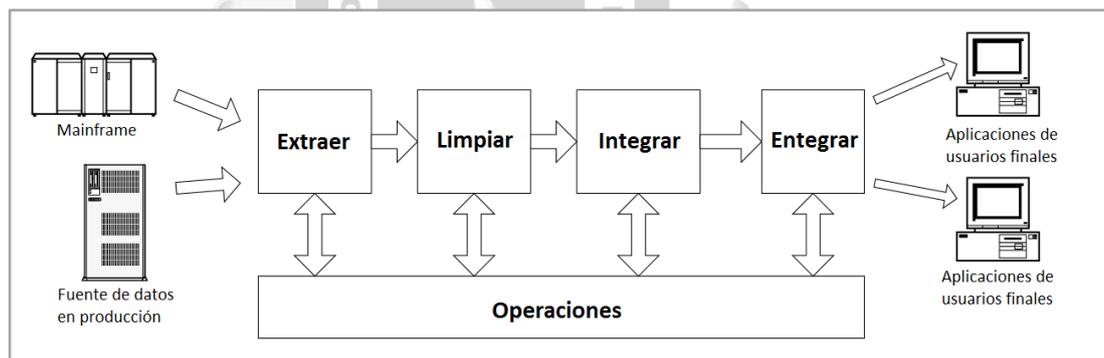


Figura 16: Flujo del ETL (Extract, Transform and Load)
Fuente: Kimball, Ralph y Caserta, Joe (2004).

h) Especificación de aplicaciones de BI

- Especificación de la aplicación analítica: Antes de comenzar a diseñar las aplicaciones iniciales, es útil establecer estándares para las aplicaciones, tales como menús desplegables comunes y apariencia y sensación de salida consistentes. Usando los estándares, especificamos cada plantilla de aplicación, capturando información suficiente sobre el diseño, las variables de entrada, los cálculos y los descansos, de modo que tanto el desarrollador de la aplicación como los representantes comerciales compartan un entendimiento común. Durante la actividad de especificación de la aplicación, también debemos considerar la organización de las aplicaciones. Necesitamos identificar caminos de navegación estructurados para acceder a las aplicaciones, lo que refleja la forma en que los usuarios piensan sobre su negocio. Aprovechar la Web y los portales de

información personalizables son las estrategias dominantes para difundir el acceso a las aplicaciones.

- Desarrollo de aplicaciones analíticas. La actividad de desarrollo de aplicaciones puede comenzar una vez que se completa el diseño de la base de datos, se instalan las herramientas de acceso a los datos y los metadatos y se ha cargado un subconjunto de datos históricos. Las especificaciones de la plantilla de la aplicación deben revisarse para tener en cuenta los inevitables cambios en el modelo de datos desde que se completaron las especificaciones. Cada herramienta en el mercado tiene trucos específicos del producto que pueden hacer que salte a través de los aros hacia atrás. En lugar de tratar de aprender las técnicas a través de prueba y error, debe invertir en la educación específica de la herramienta o recursos suplementarios para el equipo de desarrollo. Mientras se desarrollan las aplicaciones, se obtienen varios beneficios auxiliares. Los desarrolladores de aplicaciones, armados con una herramienta robusta de acceso a datos, encontrarán rápidamente problemas de punción en el pajar de datos a pesar de la garantía de calidad realizada por la aplicación provisional. Esta es una de las razones por las que preferimos comenzar con la actividad de desarrollo de aplicaciones antes de la supuesta finalización de la puesta en escena.

i) Despliegue

Al igual que un lanzamiento de un producto de software pasa por una serie de fases antes de la disponibilidad general, también lo debe el almacén de datos. La fase de prueba alfa consiste en que el equipo central del proyecto realiza una prueba de sistema de extremo a extremo. Al igual que con cualquier prueba del sistema, seguramente encontrará problemas, así que asegúrese de que haya un tiempo adecuado en el cronograma para la repetición inevitable. Con la prueba beta, involucramos a un grupo limitado de usuarios comerciales para realizar una prueba de aceptación del usuario, especialmente en lo que se refiere a la relevancia y la calidad empresarial de los productos entregables del almacén. Finalmente, el almacén de datos se publica para la disponibilidad general, aunque como un despliegue controlado.

j) Mantenimiento y crecimiento

Lo logramos a través del despliegue, así que ahora estamos listos para descansar y relajarse. ¡No tan rápido! Nuestro trabajo está lejos de completarse una vez que hemos implementado. Necesitamos continuar invirtiendo recursos en las siguientes áreas:

- Apoyo: La asistencia al usuario es crucial inmediatamente después del despliegue para garantizar que la comunidad empresarial se enganche. Durante las primeras semanas posteriores a la educación del usuario, el equipo de soporte debe trabajar proactivamente con los usuarios. No podemos sentarnos en nuestros cubículos y asumir que ninguna

noticia de la comunidad comercial es una buena noticia. Si no tenemos noticias de ellos, entonces es probable que nadie esté usando el almacén de datos.

- **Educación:** Necesitamos proporcionar un programa de educación continua para el depósito de datos. El plan de estudios debe incluir cursos de actualización y cursos avanzados, así como repetir cursos introductorios. Se puede ofrecer más educación informal a los desarrolladores y usuarios avanzados para fomentar el intercambio de ideas.
- **Soporte técnico:** El almacén de datos ya no es un programa agradable, pero debe tratarse como un entorno de producción, completo con acuerdos de nivel de servicio. Por supuesto, el soporte técnico debe monitorear de manera proactiva el desempeño y las tendencias de la capacidad del sistema. No queremos confiar en la comunidad empresarial para decirnos que el rendimiento se ha degradado.
- **Apoyo al programa:** Si bien la implementación de una fase específica del depósito de datos puede estar terminando, el programa de almacenamiento de datos continúa. Debemos continuar monitoreando el progreso en comparación con los criterios de éxito acordados.

2.2.5 Demanda Eléctrica

Como menciona Espina Alvarado, José (2017) podemos definir demanda eléctrica como la intensidad de corriente, o potencia eléctrica, relativa a un intervalo de tiempo específico, que absorbe su carga para funcionar. Ese lapso se denomina intervalo de demanda, y su indicación es obligatoria a efecto de interpretar un determinado valor de demanda. Los intervalos de demanda son típicamente de 15, 30 o 60 minutos. Los lapsos de 15 o 30 minutos se aplican comúnmente en facturación, selección de la capacidad de equipos, estudios de balanceo y transferencia de carga. El intervalo de 60 minutos permite construir “Perfiles de Carga Diarios” para el análisis de consumo de energía, determinar el rendimiento de dispositivos, y también para elaborar un completo plan de expansión del sistema de distribución de energía eléctrica. No obstante, en la selección de fusibles y el ajuste de protecciones también es importante conocer las demandas máximas instantáneas. Ahora bien, la demanda es una cantidad cuya medida depende del caso de estudio: amperios para la selección o reemplazo de conductores, fusibles, o de interruptores, ajuste de protecciones y balanceo de carga; kilovatios para la planificación del sistema, estudios de energía consumida, energía no vendida, y energía pérdida; kilovoltamperios para la selección de la capacidad de transformadores y alivio de carga. La mayoría de las compañías de servicio eléctrico distinguen el comportamiento de la demanda sobre una base de clases, caracterizando cada clase por un PCD “típico”, el cual representa el patrón de uso esperado de la carga para un cliente de esa clase el día de la demanda máxima del sistema, esto es, el día pico (durante el cual el sistema de distribución debe soportar la mayor exigencia del año calendario). Tales perfiles describen los aspectos más importantes desde el punto de vista del planificador de distribución: la

magnitud de la demanda máxima o pico de carga por cliente, la duración del pico, y la energía total consumida. Con propósitos de planificación, es una práctica común suponer que el comportamiento de un perfil típico es cíclico. El pico de carga por cliente, o demanda máxima diversificada, cae a medida que se agregan nuevos clientes al grupo. Cada residencia tiene un breve, pero pronunciado pico que rara vez es coincidente con otro dentro del mismo grupo, especialmente si es numeroso. En fin, el pico del grupo ocurre cuando de la combinación de las curvas de carga individuales se obtiene un máximo que es sustancialmente menor que la suma de los picos individuales como puede verificarse en la ilustración anterior.

Según el Organismo Supervisor de la Inversión en Energía y Minería (2016), la demanda eléctrica se caracteriza por registrar un comportamiento variable durante el día. Al esquema que muestra esta particularidad se le denomina curva o diagrama de carga, el cual permite la identificación de periodos de alta o baja demanda denominados horas punta y horas fuera de punta, respectivamente. Asimismo, si se ordena la demanda de potencia eléctrica de forma descendente, se obtiene una curva de pendiente negativa denominada diagrama de duración, cuya utilidad contribuye al proceso de planificación de las inversiones en generación eléctrica.

2.2.6 Modelo de calidad de producto de datos: ISO/IEC 25012

Tal y como menciona Hernández González (2017), este es un estándar internacional que define un modelo general de calidad de datos para mantener un formato estructurado dentro de los sistemas de cómputo. Este estándar se enfoca en la calidad de los datos como parte de los sistemas de cómputo y define características de calidad con objetivo en el uso de los datos por humanos y por sistemas. Este estándar toma en cuenta todos los tipos de datos, para asignar valores y relaciones entre los datos, no está definido para una organización específica de los datos, todos los procesos y envíos de datos relacionados cuentan con los beneficios al aplicar el estándar.

El estándar define la calidad de datos, como el grado en que las características de los datos guardan las condiciones y sugiere las necesidades cuando es usado bajo condiciones específicas. Indica también que las características son las categorías de los atributos que llevan a la calidad de datos; y que el modelo es el conjunto de características de calidad que proveen un marco de trabajo con requerimientos específicos de calidad de datos y su evaluación. Este modelo define un conjunto de características de calidad que son consideradas desde dos puntos de vista: inherente y dependiente del sistema. La calidad de datos inherente se refiere al grado en el cual las características de calidad de datos tienen el potencial intrínseco para satisfacer las necesidades implicadas cuando el dato es usado bajo condiciones específicas, las cuales apreciamos en la Tabla 5. Por otra parte, la calidad de datos dependiente del sistema se refiere al grado en el cual la calidad de los datos es enriquecida y preservada dentro de un sistema de cómputo cuando el dato

es usado bajo condiciones específicas, cuyas dimensiones se presentan en la Tabla 6. (Hernández Gonzáles, 2017)

Características inherentes	
Dimensión	Descripción
Exactitud (Accuracy)	El grado en el cual el dato tiene atributos que representan correctamente el valor del atributo intencionado de un concepto o evento en un contexto específico de empleo.
Compleitud (Completeness)	El grado al cual el dato del sujeto asociado con una entidad tiene valores para todos los atributos esperados e instancias de entidad relacionadas en un contexto específico de uso.
Consistencia (Consistency)	El grado en el cual el dato tiene los atributos que son libres de contradicción y son coherentes con otros datos en un contexto específico de uso.
Credibilidad (Credibility)	El grado en el cual el dato tiene atributos que son considerados verdaderos y creíbles por usuarios en un contexto específico de uso.
Actualidad (Currentness)	El grado en el cual el dato tiene los atributos que son del período correcto en un contexto específico de uso.

Tabla 5: Características inherentes del sistema según ISO/IEC 25012
Fuente: Guerra García, César (2009)

Características dependientes del sistema	
Dimensión	Descripción
Disponibilidad (Availability)	El grado en el cual el dato tiene atributos que le permiten ser recuperados por usuarios autorizados o por aplicaciones en un contexto específico de uso.
Portabilidad (Portability)	El grado en el cual el dato tiene los atributos que le permiten ser instalado, substituido o movido de un sistema a otro conservando la calidad existente en un contexto específico de uso.
Fiabilidad (Reliability)	El grado en el cual el dato puede mantener y conservar un nivel especificado de operaciones y calidad, aún en caso de falla.

Tabla 6: Características dependientes del sistema según ISO/IEC 25012
Fuente: Guerra García, César (2009)

Características inherentes y dependientes del sistema	
Dimensión	Descripción
Accesibilidad (Accessibility)	El grado en el cual el dato puede ser accesado en un contexto específico de uso, en particular por la gente que necesita el soporte de tecnología o una configuración especial debido a alguna inhabilidad (incapacidad).
Conformidad (Compliance)	El grado en el cual el dato tiene atributos que se adhieren a normas, convenciones o regulaciones vigentes y reglas similares relacionadas con la calidad de datos en un contexto específico de uso.
Confidencialidad (Confidentiality)	El grado en el cual el dato tiene los atributos que aseguran que solo es accesible e interpretable por usuarios autorizados en un contexto específico de uso.
Eficiencia (Efficiency)	El grado en el cual el dato tiene los atributos que pueden ser procesados, y proporciona los niveles esperados de funcionamiento (desempeño) usando las cantidades y los tipos de recursos apropiados en un contexto específico de uso.
Precisión (Precision)	El grado en el cual el dato tiene atributos que son exactos o que proporcionan la discriminación en un contexto específico de uso.
Trazabilidad (Traceability)	El grado en el cual el dato tiene atributos que proporcionan un rastro de auditoría de acceso a los datos y de cualquier cambio hecho a los datos en un contexto específico de uso.
Entendibilidad (Understandability)	El grado en el cual el dato tiene atributos que le permiten ser leído e interpretado por usuarios, y es expresado en lenguajes apropiados, símbolos y unidades en un contexto específico de uso.

Tabla 7: Características inherentes y dependientes del sistema según ISO/IEC 25012
Fuente: Guerra García, César (2009)

ISO 9126		ISO/IEC 25012
Característica	Subcaracterística	Característica
Funcionalidad	Idoneidad	Consistencia Actualidad Compleitud Precisión
	Exactitud	Exactitud
	Interoperabilidad	
	Seguridad	Seguridad
Fiabilidad		Disponibilidad
	Madurez	
	Tolerancia a fallos Facilidad de recuperación	Recuperabilidad
Usabilidad	Facilidad de comprensión	Entendibilidad
	Facilidad de aprendizaje	
	Operatividad	Accesibilidad Manejabilidad
Eficiencia	Tiempo de uso Recursos utilizados	Eficiencia
Mantenibilidad	Facilidad de análisis	
	Facilidad de cambio	Facilidad de cambio
	Estabilidad Facilidad de prueba	
Portabilidad	Facilidad de instalación Facilidad de ajuste Facilidad de adaptación al cambio	Portabilidad

Tabla 8: Correspondencia entre las características del ISO 9126 e ISO/IEC 25012

Fuente: González Pinzón, Miguel y Gonzáles Sanabria, Juan (2013)

Cabe señalar que los indicadores utilizados en el trabajo de investigación son: accesibilidad, fiabilidad, eficiencia y disponibilidad.

2.3 Glosario de términos

- Bloques horarios: Son periodos horarios en los que los costos de generación son similares, determinados en función a las características técnicas y económicas del sistema. (Comité de Operaciones Económicas del Sistema Eléctrico Interconectado Nacional, 2000)
- Grandes Usuarios Libres: Usuarios Libres con una potencia contratada igual o superior a 10MW, o agrupaciones de Usuarios Libres cuya potencia contratada total sume por lo menos 10MW. (Organismo Supervisor de la Inversión en Energía y Minería, 2006)
- Integración (de datos): podemos definir la integración de datos como un proceso de transformación y conciliación de datos que permita una mayor agilidad en la gestión, proporcionando datos conectados, seguros y de calidad. Integrar significa combinar datos que se encuentran en diferentes fuentes para permitirle al usuario final tener una vista unificada de los mismos para una accesibilidad idónea, que sirva a las necesidades de negocio. (Power Data Especialistas en Gestión de Datos, 2015)
- Llaves artificiales: Un sustituto es un identificador asignado a la representación de un objeto por el propio sistema de base de datos, cuando esa representación del objeto se ingresa en la base de datos. Si suponemos que cada objeto está representado por una tupla en la base de datos, cuando se crea una tupla, el sistema de DB crea un sustituto nuevo que permanece invariable incluso cuando se cambian los valores en la tupla. Ese sustituto se puede ver como el DB interno representativo del objeto del mundo real. (Wieringa, R. J. y De Jonge, W., 1991)
- Medidores de bornes de generación: La demanda de kW se mide por medio de su medidor eléctrico como el promedio más alto de la demanda durante un período de 15 minutos del mes. (Mike M., s.f.)
- Power BI: es un conjunto de aplicaciones de análisis de negocios que permite analizar datos y compartir información. También permite acceder a sus datos e informes desde cualquier lugar, y estos se actualizan automáticamente con los cambios que se realizan en los datos. (Microsoft, 2017)
- Power Pivot: es una tecnología que permite procesar y analizar información compleja de forma muy eficiente y rápida. Su principal característica y la que le aporta una enorme velocidad es que es una tecnología “In-Memory”, es decir que tiene todos los datos en memoria con una alta compresión, para disminuir el espacio que ocupa y aumentar la

velocidad de análisis. Estos modelos de datos están basados en tablas y relaciones. (Salvador Ramos, 2015)

- SCADA: es una aplicación o conjunto de aplicaciones de software especialmente diseñadas para funcionar sobre ordenadores de control de producción, con acceso a la planta mediante la comunicación digital con instrumentos y actuadores, e interfaz gráfica de alto nivel para el operador (pantallas táctiles, ratones o cursores, lápices ópticos, etc.). (Pérez López, Esteban, 2015)
- SEIN: Sistema Eléctrico Interconectado Nacional. Conjunto de líneas de transmisión y subestaciones eléctricas conectadas entre sí, así como los respectivos centros de despacho de carga, que permite la transferencia de energía eléctrica entre dos o más sistemas de generación. (Comité de Operaciones Económicas del Sistema Eléctrico Interconectado Nacional, 2000)
- SQL Server Analysis Services: es un producto que cubre el área de Business Inteligence, permite realizar modelos multidimensionales, tabulares y con Power Pivot. (J. Morán, Luis, 2016)
- SQL Server Integration Services: es un servicio que se utiliza para realizar procesos ETL (Extracción, Transformación y Carga (Load) que por regla general ayudan a cargar modelos multidimensionales o tabulares, aunque también para procesos relacionales. (J. Morán, Luis, 2016)
- SQL Server Reporting Services: es un servicio que incluyen un conjunto de herramientas que permite el diseño, la administración y el envío de informes. (J. Morán, Luis, 2016)
- Usuarios Libres: Usuarios conectados al SEIN no sujetos a regulación de precios por la energía o potencia que consumen. (El Peruano Diario Oficial, 2009)

CAPÍTULO 3: VARIABLES E HIPÓTESIS

3.1 Variables e Indicadores

a. Identificación de Variables

- Variable Independiente: Implementación de un Datamart.
- Variable Dependiente: Mejora en la toma de decisiones en el control de la demanda eléctrica

b. Operacionalización de Variables

Indicadores Variable Independiente:

- Nivel de accesibilidad
- Nivel de fiabilidad
- Nivel de eficiencia
- Nivel de disponibilidad

Indicadores Variable Dependiente:

- Cantidad de reportes de comportamiento de la demanda eléctrica.
- Tiempo en la generación de reportes del crecimiento de la demanda eléctrica.
- Volumen de información histórica para analizar sobre demanda eléctrica.
- Número de empresas que cumplen con enviar su información de demanda eléctrica.

3.2 Hipótesis

Hipótesis General

La implementación de un datamart influye positivamente en la mejora en la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional.

Hipótesis Específicas

- El nivel de accesibilidad de un datamart influye positivamente en la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional.
- El nivel de fiabilidad de un datamart influye positivamente en la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional.

- El nivel de eficiencia de un datamart influye positivamente en la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional.
- El nivel de disponibilidad de un datamart influye positivamente en la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional.

3.3 Matriz de coherencia interna

En el Anexo VI se encuentra la matriz de coherencia interna.



CAPÍTULO 4: METODOLOGÍA DE DESARROLLO

Para la implementación del datamart se utilizó la metodología de Ralph Kimball debido al nivel de adaptabilidad y madurez que muestra en relación con el desarrollo de la solución tecnológica. En el presente capítulo, se procedió a evaluar distintas metodologías para el desarrollo de un modelo dimensional, con lo que se obtuvo el análisis necesario para definir y elegir la más apropiada para el desarrollo del datamart. Por último, se presenta las etapas, actividades y artefactos propuestas por la metodología, indicando las utilizadas para el desarrollo de la solución tecnológica.

4.1 Comparación de metodologías para desarrollo de un Datamart

Para seleccionar la metodología adecuada en el desarrollo de la tesis, nos apoyamos en la Tabla 9 que muestra un cuadro comparativo, en el cuál se observan 4 metodologías que definen distintos pasos para lograr desarrollar un modelo dimensional:

Metodología	Descripción	Tipo de Proyecto	Características propias de la Metodología
Ralph Kimball	Metodología bottom-up para desarrollo de estructuras dimensionales.	Datawarehouse y Datamart	Enfocado en la elaboración de un datawarehouse mediante los datamarts
Bill Inmon	Metodología de desarrollo top-down, para el desarrollo de estructuras dimensionales.	Datawarehouse y Datamart	Propone la creación de un datawarehouse, y posterior a ello, usar el almacén de datos para generar los datamarts
Business Intelligence Road Map	Metodología para la implementación de Business Intelligence.	Datawarehouse	Define un proceso mediante el cual se generan KPI's que permitan generar una aplicación para el soporte de decisiones. Se asocia a la metodología RUP.
Hefesto	Metodología para el desarrollo de un datawarehouse o un datamart	Datawarehouse y Datamart	Es una metodología que se encuentra en continuo crecimiento. Su fase final es el ETL.

Tabla 9: Comparación de metodologías para el desarrollo de estructuras dimensionales
Fuente: Elaboración propia

Complementando a los mostrados anteriormente, se realizó un análisis para determinar la metodología idónea para el desarrollo del presente trabajo de investigación; como niveles de evaluación se tomaron los establecidos en la Tabla 10, los cuales se emplearon para establecer los valores de cada metodología según los criterios definidos en la Tabla 11 y Tabla 12.

Nivel de evaluación	Puntaje
Bajo	1
Medio	2
Alto	3

Tabla 10: Niveles de evaluación para el análisis comparativo de metodologías
Fuente: Elaboración propia

Criterio de evaluación	Ralph Kimball	Bill Inmon	Business Intelligence Road Map	Hefesto
Frecuencia de uso de la metodología en el mercado.	3	3	3	1
Relación con la adaptabilidad al desarrollo de datamart.	3	2	2	3
Adaptabilidad de fases, herramientas y buenas prácticas.	3	3	3	3
Tiempo que dura el proceso de la metodología.	2	2	3	2
Nivel de documentación.	2	2	3	2
Nivel de comunicación con el cliente.	3	3	3	3
Nivel de relevancia en generar un datamart mediante sus fases.	3	2	1	1
Total	19	17	18	15

Tabla 11: Evaluación general de metodologías propuestas
Fuente: Elaboración propia

Característica	Descripción	Ralph Kimball	Bill Inmon	Business Intelligence Road Map	Hefesto
Flexibilidad	¿Cómo es el nivel de reacción frente a cambio que pueden darse durante el proceso?	3	3	3	2
Calidad	¿Cuál es el nivel de calidad del producto obtenido?	3	3	3	3
Riesgos	¿Cuál es el nivel de riesgos?	2	2	2	3
Velocidad	¿El método produce resultados de forma rápida?	3	3	2	3
Integración	¿Permite integrar la información general de la organización?	3	3	3	3
Total		14	14	13	14

Tabla 12: Evaluación de metodologías propuestas según características
Fuente: Elaboración propia

Así mismo, se analizó el nivel de respuesta de cada metodología frente a criterios relacionados al desarrollo de la tesis como proyecto tecnológico, de tal forma de observar que metodología se acoplaría mejor para la realización de la tesis, con el fin de obtener un nivel de respuesta idóneo frente a posibles eventos relevantes que influyen en la gestión o planificación de éste. Por ello, en la Tabla 13, se definen los niveles de respuesta para evaluar los distintos criterios presentados en la Tabla 14 y Tabla 15

Nivel de respuesta	Puntaje
Bajo	1
Medio	2
Alto	3

Tabla 13: Niveles de evaluación de repuesta para el análisis comparativo de metodologías
Fuente: Elaboración propia

Criterio de evaluación	Ralph Kimball	Bill Inmon	Business Intelligence Road Map	Hefesto
El proyecto puede verse afectado a cambios, pero generalmente son muy pequeños y ligeros.	3	3	3	2
El alcance del proyecto implica que el desarrollo de éste sea por fases.	3	3	3	3
Se necesita mantener un control continuo durante y a término de cada etapa.	3	3	3	2
Total	9	9	9	7

Tabla 14: Evaluación de metodología propuesta según criterio en la relación con el proyecto
Fuente: Elaboración propia

Criterio de evaluación	Ralph Kimball	Bill Inmon	Business Intelligence Road Map	Hefesto
El proyecto consta de un desarrollo complejo, ¿cuál es el nivel de la metodología en cuanto a este punto?	3	3	3	1
El equipo de desarrollo es muy corto para el proyecto, ¿cuáles serían de acuerdo con lo visto los niveles de relación a este punto?	3	3	2	3
La evolución del proyecto debe darse en relación con un conjunto de fase, ¿cuáles serían los niveles de desarrollo?	3	3	3	3
Nivel de la metodología con relación al desarrollo del presente proyecto.	3	1	2	1
Total	12	10	10	8

Tabla 15: Evaluación de metodología propuesta según criterios generales del proyecto
Fuente: Elaboración propia

Según como lo mencionan Leonard Brizuela, Eric Ismael y Castro Blanco, Yudi (2013), para seleccionar una metodología que nos permita el desarrollo de un proyecto de estas magnitudes, no es recomendable utilizar aquellas que necesiten una extensa fase para reunir los requerimientos; como objetivo se debe trazar la entrega de una primera implementación que demuestre las ventajas de un almacén de datos, con el fin de mejorar la calidad, buscando la satisfacción de los usuarios, y que éstos se vean beneficiados con la implementación de dicho proyecto.

4.2 Metodología de Ralph Kimball

La metodología de Ralph Kimball es un estándar para el desarrollo de almacenes de datos (datawarehouse), que busca la creación de una estructura dimensional mediante la recolección de información que una empresa maneje en sus distintos repositorios de datos. Dicha metodología presenta un enfoque centrado en el negocio, y por ello se requiere la construcción de una estructura adecuada en la cual la información pueda ser correctamente almacenada y consumida.

Se ha utilizado la presente metodología debido a que su desarrollo, en comparación a otras, se encuentra orientado a la elaboración de un datamart al ser un desarrollo bottom-up (de abajo hacia arriba); a diferencia de otras metodologías que buscan realizar únicamente la construcción de un datawarehouse; mediante el ciclo de vida dimensional que propone Ralph Kimball, promueve que la elaboración de un datawarehouse debe estar encaminado mediante la creación de los datamarts, al encontrar dicha diferencia entre las distintas metodologías para el desarrollo de un modelo dimensional, se identificó que ésta es propicia para lograr el desarrollo de la tesis según los alcances definidos y con ello cumplir los objetivos planteados.

4.2.1 Etapas de la metodología

La metodología Kimball presenta un conjunto de pasos los cuales direccionan el desarrollo del presente trabajo de investigación, con el fin de cumplir los objetivos establecidos en la tesis. Cabe resaltar que el alcance para el desarrollo abarca la elaboración de un datamart, por ello, la presente metodología ha sido adecuada a la solución propuesta con el fin de emplear los artefactos y procedimientos establecidos con relación a la necesidad de la tesis.

Para definir las etapas que conlleva la metodología Kimball, los autores Kimball, Ralph y Ross, Margie (2013), proponen una arquitectura de desarrollo la cual se presenta en la Figura 17, en la que se observa las etapas de la metodología seleccionada.

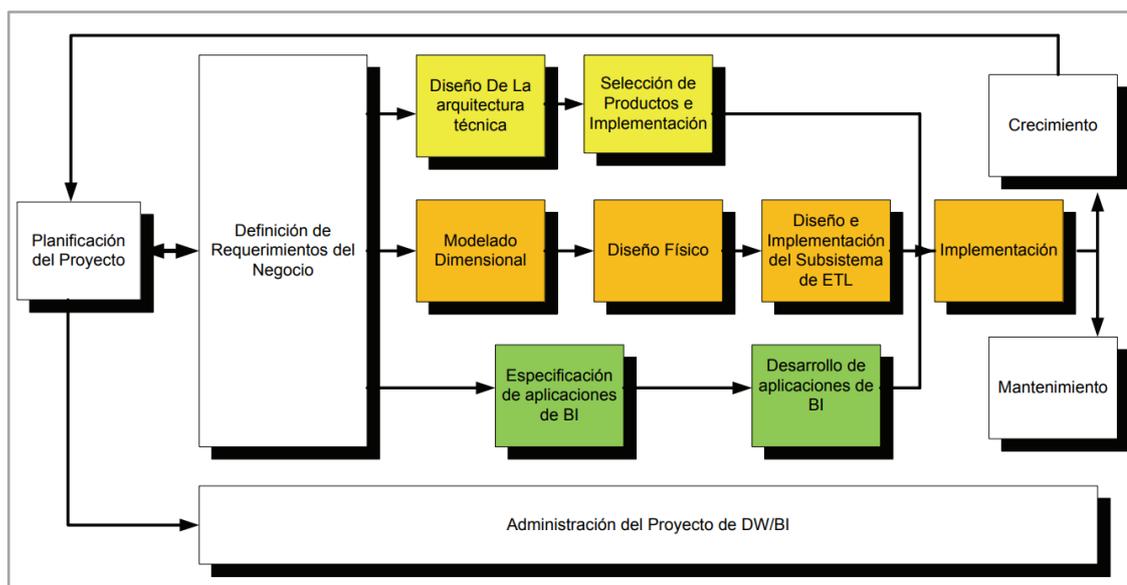


Figura 17: Diagrama del ciclo de vida de la metodología Kimball
Fuente: Kimball, Ralph y Ross, Margy. (2013)

A continuación, se procede a describir las etapas de la metodología Kimball, las cuales se han adaptado al desarrollo según las necesidades de la presente tesis:

- **Planificación del proyecto:** Definir cuál es el propósito del proyecto, los objetivos específicos y el alcance del mismo. Para ello se tomó en cuenta las siguientes tareas:
 - Definir el alcance.
 - Identificar las tareas a realizar.
 - Programar las tareas identificadas.
- **Definición de requerimientos del negocio:** Conocer las necesidades que tienen, tanto el personal de negocio, así como los especialistas de la empresa (técnicos). Para ello existe la necesidad de aprender del negocio, los competidores, la industria y los clientes, por lo que se tomó en cuenta lo siguiente:
 - Realizar entrevistas del personal de interés.
- **Diseño de la arquitectura técnica:** Definir las tecnologías que se van a integrar para el desarrollo del datamart. Para ello se tomó en cuenta las siguientes tareas:
 - Realizar el diagrama que muestre la arquitectura a utilizar para el desarrollo del proyecto.
 - Realizar el diagrama que describa la interacción de las tecnologías implementadas.
- **Diseño físico:** Definir los requerimientos físicos. Para ello se tomó en cuenta las siguientes tareas
 - Determinar el tamaño del sistema de DW/BI.

- Seleccionar las herramientas a instalar en las estaciones de trabajo del equipo de desarrollo.
- **Modelado dimensional:** Definir el modelo que usará la solución, señalando las medidas y las dimensiones a utilizar. Para ello se tomó en cuenta las siguientes tareas:
 - Seleccionar el proceso de negocio.
 - Delimitar el nivel de granularidad.
 - Definir las dimensiones, identificar las medidas y tablas de hechos.
- **Diseño e implementación del subsistema de ETL:** Diseñar adecuadamente el subsistema de extracción, transformación y carga de la información que se llevará al repositorio de datos dimensional, desde las fuentes de origen de datos definidas. Para ello se tomó en cuenta las siguientes tareas:
 - Realizar la extracción de información necesaria para el proyecto.
 - Transformar y preparar la data para cargar en el repositorio dimensional.
 - Realizar la cargar desde los archivos fuentes hacia el repositorio dimensional.
- **Especificación de aplicaciones de BI:** Especificar como se va a proporcionar acceso a la información que se encuentra disponible en el datamart. Para ello se tomó en cuenta las siguientes tareas:
 - Elaborar informes estandarizados.
 - Definir herramientas para acceso de lectura al datamart.
- **Desarrollo de aplicaciones de BI:** Desarrollo de las aplicaciones que especificadas para que el usuario tenga acceso a la información.
- **Implementación:** Pase a producción de las aplicaciones establecidas para los usuarios finales, con las cuales podrán acceder y utilizar el datamart.

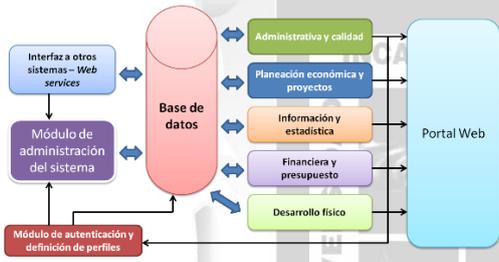
4.2.2 Artefactos

Para el correcto desarrollo de la metodología seleccionada, se ve por conveniente generar los entregables que Kimball establece en sus etapas. Por ello, a continuación, se presentan los artefactos a emplear:

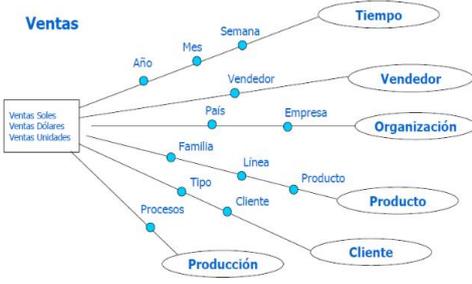
- **Entrevista**

	<p>Ésta herramienta, consiste en la interacción con los actores del negocio, con la finalidad de generar un feedback para el desarrollo del proyecto, y que ello permita detectar los requerimientos funcionales y no funcionales.</p>
---	--

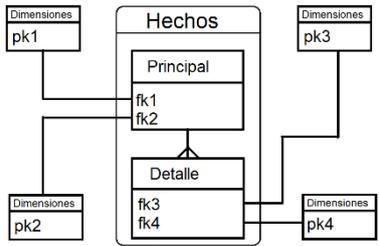
- **Diagrama de arquitectura del sistema**

	<p>Éste diagrama permite entender, de manera simplificada, la distribución de tecnologías que el proyecto va a utilizar.</p>
--	--

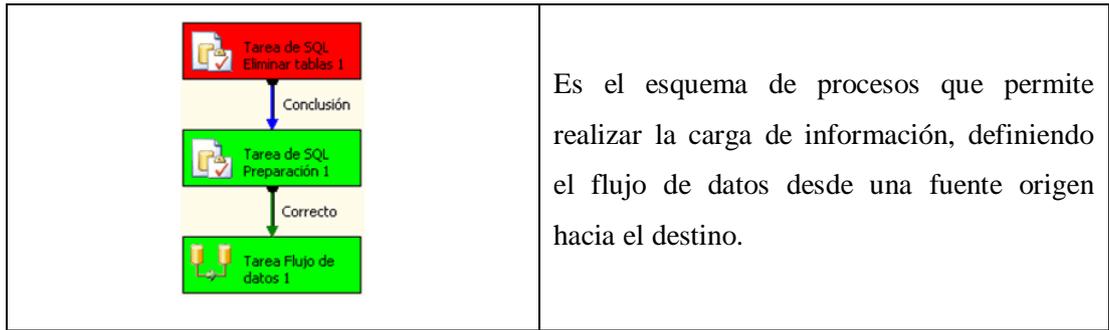
- **Esquema de granularidad**

	<p>Éste esquema presenta los niveles de granularidad (detalle) que manejan cada una de las dimensiones en base a las medidas que se han identificado.</p>
---	---

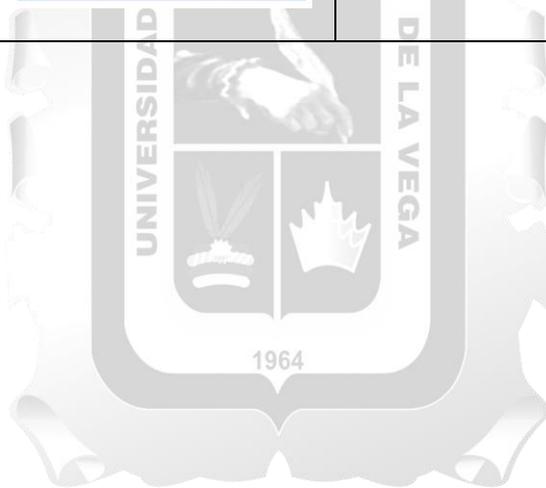
- **Modelo de Datos**

	<p>El modelo de datos muestra el esquema dimensional definido para el desarrollo del proyecto. En éste se observarán las dimensiones y las tablas hechos, así como sus relaciones.</p>
---	--

- **Paquete ETL**



- **Informes estándar**



CAPÍTULO 5: SOLUCIÓN TECNOLÓGICA

El capítulo 5 contiene el desarrollo de la solución tecnológica mediante la aplicación de la metodología seleccionada, en este caso se observa el proceso de desarrollo del datamart mediante la tecnología de Ralph Kimball según lo definido en el capítulo 4.

5.1 Planificación del proyecto

- Alcance: El desarrollo del datamart tiene como alcance la información que el COES maneja para analizar la demanda eléctrica de los Usuarios Libres desde enero de 2010 a junio de 2017.

La información fuente a utilizar para el desarrollo del proyecto, es la siguiente:

- Demanda ejecutada de Usuarios Libres.
- Demanda programada de Usuarios Libres.
- Demanda ejecutada en el SEIN.
- Demanda programada en el SEIN.
- Producción de electricidad de medidores de generación.
- Máxima demanda en el SEIN.
- Producto bruto interno (PBI).
- Desviación estándar de la demanda en el SEIN.
- Error porcentual absoluto medio para el pronóstico de la demanda.
- Información de Usuarios Libres de OSINERGMIN (Puntos de suministros y máxima demanda).

Cabe resaltar que para el desarrollo de la solución tecnológica no se tuvo acceso a la base de datos de manera directa para poder extraer la información mencionada, sin embargo, se tiene un sistema web (página web) que generan reportes en hojas de cálculo que permitieron obtener los datos necesarios para el desarrollo, pero con ello se debe contemplar que los sistemas en mención no permiten descargar grandes intervalos de tiempo, en su mayoría permiten obtener la información en formatos de mes a mes.

