

UNIVERSIDAD SAN FRANCISCO DE QUITO

Colegio de postgrado

Desarrollo de un modelo de análisis y geoprocesamiento de información espacio-temporal aplicado a un Data Warehouse, para el departamento de Análisis y Sistemas Geográficos de Información de la Empresa Eléctrica Regional Centro Sur

Mario Alberto Flor Ambrosi

Tesis de grado presentada como requisito para la obtención del título de Magíster en
Sistemas de Información Geográfica.

Quito, septiembre de 2010.

© Derechos de autor

Mario Alberto Flor Ambrosi

2010

AGRADECIMIENTOS

Quiero agradecer a la Universidad San Francisco de Quito y particularmente a Richard Resl por su apoyo y conocimientos compartidos.

DEDICATORIA

Quiero dedicar este trabajo a mi esposa y mis hijos quienes me han apoyado siempre para cumplir con mis metas.

RESUMEN

Este trabajo de tesis presenta los geoprocesos para el análisis, extracción y transformación de la información espacial de las redes almacenadas en la geodatabase del Sistema de Información Geográfico (SIG) y su combinación con la información tabular del Sistema de Gestión de Proyectos para almacenarla en un Data Mart aplicado al departamento de Análisis y Sistemas Geográficos de Información.

El trabajo lleva hasta la implementación de un prototipo para demostrar que los procesos pueden ser aplicables y mostrar el potencial de las herramientas de geoprocesamiento de los SIG integrados con herramientas de inteligencia de negocios

En el capítulo dos, se indican los conceptos Sistemas de Información Geográficos, elementos de la inteligencia de negocios como son los Data Warehouse tradicionales y espaciales, los metadatos, así como elementos importantes en un proyecto de desarrollo de un sistema como son los procesos, requerimientos y lenguajes UML.

El capítulo tres trata sobre el levantamiento de requerimientos basado en el estándar 830-1998 aplicado al caso particular del prototipo a implementar, donde se indican los requerimientos espacial, técnicos y de negocios para el mismo.

El capítulo cuatro presenta el diseño lógico del Data Mart y los diagramas de Mapeo. El capítulo cinco se centra completamente en la documentación de los procesos. Esta dividió en dos secciones; una dedicada a los procesos de extracción, transformación y carga de la información espacial desde el SIG a la bodega de datos transitoria, y la otra sección dedicada a los procesos de extracción, transformación y carga de la información tabular para el llenado de las dimensiones y tablas de hechos del Data Mart. En ambas secciones

se detalla minuciosamente los procesos, los responsables, el objetivo, alcance, los clientes y proveedores con sus requisitos, al igual que sus indicadores y un glosario de términos para cada uno de los procesos.

El capítulo seis hace una descripción del prototipo desarrollado, detallando sus componentes junto con su arquitectura, procedencia y ámbito de servicio. Además se indica la metodología adoptada junto con las fases y procesos aplicados, y los beneficios del mismo.

Los posteriores capítulos hacen referencia a los factores críticos de éxito para un proyecto de este tipo, junto con los hallazgos, conclusiones y recomendaciones producto del presente trabajo de tesis.

ABSTRACT

This thesis presents the geoprocessing for analysis, extraction and processing of spatial information networks geodatabase stored in Geographic Information System (GIS) combined with tabular information on Project Management System for storage in a Data Mart applied to the Department of Analysis and Geographic Information Systems.

The work leads to the implementation of a prototype to demonstrate that process can be applied and show the potential of geoprocessing tools of GIS integrated with business intelligence tools

In chapter two, identified the Geographic Information Systems concepts, elements of business intelligence such as traditional and spatial data warehouse, metadata, as well as important elements in a project to develop a system such as the processes, requirements and UML.

The third chapter discusses the requirements gathering based on the standard 830-1998 applied to particular case of prototype to be implemented, which shows the space requirements, technical and business for himself.

Chapter four presents the logical design of the Data Mart and Mapping diagrams. Chapter five focuses entirely on the documentation of processes. This split into two sections, one dedicated to the extraction, transformation and loading of spatial information from the GIS to the data warehouse, and another section dedicated to the extraction, transformation and loading tabular information filling of the dimensions and fact tables in the Data Mart. Both sections meticulously details the processes, responsible, objective, scope, customers and suppliers with their requirements, as well as indicators and a glossary of terms for each of the processes.

Chapter six is a description of the prototype developed, detailing its components along with their architecture, origin and scope of service. Also describes the methodology adopted with the stages and processes and benefits.

Subsequent chapters refer to the critical success factors for a project of this type, together with the findings, conclusions and recommendations issued this thesis.

Tabla de contenido

| | |
|--|----|
| Lista de Tablas..... | 11 |
| Lista de Gráficos..... | 12 |
| Introducción..... | 14 |
| Necesidades y/o problemas a solucionar | 2 |
| Objetivos..... | 3 |
| Objetivos específicos..... | 4 |
| Alcance | 4 |
| Método a aplicar | 6 |
| 2. Fundamentos Teóricos..... | 7 |
| SIG..... | 7 |
| Data Mart..... | 11 |
| Data Warehouse..... | 12 |
| Base de datos transitoria (o base de datos integradora)..... | 15 |
| ODS (Operational Data Stored)..... | 16 |
| Data Warehouse Espacial | 17 |
| Esquema general del diseño de un proyecto de Data Warehouse Espacial..... | 19 |
| Representación de la información Geográfica en un Data Warehouse Espacial..... | 23 |
| Especificación de requisitos del software..... | 23 |
| Proceso | 24 |

| | |
|--|-----|
| RUP (Rational Unified Process)..... | 24 |
| Lenguaje Unificado de Modelado (UML)..... | 25 |
| 3. Especificación de Requerimientos..... | 31 |
| 3.1 Introducción | 31 |
| 3.2 Descripción global | 34 |
| 3.3 Los requisitos específicos | 38 |
| 4. Análisis y Diseño Lógico del prototipo..... | 41 |
| Misión del Área de Negocios: Departamento SIGADE..... | 41 |
| Estrategia | 41 |
| Macro procesos beneficiados con la arquitectura..... | 41 |
| Usuario con Acceso | 41 |
| Tipo de información para cada usuario | 42 |
| Modelo Lógico del Data Mart | 44 |
| Diagramas de Mapeo | 49 |
| 5. Procesos para el manejo de la información | 53 |
| 5.1 Proceso de manejo de la información tabular y espacial en la etapa ETL para el Operacional Data Store (ODS)..... | 54 |
| 5.2 Procesos ETL para la carga de las tablas de hechos en el Data Mart. | 133 |
| 6. Descripción del prototipo | 145 |
| 6.1 Origen y reseña de acontecimientos..... | 145 |
| 6.2 Necesidad de la implementación..... | 145 |

| | | |
|-----|-----------------------------------|--------------------------------------|
| 6.3 | Principales componentes | 146 |
| 6.4 | Arquitectura del Sistema | 158 |
| 6.5 | Implementación del Data Mart..... | 161 |
| 6.6 | Beneficios..... | 165 |
| 7. | Factores críticos de éxito | 166 |
| 8. | Hallazgos | 167 |
| 9. | Conclusiones..... | 169 |
| 10. | Recomendaciones | 170 |
| | Bibliografía..... | 172 |
| | Glosario | 174 |
| | Anexos..... | ¡Error! Marcador no definido. |

Lista de Tablas

| | |
|--|-----|
| Tabla 1. Diferencias entre Data Warehouse y Data Mart..... | 12 |
| Tabla 2. Valores asignados a los campos Aereo_Subt_ID y BajaMedia_ID de acuerdo al nombre de las tablas fuente. | 61 |
| Tabla 3. Valor asignado a lo campo Fase_ID de acuerdo al campo "Subtipo" de la tabla fuente. | 62 |
| Tabla 4 . Valor asignado a lo campo Acom_Bajante_ID de acuerdo al campo "Subtipo" de la tabla fuente..... | 62 |
| Tabla 5. Valores asignados a los campos Fase_ID de acuerdo al campo Subtipo de la tabla fuente. | 79 |
| Tabla 6. Valor asignado a lo campo BajaMedia_ID de acuerdo al campo "Subtipo" de la tabla fuente. | 94 |
| Tabla 7. Metadatos para el proceso ETL de carga de la tabla de hechos WH_CUBO_TRAMOS en el Data Mart..... | 135 |
| Tabla 8. Metadatos para el proceso ETL de carga de la tabla de hechos WH_CUBO_TRANSFORMADORES en el Data Mart..... | 136 |
| Tabla 9. Metadatos para el proceso ETL de carga de la tabla de hechos WH_CUBO_LUMINARIA en el Data Mart. | 137 |
| Tabla 10. Metadatos para el proceso ETL de carga de la tabla de hechos WH_CUBO_CLIENTES en el Data Mart. | 138 |
| Tabla 11. Metadatos para el proceso ETL de carga de la tabla de hechos WH_CUBO_POSTES en el Data Mart..... | 139 |
| Tabla 12. Metadatos para el proceso ETL de carga de la tabla de hechos WH_CUBO_ESTRUCTURA_POSTE en el Data Mart..... | 140 |

