**BUSINESS INTELLIGENCE PARA ANALIZAR**

**LA DEMANDA DE ENERGÍA ELÉCTRICA**

**POR:**

**HUGO DE JESÚS GONZÁLEZ OLAYA**



**BUSINESS INTELLIGENCE PARA ANALIZAR**

**LA DEMANDA DE ENERGÍA ELÉCTRICA**

**POR:**

**HUGO GONZÁLEZ OLAYA**



**TRABAJO REALIZADO COMO REQUISITO PARCIAL PARA**

**OPTAR AL TÍTULO DE INGENIERO DE SISTEMAS**

**ASESOR:**

**ROBERTO CARLOS GUEVARA CALUME**

**CORPORACIÓN UNIVERSITARIA REMINGTON**

**FACULTAD DE INGENIERÍAS**

**MEDELLÍN - ANTIOQUIA**

**2020**

**NOTA DE ACEPTACIÓN**

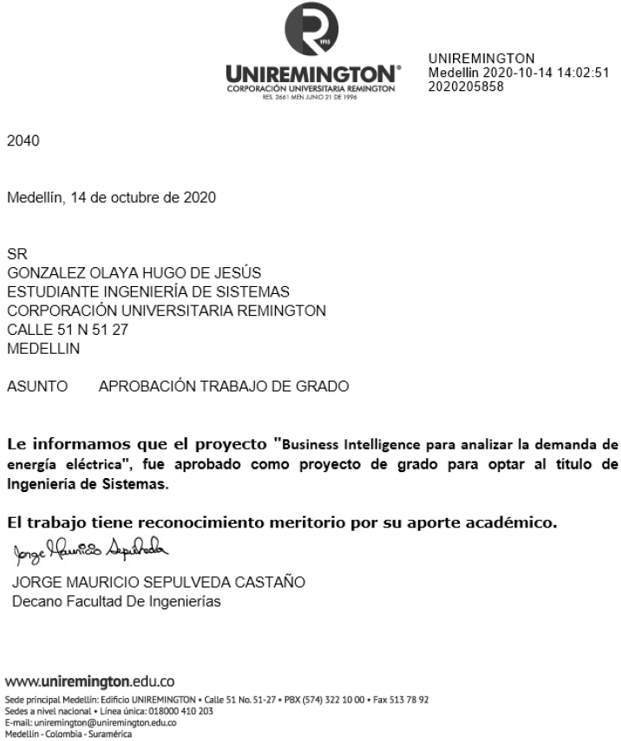
Conceptúo que el presente trabajo de grado ha cumplido con todas las normas e indicaciones por el asesor, así como también ha tenido en cuenta la metodología que la Corporación Universitaria Remington de Colombia exige. Por lo tanto, puede continuar con los subsiguientes pasos para su valoración, revisión y sustentación.

Piedad María Metaute Paniagua

Jurado

Edison Valencia Díaz

Jurado



**AGRADECIMIENTOS**

Muchas gracias a la empresa XM, y en su nombre a la arquitecta de datos Sandra Inés Betancur Muñoz, quien facilitó los datos para este proyecto.

Muchas gracias a la universidad Uniremington por la formación brindada para ser un Ingeniero de Sistemas.

Agradecimientos para mi asesor en este proyecto de grado Roberto Carlos Guevara Calume, sus recomendaciones fueron muy valiosas en la creación de este trabajo. Muchas gracias “Calume”.

Muchas gracias a Dios y a mi familia que me dan la fuerza y me animaron para continuar estudiando. En el año 1999 obtuve el título como Ingeniero Químico y ahora en el año 2020 a los 52 años de vida, estoy realizando este proyecto de grado para obtener un segundo título como Ingeniero de Sistemas. Gracias por su motivación para continuar avanzando.

**DEDICATORIA**

Dedico este trabajo a mi esposa Sandra Betancur Aguirre y a mis dos hijos Cristian González Betancur y Katerin González Betancur, que este esfuerzo sea de inspiración para que ellos también tengan un título profesional y que aprendan que la forma de progresar es estudiando y dedicando todo el empeño en estar dentro de los mejores.

Hugo de Jesús González Olaya

Qué este trabajo sea de gran utilidad para las personas que inician en el conocimiento de Business Inteligence, éxitos y disfruten el código incluido en el proyecto.

**DECLARACIÓN SOBRE DERECHOS DE AUTOR**

Las ideas expuestas en el presente trabajo de grado son de responsabilidad del autor. Los derechos de la información desarrollada durante el proceso de investigación pertenecen a la Corporación Universitaria Remington de Colombia y al creador de este trabajo. El material y código desarrollado en este trabajo es de uso libre. El autor no se hace responsable por su uso.

**AUTOR**

Hugo de Jesús González Olaya, Ingeniero Químico, MCSE: Microsoft certified solutions expert business intelligence, estudiante de Ingeniería de Sistemas en la universidad Uniremington.

[hugo14.gonzalez@gmail.com](mailto:hugo14.gonzalez@gmail.com)

[ing\_hgonzalez@hotmail.com](mailto:ing_hgonzalez@hotmail.com)

**CÓDIGO**

El código puede ser descargado del siguiente link. Solo fue subido una muestra de datos, puesto que el volumen de datos de demanda y pérdida de energía eléctrica de la empresa XM es alrededor de 2,000 millones de filas en 20 años. La muestra son unos 1,373 archivos con un tamaño de 300 MB, esta muestra contiene una fracción de los datos.

<https://github.com/hugo14gonzalez/BI_Demanda>

**CONTENIDO**

[1 INTRODUCCIÓN 1](#_Toc44421274)

[2 PLANTEAMIENTO DEL PROBLEMA 2](#_Toc44421275)

[2.1 ANTECEDENTES DEL PROBLEMA 2](#_Toc44421276)

[2.1.1 Agentes del mercado de energía eléctrica 2](#_Toc44421277)

[2.1.2 Empresa de apoyo para el proyecto 2](#_Toc44421278)

[2.1.3 Soluciones de BI 3](#_Toc44421279)

[2.2 JUSTIFICACIÓN 4](#_Toc44421280)

[3 OBJETIVOS 6](#_Toc44421281)

[3.1 OBJETIVO GENERAL 6](#_Toc44421282)

[3.2 OBJETIVOS ESPECÍFICOS 6](#_Toc44421283)

[4 MARCO TEÓRICO 7](#_Toc44421284)

[4.1 Cadena productiva DE energía eléctrica 7](#_Toc44421285)

[4.1.1 Generación de energía eléctrica 7](#_Toc44421286)

[4.1.2 Transmisión y distribución de energía eléctrica 8](#_Toc44421287)

[4.1.3 Comercialización de energía eléctrica 8](#_Toc44421288)

[4.1.4 Operadores de la red de energía eléctrica 8](#_Toc44421289)

[4.1.5 Administración del mercado de energía mayorista 8](#_Toc44421290)

[4.1.6 Sistema interconectado nacional – SIN 9](#_Toc44421291)

[4.2 DEMANDA DE ENERGÍA 9](#_Toc44421292)

[4.3 PÉRDIDAS DE ENERGÍA 9](#_Toc44421293)

[4.4 SOLUCIONES DE BUSINESS INTELLIGENCE 9](#_Toc44421294)

[4.5 Ciclo de vida de proyectos BI 10](#_Toc44421295)

[4.6 METODOLOGÍA ÁGIL PARA SOLUCIONES DE BI 12](#_Toc44421296)

[4.6.1 Enfoque tradicional de planificación en BI/DW 12](#_Toc44421297)

[4.6.2 Recomendaciones para proyectos ágiles BI/DW 13](#_Toc44421298)

[4.7 METODOLOGÍA KIMBALL 14](#_Toc44421299)

[4.7.1 Hechos y dimensiones 14](#_Toc44421300)

[4.7.2 Modelo estrella y copo de nieve 15](#_Toc44421301)

[4.7.3 Integración de datos con ETLs 15](#_Toc44421302)

[4.7.4 Data warehouse 16](#_Toc44421303)

[4.7.5 Cubos OLAP 16](#_Toc44421304)

[4.7.6 Herramientas clientes 16](#_Toc44421305)

[4.8 INTEGRACIÓN DE DATOS CON ETL 16](#_Toc44421306)

[4.8.1 ¿Qué es una ETL? 16](#_Toc44421307)

[4.8.2 Evolución de soluciones de ETL 17](#_Toc44421308)

[4.8.3 Bloques para construir ETLs 17](#_Toc44421309)

[4.8.4 Arquitectura clasica de sistemas BI 17](#_Toc44421310)

[5 MARCO METODOLÓGICO 19](#_Toc44421311)

[5.1 PLAN DE PROYECTO 19](#_Toc44421312)

[5.2 ALCANCE DEL PROYECTO 20](#_Toc44421313)

[5.3 METODOLOGÍA PARA LA CONSTRUCCIÓN DE LA SOLUCIÓN DE BI 20](#_Toc44421314)

[5.4 FASE DE CONSTRUCCIÓN 21](#_Toc44421315)

[5.5 SELECCIÓN DE HERRAMIENTAS DE DESARROLLO 22](#_Toc44421316)

[6 REQUISITOS DEL SISTEMA 25](#_Toc44421317)

[6.1 PREGUNTAS DEL NEGOCIO 25](#_Toc44421318)

[6.2 DISEÑO GENERAL DEL SISTEMA 26](#_Toc44421319)

[6.3 REQUISITOS 27](#_Toc44421320)

[6.4 Fuentes de datos 27](#_Toc44421321)

[6.5 DISEÑOS PARA LA SOLUCIÓN 28](#_Toc44421322)

[6.6 DICCIONARIO DE DATOS 28](#_Toc44421323)

[6.7 DIAGRAMAS DE casos de uso Y DE SECUENCIAS 29](#_Toc44421324)

[6.8 DESEMPEÑO DEL SISTEMA 31](#_Toc44421325)

[7 ARQUITECTURA DEL SISTEMA 32](#_Toc44421326)

[7.1 Herramientas de desarrollo 32](#_Toc44421327)

[7.2 MINERÍA DE DATOS Y MACHING LEARNING 32](#_Toc44421328)

[7.3 ARQUITECTURA CONCEPTUAL 32](#_Toc44421329)

[7.3.1 Fuentes de datos 33](#_Toc44421330)

[7.3.2 Modelo de datos 33](#_Toc44421331)

[7.3.3 Capa intermedia 34](#_Toc44421332)

[7.3.4 Capa de presentación 34](#_Toc44421333)

[7.3.5 Servicios transversales 34](#_Toc44421334)

[7.4 Casos de uso relevantes para la arquitectura 35](#_Toc44421335)

[7.4.1 Carga de datos 36](#_Toc44421336)

[7.5 ARQUITECTURA DE DESPLIEGUE 36](#_Toc44421337)

[8 DATA WAREHOUSE 38](#_Toc44421338)

[8.1 Scripts de base de datos 38](#_Toc44421339)

[8.2 Instalación de base de datos 39](#_Toc44421340)

[8.3 Esquemas de base de datos 39](#_Toc44421341)

[8.4 Modelo estrella 40](#_Toc44421342)

[8.5 Desempeño de base de datos 40](#_Toc44421343)

[8.6 MODELADO DE TABLAS DE DIMENSIONES 41](#_Toc44421344)

[9 BITÁCORA DE PROCESOS 43](#_Toc44421345)

[9.1 Logs de carga de datos 43](#_Toc44421346)

[9.1.1 Auditoria de ETLs 43](#_Toc44421347)

[9.1.2 Auditoría de Jobs 44](#_Toc44421348)

[9.1.3 Reporte de ejecución de ETLs 44](#_Toc44421349)

[9.1.4 Reportes de bitácora 45](#_Toc44421350)

[9.2 Logs SQL Server 45](#_Toc44421351)

[10 ETL PARA GENERAR ARCHIVOS PLANOS 46](#_Toc44421352)

[11 ETL PARA CARGA DEL DATA WAREHOUSE 48](#_Toc44421353)

[11.1 ETL DEMANDA BI 48](#_Toc44421354)

[11.2 INSTALACIÓN Y CONFIGURACIÓN DE ETLS 53](#_Toc44421355)

[12 CUBOS OLAP 54](#_Toc44421356)

[12.1 CUBOS DEMANDA BI 54](#_Toc44421357)

[12.2 DISEÑO DE JERARQUÍAS PARA DIMENSIONES 55](#_Toc44421358)

[12.3 CUBO DEMANDA 56](#_Toc44421359)

[12.4 NOMBRAMIENTOS DE OBJETOS 56](#_Toc44421360)

[12.5 DESEMPEÑO EN CUBOS olap 57](#_Toc44421361)

[12.5.1 Particiones en cubos 57](#_Toc44421362)

[12.5.2 Agregados en cubos 59](#_Toc44421363)

[12.5.3 Modo de almacenamiento 59](#_Toc44421364)

[12.6 INSTALACIÓN Y CONFIGURACIÓN DE CUBOS 59](#_Toc44421365)

[13 HERRAMIENTA CLIENTE DE REPORTES 60](#_Toc44421366)

[13.1 HERRAMIENTAS DE REPORTES 60](#_Toc44421367)

[13.2 OTRAS HERRAMIENTAS PARA REPORTES 61](#_Toc44421368)

[13.3 REPORTE TIPO DASHBOARD 61](#_Toc44421369)

[13.4 REPORTES GRÁFICOS 62](#_Toc44421370)

[13.5 REPORTES PÁGINADOS 65](#_Toc44421371)

[13.6 ¿CÓMO CONSULTAR LOS REPORTES? 66](#_Toc44421372)

[13.7 DISEÑO DE REPORTES 67](#_Toc44421373)

[13.8 INSTALACIÓN Y CONFIGURACIÓN DE REPORTES 67](#_Toc44421374)

[14 HERRAMIENTA CLIENTE APLICACIÓN WEB 68](#_Toc44421375)

[14.1 HERRAMIENTA DE DESARROLLO 68](#_Toc44421376)

[14.2 MOSTRAR LOS REPORTES EN LA APLICACIÓN WEB 68](#_Toc44421377)

[14.3 MENÚS DE LA APLICACIÓN 68](#_Toc44421378)

[14.4 CONFIGURACIÓN DE LA APLICACIÓN 69](#_Toc44421379)

[14.5 INSTALACIÓN DE LA APLICACIÓN 69](#_Toc44421380)

[15 RESULTADOS Y ENTREGABLES DEL PROYECTO 70](#_Toc44421381)

[15.1 CONSTRUCCIÓN DE UN BI 70](#_Toc44421382)

[15.2 modelo estrella 70](#_Toc44421383)

[15.3 SOLUCIÓN DE BI 70](#_Toc44421384)

[15.4 ANÁLISIS DE LA DEMANDA Y PÉRDIDAS DE ENERGÍA 71](#_Toc44421385)

[15.5 RESULTADOS PARA LOS AGENTES 71](#_Toc44421386)

[15.6 utilidad del sistema de bi para el negocio de energía 71](#_Toc44421387)

[15.7 APLICACIÓN WEB 72](#_Toc44421388)

[15.8 ENTREGABLES DEL PROYECTO 72](#_Toc44421389)

[16 CONCLUSIONES 76](#_Toc44421390)

[17 RECOMENDACIONES PARA TRABAJOS FUTUROS 78](#_Toc44421391)

[18 ANEXOS 79](#_Toc44421392)

[18.1 ANEXO 1: SOLICITUD DE DATOS A LA EMPRESA XM 79](#_Toc44421393)

[18.2 ANEXO 2: MANUAL DE INSTALACIÓN 79](#_Toc44421394)

[18.3 ANEXO 3: MANAUL DE OPERACIÓN 79](#_Toc44421395)

[18.4 ANEXO 4: HERRAMIENTAS DE SOFTWARE 79](#_Toc44421396)

[19 REFERENCIAS 80](#_Toc44421397)

**INDICE DE FIGURAS**

[Figura 1. Portal BI de la empresa XM 3](#_Toc44421398)

[Figura 2. Cadena productiva de energía eléctrica 7](#_Toc44421399)

[Figura 3. Modelos para construcción de BI creados por Kimball, Inmon y Lidstedt 13](#_Toc44421400)

[Figura 4. A la izquierda modelo estrella y a la derecha modelo copo de nieve 15](#_Toc44421401)

[Figura 5. Cuadrante mágico de Gartner para Analítica y BI 23](#_Toc44421402)

[Figura 6. Diseño general del sistema 26](#_Toc44421403)

[Figura 7. Diagrama general de casos de uso 30](#_Toc44421404)

[Figura 8. Diagrama de secuencia carga automática DW 30](#_Toc44421405)

[Figura 9. Diagrama de secuencia recarga manual DW 31](#_Toc44421406)

[Figura 10. Arquitectura conceptual DemandaBI 33](#_Toc44421407)

[Figura 11. Grupos funcionales DemandaBI 35](#_Toc44421408)

[Figura 12. Arquitectura de despliegue 37](#_Toc44421409)

[Figura 13. Modelo estrella demanda y pérdidas de energía 40](#_Toc44421410)

[Figura 14. Función de partición y esquemas de partición 41](#_Toc44421411)

[Figura 15. A la izquierda dimensión tipo 1, a la derecha dimensión tipo 2 42](#_Toc44421412)

[Figura 16. Reporte propio de SQL de ejecución de ETLs 44](#_Toc44421413)

[Figura 17. Reporte de bitácora del sistema 45](#_Toc44421414)

[Figura 18. Carpetas para la carga de datos 46](#_Toc44421415)

[Figura 19. ETL para generar archivos planos 47](#_Toc44421416)

[Figura 20. ETL para carga de tabla de dimensión 49](#_Toc44421417)

[Figura 21. ETL para carga de tabla fact de demanda y pérdidas 50](#_Toc44421418)

[Figura 22. ETL para procesar dimensiones de la base de datos multidimensional 51](#_Toc44421419)

[Figura 23. ETL para crear particiones para cubos 51](#_Toc44421420)

[Figura 24. ETL para procesar particiones de cubos 52](#_Toc44421421)

[Figura 25. ETL completa ejecuta todo para cargar el DW 52](#_Toc44421422)

[Figura 26. Base de datos multidimensional Demanda\_OLAP 55](#_Toc44421423)

[Figura 27. Jerarquías para la dimensión fecha 55](#_Toc44421424)

[Figura 28. Cubo demanda. A la izquierda métricas naturales. A la derecha relación del cubo con dimensiones 56](#_Toc44421425)

[Figura 29. Diseño de particiones para el cubo demanda 57](#_Toc44421426)

[Figura 30. Particiones implementadas para el cubo demanda 58](#_Toc44421427)

[Figura 31. Definición de consulta para crear particiones en cubos 58](#_Toc44421428)

[Figura 32. Reporte tipo dashboard con KPIs 62](#_Toc44421429)

[Figura 33. Reporte demanda real, comercial y pérdidas 63](#_Toc44421430)

[Figura 34. Reporte demanda comercial por agente 64](#_Toc44421431)

[Figura 35. Reporte demanda por geografía 65](#_Toc44421432)

[Figura 36. Reporte paginado de bitácora 66](#_Toc44421433)

[Figura 37. Servidor SQL Reporting Services 67](#_Toc44421434)

[Figura 38. Página de inicio de la aplicación DemandaWeb 69](#_Toc44421435)

**INDICE DE TABLAS**

[Tabla 1. Ciclo de vida proyectos BI 11](#_Toc44421436)

[Tabla 2. Flujo de pasos para el proyecto 19](#_Toc44421437)

[Tabla 3. Preguntas del negocio 25](#_Toc44421438)

[Tabla 4. Diccionarios de datos 29](#_Toc44421439)

[Tabla 5. Proyectos de base de datos de la solución SSMS 38](#_Toc44421440)

[Tabla 6. Esquemas de bases de datos 39](#_Toc44421441)

[Tabla 7. ETLs del proyecto 48](#_Toc44421442)

[Tabla 8. Artefactos del sistema 72](#_Toc44421443)

**GLOSARIO**

**Agente**: Representan las empresas que hacen parte del sistema eléctrico, los agentes pueden ser: generador, transmisor, distribuidor o comercializador.

**AS**: SQL Analysis Services, motor de base de datos multidimensional, donde están almacenados los cubos de datos.

**BI:** Bussiness Intelligence, inteligencia de negocios.

**Bitácora**: Registro de auditoría y de ejecución de procesos.

**Dashboard**: Tablero de mando o tablero de control, es un reporte gráfico con métricas del sistema o KPIs (Key Performace Indicators: Indicadores claves de desempeño).

**Data mart**: Un data mart es similar a un data warehouses, es una bodega de datos, la diferencia es que son temáticos, por ejemplo las ventas, compras, producción, etc.

**Data warehouse (DW)**: Bodega de datos, es una base de datos relacional, donde son consolidados los datos de la compañía en modelos (estrella o copo de nieve) para el análisis de la información.

**Demanda**: Necesidad de energía eléctrica para el consumo de los diferentes sectores: hogares, agrícola, industrial, comercio, etc.

**ETL**: Proceso de extracción, transformación y cargar, encargado de pasar datos de una fuente de datos hacia otra.

**Fact**: Tablas de hechos de un modelo estrella.

**OLTP**: Las bases de datos relacionales son sistemas (On-line transaction process) que realizan procesos transaccionales de: consulta, y modificación de datos.