- Tareas identificadas: Para el desarrollo del datamart, se establecieron las siguientes tareas principales:
 - Identificación de las fuentes de información para obtener los datos a utilizar en el datamart.
 - Extracción y descarga de los archivos fuente para cargar al motor de base de datos SQL Server.
 - Transformación de los archivos fuente en los formatos requeridos y en el tipo de archivo (.txt) necesario para cargar al motor de base de datos SQL Server.

- Carga de archivos al motor de base de datos SQL Server mediante el Microsoft SQL Server Management Studio.
 - Creación de las bases de datos relacionales consolidadas con la información cargada. Para ello se crearon 2 bases de datos: COES y Provisional COES; la primera es una réplica de la base de datos del caso de estudio elaborada con la información obtenida, y la segunda es un repositorio de datos con la información externa que el COES utiliza en el control de la demanda eléctrica.
 - Integración de las bases de datos en un modelo dimensional (datamart), mediante el uso de queries de carga y la herramienta Microsoft SQL Server Integration Services.
 - Creación del cubo multidimensional mediante la herramienta Microsoft SQL Server Analysis Services.
 - Elaboración de informes personalizados mediante la herramienta Power BI.
- Programar las tareas identificadas: En la Tabla 16 se presenta la planificación establecida para el desarrollo de las actividades establecidas, así mismo la Figura 18 muestra las etapas del proceso de desarrollo del datamart:



Actividad	Semana											
	1	2	3	4	5	6	7	8	9	10	11	12
Identificación de las fuentes de información.												
Extracción y descarga de los archivos fuente.												
Transformación de los archivos fuente en los formatos requeridos y en el tipo de archivo necesario.												
Carga de archivos al motor de base de datos SQL Server.												
Creación de las bases de datos relacionales consolidadas con la información cargada.												
Integración de las bases de datos en un modelo dimensional.												
Creación del cubo multidimensional.												
Elaboración de informes personalizados.												

Tabla 16: Cronograma de tareas identificadas para el desarrollo del Datamart
Fuente: Elaboración propia

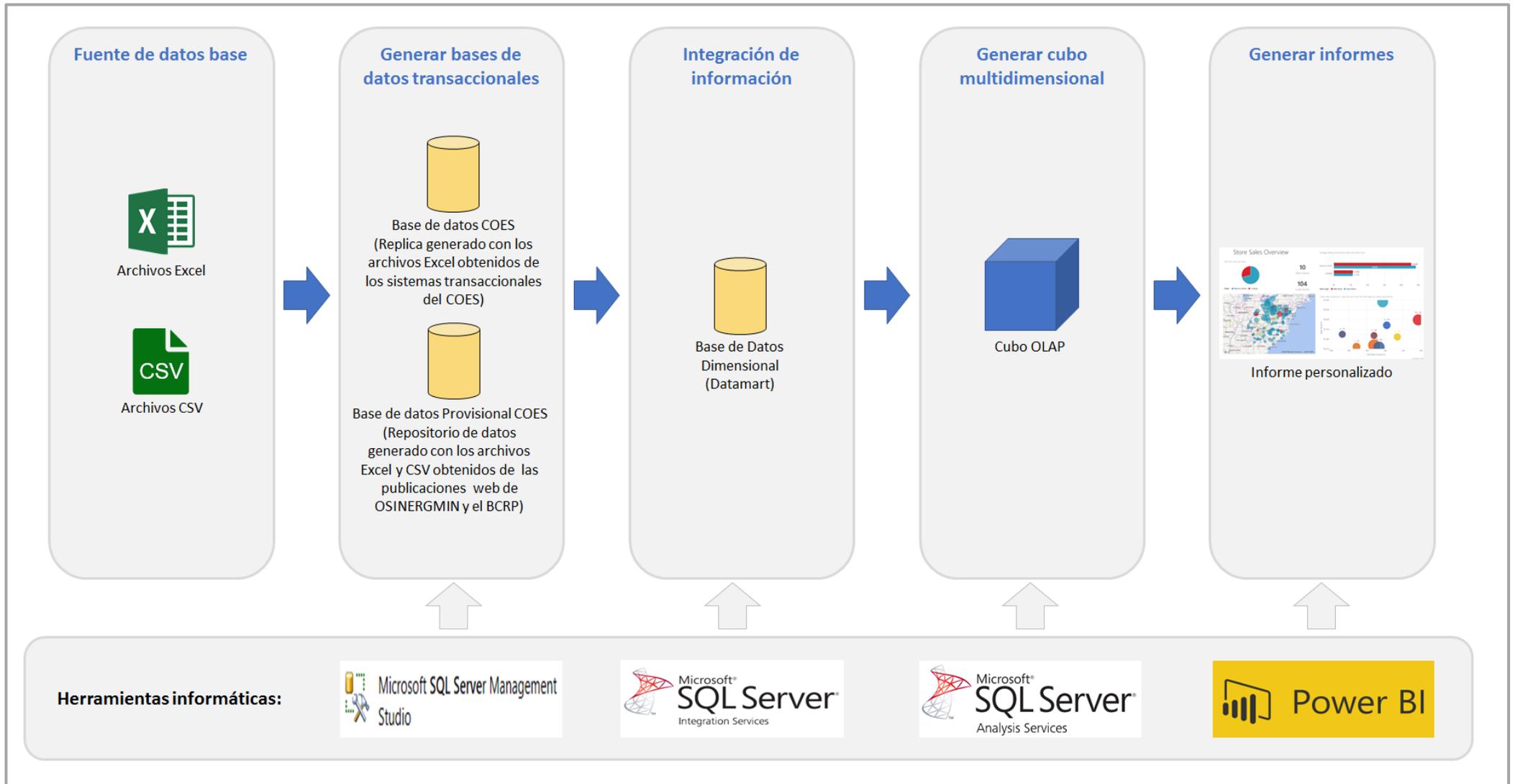


Figura 18: Flujo de las etapas para el desarrollo del proyecto

Fuente: Elaboración propia

5.2 Definición de requerimientos del negocio

Como se ha mencionado en los capítulos anteriores, el COES es una empresa que maneja información diaria, con ello se entiende que la cantidad de registros que se puede obtener en 1 año es amplia; debido a que la información es remitida por muchas empresas que cuentan con distintas cantidades de puntos de medición, y si a ello se agrega que la información remitida por dichas empresas es registrada en intervalos de tiempos entre 15 a 30 minutos, ello aumenta en una gran cantidad los datos que se almacenan. Se tiene conocimiento del negocio, los procesos y las dificultades que se tienen, así mismo se analizó la entrevista del Anexo VII. Para definir los requerimientos se determinaron los siguientes puntos:

- **¿Cuál es la finalidad del datamart?**

La implementación del datamart tiene como finalidad mantener una estructura acorde a las nuevas tendencias tecnológicas, para aprovechar el almacenamiento de los datos históricos en una estructura de almacenamiento organizada y ordenada bajo los parámetros de análisis para la toma de decisiones del control de la demanda eléctrica.

- **¿Cómo impactará en la organización?**

Se obtendrá un amplio beneficio en cuanto a la disponibilidad de información, con la cual se podrá trabajar. Con ello la elaboración de indicadores y reportes gráficos también se verá beneficiada, gracias a la estructura y dimensiones de análisis que se emplea en el datamart.

- **¿Cuáles son los riesgos al elaborar el datamart?**

El principal riesgo de la elaboración de un datamart, es definir las características del proceso de carga, esto se debe a que el proceso de extracción, transformación y carga es muy relevante para obtener data confiable y en los parámetros establecidos.

Como parte de la definición de requerimientos se obtuvo lo siguiente:

- La información de demanda eléctrica puede ser actualizada, por ello se vio la necesidad de realizar cargas completas y no cargas incrementales.
- Las principales dimensiones seleccionadas para analizar la demanda fueron: ubicación, actividad económica, tiempo, hora y empresa.
- En la mayoría de dimensiones existen niveles de jerarquías, por ello se requirió especificar el nivel de granularidad de cada una de estas.

5.3 Diseño de la arquitectura técnica

- Diagrama arquitectura para el desarrollo del proyecto: Para comprender el proceso realizado para la elaboración del modelo dimensional y los informes personalizados, en la Figura 19 se muestra el flujo para el desarrollo del proyecto, y como éstas se relacionan con las distintas herramientas tecnológicas:

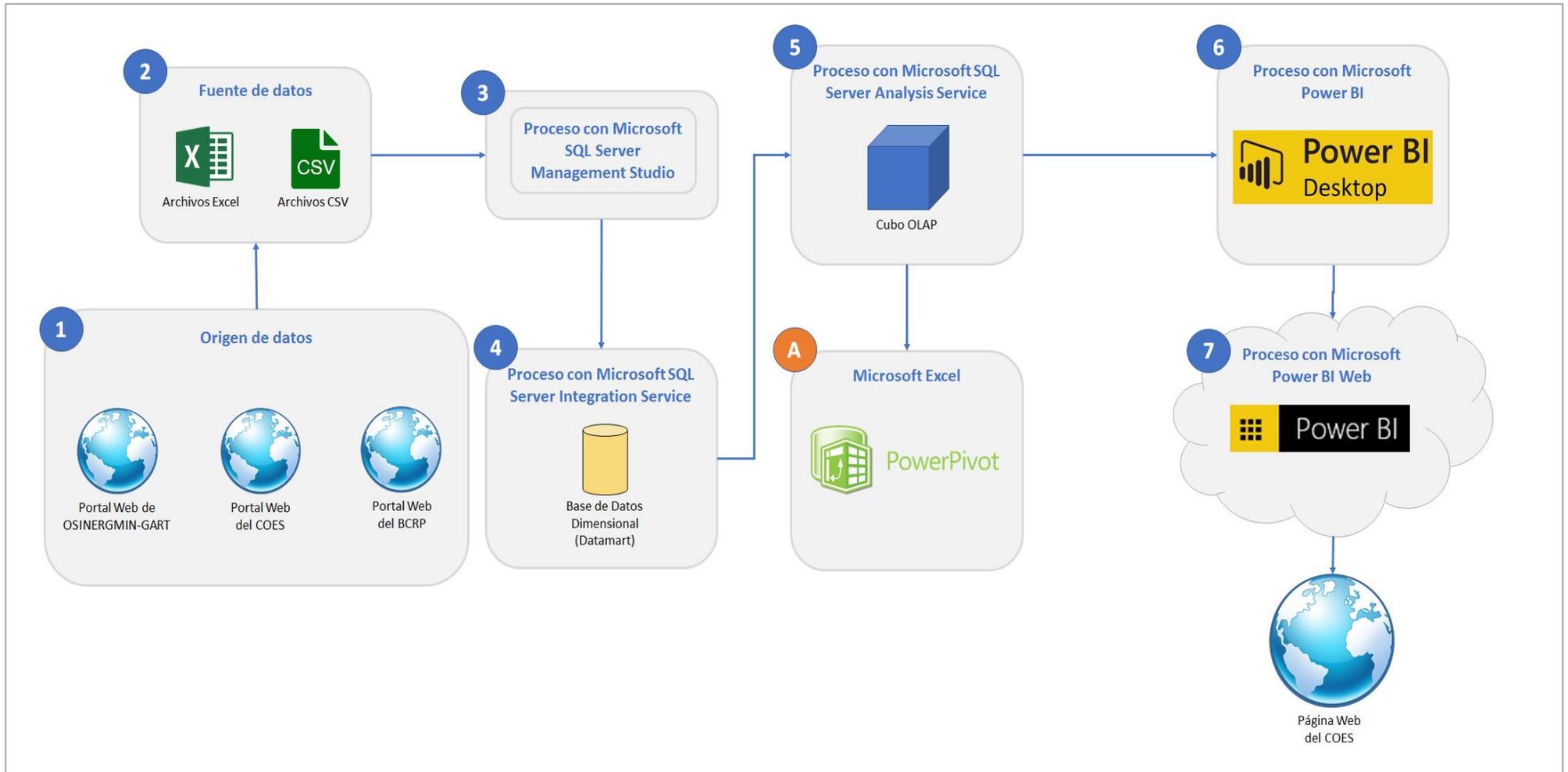


Figura 19: Diagrama de arquitectura para el desarrollo del proyecto
Fuente: Elaboración propia

A continuación, en la Tabla 17 se explica cada punto del diagrama de arquitectura para el desarrollo del proyecto, de tal manera que sirva para comprender el desarrollo realizado:

Nro.	Titulo	Descripción
1	Origen de datos	Primero se definió los sistemas y páginas web de las cuáles se iban a obtener los datos necesarios para generar el datamart.
2	Fuente de datos	Una vez definidos los orígenes de datos, se obtuvieron los archivos fuentes que contenían la información a utilizar en el datamart. La gran mayoría de archivos eran hojas de cálculo Excel, cuyos formatos fueron adaptados para realizar la carga al motor de base de datos. La información que se encontraba en los archivos Excel, fueron replicada en archivos “.txt”, a fin de realizar la carga sin inconvenientes.
3	Proceso con Microsoft SQL Server Management Studio	Posterior a la obtención de archivos de cargas, se procedió a utilizar la herramienta Microsoft SQL Server Management Studio para realizar la carga de datos por meses en distintas tablas que posteriormente se consolidaron en una sola tabla para que puedan ser integradas en el modelo dimensional, para ello se generaron 2 bases de datos OLTP: COES y Provisional.
4	Proceso con Microsoft SQL Server Integration Service	Una vez obtenidas las bases OLTP con los datos cargados y estructurados correctamente, se procedió a generar la base de datos dimensional (OLAP), y mediante la herramienta Microsoft SQL Server Integration Service se procedió a realizar el proceso de carga en el datamart.
5	Proceso con Microsoft SQL Server Analysis Service	Con la obtención del datamart, se procedió a generar los cubos OLAP, cada tabla hecho fue procesada mediante las distintas dimensiones a las que se les asoció.

6	Proceso con Microsoft Power BI Desktop	Para elaborar los distintos informes estandarizados, se utilizó la herramienta Power BI Desktop, se diseñaron distintos reportes los cuales permiten la toma de decisiones para mejorar el control de la demanda eléctrica.
7	Proceso con Microsoft Power BI Web	Una vez generados los informes estandarizados, se generó un link mediante el Power BI en Web, para que puedan ser incrustados en la página web del COES y los agentes del SEIN puedan acceder a dicha información.
A	Microsoft Excel (Power Pivot)	Como complemento de la solución tecnológica, se propuso la inclusión del complemento Power Pivot en las estaciones de trabajo de los especialistas, de tal manera que tengan acceso a la información del datamart y puedan trabajar con dicha fuente de datos.

Tabla 17: Descripción del diagrama de arquitectura para el desarrollo del Datamart
Fuente: Elaboración propia

- Diagrama de arquitectura de las tecnologías implementadas: Para la implementación del proyecto fue necesario comprender la interacción de las distintas tecnologías que se utilizan. Para ello en la Figura 20, podemos observar cómo interactúan los distintos componentes tecnológicos empleados en el desarrollo del proyecto:

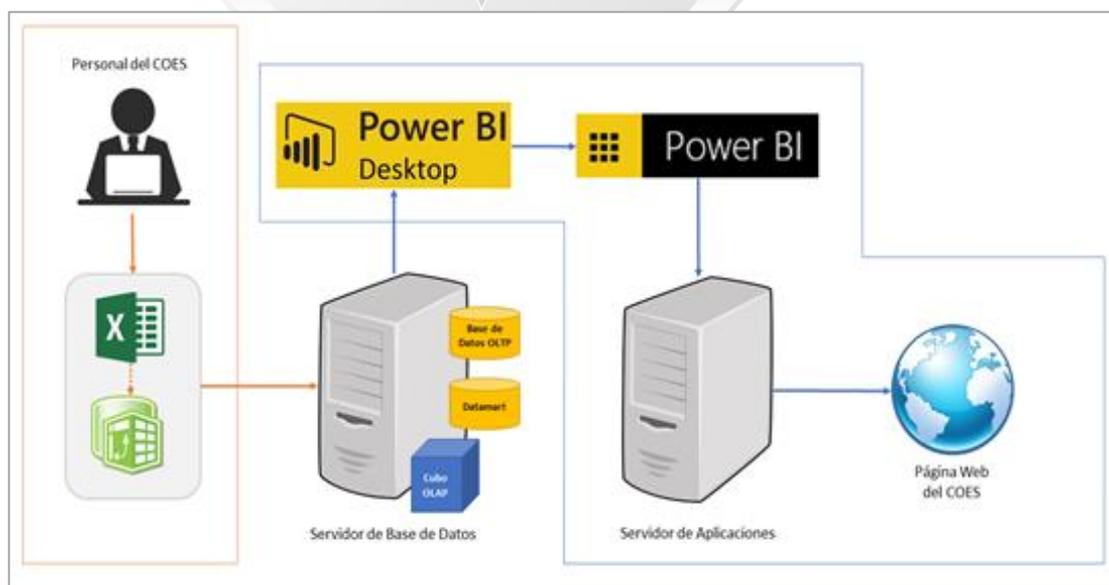


Figura 20: Diagrama de arquitectura de las tecnologías implementadas
Fuente: Elaboración propia

5.4 Diseño físico

- Tamaño del sistema de DW/BI: El tamaño del datamart ha sido definido por la cantidad de meses que se han obtenido. Para este caso se utilizó información disponible desde Enero del 2010 hasta Junio del 2017 de las distintas fuentes de información que servirán como medidas en el modelo dimensional (datamart).
- Herramientas tecnológicas para el desarrollo del proyecto: Para el desarrollo e implementación del datamart y los distintos informes personalizados, se utilizaron las herramientas tecnológicas presentadas en la Figura 21:



Figura 21: Herramientas tecnológicas utilizadas para el desarrollo del datamart
Fuente: Elaboración propia

Para el proceso de selección de herramientas se analizaron diversas propuestas tecnológicas, finalmente se optó por alinear la solución a las herramientas que Microsoft ofrece, ello debido al alto nivel en el que se maneja en el mercado, tal y como se puede apreciar en los Anexos VIII y IX. Así mismo, cabe resaltar que la tecnología utilizada por el COES se encuentra alineada a Microsoft, siendo ello un factor relevante para la selección de las herramientas tecnológicas a empleadas en el desarrollo de la presente tesis. Por último, cabe señalar que el factor económico ha sido otra variable que se ha tomado en cuenta para realizar la selección de las herramientas tecnológicas

5.5 Modelado dimensional

- Seleccionar el proceso de negocio: El datamart tiene como principal abarcar el proceso de control de la demanda eléctrica en el SEIN. Cabe resaltar que la información con la que va a contar el datamart también servirá de apoyo para otros procesos en los cuales, la demanda eléctrica influya, un caso de ejemplo es el mercado mayorista de electricidad que se presenta en el Procedimiento Técnico N°2 del COES.
- Nivel de granularidad: Se presenta el nivel de granularidad que se utilizó para las dimensiones Horario, Tiempo, Ubigeo, Sector, Actividad Económica, Usuarios Libres OS, Usuarios Libres COES y Generador, en las Figuras 22, 23, 24, 25 ,26 ,27,28 y 29 respectivamente:

- Dimensión Horario:

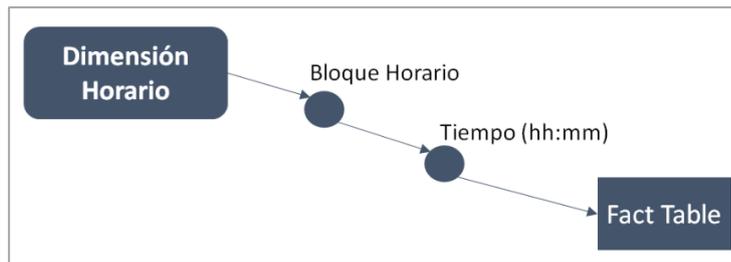


Figura 22: Nivel de granularidad de la dimensión Horario
Fuente: Elaboración propia

- Dimensión Tiempo:



Figura 23: Nivel de granularidad de la dimensión Tiempo
Fuente: Elaboración propia

- Dimensión Ubigeo:

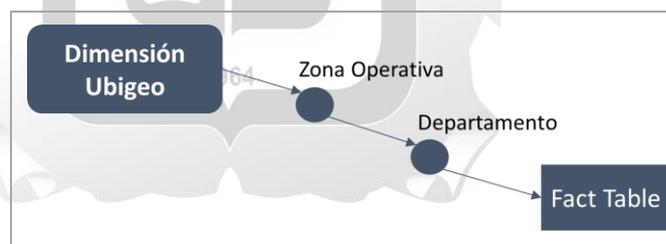


Figura 24: Nivel de granularidad de la dimensión Ubigeo
Fuente: Elaboración propia

- Dimensión Sector:

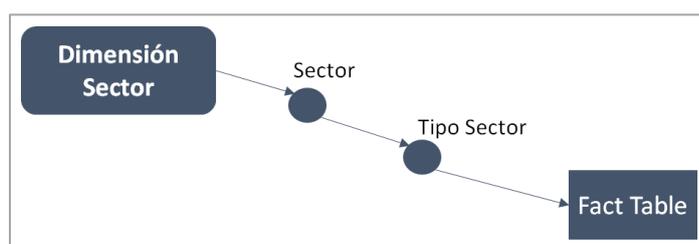


Figura 25: Nivel de granularidad de la dimensión Sector
Fuente: Elaboración propia

- Dimensión Actividad Económica:

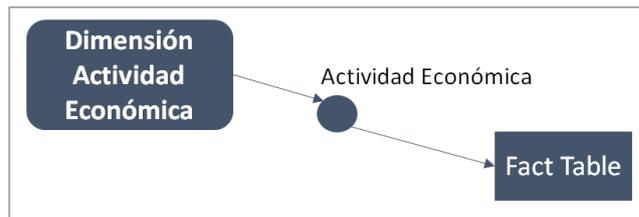


Figura 26: Nivel de granularidad de la dimensión Actividad Económica
 Fuente: Elaboración propia

- Dimensión Usuario Libre OS (OSINERGMIN):

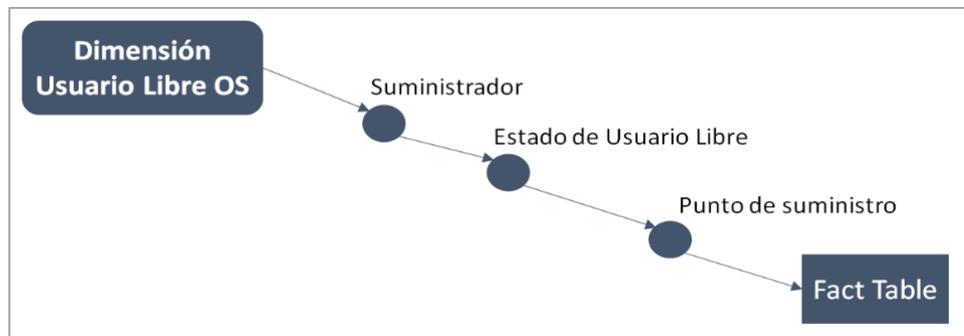


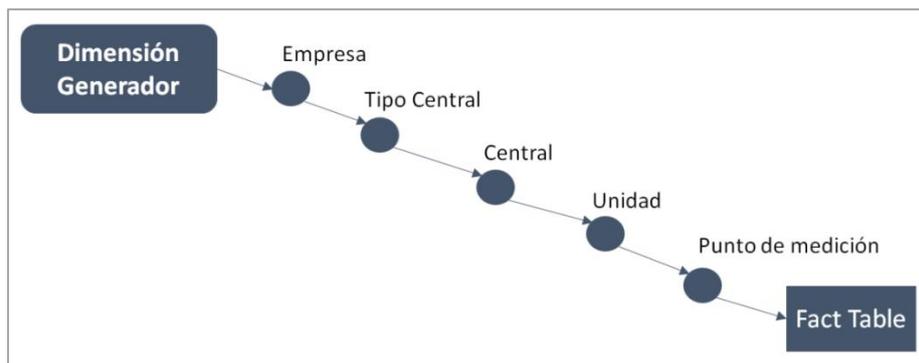
Figura 27: Nivel de granularidad de la dimensión Usuario Libre OS
 Fuente: Elaboración propia

- Dimensión Usuario Libre COES: 64



Figura 28: Nivel de granularidad de la dimensión Usuarios Libres COES
 Fuente: Elaboración propia

- Dimensión Generador:

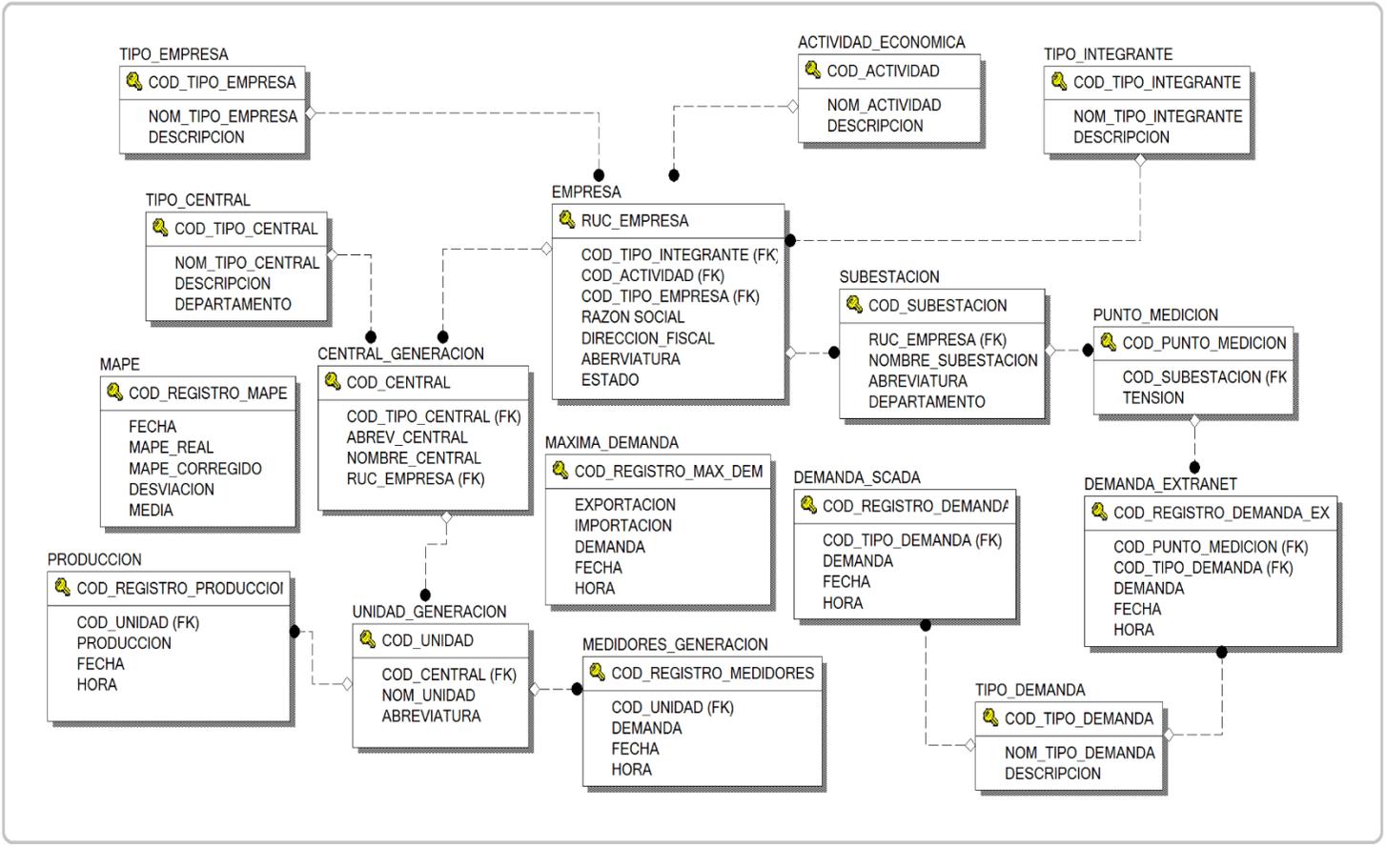


*Figura 29: Nivel de granularidad de la dimensión generador
Fuente: Elaboración propia*

- Modelo lógico: Previamente para definir las dimensiones y tablas de hechos, fue necesario realizar un modelo transaccional con la información recolectada para el datamart, el cual se presenta en la Figura 30; posterior a ello, en la Figura 31 se presenta el modelo de datos del datamart, en el que se puede observar los campos de cada tabla, así como las relaciones entre las tablas de hechos y las dimensiones.

Cabe señalar que se ha empleado el modelo estrella, debido al mejor nivel de respuesta que presenta en cuanto a la velocidad de consultas en comparación a los otros modelos (copo de nieve, constelación, etc.). Así mismo, cabe señalar que para la elaboración del datamart, se optó por utilizar llaves artificiales para las dimensiones, las cuales van a ingresar como llaves foráneas en las tablas de hechos, con ello se permitió mejorar el desempeño a nivel de consultas optimizando el uso del datamart.

Base de Datos COES (Replica)



Base de Datos Provisional

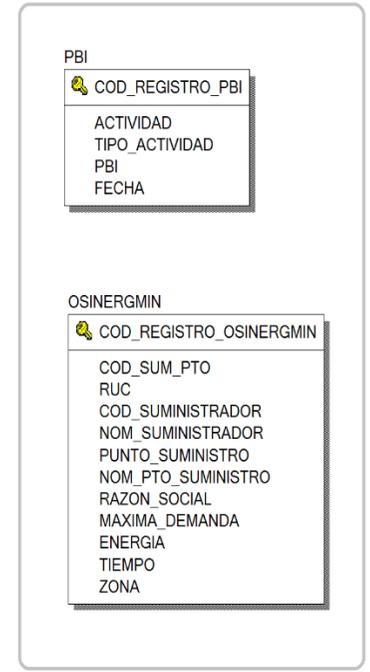


Figura 30: Modelo físico de la Base de Datos COES – Replica y Provisional
Fuente: Elaboración propia

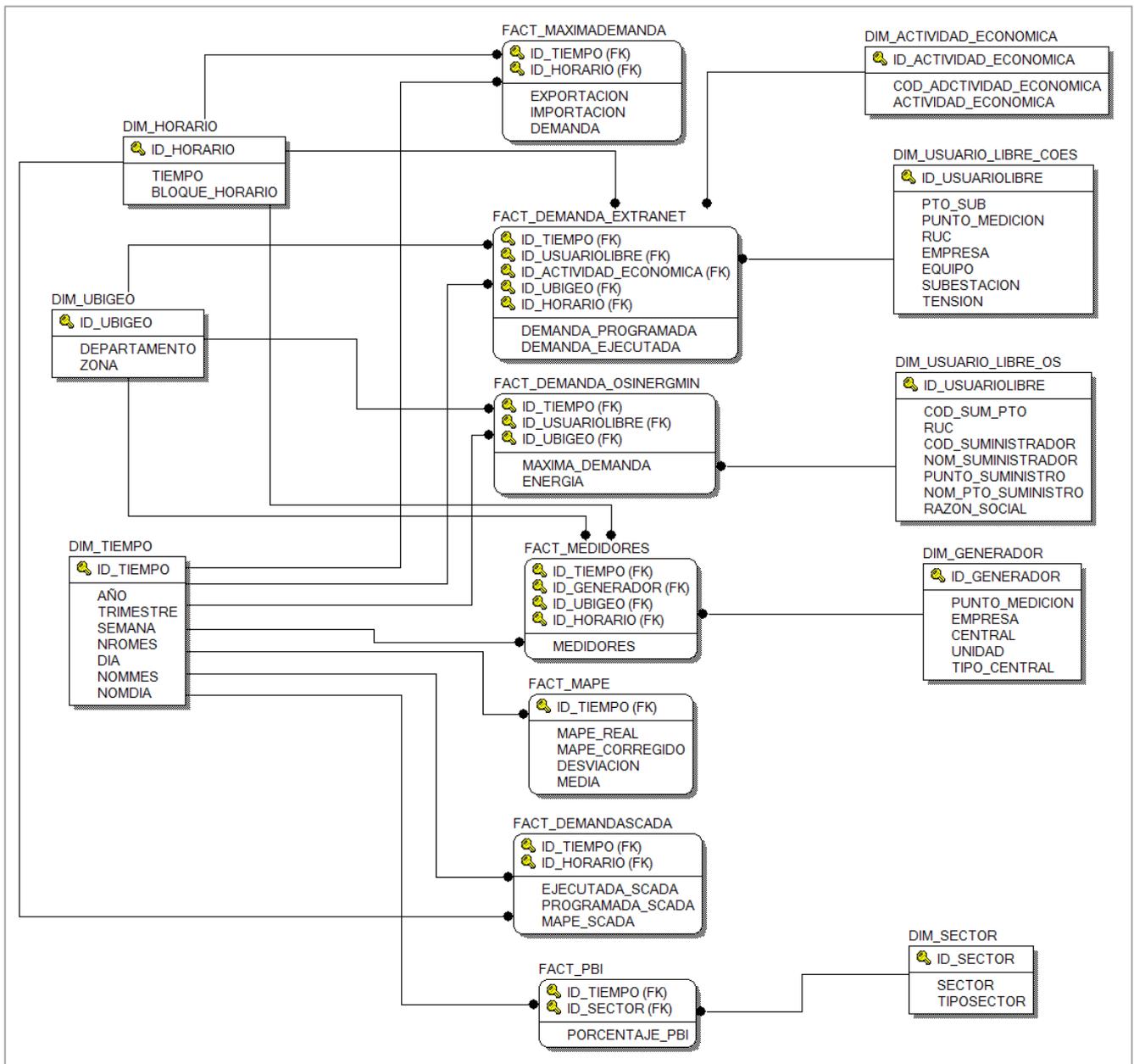


Figura 31: Modelo físico de Datamart
Fuente: Elaboración propia

5.6 Diseño e implementación del subsistema de ETL

- Extracción de información: Para el presente trabajo de investigación, al no tener acceso directo a la base de datos del COES, se obtuvo la información de los distintos sistemas transaccionales que generan reportes. En la Tabla 18, se muestra la relación entre la información necesaria para el desarrollo del datamart y los sistemas fuentes involucrados:

Información	Sistema	Descripción
Demanda ejecutada de Usuarios Libres.	Portal Web COES – Demanda Usuarios Libres	Se descargó la información del periodo establecido en el alcance. Esta información es declarada al COES por los Usuarios Libres vía Extranet en lapsos de 30 minutos para el cumplimiento del Procedimiento Técnico N° 3 del COES.
Demanda programada de Usuarios Libres.	Portal Web COES – Demanda Usuarios Libres	Cabe resaltar que dicho sistema solo permite la descarga de reportes mensuales.
Demanda ejecutada en el SEIN	Portal Web COES – Portal de Indicadores	Esta información es obtenida desde el sistema SCADA que maneja el COES, y presenta información en lapsos de 30 minutos. Dicha información se basa en el cumplimiento de la Norma Técnica para la coordinación de la operación en tiempo real de los sistemas interconectados.
Demanda programada en el SEIN	Portal Web COES – Portal de Indicadores	El sistema no permite realizar descargas masivas de grandes periodos de tiempo (meses).

<p>Producción de electricidad de medidores de generación.</p>	<p>Portal Web COES – Medidores de Generación</p>	<p>Esta información es declarada por los generadores integrantes del COES de manera mensual.</p> <p>El sistema permite realizar descargas en periodos trimestrales.</p>
<p>Máxima demanda en el SEIN.</p>	<p>Portal Web COES – Reporte de Máxima Demanda</p>	<p>La información sobre máxima demanda en el SEIN que el COES publica en su página web, es obtenida desde los archivos de medidores que remiten los agentes.</p> <p>El sistema permite realizar descargas en periodos mensuales.</p>
<p>Error porcentual absoluto medio para el pronóstico de la demanda (MAPE).</p>	<p>Portal Web COES – Portal de Indicadores</p>	<p>Mediante la información de demanda ejecutada y programada en el SEIN, se realizó el cálculo del MAPE para cada intervalo de tiempo.</p>
<p>Información de Usuarios Libres de OSINERGMIN (Puntos de suministros y máxima demanda).</p>	<p>Portal Web OSINERGMIN GART – Usuarios Libres</p>	<p>Se presenta los Usuarios Libres pertenecientes al SEIN, así como información de sus contratos con suministradores, máxima demanda mensual, potencia en hora punta y fuera de punta.</p> <p>La información se presenta en archivos de cálculo mensuales.</p>

<p>Producto bruto interno (PBI).</p>	<p>Portal Web Banco de Central de Reserva del Perú</p>	<p>Para temas de pronósticos a largo plazo, el PBI es un factor relevante para la demanda.</p> <p>El sistema permite descargar un Excel consolidando la información en periodos mensuales.</p>
--------------------------------------	--	--

Tabla 18: Relación entre la información y los sistemas transaccionales
Fuente: Elaboración propia

- Transformación y preparación para la carga en el repositorio dimensional: En la Figura 32, se muestra el proceso que se realizó para preparar la información con la estructura necesaria y en los formatos requeridos para realizar la carga de datos en el modelo dimensional:



Figura 32: Proceso de preparación de la información para la carga en el modelo dimensional
Fuente: Elaboración propia

En la Figura 32, podemos observar las 5 etapas de preparación de la información; primero tenemos los archivos fuentes en los formatos de descarga que manejan los distintos sistemas transaccionales. Como segundo paso, se utilizó hojas de cálculo Excel para adecuar dicha información en los formatos necesarios para la carga de datos al SQL, y con la finalidad de realizar un proceso limpio y sin errores, se volcaron los datos que se tenían en las hojas de cálculos hacia archivos de extensión “.txt” (flat file). Una vez con los archivos “.txt” generados, se procedió a realizar la carga en el motor de base de datos, donde todas las tablas autogeneradas a partir de la carga de los archivos “.txt” fueron consolidadas en otras tablas con la finalidad de manejar una única tabla con la información

integrada, a partir de estas tablas se realizó la carga de información a las 2 bases de datos creadas (COES y Provisional), las cuales fueron las fuentes de datos para realizar la carga en el datamart.

- Realizar la carga hacia el repositorio dimensional: Para realizar la carga de datos al datamart, se utilizó la herramienta de SQL Server Data Tools para generar un archivo de Integration Services. En la Figura 33 se muestra el paquete ETL que se elaboró para cargar la información en el datamart.