Lista de Gráficos

| | |
|---|----|
| Gráfico 1. Componentes del SIG | 7 |
| Gráfico 2. Dimensión geográfica..... | 15 |
| Gráfico 3. Esquema de estructura de Data Warehouse Espacial..... | 17 |
| Gráfico 4. Esquema de un proyecto de Data Warehouse Espacial..... | 19 |
| Gráfico 5. Proceso de Extracción, Transformación y Carga de los datos para alimentar la bodega..... | 21 |
| Gráfico 6. Jerarquía de los diagramas UML 2.0..... | 26 |
| Gráfico 7. Diagrama de actividades. | 28 |
| Gráfico 8. Diagrama de componentes. | 29 |
| Gráfico 9. Diagrama de Despliegue. | 30 |
| Gráfico 10. Organigrama de la Empresa. | 32 |
| Gráfico 11. Usuario con acceso al Data Mart..... | 42 |
| Gráfico 12. Modelo Lógico del cubo Tramos. | 44 |
| Gráfico 13. Modelo Lógico del cubo Transformadores. | 45 |
| Gráfico 14. Modelo Lógico del cubo Luminarias. | 46 |
| Gráfico 15. Modelo Lógico del cubo Clientes. | 47 |
| Gráfico 16. Modelo Lógico del cubo Postes. | 48 |
| Gráfico 17. Modelo Lógico del cubo Estructuras en Poste..... | 49 |
| Gráfico 18. Mapeo de las tabas de hechos del Data Mart. (1/2)..... | 50 |
| Gráfico 19. Mapeo de las tabas de hechos del Data Mart (2/2)..... | 51 |
| Gráfico 20. Mapeo de las dimensiones del Data Mart. | 52 |
| Gráfico 21. Procesos ETL para el Data Mart | 53 |
| Gráfico 22. Diagrama de actividades del proceso ETL para los tramos del ODS. | 71 |

| | |
|---|-----|
| Gráfico 23. Diagrama de actividades del proceso ETL para los transformadores del ODS. | 86 |
| Gráfico 24. Diagrama de actividades del proceso ETL para los clientes del ODS. | 101 |
| Gráfico 25. Diagrama de actividades del proceso ETL para las luminarias del ODS..... | 116 |
| Gráfico 26. Diagrama de actividades del proceso ETL para los postes del ODS. | 131 |
| Gráfico 27. Diagrama de actividades del proceso ETL para carga de las tablas de hechos en el Data Mart. | 143 |
| Gráfico 28. Pantallas del Sistema de Gestión de Proyectos. | 146 |
| Gráfico 29. Pantalla del ArcGIS Data Interoperability. (Spatial ETL de Luminarias) | 148 |
| Gráfico 30. Arquitectura de software de ESRI. | 149 |
| Gráfico 31. Pantalla de software SIG de EERCS. | 150 |
| Gráfico 32. Diagrama de correspondencias de Oracle Warehouse Builder. | 152 |
| Gráfico 33. Oracle BI Administration. | 153 |
| Gráfico 34. Pantalla de Oracle Answer. | 154 |
| Gráfico 35. Pantalla de Oracle Interactive Dashboard. | 155 |
| Gráfico 36. Diagrama de componentes del prototipo de Data Mart. | 158 |
| Gráfico 37. Diagrama de despliegue del prototipo de Data Mart. | 159 |
| Gráfico 38. Fases y procesos de la implementación del prototipo de Data Mart. | 164 |

Introducción

La importancia de los datos espaciales está dada por el hecho de que al menos el 80% de la información que se maneja en un sistema tiene un componente espacial. Por ejemplo, en los datos de un clientes, se almacena el lugar de nacimiento, en una venta se almacena el lugar de la venta, en una interrupción de servicios se registra la dirección, entre otros ejemplos.

Por lo tanto, los sistemas de información deben cambiar de ser simplemente aplicaciones que almacenan y procesan transacciones, a sistemas que soporten a los negocios por medio de grandes bases de datos que integren y analicen la globalidad del negocio, teniendo en cuenta los datos alfanuméricos que representan comportamientos del negocio junto con su información espacial, todo dentro de una estructura holística, dinámica y variante en el tiempo.

Entre los campos de aplicación de los Data Warehouse con componentes espaciales están entre otros:

- Demográfico (desplazamiento de poblaciones, crecimiento/decrecimiento de ciudades).
- Técnico (Frecuencia de fallas en las redes de distribución).
- Gerenciales (Información de cumplimientos, avances y previsión de sus proyectos, presupuesto, crecimiento de infraestructura).
- Ecológico (crecimiento/decrecimiento de bosques).
- Climatológicos o Geológicos (inundaciones o terremotos).

Necesidades y/o problemas a solucionar

Uno de los departamentos claves donde se maneja la información técnica y de índices de la empresa Eléctrica Regional Centro Sur (EERCS), es el Departamento de Análisis y Sistemas Geográficos de Distribución (SIGADE), cuya misión es recopilar, procesar, actualizar, analizar y difundir la información técnica, administrativa, económica y espacial, con calidad y oportunidad, para facilitar la toma de decisiones de la dirección y de la administración, a fin de optimizar la gestión técnica y administrativa que beneficie al cliente externo e interno.

Por otro lado la Empresa Eléctrica, se encuentra implementando su Plan Informático, que fue aprobado por el Directorio de la Empresa en el año 2006 y que incluye en uno de sus ejes la “Inteligencia de Negocios”, cuyo objetivo es lograr que las decisiones que se tomen en la Empresa estén soportadas en conocimiento e información fácil y ágilmente disponible para el personal directivo y la Presidencia Ejecutiva.

La situación de la información geográfica de las redes de distribución que se maneja en el SIGADE está influenciada por factores dinámicos como son: nuevos proyectos de electrificación, replanteo de proyectos, transferencia de carga, cambios demográficos, y plan de expansión de las redes. Sin embargo; cada una de éstas dinámicas se manifiesta en procesos de contraste espacio-temporal, cuyo análisis, de manera generalizadora y sistémica, presenta un reto notorio para la investigación en esta área.

La demanda de servicios de electricidad supone un gran desembolso presupuestario en el ámbito gubernamental, por lo que la capacidad de los Sistema de Información Geográficos para representar y gestionar información relevante sobre la oferta, la demanda, y los servicios, los convierte en un medio idóneo para la provisión eficiente de los mismos. Sin embargo, aunque un SIG ha sido implantado en esta empresa, principalmente para la

gestión de infraestructura, el potencial de los modelos geográficos permanece subutilizado en la provisión de servicios y en la utilización de su información para tener una visión global de la compañía y así tomar mejores decisiones.

Entre los principales inconvenientes del SIGADE con respecto a este tema, tenemos a continuación los más importantes:

- Se necesita saber manejar las herramientas de análisis espacial para obtener los resultados.
- La recopilación de información específica de redes y proyectos tiene un alto costo y toma tiempo.
- No dispone de información resumida y sumariada de los proyectos de electrificación que le faciliten la toma de decisiones.

Entre las necesidades del departamento SIGADE tenemos a continuación las más importantes:

- Mejorar el acceso a los datos espaciales con información resumida y sumariada.
- Necesidad de información para medir sus avances.
- Satisfacer consultas de la alta gerencia, direcciones, contratistas y organizamos externos de regulación como el Ministerios de Electricidad, CONELEC, entre otros.