**OLAP**: Las bases de datos multidimensionales o analítica son solo de consulta (On-line analytic process).

**Pérdidas**: Pérdidas de energía eléctrica por causas técnicas, robo o no pago.

**Submercado**: Lugar de intercambios energéticos entre agentes comercializadores, generadores y distribuidores de energía eléctrica.

**SSIS**: SQL Integration Services, herramienta para intercambio de información entre fuentes de datos. Realiza el proceso ETL para carga de datos.

**SSMS**: SQL Server Management Studio, herramienta para administración de bases de datos SQL Server.

**SSRS**: SQL Reporting Services, herramienta para construcción de reportes

**Staging Area (SA)**: Área de almacenamiento temporal, utilizada por las ETLs de carga de datos.

**XM**: “XM Compañía de expertos en mercados S.A. E.S.P.” Empresa coordinadora del sistema eléctrico en Colombia.

**RESUMEN**

La demanda y pérdidas de energía eléctrica son variables muy importantes para los agentes del mercado de energía eléctrica. Para enfrentar este problema fue creada una solución de Business intelligence (BI). La metodología fue basada en la “arquitectura bus” definida por Kimball, y fue construida una solución de software para analizar la información histórica de demanda y pérdidas de energía eléctrica.

El software construido es un data warehouse (DW) con un solo modelo estrella con las métricas de demanda y pérdidas de energía. Los datos para el sistema fueron facilitados por la empresa XM que coordina el mercado de energía eléctrica en Colombia. fueron creadas ETLs para cargar el DW y procesar los cubos OLAP. Fueron creados reportes tipo dashboar y gráficos y una aplicación web. También fue incluido bitácora para seguimiento a los procesos de carga y utilidades para la carga del sistema. Las herramientas seleccionadas fueron Microsoft SQL Server para la base de datos DW, cubos OLAP y reportes Reporting Services, y para la aplicación web Visual Studio .Net. También fueron empleados otros frameworsk como Bootstrap, JQuery y JavaScript. El proyecto incluye el código construido, manuales de instalación y operación.

El proyecto realiza un recorrido explicando la construcción de cada uno de los artefactos de la “arquitectura bus” para BI. El modelo estrella representa las “preguntas del negocio”, pero para que el proyecto de BI sea exitoso, es decir “que el usuario lo use” es porque le ofrece “lo que necesita”. El BI tiene desafíos como la calidad de datos y el desempeño, para enfrentar estos retos fue realizado un buen diseño de ETLs y cubos, e incluidos aspectos como bitácora para seguimiento a la carga de procesos, particiomiento de tablas y cubos, y utilidades para la carga diaria en forma automática y recarga de datos históricos.

**PALABRAS CLAVES**

BI, business intelligence, inteligencia de negocios, bodega de datos, ETL, demanda, energía eléctrica.

**ABSTRACT**

The demand and losses of electrical energy are very important variables for the agents of the electrical energy market. To deal with this problem, a Business intelligence (BI) solution was created. The methodology was based on the “bus architecture” defined by Kimball, and a software solution was built to analyze the historical information on demand and losses of electrical energy.

The built software is a data warehouse (DW) with a single star model with the metrics of demand and energy losses. The data for the system was provided by the XM company that coordinates the electricity market in Colombia. ETLs were created to load the DW and process the OLAP cubes. Dashboar reports and graphs and a web application were created. A bitacora was also included to track the loading processes and utilities for loading the system. The selected tools were Microsoft SQL Server for the DW database, OLAP cubes and Reporting Services reports, and for the web application Visual Studio .Net. Other frameworsk as Bootstrap, JQuery and JavaScript were also used. The project includes the built code, installation and operation manuals.

The project takes a tour explaining the construction of each of the artifacts of the "bus architecture" for BI. The star model represents the “business questions”, but for the BI project to be successful, that is, “for the user to use” it is because it offers what they need. The BI has challenges such as data quality and performance, in order to face these challenges, a good design of ETLs and cubes was carried out, and included aspects as bitacora for monitoring process loading, participation of tables and cubes, and utilities for the daily automatic loading and reloading of historical data.

**KEY WORDS**

Business intelligence, data warehouse, ETL, demand, electric energy.

# INTRODUCCIÓN

Para los agentes del mercado de energía eléctrico la demanda y pérdidas de energía son variables muy importantes. En este proyecto fue construido una solución de Business intelligence (BI) para resolver la necesidad de los agentes de contar un información confiable y rápida para la toma de decisiones.

El proyecto BI está formado por una bodega de datos (DW: Data warehouse), ETLs para carga de datos, cubos OLAP, reportes y una aplicación Web. Las herramientas utilizadas son SQL Server para el back-end y Visual Studio .Net para el front-end.

Los datos para alimentar el sistema fueron suministrados por la empresa: “XM Compañía de expertos en mercados S.A. E.S.P.”, la cual coordina el mercado de energía eléctrica en Colombia, XM tiene un BI de acceso público, con variables e indicadores del mercado eléctrico.

BI es una tecnología para analizar información y tomar decisiones, cada día es más utilizada, pero existe poco personal calificado. El proyecto utiliza la metodología Kimball para BI, y explicación cada elemento de la metodología de desarrollo. El proyecto fue creado para resolver el problema de los agentes contar con herramientas para analizar y tomar decisiones, y como un aporte en el conocimiento de BI.

# PLANTEAMIENTO DEL PROBLEMA

## ANTECEDENTES DEL PROBLEMA

El mercado de energía eléctrica en Colombia es administrado por la empresa XM, los agentes son empresas registradas que intervienen en el mercado de energía mayorista que realizan actividad de: comercialización, generación, transmisión, distribución y operación de la red de energía eléctrica (Memorias seminarios 2018, XM).

### Agentes del mercado de energía eléctrica

El conocimiento de la demanda de energía es importante para diferentes tipos de agentes del sistema eléctrico, por ejemplo, los comercializadores pagan por anticipado la energía que van a vender a los consumidores finales, para ello consignan dinero denominado “garantías” al administrador del sistema eléctrico (en Colombia el mercado eléctrico es coordinado por la empresa XM). Los generadores tienen los recursos para generar según la demanda. El administrador del sistema (XM) utiliza la información de demanda, para realizar la programación del despacho de energía. Para estos agentes el conocimiento de la demanda y su pronóstico (diario, semanal, mensual y anual) es muy importante para la toma de decisiones (Memorias seminarios 2018, XM).

Los diferentes agentes del mercado envían datos a XM, que es responsable de centralizar y entregar información oportuna sobre las diferentes variables del sistema eléctrico como: la demanda y pérdidas de energía eléctrica, para ello cuenta con una bodega de datos (DW – Data Warehouse) y un portal Web, el cual es mostrado en la Figura 1 (Información suministrada por expertos del mercado de energía de XM).

### Empresa de apoyo para el proyecto

La empresa “XM Compañía de expertos en mercados S.A. E.S.P.”, cuenta con los datos de demanda y pérdidas de energía para el proyecto, algunas de sus fuentes de datos son públicas. Pero para las necesidades del proyecto fueron extraídos en forma directa de su base de datos. Para obtener los datos fueron creadas consultas de selección de datos y fue solicitado autorización a XM, quién facilitó la extracción de datos, en los anexos incluimos la autorización.

La Figura 1 muestra el portal BI de la empresa XM, en el cual pueden ser consultadas las variables mencionadas:

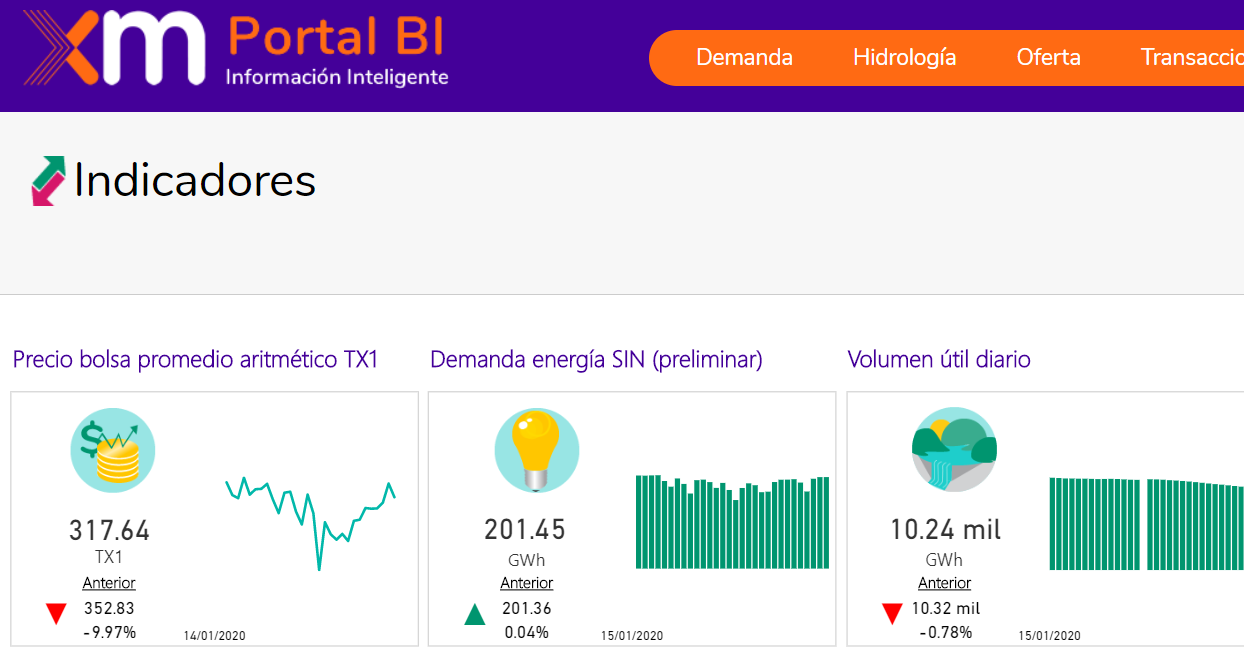


Figura 1. Portal BI de la empresa XM

Fuente: <http://portalbissrs.xm.com.co>

### Soluciones de BI

La tecnología de estudio es denominada BI. Los BI son soluciones de mucho movimiento, todos los días o con una periodicidad establecida, carga el sistema con los datos transaccionales de los agentes. Esto hace que la solución sea de mucho volumen, lo que implica desafíos como el desempeño y la necesidad de sistemas robustos que entreguen la información en forma oportuna.

Otro de los desafíos que enfrenta el BI es la calidad de los datos, “los datos en el sistema DW / BI deben ser creíbles. Los datos deben reunirse cuidadosamente a partir de una variedad de fuentes, limpiarse, garantizarse la calidad y divulgarse solo cuando sean aptos para el consumo del usuario (Kimball y Ross, 2013, p.3). Otros autores también afirman sobre el problema de calidad de datos Ferrai & Russo (2008): “Normalmente en sistemas OLTP, existen datos históricos que no son correctos. Esta es una regla, no una excepción. Un sistema que se ejecuta desde hace años a menudo tiene datos que no son correctos y nunca lo serán” (p. 9). Para el BI de XM es necesario mejorar los procesos de auditoría y resolver errores por calidad de datos, cargas de datos incompletas y datos inconsistentes. Este es uno de los retos para este proyecto, la creación de sistemas de seguimiento y tratamiento de datos inconsistentes para garantizar calidad en la información.

## JUSTIFICACIÓN

Desde el punto de vista de negocio de energía eléctrica, el conocimiento y análisis de la demanda y pérdida de energía eléctrica, es muy importante para los agentes del mercado, este proyecto ilustra las ventajas de las soluciones de BI para estos casos donde es necesario analizar información histórica para la toma de decisiones. Contar con una solución de BI, va a entregar información confiable en muy corto tiempo.

Una solución de BI une fuentes de datos diversas, en diferentes formatos, entregando resultados en forma rápida y estructurada para satisfacer los requerimientos de los agentes para tomar decisiones rápidas. Las empresas cada día están empezando a crear soluciones de BI, como lo mencionan Curto y Conesa (2015): “La sociedad de la información ha propiciado la necesidad de tener mejores, más rápidas y más eficientes métodos para extraer y transformar los datos de una organización en información y distribuirla a lo largo de la cadena de valor” (p. 18).

En la tesis de grado de predicción de demanda de energía eléctrica en Colombia, Rueda (2011), referencia otros estudios que indica la necesidad de contar con una predicción precisa de la demanda de electricidad como “un insumo fundamental para la toma de decisiones operativas y estratégicas de los agentes de un mercado de electricidad” (p. 2). En este trabajo menciona estudios que demuestran que: “las imprecisiones en el pronóstico tienen altos costos económicos; por ejemplo, Bunn y Farmer (1985) demuestran que un incremento de 1% en el error de pronóstico, puede causar un incremento de hasta diez millones de libras esterlinas en la operación del sistema” (p. 2).

Para una empresa tener un BI trae ventajas, incluyo algunas de ellas tomadas de Curto (2015):

* Transformación de Los datos en información que generan conocimiento para tomar decisiones que se traducen en mejores resultados.
* Una única visión de los datos con calidad.
* Contar con métricas e indicadores claves de desempeño (KPI).
* Cruce de métricas con diferentes datos maestros.
* Mejora en la competitividad de la organización.

Otros de los beneficios de realizar un proyecto de BI, es porque investiga otros temas no tan comunes en la ingeniería de software a nivel de pregrado y aporta en el conocimiento de bases de datos de datos y de inteligencia de negocios (BI). El proyecto va a ser de utilidad para las personas que inician o que ya han trabajado con BI, serán utilizadas buenas prácticas y una breve descripción de las fases de desarrollo, diseño y construcción, con el objeto de motivar a los lectores en el tema de BI.

# OBJETIVOS

## OBJETIVO GENERAL

Construir un prototipo funcional para analizar la demanda y pérdidas de energía eléctrica utilizando tecnología de Business Inteligence (BI) como solución de software.

## OBJETIVOS ESPECÍFICOS

* Construir una solución de BI que le sirva a los agentes del mercado para analizar la información de demanda y pérdidas de energía eléctrica para la toma de decisiones en forma oportuna.
* Mejorar la calidad de datos por medio de procesos de auditoría de procesos y reportes de auditoría para garantizar la confiabilidad de los agentes en la información que están consultado.
* Crear un sitio de consultas y reportes, rápido, de agradable presentación y sin costos adicionales, para que los agentes (usuarios) analicen la información.
* Contribuir en la generación de conocimiento en BI, con la construcción de código robusto para soluciones de BI.

# MARCO TEÓRICO

Este capítulo explica dos bloques como marco teórico para el proyecto:

* Demanda de energía eléctrica
* Soluciones de business intelligence (BI).

## Cadena productiva DE energía eléctrica

Para satisfacer las necesidades de energía eléctrica existe un mercado de energía. La Figura 2 muestra los elementos que hacen parte de la cadena productiva de energía eléctrica, los elementos son explicados a continuación.

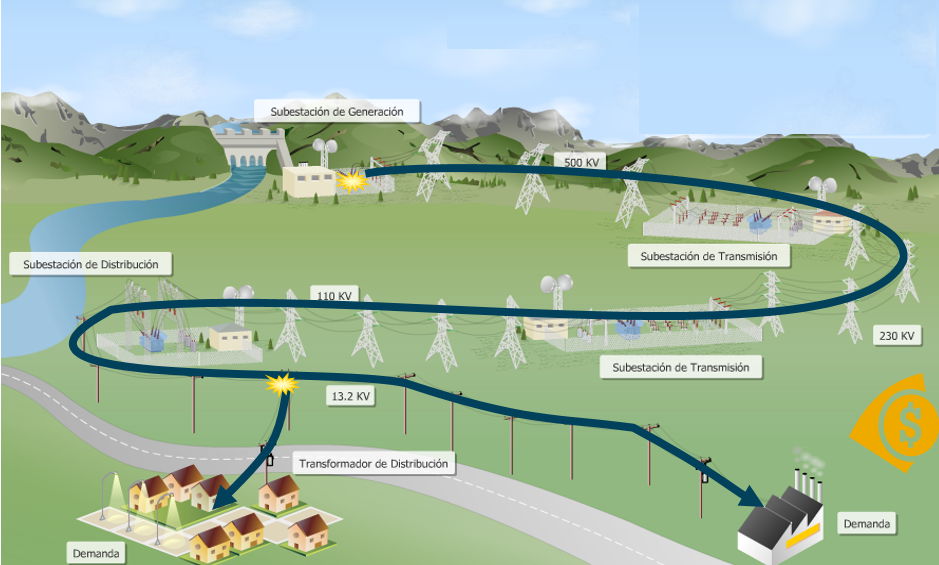


Figura 2. Cadena productiva de energía eléctrica

Fuente: <https://www.xm.com.co>

### Generación de energía eléctrica

“Persona natural o jurídica que produce energía eléctrica y tiene por lo menos una planta y/o unidad de generación conectada al Sistema Interconectado Nacional (SIN), bien sea que desarrolle esa actividad en forma exclusiva o en forma combinada con otra u otras actividades del sector eléctrico, cualquiera de ellas sea la actividad principal” (Memorias seminarios 2018, XM).

### Transmisión y distribución de energía eléctrica

Los transmisores son agentes que transportan la energía eléctrica. Los transmisores transportan la energía desde largas distancias desde el sitio de generación hasta las regiones y los distribuidores realizan el transporte en menores distancia en las regiones hasta el consumidor final (Memorias seminarios 2018, XM).

### Comercialización de energía eléctrica

Los agentes comercializadores son las empresas encargadas de la venta de energía eléctrica. Venden la energía a dos mercados macro conocidos como mercado regulado y no regulado, también vende energía para el alumbrado público y exportaciones a otros países. El mercado regulado son los hogares y pequeños consumidores del sector comercial e industrial y el mercado no regulado son las ventas de energía a grandes empresas o lugares con altos consumos de energía (Memorias seminarios 2018, XM).

### Operadores de la red de energía eléctrica

Los operadores de red son los responsables de mantener las redes de distribución de energía en una región. En una región pueden existir varios distribuidores, pero solo uno de ellos es el operador de la red local (Memorias seminarios 2018, XM).

### Administración del mercado de energía mayorista

XM S.A. E.S.P. Es el operador del sistema y administrador del mercado. Es el encargado de la planeación y coordinación de la operación de los recursos de generación y transmisión del SIN. Administra los contratos y la liquidación de cargos por uso de las redes de energía del sistema interconectado nacional, realiza la liquidación, facturación, cobro y pagos de los contratos y compras en la bolsa por generadores y comercializadores (Memorias seminarios 2018, XM).

### Sistema interconectado nacional – SIN

El SIN es el sistema interconectado nacional, es una red de transmisión que permite llevar la energía desde las centrales de generación hasta los centros de consumo. Está formado por líneas, torres, transformadores, subestaciones y otros equipos. El SIN está divido en tres subsistemas: STN: Sistema de transmisión nacional encargado de transmisión de energía a nivel nacional, el STR: Sistema de transmisión regional recibe la energía del STN y la distribuye a nivel regional y el SDL: Sistema de transmisión local es la última etapa de la cadena encargad de distribuir la energía a nivel local (Memorias seminarios 2018, XM).

## DEMANDA DE ENERGÍA

La demanda es referida al consumo de energía en una región y es uno de los insumos para programar la generación de energía eléctrica. La demanda depende de elementos como geografía, día de la semana, hora, etc.

## PÉRDIDAS DE ENERGÍA

En el año 2018 Colombia perdió alrededor de mil billones de dólares anuales por robo de energía y por pérdidas no técnicas, unos 5,700 millones de kWh, que representa 10% de la energía distribuida en el país. Tasa muy alta respecto al nivel global que están entre el 1 y 3% (El universal, 2018).

## SOLUCIONES DE BUSINESS INTELLIGENCE

¿Qué es BI? La sociedad de la información requiere métodos más rápidos para extraer y transformar datos de una organización en información y difundirlo en la cadena de valor. BI responde a esta necesidad. BI es un sistema de soporte a la organización basada en hechos históricos. (Curto Diaz, 2015).

En el libro “The data warehouse toolkit” (Kimball, 2013) dice que el volumen de datos de las empresas ha crecido de megabytes a gigabytes, terabytes, incluso a petabytes, sin embargo, el desafío básico de los sistemas DW / BI ha sido mantenido constante. El trabajo de BI es tomar estos datos históricos y transformarlos para la toma de decisiones. Este tipo de soluciones es conocido como inteligencia de negocios o Business Intelligence (BI).

BI es un grupo de técnicas y tecnologías para: (1) crear un Data Warehouse (DW) o bodega de datos que almacena los datos para facilitar la extracción desde las fuentes de datos de la empresa. El DW ayuda en la toma de decisiones. Es un historial de la organización modelado como estrellas o copos de nieve para favorecer el análisis y la divulgación de la información, (2) la carga de datos es realizada por ETL (Extraction – Transformation – Load: Carga) que extrae, transforma y carga los datos desde las bases de datos operativas hasta el DW, (3) luego son creados sistemas de análisis (OLAP) con métricas e indicadores del negocio y (4) finalmente los datos son presentados al usuario en diferentes herramientas como informes, cuadros de mando (dashboard), hojas de cálculos, aplicaciones cliente, etc. (Kimball, 2013).