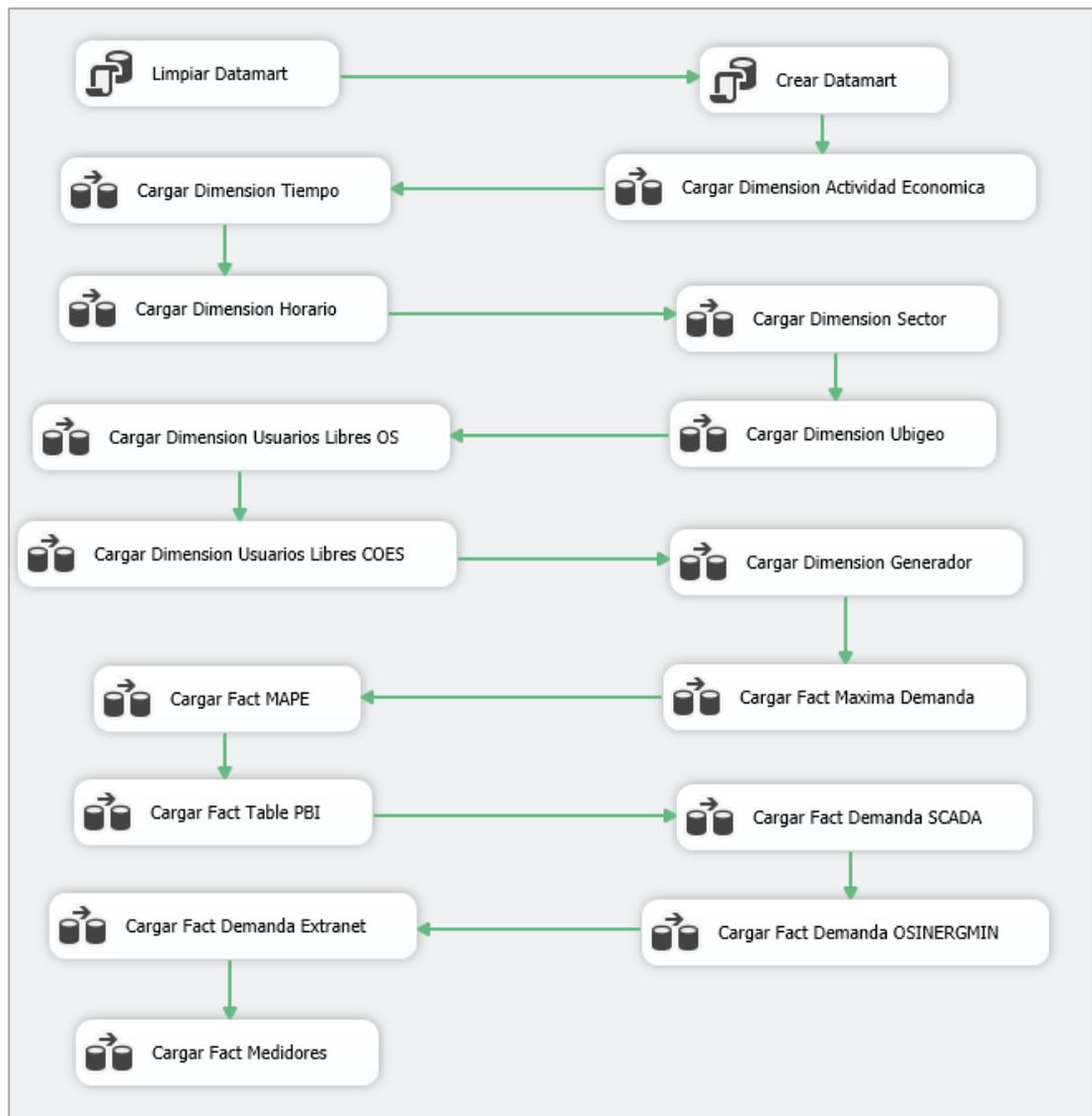


Figura 33: Proceso ETL con la herramienta SQL Server Data Tools - Inicio
 Fuente: Elaboración propia

El proceso de ETL que muestra en le Figura 33, inicia con la ejecución de un script que es “Limpiar Datamart”, debido a que algunos registros pueden sufrir actualizaciones, se realizará una carga completa, por ello para iniciar el proceso, se procede a eliminar por completo el datamart mediante el script SQL que se aprecia en le Figura 34. Una vez completo el proceso de “limpieza”, se procede a generar el datamart, para ello en “Crear

Datamart” se ejecuta, el script SQL, el cual crea las tablas de dimensiones y las tablas hechos, así como las relaciones que existen entre ellas. En la Figura 35 se aprecia parte del script SQL para “Crear Datamart”, donde se observa la creación de la Dim_Usuario_Libre_OS, la Fact_Demanda_Extranet, así como la creación de la relación de entre la Fact_Demanda_Extranet y la Dim_Actividad_Economica.

```

DROP TABLE [dbo].[FACT_DEMANDA_EXTRANET];
DROP TABLE [dbo].[FACT_DEMANDA_OSINERGMIN];
DROP TABLE [dbo].[FACT_DEMANDA_SCADA];
DROP TABLE [dbo].[FACT_MAPE];
DROP TABLE [dbo].[FACT_MAXIMA_DEMANDA];
DROP TABLE [dbo].[FACT_MEDIDORES];
DROP TABLE [dbo].[FACT_PBI];
DROP TABLE [dbo].[DIM_ACTIVIDAD_ECONOMICA];
DROP TABLE [dbo].[DIM_GENERADOR];
DROP TABLE [dbo].[DIM_HORARIO];
DROP TABLE [dbo].[DIM_SECTOR];
DROP TABLE [dbo].[DIM_TIEMPO];
DROP TABLE [dbo].[DIM_UBIGEO];
DROP TABLE [dbo].[DIM_USUARIO_LIBRE_COES];
DROP TABLE [dbo].[DIM_USUARIO_LIBRE_OS];
  
```

Figura 34: Script SQL para limpiar datamart en el proceso de ETL
Fuente: Elaboración propia

```

CREATE TABLE [dbo].[DIM_USUARIO_LIBRE_OS](
  [ID_USUARIOLIBRE] [int] IDENTITY(1,1) NOT NULL,
  [COD_SUM_PTO] [varchar](10) NULL,
  [RUC] [varchar](20) NULL,
  [COD_SUMINISTRADOR] [varchar](5) NULL,
  [NOM_SUMINISTRADOR] [varchar](80) NULL,
  [PUNTO_SUMINISTRO] [varchar](10) NULL,
  [NOM_PTO_SUMINISTRO] [varchar](200) NULL,
  [RAZON_SOCIAL] [varchar](150) NULL,
  PRIMARY KEY CLUSTERED
  (
    [ID_USUARIOLIBRE] ASC
  )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF, ALLOW_ROW_LOCKS = ON,
  ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
GO

CREATE TABLE [dbo].[FACT_DEMANDA_EXTRANET](
  [ID_USUARIO_LIBRE] [int] NULL,
  [ID_TIEMPO] [int] NULL,
  [ID_HORARIO] [int] NULL,
  [ID_UBIGEO] [int] NULL,
  [ID_ACTIVIDAD_ECONOMICA] [int] NULL,
  [DEMANDA_PROGRAMADA] [real] NULL,
  [DEMANDA_EJECUTADA] [real] NULL
) ON [PRIMARY]
GO

ALTER TABLE [dbo].[FACT_DEMANDA_EXTRANET] WITH CHECK ADD FOREIGN KEY([ID_ACTIVIDAD_ECONOMICA])
REFERENCES [dbo].[DIM_ACTIVIDAD_ECONOMICA] ([ID_ACTIVIDAD_ECONOMICA])
GO
  
```

Figura 35: Script SQL para crear datamart en el proceso de ETL (Fact_Demanda_Extranet y Dim_Usuario_Libre_OS)
Fuente: Elaboración propia

Una vez creada las tablas para el datamart, se procede a cargar las dimensiones; la primera dimensión en ser cargada es la de Actividad Económica, en la Figura 36 se aprecia la relación entre la fuente de datos y el destino en la herramienta SQL Server Data Tools, mientras que en la Figura 37, se observa la relación entre los campos de entrada y los campos de destino:

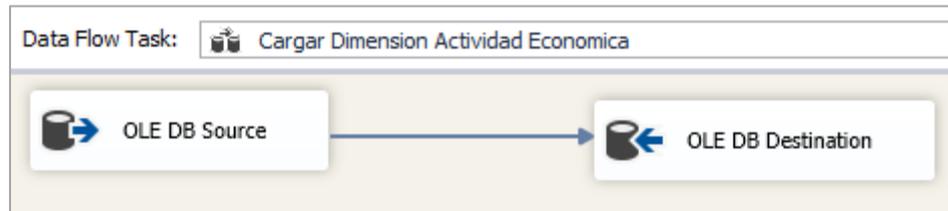


Figura 36: Data flow para cargar dimensión Actividad Económica
Fuente: Elaboración propia

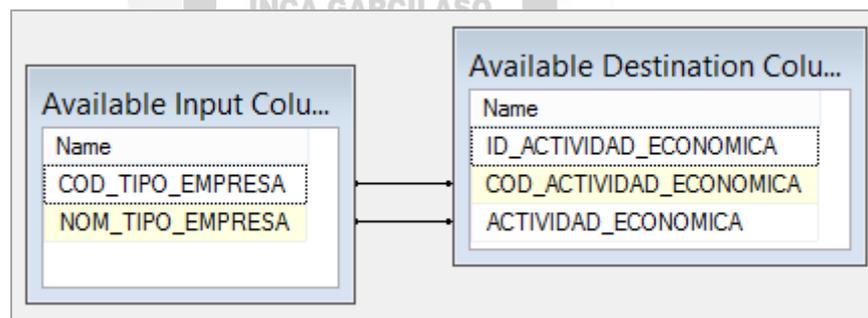


Figura 37: Mapeo de origen a destino en la dimensión actividad económica
Fuente: Elaboración propia

El siguiente paso del proceso, es la carga de la dimensión Tiempo, en la Figura 38 se aprecia la relación entre la fuente de datos y el destino en la herramienta SQL Server Data Tools, mientras que en la Figura 39, se observa la relación entre los campos de entrada y los campos de destino:

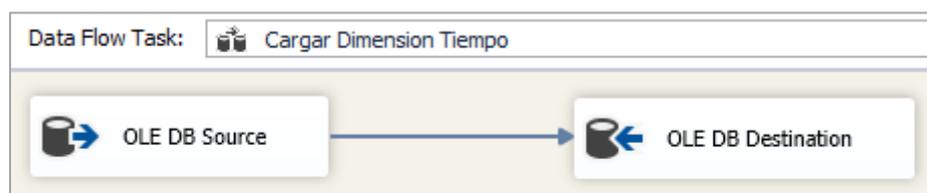


Figura 38: Data flow para cargar dimensión Tiempo
Fuente: Elaboración propia

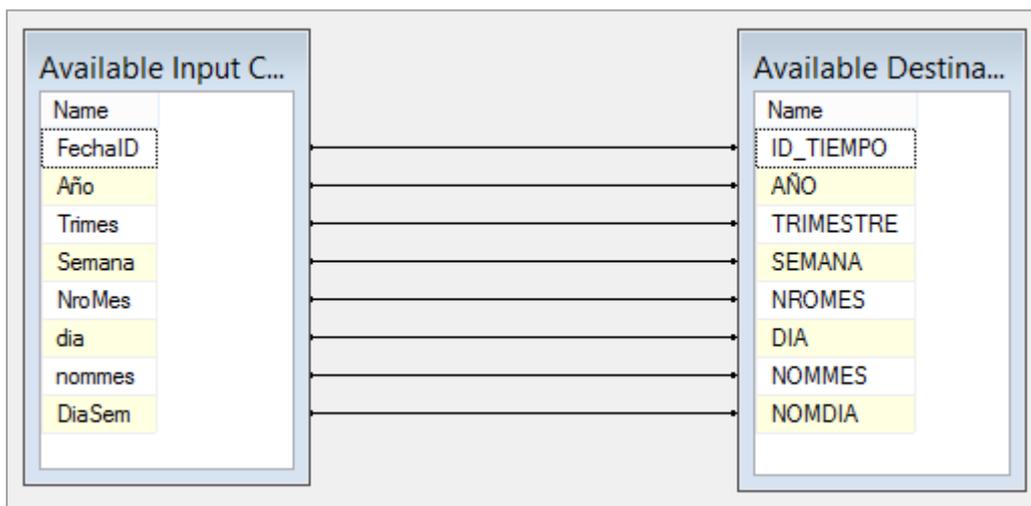


Figura 39: Mapeo de origen a destino en la dimensión Tiempo
Fuente: Elaboración propia

Una vez culminada la carga de la dimensión Tiempo, se procede con la carga de la dimensión Horario, en la Figura 40 se aprecia la relación entre la fuente de datos y el destino en la herramienta SQL Server Data Tools, mientras que en la Figura 41, se observa la relación entre los campos de entrada y los campos de destino:



Figura 40: Data flow para cargar dimensión Horario
Fuente: Elaboración propia

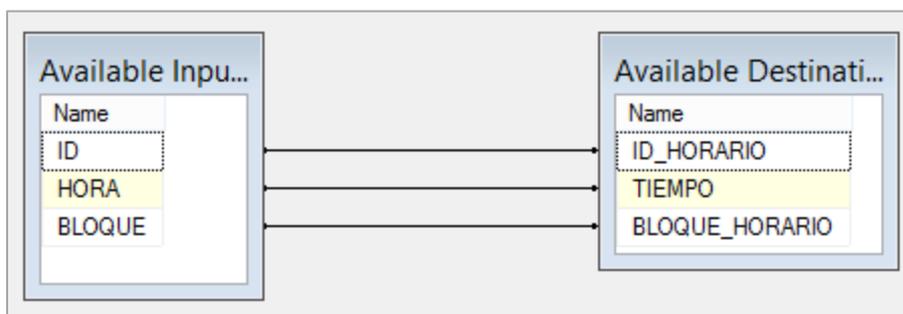


Figura 41: Mapeo de origen a destino en la dimensión Horario
Fuente: Elaboración propia

Siguiendo el flujo del ETL, se procede con la carga de la dimensión Sector, en la Figura 42 se aprecia la relación entre la fuente de datos y el destino en la herramienta SQL Server Data Tools, mientras que en la Figura 43, se observa la relación entre los campos de entrada y los campos de destino:



Figura 42: Data flow para cargar dimensión Sector
Fuente: Elaboración propia

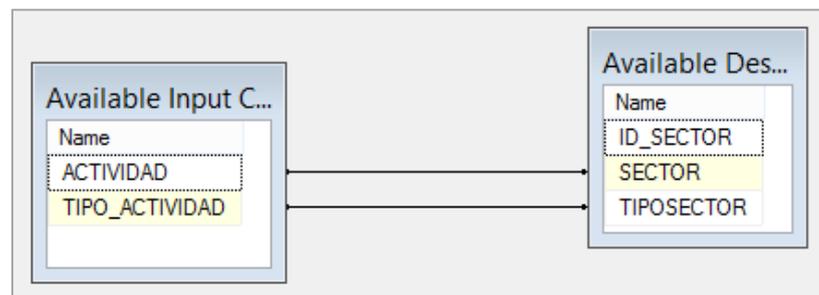


Figura 43: Mapeo de origen a destino en la dimensión Sector
Fuente: Elaboración propia

El siguiente proceso de carga pertenece a la dimensión Ubigeo, en la Figura 44 se aprecia la relación entre la fuente de datos y el destino en la herramienta SQL Server Data Tools, mientras que en la Figura 45, se observa la relación entre los campos de entrada y los campos de destino:



Figura 44: Data flow para cargar dimensión Ubigeo
Fuente: Elaboración propia

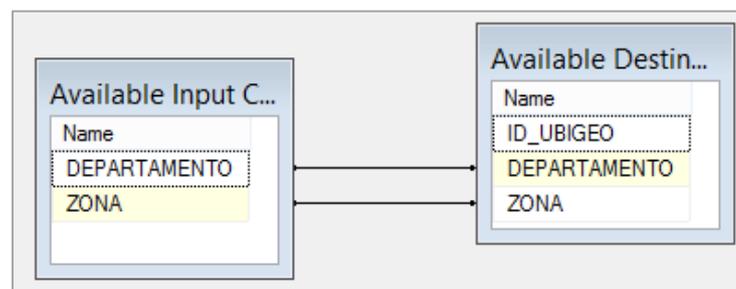


Figura 45: Mapeo de origen a destino en la dimensión Ubigeo
Fuente: Elaboración propia

El proceso que continúa es la carga de la dimensión Usuarios Libres OS, en la Figura 46 se aprecia la relación entre la fuente de datos y el destino en la herramienta SQL Server Data Tools, mientras que en la Figura 47, se observa la relación entre los campos de entrada y los campos de destino:

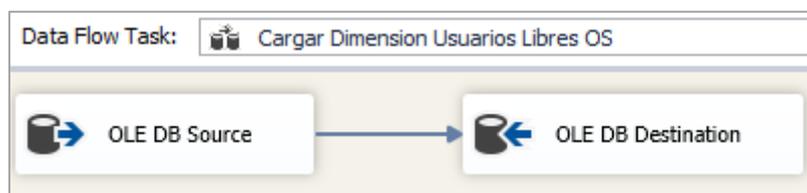


Figura 46: Data flow para cargar dimensión Usuarios Libres OS
Fuente: Elaboración propia

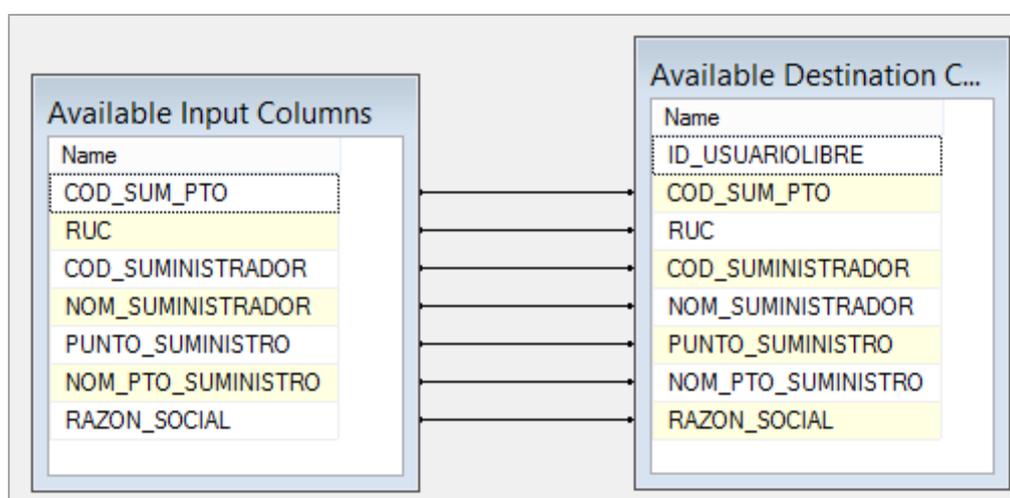


Figura 47: Mapeo de origen a destino en la dimensión Usuarios Libres OS
Fuente: Elaboración propia

El siguiente proceso de carga es la dimensión Usuarios Libres COES, en la Figura 48 se aprecia la relación entre la fuente de datos y el destino en la herramienta SQL Server Data Tools, así como un subproceso para convertir los datos, ello debido a que se necesitaba la misma longitud entre los campos de razón social, en la Figura 49 se aprecia la conversión; mientras que en la Figura 50, se observa la relación entre los campos de entrada y los campos de destino:

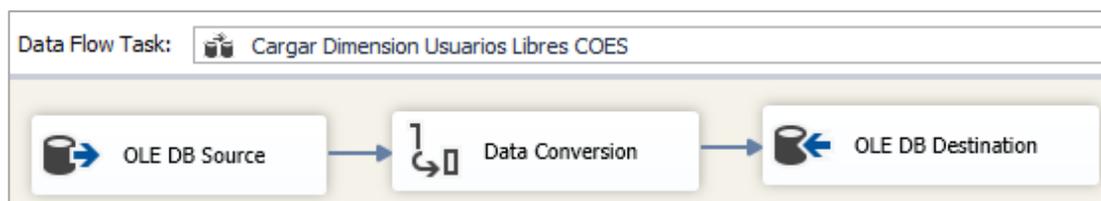


Figura 48: Data flow para cargar la dimensión Usuarios Libres COES
Fuente: Elaboración propia

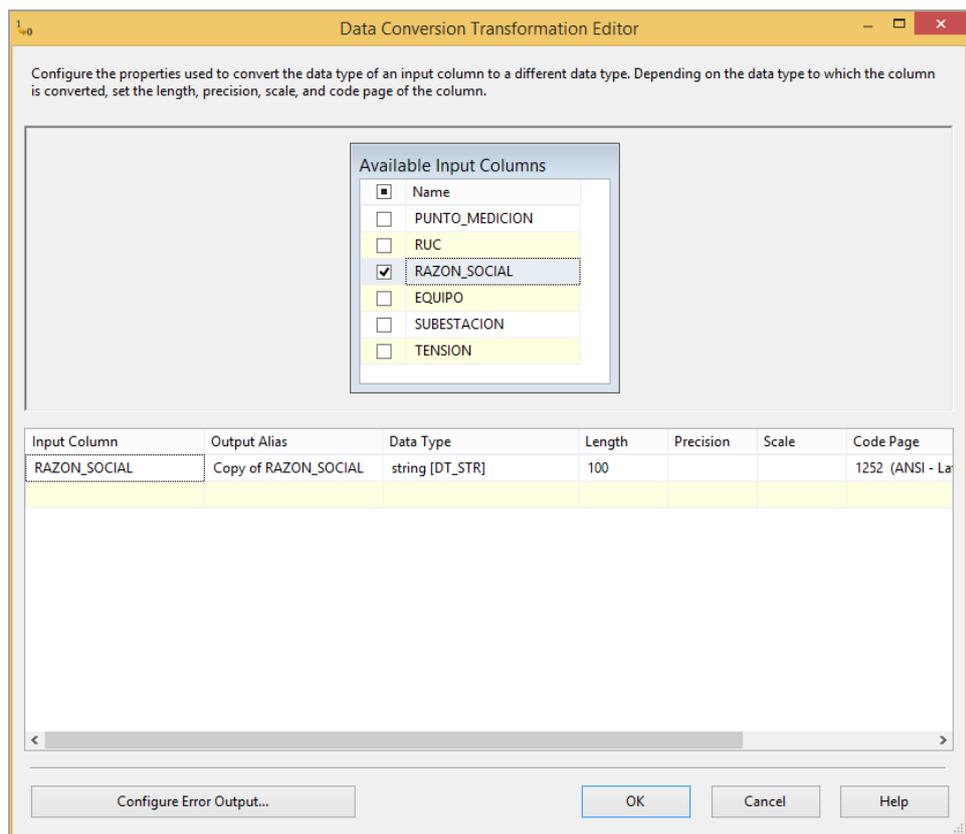


Figura 49: Sub proceso de conversión para el campo razón social en la dimensión Usuarios Libres COES
Fuente: Elaboración propia

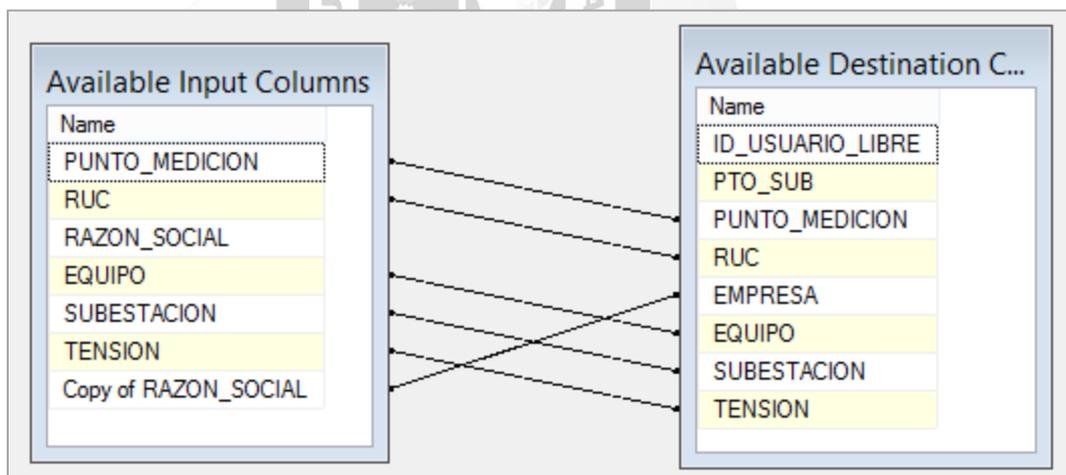


Figura 50: Mapeo de origen a destino en la dimensión Usuarios Libres COES
Fuente: Elaboración propia

El proceso que continúa es la carga de la dimensión Generador, en la Figura 51 se aprecia la relación entre la fuente de datos y el destino en la herramienta SQL Server Data Tools, mientras que en la Figura 52, se observa la relación entre los campos de entrada y los campos de destino:

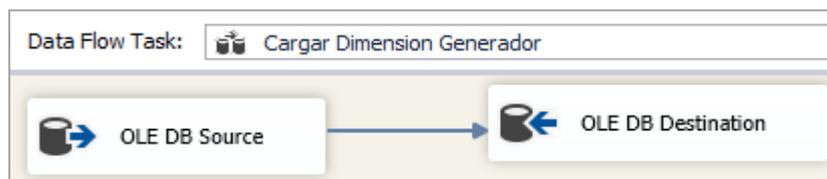


Figura 51: Data flow para cargar la dimensión Generator
Fuente: Elaboración propia

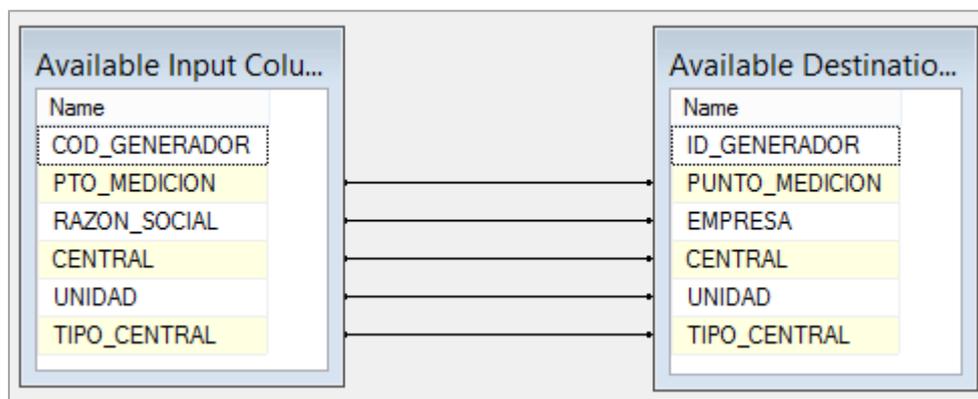


Figura 52: Mapeo de origen a destino en la dimensión Generator
Fuente: Elaboración propia

Una vez culminada la carga de las dimensiones, el proceso carga las fact tables, iniciando con la Fact Máxima Demanda, en la Figura 53 se aprecia la relación entre la fuente de datos y el destino en la herramienta SQL Server Data Tools, mientras que en la Figura 54, se observa la relación entre los campos de entrada y los campos de destino:



Figura 53: Data flow para cargar la tabla de hechos Máxima Demanda
Fuente: Elaboración propia

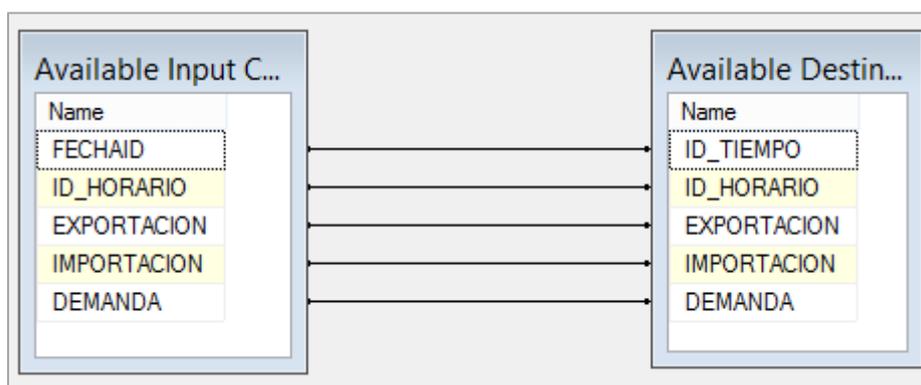


Figura 54: Mapeo de origen a destino en la tabla de hechos Máxima Demanda
Fuente: Elaboración propia

Continuando con el proceso de carga, se prosigue con la Fact MAPE, en la Figura 55 se aprecia la relación entre la fuente de datos y el destino en la herramienta SQL Server Data Tools, mientras que en la Figura 56, se observa la relación entre los campos de entrada y los campos de destino:



Figura 55: Data flow para cargar la tabla de hechos MAPE
Fuente: Elaboración propia

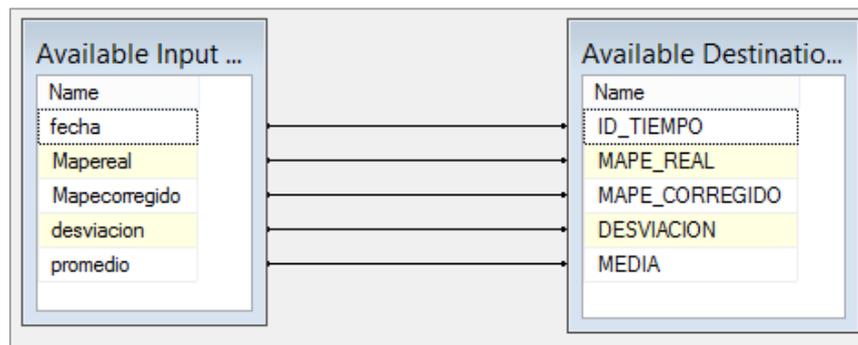


Figura 56: Mapeo de origen a destino en la tabla de hechos MAPE
Fuente: Elaboración propia

El proceso que continúa en la carga el de la Fact PBI, en la Figura 57 se aprecia la relación entre la fuente de datos y el destino en la herramienta SQL Server Data Tools, mientras que en la Figura 58, se observa la relación entre los campos de entrada y los campos de destino:

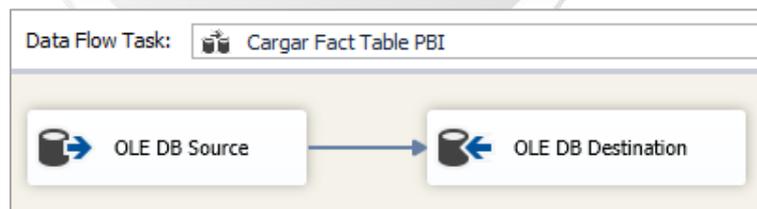


Figura 57: Data flow para cargar la tabla de hechos PBI
Fuente: Elaboración propia

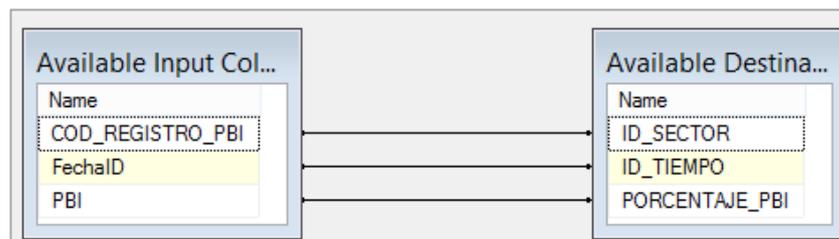


Figura 58: Mapeo de origen a destino en la tabla de hechos PBI
Fuente: Elaboración propia

Se prosigue la carga de la Fact Demanda SCADA, en la Figura 59 se aprecia la relación entre la fuente de datos y el destino en la herramienta SQL Server Data Tools, mientras que en la Figura 60, se observa la relación entre los campos de entrada y los campos de destino:

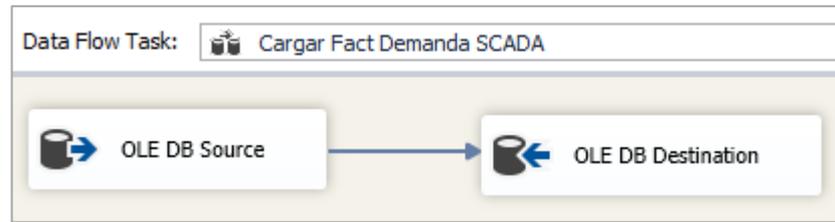


Figura 59: Data flow para cargar la tabla de hechos Demanda SCADA
Fuente: Elaboración propia

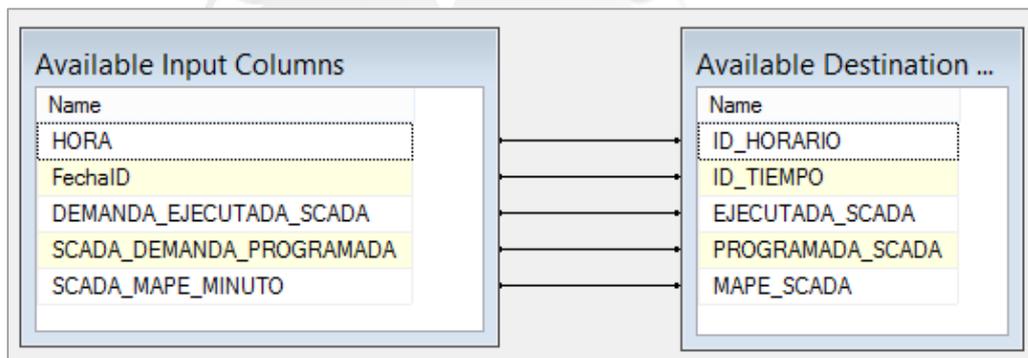


Figura 60: Mapeo de origen a destino en la tabla de hechos Demanda SCADA
Fuente: Elaboración propia

El proceso de ETL continua con la carga de la Fact Demanda OSINERGMIN, en la Figura 61 se aprecia la relación entre la fuente de datos y el destino en la herramienta SQL Server Data Tools, mientras que en la Figura 62, se observa la relación entre los campos de entrada y los campos de destino:



Figura 61: Data flow para cargar la tabla de hechos Demanda OSINERGMIN
Fuente: Elaboración propia

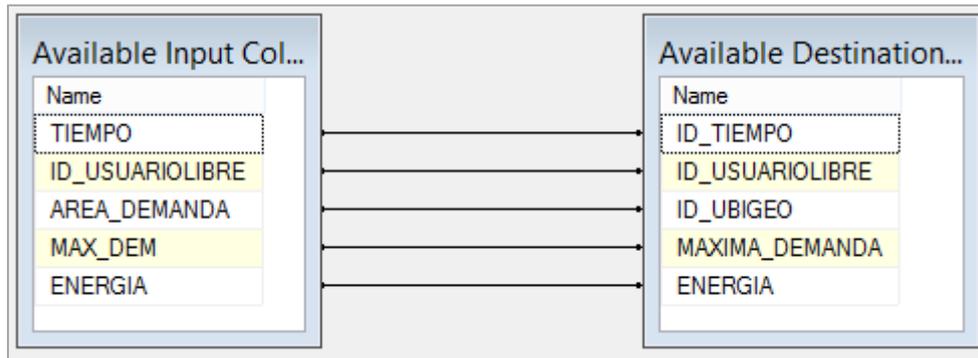


Figura 62: Mapeo de origen a destino en la tabla de hechos Demanda OSINERGMIN
Fuente: Elaboración propia

El proceso de ETL continua con la carga de la Fact Demanda Extranet en la Figura 63 se aprecia la relación entre la fuente de datos y el destino en la herramienta SQL Server Data Tools, mientras que en la Figura 64, se observa la relación entre los campos de entrada y los campos de destino:

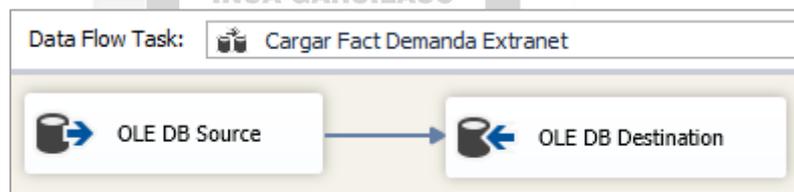


Figura 63: Data flow para cargar la tabla de hechos Demanda Extranet
Fuente: Elaboración propia

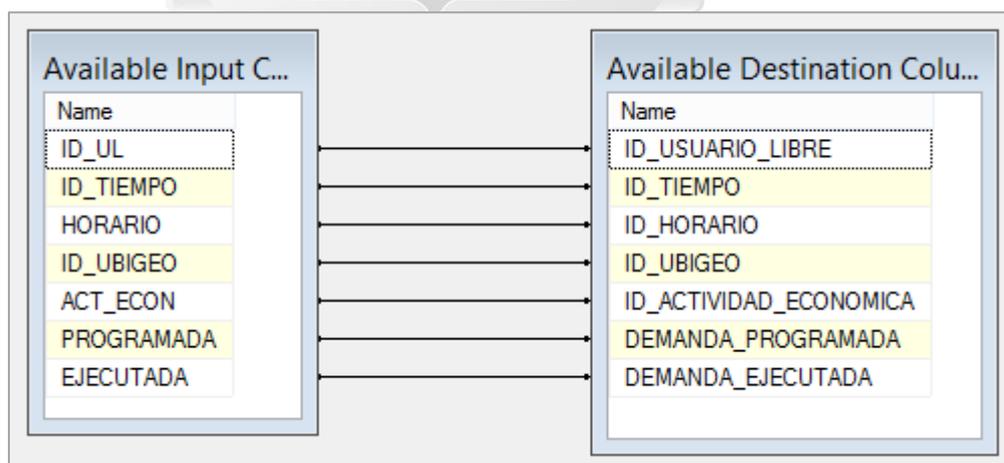


Figura 64: Mapeo de origen a destino en la tabla de hechos Demanda Extranet
Fuente: Elaboración propia

Finalmente, se realiza el último proceso de carga que es el perteneciente a la Fact Medidores, en la Figura 65 se aprecia la relación entre la fuente de datos y el destino en la herramienta SQL Server Data Tools, mientras que en la Figura 66, se observa la relación entre los campos de entrada y los campos de destino:



Figura 65: Data flow para cargar la tabla de hechos Medidores
Fuente: Elaboración propia

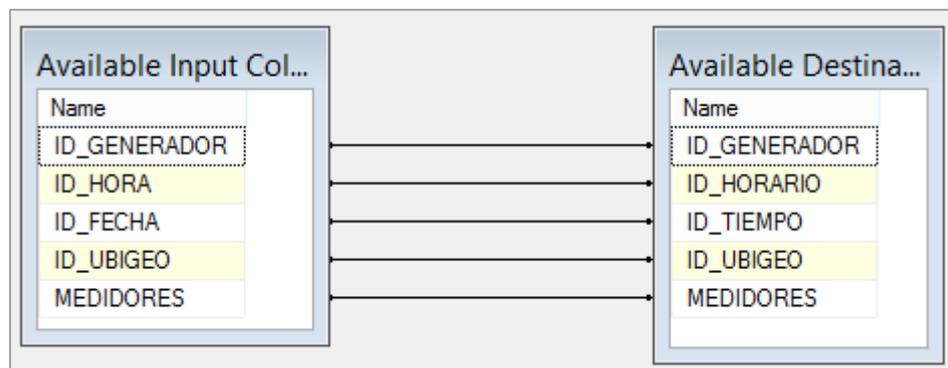


Figura 66: Mapeo de origen a destino en la tabla de hechos Medidores
Fuente: Elaboración propia

Al finalizar el proceso de ETL, se obtuvo como resultado, cargar el datamart con la información de las fuentes de datos; en la Figura 67, se observa como todos los paquetes del ETL han sido correctamente validados y cargados sin problemas:



Figura 67: Proceso ETL con la herramienta SQL Server Data Tools al finalizar
Fuente: Elaboración propia

Una vez culminado el proceso de ETL, se obtiene la creación del datamart, el cual cuenta con las siguientes dimensiones y tablas hechos:

- Dimensión Actividad Económica: Comprende las actividades económicas de los Usuarios Libres registrados. En la Figura 68 se observa el diseño de la tabla Dim_Actividad_Economica.