Objetivos

El objetivo general de esta tesis es elaborar un modelo de análisis y geoprocesamiento de datos espaciales y temporales, para luego cargar la información resultante de estos análisis en un Data Mart que permitirá al departamento de Análisis y Sistemas Geográficos de

Información de la Empresa Eléctrica Regional Centro Sur tener una herramienta prototipo que le ayudara a la tomar las decisiones empresariales

Objetivos específicos

Entre los objetivos específicos de esta tesis tenemos:

- Implementar un prototipo de Data Mart a través del cual se pueda demostrar las ventajas y potencialidad para la toma de decisiones que representa una integración de una herramienta SIG con una de inteligencia de negocios.
- Elaborar los geoprocenos para el análisis, la extracción y transformación de la información espacial del SIG de la EERCS en información alfanumérica para ser cargada en un Data Mart

Alcance

El alcance de esta tesis es definir los geoprocenos para el análisis, extracción y transformación de la información espacial de las redes almacenadas en la geodatabase del Sistema de Información Geográfico (SIG) y combinarla con la información tabular del Sistema de Gestión de Proyectos (SGP) para almacenarla en un Data Mart aplicado al departamento de Análisis y Sistemas Geográficos de Información (SIGADE).

Se implementara un prototipo de Data Mart para resaltar las ventajas de este proyecto, utilizando herramientas de geoprocensamiento de ArcGIS y de Inteligencia de Negocios de Oracle. No se analizara ni desarrollara la creación de un spatial OLAP (On Line Analytical Process) ni la creación de herramientas DSS (Decision Support System).

Esta tesis se basa específicamente en los geoprocenos y sus datos tabulares obtenidos de la información espacial de las redes de distribución incorporada con la información

transaccional de los proyectos de electrificación y sus costos almacenadas en el Sistema de Gestión de Proyectos (SGP), y de esta manera valorar los proyectos de acuerdo a la información representada en el Sistema de Información Geográfico (SIG).

Es importante acotar con respecto al presente trabajo de tesis, que el mismo no está centrado en la implementación del software para el Data Mart o SIG, sino en los geoprocesos para la extracción, transformación y carga de la información fuente desde el SIG y el SGP de la Centro Sur en el prototipo de Data Mart. Para demostrar que los procesos pueden ser aplicables y mostrar el potencial de las herramientas de geoprocesamiento de los SIG integradas con herramientas de inteligencia de negocios se realizó un prototipo de software para el Data Mart, pero por tratarse de un prototipo, no se incluye las especificaciones de los requerimientos funcionales estrictamente, ni tampoco se definió los requerimientos no funcionales los que deberán ser identificados y cumplidos una vez que se definan los recursos y el entorno de operación definitivo.

Método a aplicar

No existe una metodología formal, por lo tanto este trabajo de tesis se basará en la aproximación metodológica sugerida por Bohorquez [6], ESRI [14] y Morgan [11] y la metodología de diseño de Data Marts propuesta por Oracle[8] y Geer[12], aplicando las adecuaciones y mejoras que se acerquen más a la realidad específica de la Empresa Eléctrica y la estructura de información geográfica que se maneja. En la sección 6.5 “Implementación del Data Mart” de este trabajo de tesis se detallaran las fases y procesos que se deberían adoptar como metodología para la implementación de un Data Mart.

Para el caso de este prototipo solo se adopto las fases de la metodología de implementación que involucra el aspecto didáctico de este trabajo es decir fases como la de transición, revisión y evaluación del proyecto de implantación, no se realizaron por ser solo un prototipo, pero si se aplicaron las otras fases como son el establecimiento de la estrategia de operación y administración de los datos y metadatos, la definición del alcance del proyecto, la realización de la fase de análisis, la fase de diseño de los modelos lógicos, físicos y los procesos y geoprocesos ETL, para luego concretar con la construcción, implementar y evaluación de los procesos ETL. Dentro de estas fases se realizaron únicamente los procesos de definición de requisitos de la aplicación, definición de la arquitectura tecnológica, definición de la operación de datos, administración de los metadatos y el Diseño y construcción del Data Mart.

2. Fundamentos Teóricos

En este capítulo se presentan conceptos de SIG, Data Mart y Data Warehouse y a pesar de que el tema de esta tesis no es hacer un Data Warehouse espacial, se mencionan las características que debería tener un proyecto de este tipo.

SIG

Una definición técnica de un Sistema de Información Geográfica (SIG) lo presenta como un conjunto de componentes interrelacionados que sirven para gestionar datos georeferenciados¹, lo que llamamos también datos espaciales. Entendiéndose por gestionar a capturar, integrar, almacenar, analizar y visualizar información georeferenciada.

Componentes de un SIG

Los componentes de un SIG son:



Gráfico 1. Componentes del SIG

Fuente: <http://www.grafosistemas.com>.

¹ Por georeferenciados se entiende que estos datos o mapas tienen unas coordenadas geográficas reales asociadas, las cuales nos permiten manejar y hacer análisis con información reales como longitudes, perímetros o áreas.

Equipos (Hardware)

Todos los recursos tangibles donde opera el SIG. Hoy por hoy, programas de SIG se pueden ejecutar en un amplio rango de equipos, desde servidores hasta computadores equipos móviles como GPS, PDA, Tablet PC, entre otros.

Programas (Software)

Los programas de SIG proveen las funciones y las herramientas necesarias para almacenar, analizar y desplegar la información geográfica. Los principales componentes de los programas son:

- ◆ Herramientas para la entrada y manipulación de la información geográfica.
- ◆ Un sistema de manejador de base de datos (DBMS).
- ◆ Herramientas que permitan búsquedas geográficas, análisis y visualización.
- ◆ Interfaz gráfica para el usuario (GUI) para acceder fácilmente a las herramientas.

Datos

Probablemente la parte más importante de un Sistema de Información Geográfico son sus datos. Los datos geográficos y tabulares pueden ser adquiridos por quien implementa el sistema de información, así como por terceros que ya los tienen disponibles. El Sistema de Información Geográfico integra los datos espaciales con otros recursos de datos y puede, incluso, utilizar los manejadores de base de datos más comunes para usar la información geográfica.

Recurso Humano

La tecnología del SIG está limitada si no se cuenta con el personal adecuado y capacitado para operar, desarrollar y administrar el sistema, que establece planes para aplicarlo en problemas del mundo real.

Procedimientos

Un SIG operará acorde con un plan bien diseñado y con unas reglas claras del negocio, que son los procesos y los procedimientos operativos característicos de cada organización.

El diseño de la geodatabase

El objetivo del diseñador de un SIG es normalizar y unificar esa información, construyendo un tipo de base de datos que se denomina **base de datos geográfica**. Tras este diseño pasamos a un proceso que casi siempre está en los proyectos de Bussines Intelligence: el ETL. Dentro de la base de datos geográfica se integra información cartográfica junto con información proveniente de otros sistemas, como pueden ser CRM, ERP, Sistemas de análisis técnicos, entre otros.

En una Base de Datos geográfica queda modelado el mundo real como objetos con su representación gráfica en términos de tamaño y dimensión relativa a la superficie de la Tierra. Estos datos, que llamamos cartográficos, quedan integrados a su vez con lo que se denominan atributos temáticos. Ejemplos típicos de atributos temáticos son los datos económicos y demográficos.

Los objetos geográficos están organizados en capas de información. Objetos que físicamente tendrían la misma ubicación quedan organizados en distintas capas.

Por ejemplo, si una empresa desea representar la ubicación física de sus clientes actuales y potenciales para hacer una campaña de marketing, tendríamos en un SIG el mapa

cartográfico de la ciudad en cuestión por un lado y por ejemplo representaríamos los atributos temáticos disponibles sobre tipo de compra que efectúan los clientes actuales y la capacidad adquisitiva de los potenciales. El análisis de esta información puede, por ejemplo, delimitarnos zonas geográficas con denominadores comunes interesantes entre clientes potenciales y actuales. Con este conocimiento podríamos crear una campaña muy localizada, a costo bajo y por tanto muy productiva.

Para ello el SIG provee de la capacidad analítica para gestionar los datos georeferenciados teniendo en cuenta que lo que nos interesa es que se estén referenciando objetos con una componente espacial sobre los que se tiene información asociada.

Se suele clasificar los SIG en los de tipo Vectorial y los de tipo Raster. Los SIG vectoriales utilizan puntos y vectores definidos en términos de pares de coordenadas y altitudes respecto a un sistema cartográfico. Por otra parte, los SIG Raster están especialmente indicados en aquellos casos en los que los objetos a representar son difusos o difíciles de delimitar como puede ser una mancha de petróleo o una nube. Delimitan la base de datos mediante mallas de píxeles. En caso de que sea necesaria una resolución más alta y por tanto subir el número de píxeles que representan el elemento, el rendimiento del sistema se ve penalizado.