“BI concentra alto rendimiento, capacidades y beneficios como la analítica acelerada e información, BI utiliza técnicas como ETL para analizar de forma rápida grandes volúmenes de datos, dando paso al análisis en tiempo real, y la reducción de tiempos de espera para reportes” (Bustamante, 2018). Para el proyecto el desempeño será uno de los retos, como lo mencionan los autores la necesidad de procesar altos volúmenes de datos y las exigencias de los usuarios para consultar reportes, requieren de un sistema BI robusto capaz de entregar la información en forma oportuna.

## Ciclo de vida de proyectos BI

Morales (2019) en su tesis de grado sobre BI, menciona que el ciclo de vida de proyectos de BI es similar a la de otros proyectos informáticos, formado por cinco fases:

* Fase 1: Dirección y planeación
* Fase 2: Recolección de información
* Fase 3: Procesamiento de datos
* Fase 4: Análisis y construcción
* Fase 5: Difusión

En la Tabla 1, hay otras consideraciones en el ciclo de vida utilizado en proyectos de BI, la tabla es basada en Vanrell, Bertone, & Ramón (Citados por Morales, 2019). Esta tabla adiciona un componente diferenciador respecto a las fases anteriores y es la importancia de los datos.

Tabla 1. Ciclo de vida proyectos BI

|  |  |  |
| --- | --- | --- |
| **Subprocesos** | **Tarea** | **Salida** |
| Entendimiento del negocio | * Determinar las metas del BI. | * Criterios de éxito * Preguntas del negocio |
| Entendimiento de los datos | * Reunir datos iniciales * Describir los datos * Explorar los datos | * Reporte de datos iniciales, descripción y exploración * Reporte de calidad |
| Preparación de datos | * Selección de datos * Limpieza de datos | * Conjuntos de datos descritos * Justificación de inclusión y exclusión * Reporte de limpieza |
| Modelado | * Seleccionar técnica de modelado | * Técnica de modelado * Suposiciones de modelado * Diseño de pruebas * Establecer parámetros * Modelado de los datos |
| Evaluación | * Evaluar resultados * Revisar procesos * Determinar próximos pasos | * Evaluación de resultados hacia los criterios de éxito * Modelos aprobados * Lista de decisiones |
| Entrega | * Producir un reporte final | * Reporte final * Presentación final |

## METODOLOGÍA ÁGIL PARA SOLUCIONES DE BI

El portal TodoBI (2019), en su artículo sobre metodologías ágiles para BI, da recomendaciones para la construcción de proyectos: Las metodologías ágiles han sido aplicadas en proyectos donde el componente de ‘desarrollo’ tiene un peso muy importante y es difícil aplicarlo en proyectos de BI.

### Enfoque tradicional de planificación en BI/DW

Las metodologías en cascada han demostrado deficiencias para crear un proyecto de BI exitoso. Un proyecto de BI puede considerar exitoso cuando es utilizado por la mayor parte de la organización porque ofrece ‘lo que necesitan’). La Figura 3 muestra modelos teóricos para construcción de BI creados por Kimball , Inmon y Lindstedt. Estos modelos son muy útiles para construir un BI, pero es necesario enfoques ágiles para llevarlos a la práctica.

Bill Inmon fue uno de los primeros autores en escribir sobre DW (1990), donde el DW es orientado por temas (por ejemplo, ventas, producción, compras, etc.) que son variantes en el tiempo, define una metodología descendente (top-down) considera mejor crear toda la implementación corporativa, y los data marts pueden ser creados después de terminar el DW completo de la organización. La propuesta de Lindstedt, es similar a la de Inmon con una capa adicional (Morales, 2019).

Ralph Kimbal (1996), define el DW como la unión de todos los data marts y define múltiples bases de datos llamadas data marts organizadas por procesos de negocio (Morales, 2019).

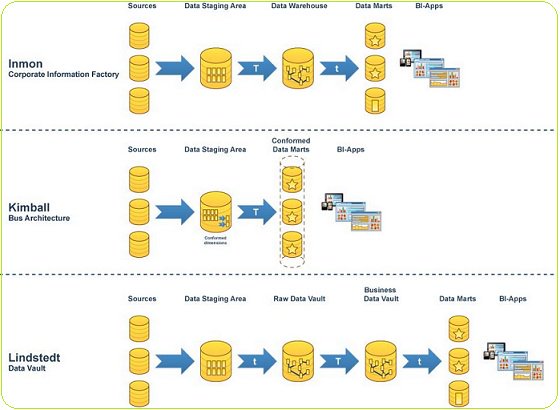


Figura 3. Modelos para construcción de BI creados por Kimball, Inmon y Lidstedt

Fuente: <https://todobi.com/metodologias-agiles-para-bidw/>

Los problemas ocurren porque estos proyectos tienen meses de planificación y cuando surgen problemas de arquitectura, volumen de datos, cambios de requisitos y mejoras, es muy difícil una respuesta rápida. Los usuarios e interesados en el proyecto no se sienten parte del proyecto porque sus sugerencias llegan tarde.

### Recomendaciones para proyectos ágiles BI/DW

Las siguientes son algunas recomendaciones del artículo de TodoBI (2019), para implementar proyectos de BI utilizando metodologías ágiles:

* Crear prototipos para establecer acuerdos con el usuario.
* Utilizar una metodología ágil, por ejemplo, Scrum, lo más importante es un cambio de mentalidad.
* La regla de oro: mejor rehacer 30% ahora que 100% dentro de 6 meses. No tener miedo al cambio en los prototipos.
* Tener en cuenta la opinión de todo el equipo, que sientan que sus opiniones cuentan.
* Realiza los planes de prueba desde el inicio.
* Involucra a los usuarios en todas las fases de desarrollo.
* No dejar el diseño y la usabilidad para el final.
* Un proyecto BI/DW nunca se acaba.

## METODOLOGÍA KIMBALL

La Figura 3 muestra la “arquitectura bus” definida por Kimball, existen otros enfoques, pero esta arquitectura es la utilizada en este proyecto. A continuación, una breve descripción. Para mayor detalle consulte en Ferrai & Russo (2008), Kimball. & Ross (2013).

La arquitectura para las soluciones BI es construida estableciendo las fuentes de datos a las cuales son realizadas procesos de transformación para ser llevados los datos al DW en los modelos estrella, los datos luegos son procesados en los cubos de información para ser presentada la información a los usuario en las herramientas cliente.

El diseño dimensional y la arquitectura propuesta por Kimball (2014), es similar a otros autores como Curto (2015) y Bustamante (2018), este tipo de diseño es la base de este proyecto, en los siguientes capítulos será construida la solución de BI, utilizando los lineamientos para el modelo dimensional.

### Hechos y dimensiones

El core de la metodología Kimball separa las bases de datos OLAP en dos tipos de entidades:

* **Hechos (fact)**: Almacenan las métricas del sistema, para el proyecto la demanda y las pérdidas de energía.
* **Dimensiones**: En este proyecto son creadas dimensiones como: fecha, geografía y agente. Las dimensiones son las entidades por las cuales son consultadas las métricas.

En el mundo de BI encontramos otros conceptos como los siguientes, pero queda por fuera del alcance del proyecto sus detalles:

* Dimensiones: Junk dimensions (dimensión basura), dimensiones degeneradas, dimensiones de cambio lento tipo 1, 2, 3 y 6.
* Hechos: Tablas de hechos sin métricas (factless fact tables)

### Modelo estrella y copo de nieve

La Figura 4 muestra el modelo estrella, el centro de la estrella es una tabla de hechos y las puntas de la estrella son las tablas con las cuales está relacionada la tabla de hechos. Los modelos estrellas son los preferibles, pero existen otros tipos como el modelo copo de nieve, en estos las dimensiones de las puntas de la estrella son relacionadas con otras tablas, por ejemplo, una dimensión producto podría tener relación con otras tablas de categoría y subcategoría de productos.

|  |  |
| --- | --- |
|  |  |

Figura 4. A la izquierda modelo estrella y a la derecha modelo copo de nieve

Fuente: Autor

### Integración de datos con ETLs

Las ETLs son la herramienta que nos ayudan a mover los datos de un sistema a otro. Para el proyecto las fuentes de datos son archivos planos, estos datos son pasados a las tablas temporales o sataging area (SA), son temporales porque los datos de las tablas son borrados al inicio de cada ejecución, en el proceso son pobladas, luego aplicados procesos de transformación y limpieza y finalmente son pasados los datos al DW.

### Data warehouse

El DW es el más importe en el modelo Kimball, en la literatura hay mucha información, los DW son modelados como estrellas o copos de nieve. Las ETLs llenan los datos y luego otras ETLs toman los datos del DW y los llevan a los cubos OLAP.

### Cubos OLAP

Los cubos son procesados con ETLs que llevan los datos de la base de datos DW. Los dos elementos importantes son el diseño de las dimensiones con jerarquías y los cálculos para la creación de las métricas.

### Herramientas clientes

Las siguientes son algunas herramientas clientes para BI: reportes, dashboard, aplicaciones web, tablas dinámicas en Excel, servicios web para consultar los cubos.

## INTEGRACIÓN DE DATOS CON ETL

En el libro “Pentaho Kettle solutions”, los autores Casters, Bouman, y Dongen (2010) aclaran la diferencia entre los sistemas BI de los sistemas transaccionales OLTP (Procesos de transacciones en línea), para ellos la gran diferencia entre un OLTP y una base de datos BI (la bodega de datos o DW) es la cantidad de datos analizada en una transacción. Los sistemas OLTP son manajados por varios usuarios concurrentes modificando uno o un grupo de registros, mientras que en un DW tiene la capacidad de operar en millones de registros para responder a una consulta. Un OLTP es utilizado para la parte operativa de los datos mientras que un BI es usado para análisis de información.

### ¿Qué es una ETL?

ETL son procesos de extracción transformación y carga. “Una definición simple puede ser el conjunto de procesos para llevar datos de un sistema OLTP a un data warehouse” (Casters, 2010, p. 53). En realidad las fuentes de datos no solo son de sistemas OLTP, los DW también son alimentados de otras fuentes de datos como servicios web, archivos planos, hojas de cálculo, etc. Los tres pasos más importante de una ETL son:

* Extracción: Obtener datos de las diferentes fuentes de datos
* Transformación: Procesos aplicados a los datos como: validación, modificación, operaciones de calidad de datos, etc.
* Carga: Llevar los datos al desfino final, que para un BI es la bodega de datos DW.

### Evolución de soluciones de ETL

Para realizar la integración de datos es necesario contar con herramientas, antes de la existencias de las ETLs esta integración era realizada por por scripts y por código creado en diferentes lenguajes. Hoy existen herramientas gratuitas como Penthaho y otras herramientas pagas con funcionalidades adicionales para crear soluciones que incrementan la productividad y mantenibilidad (Casters, 2010).

En este proyecto las ETLs son una de las herramientas más importantes, porque son las que alimentan día a día el sistema con los datos para la bodega de datos.

### Bloques para construir ETLs

El sistema debe funcionar de la siguente forma, cada día es alimentado los datos de las tablas de dimensiones por las cuales consultar la información, estos datos maestros son transformados de los sistemas OLTP a esquemas dimensiones en el DW. También es extraído los datos de las tablas de movimiento en un rango de fechas, estos datos son cargadas en las tablas de hechos aplicando reglas de negocio y realizando tareas de calida de datos para que la informacón sea confiable y por último los datos son procesados en cubos multidimensionales para responder a las preguntas del negocio (Casters, 2010).

### Arquitectura clasica de sistemas BI

La arquitectura clásica de un BI puede ser resumida como: Un grupo de ETLs toma los datos de las fuentes de datos, a estos datos son aplicadas transformaciones y son llevados a la base de datos DW, luego son ejecutadas otras ETLs para procesar los cubos, los reportes consumen las métricas e indicadores de los cubos y las aplicaciones presentan la información a los usuarios finales (Casters, 2010).

Esto es lo que será construido en este proyecto para analizar la información de demanda y pérdias de energía, en los siguientes capítulos será explicado cada una de los elementos descritos en esta arquitectura base utilizando las herramientas seleccionadas para el proyecto y aplicando buenas prácticas en el proceso de construcción. En la Figura 6. Diseño general del sistema, es el diseño para para este proyecto, el diseño sigue estos lineamientos de arquitectura para un BI, el cual es modelado de forma muy similar (casi identica – apreciación personal) por autores como Casters (2010), Kimball (2013) y Curto (2015).

# MARCO METODOLÓGICO

El tipo de investigación del proyecto es: investigación aplicada, el resultado final será un producto software para “analizar la demanda y pérdidas de energía eléctrica”. A continuación, será explicado el flujo de pasos para la construcción del sistema, la metodología para la construcción de soluciones de BI y la selección de herramientas para la construcción del proyecto.

## PLAN DE PROYECTO

La Tabla 2 presenta el flujo de pasos para la construcción del sistemas.

Tabla 2. Flujo de pasos para el proyecto

|  |
| --- |
| 1. Fase de elaboración    * Selección de herramientas de desarrollo    * Requisitos      1. Establecer las preguntas del negocio      2. Definir las fuentes de datos    * Diseños para el sistema 2. Fase de construcción    * Construcción de artefactos      1. Data warehouse      2. Bitácora de procesos      3. ETL para generar los archivos planos para el proyecto      4. ETLs del sistema      5. Cubos OLAP      6. Reportes      7. Sitio Web    * Pruebas de validación del sistema    * Documentación      1. Crear de manual de instalación del sistema      2. Crear manual de operación del sistema 3. Fase de implementación    * Instalación de la solución    * Resultados y entregables del proyecto    * Conclusiones del proyecto    * Recomendaciones para trabajos futuros |

## ALCANCE DEL PROYECTO

El alcance del proyecto es la construcción de un producto de software para analizar la demanda y pérdidas de energía eléctrica por conceptos como fecha, periodo, agente, tipo de mercado (regulado y no regulado) y geografía. Como fuente de datos serán utilizados archivos planos, estos archivos serán obtenidos de las bases de datos de la empresa XM. Para contar con información confiable serán incluido objetos de bitácora para seguimiento a los procesos de carga de datos. También serán creado objetos de utilidades para automatizar los procesos de carga y recarga de datos.

Como el número de combinaciones de cruces de las métricas con las dimensiones es amplio, así como la granularidad para presentar los datos (por ejemplo, mostrar la demanda de energía por hora, día, mes o año), solo serán creados algunos reportes para las métricas del sistema y de la bitácora de información del sistema.

No serán creado pronósticos futuros de las métricas de demanda y pérdidas de energía puesto que esto es del alcance de un BA, esta es una recomendación para proyectos futuros, solo será realizada análisis de la información histórica.

## METODOLOGÍA PARA LA CONSTRUCCIÓN DE LA SOLUCIÓN DE BI

El diseño del sistema será realizado utilizando la metodología Kimball, la cual es apropiada para pequeños proyectos y es más simple (Morales, 2019), el cual es nuestro caso que solo cuenta con una modelo estrella con dos métricas: demanda y pérdidas de energía.

En cuento a la metodología de construcción de proyectos como cascada o ágil, algunos aspectos no aplican, puesto que no existe un equipo de proyecto y no hay riesgos por cambios en las especificaciones puesto que no existe un usuario final “real”. En proyectos que he participado de BI, han sido realizado con una metodología en cascada. “El uso de metodologías ágiles en proyectos de BI es difícil, pero es necesario tener un cambio de mentalidad para aprovechar las ventajas que estas metodologías ofrecen” (TodoBI, 2019).

## FASE DE CONSTRUCCIÓN

La construcción del sistema inicia después de establecer los requisitos, realizar los diseños y seleccionar las herramientas de desarrollo. En los siguientes capítulos será realizado una explicación de los aspectos más relevantes en la construcción.

Ferrari (2019) en su libro de metodología BI, selecciona como herramientas para BI Microsoft SQL (base de datos, ETLs, cubos OLAP y reportes Reporting Services), sigue “La arquitectura bus” de Kimball para BI. El libro ha sido una buena referencia porque este proyecto también utiliza estas herramientas y el modelo de Kimball para la construcción del BI/DW.

Los siguientes son los artefactos a construir:

* Fuentes de datos: Las fuentes de datos son archivos planos, y es necesario construir una ETL para obtener los datos de las bases de datos de XM.
* Staging Area: tablas de almacenamiento temporal utilizadas para el paso de datos y aplicación de tareas de limpieza y transformación antes de ser subidos los datos al DW.
* Data warehouse: La bodega de datos es una base de datos modelada como una estrella.
* Cubos OLAP: Estructuras multidimensionales con las métricas del negocio y las dimensiones por las cuales consultarlas
* Reportes: El proyecto tendrá dos grupos de reportes: bitácora de procesos y los propios del sistema para analizar la demanda y pérdidas de energía eléctrica.
* Herramienta cliente: Aplicación Web, para consultar los reportes del sistema.
* Aspectos no funcionales: Los aspectos no funcionales más importantes que van a ser incorporados en el proyecto son:
  + Bitácora para seguimiento de ejecución procesos y calidad de datos
  + Particionamiento de tablas y cubos para manejar grandes volúmenes de datos
  + Creación de objetos de utilidades para la carga diaria en forma automática y recarga de datos históricos.

## SELECCIÓN DE HERRAMIENTAS DE DESARROLLO

Al comparar las herramientas de BI, cada una tiene aspectos favorables y desfavorables, a continuación, algunos aspectos que fueron tenidos en cuenta en la selección de las herramientas para la construcción del proyecto.

La Figura 5 muestra el “cuadrante Mágico de Gartner” creada por la empresa Gartner, en el cual “posiciona jugadores de tecnología”. Para entender el gráfico el eje X evalúa el conocimiento y la integridad de visión de proveedores de software para generar valor, para ellos mismos y para sus clientes. El eje Y evalúa la capacidad de ejecución de su visión. El siguiente gráfico es el cuadrante de la empresa Gartner para Analítica y BI para el año 2020. El gráfico muestra a Microsoft como líder en el 2020 para plataformas de Análisis y BI. Ferrari (2019) también presenta el cuadrante de Gartner en el proceso de selección de las herramientas de desarrollo.



Figura 5. Cuadrante mágico de Gartner para Analítica y BI

Fuente: <https://www.gartner.com/>

Las herramientas de uso libre son una buena opción a nivel académico, pero tienen funcionalidades reducidas y menor desempeño comparada con las pagas. Por lo general los proyectos a nivel empresariales de BI son de grandes volúmenes de datos, con necesidades de alto desempeño. En este proyecto los datos son tomados de la bodega de datos de XM, el número de filas de demanda y pérdidas de energía es alrededor de 2,000 millones de filas en 20 años, por lo tanto, es necesario contar con herramientas robustas.

Otra dificultad de las herramientas gratuitas es mencionada por Dingeldein (2019): “Desafortunadamente, ningún software de BI de uso libre aprobó nuestra metodología rigurosa (principalmente debido a la falta de revisiones suficientes). Si bien aún no tienen muchas reseñas, parecen prometedoras”.

Entre las herramientas gratuitas de BI está: Pentaho, la cual es utilizada en la maestría “Business Intelligence y Big data”, dictada en la Universidad Oberta de Catalunya, también aparece en reseñas de herramientas de BI de uso libre y en libros sobre BI: “De todas las herramientas libres de BI, Penthaho es posiblemente la suite más completa y madura del mercado en el momento de la redacción de este libro”. (Curto Diaz, 2015, p.14).

Entre las herramientas pagas, las más conocidas en nuestro medio son: Microsoft SQL Server, Oracle BI, IBM Cognos Analytics y SAP BI.

Para el proyecto fueron seleccionado las herramientas Microsoft para la construcción del BI y el sitio Web de consultas, las razones más importantes son:

* La valoración dada por Garther la cual posiciona a Microsoft en los últimos años como como líder para plataformas de Análisis y BI.
* La necesidad de trabajar con grandes volúmenes de datos y el tiempo de respuesta requerido por los agentes de mercado para el análisis de información.
* Por ser una tecnología muy difundida en nuestro medio, en particular, el BI de la empresa XM es construido con Microsoft SQL Server.

# REQUISITOS DEL SISTEMA

Los requisitos son el punto de partida para establecer las funcionalidades del sistema, recursos y tiempo de implementación. Sería muy extenso redactar todos los requisitos, diseños y arquitectura del sistema. En su lugar será realizada una descripción general de los aspectos importantes para la implementación del BI.

## PREGUNTAS DEL NEGOCIO

En el ámbito académico utilizamos la “pregunta problematizadora” para establecer la meta a lograr. Para BI son utilizadas las “preguntas del negocio” que el sistema va a responder, la Tabla 3 es un ejemplo de los campos que puede tener el listado de preguntas del negocio. Generalmente son obtenidas en reuniones con los expertos del negocio. Este listado es categorizado y analizado para llegar a acuerdos del alcance del sistema a construir.

Tabla 3. Preguntas del negocio

|  |  |  |
| --- | --- | --- |
| **Campo** | **Descripción** | **Ejemplo** |
| Id | Identificador, por ejemplo, consecutivo | 1, 2, 3 |
| Pregunta | Pregunta del negocio | ¿Cuál es la demanda de energía eléctrica por geografía? |
| Temática | Tema que agrupa la pregunta | Demanda |
| Existen informes | Indica si existe informes construidos, esto sirve para tener un reporte de referencia | Si, en el porta BI de XM hay reportes gráficos y archivos Excel. |
| Fuente de datos | Indica cual es la fuente de datos para obtener la información | Base de datos de XM creada en SQL Server. |
| Observaciones | Observaciones sobre la pregunta. | Los datos pueden ser descargados del portal BI de XM o del DW. |

Para este proyecto en particular las preguntas del negocio a resolver son:

“Obtener información de demanda y pérdidas de energía eléctrica, que pueda ser consultada por fecha, geografía, agente y tipo de mercado”.