	Column Name	Data Type
🔑	ID_ACTIVIDAD_ECONOMICA	int
	COD_ACTIVIDAD_ECONOMICA	varchar(5)
	ACTIVIDAD_ECONOMICA	varchar(200)

Figura 68: Tabla Dim_Actividad_Economica del Datamart
Fuente: Elaboración propia

- Dimensión Tiempo: Comprende las fechas registradas entre el periodo establecido en el alcance para el desarrollo del datamart. En la Figura 69 se observa el diseño de la tabla Dim_Tiempo:

	Column Name	Data Type
🔑	ID_TIEMPO	int
	AÑO	smallint
	TRIMESTRE	tinyint
	SEMANA	smallint
	NROMES	smallint
	DIA	smallint
	NOMMES	varchar(15)
	NOMDIA	varchar(15)

Figura 69: Tabla Dim_Tiempo del Datamart
Fuente: Elaboración propia

- Dimensión Horario: Engloba los intervalos de tiempo en los que se registran datos dentro las 24 horas de un día. En algunos casos los registros de las fact tables son en periodos 15 minutos mientras que en otros casos son cada 30 minutos. En la Figura 70 se observa el diseño de la tabla Dim_Horario:

	Column Name	Data Type
🔑	ID_HORARIO	int
	TIEMPO	varchar(5)
	BLOQUE_HORARIO	varchar(15)

Figura 70: Tabla Dim_Horario del Datamart
Fuente: Elaboración propia

- Dimensión Sector: Comprende los sectores y los tipos de sectores en los que se observan los registros del porcentaje del PBI. En la Figura 71 se observa el diseño de la tabla Dim_Sector:

	Column Name	Data Type
🔑	ID_SECTOR	int
	SECTOR	varchar(50)
	TIPOSECTOR	varchar(50)

Figura 71: Tabla Dim_Sector del Datamart
Fuente: Elaboración propia

- Dimensión Ubigeo: Engloba los departamentos y las zonas geográficas de los distintos puntos de medición, ya sea de los Usuarios Libres o de los Generadores. En la Figura 72 se observa el diseño de la tabla Dim_Ubigeo:

	Column Name	Data Type
🔑	ID_UBIGEO	int
	DEPARTAMENTO	varchar(50)
	ZONA	varchar(50)

Figura 72: Tabla Dim_Ubigeo del Datamart
Fuente: Elaboración propia

- Dimensión Usuario Libre COES: Comprende todos los Usuarios Libres que envían información al COES, con sus respectivas subestaciones y características de sus puntos de suministro. En la Figura 73 se observa el diseño de la tabla Dim_Usuario_Libre_COES:

	Column Name	Data Type
🔑	ID_USUARIO_LIBRE	int
	PTO_SUB	varchar(150)
	PUNTO_MEDICION	varchar(50)
	RUC	varchar(20)
	EMPRESA	varchar(100)
	EQUIPO	varchar(50)
	SUBESTACION	varchar(50)
	TENSION	varchar(20)

Figura 73: Tabla Dim_Usuario_Libre_COES del Datamart
Fuente: Elaboración propia

- Dimensión Usuario Libre OS: Engloba los Usuarios Libres cuya información es publicada por OSINERGMIN. En la Figura 74 se observa el diseño de la tabla Dim_Usuario_Libre_OS:

	Column Name	Data Type
🔑	ID_USUARIOLIBRE	int
	COD_SUM_PTO	varchar(10)
	RUC	varchar(20)
	COD_SUMINISTRADOR	varchar(5)
	NOM_SUMINISTRADOR	varchar(80)
	PUNTO_SUMINISTRO	varchar(10)
	NOM_PTO_SUMINISTRO	varchar(200)
	RAZON_SOCIAL	varchar(150)

Figura 74: Tabla Dim_Usuario_Libre_OS del Datamart
Fuente: Elaboración propia

- Dimensión Generador: Comprende las empresas generadoras que remiten información al COES en base a los medidores de las distintas unidades de

generación por cada central que les pertenece. Cabe señalar que cada central también es categorizada por el tipo de central. En la Figura 75 se observa el diseño de la tabla Dim_Generador:

	Column Name	Data Type
🔑	ID_GENERADOR	int
	PUNTO_MEDICION	varchar(10)
	EMPRESA	varchar(100)
	CENTRAL	varchar(100)
	UNIDAD	varchar(50)
	TIPO_CENTRAL	varchar(25)

Figura 75: Tabla Dim_Generador del Datamart
Fuente: Elaboración propia

- Fact Demanda Extranet: Comprende los registros de demanda declarados al COES vía Extranet por los Usuarios Libres categorizado por las distintas dimensiones que se relacionan a ésta. En la Figura 76 se observa el diseño de la tabla Fact_Demanda_Extranet:

Column Name	Data Type
ID_USUARIO_LIBRE	int
ID_TIEMPO	int
ID_HORARIO	int
ID_UBIGEO	int
ID_ACTIVIDAD_ECONOMICA	int
DEMANDA_PROGRAMADA	real
DEMANDA_EJECUTADA	real

Figura 76: Tabla Fact_Demanda_Extranet del Datamart
Fuente: Elaboración propia

- Fact Demanda OSINERGMIN: Engloba los registros de máxima demanda y energía que OSINERGMIN dispone en sus hojas de cálculo de los distintos Usuarios Libres. En la Figura 77 se observa el diseño de la tabla Fact_Demanda_OSINERGMIN:

Column Name	Data Type
ID_TIEMPO	int
ID_USUARIOLIBRE	int
ID_UBIGEO	int
MAXIMA_DEMANDA	real
ENERGIA	real

Figura 77: Tabla Fact_Demanda_OSINERGMIN del Datamart
Fuente: Elaboración propia

- Fact Demanda SCADA: Comprende los registros de demanda instantánea que se obtienen vía SCADA en relación al tiempo y horario. En la Figura 78 se observa el diseño de la tabla Fact_Demanda_SCADA:

Column Name	Data Type
ID_HORARIO	int
ID_TIEMPO	int
EJECUTADA_SCADA	real
PROGRAMADA_SCADA	real
MAPE_SCADA	real

Figura 78: Tabla Fact_Demanda_SCADA del Datamart
Fuente: Elaboración propia

- Fact MAPE: Engloba los registros obtenidos del error porcentual absoluto medio que se ha obtenido mediante de la evaluación la calidad en el pronóstico de la demanda de manera diaria. En la Figura 79 se observa el diseño de la tabla Fact_MAPE:

Column Name	Data Type
ID_TIEMPO	int
MAPE_REAL	real
MAPE_CORREGIDO	real
DESVIACION	real
MEDIA	real

Figura 79: Tabla Fact_MAPE del Datamart
Fuente: Elaboración propia

- Fact Máxima Demanda: Comprende los registros obtenidos de los días de la máxima demanda mensual, así como la importación o exportación que se dio en ese día. En la Figura 80 se observa el diseño de la tabla Fact_Maxima_Demanda:

Column Name	Data Type
ID_TIEMPO	int
ID_HORARIO	int
EXPORTACION	real
IMPORTACION	real
DEMANDA	real

Figura 80: Tabla Fact_Maxima_Demanda del Datamart
Fuente: Elaboración propia

- Fact Medidores: Engloba todos los registros de los medidores que los generadores remiten al COES, estos registros se encuentran categorizados bajo las distintas dimensiones con las que se relaciona la tabla hechos. En la Figura 81 se observa el diseño de la tabla Fact_Medidores:

Column Name	Data Type
ID_GENERADOR	int
ID_HORARIO	int
ID_TIEMPO	int
ID_UBIGEO	int
MEDIDORES	real

Figura 81: Tabla Fact_Medidores del Datamart
Fuente: Elaboración propia

Para poder desarrollar las aplicaciones de BI, se empleó el repositorio dimensional (datamart), para generar los cubos OLAP, los cuales están conformados por las fact table y sus dimensiones asociadas. Mediante la herramienta SQL Server Data Tools en el Analysis Services se definió la jerarquía que iba a existir dentro de cada dimensión, como el caso de la dimensión Tiempo que se aprecia en la Figura 82, dimensión Horario en la Figura 83, la dimensión Sector en la Figura 84, la dimensión Ubigeo en la Figura 85, la dimensión Actividad Económica en la Figura 86, la dimensión Usuarios Libres OS en la Figura 87, la dimensión Usuarios Libres COES en la Figura 88 y la dimensión Generador en la Figura 89:



Figura 82: Jerarquía de la dimensión Tiempo
Fuente: Elaboración propia

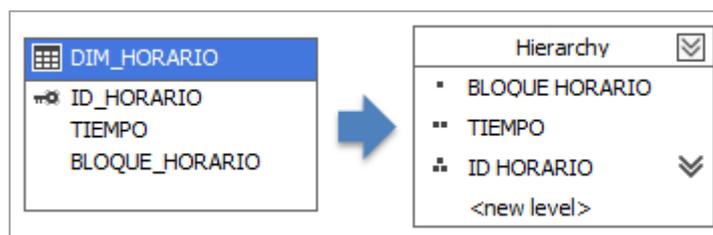


Figura 83: Jerarquía de la dimensión Horario
Fuente: Elaboración propia

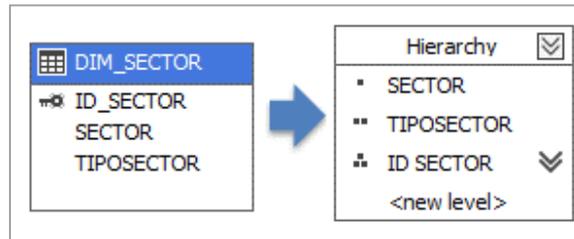


Figura 84: Jerarquía de la dimensión Sector
Fuente: Elaboración propia

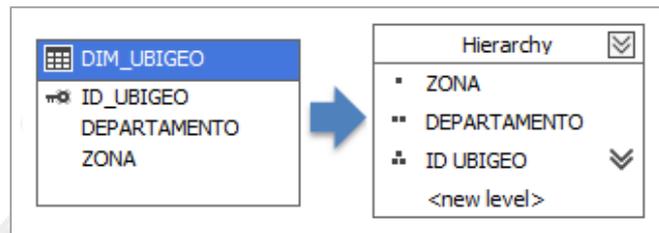


Figura 85: Jerarquía de la dimensión Ubigeo
Fuente: Elaboración propia



Figura 86: Jerarquía de la dimensión Actividad Económica
Fuente: Elaboración propia

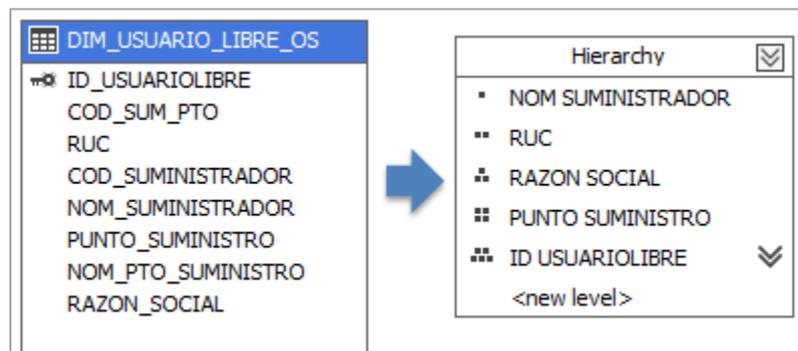


Figura 87: Jerarquía de la dimensión Usuario Libre OS
Fuente: Elaboración propia

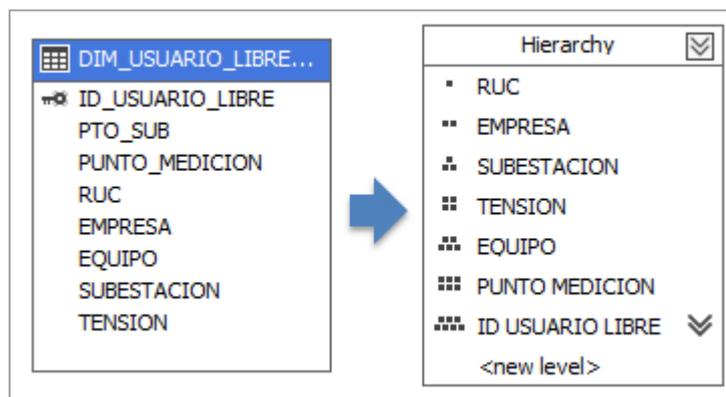


Figura 88: Jerarquía de la dimensión Usuario Libre COES
Fuente: Elaboración propia

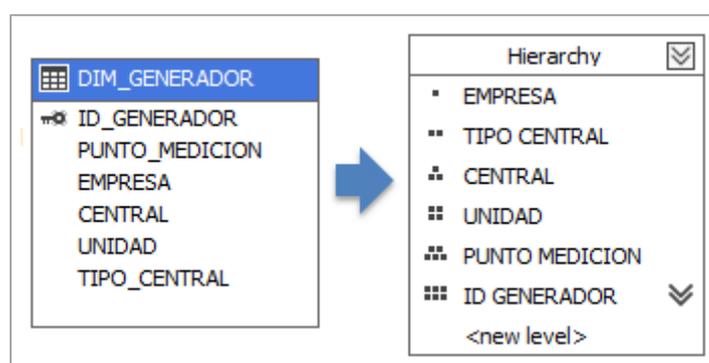


Figura 89: Jerarquía de la dimensión Generador
Fuente: Elaboración propia

Finalmente se obtuvieron 7 cubos que serán utilizados como fuente de información, estos son:

- Cubo_PBI: En la Figura 90, se muestra el cubo relacionado al Peso Bruto Interno el cual contiene los registros mensuales del PBI, los cuales también pueden ser consultados según el sector (general) y el tipo de sector, también se aprecia que el cubo trabaja con las dimensiones Tiempo y Sector; éste se ha generado mediante la herramienta de SQL Server Data Tools en el Analysis Services.

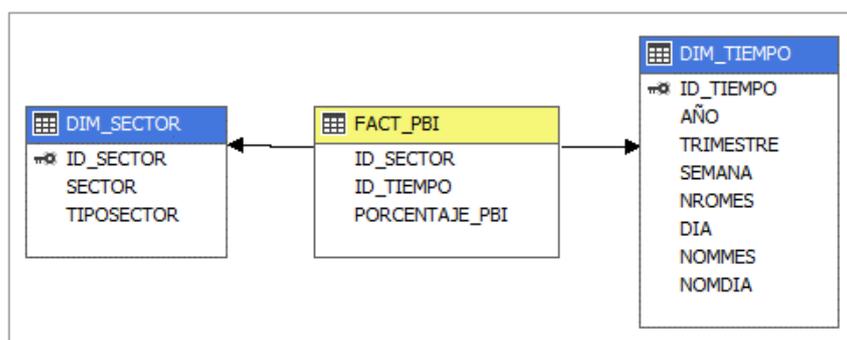


Figura 90: Cubo PBI en la herramienta SQL Server Data Tools
Fuente: Elaboración propia

- Cubo_Demanda_SCADA: En la Figura 91, se observa el cubo de la demanda SCADA, que comprende los registros diarios y en intervalos de tiempo de la demanda instantánea ejecutada y programada en el SEIN, se observa que las dimensiones que la componen son Tiempo y Horario; éste se ha generado mediante la herramienta de SQL Server Data Tools en el Analysis Services.

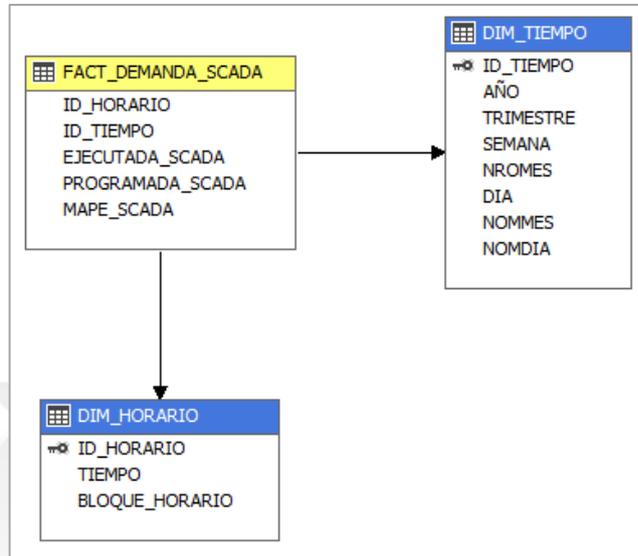


Figura 91: Cubo Demanda SCADA en la herramienta SQL Server Data Tools
Fuente: Elaboración propia

- Cubo_MAPE: En la Figura 92 tenemos el cubo del MAPE que comprende los registros porcentuales al evaluar la demanda SCADA ejecutada y programada diariamente, donde se observa que se relaciona con la dimensión Tiempo; éste se ha generado mediante la herramienta de SQL Server Data Tools en el Analysis Services.

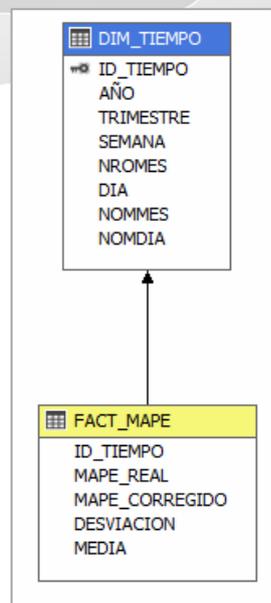


Figura 92: Cubo MAPE en la herramienta SQL Server Data Tools
Fuente: Elaboración propia

- **Cubo_Demanda_OSINERGMIN:** En la Figura 93 tenemos el cubo en relación con la demanda de OSNERGMIN que muestra los registros de máxima demanda y energía de los Usuarios Libres de manera mensual y a su vez se define por zonas geográfica, se observa que las dimensiones que la componen son Tiempo, Usuario Libre Osinergmin y Ubigeo; éste se ha generado mediante la herramienta de SQL Server Data Tools en el Analysis Services.

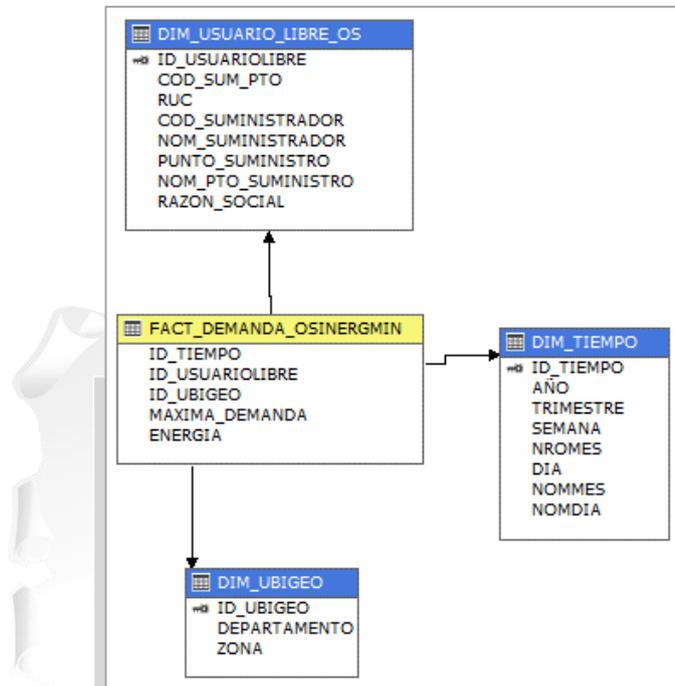


Figura 93: Cubo Demanda OSINERGMIN en la herramienta SQL Server Data Tools
Fuente: Elaboración propia

- **Cubo_Maxima_Demanda:** En la Figura 94 se presenta el cubo relacionado a la máxima demanda que contiene los registros de demanda, exportación e importación que se ejecutaron en el día y hora de la máxima demanda mensual, se observa que las dimensiones que la componen son Tiempo y Horario; éste se ha generado mediante la herramienta de SQL Server Data Tools en el Analysis Services.

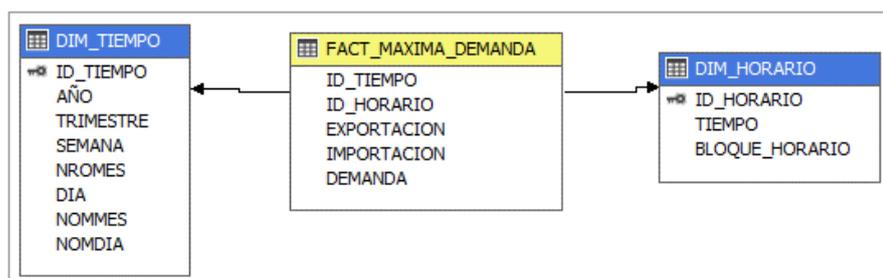


Figura 94: Cubo Máxima Demanda en la herramienta SQL Server Data Tools
Fuente: Elaboración propia

- Cubo_Demanda_Extranet: En la Figura 95 tenemos el cubo en relación con la demanda declarada vía Extranet al COES por los Usuarios Libres de manera diaria y en los intervalos de tiempo determinados, así mismo se puede apreciar dichos registros por la actividad económica del Usuario Libre y también por la ubicación de cada punto de medición; se observa que las dimensiones que la componen son Tiempo, Horario, Actividad Económica, Usuario Libre COES y Ubigeo; éste se ha generado mediante la herramienta de SQL Server Data Tools en el Analysis Services.

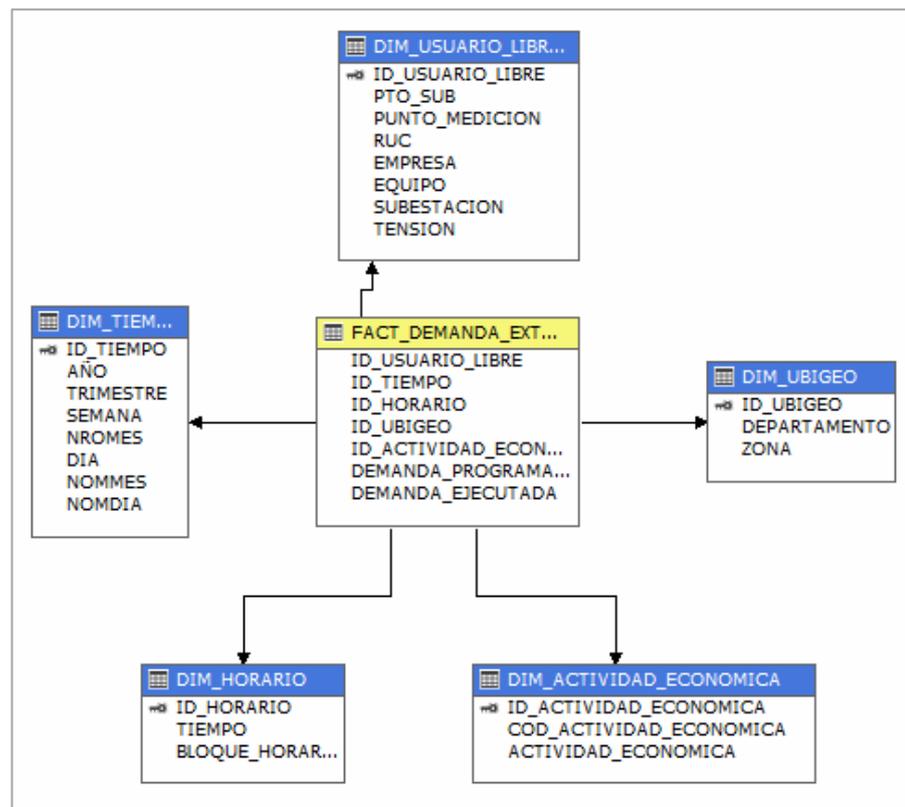


Figura 95: Cubo Demanda Extranet en la herramienta SQL Server Data Tools
Fuente: Elaboración propia

- Cubo_Medidores: En la Figura 96 se presenta el cubo relacionado a los medidores de generación que comprende los registros remitidos por los Generadores en periodos diarios y en los intervalos de tiempo establecidos, dichos registros son generados a partir de sus unidades de generación las cuales se ubican en las centrales pertenecientes a la empresas generadora, las cuales tienen una ubicación geográfica y a su vez cada central puede variar en su tipo (eólica, solar, termoeléctrica o hidroeléctrica) ; se observa que las dimensiones que la componen son Tiempo, Horario, Generador y Ubigeo; éste se ha generado mediante la herramienta de SQL Server Data Tools en el Analysis Services.

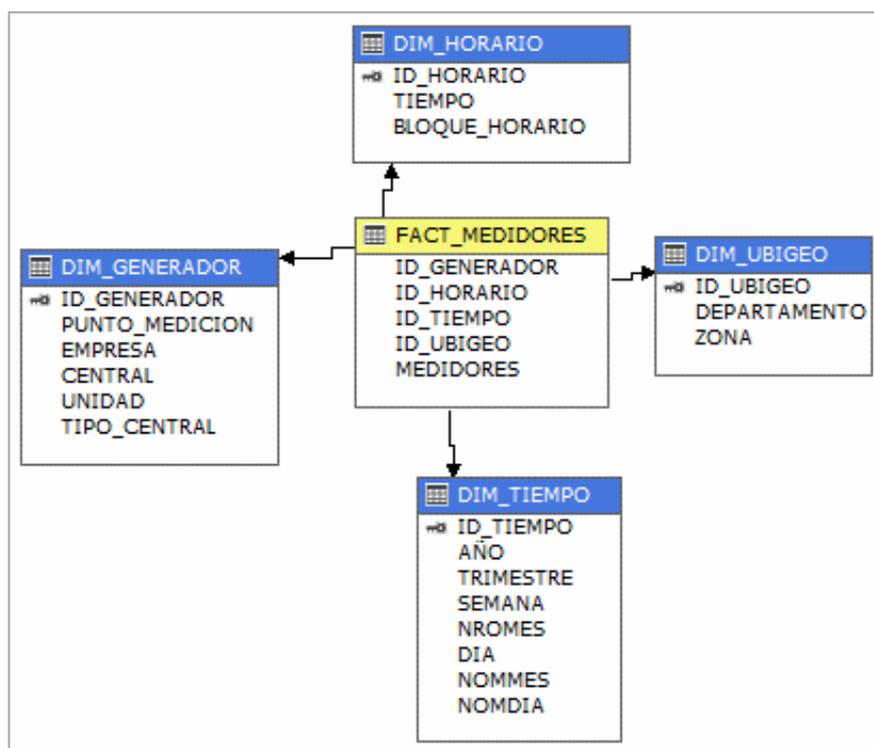


Figura 96: Cubo Medidores en la herramienta SQL Server Data Tools
Fuente: Elaboración propia

5.7 Especificación de aplicaciones de BI

- Informes estándar: Para mostrar la información que se maneja en el datamart, se definió que se iba a mostrar en dichos gráficos, a fin de entregar información que sea útil para los objetivos del presente trabajo de investigación.
- Herramientas para acceso de lectura al datamart: Para la lectura del datamart, se presentan 2 opciones de acuerdo con el tipo de rol del usuario que visualizará la información y mediante que herramienta, lo cual podemos observar en la Tabla 19:

Herramienta	Rol	Descripción
Informes Personalizados en Web	Todos (Usuarios Libres, Generadores, Distribuidores, Transmisores, inversionistas, consultores, estudiantes, etc.)	Para los informes gráficos que se publiquen en la web, el acceso es público. Como objetivo de información se tiene a los Usuarios Libres y también a otras empresas que desean invertir en el sector eléctrico nacional.

<p>Conexión a Datamart con Power Pivot</p>	<p>Especialistas del COES con acceso autorizado</p>	<p>Los especialistas que requieran de tener acceso a la información del datamart podrán solicitar un usuario y contraseña para obtener la información en Excel mediante el complemento Power Pivot.</p>
--	---	---

Tabla 19: Herramientas para el acceso a la información del datamart
Fuente: Elaboración propia

5.8 Desarrollo de aplicaciones de BI

Con la finalidad de que el usuario cuente con acceso a la información del datamart, se generaron 2 alternativas de ingreso y consulta de la data, éstas son mediante 2 herramientas tecnológicas: Power Pivot, que permitirá al usuario acceder a la información desde un archivo Excel y poder trabajarla mediante tablas dinámicas con las cuales puede generar gráficos en las hojas de cálculo; y Power BI, herramienta que permite la explotación de información mediante gráficos dinámicos, los cuales se realizaron para que el usuario pueda visualizar la información de manera amigable. A continuación, se presentan el desarrollo de ambas aplicaciones de BI definidas:

5.8.1 Desarrollo mediante la herramienta Power Pivot

Para poder entregar al usuario la información a la cual pueda conectarse, fue necesario generar el modelo dimensional (datamart) y también los cubos OLAP. Ello debido a que en el caso un usuario desee obtener la estructura del datamart, va a poder observarla al obtener los datos mediante Power Pivot, en el caso de los cubos OLAP, el usuario utilizará Power Pivot para generar las consultas MDX mediante las cuales podrá acceder a la información que requiere analizar.

5.8.2 Desarrollo mediante la herramienta Power BI

Con los datos almacenados en un modelo dimensional y en cubos OLAP, se genera la conexión desde el Power BI Desktop hacia una de las fuentes de datos, para poder desarrollar los reportes gráficos. En la Figura 97, 98 y 99 se observa el proceso de conexión que se realizó desde Power BI al modelo de datos dimensional (datamart), y finalmente en la Figura 100 se observan las tablas cargadas en el Power BI Desktop.

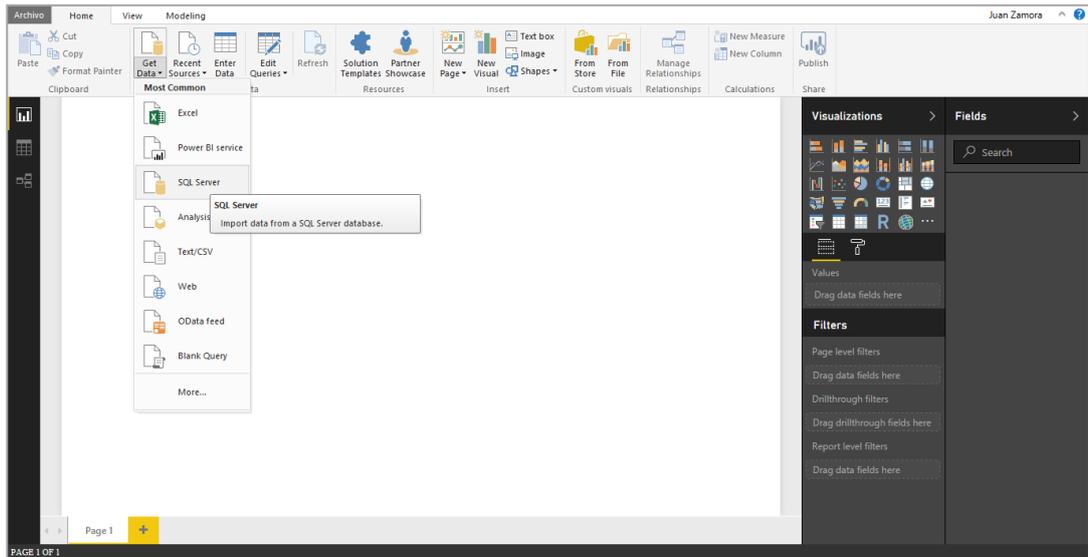


Figura 97: Selección de fuente de datos en Power BI Desktop
Fuente: Elaboración propia

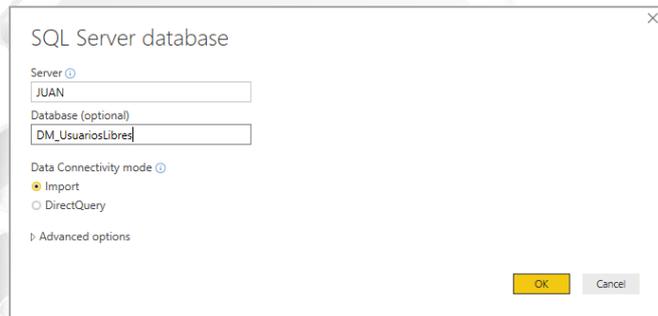


Figura 98: Conexión al datamart desde Power BI Desktop
Fuente: Elaboración propia

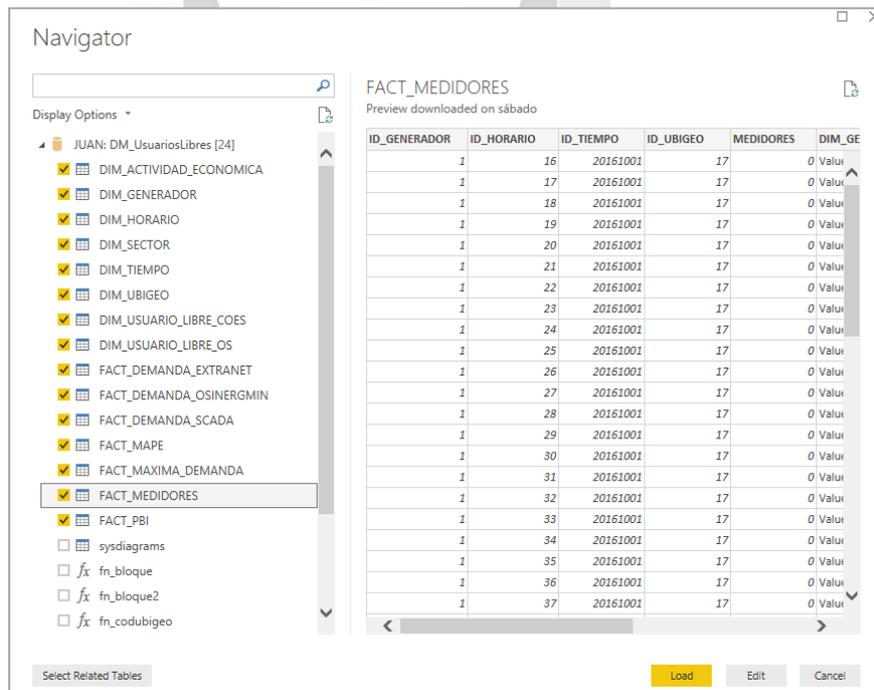


Figura 99: Selección de tablas para obtener datos en Power BI Desktop
Fuente: Elaboración propia

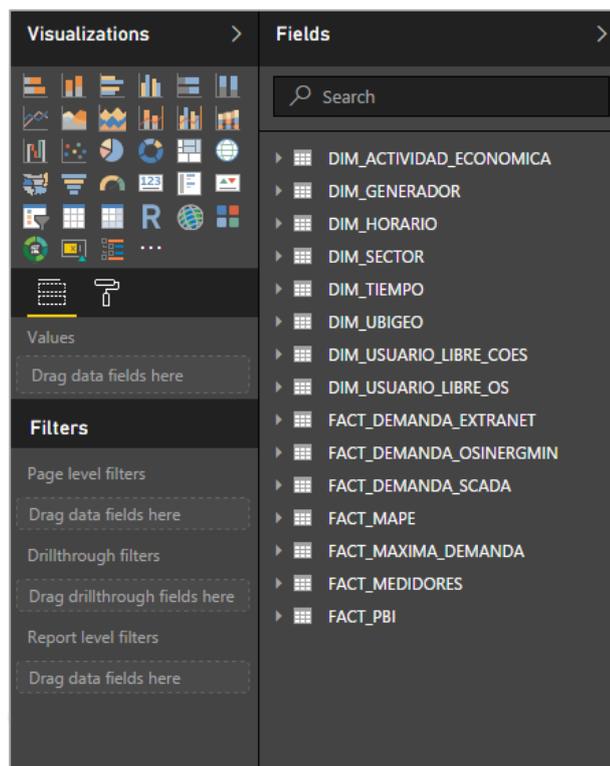


Figura 100: Tablas cargadas en Power BI Desktop
Fuente: Elaboración propia

Una vez con las tablas cargadas a la herramienta Power BI, se procede a elaborar los reportes gráficos, en la Figura 101 se muestra un reporte gráfico sobre la demanda SCADA y el MAPE:

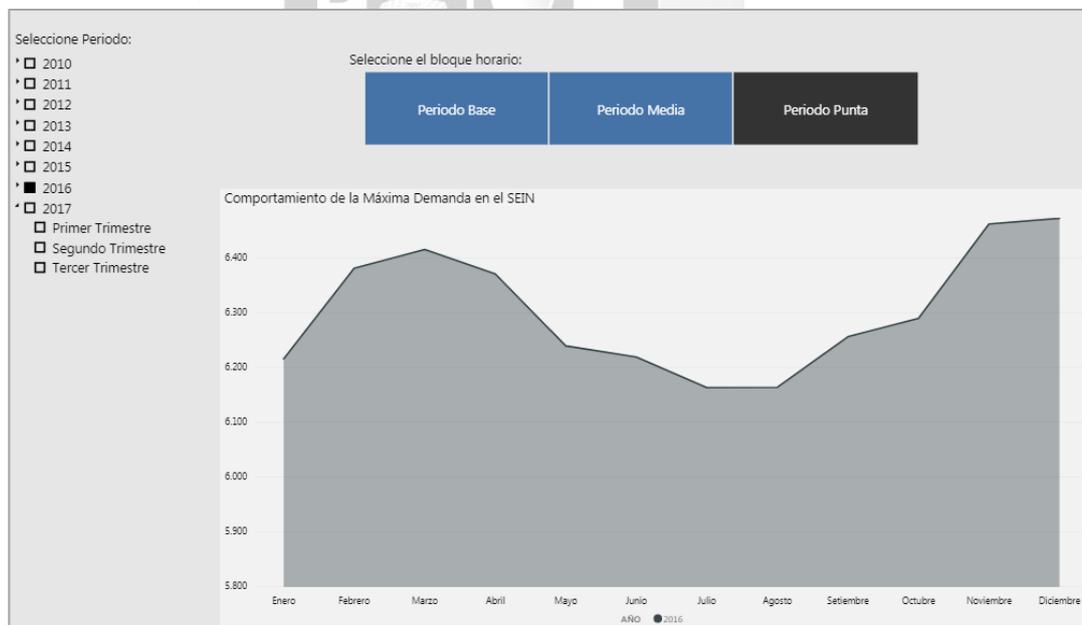


Figura 101: Reporte gráfico en relación a los registros de máxima demanda en el SEIN
Fuente: Elaboración propia

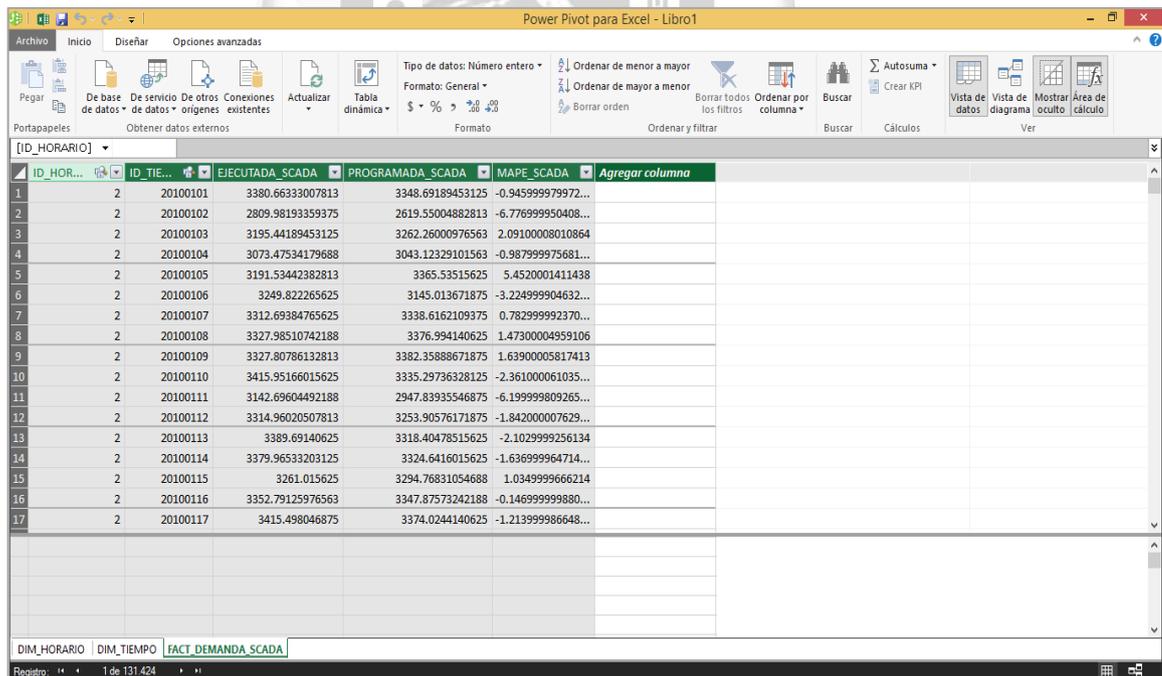
En el Anexo X, se pueden observar los reportes gráficos desarrollados mediante Power BI.