La interfaz de visualización principal de un SIG es el mapa sobre el que interactuamos para acceder, visualizar y analizar la información, así como generar informes y mejorar la toma de decisiones. Los análisis geográficos son especialmente potentes al poder superponer en el mapa las características de los elementos de una misma categoría y poder hacer operaciones en las que se tenga en cuenta la conectividad, la coincidencia, la sobreposición y la vecindad de objetos.

Tras construir el SIG, los usuarios disponen de un sistema fiable que les permite acceder de forma rápida, homogénea y organizada a conjuntos de información que han sido integrados según las necesidades de la empresa y a la que antes no tenían acceso, tenían una visión limitada y parcial o que el proceso de conseguirla era lento, costoso en cantidad de horas-hombre invertidas, proclive a errores o muy difícil de conseguir de forma integrada, por ejemplo, por diferencias de formatos.

Como el mayor valor agregado de un SIG destaca su capacidad, al igual que en Business Intelligence, de realizar análisis complejos, aunque sobre entornos geográficos y la mejora de la toma de decisiones asociada, el acceso fácil y rápido a información visual, completa, relacionada y actualizada además de la generación de mapas e informes de gran calidad. En general, al implantar un SIG, se reducen los costes operativos, se incrementa la productividad y se ahorra mucho tiempo en la gestión de información espacial.

Data Mart

Un Data Mart es una forma simplificada de un Data Warehouse que está enfocada en un tema individual (o área funcional), tal como Ventas, Finanzas, o Marketing. Los Data Mart a menudo son construidos y controlados por departamento individual en una organización. Debido a su enfoque a temas individuales, los Data Mart a menudo toman datos desde un número pequeño de fuentes. Estas fuentes pueden ser sistemas operacionales internos o externos.

A diferencia del Data Mart un Data Warehouse, trata con múltiples áreas temáticas y es típicamente implementada y controlada por una unidad central de la organización como el grupo de Tecnologías de Información (TI) corporativo. A menudo este es llamado el Data Warehouse central o empresarial. Típicamente un Data Warehouse integra datos de múltiples sistemas fuente.

Ninguno de estas definiciones básicas limita el tamaño de un Data Mart ni la complejidad de los datos para soportar decisiones que este contiene. No obstante, normalmente los Data Mart son más pequeños y menos complejos que los Data Warehouse; por lo tanto son típicamente más fáciles de construir y mantener. La tabla a continuación resume las diferencias básicas entre una Data Warehouse y un Data Mart.

| Categoría | Data Warehouse | Data Mart |
|---------------------------------|-------------------|--------------------------------------|
| Alcance | Corporativo | Departamental (Línea del Negocio) |
| Temática | Múltiples temas | Temática Individual |
| Fuente de Datos | Muchas | Pocas |
| Tamaño | 100Gb a varios TB | < 100Gb |
| Tiempo de Implementación | Meses o Años | Meses |

Tabla 1. Diferencias entre Data Warehouse y Data Mart.

Data Mart Dependientes y Data Mart Independientes

Hay dos tipos básicos de Data Mart: los dependientes y los independientes. La categorización está basada principalmente en la fuente de datos que alimenta al Data Mart. Los Data Mart Dependientes toman los datos desde un Data Warehouse central que ya ha sido creado. El Data Mart Independiente, en contraste, son sistemas independientes contruidos para tomar datos directamente desde fuentes de datos operacionales o externas.

Data Warehouse

Un Data Warehouse es una colección de datos orientada a un determinado ámbito (empresa, organización, entre otros), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza.

Existen dos diferentes definiciones de dos autores que son:

Según Bill Inmon que fue uno de los primeros autores en escribir sobre el tema de los almacenes de datos, define un Data Warehouse en términos de las características del repositorio de datos:

Orientado a temas.- Los datos en la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.

Variante en el tiempo.- Los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones.

No volátil.- La información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en información de sólo lectura, y se mantiene para futuras consultas.

Integrado.- La base de datos contiene los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes.

Según el enfoque de Ralph Kimball que es otro conocido autor en el tema de los Data Warehouse, define un almacén de datos como: "una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis" y determinó que un Data Warehouse es en realidad: "la unión de todos los Data Mart de una entidad".

Para abordar un proyecto de Data Warehouse es necesario hacer un estudio de algunos temas generales de la organización o empresa, los cuales se describen a continuación:

- Situación actual de partida.- Cualquier solución propuesta de Data Warehouse debe estar muy orientada por las necesidades del negocio y debe ser compatible con la arquitectura técnica existente y planeada de la compañía.
- Tipo y características del negocio.- Es importante tener el conocimiento claro del tipo de negocios de la organización y el soporte que representa la información dentro de todo su proceso de toma de decisiones.
- Entorno técnico.- Se debe incluir tanto el aspecto del hardware, software, procesos manuales y automatizados.
- Expectativas de los usuarios.- Un proyecto de Data Warehouse no es únicamente un proyecto tecnológico, es un elemento que involucra a toda la empresa y como tal, tiene que contar con el apoyo y confianza de sus usuarios.
- Etapas de desarrollo.- Una vez concluidas las etapas anteriores se trabaja en el desarrollo de un modelo conceptual para la construcción del Data Warehouse.
- Prototipo.- El prototipo debe simular tanto como sea posible el producto final que será presentado a los usuarios.

Una Data Warehouse se modela en forma multidimensional. Esta organización de la información facilita el análisis de los datos y la concepción de las consultas. Un modelo multidimensional consiste en un conjunto de dimensiones que son asociadas a un fenómeno medible de interés para una organización. Este fenómeno es denominado “hecho”. Las dimensiones se componen de niveles, los cuales se estructuran jerárquicamente. El gráfico 1 muestra una dimensión geográfica con niveles Ciudad, Estado, Región y País. Cada nivel puede tener atributos, por ejemplo, el nivel País puede tener atributos nombre, idioma, área etc.

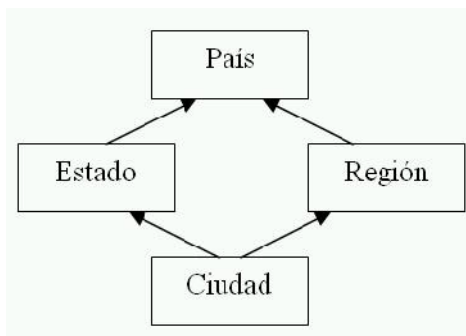


Gráfico 2. Dimensión geográfica.

Tradicionalmente, las dimensiones han sido tratadas de forma alfanumérica y las medidas de un hecho de forma numérica. En la actualidad, los modelos multidimensionales se han enriquecido con elementos traídos de las bases de datos espaciales. Una base de datos espacial está orientada al soporte de geometrías (puntos, líneas, regiones) asociadas a objetos del mundo real. Una base de datos espacio temporal soporta geometrías que cambian con el tiempo, ya sea en forma o posición.

Para la implementación de un Data Warehouse se debe diseñar la arquitectura dependiendo de la estructura interna de los datos y del tipo de consultas a realizar, luego se necesita construir los procesos para utilizar con las herramientas de migración y transformación de los datos hacia el almacén.

Base de datos transitoria (o base de datos integradora)

Es donde se realiza la limpieza, unificación, integración y organización de los datos para el Data Warehouse.

ODS (Operational Data Stored)

Es una base de datos diseñada para integrar datos de múltiples fuentes para hacer análisis y reportes más fácilmente. Debido a que los datos proceden de múltiples fuentes, la integración implica a menudo hacer la limpieza de los datos y contrastar la redundancia y comprobación de reglas de negocio en contra de la integridad. El ODS es por lo general diseñados para contener datos de bajo nivel o indivisible que se captura en "tiempo real" o "casi en tiempo real" en comparación con los volúmenes mucho mayores de datos almacenados en los Data Warehouse que son generalmente con menos frecuencia.

Un ODS no es un sustituto de un Data Warehouse, pero si podría convertirse en una fuente para el Data Warehouse.

Generalmente las implementaciones de los "ODS" son para la integración de datos de distintos sistemas fuente en una única estructura. Esto permite el acceso a los datos de los informes operacionales, datos maestros o datos de referencia de gestión

La tareas del ODS es estructurar y soportar todos los datos de análisis, organizar los datos por áreas para la bodega, realizar las sumalizaciones de los datos; de igual manera en el ODS se dimensionan los datos geográficos y se desarrollan los análisis preliminares que el cliente no debe hacer con respecto a la espacialidad del dato, tales como joins espaciales, buffers entre otros.