## DISEÑO GENERAL DEL SISTEMA

La Figura 6 es un diseño típico para solución de BI, de izquierda a derecha encuentra: preguntas del negocio (especificaciones), fuentes de datos, área de almacenamiento temporal, DW, cubos OLAP y herramientas clientes. Uno de los componentes más importantes son las ETLs que alimentan el sistema en forma periódica (diaria). Como herramientas clientes serán creados reportes SQL Reporting Services y una aplicación web para consultarlos. Este diseño sigue la metodología de Kimball para proyectos de BI.

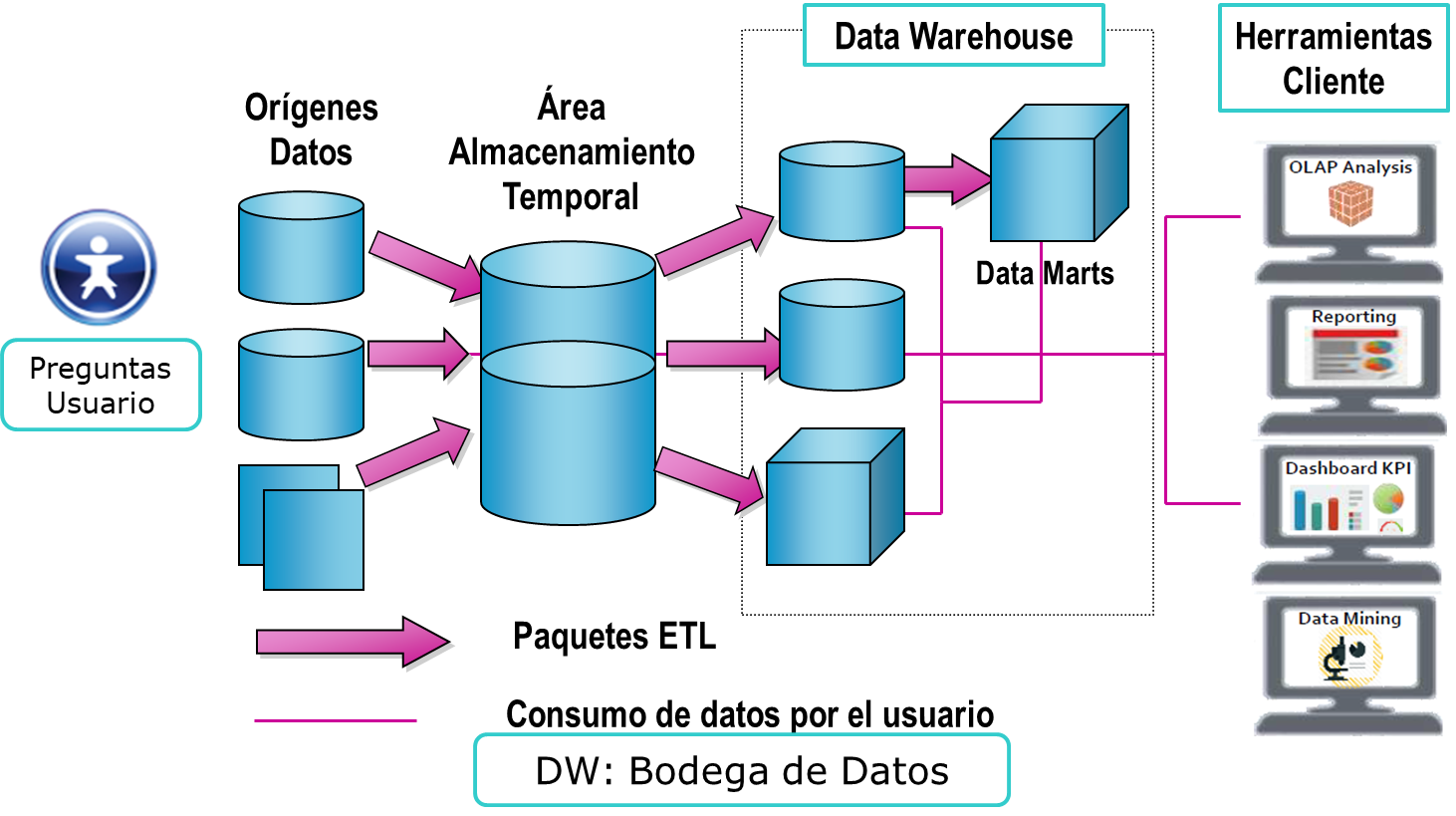


Figura 6. Diseño general del sistema

Fuente: Autor

## REQUISITOS

Una buena definición de los requisitos del sistema (funcionales y no funcionales) son claves para el éxito del proyecto. En particular para proyectos de BI es fundamental tener claro los orígenes de datos y las consultas de selección de datos. Los requisitos son creados después de analizar las preguntas del negocio.

Para implementar el sistema fueron establecido un modelo (estrella) formado por la tabla de movimiento con las métricas que van a responder las preguntas del negocio (centro de la estrella) y las dimensiones (puntas de la estrella) que representan los elementos por los cuales consultar los datos como: fecha, periodo, geografía, agente, compañía y mercado.

## Fuentes de datos

Para este sistema las fuentes de datos son establecidas como archivos planos, estos archivos fueron obtenidos de la empresa XM. El anexo 1 contiene la autorización por parte de XM.

Fue construida una ETL de apoyo para extraer los archivos planos para el proyecto. En el anexo 1 están las consultas de selección de datos para las siguientes tablas del DW:

* Dimensión agente
* Dimensión compañía
* Dimensión geografía
* Dimensión mercado
* Hechos demanda y pérdidas de energía

Un mayor detalle de estos archivos y su formato puede ser ampliado en el Anexo 3 manual de operación del sistema.

## DISEÑOS PARA LA SOLUCIÓN

Los diseños son parte de la fase de elaboración, estos son algunos diseños que pueden ser creados para un proyecto de BI. En los siguientes capítulos puede consultar algunos de estos diseños.

* Diseño de la arquitectura de la solución
* Diseño de tablas de la bodega de datos (DW)
* Diseño de métricas para los cubos de información
* Diseño de ETLs para carga de datos
* Diseño de bitácora de ejecución de procesos
* Diseño de reportes
* Diseño del sitio Web

## DICCIONARIO DE DATOS

El diccionario de datos contiene el diseño de las tablas del sistema, para este BI fue creado el siguiente grupo de tablas:

* Tablas de dimensiones: Entidades por las cuales consultar los datos, por ejemplo, fecha, geografía, agente. En sistemas BI son denominadas tablas de dimensión y en otros sistemas tablas maestras.
* Tablas de hechos: Contienen las métricas que responden las preguntas del negocio, por ejemplo, demanda de energía. En sistemas BI son denominadas tablas Fact (hechos) y en otros sistemas son tablas de movimiento.
* Tablas de bitácora: Almacenan información de auditoría de ejecución de procesos.
* Tablas de utilidades: Tablas auxiliares para manejar el sistema.
* Tablas staging: Tablas temporales utilizadas en la carga de datos.

La Tabla 4 muestra los diccionarios de datos creados en el proyecto

Tabla 4. Diccionarios de datos

|  |  |
| --- | --- |
| **Nombre** | **Descripción** |
| \DemandaBI\Doc\Analisis\DicDatos\_DemandaDW.xlsm | Diccionario de datos del proyecto |
| \DemandaBI\Doc\Analisis\DicDatos\_DemandaXM.xlsm | Diccionario de datos con las tablas orígenes de datos de la empresa XM. |
| \DemandaBI\Doc\Analisis\DisenoOLAP\_DemandaBI.xlsm | Diseño de cubos y matriz de bus con cruce de métricas con dimensiones |

## DIAGRAMAS DE casos de uso Y DE SECUENCIAS

El sistema puede ser dividido en los siguientes grupos funcionales:

* Carga de datos
* Bitácora de procesos
* Reportes
* Sitio Web

Para un BI la carga de datos es la tarea más importante, esta es realizada de dos formas:

* Carga automática: Ejecutada en forma diaria. Para cada modelo (\*) debe ser establecido el día de ejecución y el rango de fechas a procesar
* Recarga manual: Requiere especificar los modelos a cargar y los rangos de fechas de procesos.

**MODELO**: En este contexto un modelo es una estrella, es decir una tabla fact y las tablas de dimensiones relacionadas.

La Figura 7 muestra los casos de uso para el sistema, para el diseño de las actividades de carga de datos, bitácora de proceso, procesamiento de dimensiones u cubos, creación de reportes y sitio web.

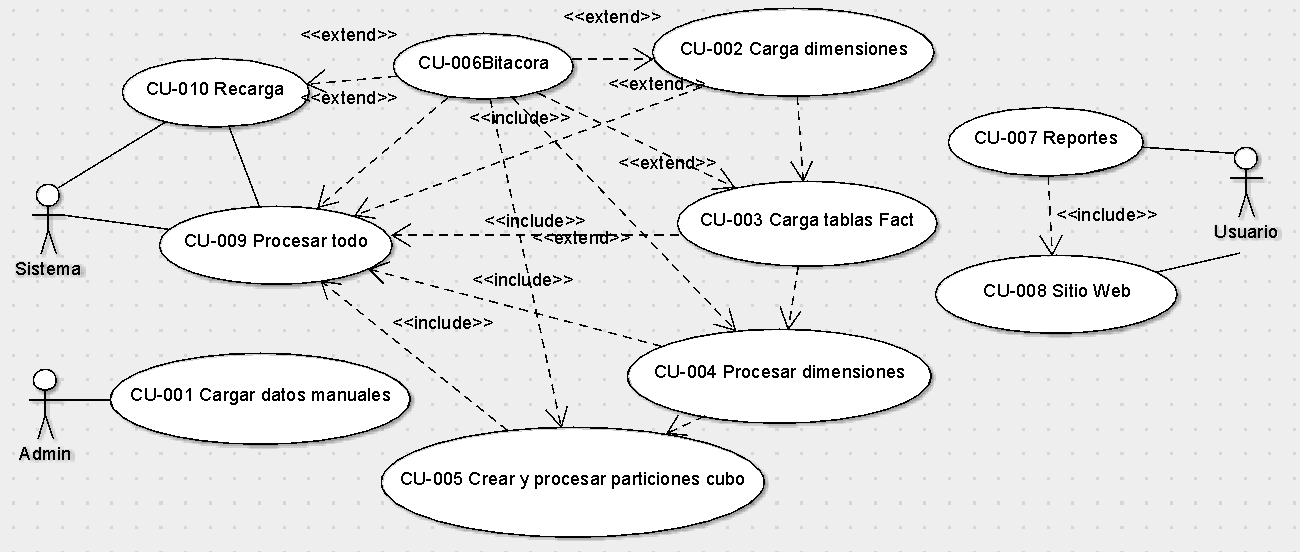


Figura 7. Diagrama general de casos de uso

Fuente: Autor

La Figura 8 es el diagrama de secuencia de las actividades realizadas para el carga automática del DW, esto es lo que ejecuta el job SQL automático para la carga diaria del sistema.

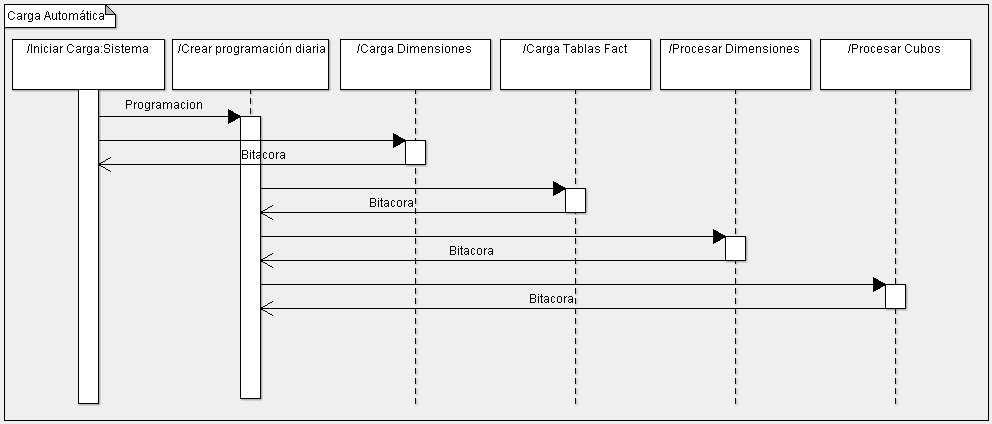


Figura 8. Diagrama de secuencia carga automática DW

Fuente: Autor

La Figura 9 es el diagrama de secuencias para la recarga del sistema cuando requiera ejecutar los procesos de carga en fechas del pasado, por ejemplo cuando los datos en las fuentes de datos sean modificados y requiera actualizar la información del BI.

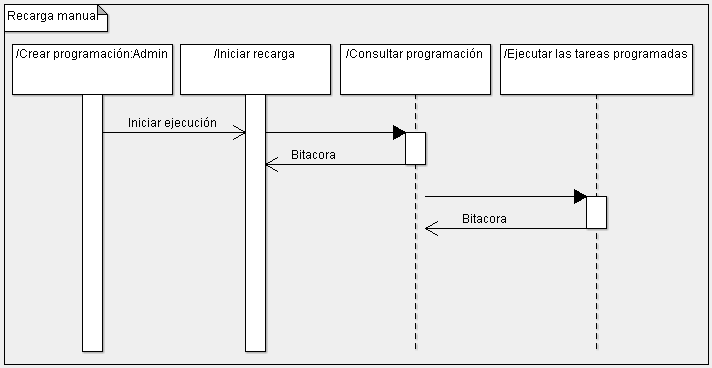


Figura 9. Diagrama de secuencia recarga manual DW

Fuente: Autor

## DESEMPEÑO DEL SISTEMA

Desde la etapa de diseño es definido la granularidad de los datos, también es conveniente estimar el volumen y la tendencia de crecimiento de los datos, esta información es importante para establecer necesidades de hardware y de construcción de artefactos. Este tema será ampliado en los próximos capítulos.

# ARQUITECTURA DEL SISTEMA

El objetivo de la arquitectura es identificar los componentes que harán parte de la solución final. En proyectos de software, los miembros del equipo de arquitectura y diseño utilizan este documento como línea base en el momento de validar cualquier decisión de diseño y los desarrolladores lo utilizan en el momento de implementar los artefactos. En este capítulo serán creados los diseños del sistema.

## Herramientas de desarrollo

En la implementación, fueron utilizado las siguientes herramientas de Microsoft: SQL Server para el motor de bases de datos para la bodega de datos (DW); el motor de integración SQL Integration Services (SSIS) para ETLs de carga de datos y procesamiento de los cubos OLAP; SQL Analysis Services (SSAS), como motor de base de datos multidimensional para los data marts; SQL Reporting Services (SSRS) para reportes; y Microsoft Visual Studio .Net para la creación del sitio Web.

## MINERÍA DE DATOS Y MACHING LEARNING

Un complemento del BI es denominado BA (Business Analytics). BA tiene que ver con minería de datos e inteligencia artificial. BA no hacen parte de este proyecto, pero son incluidos en algunos diagramas. Un complemento para el proyecto sería la construcción de pronósticos de demanda y pérdidas de energía, esta es una de las recomendaciones para siguientes proyectos.

## ARQUITECTURA CONCEPTUAL

La Figura 10 representa la arquitectura conceptual de la solución de BI, a continuación una descripción de los elementos de este diagrama.

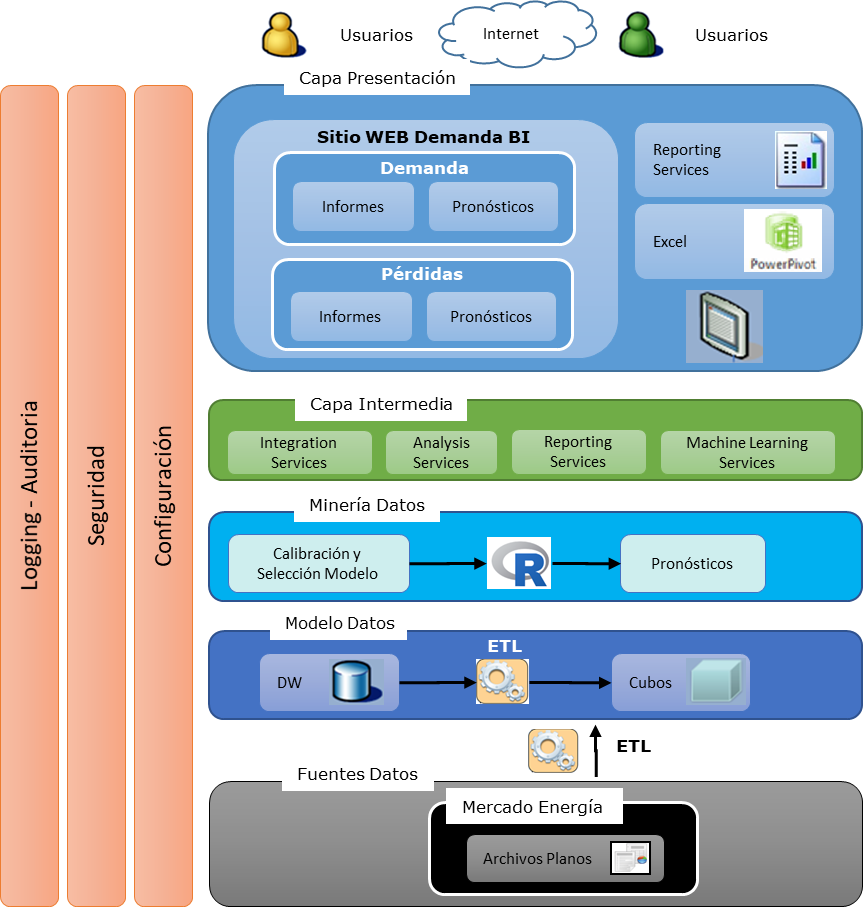


Figura 10. Arquitectura conceptual DemandaBI

Fuente: Autor

### Fuentes de datos

Las fuentes de datos para las ETLs de carga son archivos planos tipo CSV.

### Modelo de datos

El modelo de datos incluye la base de datos de la bodega de datos DW, los objetos de base de datos temporales SA y las estructuras multidimensionales de los data marts (cubos OLAP). EL DW es alimentado por ETLs utilizando los archivos planos como fuentes de datos. El procesamiento de cubos OLAP y sus dimensiones es realizado por ETLs.

### Capa intermedia

La capa intermedia contiene herramientas de back-end:

* SQL Integration Services: Componente de SQL Server, utilizado en las ETLs para carga de la bodega de datos DW y el procesamiento de los cubos.
* SQL Analisys Services: Componente de SQL Server, utilizado para almacenar las bases de datos multidimensionales (cubos OLAP), para el análisis de la información.
* SQL Reporting Services: Componente de SQL Server, utilizado para reportes.
* Machine learning: Componentes para realizar pronósticos (no hace parte de este proyecto).

### Capa de presentación

En la capa de presentación están las herramientas de la parte visual del sistema:

* Microsoft Visual Studio .Net: El sitio fue creado como una aplicación Web, en el sitio pueden ser consultados los reportes.
* SQL Reporting Services: Herramienta para creación de reportes.
* Excel - Power Pivot: Excel es una herramienta de usuarios para la creación de reportes personalizados utilizando la información de los cubos, en este proyecto no fueron creados reportes con Excel.
* Aplicaciones personalizadas: Están por fuera del alcance definido, pero puede expandir el sistema con otras aplicaciones.

### Servicios transversales

Como servicios transversales son identificados los siguientes componentes:

* Logging - Auditoria: Para registro de ejecución y errores de proceso. También permite realizar trazabilidad de la ejecución de los procesos de carga de datos. Para visualizar de forma fácil el log fueron construidos reportes de bitácora.
* Seguridad: Seguridad y perfilamiento del contenido presentado, para nuestro caso los permisos estarán a nivel de motores de base de datos, pero no existen restricciones en la aplicación Web de consulta.
* Configuración: Configuración de los procesos de carga de datos.

## Casos de uso relevantes para la arquitectura

Después de establecer los requisitos funcionales y no funcionales, creamos los casos de uso, la Figura 11 presenta cuatro grupos funcionales de los aspectos más relevantes del proyecto.