5.9 Implementación

Finalmente, se procede a realizar la implementación de la solución tecnológica. Como se mencionó anteriormente en el punto 5.8 del presente trabajo de investigación, se ha utilizado 2 formas de entregar al usuario el acceso a la información del datamart que son con Power Pivot y Power BI.

5.9.1 Implementación con la herramienta Power Pivot

Para el uso de la herramienta Power Pivot, se debe instalar el complemento respectivo, dicho proceso se puede apreciar en el Anexo XI. Con la herramienta Power Pivot implementada en las computadoras de los trabajadores del COES, se puede consumir la información de los cubos OLAP generados que se encuentran en el Analysis Services o también del modelo dimensional (datamart), cuyo proceso de conexión se observa en el Anexo XII. Una vez realizada la conexión a la fuente de datos seleccionada, mediante la interfaz de Power Pivot se observan los registros almacenados en el repositorio seleccionado como se muestra en la Figura 102 y también la estructura utilizada tal como se observa en la Figura 103.



ID_HORARIO	ID_TIE...	EJECUTADA_SCADA	PROGRAMADA_SCADA	MAPE_SCADA	Agregar columna
1	2	20100101	3380.66333007813	3348.69189453125	-0.94599979972...
2	2	20100102	2809.38193359375	2619.55004882813	-6.776999950408...
3	2	20100103	3195.44189453125	3262.26000976563	2.09100008010864
4	2	20100104	3073.47534179688	3043.12329101563	-0.987999975681...
5	2	20100105	3191.53442382813	3365.53515625	5.4520001411438
6	2	20100106	3249.822265625	3145.013671875	-3.224999904632...
7	2	20100107	3312.69384765625	3338.6162109375	0.782999992370...
8	2	20100108	3327.98510742188	3376.994140625	1.47300004959106
9	2	20100109	3327.80786132813	3382.35888671875	1.63900005817413
10	2	20100110	3415.95166015625	3335.29736328125	-2.361000061035...
11	2	20100111	3142.69604492188	2947.83935546875	-6.199999809265...
12	2	20100112	3314.96020507813	3253.90576171875	-1.842000007629...
13	2	20100113	3389.69140625	3318.40478515625	-2.1029999256134
14	2	20100114	3379.96533203125	3324.6416015625	-1.636999964714...
15	2	20100115	3261.015625	3294.76831054688	1.0349999666214
16	2	20100116	3352.79125976563	3347.87573242188	-0.146999999880...
17	2	20100117	3415.498046875	3374.0244140625	-1.213999986648...

Figura 102: Datos del modelo dimensional (datamart) desde Power Pivot
Fuente: Elaboración propia

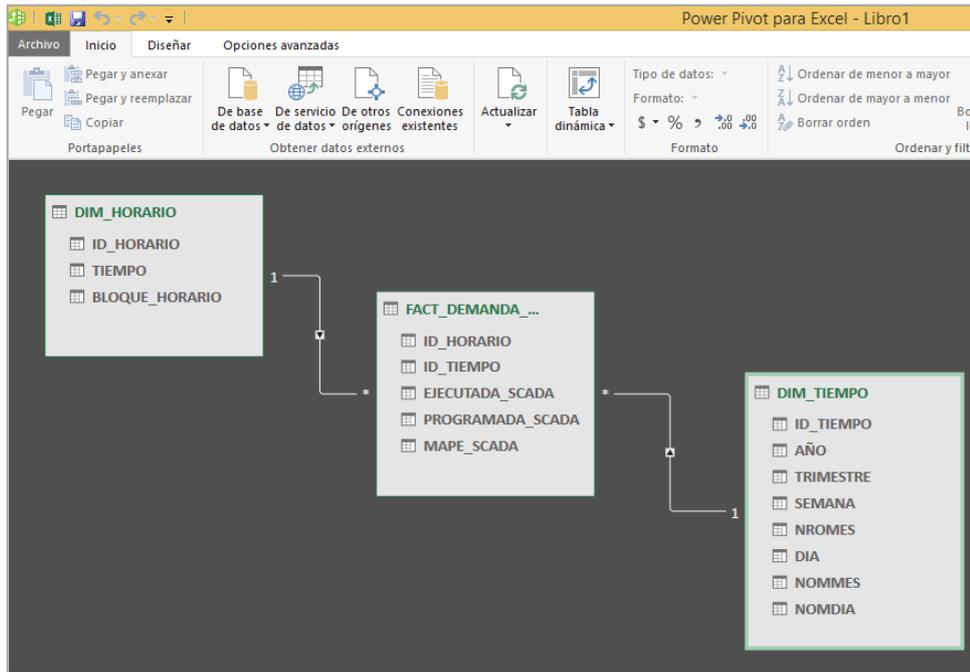


Figura 103: Modelo de datos de las tablas obtenidas desde Power Pivot
Fuente: Elaboración propia

Una vez que los datos han sido obtenidos desde la fuente de datos seleccionada, Power Pivot permite al usuario generar una tabla dinámica para explotar la información obtenida y obtener una mejora facilidad al realizar el análisis respectivo, lo que le permite generar reportes en Excel como se muestra en la Figura 104.

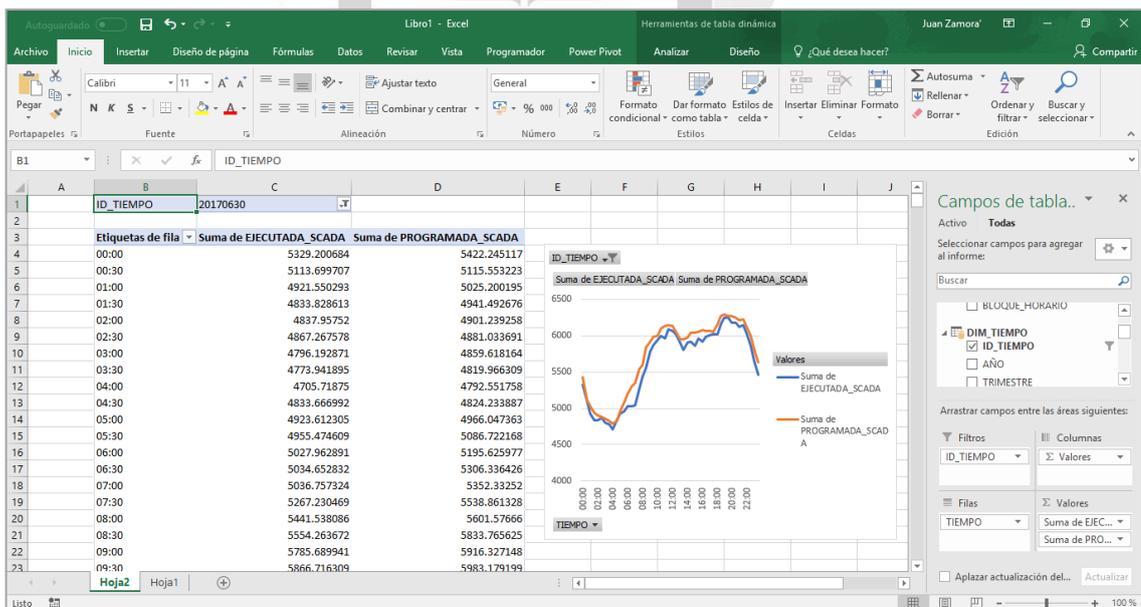


Figura 104: Reporte en tablas dinámicas del datamart desde Power Pivot
Fuente: Elaboración propia

5.9.2 Explotación de datos con Power BI

Para publicar los informes en la web, se procedió a realizar el proceso de “publicación” de Power BI Desktop, el cual se muestra en el Anexo XIII, en el cual también se observa cómo generar los enlaces de acceso para visualizar los reportes gráficos de Power BI que se han realizado, tanto para web como para dispositivos móviles.

Para visualizar los reportes se utilizó la etiqueta HTML generada por la herramienta Power BI, para embeber el reporte gráfico a una página web, como se muestra en le Figura 105.

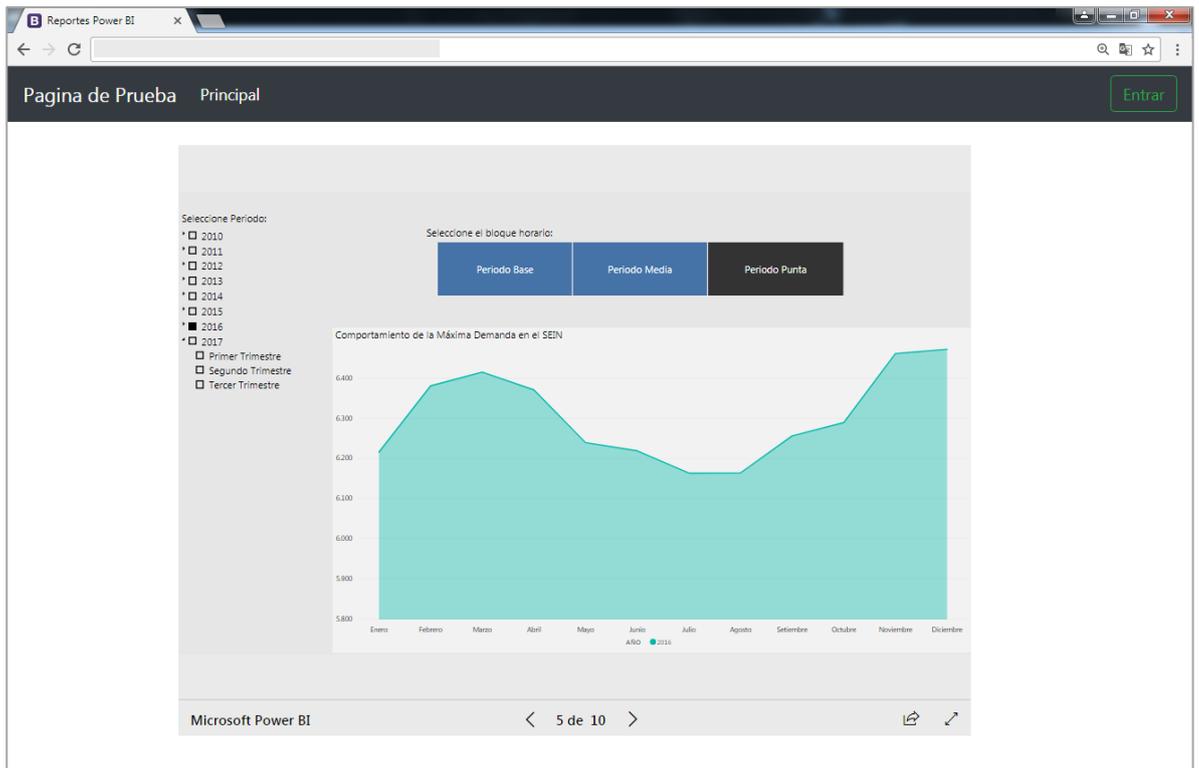


Figura 105: Reporte gráfico de Power BI embebido en HTML
Fuente: Elaboración propia

Así mismo, se utilizó el link de acceso directo al reporte gráfico, generado por la herramienta Power BI, el cual permite ingresar al URL y nos muestra los informes en la web, como se aprecia en la Figura 106.

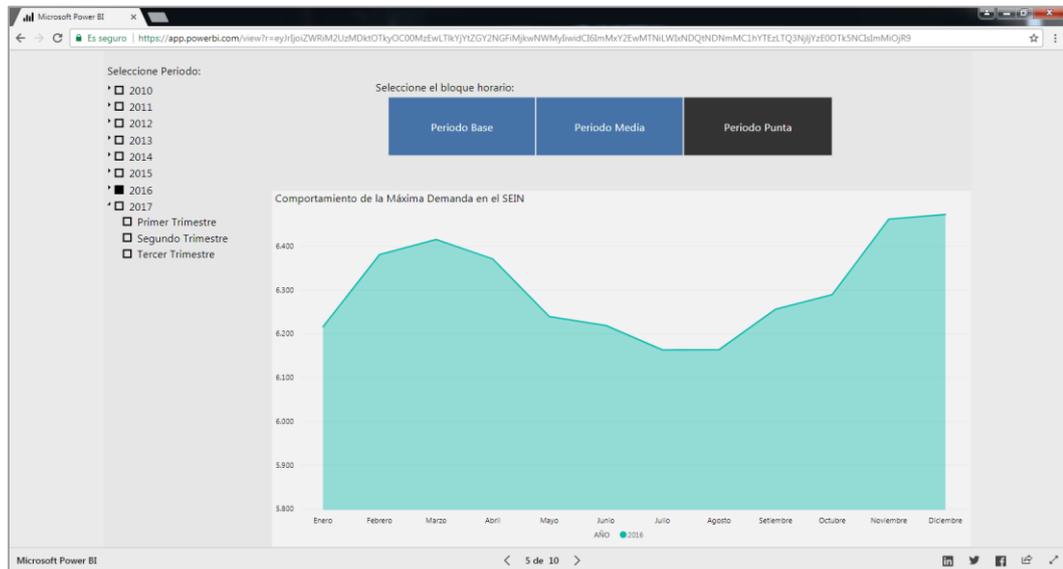


Figura 106: Reporte gráficos de Power BI en la web
Fuente: Elaboración propia

Finalmente, se utilizó el aplicativo de Power BI Mobile para acceder a los reportes gráficos mediante la lectura del código QR generado con la herramienta Power BI, como se aprecia en la Figura 107.

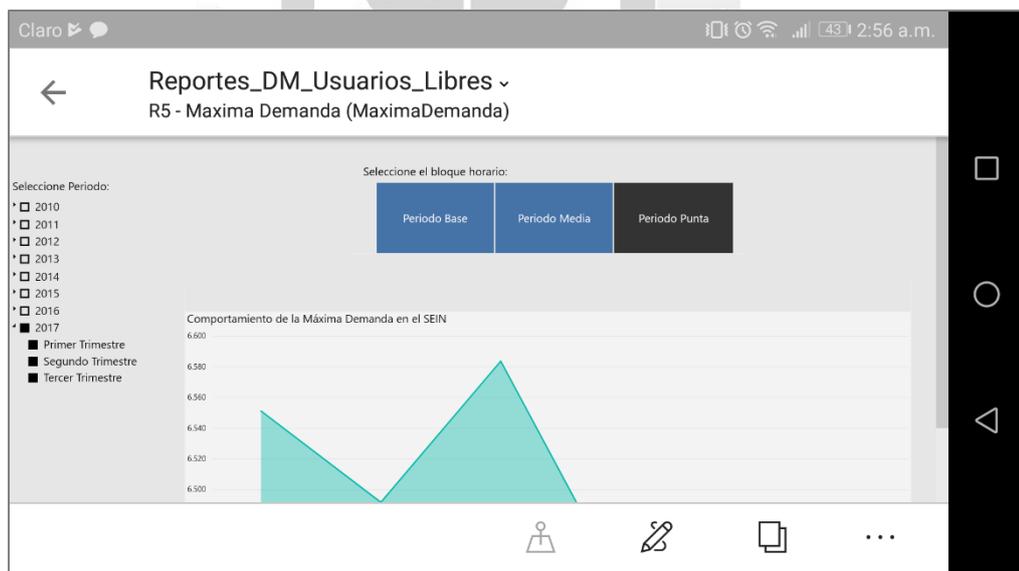


Figura 107: Reporte de registros de máxima demanda desde Power BI Mobile
Fuente: Elaboración propia

CAPÍTULO 6: RESULTADOS

Este capítulo tiene como finalidad, validar el cumplimiento de los indicadores planteados en el capítulo 3, los cuales han sido definidos en el presente trabajo de investigación. Con el fin de obtener el sustento necesario se elaboró una encuesta (Anexo XIV).

La población general de la Sub Dirección de Gestión de Información en el COES que utilizará el datamart es de 10 personas, de las cuales se ha determinado utilizar una muestra de 9 encuestados para poder obtener los resultados. La encuesta fue respondida por 9 personas, a quienes se les presentaron los reportes gráficos generados mediante la herramienta Power BI, y a su vez se les mostró la conexión al datamart mediante Power Pivot, de los cuales se obtuvieron los siguientes resultados:

La primera pregunta, ayudó a conocer si la información del datamart era utilizada por los usuarios, por ello en la Figura 108, se observa la relación entre la cantidad de encuestados y la información del datamart:

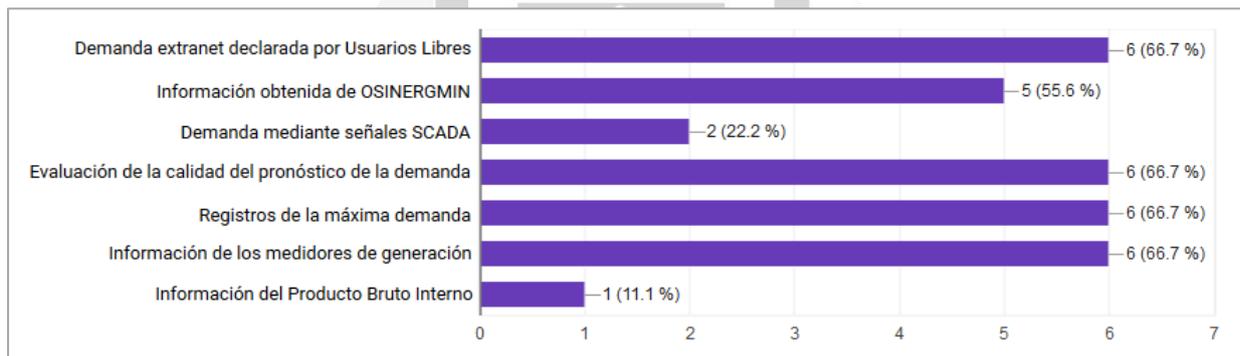


Figura 108: Gráfico estadístico de la relación entre los encuestados y la información utilizada
Fuente: Elaboración Propia

- **Resultados en base al indicador de accesibilidad**

- En la Tabla 20 se observan las respuestas según la escala de Lickert, para determinar si el nivel de accesibilidad permite reducir el tiempo que le generaba descargar los archivos para obtener la data histórica:

Escala de Lickert	Resultados
Muy de acuerdo	5
De acuerdo	4
Ni de acuerdo ni en desacuerdo	0
Desacuerdo	0
Muy en desacuerdo	0

Tabla 20: Respuestas de la encuesta para el indicador de accesibilidad – Pregunta 2
Fuente: Elaboración Propia

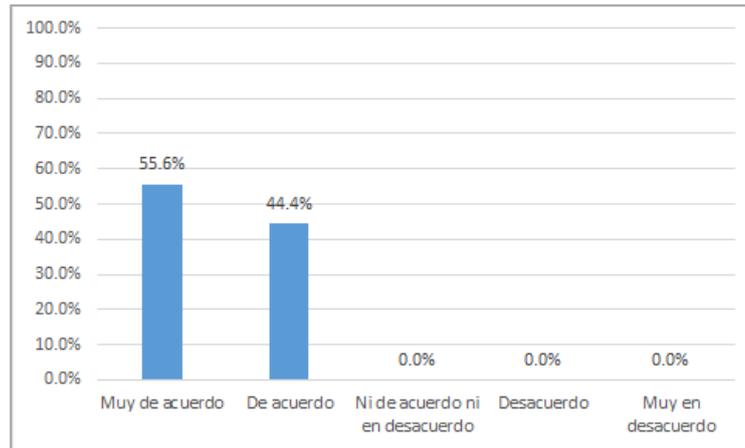


Figura 109: Gráfico estadístico de respuestas de la segunda pregunta de la encuesta
Fuente: Elaboración Propia

Como se aprecia en la Figura 109, el 55.6% se encuentra muy de acuerdo en que el datamart permitió reducir el tiempo que se generaba en descargar los archivos para obtener la data histórica, mientras que el otro 44.4% está de acuerdo.

- En la Tabla 21 se observan las respuestas según la escala de Lickert, para determinar si tener acceso a la información del datamart mediante las tecnologías implementadas, permite mejorar las consultas sobre información histórica de años anteriores:

Escala de Lickert	Resultados
Muy de acuerdo	4
De acuerdo	5
Ni de acuerdo ni en desacuerdo	0
Desacuerdo	0
Muy en desacuerdo	0

Tabla 21: Respuestas de la encuesta para el indicador de accesibilidad – Pregunta 3
Fuente: Elaboración Propia

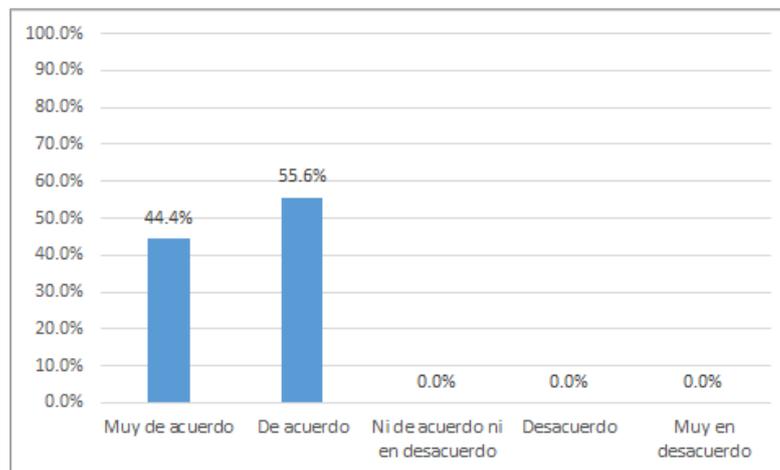


Figura 110: Gráfico estadístico de respuestas de la tercera pregunta de la encuesta
Fuente: Elaboración Propia

Como se aprecia en la Figura 110 el 55,6% está de acuerdo en que el datamart permitió mejorar las consultas sobre información histórica de años anteriores, y un 44,4% está muy de acuerdo.

- **Resultados en base al indicador de fiabilidad**

- En la Tabla 22 se observan las respuestas según la escala de Lickert, para determinar si se considera que los reportes de comportamiento de la demanda eléctrica funcionan en un óptimo estado bajo cualquier circunstancia:

Escala de Lickert	Resultados
Muy de acuerdo	4
De acuerdo	5
Ni de acuerdo ni en desacuerdo	0
Desacuerdo	0
Muy en desacuerdo	0

Tabla 22: Respuestas de la encuesta para el indicador de fiabilidad – Pregunta 4
Fuente: Elaboración Propia

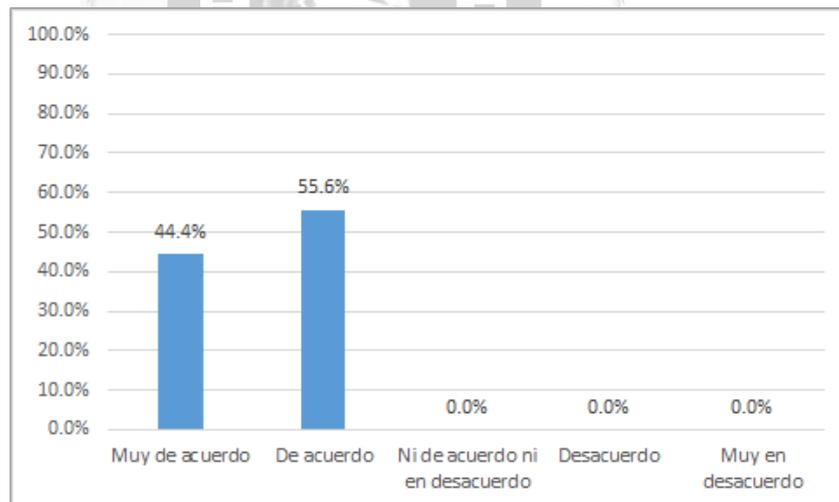


Figura 111: Gráfico estadístico de respuestas de la cuarta pregunta de la encuesta
Fuente: Elaboración Propia

Como se aprecia en la Figura 111, un 55,6% está de acuerdo en que el datamart permitió obtener reportes de comportamiento de la demanda eléctrica que funcionen en un estado óptimo bajo cualquier circunstancia, y un 44,4% está muy de acuerdo.

- En la Tabla 23 se observan las respuestas según la escala de Lickert, para determinar si la información que se presenta en los reportes gráficos en relación con el datamart tiene información fiable para el análisis de la demanda eléctrica:

Escala de Lickert	Resultados
Muy de acuerdo	4
De acuerdo	5
Ni de acuerdo ni en desacuerdo	0
Desacuerdo	0
Muy en desacuerdo	0

Tabla 23: Respuestas de la encuesta para el indicador de fiabilidad – Pregunta 5
Fuente: Elaboración Propia

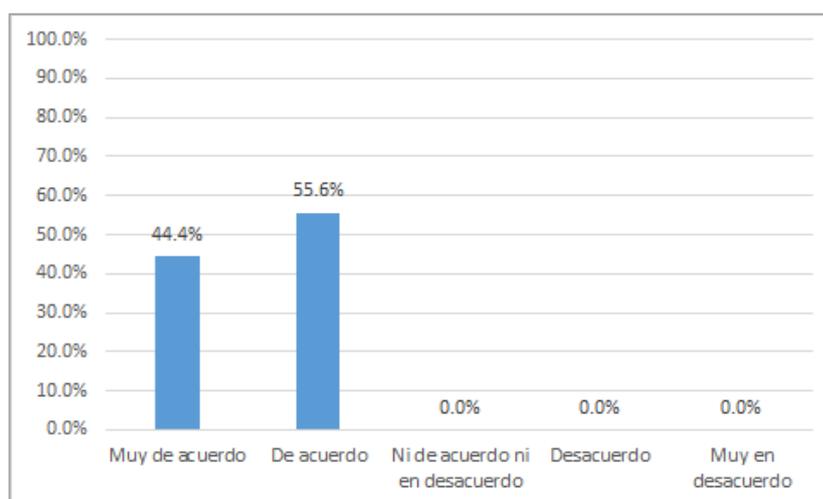


Figura 112: Gráfico estadístico de respuestas de la quinta pregunta de la encuesta
Fuente: Elaboración Propia

Como se aprecia en la Figura 112, un 55,6% está de acuerdo en que el datamart permitió que los reportes gráficos cuenten con información fiable para el análisis de la demanda eléctrica, y un 44,4% está muy de acuerdo.

• **Resultados en base al indicador de eficiencia**

- En la Tabla 24 se observan las respuestas según la escala de Lickert, para determinar si se considera que el uso de un datamart para consultar grandes volúmenes de datos es más eficiente con la tecnología actual (datamart) que mediante el uso de hojas de cálculo:

Escala de Lickert	Resultados
Muy de acuerdo	5
De acuerdo	4
Ni de acuerdo ni en desacuerdo	0
Desacuerdo	0
Muy en desacuerdo	0

Tabla 24: Respuestas de la encuesta para el indicador de eficiencia – Pregunta 6
Fuente: Elaboración Propia

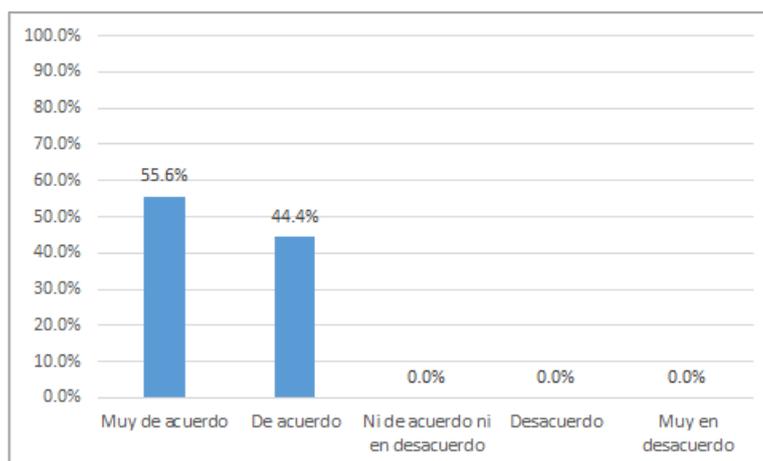


Figura 113: Gráfico estadístico de respuestas de la sexta pregunta de la encuesta
Fuente: Elaboración Propia

Como se aprecia en la Figura 113, el 55,6% se encuentra muy de acuerdo en que el datamart permitió consultar grandes volúmenes de datos de una forma más eficiente que mediante el uso de hojas de cálculo, mientras que el otro 44,4% está de acuerdo.

- En la Tabla 25 se observan las respuestas según la escala de Lickert, para determinar si el nivel de eficiencia ha mejorado en la disminución de tiempos que le tomaba para preparar la información en la estructura requerida para poder ser analizada:

Escala de Lickert	Resultados
Muy de acuerdo	4
De acuerdo	5
Ni de acuerdo ni en desacuerdo	0
Desacuerdo	0
Muy en desacuerdo	0

Tabla 25: Respuestas de la encuesta para el indicador de eficiencia – Pregunta 7
Fuente: Elaboración Propia

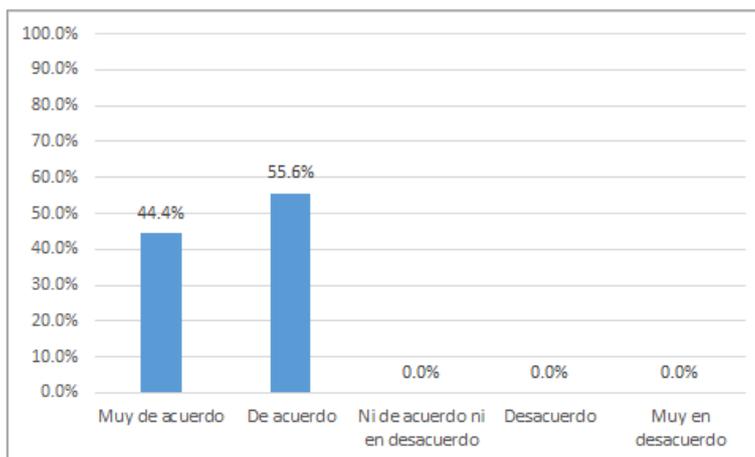


Figura 114: Gráfico estadístico de respuestas de la séptima pregunta de la encuesta
Fuente: Elaboración Propia

Como se aprecia en la Figura 114, el 55,6% se encuentra de acuerdo en que el datamart permitió disminuir el tiempo que se tomaba en preparar la información en la estructura requerida para poder ser analizada, y un 44,4% está muy de acuerdo.

- En la Tabla 26 se observan las respuestas según la escala de Lickert, para determinar si el datamart permite reducir el tiempo de horas hombre que toma en preparar la información para generar reportes o analizar los datos históricos:

Escala de Lickert	Resultados
Muy de acuerdo	5
De acuerdo	4
Ni de acuerdo ni en desacuerdo	0
Desacuerdo	0
Muy en desacuerdo	0

Tabla 26: Respuestas de la encuesta para el indicador de eficiencia – Pregunta 8
Fuente: Elaboración Propia

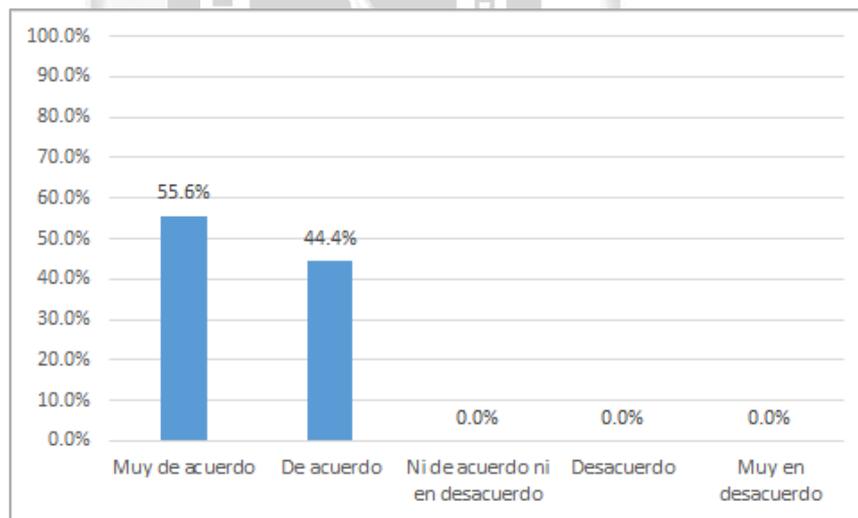


Figura 115: Gráfico estadístico de respuestas de la octava pregunta de la encuesta
Fuente: Elaboración Propia

Como se aprecia en la Figura 115, el 55,6% se encuentra muy de acuerdo en que el datamart permitió reducir el tiempo de horas hombre que toma en preparar la información para generar reportes o analizar los datos históricos, mientras que el otro 44,4% sólo está de acuerdo.

• Resultados en base al indicador de disponibilidad

- En la Tabla 27 se observan las respuestas según la escala de Lickert, para determinar si la disponibilidad de información histórica que ofrece el datamart permite satisfacer las necesidades en cuánto a la cantidad de datos para el análisis de demanda eléctrica:

Escala de Lickert	Resultados
Muy de acuerdo	4
De acuerdo	5
Ni de acuerdo ni en desacuerdo	0
Desacuerdo	0
Muy en desacuerdo	0

Tabla 27: Respuestas de la encuesta para el indicador de disponibilidad – Pregunta 9
Fuente: Elaboración Propia

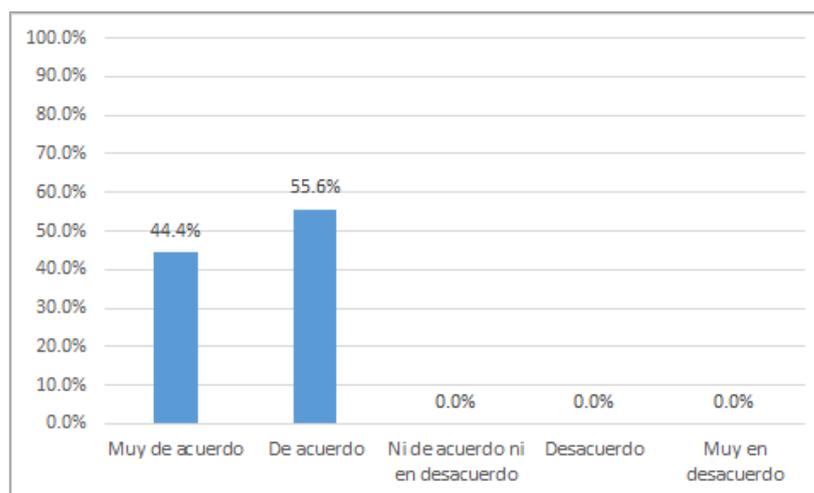


Figura 116: Gráfico estadístico de respuestas de la novena pregunta de la encuesta
Fuente: Elaboración Propia

Como se aprecia en la Figura 116, el 55,6% se encuentra de acuerdo en que el datamart permitió satisfacer las necesidades en cuanto a la cantidad de datos para en análisis de demanda eléctrica, y un 44,4% está muy de acuerdo.