La finalidad del ODS, no es guardar el histórico del dato, si no simplemente es un área de organización para la bodega y no desempeña las tareas de la bodega, o una instantánea de la bodega en un momento dado.

Data Warehouse Espacial

Un Data Warehouse espacial es similar a un Data Warehouse pero con la adición del componente geográfico y la presentación de resultados en un ambiente geográfico, por lo que, resumiendo, diremos que es una colección de datos:

- Orientados a temas
- Integrada
- Variante en el tiempo
- No volátil

Este componente geográfico puede presentarse como una dimensión o como un valor agregado o de referencia, donde se pueden visualizar y analizar espacialmente los resultados.

A continuación se presenta una breve explicación de estas características, poniendo énfasis en su componente espacial.



Gráfico 3. Esquema de estructura de Data Warehouse Espacial.

Fuente: Bohorquez [6]

Orientado a temas

Cuando nos referimos a que los datos deben ser orientado a temas nos referimos a que estos son categorizados, almacenados y analizados por áreas temáticas del negocio y no por aplicaciones que se manejan en la organización.

Integrado

Los datos en el Data Warehouse son integrados y estructurados como un solo ente en la organización, esto nos permite garantizar:

- Un juego de información estándar
- Consistencia
- Confiabilidad

Y todo esto aprovechado para todas las aplicaciones, procesos y análisis del negocio. Esto se logra a través de los procesos ETL del Data Warehouse.

Variable en el tiempo

En un Data Warehouse los datos son almacenados con sus respectivos históricos, lo que garantiza poder desarrollar análisis de la dinámica de los mismos y nos permite:

- Desarrollar pronósticos
- Análisis de tendencias
- Patrones

No volátil

Como en un Data Warehouse tradicional en un Data Warehouse espacial los nuevos datos espaciales son insertados como nuevos registros. No hay borrado, ni actualización de registros en un Data Warehouse espacial, solo se permiten las consultas e inserciones.

Modelo datos geográficos

Como todo modelo, un modelo de datos geográficos es una abstracción del mundo real que emplea un conjunto de objetos, para soportar el despliegue de mapas, consulta, edición y

análisis, donde se presentan la información a través de mapas y símbolos, que representan la geografía como:

- Formas geométricas: punto para representar postes, líneas para representar carreteras, y polígonos para representar manzanas o predios.
- Redes: Para representar sistemas eléctricos o de agua potables
- Superficies: Para representar la forma del terreno en tres dimensiones
- Ubicaciones: Para representar la posición de un vehículo con GPS.
- Imágenes: Como fotos aéreas o satelitales.

Esquema general del diseño de un proyecto de Data Warehouse Espacial

El esquema para un proyecto de un Data Warehouse espacial depende de la implantación de una metodología adecuada, las consideraciones de diseño y modelamiento, la estandarización y automatización de los procesos de extracción, transformación y carga de los datos tabulares y espaciales (ETL), además de la georeferenciación y geocodificación de la información espacial, las consideraciones de administración, seguridad y afinamiento de los datos, el desarrollo de procesos y aplicaciones para el accesos y análisis de los datos tabulares y espaciales y durante todas estas etapas la actualización y mantenimiento de los metadatos.



Gráfico 4. Esquema de un proyecto de Data Warehouse Espacial.

Referencia: Bohorquez [6]

A continuación detallaremos cada una de estas etapas:

Metodología

La metodología propuesta debería ser incremental para que se ajuste a los nuevos requerimientos y al crecimiento de la organización.

Lo recomendable debería ser utilizar una metodología híbrida que utilice las utilizadas en los Data Warehouse tradicionales y los proyectos de Sistemas de Información Geográfica, enmarcados en estándares y procesos claros y bien documentados para el acceso a los datos.

Diseño y Modelamiento

El diseño y modelamiento de un Data Warehouse espacial involucra las etapas también utilizadas en los sistemas de información geográficos conocidos como etapa de conceptualización que involucra los requerimientos y expectativas de los usuarios, el entendimiento claro del flujo operativo y administrativo de datos, la identificación de involucrados y su relaciones además de también la correcta identificación de temas, categorías y jerarquías. Todo esto sin olvidar la orientación a los temas y no a las funciones o aplicaciones.

ETL

En Data Warehouse los procesos de extracción, transformación y carga (ETL) garantizan que la información tabular y espacial sea válida, estándar, esté integrada y tenga su sello de tiempo para que pueda ser transportada a la bodega transitoria (que sirve también como ODS). Solamente la información que no posea o no deba tener componente geográfica

puede pasar directamente a la bodega de datos definitivos. Luego es en la bodega transitoria donde se realiza la validación de la información en un contexto geográfico.

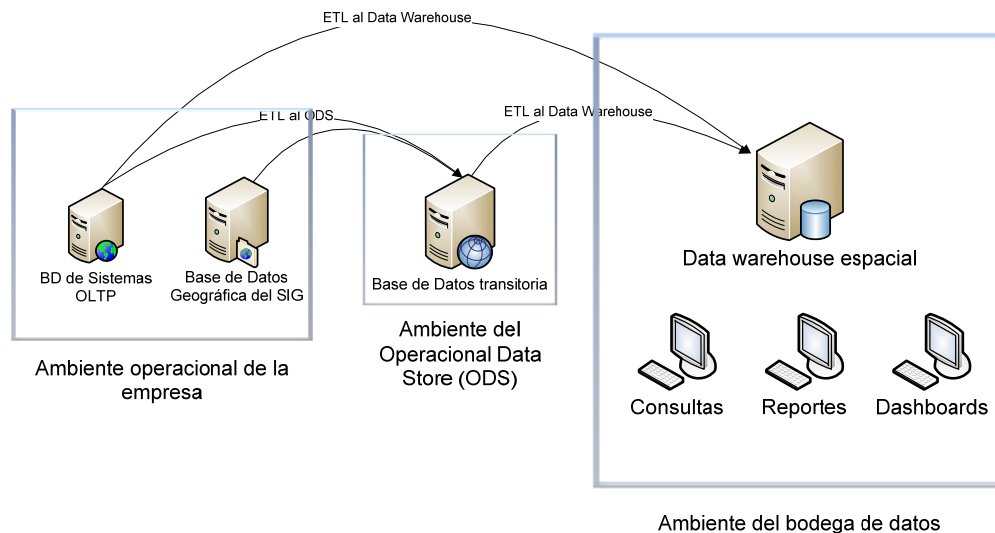


Gráfico 5. Proceso de Extracción, Transformación y Carga de los datos para alimentar la bodega.

Georeferenciación y geoprocesamiento de los datos

Consiste en georeferenciar y geocodificar la información mapificable. Para lo cual será imprescindible la utilización de la cartografía digital base y de herramientas tipo SIG.

Si la información espacial existe, se traslada al Data Warehouse definitivo, pero si no existe, es necesario el establecimiento de tareas para construirla, adquirirla y modelar los geoprocesos necesarios para llegar a la información necesaria como:

- Proximidad
- Overlapping
- Clip

Administración y seguridad de los datos.

Se presentan de acuerdo al ciclo de refrescamiento del Data Warehouse espacial y se deben desarrollar aplicaciones y procesos en lotes para automatizar todas estas tareas que deben

ser ejecutadas periódicamente. Y deben establecerse mecanismos de seguridad del lado de la carga y del lado del usuario.

Acceso y Análisis de los datos espaciales y tabulares

Se debe tener en cuenta que el Data Warehouse espacial no es como un base de datos común, ya que tiene un tamaño superior, que la estructura de los datos no es común, y que la información geográfica no es tabular y que por ende requerimos herramientas o componentes tipo SIG, para que se consulten los datos espaciales, se navegue, visualicen y analicen para que el cliente los interprete y procese.

Metadatos

Es importante en el ambiente de los Data Warehouse documentar todos los procesos, tipos de datos y estructuras de datos a través de metadatos, es decir los datos de los datos, los metadatos brindan información de localización, estructura, y significado de los datos, básicamente mapean los datos; en el Data Warehouse espacial encontramos diferentes tipos de Metadatos:

Los Metadatos de los procesos ETL: contienen las reglas de extracción, depuración y traslado de los datos a la bodega.