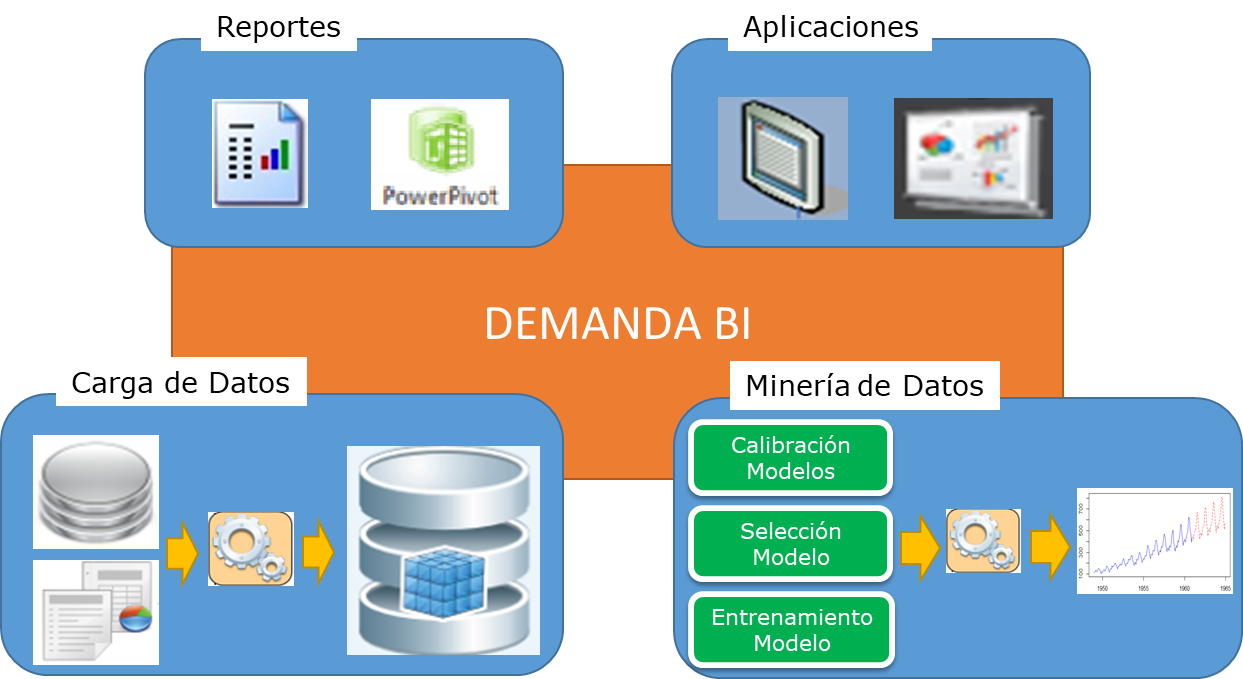


Figura 11. Grupos funcionales DemandaBI

Fuente: Autor

**NOTA**: La minería de datos no hace parte del alcance de este proyecto, es incluido por ser una tendencia que cada día toma más importancia el contar con inteligencia artificial para predecir comportamientos futuros, este es una de las recomendaciones para proyectos futuros.

### Carga de datos

La carga de datos es realizada con ETLs construidas con paquetes SQL Integration Services. Las fuentes de datos son archivos planos tipo CSV. A continuación, son explicados los componentes que hacen parte de la carga de datos.

#### Base de datos relacional Data Warehouse

DemandaDW: Es el nombre de la base de datos relacional DW, contiene los objetos de base de datos como: tablas de hechos (Fact) y tablas de dimensiones de los modelos estrella del sistema. El motor de base de datos es Microsoft SQL Server.

#### Área de almacenamiento temporal Staging Area

En la base de datos del DW estará alojada el Staging Area (SA), está es un grupo de tablas para almacenamiento temporal de datos utilizadas en las ETLs de carga de datos.

#### ETLs para carga de la bodega de datos DW

Las ETLs son el corazón del BI, su función es la carga de datos. También es construida una ETL que agrupe y sincronice la ejecución de todas las ETLs. Las ETLs tendrán bitácora para la auditoría de los procesos. La ejecución de estas generalmente es en forma diaria, la ejecución es realizada en forma automática por Jobs SQL. Para cada tabla será construida una ETL independiente.

#### ETLs para procesamiento de cubos

También fueron creadas tres ETLs encargadas de procesar todas las dimensiones, crear particiones y procesar particiones de los cubos. Para ampliar el tema consulte el manual de operación.

## ARQUITECTURA DE DESPLIEGUE

La Figura 12 presentan la arquitectura de despliegue recomendada para el sistema DemandaBI. Otra alternativa es unir los componentes en un único servidor, en este caso el servidor debe contar con los recursos suficientes para atender los diferentes roles. La decisión de contar con un único o diferentes servidores depende de los recursos económicos y las necesidades de disponibilidad y desempeño.

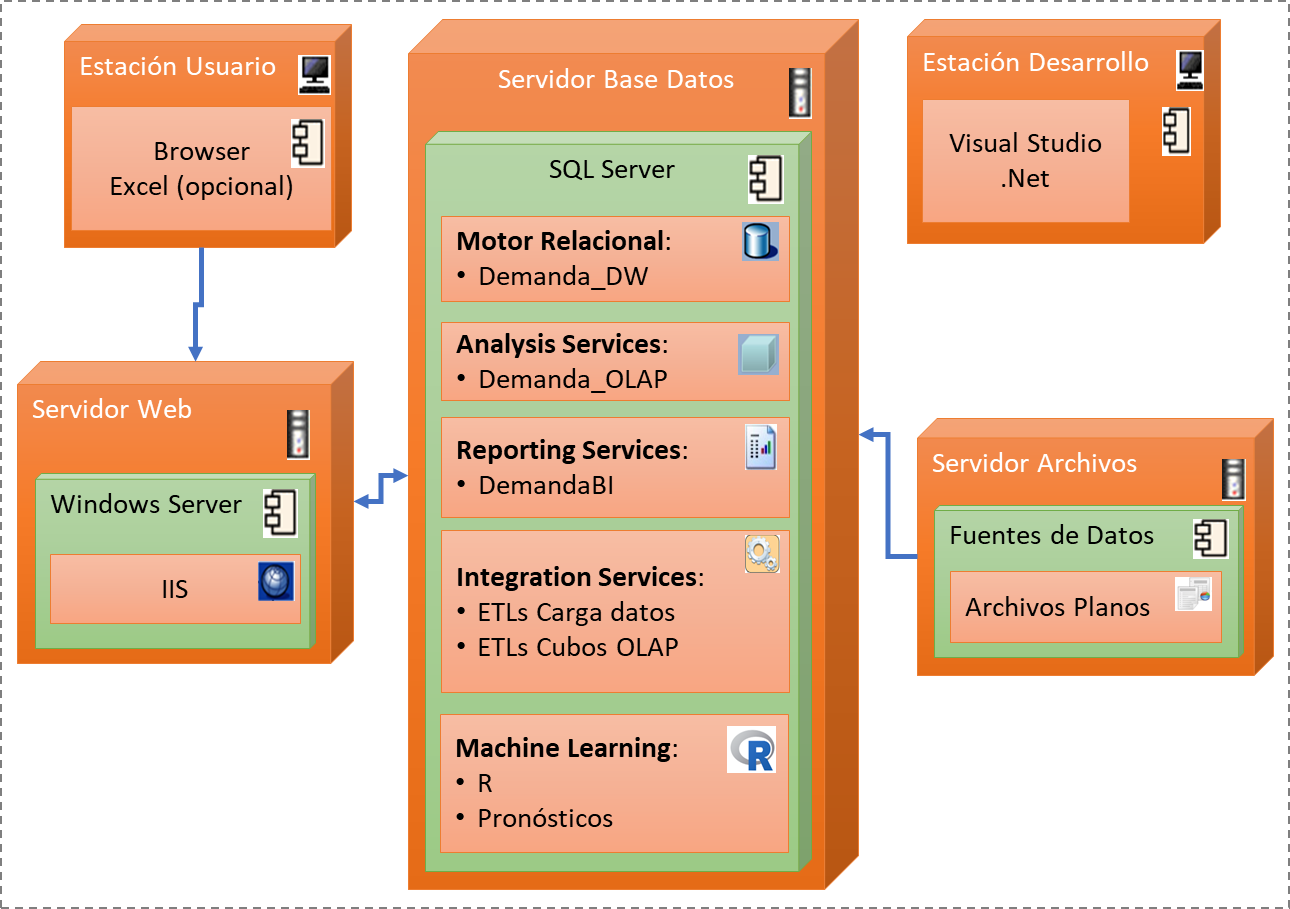


Figura 12. Arquitectura de despliegue

Fuente: Autor

# DATA WAREHOUSE

## Scripts de base de datos

Un DW es una base de datos, pero para BI, la base de datos es diseñada como “**modelo estrella**”. En los diccionarios de datos puede consultar las tablas y sus campos.

La Tabla 5 muestra los proyectos de la solución de base de datos creada con SSMS. Los scripts para crear los objetos de base de datos de este proyecto están en una solución de base de datos creada con SQL Server Management Studio (SSMS), la cual contiene proyectos que agrupan los scripts por temas, esta solución está en la ruta:

\DemandaBI\Src\Script\DemandaDB\DemandaDB.ssmssln

Tabla 5. Proyectos de base de datos de la solución SSMS

|  |  |
| --- | --- |
| **Proyecto** | **Descripción** |
| A00Deployment | Variables para instalar la base de datos utilizando Power Shell |
| A01Uti\_DB | Scripts para crear objetos de base de datos de utilidades y bitácora. |
| A02DemandaDW\_DB | Scripts para crea objetos de base de datos para el DW y SA |
| A03SSISDB | Scripts para configurar las ETLs. |
| B01Uti\_Data | Scripts para insertar datos iniciales a tablas de utilidades y bitácora. |
| B02DemandaDW\_Data | Scripts para insertar datos iniciales a tablas del DW y SA |
| C01Uti\_Security | Scripts para manejo de seguridad en objetos de base de datos de utilidades y bitácora. |
| C02DemandaDW\_Security | Scripts para manejo de seguridad en objetos de base de datos DW y SA. |
| C03SSISDB\_Security | Scripts para manejo de seguridad en objetos de base de datos de ETLs |
| D01SQLAgent | Scripts para el manejo de Jobs encargado de ejecutar tareas automáticas. |
| Z01Test | Scripts de pruebas para la base de datos relacional DW |
| Z02Test | Scripts de pruebas para la base de datos multidimensional (cubos OLAP) |

## Instalación de base de datos

Para instalar la base de datos y los otros objetos del sistema consulte el Anexo 2 manual de instalación. En el Anexo 4 encuentra sitios de descarga de SQL y SSMS, la versión recomendada para aprendizaje y desarrollo de SQL Server es SQL Server Developer, la cual es de “uso libre” (consulte Microsoft para más detalle de las condiciones de licenciamiento).

## Esquemas de base de datos

La Tabla 6 muestra los esquemas de base de datos en los cuales fueron agrupados los objetos de base de datos del DW.

Tabla 6. Esquemas de bases de datos

|  |  |
| --- | --- |
| **Esquema** | **Descripción:** |
| Audit | Objetos de bitácora de procesos |
| DW | Objetos de la bodega de datos DW |
| Staging | Objetos del Staging Area para la carga de datos |
| Utility | Objetos utilidades |

## Modelo estrella

La Figura 13 muestra el modelo de datos de demanda y pérdidas de energía. En la tecnología de BI son muy utilizados la representación estrella donde la tabla de hechos (Fact) es el centro de la estrella y las tablas de dimensiones son las puntas de la estrella.

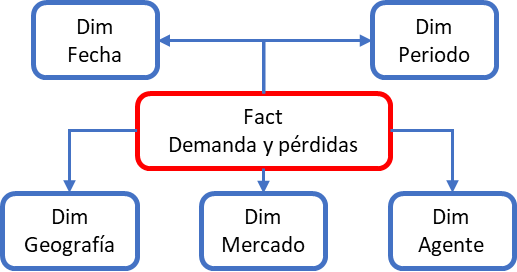


Figura 13. Modelo estrella demanda y pérdidas de energía

Fuente: Autor

## Desempeño de base de datos

Para el diseño de base de datos y dimensionamiento de recursos de hardware necesitamos información de tamaño y tendencia de crecimiento de las tablas de datos. En particular, los BI son sistemas que manejan muchos datos, por ejemplo, la tabla de demanda y pérdidas de la empresa XM, tiene más de 2,000 millones de filas en 20 años. Para un buen desempeño fueron creados índices en las tablas de la base de datos, pero lo más importante fue la creación de un esquema de particionamiento para tablas, en los scripts de base de datos encuentra este esquema para partir la tabla de hechos. Para esta tabla fue definido un esquema de particionamiento por mes. Las tablas de dimensiones son de menor tamaño y no requieren particiones, para estas fueron creado algunos índices.

El particionamiento de tablas es una funcionalidad de algunos motores de base de datos, para nuestro caso SQL Server, para definir las particiones de la tabla Fact del DW fue creado una función de partición y dos esquemas de partición. En los scripts del sistema puede consultar estos objetos.

* Función de partición: Son divisiones para los datos. El proveedor de base de datos establece unas recomendaciones del tamaño máximo de filas o de espacio físico en disco que debe tener una partición. Para nuestro sistema fue definido el tamaño de particiones en forma mensual.
* Esquema de partición: Mapea cada una de las divisiones de la función de partición con un lugar de almacenamiento físico en disco.

La Figura 14 muestra la función de partición y los esquemas de partición para la tabla Fact.

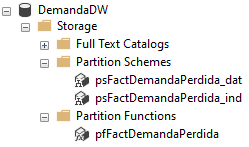


Figura 14. Función de partición y esquemas de partición

Fuente: Autor

## MODELADO DE TABLAS DE DIMENSIONES

En BI existen tipos de dimensiones: 1, 2, 3 y 6 (Kimball, 2013). En este proyecto fueron creadas dimensiones tipo 1 y tipo 2, que son las más comunes.

* Tipo 1 “sobrescribir”: Utilizado cuando hay cambios en la fuente de datos y en el DW los datos son actualizados sin conservar la historia.
* Tipo 2 “añadir fila”: Utilizado cuando hay un cambio en la fuente de datos y en el DW es adicionada una nueva fila para conservar la historia de los datos.

La Figura 15 es un ejemplo de modelado de tablas de dimensiones como tipo 1 y tipo 2, note que las dimensiones tipo 2 tiene campos de fecha, para almacenar los cambios históricos de los datos maestros.

|  |  |
| --- | --- |
|  |  |

Figura 15. A la izquierda dimensión tipo 1, a la derecha dimensión tipo 2

Fuente: Autor

# BITÁCORA DE PROCESOS

La calidad de datos, las fallas en los procesos de carga son algunos de los desafíos que tiene que afrontar un BI. No solo para un BI, sino para otros tipos de sistemas es necesario contar con una bitácora para realizar seguimiento a la ejecución de procesos para poder resolver los problemas. Para el BI fue creado un conjunto de tablas, procedimientos almacenados y reportes para realizar seguimiento a las tareas, adicional hay otras herramientas para consultar el estado del sistema. A continuación, será descrito como consultar el log de procesos. En el Anexo 3 manual de operación puede consultar los aspectos técnicos para consultar el log de procesos de este proyecto.

## Logs de carga de datos

A continuación, son explicados los diferentes tipos de registro (log) utilizadas en el proyecto.

### Auditoria de ETLs

Las ETLs dejan log de las tareas en los siguientes lugares:

* Tablas personalizadas de auditoría
* Archivo plano CSV, en disco

La auditoría es almacenada en las siguientes tablas:

* Bitácora: Información general de ejecución
* Bitácora archivo: Información de archivos procesados
* Bitácora tabla: Información de inconsistencias en filas de tabla procesada
* Bitácora estadística: Cuenta filas procesadas
* Bitácora detalle: Detalles de ejecución

Estos son algunos de los datos almacenado en las tablas de bitácora:

* Fecha y hora de inicio de ejecución
* Fecha y hora de finalización de ejecución
* Estado de ejecución: Éxito, fallo, en ejecución, cancelado, etc.
* Datos de los archivos procesados: Ruta, nombres, etc.
* Tarea que presenta fallos
* Mensaje de error

### Auditoría de Jobs

Para la tecnología utilizada Microsoft SQL Server, las tareas automáticas son ejecutadas por Jobs SQL, para el proyecto fue creado un job para ejecutar las ETLs del sistema. Los Jobs tienen un historial de ejecución, el cual es un punto inicial para detectar fallas de ejecución.

### Reporte de ejecución de ETLs

Microsoft SQL Server también cuenta con un reporte que muestra todas las actividades realzada por la ETL y los mensajes de error en caso de fallos. La Figura 16 muestra el reporte propio SQL para revisar el estado de ejecución de ETLs.

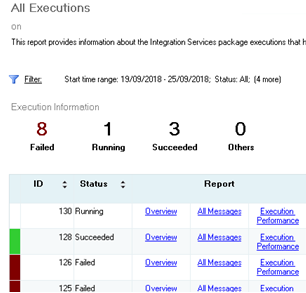


Figura 16. Reporte propio de SQL de ejecución de ETLs

Fuente: Autor

### Reportes de bitácora

Para consultar las tablas de bitácora de procesos fueron creados reportes SQL Reporting Services, la Figura 17 muestra uno de los reportes de bitácora, consultado desde la aplicación web.

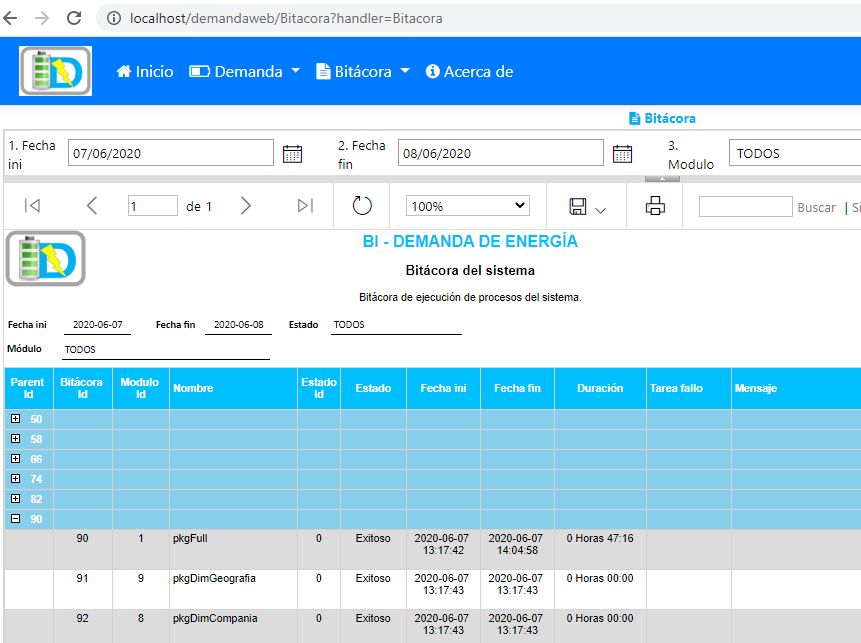


Figura 17. Reporte de bitácora del sistema

Fuente: Autor

## Logs SQL Server

A nivel de bases de datos relacional y multidimensional y de reportes, Microsoft SQL cuenta con archivos de log en los cuales registra fallos, estos logs son del dominio de los administradores de bases de datos.

# ETL PARA GENERAR ARCHIVOS PLANOS

La empresa XM facilitó los datos de su base de datos de BI, en el Anexo 1 están las consultas de selección de datos. Para generar los archivos planos fue creada una ETL que retorna un archivo para cada dimensión, mientras que para la tabla de hechos genera un archivo con datos para un cada día (organizados en carpetas por mes y año). La bodega de datos de XM, tiene datos desde el año 2000, en 20 años son alrededor de 24 GB, 7,300 archivos, 300 carpetas y 2,000 millones de filas. Cada archivo tiene alrededor de 2,200 KB y 23,500 filas. La Figura 18 muestra la estructura de carpetas con los archivos para la carga de datos del DW.

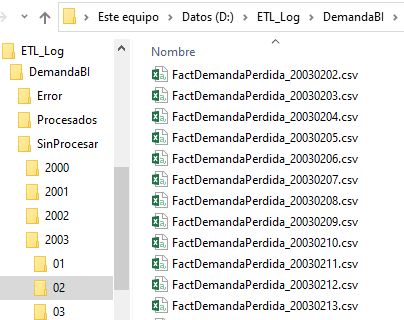


Figura 18. Carpetas para la carga de datos

Fuente: Autor

**NOTA**: Por el volumen, solo una muestra está en la documentación. Pero puede obtener más datos de demanda y pérdidas de energía del portal BI de la empresa XM, ya que está información es pública.