- En la Tabla 28 se observan las respuestas según la escala de Lickert, para determinar si el datamart y la solución tecnológica implementada permite un mayor nivel de disponibilidad en comparación a otros métodos utilizados:

Escala de Lickert	Resultados
Muy de acuerdo	3
De acuerdo	6
Ni de acuerdo ni en desacuerdo	0
Desacuerdo	0
Muy en desacuerdo	0

Tabla 28: Respuestas de la encuesta para el indicador de disponibilidad – Pregunta 10
Fuente: Elaboración Propia

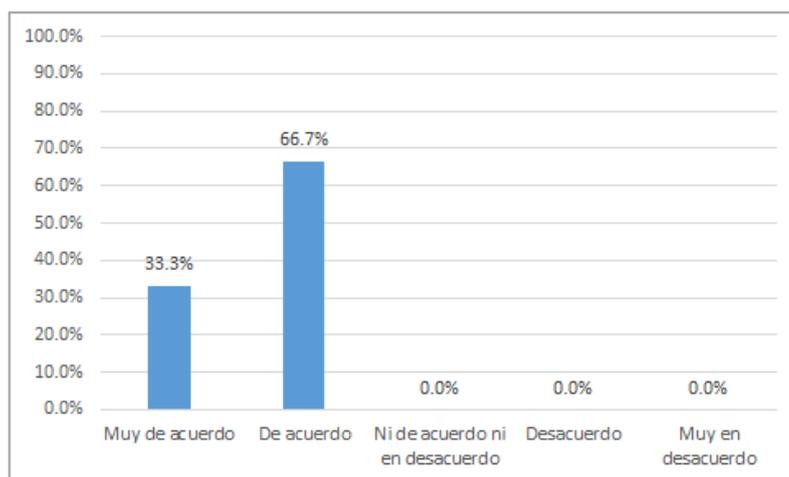


Figura 117: Gráfico estadístico de respuestas de la décima pregunta de la encuesta
Fuente: Elaboración Propia

Como se aprecia en la Figura 117, el 66,7% se encuentra de acuerdo en que el datamart permitió un mayor nivel de disponibilidad en comparación a otros métodos utilizados (hojas de cálculo) y un 33,3% está muy de acuerdo.

- En la Tabla 29 se observan las respuestas según la escala de Lickert, para determinar si hubo momentos en el que la información del datamart no se encontró disponible:

Escala de Lickert	Resultados
Muy de acuerdo	0
De acuerdo	0
Ni de acuerdo ni en desacuerdo	0
Desacuerdo	9
Muy en desacuerdo	0

Tabla 29: Respuestas de la encuesta para el indicador de disponibilidad – Pregunta 11
Fuente: Elaboración Propia

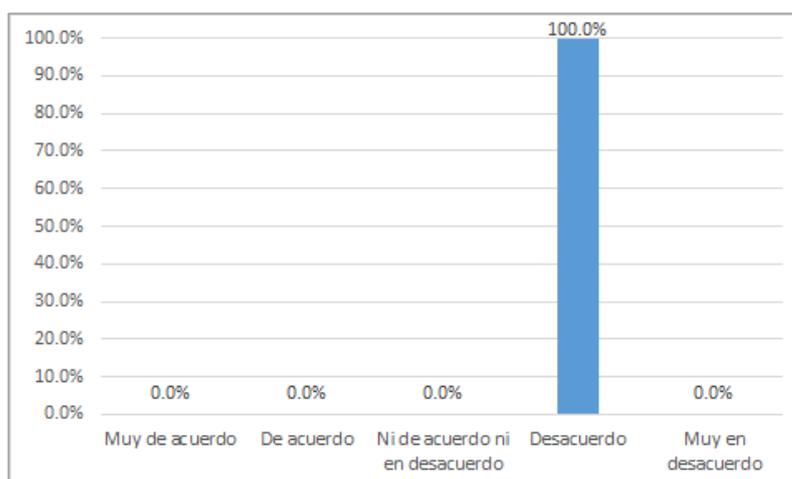


Figura 118: Gráfico estadístico de respuestas de la undécima pregunta de la encuesta
Fuente: Elaboración Propia

Como se aprecia en la Figura 118, el 100% afirma que no hubo ningún momento en que la información del datamart no estuvo disponible.



CONCLUSIONES

- La implementación del datamart influyó satisfactoriamente en la toma de decisiones ya que permitió a los usuarios acceder a un repositorio que cuente con la información necesaria reduciendo el tiempo que tomaba la descarga de información histórica.
- Se determinó que la implementación del datamart influyó positivamente en cuanto a la satisfacción de los usuarios con el funcionamiento de los reportes en un estado óptimo bajo cualquier circunstancia con data fiable para el análisis de la demanda eléctrica y una posterior toma de decisiones.
- Por medio de la implementación del datamart, la eficiencia en consultar grandes volúmenes de datos y la reducción de horas/hombre en la elaboración de reportes gráficos permitieron mejorar la toma de decisiones para el control de la demanda eléctrica.
- Mediante la implementación del datamart se mejoró notablemente la disponibilidad de información, permitiendo recuperar la cantidad de datos necesarios para el análisis de la demanda eléctrica.
- Finalmente, según los objetivos planteados en el presente trabajo de investigación, y así mismo, tomando en cuenta los resultados obtenidos, se concluye en que el datamart influyó satisfactoriamente en la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional.

RECOMENDACIONES

- Implementar herramientas que permitan a los trabajadores, utilizar la información directa de los repositorios de datos con la data actualizada, y sin probabilidades de alterar los registros de las bases de datos que se almacenan en el servidor, y no únicamente del datamart.
- Implementar un servidor para herramientas de explotación de datos, a fin de generar reportes gráficos con los cuales los agentes del Sistema Eléctrico Interconectado Nacional puedan realizar consultas históricas que les permitan obtener una retroalimentación; ello también permitirá a inversionistas observar el comportamiento eléctrico histórico del país, a fin de generar una visión próspera para las inversiones en el sector eléctrico peruano.
- Configurar una red interna, la cual permita establecer la confidencialidad en la visualización de los reportes que exploten la información del datamart de la organización.
- Enfatizar en la planificación del proceso de extracción, transformación y carga (ETL), debido a que es uno de los procesos más relevantes para la elaboración de un datamart, por ello realizar un proceso óptimo de extracción y carga de datos garantiza adecuados informes o reportes propios de la implementación de herramientas para inteligencia de negocios.
- Evaluar el desarrollo de datamarts para las áreas y procesos faltantes del COES, a fin de generar un datawarehouse que permita a la organización obtener un mejor nivel de almacenamiento de los datos con los que cuenta en sus diversos sistemas transaccionales.
- Actualizar la tabla de hechos, a fin de implementar la información de proyecciones a largo plazo con el fin de realizar el análisis de crecimiento a futuro de la demanda.

REFERENCIAS BIBLIOGRÁFICAS

- Ballard, Chuck et al. (2005). *Data Mart Consolidation: Getting Control of Your Enterprise Information*. Estados Unidos.
- Clay, C. y Smith, Bryan. (2008). *Microsoft SQL Server 2008 MDX Step by Step*. Estados Unidos.
- Comité de Operaciones Económicas del Sistema Interconectado Nacional. (2000). *Glosario de Términos*. Recuperado de Página del COES (<http://www.coes.org.pe/portal/>)
- Comité de Operaciones Económicas del Sistema Interconectado Nacional. (2016). *Memoria Anual 2015*. Lima, Perú.
- Comité de Operaciones Económicas del Sistema Interconectado Nacional. (2017). *Portal de Indicadores: Demanda*. Recuperado de Portal Web del COES (<http://www.coes.org.pe/Portal/portalinformacion/demanda>)
- Conesa Caralt, Jordi y Curto Díaz, Josep. (2010). *Introducción al Business Intelligence*. España.
- De los Angeles Ibarra, María. (2006). *Procesamiento Analítico en Línea (OLAP)*. Corrientes, Argentina.
- Espina Alvarado, José. (2017). *Carga, demanda y energía eléctrica: Conceptos fundamentales para la distribución de electricidad*. Recuperado de Sector Electricidad (<http://www.sectorelectricidad.com/17597/carga-demanda-y-energia-electrica-conceptos-fundamentales-para-la-distribucion-de-electricidad/>)
- Galarza Torres, Wilfredo y Valdivieso Zavala, Diego. (2015). *Implementación de Inteligencia de Negocios para la mejora de calidad de suministro de Edelnor*. Perú.
- González Marroquín, Horacio. (2012). *Inteligencia de Negocios en el Desarrollo de Sistemas de Monitoreo de Mercado para el Sector Eléctrico*. Chile.
- González Pinzón, Miguel y González Sanabria, Juan. (2013). *Aplicación del estándar ISO/IEC 9126-3 en el modelo de datos conceptual entidad-relación*. Colombia.

- Guerra García, César A. (2009). *Calidad y Medición de Sistemas de Información, Trabajo: Intruducción de Calidad de Datos (DQ) como requisitos añadidos en el proceso de desarrollo de u producto software*. Castilla-La Mancha, España.
- Han, Jiawei; Kamber, Micheline & Pei, Jian. (2012). *Data Mining Concepts and Techniques (Third Edition)*. Estados Unidos.
- Hernández Gonzáles, Amulfo Napoleón. (2017). *Características de Calidad de Datos de los Almacenes de Datos*. Castilla-La Mancha, España.
- Inmon, William. (2005). *Building the Data Warehouse, Fourth Edition*. Estados Unidos.
- J. Morán, Luis. (2016). *Administración de SQL Server: Integration Services, Analysis Services, Reporting Services*. España.
- Kimball, Ralph y Ross, Margy. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3rd Edition*. Estados Unidos.
- Kimball, Ralph y Caserta, Joe (2004). *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. Estados Unidos.
- Koo Labrín, Alan. (2014). *Introducción a Multidimensional Expresiones (MDX) Libera el poder total de Analysis Services Multidimensional*. Recuperado de Página DOCPLAYER (<http://docplayer.es/24345270-Introduccion-a-multidimensional-expressions-mdx.html>).
- Leonard Brizuela, Eric Ismael y Castro Blanco, Yudi. (2013). *Metodologías para desarrollar Almacén de Datos*. Cuba.
- Lluís Cano, Josep. (2007). *Business Intelligence: Competir con Información*. Barcelona, España.
- Matallana Caffo, Carlos y Vivanco Valdez, Edwin. (2014). *Diseño de un Datamart para el proceso de inscripción en el registro de hidrocarburos de Osinergmin utilizando la metodología Kim Ball*. Perú.
- Martínez Orol, Jorge Alfredo. (2007). *OLAP y el diseño de cubos*. Recuperado de Gestipolis (<https://www.gestipolis.com/olap-y-el-diseno-de-cubos/>)

Mencías Imbaquingo, James y Mencías Imbaquingo, Amy. (2012). *Desarrollo de un aplicativo Business Intelligence para el área administrativo financiera de la Empresa Eléctrico “Quito” S.A.* Ecuador.

Microsoft Technet. (2017). *Funciones (Sintaxis de MDX)*. Recuperado de Página Oficial de Microsoft Technet ([https://technet.microsoft.com/es-es/library/ms145563\(v=sql.1110\).aspx#Anchor_0](https://technet.microsoft.com/es-es/library/ms145563(v=sql.1110).aspx#Anchor_0))

Microsoft. (2017). *¿Qué es Power BI?*. Recuperado de Página Oficial de Microsoft (<https://powerbi.microsoft.com/es-es/what-is-power-bi/>)

Mike M. (s.f.). *¿A qué se refiere el cobro por demanda que aparece en mi factura de electricidad de mi cuenta comercial? ¿Qué es demanda?*. Recuperado de Página de ONCOR (<http://www.askoncor.com/ES/Paginas/FAQs/Billing-and-Rates-8.aspx>)

Modak Analytics. (2014). *Data Warehousing and Business Intelligence*. Recuperado de Página Oficial de Modak Analytivs (<http://www.modakanalytics.com/business-intelligence-and-data-warehousing.html>)

Organismo Supervisor de la Inversión en Energía y Minería. (2016). *La industria de la electricidad en el Perú: 25 años de aportes al crecimiento económico del país*. Perú.

Organismo Supervisor de la Inversión en Energía y Minería. (2006). *Ley 28832: Ley para Asegurar el Desarrollo Eficiente de la Generación Eléctrica*. Perú.

El Peruano Diario Oficial. (2009). *Decreto Supremo N° 022-2009-EM, Reglamento de Usuarios Libres de Electricidad*. Perú.

Peña Gutiérrez, Gonzalo y Pincheira Henríquez, Iván. (2014), *Implementación de Business Intelligence para una Pyme local del rubro eléctrico*, Chile.

Pérez López, Esteban. (2015). *Los sistemas SCADA en la automatización industrial. Tecnología en Marcha Vol. 28, N°4*. Costa Rica Rodríguez, A. (2008). *Sistemas SCADA*. 2 ed. Barcelona: Editorial Marcombo.

Power Data Especialistas en Gestión de Datos. (2015). *El valor de la gestión de datos*. Recuperado de Página Oficial de Power Data (<http://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/405060/qu-significa-la-integraci-n-de-datos>)

R. Darío, Bernabeu (2009). *Datawarehouse manager*. Recuperado de Dataprix (<http://www.dataprix.com/datawarehouse-manager#x1-520003.4.5.3>)

Salvador Ramos (2016). *Data Warehouse, Data Marts y Modelos Dimensionales. Un pilar fundamental para la toma de decisiones*. España.

Salvador Ramos (2015). *Excel 2013, Power Pivot y DAX. Tus análisis elevados a la enésima potencia*. España.

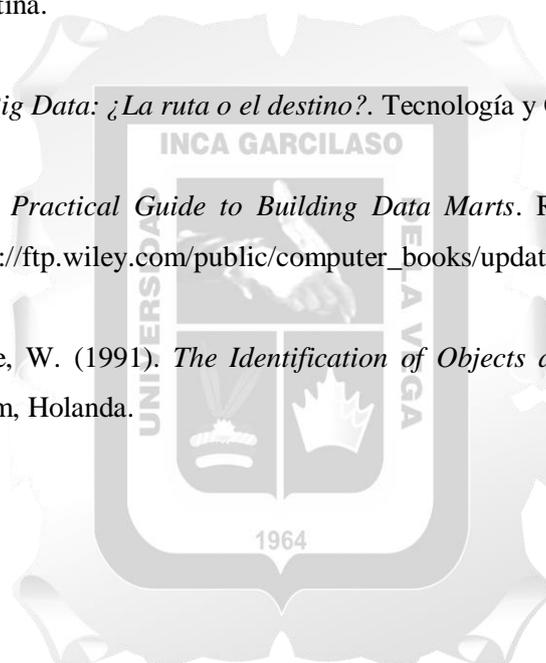
Salvador Ramos (2011). *Microsoft Business Intelligence: vea el cubo medio lleno*. España.

Rivadiera, Gustavo R. (2010). *La metodología de Kimball para el diseño de almacenes de datos (Data warehouses)*. Argentina.

Salvador, Fabrizio (2014). *Big Data: ¿La ruta o el destino?*. Tecnología y Crecimiento. España.

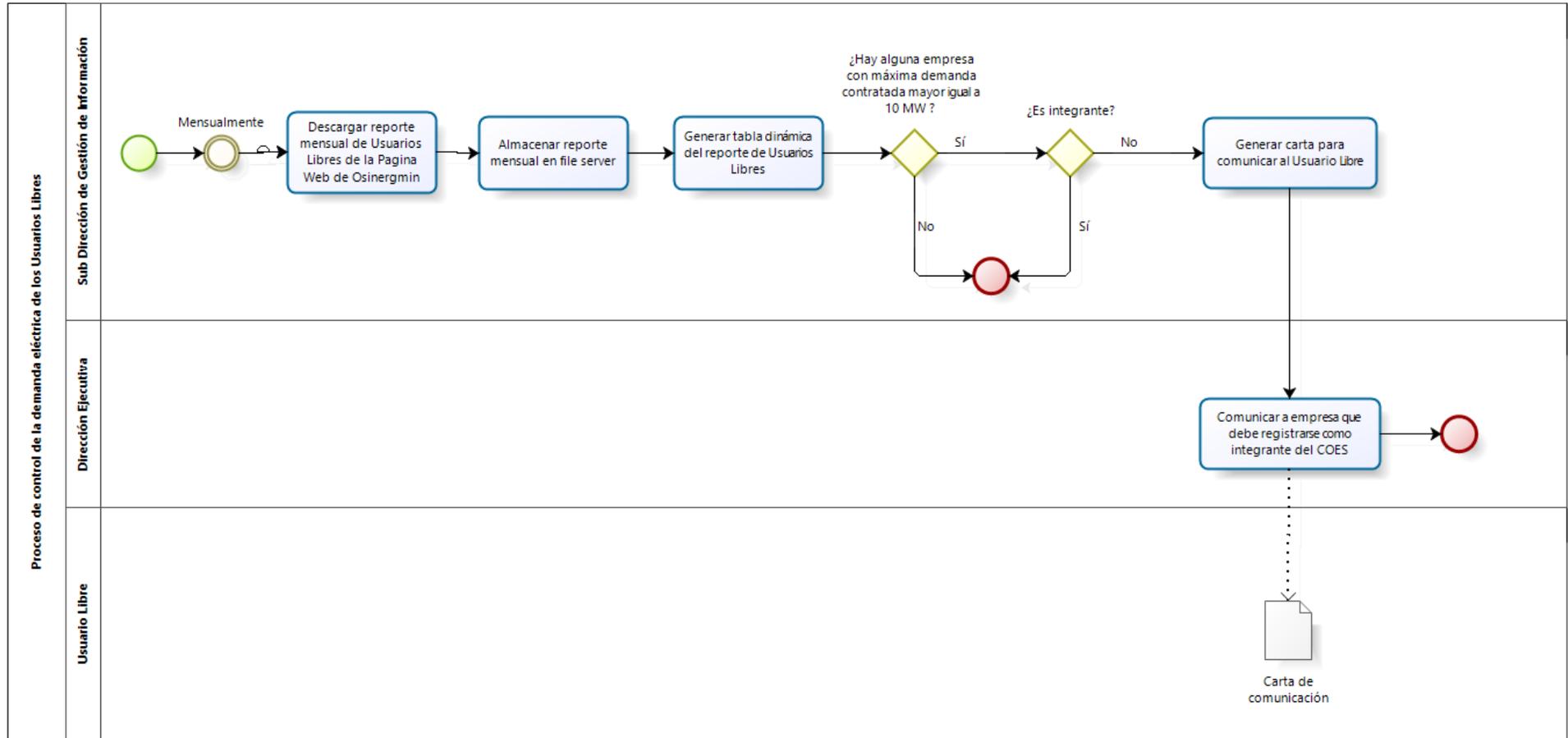
Talus Company. (1998). *A Practical Guide to Building Data Marts*. Recuperado de Wiley, public computer books (ftp://ftp.wiley.com/public/computer_books/updates/guide.pdf.)

Wieringa, R. J. y De Jonge, W. (1991). *The Identification of Objects and Roles - object identifiers revisited*. Amsterdam, Holanda.



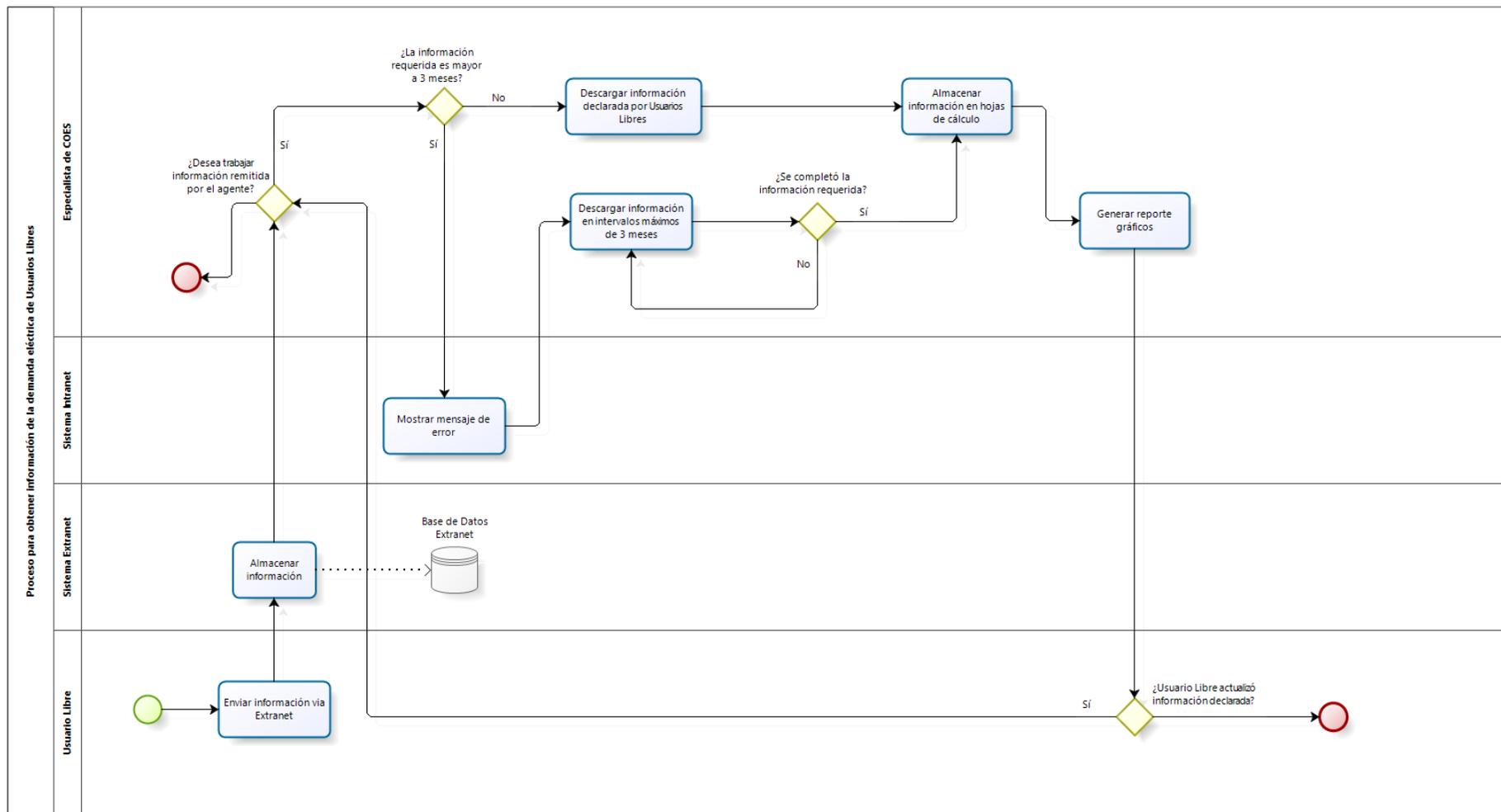
ANEXO

ANEXO I: FLUJOGRAMA DEL PROCESO DE CONTROL DE LA DEMANDA DE LOS USUARIOS LIBRES



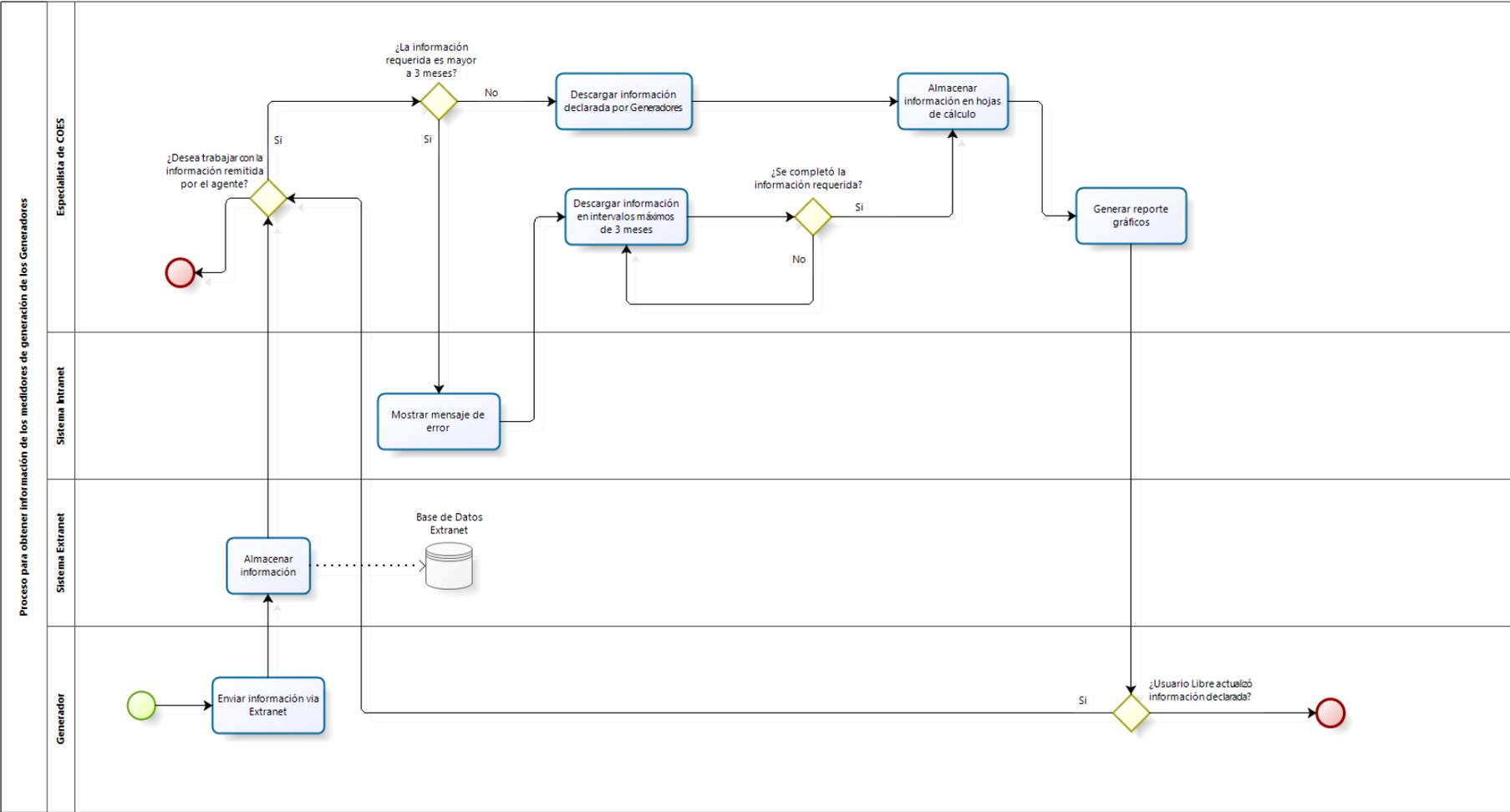
Fuente: Elaboración Propia

ANEXO II: FLUJOGRAMA DEL PROCESO PARA OBTENER INFORMACIÓN DE LA DEMANDA ELÉCTRICA DE LOS USUARIOS LIBRES



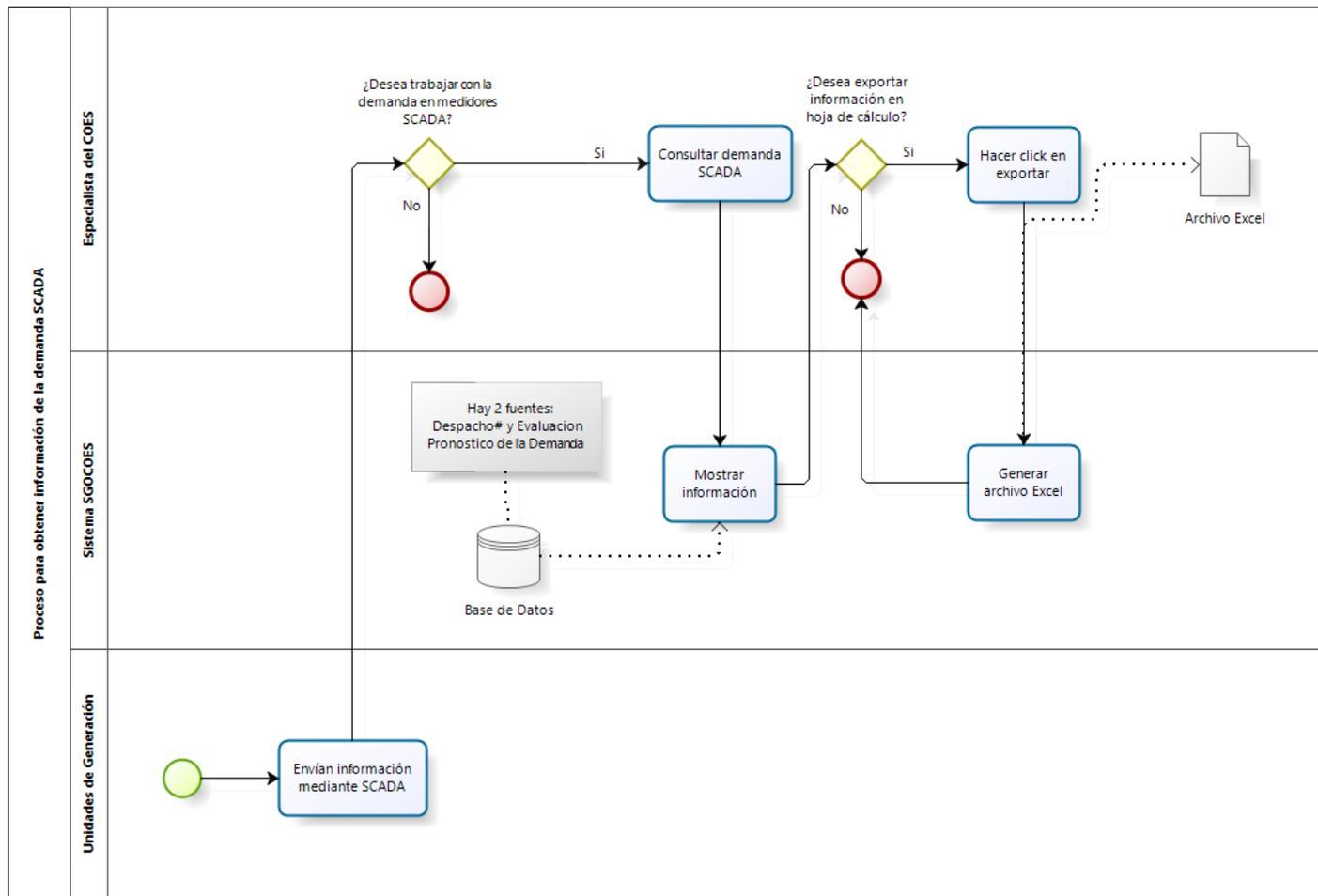
Fuente: Elaboración Propia

ANEXO III: FLUJOGRAMA DEL PROCESO PARA OBTENER INFORMACIÓN DE LOS MEDIDORES DE GENERACIÓN DE LOS GENERADORES



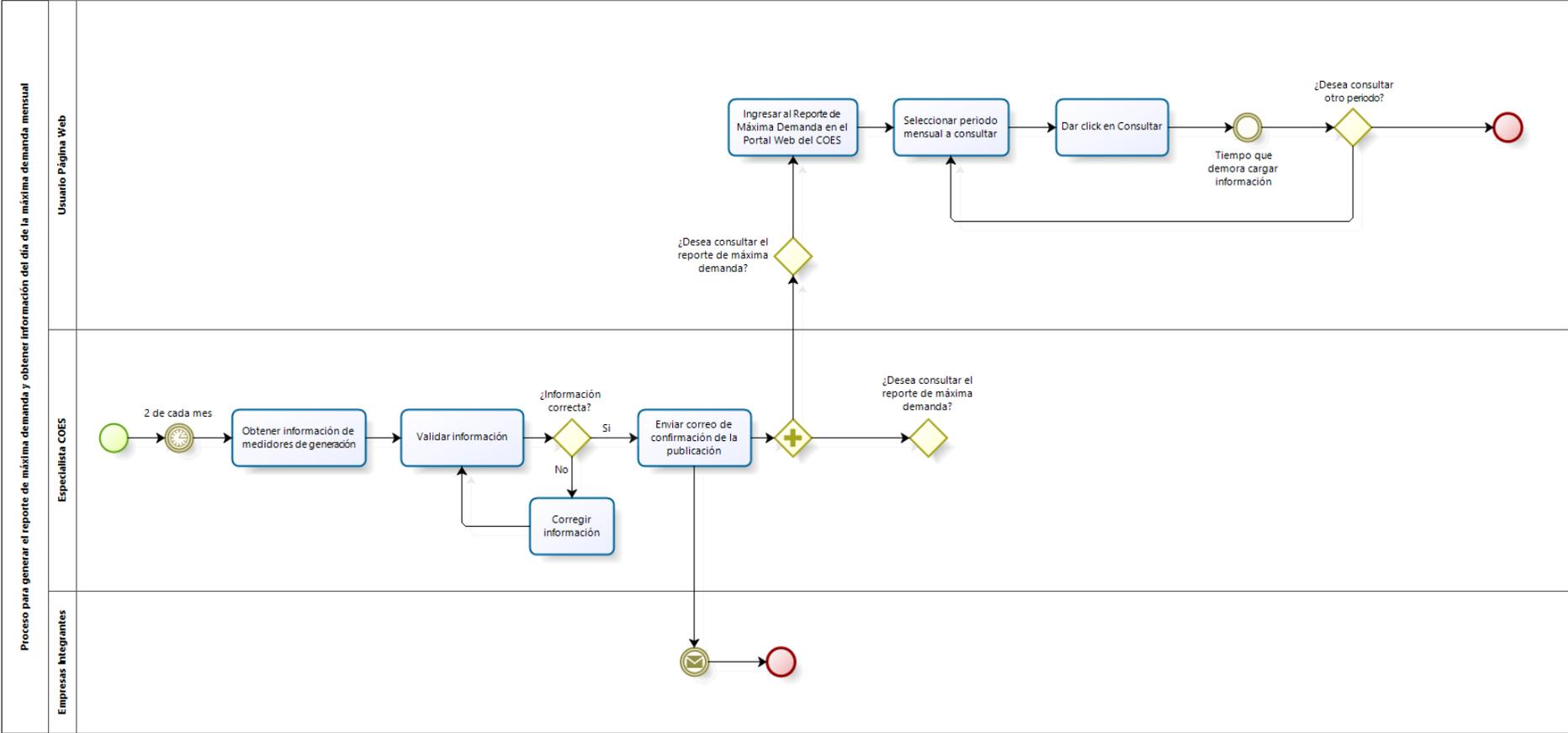
Fuente: Elaboración Propia

ANEXO IV: FLUJOGRAMA DEL PROCESO PARA OBTENER INFORMACIÓN DE LA DEMANDA SCADA



Fuente: *Elaboración Propia*

ANEXO V: FLUJOGRAMA DEL PROCESO PARA GENERAR EL REPORTE DE MÁXIMA DEMANDA Y OBTENER INFORMACIÓN DEL DÍA DE LA MÁXIMA DEMANDA MENSUAL



Fuente: Elaboración Propia

ANEXO VI: MATRIZ DE COHERENCIA INTERNA

	PROBLEMAS	OBJETIVOS	HIPÓTESIS	VARIABLES	INDICADORES
GENERAL	¿En qué medida la Implementación de un Datamart, influye en la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional?	Determinar la influencia de un Datamart para la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional.	La implementación de un Datamart, influye positivamente en la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional.	Datamart	<ul style="list-style-type: none"> • Nivel de accesibilidad. • Nivel de fiabilidad. • Nivel de eficiencia. • Nivel de disponibilidad.
ESPECÍFICO	¿En qué medida el nivel de accesibilidad de un Datamart, influye en la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional?	Determinar el nivel de influencia de la accesibilidad de un Datamart para la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional.	El nivel de accesibilidad de un Datamart influye positivamente en la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional.	Mejora en la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional	<ul style="list-style-type: none"> • Cantidad de reportes de comportamiento de la demanda eléctrica. • Tiempo en la generación de reportes del crecimiento de la demanda eléctrica.

<p>¿En qué medida el nivel de fiabilidad de un Datamart, influye en la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional?</p>	<p>Determinar el nivel de influencia de la fiabilidad de un Datamart para la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional.</p>	<p>El nivel de fiabilidad de un Datamart influye positivamente en la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional.</p>		<ul style="list-style-type: none"> • Volumen de información histórica para analizar sobre demanda eléctrica. • Número de empresas que cumplen con enviar su información de demanda eléctrica.
<p>¿En qué medida el nivel de eficiencia de un Datamart, influye en la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional?</p>	<p>Determinar el nivel de influencia de la eficiencia de un Datamart para la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional.</p>	<p>El nivel de eficiencia de un Datamart influye positivamente en la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional.</p>		
<p>¿En qué medida el nivel de disponibilidad de un Datamart, influye en la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional?</p>	<p>Determinar el nivel de influencia de la disponibilidad de un Datamart para la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional.</p>	<p>El nivel de disponibilidad de un Datamart influye positivamente en la mejora de la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional.</p>		

ANEXO VII: ENTREVISTA PARA LA DEFINICIÓN DE REQUERIMIENTOS

Nombre:	Ing. César Butrón
Empresa:	Comité de Operaciones Económicas del Sistema Interconectado Nacional (COES)
Cargo en la empresa:	Presidente del COES

1. Desde su punto de vista ¿cuál es la situación actual del sector eléctrico peruano?

Hay una situación de sobre oferta de generación dado que la demanda no ha crecido como estuvo proyectada y eso hace que sean los ciclos combinados los que están fijando costos marginales casi todo el tiempo. El resultado es que los costos marginales en operación normal sean consistentemente inferiores a 15 USD/Mwh que es una cifra realmente baja e incapaz de sustentar cualquier nueva inversión. Esto hace que la única manera que los generadores puedan mantener el balance entre sus ingresos y sus gastos y tener rentabilidad es mediante contratos bilaterales de largo plazo y, por tanto, hay una feroz competencia para conseguir esos contratos con los clientes libres y hacer que clientes regulados se pasen a ser libres. Con ello, los precios promedio de clientes libres han bajado de 45 a 22 USD/Mwh y hay muchos clientes regulados que han migrado a libres dejando a los distribuidores con sobrecostos por potencia contratada y, en consecuencia, empeorando sus estados financieros. Entonces el sector es momentos de mucha competencia y reducción de ingresos tanto para generadores como distribuidores. En contrapartida, la transmisión, merced al Plan de Transmisión elaborado por el COES, está creciendo de una manera robusta.