Metadatos para usuarios finales que les permitan navegar por los datos y sirven de documentación para las herramientas de análisis que se empleen.

Metadatos operacionales contienen información de tareas de carga, y procesos de acceso. Es usado para análisis sobre las condiciones de la bodega.

Metadatos geográficos documentar todos los geo-objetos, para poder desarrollar las diferentes operaciones y análisis de los mismos, por ejemplo

- Escala del mapa
- Sistema de proyección
- Fuente de los datos espaciales

Representación de la información Geográfica en un Data Warehouse Espacial

La información geográfica dentro de un Data Warehouse espacial puede ser representada:

Como un dato agregado o de referencia.

Son los datos espaciales que durante todos los procesos y operaciones de la empresa son estáticos ya que su dinámica de cambio es mínima y por lo general no depende de la organización. Por ejemplo los límites espaciales de una parroquia o una provincia.

Como una variable de acceso y análisis

Son los datos espaciales donde la dinámica de cambio depende de la dinámica del negocio. Por ejemplo la forma del crecimiento de la red de distribución eléctrica.

Especificación de requisitos del software

Una especificación de requisitos del software es una descripción completa del comportamiento del sistema a desarrollar. Las estrategias recomendadas para la especificación de los requisitos del software están descritas por IEEE 830-1998. Este estándar describe las estructuras posibles, contenido deseable, y calidades de una especificación de requisitos del software.

Los requisitos se dividen en tres tipos:

- Funcionales: son los que el usuario necesita que efectúe el software. Ej.: el sistema debe emitir un comprobante al asentar la entrega de mercadería.

- No funcionales: son los requisitos que imponen restricciones al diseño o funcionamiento del sistema (tal como requisitos de funcionamiento, estándares de calidad, o requisitos del diseño).
- Empresariales u Organizacionales: son el marco contextual en el cual se implantará el sistema para conseguir un objetivo macro. Ej.: abaratar costos de expedición.

Proceso

Un proceso puede ser definido como un conjunto de actividades enlazadas entre sí que, partiendo de una o más entradas (inputs) las transforma, generando un resultado (output).

Las actividades de cualquier organización pueden ser concebidas como integrantes de un proceso determinado. Desde este punto de vista, una organización cualquiera puede ser considerada como un sistema de procesos, con mayor o menor relación entre sí, en los que buena parte de los inputs serán generados por proveedores internos, y cuyos resultados irán frecuentemente dirigidos hacia clientes también internos.

Documentación de procesos.

Es un método estructurado que utiliza un manual preciso para comprender el contexto y los detalles de los procesos clave. Siempre que un proceso vaya a ser rediseñado o mejorado, su documentación es esencial como punto de partida.

RUP (Rational Unified Process)

El Proceso Unificado de Rational (RUP) es un proceso de desarrollo de software y junto con el Lenguaje Unificado de Modelado UML, constituye la metodología estándar más utilizada para el análisis, implementación y documentación de sistemas orientados a

objetos. El RUP no es un sistema con pasos firmemente establecidos, sino un conjunto de metodologías adaptables al contexto y necesidades de cada organización.

Lenguaje Unificado de Modelado (UML)

Es el lenguaje de modelado de sistemas de software más conocido y utilizado en la actualidad; está respaldado por el OMG (Object Management Group). Es un lenguaje gráfico para visualizar, especificar, construir y documentar un sistema. UML ofrece un estándar para describir un "plano" del sistema (modelo), incluyendo aspectos conceptuales tales como procesos de negocio y funciones del sistema, y aspectos concretos como expresiones de lenguajes de programación, esquemas de bases de datos y componentes reutilizables.

Es importante resaltar que UML es un "lenguaje de modelado" para especificar o para describir métodos o procesos. Se utiliza para definir un sistema, para detallar los artefactos en el sistema y para documentar y construir. En otras palabras, es el lenguaje en el que está descrito el modelo.

UML cuenta con varios tipos de diagramas, los cuales muestran diferentes aspectos de las entidades representadas.

Diagramas

En UML 2.0 hay trece tipos diferentes de diagramas. Para comprenderlos de manera concreta, a veces es útil categorizarlos jerárquicamente, como se muestra en el gráfico siguiente:

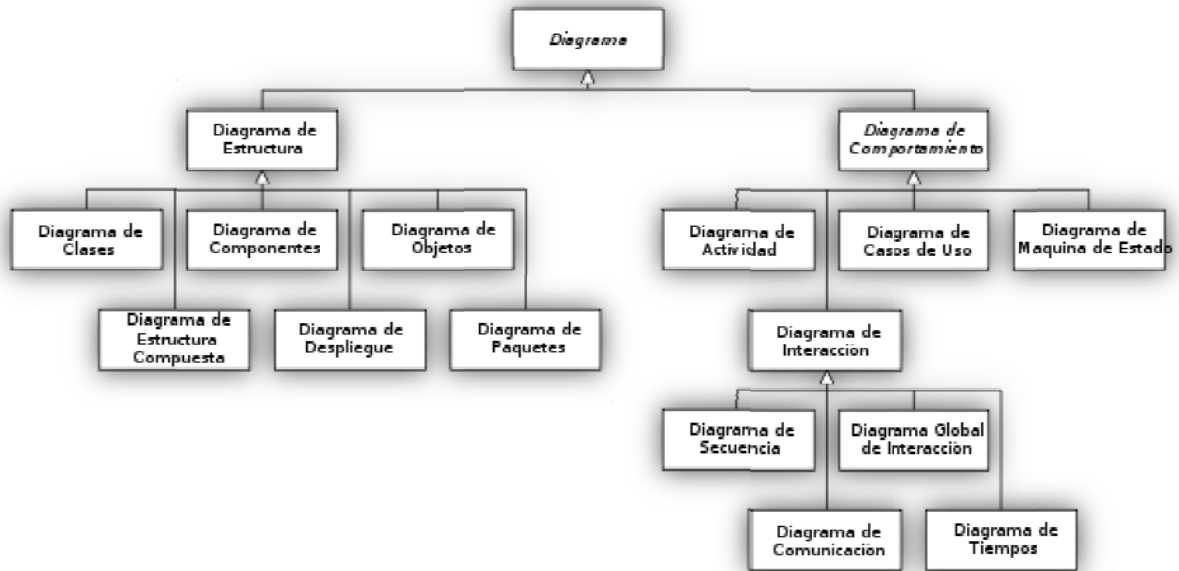


Gráfico 6. Jerarquía de los diagramas UML 2.0.

Fuente: http://es.wikipedia.org/wiki/Lenguaje_Unificado_de_Modelado

Los *Diagramas de Estructura* enfatizan en los elementos que deben existir en el sistema modelado:

- Diagrama de clases
- Diagrama de componentes
- Diagrama de objetos
- Diagrama de estructura compuesta (UML 2.0)
- Diagrama de despliegue
- Diagrama de paquetes

Los *Diagramas de Comportamiento* enfatizan en lo que debe suceder en el sistema modelado:

- Diagrama de actividades
- Diagrama de casos de uso
- Diagrama de estados

Los *Diagramas de Interacción* son un subtipo de diagramas de comportamiento, que enfatiza sobre el flujo de control y de datos entre los elementos del sistema modelado:

- Diagrama de secuencia
- Diagrama de comunicación, que es una versión simplificada del Diagrama de colaboración (UML 1.x)
- Diagrama de tiempos (UML 2.0)
- Diagrama global de interacciones o Diagrama de vista de interacción (UML 2.0)

A continuación se describirán con un poco más de detalle los tres diagramas que se utilizan en este trabajo para la documentación de los procesos ETL.

Diagrama de actividades

En el UML (Lenguaje de Modelado Unificado), un **diagrama de actividades** representa los flujos de trabajo paso a paso de negocio y operacionales de los componentes en un sistema. Un Diagrama de Actividades muestra el flujo de control general.

El propósito del diagrama de actividad es modelar un proceso de flujo de trabajo (workflow) y/o modelar operaciones. Una Operación es un servicio proporcionado por un objeto, que está disponible a través de una interfaz.

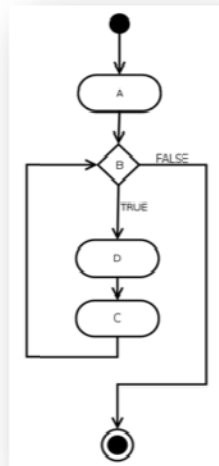


Gráfico 7. Diagrama de actividades.