La Figura 19 muestra la ETL para generar los archivos planos, la cual fue construida con SQL Server Data Tools (SSDT) para Visual Studio 2017 (también puede ser abierta con SSDT para 2019), al abrir solicita una contraseña (en minúsculas): **demandabi**. Las fuentes de ésta ETL están en la ruta:

\DemandaBI\Src\DemandaBI\_Datos\_SSIS\DemandaBI\_Datos\_SSIS.sln

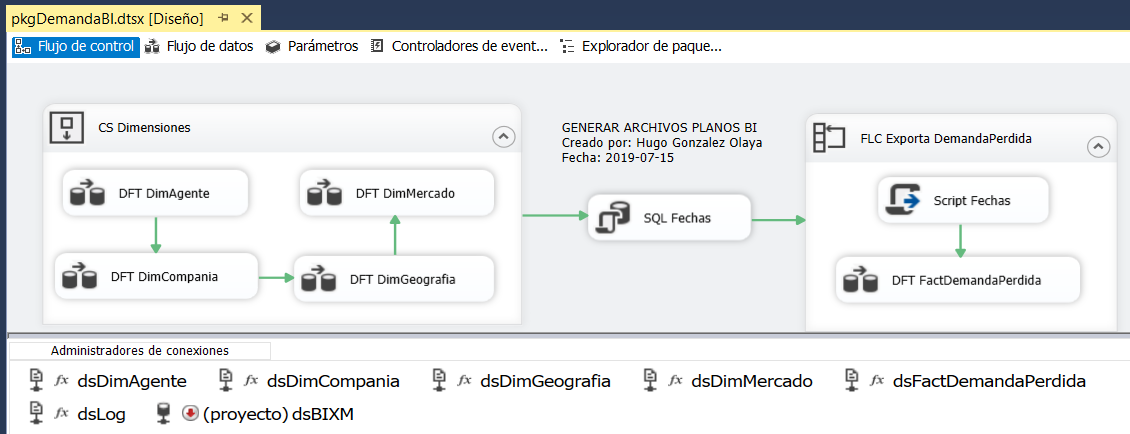


Figura 19. ETL para generar archivos planos

Fuente: Autor

# ETL PARA CARGA DEL DATA WAREHOUSE

## ETL DEMANDA BI

Las ETLs son una de los elementos más importantes para un BI, ellas alimentan el sistema con los datos de los sistemas transaccionales y otras fuentes de datos. Las ETLs fueron construidas con SQL Server Data Tools (SSDT) para Visual Studio 2017 (también puede ser abierta con los SSDT para 2019), al abrir solicita una contraseña (en minúsculas): **demandabi**. La Tabla 7 muestra las ETLs para la carga de datos y procesamiento de los cubos OLAP del sistema. En el anexo 4 encuentra links donde puede descargar las herramientas. Las fuentes de las ETLs del sistema están en la ruta:

\DemandaBI\Src\DemandaBI\_SSIS\DemandaBI\_SSIS.sln

Tabla 7. ETLs del proyecto

|  |  |
| --- | --- |
| **ETL** | **Descripción** |
| pkgCubeDimProcess.dtsx | Procesa dimensiones de la base de datos multidimensional |
| pkgCubePartitionCreate.dtsx | Crea particiones para el cubo OLAP |
| pkgCubePartitionProcess.dtsx | Procesa particiones para el cubo OLAP |
| pkgDimAgente.dtsx | Carga datos a la tabla: DW. DimAgente |
| pkgDimCompania.dtsx | Carga datos a la tabla: DW.DimCompania |
| pkgDimGeografia.dtsx | Carga datos a la tabla: DW.DimGeografia |
| pkgFactDemandaPerdida.dtsx | Carga datos a la tabla: DW.FactDemandaPerdida |
| pkgFull.dtsx | Ejecuta todas las ETLs |

El diseño de las ETLs es muy similar, la secuencia de procesos es: Auditoría inicio, configurar rutas de archivos, truncar tabla temporal (tabla SA), cargar tabla temporal, ejecutar procedimiento almacenado que copia los datos de la temporal a la tabla del DW, mover archivo a la carpeta de archivos procesados o a la carpeta de archivos con error cuando hay fallos y termina con auditoría fin. Las tablas de dimensiones utilizan un solo archivo, pero la tabla Fact, requiere una programación de carga con los grupos de archivos a procesar según el rango de fechas establecidos. La ETL full, fue creada por facilidad, tiene la secuencia y ejecución de todos los procesos.

La Figura 20 es una de las ETLs para cargar tabla de dimensión, solo utiliza un archivo como fuente de datos el cual es movido al final de la ejecución.

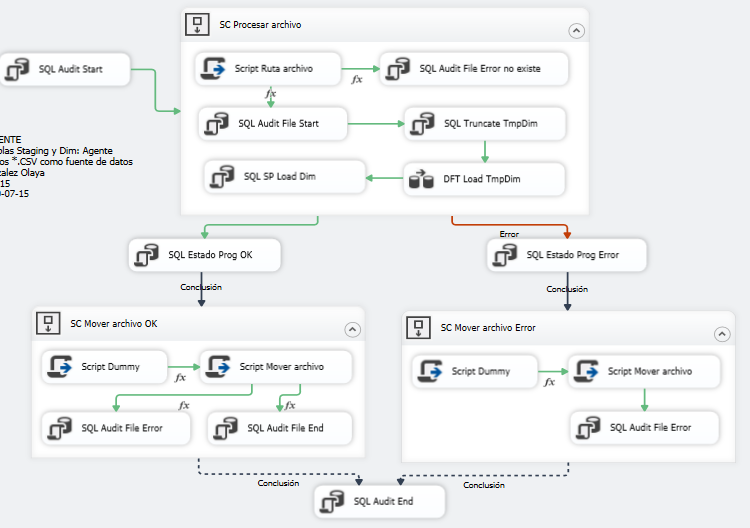


Figura 20. ETL para carga de tabla de dimensión

Fuente: Autor

La Figura 21 muestra la ETL para cargar la tabla fact que almacena los datos de demanda y pérdidas de energía, esta ETL contiene un ciclo para carga de datos por rangos de fechas, en este proyecto cada ciclo procesa un día de datos.

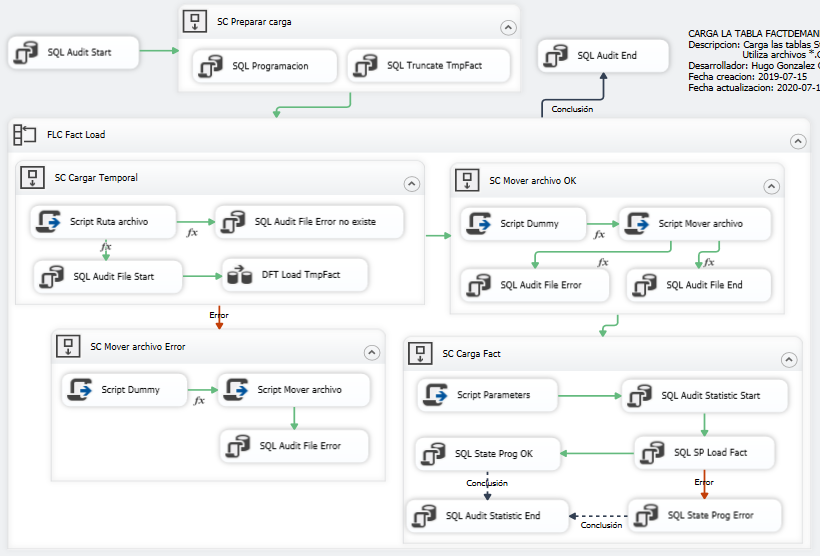


Figura 21. ETL para carga de tabla fact de demanda y pérdidas

Fuente: Autor

La Figura 22 presenta la ETL para procesar todas las dimensiones de la base de datos multidimensional, la tarea es ejecutado por código creado en el lenguaje C#.

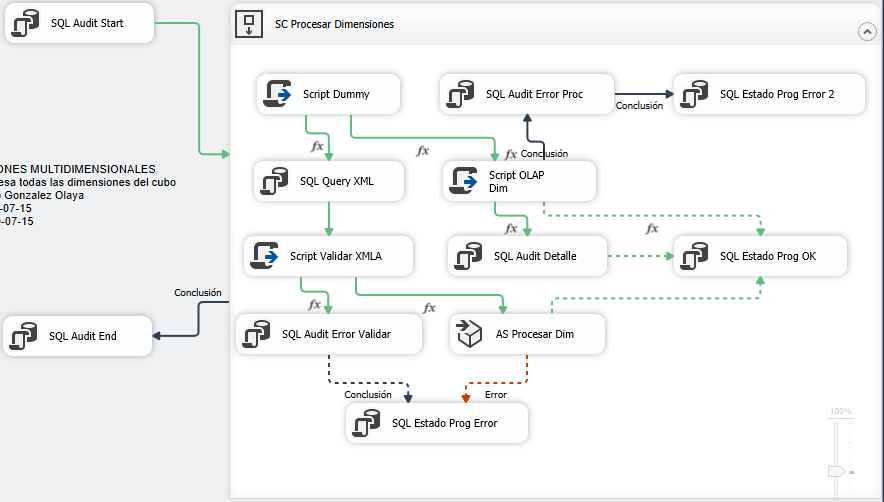


Figura 22. ETL para procesar dimensiones de la base de datos multidimensional

Fuente: Autor

La Figura 23 presenta la ETL para crear particiones para los cubos OLAP, la tarea es ejecutado por código creado en el lenguaje C#.

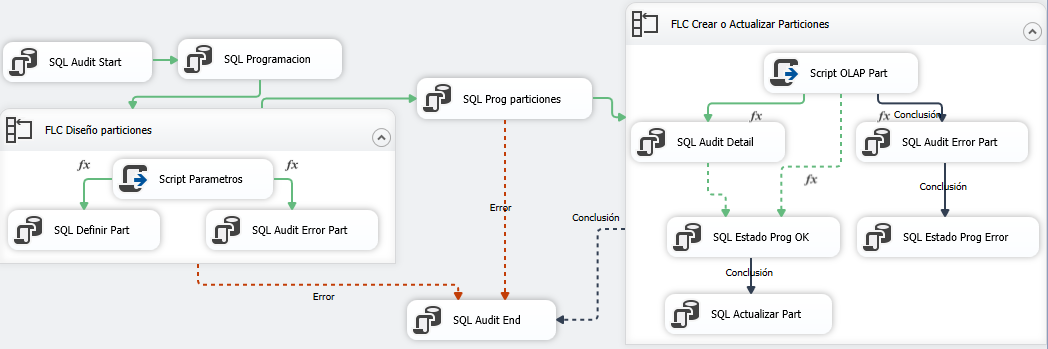


Figura 23. ETL para crear particiones para cubos

Fuente: Autor

La Figura 24 presenta la ETL para procesar particiones para los cubos OLAP, la tarea es ejecutado por código creado en el lenguaje C#.

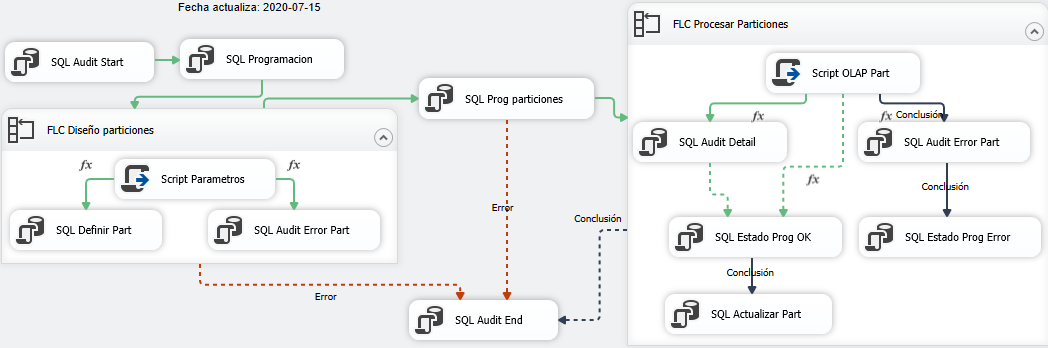


Figura 24. ETL para procesar particiones de cubos

Fuente: Autor

La Figura 25 presenta la ETL que ejecuta todas las ETLs en el orden apropiado para carga del DW, también puede ser utilizada para la recarga de datos históricos, pero antes de ejecutar debe ser establecido el rango de fechas a procesar.

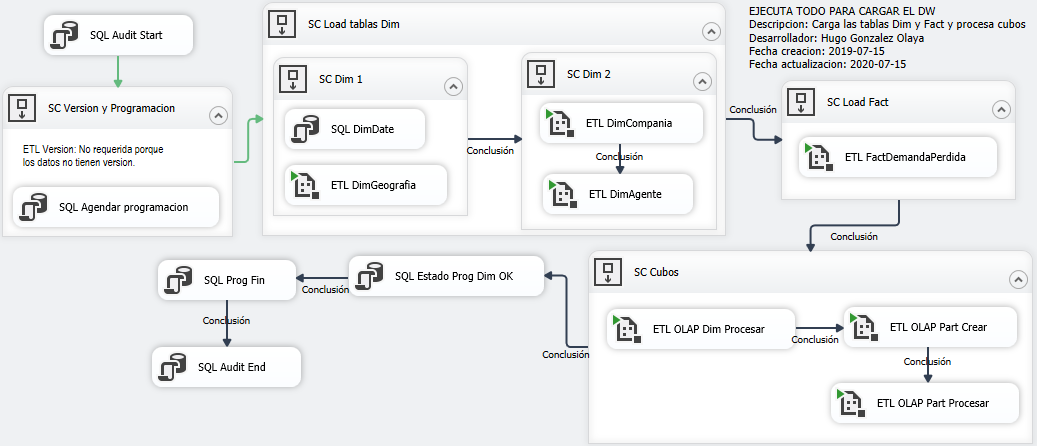


Figura 25. ETL completa ejecuta todo para cargar el DW

Fuente: Autor

## INSTALACIÓN Y CONFIGURACIÓN DE ETLS

Consulte el Anexo 2 manual de instalación y Anexo 3 manual de operación para mayor detalle. En el anexo 4 encuentra sitios de descarga de las herramientas como SSDT para la creación de ETLs la cual es de uso libre.

# CUBOS OLAP

## CUBOS DEMANDA BI

Una pregunta frecuenta en un BI es: ¿Si ya están los datos en la base de datos relacional: “bodega de datos” (DW), para qué crear cubos?

Parece como si la construcción de los cubos fuera redundante y que no aporta valor, puesto que el modelo estrella del DW es similar al de los cubos OLAP. Pero los cubos son necesario porque el volumen de datos es muy alto y la velocidad de respuesta de las consultas es mucho mayor en los cubos OLAP, ya que estas estructuras sumarizan y pre-calculan los totales para un rápido consumo.

Para la tecnología Microsoft existen dos tipos de cubos: multidimensionales y tabulares. En el proyecto fueron utilizados los multidimensionales, puesto que dan mejor desempeño para mayores volúmenes de datos, adicional los tabulares son más exigentes en hardware y faltan algunas funcionalidades que existen en los cubos multidimensionales.

La Figura 26 muestra la base de datos multidimensional creada con SQL Analysis Services. Los cubos fueron construidos con SQL Server Data Tools (SSDT) para Visual Studio 2017 (la solución también puede ser abierta con los SSDT para 2019). En el anexo 4 encuentra links donde puede descargar las herramientas. Las fuentes del proyecto de cubos OLAP está en la ruta:

\DemandaBI\Src\Demanda\_OLAP\Demanda\_OLAP.sln

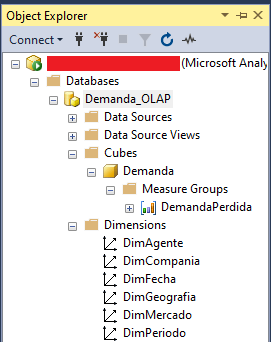


Figura 26. Base de datos multidimensional Demanda\_OLAP

Fuente: Autor

## DISEÑO DE JERARQUÍAS PARA DIMENSIONES

La Figura 27 muestra la dimensión fecha en la herramienta de desarrollo. Las dimensiones tienen jerarquías o niveles, las cuales mejoran el desempeño de consultas y la experiencia de usuario.

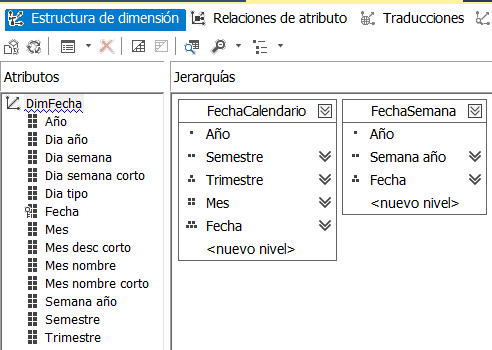


Figura 27. Jerarquías para la dimensión fecha

Fuente: Autor

## CUBO DEMANDA

Un cubo contiene uno o varios grupos de medidas y cada grupo de medidas está relacionado con una tabla Fact del DW. Un grupo de medidas contiene un conjunto de métricas. Las métricas pueden ser naturales (sumarizan campos de la tabla Fact) o calculadas (expresiones para obtener un valor), el proyecto contiene dos métricas naturales demanda y pérdidas, las cuales totalizan los campos de demanda y pérdida, y fue creada una métrica para contar filas, también fueron creadas algunas métricas calculadas a modo de ilustración. En la Figura 28 puede ver las métricas del cubo demanda y la relación entre el grupo de medidas y las dimensiones.

|  |  |
| --- | --- |
|  |  |

Figura 28. Cubo demanda. A la izquierda métricas naturales. A la derecha relación del cubo con dimensiones

Fuente: Autor

## NOMBRAMIENTOS DE OBJETOS

Para asignar nombres a los objetos en bases de datos relacionales existen estándares para un buen mantenimiento. En bases de datos multidimensionales el nombramiento es todavía más importante, puesto que los usuarios ven en forma directa los nombres de las métricas y dimensiones a ser consultadas. Estos nombres deben ser claros para el usuario. Para este sistema la palabra demanda puede pertenecer a un grupo de medidas de energía, transacciones en bolsa, impuestos, etc. En la figura de arriba, note como las métricas del grupo de medidas de demanda inician con Dem\_, esto da claridad al usuario a que modelo pertenece las medidas.

## DESEMPEÑO EN CUBOS olap

En los cubos el desempeño está generalmente en dos lugares: el procesamiento (cuando los datos del DW son cargados a dimensiones y cubos) y en las consultas de los usuarios. A continuación, algunas estrategias para mejorar el desempeño (no todas fueron implementadas).

### Particiones en cubos

Generalmente el mismo esquema de particionamiento en la base de datos relacional (DW), es definido para los cubos OLAP, para nuestro caso el cubo fue partido por mes. Las particiones son rebanadas de datos. La Figura 29 muestra el diseño de particiones en la herramienta de desarrollo y la Figura 30 muestra las particiones creadas en la base de datos.

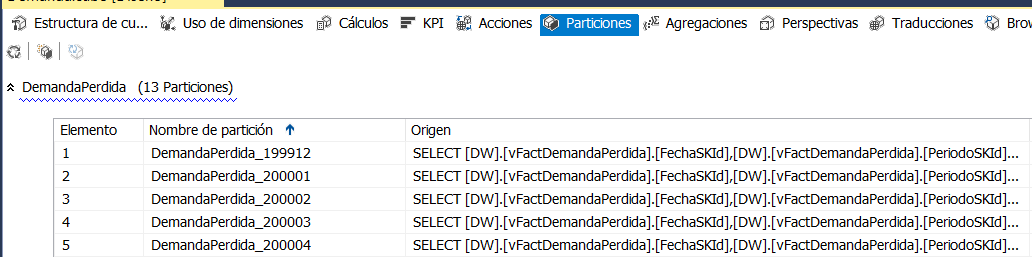


Figura 29. Diseño de particiones para el cubo demanda

Fuente: Autor

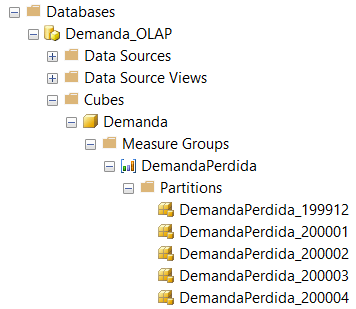


Figura 30. Particiones implementadas para el cubo demanda

Fuente: Autor

La creación de particiones en cubos es diferente a la de las bases de datos relaciones. Las particiones son creadas estableciendo un filtro a la consulta que selecciona los datos de la base de datos relacional DW. Cada partición va desde una fecha inicio hasta una fecha fin. La Figura 31 es un fragmento de la instrucción XMLA para crear una partición, note que en la instrucción de selección de datos los datos son filtrados desde una fecha inicio hasta una fecha fin, las anteriores y siguientes particiones deben ser consistentes, de lo contrario el cubo puede tener datos duplicados o datos ausentes.

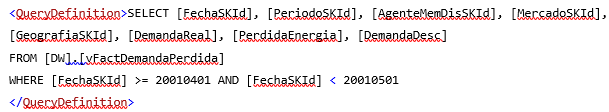


Figura 31. Definición de consulta para crear particiones en cubos

Fuente: Autor

### Agregados en cubos

Los agregados es otra estrategia para mejorar el desempeño de las consultas de los usuarios. Son cálculos almacenados en disco para mejorar el desempeño de consultas. En este proyecto no fueron creados.

### Modo de almacenamiento

Los cubos tienen tres formas de almacenamiento ROLAP, MOLAP y HOLAP. En este proyecto fue utilizado el modo de almacenamiento MOLAP, lo que significa que los datos son almacenados en disco. MOLAP fue seleccionado porque da mejor desempeño para las consultas de los usuarios.

## INSTALACIÓN Y CONFIGURACIÓN DE CUBOS

Consulte el Anexo 2 manual de instalación y Anexo 3 manual de operación para mayor detalle. En el anexo 4 encuentra sitios de descarga de las herramientas como SSDT para la creación de cubos la cual es de uso libre.

# HERRAMIENTA CLIENTE DE REPORTES

## HERRAMIENTAS DE REPORTES

Para la creación de reportes fueron analizadas dos alternativas: Microsoft SQL Reporting Services (SSRS) y Microsoft Power BI. Los reportes Power BI ofrecen mejor experiencia de usuario, pero tienen el inconveniente de costo e implementación en la nube, adicional, la versión gratuita presenta algunas limitantes para los reportes que utilizan como fuente de datos cubos. Power BI tiene diferentes tarifas dependiendo de los recursos de procesamientos, la menor tarifa, para el año 2019 es de unos U$ 800 / mes, es decir alrededor de $ COP 2,400,000 / mes (a la tasa cambio del dólar para ese año). Power BI está más enfocado a dashboard y reportes gráficos, pero no es bueno para reportes paginados. SSRS tiene como ventajas que no tiene costo adicional, viene incluido dentro de las licencias SQL, existen versiones libres, pero es conveniente consultar la literatura para establecer las condiciones de uso. La instalación y configuración es fácil y permite crear dashboard, reportes gráficos y paginados. SSRS fue la opción seleccionada para reportes.