2. ¿Cuál es su opinión sobre el crecimiento de la demanda eléctrica en los 5 últimos años? ¿Cuál es la importancia de ésta para el país?

Como dijimos en la anterior pregunta, el crecimiento de la demanda ha desacelerado. De estar creciendo cerca del 6% en los últimos 5 años, ahora el crecimiento es inferior al 2%. La importancia para esto es que refleja lo que ha sucedido con la economía del país en general, ha sufrido una reducción lo cual es mala noticia para todos.

3. Se aproxima el ingreso en vigencia del procedimiento relacionado al Mercado

Mayorista de Electricidad, ¿cómo considera el papel que tomarán los Usuarios Libres en el SEIN?

Según el nuevo reglamento del Mercado Mayorista de Electricidad, los Usuarios Libres podrán comprar hasta el 10% de su demanda en el mercado de corto plazo. En teoría eso debería causar que se conviertan en integrantes del COES más activos pues la compra la harán a costo marginal de manera de deberían estar mucho más enterados de cómo se calculan los marginales y deberían participar en varios comités de trabajo del COES. Sin embargo, mi estimado es que esto solo ocurrirá con los Usuarios Libres más grandes que son los que tienen las capacidades de afrontar esta nueva actividad. Por lo menos inicialmente.

4. ¿Considera que el uso de información histórica es un factor importante para la toma de decisiones en el COES? ¿Por qué?

El uso de información histórica es siempre un factor muy importante para la toma de decisiones de cualquier organización y de hecho lo es para el COES. El análisis de la información histórica permite detectar patrones que a su vez permite anticipar y corregir ciertas situaciones. También permite detectar fallas o anomalías, prepararnos para resolver mejor ciertas situaciones repetitivas.

5. ¿Cuál cree que sería la importancia de la implementación de un datamart en el COES?

Tiene la importancia y la utilidad que se ha indicado en la anterior pregunta. Ahora bien, parece muy importante empezar con un datamart para aprovecharlo al máximo en un área del COES y luego poder trasladar esa experiencia a otras áreas ya sea mediante otros datamarts o bien ya pasando al siguiente nivel de datawarehouse.

6. ¿Cree usted que el COES se vería beneficiado con el uso de indicadores en relación a la información histórica almacenada en un datamart? ¿Por qué?

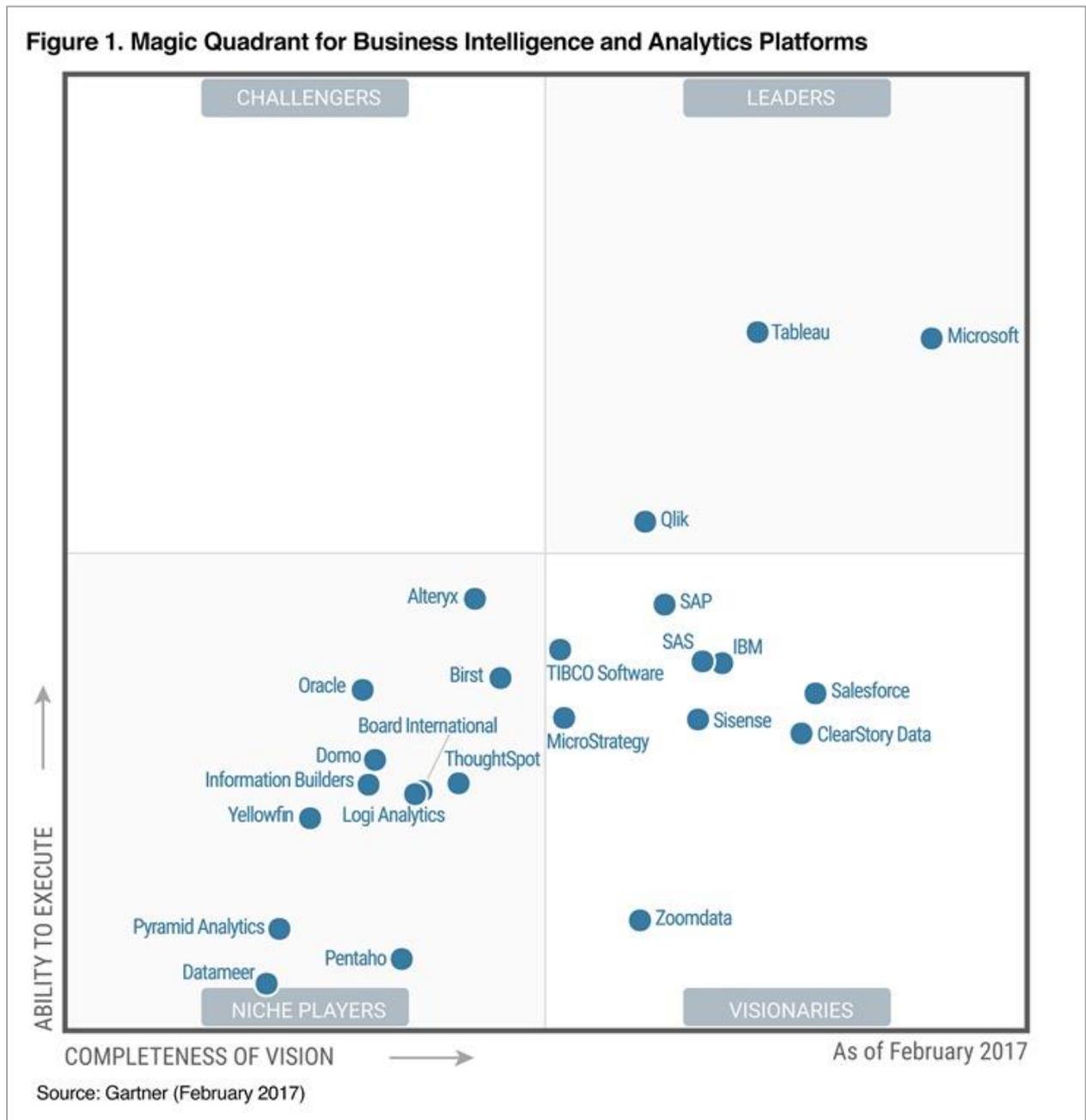
Si se vería beneficiado por las razones indicadas. Se contaría con una herramienta que permita predecir determinadas situaciones o comportamientos, mejorar la proyección de la demanda, identificar patrones, detectar fallas. Reduciría el tiempo empleado por los especialistas del COES para hacer su trabajo por la inherente facilidad de uso que tiene un datamart.

ANEXO VIII: RANKING DE HERRAMIENTAS PARA VISUALIZACIÓN DE DATOS POR SELECTHUB

	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6	Rank 7	Rank 8	Rank 9	Rank 10	Rank 11	Rank 12
Business Intelligence [BI] MATRIX VIEW <input checked="" type="checkbox"/> Collapsed <input type="checkbox"/> Expanded <input type="checkbox"/> Expand comments <input checked="" type="checkbox"/> Expand All Rows	 SAP BusinessOb... 98	 QlikView 96	 IBM Cognos Analytics 96	 Oracle BI 96	 Dundas BI 92	 Microsoft Power BI 92	 Sisense 92	 MicroStrategy Enterprise A... 91	 Tableau Server 90	 Domo 89	 Birst BI 86	 SAS Visual Analytics 80
KEY REQUIREMENTS												
1.0.1 Platform Functions	100	100	100	100	100	100	100	100	85	100	100	100
Priority: High												
1.0.2 Data Visualization	100	100	100	100	100	100	100	70	85	85	70	100
Priority: High												
1.0.3 Analytics	100	100	70	85	85	85	100	85	85	100	85	85
Priority: High												
1.0.4 Online Analytical Processing (OLAP)	100	100	100	100	100	100	100	100	100	100	100	100
Priority: High												
1.0.5 Document Management	100	100	100	100	50	50	50	100	100	50	50	50
Priority: High												
1.0.6 Decision Services	85	100	100	100	100	100	100	100	85	100	85	0
Priority: High												
1.0.7 Integrations	100	85	100	85	100	100	100	70	70	100	100	100
Priority: High												
1.0.8 Big Data Integration	100	85	100	100	100	100	85	100	100	70	85	85
Priority: High												
1.0.9 Deployment Environment	100	100	100	100	100	100	100	100	100	100	100	100
Priority: High												
TOP REVIEWS												
2.0.1 Top Analyst and Community Reviews	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Priority: N/a	View Responses	View Responses	View Responses	View Responses	View Responses	View Responses	View Responses	View Responses	View Responses	View Responses	View Responses	View Responses
2.0.2 Social Media Reviews	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Priority: N/a	View Responses	View Responses	View Responses	View Responses	View Responses	View Responses	View Responses	View Responses	View Responses	View Responses	View Responses	View Responses

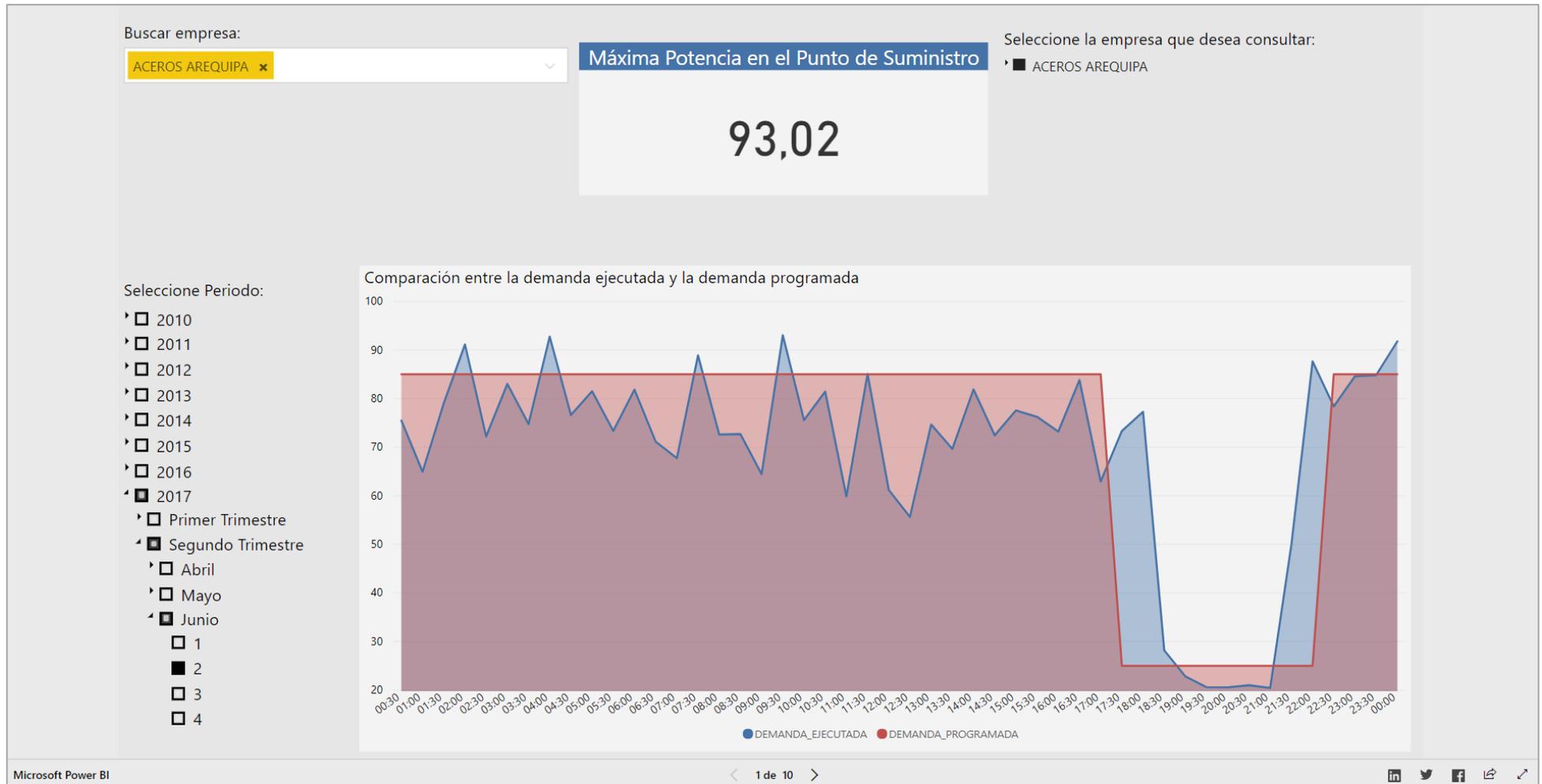
Fuente: SelectHub – Matrix Report Business Intelligence
 (https://app.selecthub.com/reports/580037b2deba89470db76486160c8f0e_9c902db7c92f95c582e2436c15d2272f)

ANEXO IX: CUADRANTE MÁGICO DE GARTNER PARA HERRAMIENTAS DE BUSINESS INTELLIGENCE (FEBRERO 2017)



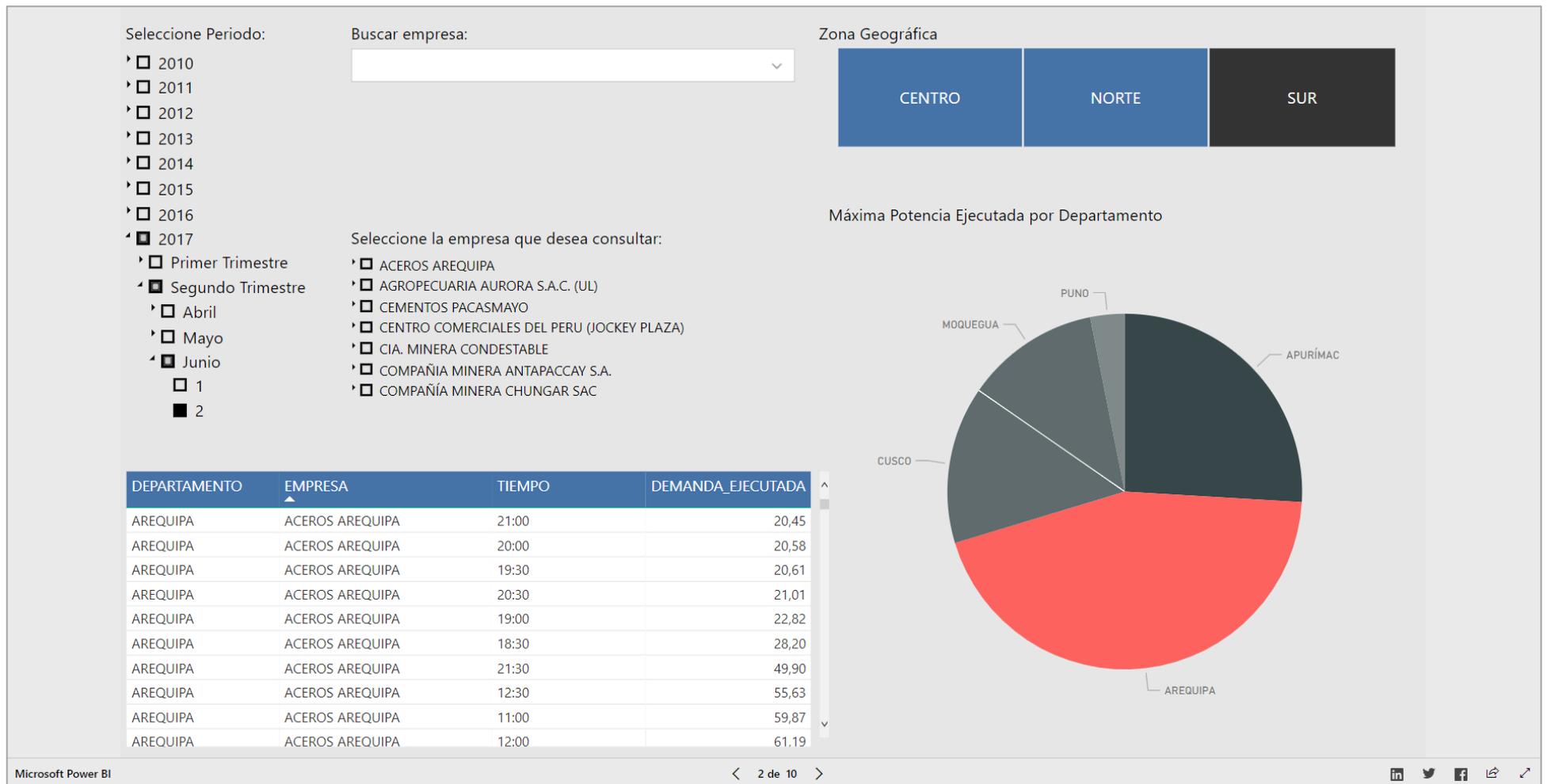
Fuente: Gartner, Febrero 2017

ANEXO X: REPORTE GRÁFICOS GENERADOS CON POWER BI

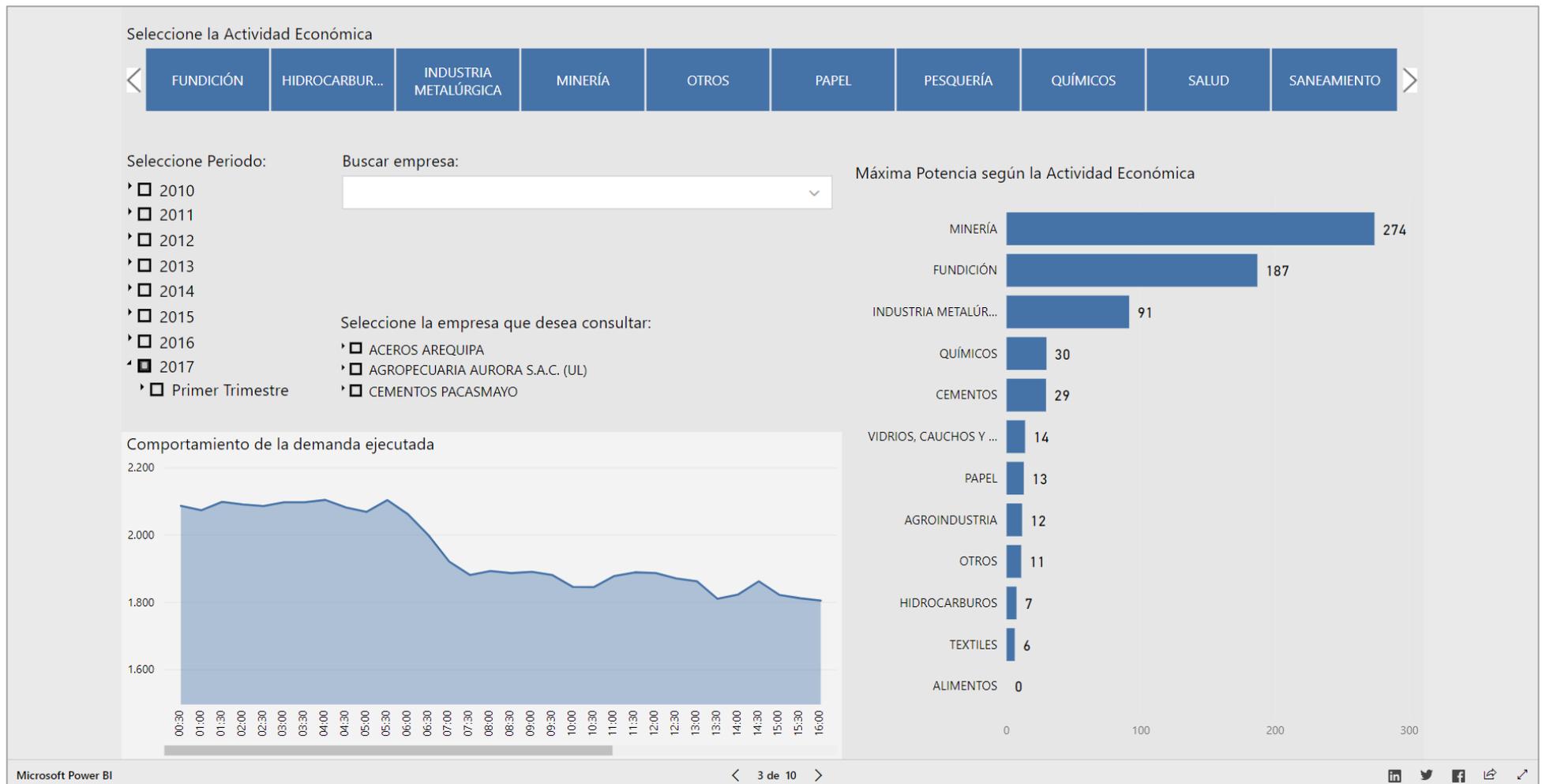


Reporte de la demanda ejecutada y programada para las empresas en el Extranet

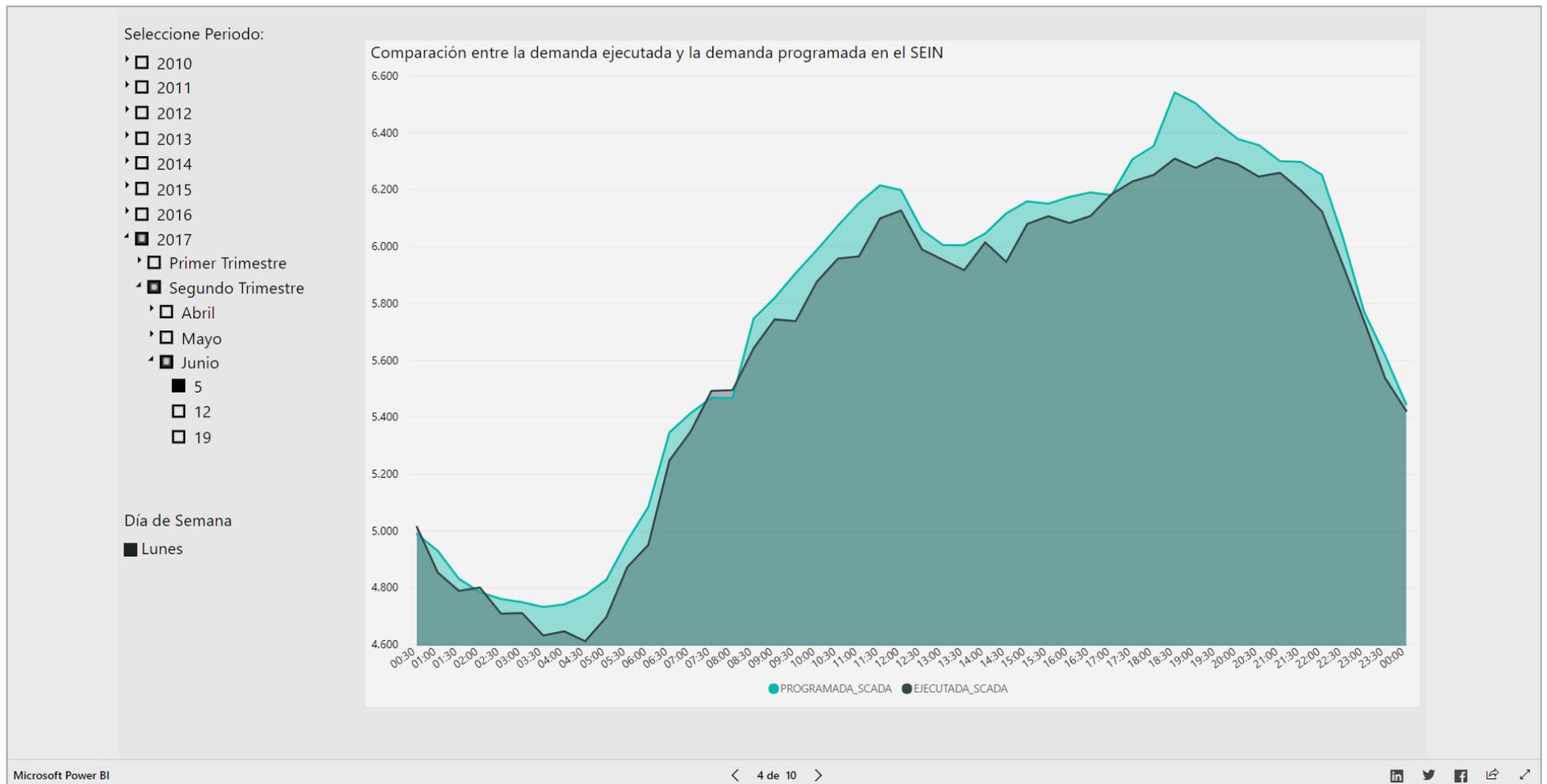
Fuente: Elaboración propia



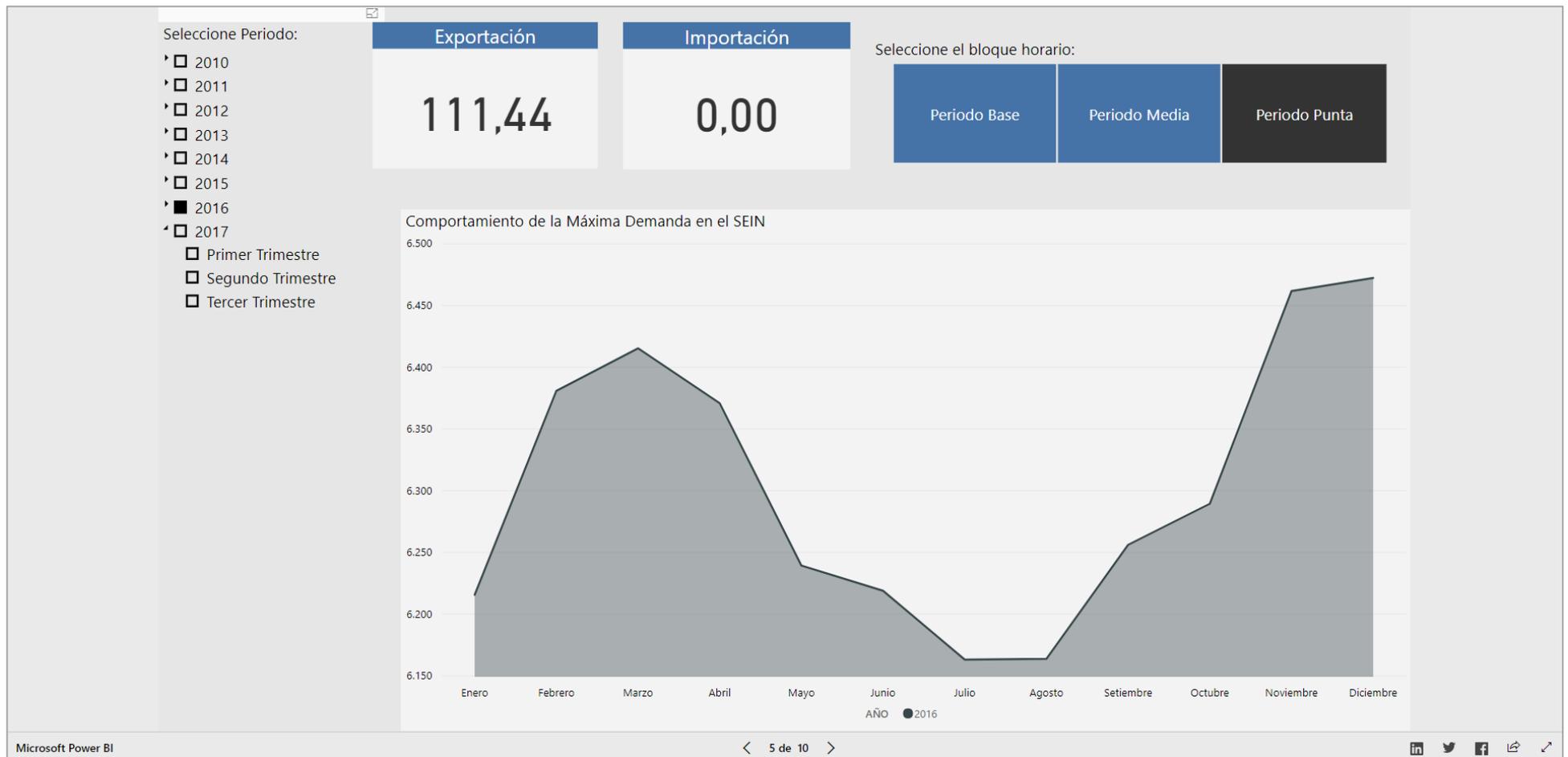
Reporte de la demanda ejecutada por zona geográfica para las empresas en el Extranet
Fuente: Elaboración propia



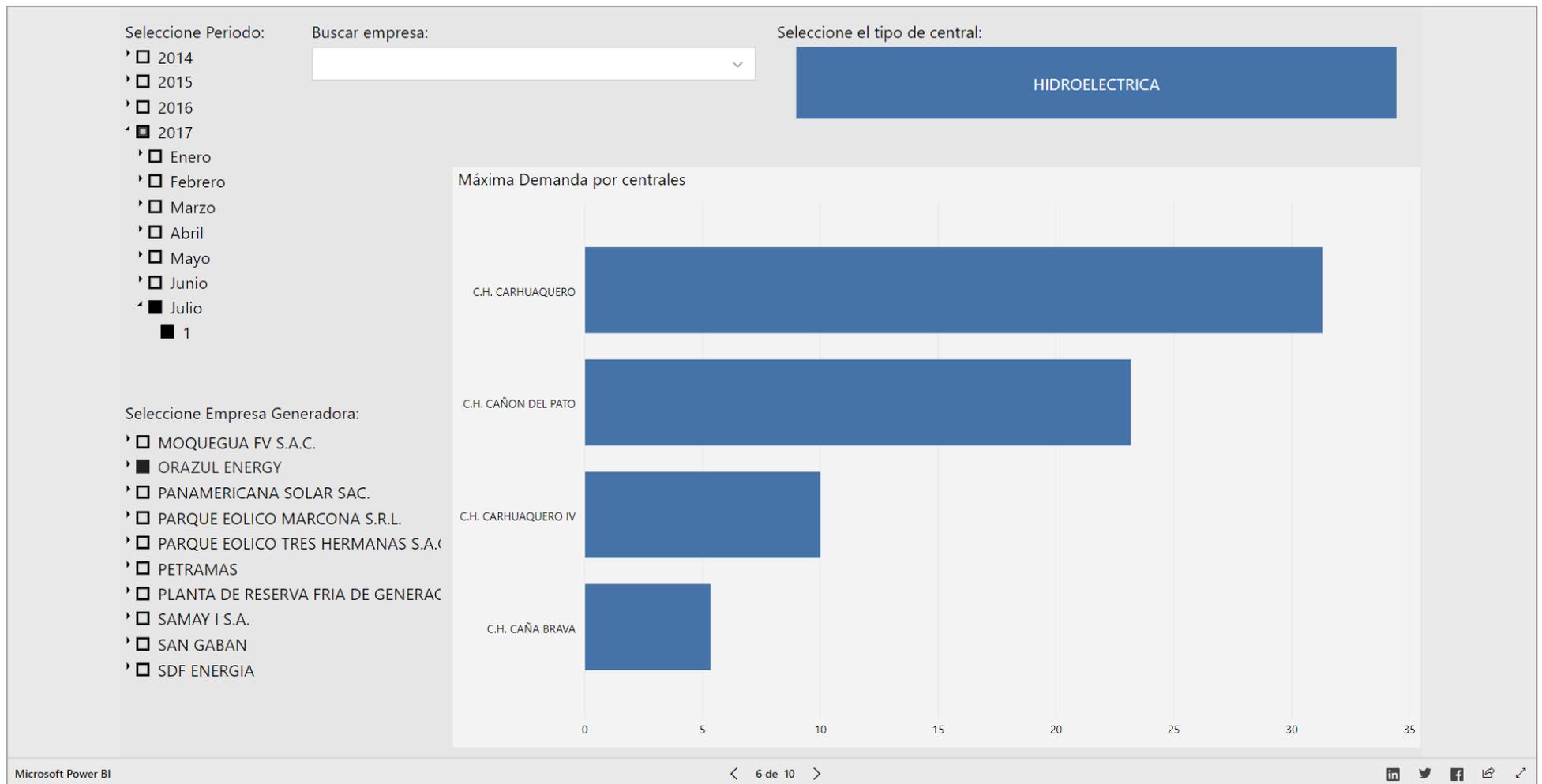
Reporte de la demanda ejecutada por actividad económica para las empresas en el Extranet
Fuente: Elaboración propia



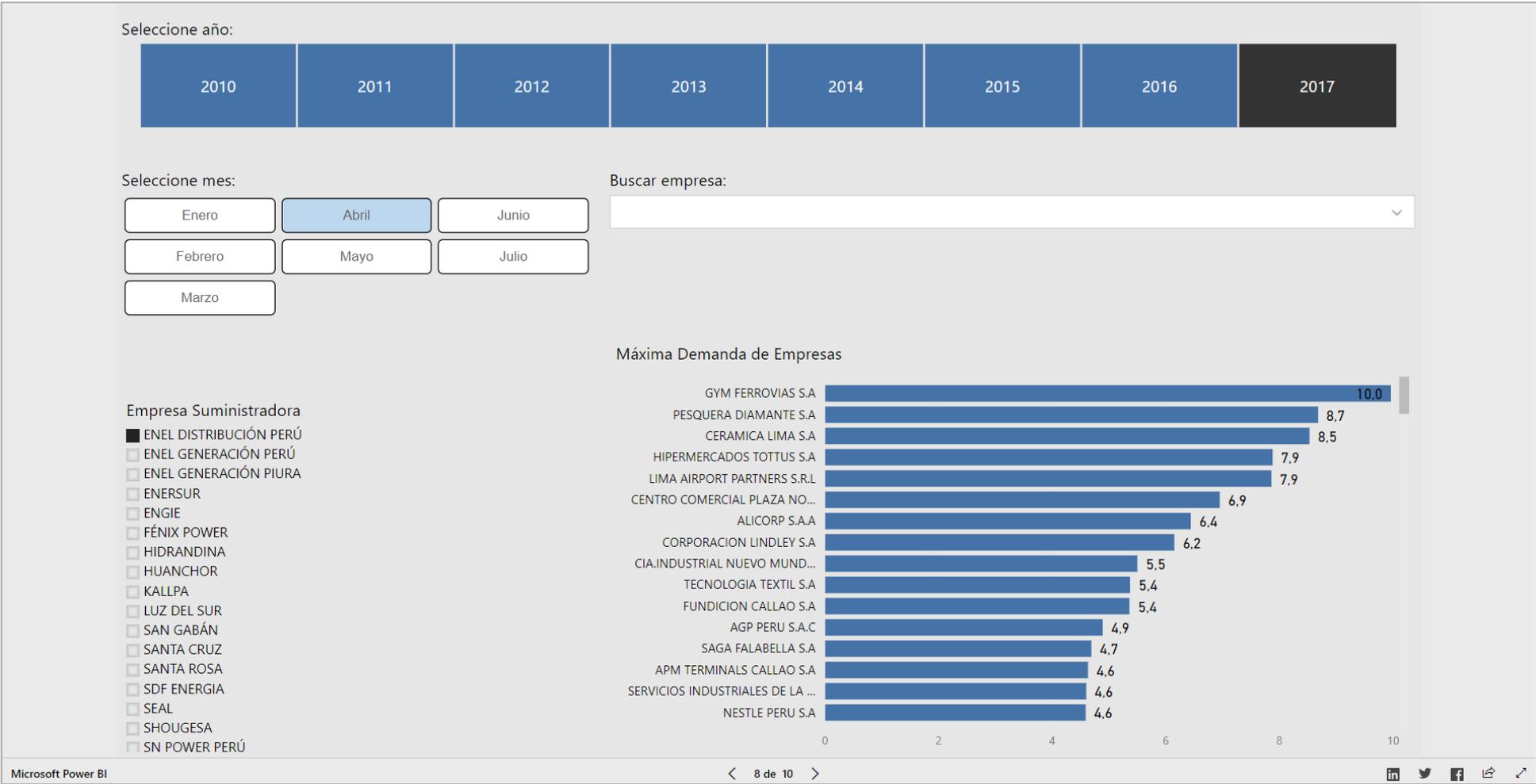
Reporte de la demanda ejecutada obtenida por señales SCADA frente a la demanda programada
Fuente: Elaboración propia