Los reportes fueron construidos con SQL Server Data Tools (SSDT) para Visual Studio 2017 (la solución también puede ser abierta con SSDT para 2019). En el Anexo 4 encuentra links donde puede descargar las herramientas.

Fueron creado dos proyectos de reportes, uno para reportes de bitácora y otro para reportes del sistema. Las fuentes están en las rutas:

* \DemandaBI\Src\Auditoria\_RS\Auditoria\_RS.sln
* \DemandaBI\Src\DemandaBI\_RS\DemandaBI\_RS.sln

Los reportes de auditoría utilizan como fuente de datos las tablas de bitácora del DW, mientras que los reportes del sistema utilizan los datos del cubo OLAP, para consultar los cubos debe ser utilizado el lenguaje de consultas MDX (lenguaje de consultas multidimensionales).

## OTRAS HERRAMIENTAS PARA REPORTES

Los reportes SSRS también pueden ser construidos con Microsoft Report Builder, la cual es una herramienta pensada en usuario final. Otra herramienta es Microsoft Excel la cual permite crear reportes utilizando tablas dinámicas o Power Pivot, un complemento que viene con Excel. Estas herramientas es lo que en BI es denominado auto servicio, es decir el usuario final puede crear sus propios reportes sin requerir de personal de TI para desarrollo.

## REPORTE TIPO DASHBOARD

Un dashboard o tablero de control, es un reporte que presenta KPI (Key Performace Indicator) o indicadores de desempeño del sistema. Los tableros presentan información agrupada, con elementos gráficos y/o números con medidas del sistema.

La Figura 32 muestra la página de inicio de la aplicación, la cual contiene dashboard con KPIs. En este proyecto el dashboard presenta datos para la última fecha que tenga datos los cubos OLAP, esta fecha aparece en la parte inferior del “tablerito”. Por ejemplo, el tablero “Demanda de energía real” presenta el valor actual para esta fecha, el valor del día anterior y un mini gráfico (spark line) con datos de 30 días atrás.



Figura 32. Reporte tipo dashboard con KPIs

Fuente: Autor

## REPORTES GRÁFICOS

Para el sistema fueron creados tres reportes gráficos. Cada reporte consta de un gráfico y una tabla de datos, puede establecer parámetros de consulta como la fecha. Los reportes pueden ser exportados a Word, Excel, PDF y otros formatos. Para mayor detalle de cómo utilizar la aplicación y ver los reportes, consulte el Anexo 3 manual de operación.

La Figura 33 muestra el reporte de demanda real, demanda comercial y pérdidas de energía, este reporte presenta los datos en forma anual, puede seleccionar el año y el número de años hacia atrás en el cual consultar los datos, a mano derecha tiene una tabla con los datos del gráfico.

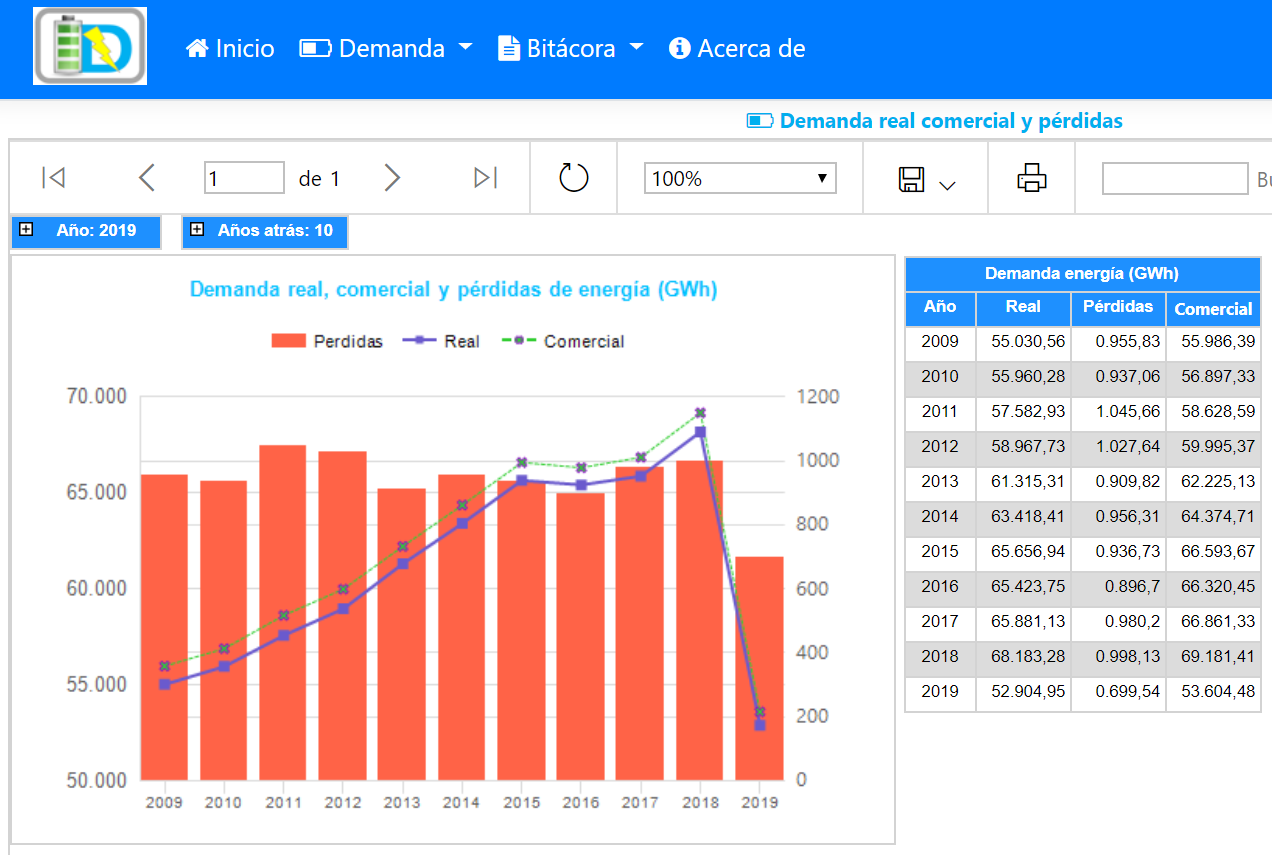


Figura 33. Reporte demanda real, comercial y pérdidas

Fuente: Autor

La Figura 34 presenta la demanda comercial para el agente seleccionado, el área representa el valor total y las barras son los valores para los tipos de mercado regulado y no regulado, también existen otros tipos de mercado, pero el valor es muy pequeño y no es presentado en esta figura. También tiene una tabla a la derecha con los datos del gráfico.

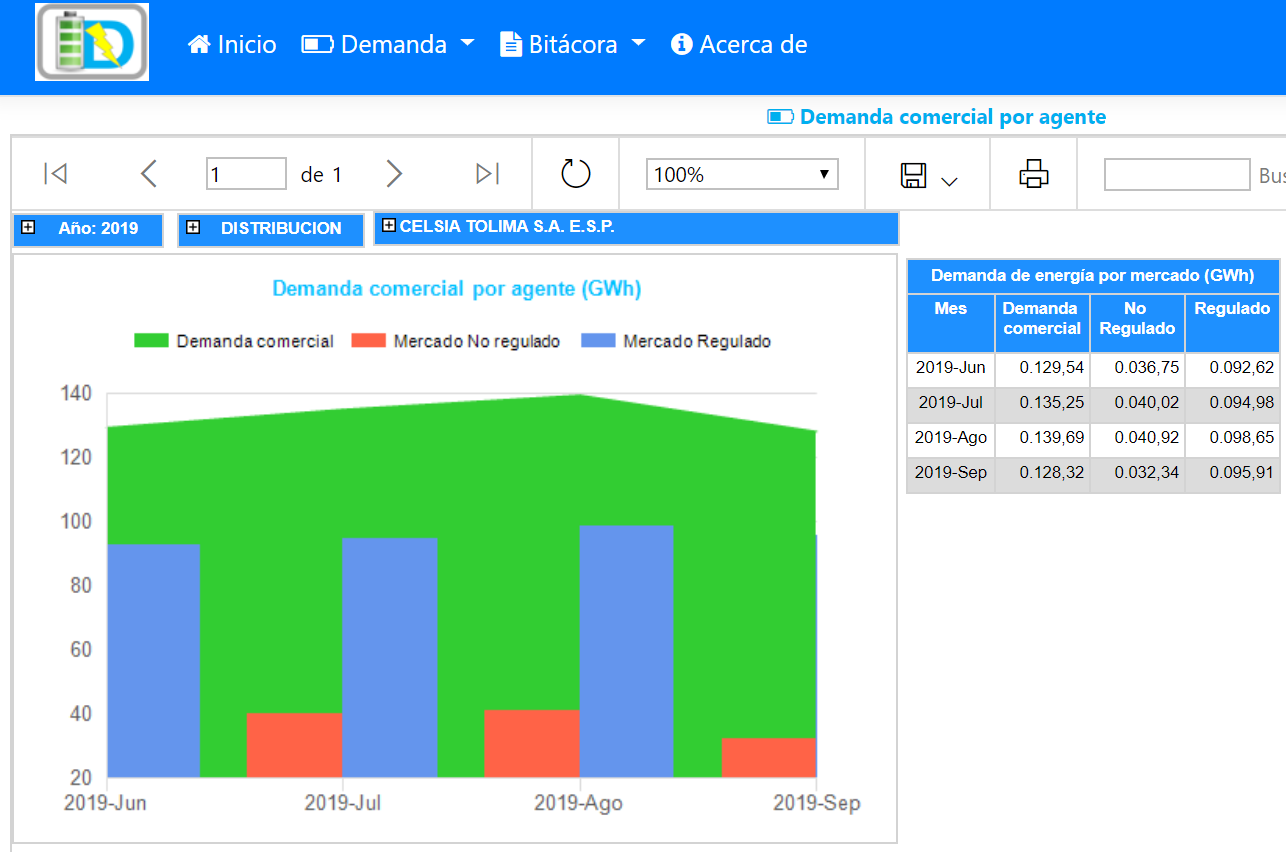


Figura 34. Reporte demanda comercial por agente

Fuente: Autor

La Figura 35 presenta un reporte de demanda comercial por departamento para el mes seleccionado. Debajo del gráfico hay una tabla con los datos del reporte.

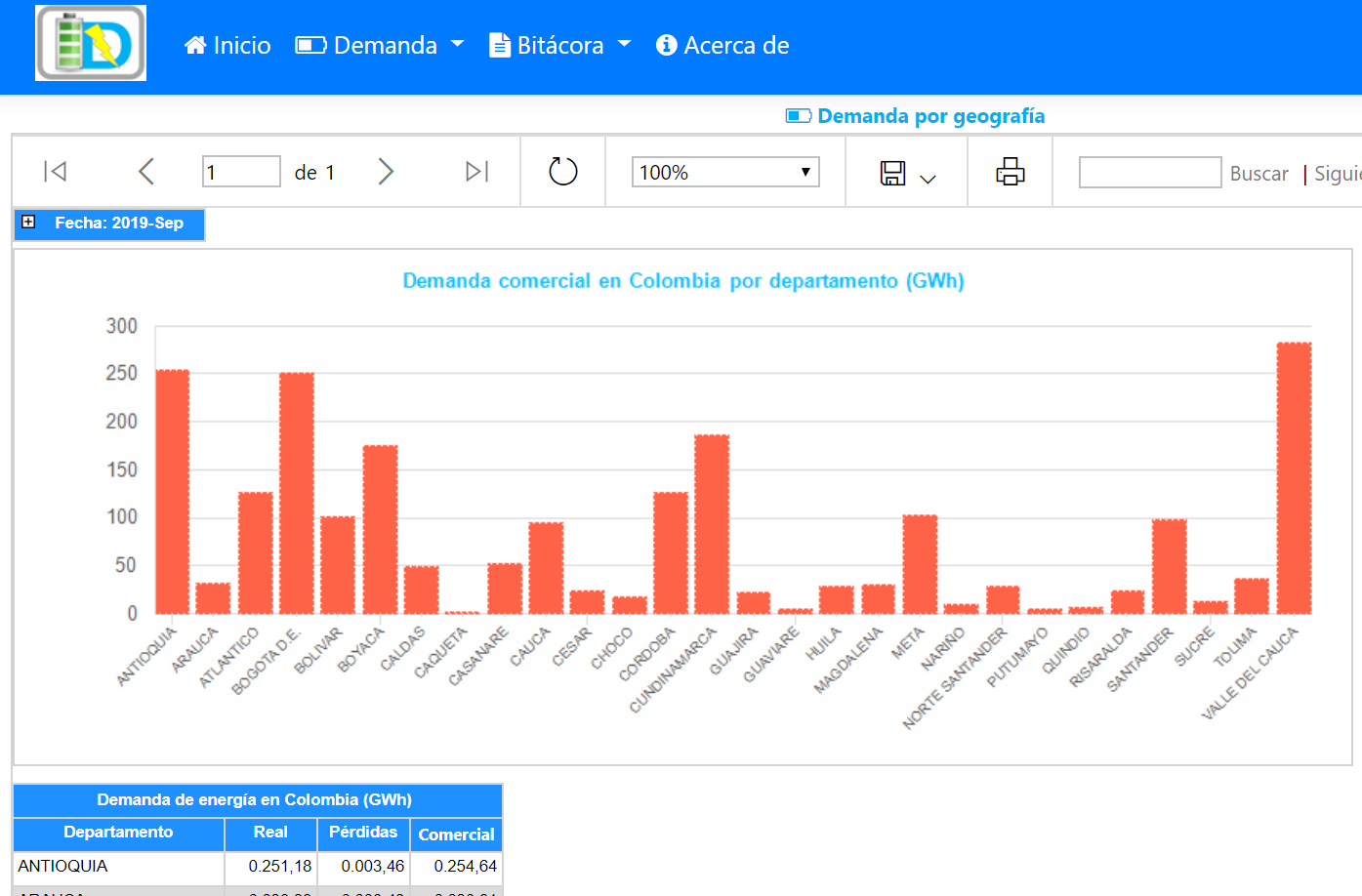


Figura 35. Reporte demanda por geografía

Fuente: Autor

## REPORTES PÁGINADOS

La Figura 36 muestra uno de los reportes paginados, en el proyecto fueron creados cinco reportes paginados para consultar la bitácora con el estado de ejecución de las cargas de datos y otros procesos del sistema. Los reportes tienen parámetros de consulta y pueden ser exportados a diferentes tipos de formatos. Los reportes creados son:

* Bitácora del sistema: Presenta el estado de ejecución de procesos.
* Bitácora tabla: Presenta las filas con inconsistencias o advertencias en la carga de datos.
* Bitácora archivos: Presenta el estado de archivos utilizados en la carga de datos.
* Bitácora estadística: Presenta datos de filas procesadas.
* Bitácora detalle: Presenta información de algunos eventos ocurridos en la carga.

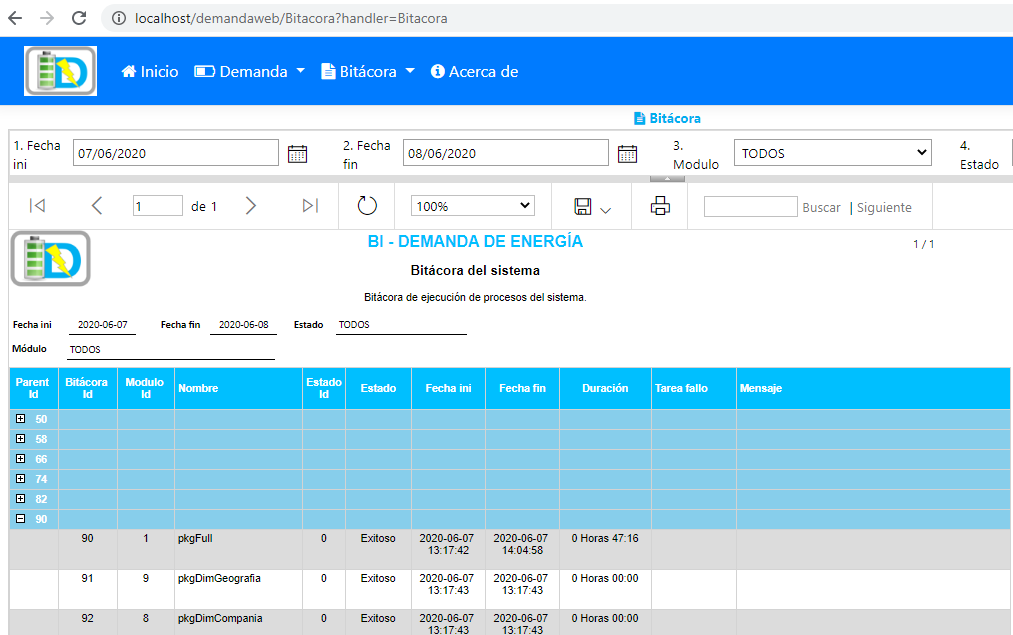


Figura 36. Reporte paginado de bitácora

Fuente: Autor

## ¿CÓMO CONSULTAR LOS REPORTES?

Los reportes pueden ser consultados dese el servidor de reportes Microsoft SSRS o en la aplicación Web. Las figuras anteriores muestran los reportes dentro de la aplicación, la Figura 37 muestra el servidor de reportes, desde el cual puede ver los reportes. Para ingresar al servidor de reportes escriba en un browser la ruta del servidor de informes, la cual es similar a las siguientes, si no está seguro consulte la configuración del SSRS instalado en su ambiente. Para mayor de detalle consulte el Anexo 2 manual de instalación y Anexo 1 manual de operación.

<http://MiServidor/Reports>

<http://localhost/Reports>

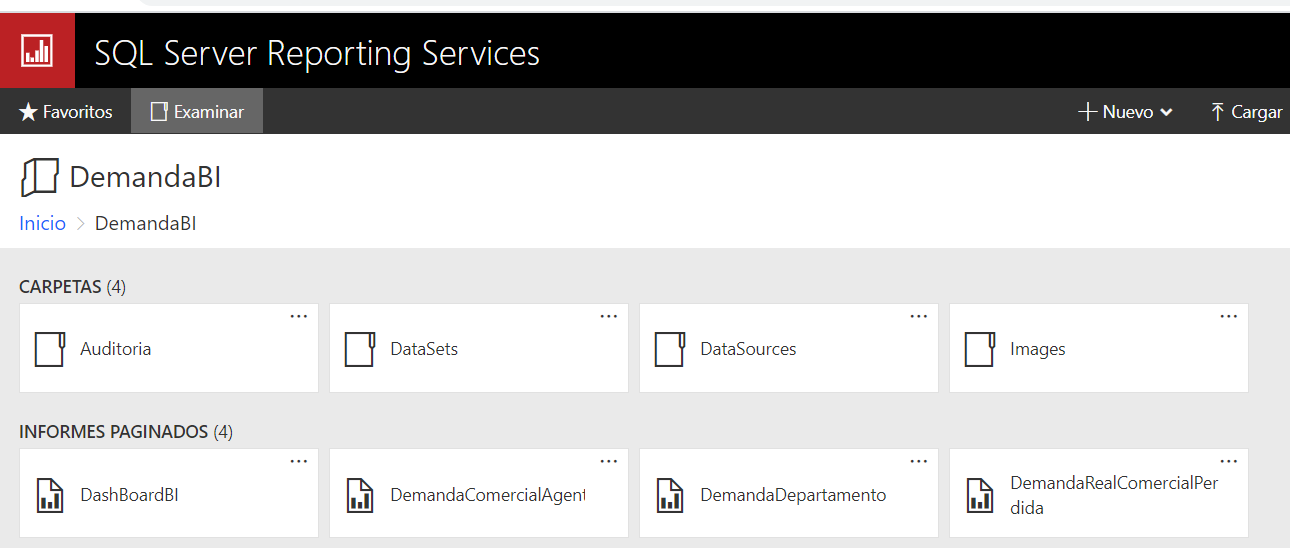


Figura 37. Servidor SQL Reporting Services

Fuente: Autor

## DISEÑO DE REPORTES

La construcción de reportes involucra aspectos estéticos, al crear reportes no solo debe tener en cuenta el tamaño de pantalla, también es conveniente la forma de impresión. A continuación, algunos aspectos a tener en cuenta para la construcción de reportes:

* Consultar con el usuario si tiene un estándar de formatos y logotipos.
* Cree un reporte plantilla y validar el aspecto visual con el usuario final.
* Generalmente utilizar un solo tipo de letra.
* Tamaños de letra pequeña, por ejemplo, 12 puntos título, 10 subtítulo, 8 para tablas. Para los gráficos 8 puntos para los ejes y leyenda y 10 – 12 para el título.
* Establezca un tamaño de página y márgenes.

## INSTALACIÓN Y CONFIGURACIÓN DE REPORTES

Para mayor detalle consulte el Anexo 2 manual de instalación y Anexo 3 manual de operación. En el Anexo 4 encuentra sitios de descarga de las herramientas como SSDT para la creación de reportes la cual es de uso libre. Como fue mencionado hay opciones para la creación de reportes como: Report Builder, Power BI y Microsoft Excel.

# HERRAMIENTA CLIENTE APLICACIÓN WEB

## HERRAMIENTA DE DESARROLLO

Para presentar la información de demanda y pérdidas de energía fue creada una aplicación web llamada: DemandaWeb. Fue construidas con Microsoft Visual Studio .Net 2019, la aplicación es de tipo MVC con el framewort .Net Core. También fueron empleados otros frameworsk como Bootstrap, JQuery y JavaScript. La aplicación es solo de consulta y muestra los reportes construido tanto para el DW como para bitácora de procesos. En el Anexo 4 encuentra links donde puede descargar las herramientas. Las fuentes de la aplicación están en la ruta:

\DemandaBI\Src\DemandaWeb\DemandaWeb.sln

## MOSTRAR LOS REPORTES EN LA APLICACIÓN WEB

La aplicación muestra los reportes SSRS, para ver los reportes existe un control llamado ReportViewer, pero este ya no funciona para la nueva versión de .Net Core. En la fase de desarrollo, con las referencias de la literatura fue simulado el comportamiento de este control gráfico y funcionó bien con reportes tipo dashboard y gráficos, pero para los reportes que tienen controles para expandir y comprimir filas de una tabla, muestra las filas de las tablas, pero no permite expandir y comprimir detalles. Por lo tanto, fue desarrollada otra alternativa, que consiste en utilizar el tag HTML: **IFrame**, al cual fueron aplicados estilos para establecer el ancho y alto para expandir el tamaño de los reportes.

## MENÚS DE LA APLICACIÓN

La aplicación tiene dos menús principales: demanda y bitácora, que agrupan los reportes de estos temas. La Figura 38 muestra la página de inicio, ésta página presenta un reporte tipo dashboard con los indicadores de demanda y pérdidas de energía en la última fecha en la cual hay datos. En un BI real cada día son cargados nuevos datos y el reporte mostrará los últimos datos procesados cada que abra la aplicación.

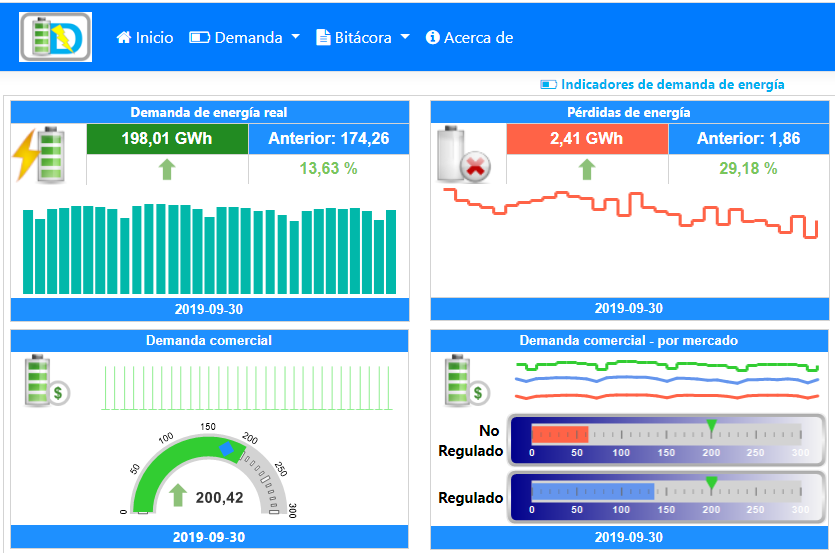


Figura 38. Página de inicio de la aplicación DemandaWeb

Fuente: Autor

## CONFIGURACIÓN DE LA APLICACIÓN

La aplicación tiene pocos parámetros de configuración. Solo debe establecer la ruta del servidor de reporte, las carpetas donde están ubicados y los nombres de los reportes.

## INSTALACIÓN DE LA APLICACIÓN

Para mayor detalle consulte el Anexo 2 manual de instalación y Anexo 3 manual de operación. En el anexo 4 encuentra sitios de descarga de las herramientas de desarrollo.

# RESULTADOS Y ENTREGABLES DEL PROYECTO

## CONSTRUCCIÓN DE UN BI

BI es una de las tecnologías utilizadas por la ingeniería de sistema para dar respuesta a necesidades empresariales de análisis de los datos históricos. El objetivo del proyecto fue la de construir un sistema de BI para analizar la demanda y pérdida de energía eléctrica y este objetivo fue cumplido, el software presenta la información de estas variables y este proyecto contribuye a la generación de conocimiento en BI.

La metodología fue basada en la “arquitectura bus” definida por Kimball para proyectos de BI, fue implementada en el diseño y construcción, el proyecto realiza un recorrido por las diferentes fases para la construcción del sistema, desde establecer las preguntas del negocio, hasta la creación de herramientas cliente para presentar la información.

Fue creado código, manuales de instalación y operación, como un aporte para las personas que consulte este trabajo como material de apoyo en el conocimiento de la tecnología de BI.

## modelo estrella

La construcción de un BI es un trabajo de equipo entre diseñadores y expertos del negocio. Los modelos estrella es una de las metas en el levantamiento de requisitos y con estos modelos son identificadas las fuentes de datos y la representación de la información para los usuarios. En el proyecto fue creado un modelo de demanda y pérdidas de energía.

## SOLUCIÓN DE BI

Los componentes macros construidos para la solución de BI son:

* Data Warehouse: Base de datos con el modelo de demanda y pérdidas de energía, bitácora de procesos, utilidades para la carga de datos y otros objetos de base de datos.
* ETLs: Para la carga de datos y procesamiento de los cubos OLAP
* Cubos OLAP: Estructuras multidimensionales con las métricas del negocio
* Reportes: Muestran la información de demanda y pérdidas de energía, y también fueron construidos reportes de bitácora de procesos
* Sitio Web: Herramienta cliente para los usuarios consultar el sistema BI.

## ANÁLISIS DE LA DEMANDA Y PÉRDIDAS DE ENERGÍA

La demanda y pérdidas de energía son las métricas creadas en este proyecto, su tendencia puede ser consultadas en la aplicación. La página de inicio contiene tableros de control con indicador de estas variables. Sistemas como BI son alimentados, en general, en forma diaria, por lo tanto, los tableros presentan la información reciente. Los gráficos presentan las métricas por fecha, geografía, agente, etc. Estos reportes son la respuesta a las preguntas del negocio establecidas para este proyecto. Para los agentes el sistema de BI le ayuda en la toma de decisiones.

## RESULTADOS PARA LOS AGENTES

Para los agentes del mercado de energía eléctrica (usuarios del sistema), el BI les permite contar con información de demanda y pérdidas de energía en forma oportuna y confiable, esto les ayuda a tomar acciones según la información consultada. El BI brinda ventajas adicionales para los agentes, conocido como “auto servicio”, ellos pueden crear en forma fácil reportes personalizados utilizando herramientas con las que están familiarizados como Excel.

## utilidad del sistema de bi para el negocio de energía

Un proyecto de BI puede considerar exitoso cuando es utilizado por la mayor parte de la organización porque ofrece “lo que necesitan”. Los proyectos de BI enfrentan desafíos como calidad de datos, desempeño, oportunidad de la información entro otros. La calidad es una necesidad, no solo para un BI, sino para otros tipos de sistemas, los agentes del mercado requieren que la información sea confiable.

Para enfrentar estos retos fue realizado un buen diseño del DW, ETLs, cubos OLAP, reportes y el sitio Web con una agradable presentación. Fueron incluidos aspectos como bitácora para seguimiento a la carga de procesos y calidad de los datos. Por el alto volumen de sistema fue creado un esquema de particiomiento para la de tabla de hechos del DW y para los cubos OLAP, fueron creadas utilidades para facilitar la carga diaria en forma automática y la recarga de datos históricos de forma eficiente.

La información de bitácora es valiosa para los administradores del sistema para detectar fallas de procesamiento y errores en los datos para poder ofrecer a los agentes información de calidad.

## APLICACIÓN WEB

Otro de los objetivos del proyecto era el de contar con un portal web para los agentes consultar la información de demanda y pérdidas de energía. El sitio construido presenta los reportes, los cuales pueden ser filtrados por parámetros como la fecha. Adicional a los reportes del sistema también pueden ser consultados los reportes de bitácora para establecer el estado de las tareas de carga del sistema, las filas de datos que presentaron problemas de calidad de datos y los archivos procesados.

En la construcción se buscó que fuera un sitio con una agradable presentación y funcionalidades como exportar e imprimir los reportes desde el sitio web.

## ENTREGABLES DEL PROYECTO

La Tabla 8 presenta las rutas de los artefactos construidos en el proyecto, los cuales son entregados como parte del proyecto de grado:

Tabla 8. Artefactos del sistema

|  |  |
| --- | --- |
| **Archivo** | **Descripción** |
| **\DemandaBI\Doc\Proyecto\** | |
| Presentacion\_DemandaBI.pptx | Presentación proyecto de grado demanda BI |
| ProyectoGrado\_DemandaBI.docx | Documento proyecto de grado |
| Anexo1\_Solicitud\_Datos\_XM.docx | Anexo 1 solicitud de datos XM |
| Anexo2\_Manual\_Instalacion.docx | Anexo 2 manual de instalación |
| Anexo3\_Manual\_Operacion.docx | Anexo 3 manual de operación |
| Anexo4\_Herramientas\_Software.docx | Anexo 4 herramientas de software utilizadas en el proyecto |
| **\DemandaBI\Doc\Formatos\** | |
| DatosAutoriza.pdf | Solicitud de datos para el proyecto de grado a la empresa XM |
| DatosAutoriza\_Firma.pdf | Solicitud de datos para el proyecto de grado a la empresa XM |
| Solicitud\_Datos\_XM.docx | Solicitud de datos para el proyecto de grado a la empresa XM |
| **\DemandaBI\Doc\Manual\** | |
| DicDatos\_DemandaDW.xlsm | Diccionario de datos del sistema |
| DicDatos\_DemandaXM.xlsm | Diccionario de datos con las tablas del BI de XM utilizadas para generar archivos planos para el sistema |
| DisenoOLAP\_DemandaBI.xlsm | Diseño de la base de datos multidimensional |
| MatrizCRUD\_DemandaDW.xlsm | Matriz CRUD con permisos para el sistema |
| **\DemandaBI\Resources\Images\** | |
| \*.\* | Archivos de imágenes utilizadas en el sistema |
| **\DemandaBI\Setup\** | |
| Cubos | Instalador de base de datos multidimensional SQL Server Analysis Services |
| DemandaWeb | Instalador de sitio Web del sistema |
| ETL | Instaladores de ETLs |
| Report | Instaladores de reportes SQL Reporting Services |
| **\DemandaBI\Src\** | |
| Auditoria\_RS | Proyecto de reportes de auditoria creados en SSDT para VS 2017 |
| Demanda\_OLAP | Proyecto de cubos creados en SSDT para VS 2017 |
| DemandaBI\_Datos\_SSIS | Proyecto de ETLs para extraer los datos de XM creados en SSDT para VS 2017 |
| DemandaBI\_RS | Proyecto de reportes del sistema creados en SSDT para VS 2017 |
| DemandaBI\_SSIS | Proyecto de ETLs del sistema creados en SSDT para VS 2017 |
| DemandaWeb | Proyecto sitio web del sistema creado con VS 2019 |
| Script | Proyecto de base de datos creado con SSMS |
| **\DemandaBI\Test\** | |
| \*.\* | Esta carpeta contiene casos de pruebas y una muestra con archivos para cargar el sistema. Solo fue subido una muestra de datos, puesto que el volumen de datos de demanda y pérdida de energía eléctrica de la empresa XM es alrededor de 2,000 millones de filas en 20 años. La muestra son unos 1,373 archivos con un tamaño total de 300 MB, esta muestra contiene una fracción de los datos. La muestra contiene datos para los años 2016 a 2019, solos los meses 2019-08 y 2019-09 contiene todos los datos, pero los demás archivos es una parte de los datos. |

# CONCLUSIONES

Los proyectos de BI son construidos siguiendo una metodología, las más conocidas son las de Inmon y Kimball, las cuales son muy similares con diferencias como crear un DW corporativo o data marts temáticos. En nuestro caso fue utilizada la metodología de Kimball, que es apropiada para proyectos pequeños.

En la “arquitectura bus” de Kimball para BI los bloques podrían considerarse como una constante, pero la implementación de cada uno de estos bloques es abierta para cada necesidad y tecnología empleada. En la construcción de estos bloques es recomendable utilizar patrones y buenas prácticas para que el sistema, adicional a cumplir con los requisitos, tenga los atributos de calidad de softwares que son muy conocidos en la ingeniería de sistemas como: funcionalidad, desempeño, compatibilidad, usabilidad, fiabilidad, seguridad, mantenibilidad y portabilidad. En el proyecto fueron incluidos estos aspectos no funcionales para crear un sistema robusto y un aporte para el uso de esta tecnología.

Establecer las preguntas de negocio y definir las fuentes de datos son fundamentales para el éxito del proyecto, un mal entendimiento de las necesidades del negocio o una mala selección de las fuentes de datos implican reprocesos o terminar con un proyecto que no satisface las necesidades de los usuarios, para el proyecto fue definido utilizar archivos planos como fuentes de datos para no depender de fuentes de datos externas de demanda y pérdidas de energía suministrada por la empresa XM, de esta forma solo es necesario crear archivos planos con nueva información descargada de la fuentes de datos pública del portal de XM.

Un proyecto de BI puede considerar exitoso si es utilizado por los usuarios porque ofrece “lo que necesitan”, para nuestro caso los usuarios son los agentes del mercado mayorista de energía eléctrica. Esto implicó retos como el desempeño, oportunidad de la información, confianza, calidad de datos y una agradable presentación. Por estas razones fue construida una solución de BI robusta, capaz de procesar el volumen de datos de demanda y pérdidas de energía que maneja la empresa XM en corto tiempo y con aspectos como bitácora para seguimiento de procesos de carga, manejo de datos inconsistentes y utilidades para facilitar el proceso de carga y recarga de datos.

En el mercado existen herramientas de BI: gratuitas, pagas, suites completas para todos los aspectos de BI o una mezcla según las necesidades. La selección de las herramientas es un proceso que depende de las condiciones económicas de los clientes, las necesidades del proyecto y otro aspecto es que estén alineadas con las herramientas utilizadas por la empresa. Para este proyecto fue seleccionada las herramientas Microsoft SQL Server, por ser muy utilizada en nuestro medio, por ejemplo, el BI de XM, que facilitó los datos para el proyecto, es construido con las herramientas de Microsoft SQL Server. Para la capa de presentación fue construida una aplicación Web creada con las herramientas de Microsoft Visual Studio .Net y otros frameworks. Pero para los reportes fueron analizadas dos opciones de la tecnología Microsoft: SQL Reporting Services (SSRS) y PowerBI, para el proyecto fue seleccionada SSRS porque no incrementa los costos para el proyecto.

BI es una tecnología que cada día es más utilizada por las empresas, pero existe poco personal calificado. Este proyecto incluye código para contribuir en la generación de conocimiento en BI y como material de enseñanza para quienes inician o ya tienen conocimientos de BI, espero que sea de mucha utilidad. El código es gratuito y puede ser utilizado sin ninguna restricción, los riesgos y responsabilidad de uso recaen sobre quién lo utilice.

# RECOMENDACIONES PARA TRABAJOS FUTUROS

El modelo construido solo tiene dos métricas: demanda y pérdida, las cuales cruzan con dimensiones como: fecha, hora, geografía, agente, etc., y fueron creados unos gráficos. Como recomendación está crear reportes tabulares con suscripciones que dejen un archivo plano con datos de demanda y pérdidas por mes y puedan ser descargados desde el sitio web, esto es útil para los agentes de mercado tengan datos para alimentar sus sistemas de información personalizados.

Otra recomendación es la creación de reportes en Excel, como plantillas, para que el usuario final tenga un punto de partida para crear sus reportes personalizados, estos reportes pueden ser subidos al sitio web o dejados en rutas compartidas en disco. Adicional existen otras herramientas pensadas en los usuarios finales como Report Builder, la cual crea reportes de forma rápida, esta herramienta es de uso libre, en las referencias encuentra el sitio donde puede descargar esta herramienta.

El BI nos ayuda a analizar la información histórica, otro paso para las empresas es el desarrollo de BA (Business Analitics). Este tipo de herramienta es conocida como minería de datos e inteligencia artificial. Un buen complemento para proyectos futuros es la creación de “pronósticos” (tipo series de tiempo) para predecir el comportamiento futuro de la demanda y pérdidas de energía utilizando como datos de “entrenamiento” la información histórica del BI.

# ANEXOS

## ANEXO 1: SOLICITUD DE DATOS A LA EMPRESA XM

## ANEXO 2: MANUAL DE INSTALACIÓN

## ANEXO 3: MANAUL DE OPERACIÓN

## ANEXO 4: HERRAMIENTAS DE SOFTWARE

# REFERENCIAS

Aspin, A. (2015). Business Intelligence with SQL Server Reporting Services. Editorial Apress. Recuperado de: <https://www.apress.com/la/book/9781484205334>

Bustamante Granda, W. X., Macas Ruiz, E. M. y Cevallos Macas, F. B. (2018). Data Warehouse: Análisis multidimensional de BAFICI utilizando Power Pivot. Revista espacios. Vol. 39 (No. 34) Pág. 24. Recuperado de: <http://www.revistaespacios.com/a18v39n34/a18v39n34p24.pdf>

Comisión de regulación de energía y gas - CREG (2011). Propuesta para remunerar planes de reducción de pérdidas no técnicas de energía eléctrica en sistemas de distribución local. Recuperado de: <https://www.creg.gov.co/>

Curto Diaz, J. y Conesa Caralt, J. (2015). Introducción al business intelligence. Recuperado de: <https://issuu.com/gian3/docs/introduccion_al_business_intelligen>

Dingeldein, T. (2019). 7 Best Free Business Intelligence (BI). Recuperado de: <https://blog.capterra.com/top-8-free-and-open-source-business-intelligence-software/>

El universal (218). Pérdidas de energía eléctrica en Colombia. Recuperado de: <http://www.eluniversal.com.co/colombia/robos-de-energia-en-colombia-generan-al-ano-mil-billones-de-dolares-en-perdidas-196797>

Ferrari, Alberto & Russo, Marco (2008). Introduction to the SQLBI methology. Recuperado de: <https://www.sqlbi.com/wp-content/uploads/Introduction-to-SQLBI-Methodology-draft-1.0.pdf>

Gorman, K., Hirt, A., & otros (2019). Introducing Microsoft SQL Server 2019. Recuperado de: <https://info.microsoft.com/ww-landing-introducing-sql-server-2019-content.html>

Guru 99 (2020). 24 Best Business Intelligence(BI) Tools List in 2020. Recuperado de: <https://www.guru99.com/business-intelligence-tools.html>

Kimball, R. & Ross, M. (2013). The data warehouse toolkit (3 ed.). Indianapolis. Editorial Wiley.

Kimball, R. & Ross, M. (2014). Dimensional modelling in depth.

Casters, M., Bouman, R. & Dongen, J. V. (2010). Pentaho Kettle Solutions. Building open source ETL solutions with Penthaho data integration. Indianapolis. Editorial Wiley.

Mora Alba, M. F. (2016). Fundamentos de deep learning y una aplicación para predecir la demanda de energía eléctrica en México. Tesis de grado. Itam México. Recuperado de: <https://sophie-germain.github.io/deeplearning_thesis.pdf>

Morales Cardoso, Santiago Leonardo (2019). Metodología para procesos de inteligencia de negocios con mejoras en la extracción y transformación de fuentes de datos, orientado a la toma de decisiones. Tesis de grado. Universidad de Alicante. Recuperado de:

<https://rua.ua.es/dspace/bitstream/10045/92767/1/tesis_santiago_leonardo_morales_cardoso.pdf>

Root, R. & Mason, C. (2012). Pro SQL Server 2012 BI Solutions. Recuperado de: <https://www.apress.com/la/book/9781430234883>

Rueda Mejía, V. M. (2011) Predicción del consumo de energía en Colombia con modelos no lineales. Tesis de grado. Universidad Nacional de Colombia. Recuperado de: <http://bdigital.unal.edu.co/4462/1/200920150.2011.pdf>

Sarka, D., Lah, M. & Jerkic, G. (2012). Training Kit (Exam 70-463) Implementing a Data Warehouse with Microsoft SQL Server 2012 (MCSA). Recuperado de: <https://www.microsoftpressstore.com/store/training-kit-exam-70-463-implementing-a-data-warehouse-9780735666092#downloads>

TodoBI (2019). Metodologías agiles para Analytics (Business Intelligence, Big Data). Recuperado de: <https://todobi.com/metodologias-agiles-para-bidw/>

XM (2018). Introducción a la Operación y Administración del Mercado. Recuperado de: <https://www.xm.com.co/corporativo/Memorias%20Seminarios/PPTFINAL_LUNES23.pdf>

XM (2020). Portal web de BI. Recuperado de: <http://portalbissrs.xm.com.co>