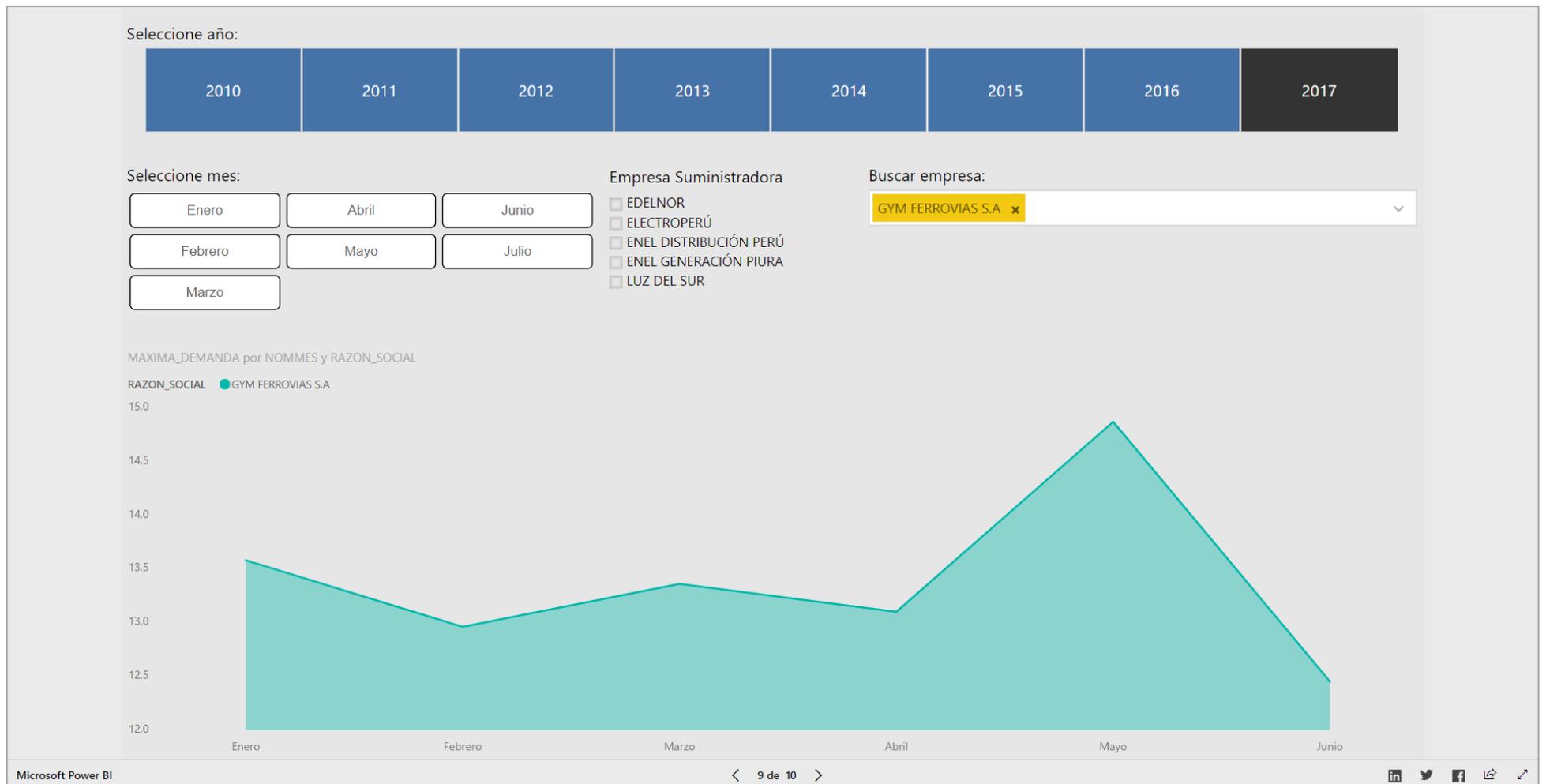
Reporte de la máxima demanda
Fuente: Elaboración propia



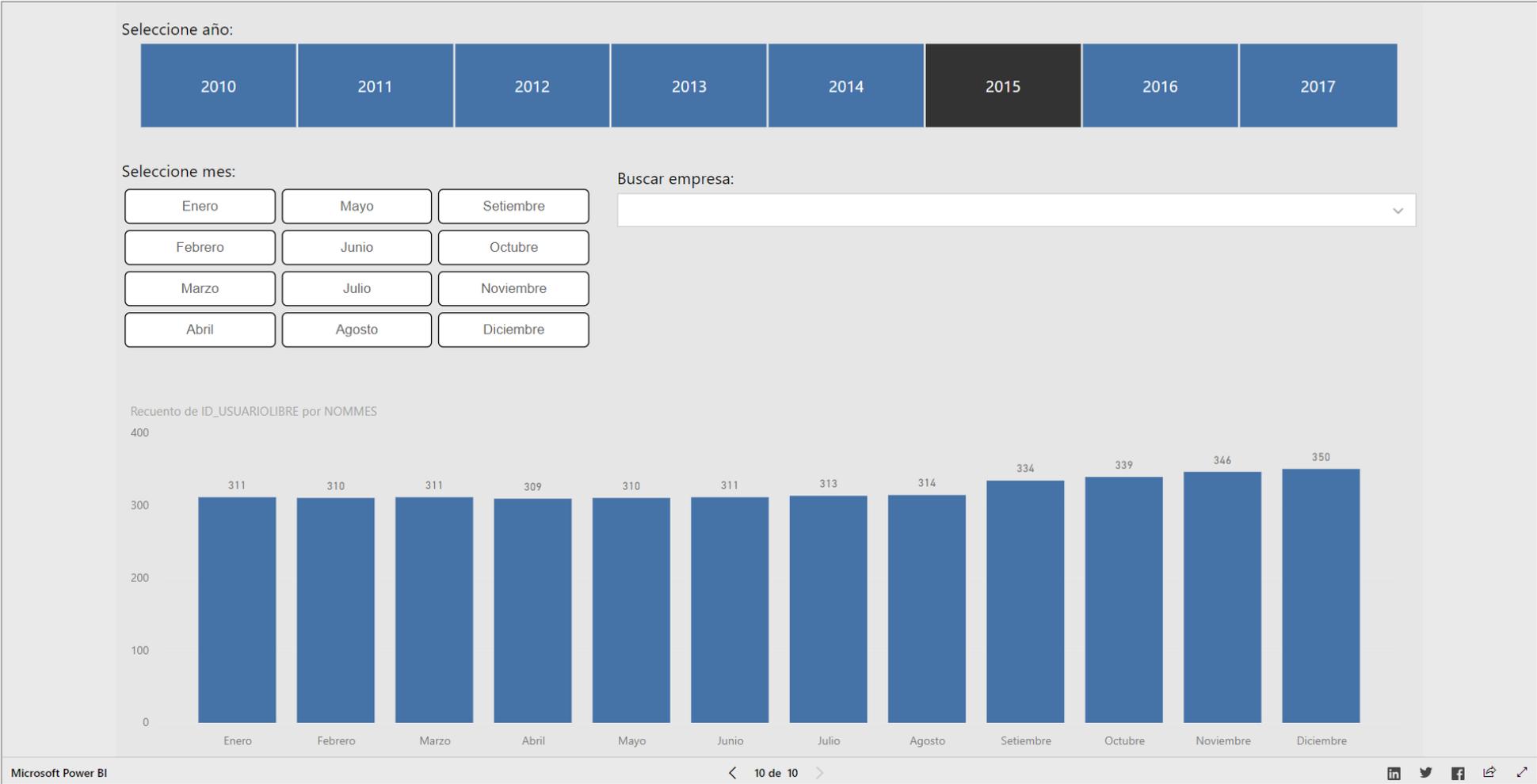
Reporte de las potencias registradas en los medidores de generación por centrales
Fuente: Elaboración propia



Reporte de máxima demanda consumida por Usuarios Libres según su suministrador
Fuente: Elaboración propia



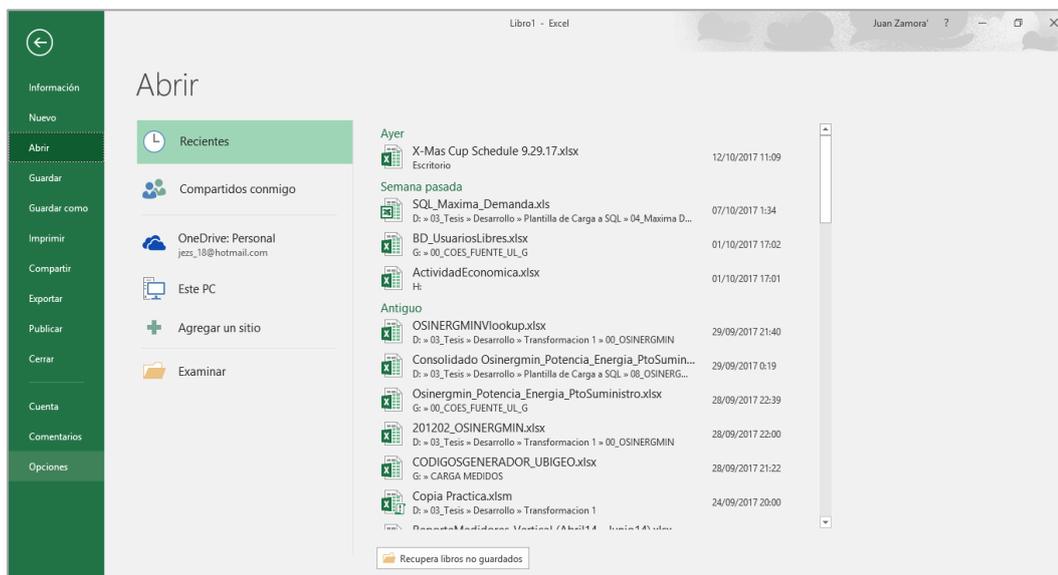
Reporte de máxima demanda consumida por Usuarios Libres según su suministrador
Fuente: Elaboración propia



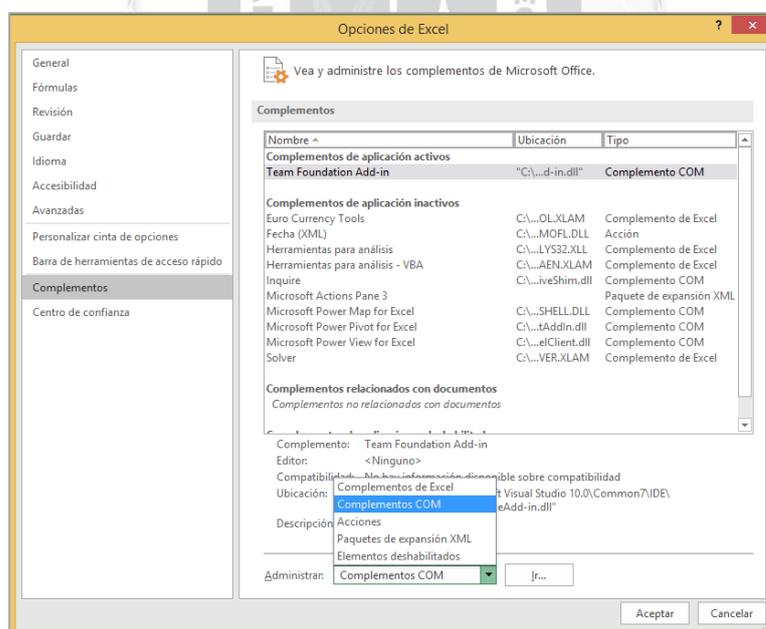
Reporte del crecimiento de puntos de suministro por empresa
Fuente: Elaboración propia

ANEXO XI: PROCESO DE INSTALACIÓN DEL COMPLEMENTO POWER PIVOT

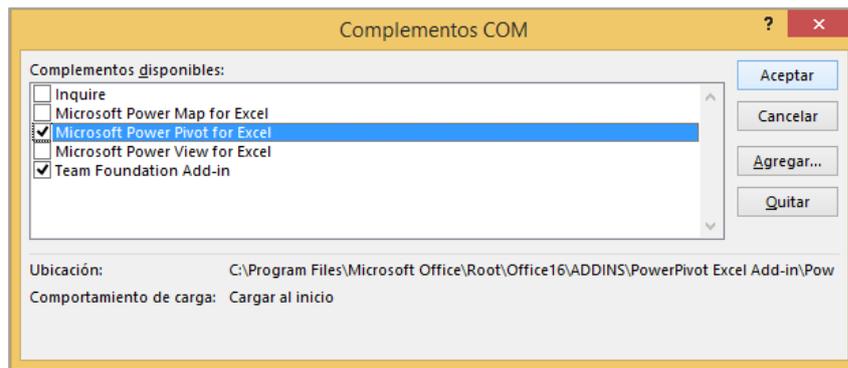
Para que la herramienta Excel en su versión 2013 y 2016 cuente con el complemento de Power Pivot, se debe realizar los pasos mostrados en las siguientes imágenes:



Proceso de instalación del complemento Power Pivot - Paso 1: Ingresar a opciones de Excel
Fuente: Elaboración propia



Proceso de instalación del complemento Power Pivot - Paso 2: Ingresar a los complementos de Excel
Fuente: Elaboración propia



Proceso de instalación del complemento Power Pivot - Paso 3: Activar el complemento de Power Pivot
Fuente: Elaboración propia

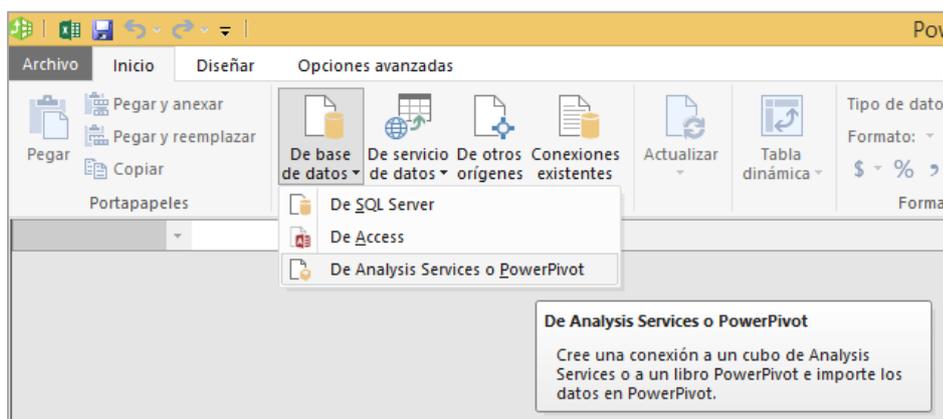


Proceso de instalación del complemento Power Pivot finalizado
Fuente: Elaboración propia

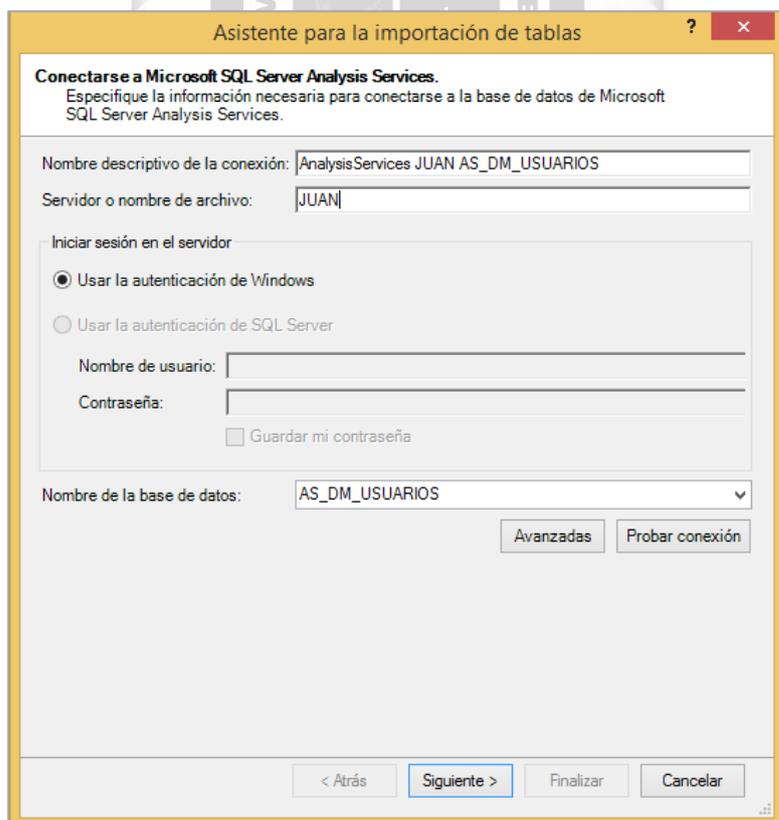


ANEXO XII: PROCESO DE CONEXIÓN AL SERVIDOR DE BASE DE DATOS CON LA HERRAMIENTA POWER PIVOT

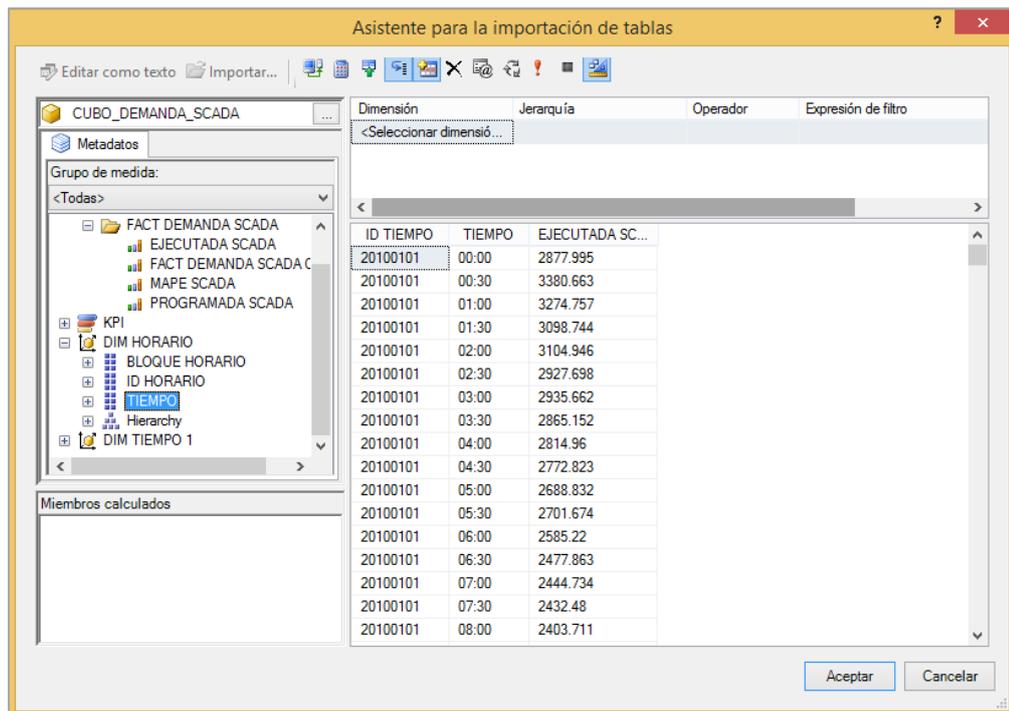
Mediante la herramienta Power Pivot, se puede realizar la conexión a distintas fuentes de datos; según los objetivos planteados en el presente trabajo de investigación, la herramienta logra conectarse al SQL Server directamente al modelo dimensional (datamart), como también al cubo OLAP que se encuentra en en el Analysis Services. Para realizar la conexión al Analysis Services, se realizan los pasos mostrados en las siguientes imágenes:



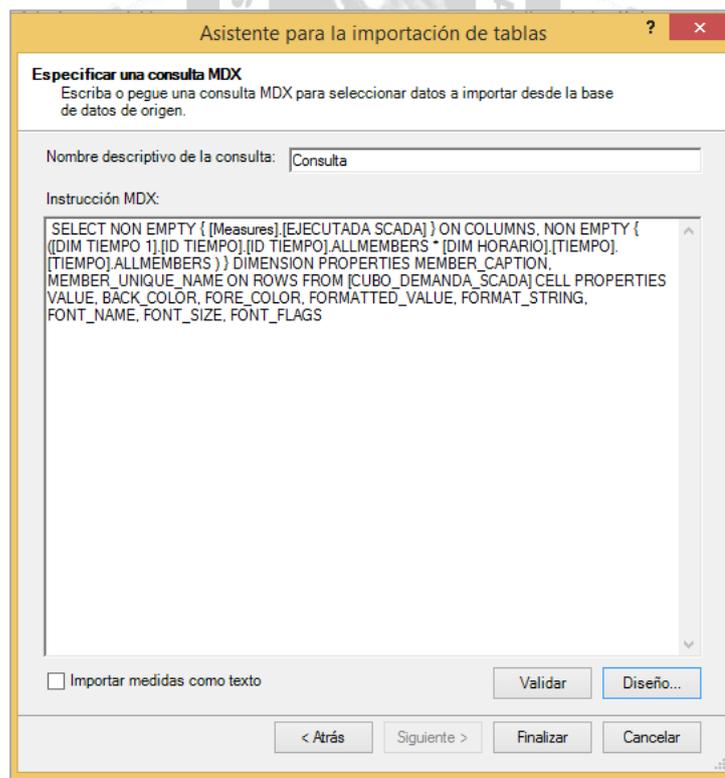
*Conexión desde Power Pivot al Analysis Services – Paso 1: Seleccionar fuente de datos
Fuente: Elaboración propia*



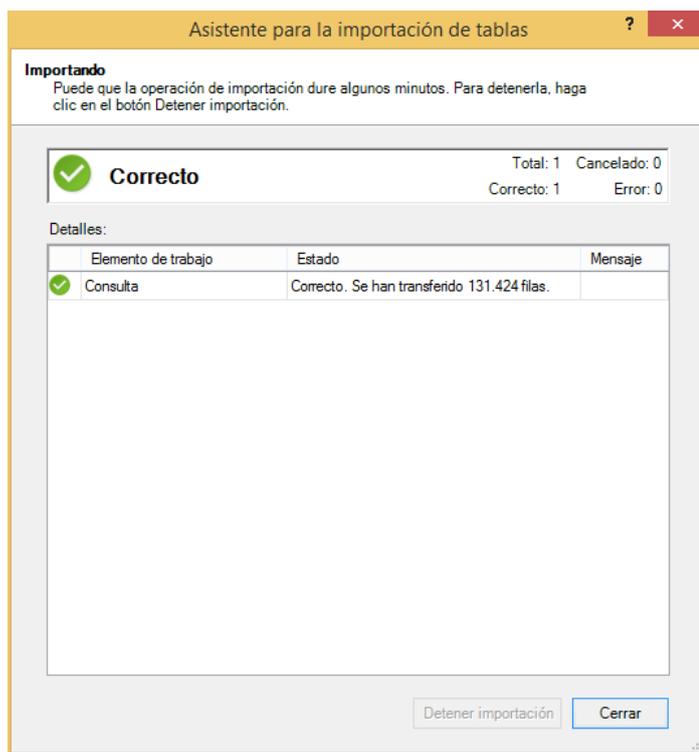
*Conexión desde Power Pivot al Analysis Services – Paso 2: Seleccionar servidor de base de datos
Fuente: Elaboración propia*



Conexión desde Power Pivot al Analysis Services – Paso 3: Diseñar la consulta a realizar mediante MDX
Fuente: Elaboración propia

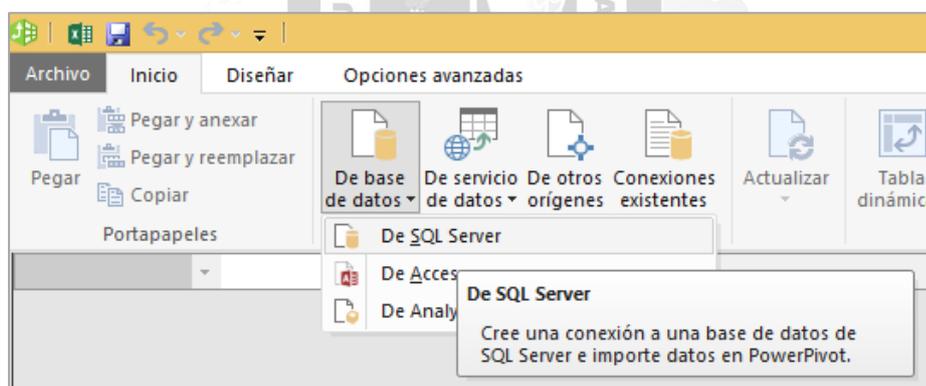


Conexión desde Power Pivot al Analysis Services – Paso 4: Consulta MDX generada
Fuente: Elaboración propia



Conexión desde Power Pivot al Analysis Services – Paso 5: Mensaje consulta realizada con éxito
Fuente: Elaboración propia

Así mismo, mediante Power Pivot el usuario puede obtener los datos del modelo dimensional (Datamart), cuyo proceso de conexión se observa en las siguientes imágenes:



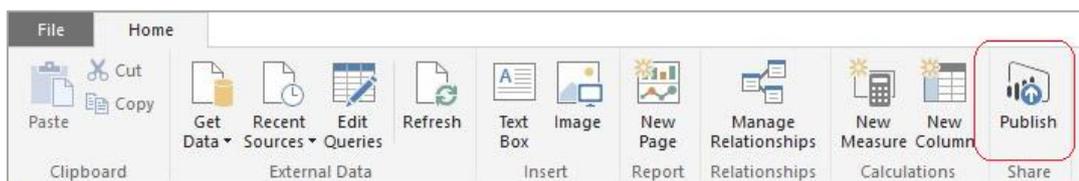
Conexión desde Power Pivot al modelo dimensional (datamart) – Paso 1: Seleccionar fuente de datos
Fuente: Elaboración propia

Conexión desde Power Pivot al modelo dimensional (datamart) – Paso 2: Seleccionar servidor de base de datos
Fuente: Elaboración propia

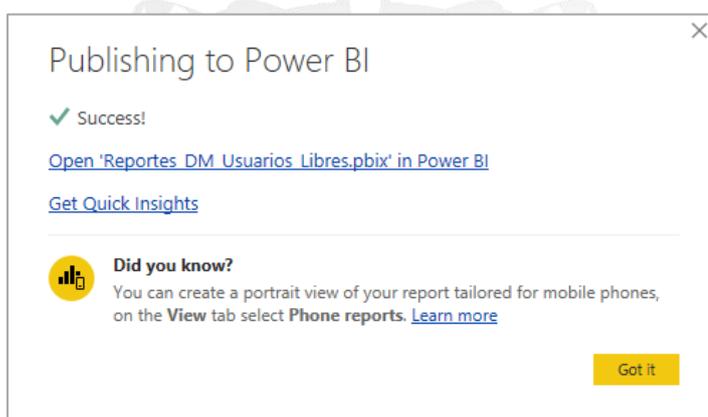
Conexión desde Power Pivot al modelo dimensional (datamart) – Paso 3: Seleccionar el tipo de importación
Fuente: Elaboración propia

ANEXO XIII: PROCESO DE PUBLICACIÓN DE REPORTES EN POWER BI

Una vez realizados los reportes gráficos en la herramienta Power BI Desktop, se procede a realizar la publicación de dichos reportes en el servidor de Power BI, con el fin de utilizar los distintos tipos acceso para visualizar los reportes generados. Para iniciar el proceso en mención, se debe realizar la publicación desde el Power BI Desktop, como se muestra en las siguientes imágenes:

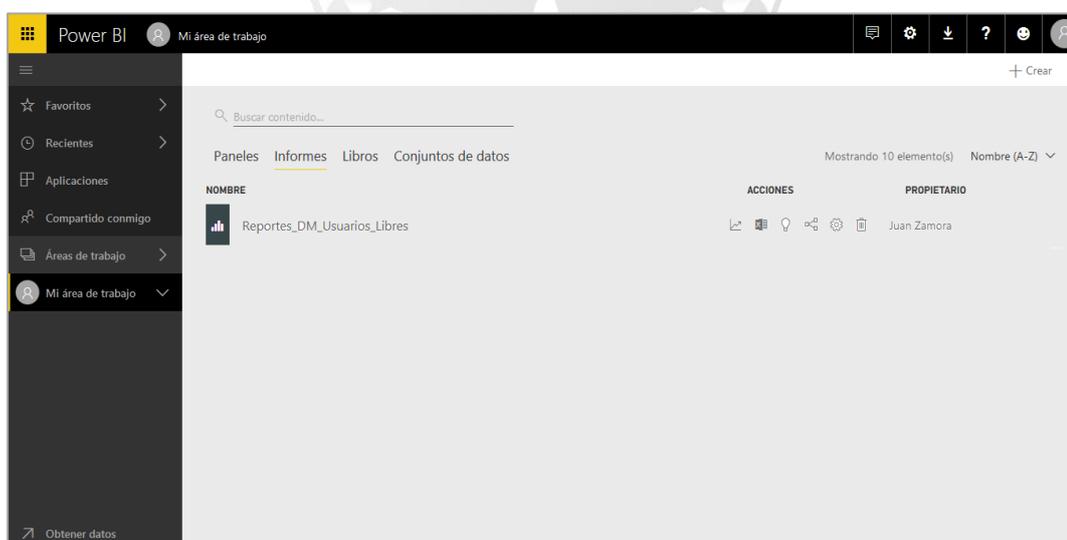


Barra de herramientas de Power BI Desktop
Fuente: Elaboración propia



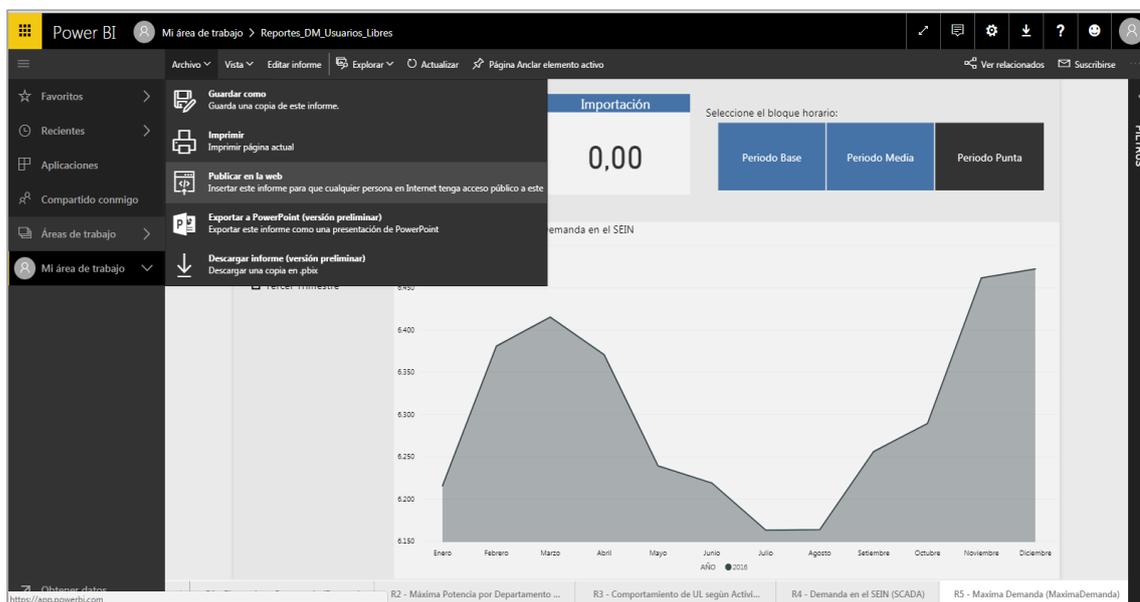
Mensaje de publicación exitosa desde Power BI Desktop
Fuente: Elaboración propia

Una vez finalizado el proceso de Power BI Desktop, el informe será cargado en el servidor de Power BI, en éste caso se ha utilizado el servidor web que la herramienta ofrece.



Informe publicado en Power BI (web)
Fuente: Elaboración propia

Una vez que el informe se encuentre publicado, Power BI permite generar un link para visualizar el reporte gráfico vía web, y generar una etiqueta HTML la cual pueda ser embebida en una página web. Para ello se debe seguir el proceso de publicar en web como muestran las siguientes imágenes:



Publicar en la web desde Power BI - Paso 1
Fuente: Elaboración propia

Operación completada correctamente. ✕

Vínculo que puede enviar por correo electrónico

<https://app.powerbi.com/view?r=eyJrljoiZWVhM2UzMDktOTkyOC00MzEw>

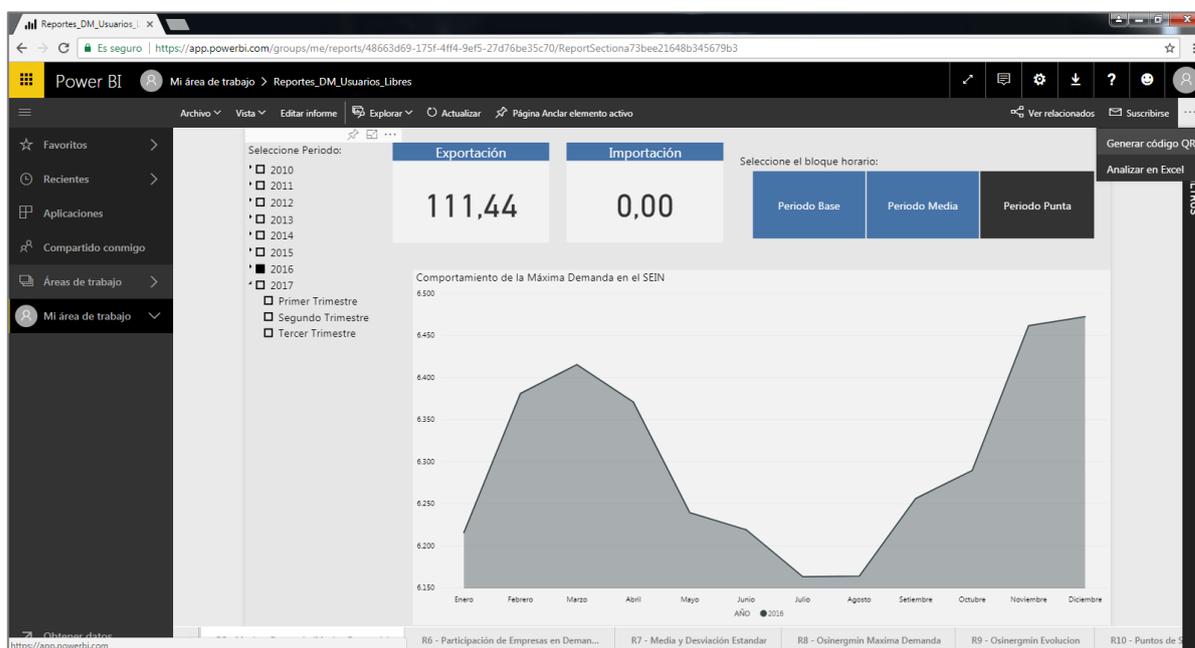
HTML que puede pegar en su blog o sitio web

```
<iframe width="933" height="700" src="https://app.powerbi.com/view?r:
```

Tamaño

Publicar en la web desde Power BI - Paso 2
Fuente: Elaboración propia

Así mismo, dichos reportes gráficos pueden ser visualizados desde un dispositivo móvil usando la aplicación de Power BI Mobile. Para ello se debe generar el código QR como se muestra en las siguientes imágenes:



Generar código QR - Paso 1
Fuente: Elaboración propia

Reportes_DM_Usuarios_Libres
✕



El código QR está listo

Ahora que ha creado el código QR, puede detectarlo con un dispositivo móvil y acceder al informe rápidamente. Para guardar la imagen del código QR de la izquierda, haga clic en el botón "Descargar".

ⓘ Sus compañeros únicamente podrán usar el código QR si disponen de los permisos pertinentes para consultar el informe.

Descargar
Listo

Código QR generado para ser visualizado en el Power BI Mobile
Fuente: Elaboración propia

ANEXO XIV: ENCUESTA DE EVALUACIÓN DE CALIDAD AL DATAMART DE DEMANDA ELÉCTRICA

Preguntas Generales

1. Seleccione que información emplea usted en su trabajo diario (Puede marcar más de una):
 - a) Demanda extranet declarada por Usuarios Libres
 - b) Información obtenida de OSINERGMIN
 - c) Demanda mediante medidores SCADA
 - d) Evaluación de la calidad del pronóstico de la demanda
 - e) Registros de la máxima demanda
 - f) Información de los medidores de generación
 - g) Información del Producto Bruto Interno

Según el indicador de accesibilidad

2. En su opinión ¿el nivel de accesibilidad permite reducir el tiempo que le generaba descargar los archivos para obtener la data histórica?
 - a) Muy de acuerdo
 - b) De acuerdo
 - c) Ni de acuerdo ni en desacuerdo
 - d) Desacuerdo
 - e) Muy en desacuerdo
3. Desde su punto de vista ¿tener acceso a la información del datamart mediante las tecnologías implementadas, permite mejorar las consultas sobre información histórica de años anteriores?
 - a) Muy de acuerdo
 - b) De acuerdo
 - c) Ni de acuerdo ni en desacuerdo
 - d) Desacuerdo
 - e) Muy en desacuerdo

Según el indicador de fiabilidad

4. Según su opinión, ¿considera que los reportes de comportamiento de la demanda eléctrica funcionan en un óptimo estado bajo cualquier circunstancia?
 - a) Muy de acuerdo
 - b) De acuerdo
 - c) Ni de acuerdo ni en desacuerdo
 - d) Desacuerdo
 - e) Muy en desacuerdo
5. A su parecer, ¿La información que se presenta en los reportes gráficos en relación con el datamart tiene información fiable para el análisis de la demanda eléctrica?
 - f) Muy de acuerdo
 - g) De acuerdo
 - h) Ni de acuerdo ni en desacuerdo
 - i) Desacuerdo
 - j) Muy en desacuerdo

Según el indicador de eficiencia

6. Según su opinión, ¿considera que el uso de un datamart para consultar grandes volúmenes de datos es más eficiente con la tecnología actual (datamart) que mediante el uso de hojas de cálculo?
 - a) Muy de acuerdo
 - b) De acuerdo
 - c) Ni de acuerdo ni en desacuerdo
 - d) Desacuerdo
 - e) Muy en desacuerdo
7. ¿El nivel de eficiencia ha mejorado en la disminución de tiempos que le tomaba para preparar la información en la estructura requerida para poder ser analizada?
 - a) Muy de acuerdo
 - b) De acuerdo
 - c) Ni de acuerdo ni en desacuerdo
 - d) Desacuerdo
 - e) Muy en desacuerdo
8. ¿Cree usted que el datamart permite reducir el tiempo de horas hombre que toma en preparar la información para generar reportes o analizar los datos históricos?

- a) Muy de acuerdo
- b) De acuerdo
- c) Ni de acuerdo ni en desacuerdo
- d) Desacuerdo
- e) Muy en desacuerdo

Según el indicador de disponibilidad

9. Para usted ¿la disponibilidad de información histórica que ofrece el datamart permite satisfacer las necesidades en cuanto a la cantidad de datos para el análisis de demanda eléctrica?
- a) Muy de acuerdo
 - b) De acuerdo
 - c) Ni de acuerdo ni en desacuerdo
 - d) Desacuerdo
 - e) Muy en desacuerdo
10. Desde su perspectiva ¿el datamart y la solución tecnológica implementada permite un mayor nivel de disponibilidad en comparación a otros métodos utilizados?
- a) Muy de acuerdo
 - b) De acuerdo
 - c) Ni de acuerdo ni en desacuerdo
 - d) Desacuerdo
 - e) Muy en desacuerdo
11. Bajo su percepción ¿hubo momentos en el que la información del datamart no se encontró disponible?
- a) Muy de acuerdo
 - b) De acuerdo
 - c) Ni de acuerdo ni en desacuerdo
 - d) Desacuerdo
 - e) Muy en desacuerdo