Fuente: http://en.wikipedia.org/wiki/Activity_diagram

Diagrama de componentes

Un **diagrama de componentes** es un diagrama tipo del Lenguaje Unificado de Modelado. Un diagrama de componentes representa cómo un sistema de software es dividido en componentes y muestra las dependencias entre estos componentes. Los componentes físicos incluyen archivos, cabeceras, librerías compartidas, módulos, ejecutables, o paquetes. Los diagramas de Componentes prevalecen en el campo de la arquitectura de software pero pueden ser usados para modelar y documentar cualquier arquitectura de sistema.

Debido a que estos son más parecidos a los diagramas de casos de usos, son utilizados para modelar la vista estática y dinámica de un sistema. Muestra la organización y las dependencias entre un conjunto de componentes. No es necesario que un diagrama incluya todos los componentes del sistema, normalmente se realizan por partes. Cada diagrama describe un apartado del sistema.

Uno de los usos principales es que puede servir para ver qué componentes pueden compartirse entre sistemas o entre diferentes partes de un sistema.

Diagrama de componentes

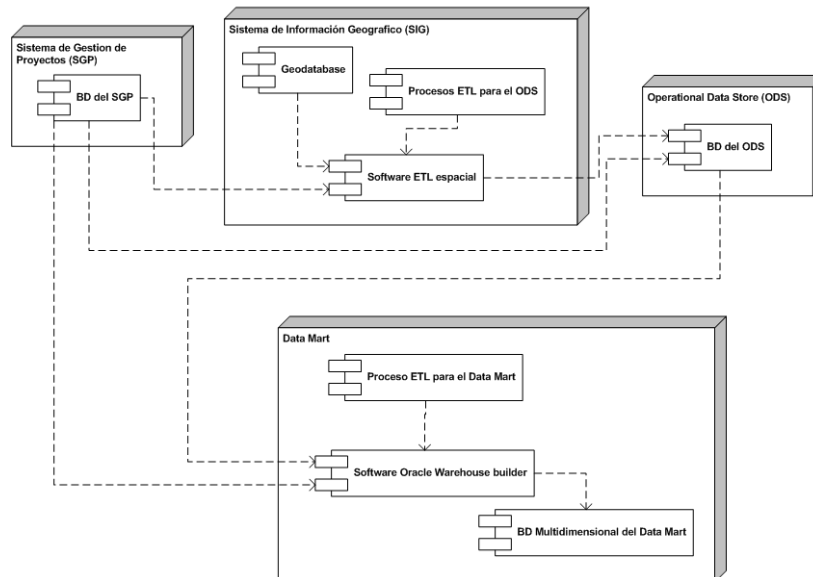


Gráfico 8. Diagrama de componentes.

Diagrama de Despliegue

El Diagrama de Despliegue es un tipo de diagrama del Lenguaje Unificado de Modelado que se utiliza para modelar el hardware utilizado en las implementaciones de sistemas y las relaciones entre sus componentes.

Los elementos usados por este tipo de diagrama son nodos (representados como un prisma), componentes (representados como una caja rectangular con dos protuberancias del lado izquierdo) y asociaciones.

La mayoría de las veces el modelado de la vista de despliegue implica modelar la topología del hardware sobre el que se ejecuta el sistema. Aunque UML no es un lenguaje de especificación hardware de propósito general, se ha diseñado para modelar muchos de los

aspectos hardware de un sistema a un nivel suficiente para que un ingeniero software pueda especificar la plataforma sobre la que se ejecuta el software del sistema.

Diagrama de despliegue

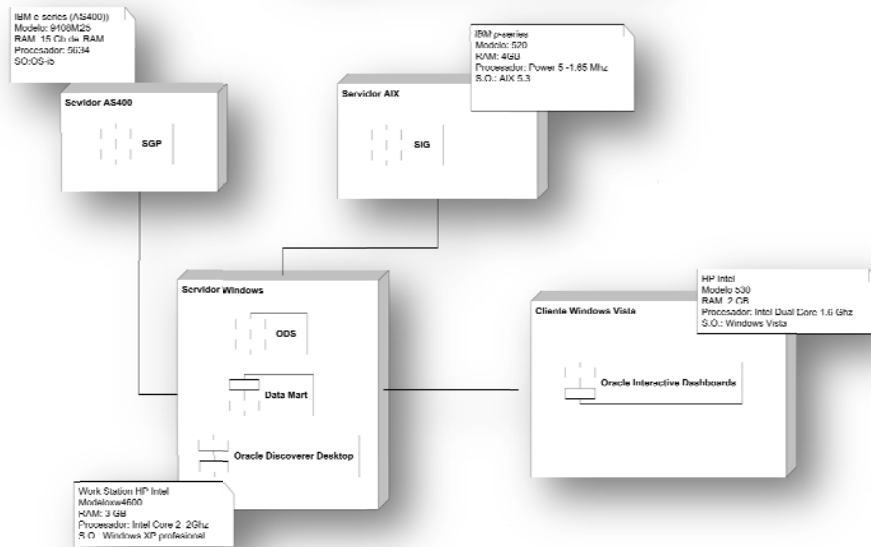


Gráfico 9. Diagrama de Despliegue.

3. Especificación de Requerimientos

A continuación se presenta el SRS (software requirements specification), levantado para la implementación del prototipo de Data Mart, basado en el estándar IEEE 830-1998 de especificación de requerimientos. Es importante aclarar que solo se han tomado los componentes del estándar IEEE 830 que se ajustan principalmente a la documentación de requerimientos para los procesos, ya que se trata de un prototipo para demostrar la funcionalidad de los geoprosesos y no hubo requisitos estrictos de usabilidad, escalabilidad o disponibilidad, sino que se centró en encontrar rápidamente una herramienta compatible con el SIG y el SGP que demuestre al personal directivo la potencialidad para la toma de decisiones que representa un Data Mart.

3.1 Introducción

3.1.1 Propósito: Los SRS son especificaciones para un producto del software en particular, programa, o juego de programas que realizan ciertas funciones en un ambiente específico. Este SRS fue desarrollado por el autor de la presente tesis y tiene el propósito de aclarar los requerimientos para el prototipo de Data Mart.

Debido a que el SRS tiene un papel específico en el proceso de desarrollo de software, al definir el SRS se debe tener el cuidado para no ir más allá de los límites de esa tarea, por lo tanto debido a las característica del proyecto, se planteo la implementación de un prototipo y no una solución de software específica y completa, el levantamiento de los requisitos fue poco estricto, es por esa raso que el presente SRS, está bastante acotado al levantamiento de requerimientos básico.

La audiencia a la que va dirigido el presente SRS, es al desarrollador del presente proyecto de tesis como insumo para la implementación del prototipo de Data Mart

y a los lectores de este trabajo de tesis, para un mejor entendimiento de los requisitos del prototipo.

3.1.2 Alcance: El alcance del prototipo de Data Mart en la organización está enmarcado en la Dirección de Distribución, específicamente en el Departamento SIGADE:

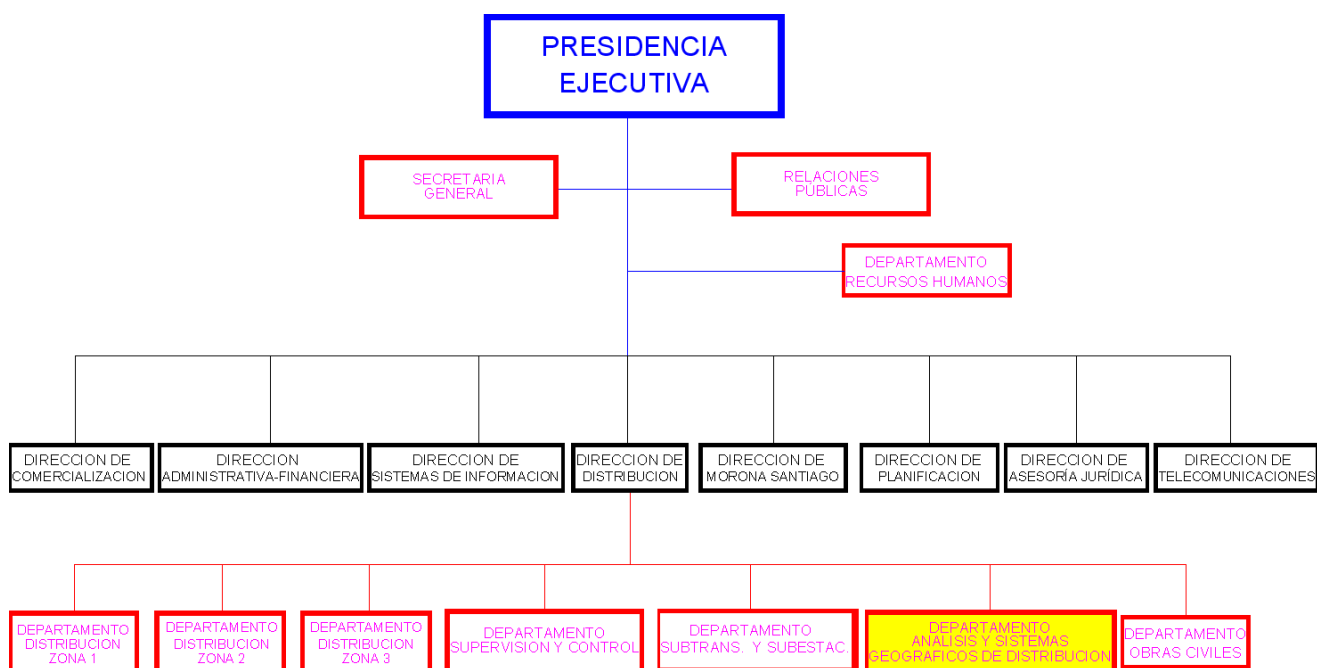


Gráfico 10. Organigrama de la Empresa.

El alcance del producto está orientado a demostrar el funcionamiento de los procesos y geoprocursos para las tareas de extracción, transformación de la información espacial del SIG en información alfanumérica para ser cargada en un Data Mart y la presentación de algunos reportes de análisis multidimensionales.

No se implementarán funciones de minería de datos, ni se presentarán los resultados de los análisis en mapas. Tampoco se pretende hacer una evaluación y selección de la herramienta de Data Mart o SIG, ni establecer requerimientos no funcionales.

La descripción de los objetivos y metas del presente prototipo al igual que los beneficios están enmarcados en el presente trabajo de tesis en las secciones 1 y 6.6 respectivamente, por lo tanto no se las repite en esta sección.

Básicamente el prototipo contara con cuatro módulos, con sus respectivas aplicaciones:

1. El ArcGIS Data Interoperability.- Es una extensión para el SIG (ArcGIS) de la CentroSur y es la herramienta utilizada para realiza los geoprosos de extracción y transformación de la información espacial para ser utilizada como fuente para el Data Mart.
2. El Oracle Warehouse Builder y Oracle BI Administration Tool.- que son las herramientas para el mantenimiento del Data Mart y los procesos ETL transaccionales.
3. El Oracle BI Answer.- Es la interfaz en ambiente web, donde los usuarios de negocios pueden armar sus reportes y análisis.
4. El Oracle BI Interactive Dashboards.- Es un tablero de control también en ambiente web, donde el personal directivo puede ver el esto actual de su área de negocio y reaccionar inmediatamente ante alguna alarma.

3.1.3 Definiciones, Siglas y Abreviaturas:

Esta sección tiene por objetivo proporcionar las definiciones de todas las condiciones, las siglas y abreviaciones que exigen interpretar el SRS propiamente. Para ver la lista de definiciones, siglas y abreviaturas refiérase a la sección “Glosario” del presente trabajo de tesis.

3.1.4 Referencias: En esta sección se lista las fuentes de información utilizadas para comprender en su totalidad el proceso a automatizar, refiérase a la sección “Bibliografía” del presente trabajo de tesis

3.2 Descripción global

3.2.1 Perspectiva del producto

Esta sección del SRS se debe poner el producto en la perspectiva con otros productos relacionados y sus restricciones, las cuales se detallan a continuación en la siguiente clasificación:

- Las interfaces del Sistema

El prototipo de Data Mart debe interactuar con los sistemas de información geográfica (SIG) y el sistema de gestión de proyectos (SGP), que serán sus fuentes de datos. Como referencia a la relación con otros sistemas véase la sección 6.4 “Arquitectura del Sistema” del presente trabajo de tesis, para ver los diagramas de componentes y de despliegue.

- Las interfaces del Usuario

Al tratarse de un prototipo, las características de la configuración (por ejemplo, formatos de la pantalla requeridos, página o esquemas de la ventana, los reportes o menús) necesarias para lograr los requisitos del software quedaron a criterio del desarrollador y no hubo detalles específicos. El único requerimiento que se solicitó

específicamente es el hecho de que se requiere que la interfaz del usuario sea en ambiente web.

- Las interfaces del Hardware

Como elementos de hardware se cuenta con:

- Un equipo de cómputo HP XW600 con las siguientes características:

- Procesador Intel Core 2 duo de 2Ghz
- 4 Gb de memoria RAM
- Sistema Operativo Windows XP
- Conexión de Red 10/100/1000 Mbps

- Un equipo de cómputo HP 530 con las siguientes características:

- Procesador Intel dual Core de 1.2Ghz
- 2 Gb de memoria RAM
- Sistema Operativo Windows XP
- Conexión de Red 10/100/1000 Mbps

- Las interfaces del Software

- En el equipo HP XW4600 se encuentran las siguientes aplicaciones:

- El SIG de ESRI cuyo nombre es ARCGIS versión 9.3 con sus módulos
 - ArcMap
 - ArcCatalog
 - ArcToolbox
- La base de datos integradora (ODS)

- Que utiliza el ArcGIS Data Interoperability 9.3
- En el equipo HP 530 se encuentran las siguientes aplicaciones:
 - El Oracle Warehouse Builder
 - El Oracle Business Intelligence Standard One con su versión 10g con sus módulos:
 - Oracle BI Administration Tool
 - Oracle BI Answer
 - Oracle BI Interactive Dashboards.

La descripción detallada de cada uno de estos componentes está en la sección 6.3 “Principales componentes” del presente trabajo de tesis.

- Las interfaces de Comunicaciones

Para el prototipo no se contempla restricciones en la interfaz de comunicación debido a que se trabaja en una red local categoría 5, con 100 Mbps de comunicación y solo con una muestra de los datos del SIG, que cumplen con todos los requerimientos de calidad.

- La Memoria

Con respecto a restricciones de memoria, está restringida a las especificadas para los equipos indicadas en la sección anterior.

- Los Funcionamientos

Con respecto al funcionamiento normal del prototipo se definió que las cargas al ODS se las realizarán semanalmente y las cargas al Data Mart mensualmente con

un periodo de inactividad del sistema de aproximadamente 15 minutos durante el proceso de carga del ODS y 45 minutos durante la carga del Data Mart.

3.2.2 Funciones del producto

Las funciones principales que se esperan del prototipo son:

La geoprocesos de extracción, transformación y carga de la información del SIG al ODS.

Que en la interfaz web para el usuario de negocio se le permita a este hacer análisis multidimensionales basados en la información tabular resultante de los procesos y geoprocesos ETL de la información espacial.

Presentar un tablero de control con dos indicadores que serán utilizados por la dirección departamental.

3.2.3 Características del usuario

Se espera que el usuario sea un usuario del negocio, es decir no necesariamente usuario con alguna especialización técnica, sino usuario con visión estratégica del negocio.

Una característica requerida del usuario es que sepa manejar y navegar en aplicaciones en ambiente web.

3.2.4 Restricciones

El desarrollo del prototipo presenta las siguientes restricciones:

- Se trabaja con software de evaluación y en un computador con prestaciones limitadas (procesamiento y almacenamiento) para un Data Mart.

- Los datos con los que se trabaja son solamente aquellos que están completos y depurados (de los sistemas SIG y SGP) para ser cargados en las tablas de hechos del Data Mart, por lo tanto la exactitud y confiabilidad de los resultados basado en los datos actuales del prototipo no está garantizada.

3.3 Los requisitos específicos

Esta sección del SRS debería contener los requisitos del software a un nivel de detalle suficiente para permitirles a los desarrolladores del prototipo diseñar un sistema para satisfacer esos requisitos, y a los auditores, probar que el sistema satisface esos requisitos; pero en este caso, por tratarse de un prototipo con solo el objetivo de demostrar que el diseño de Data Mart es funcional, al igual que los procesos y geoprosos ETL son aplicables, los requisitos específicos del software no se establecieron estrictamente. A continuación se indican los requisitos específicos tanto de negocios como los técnicos que se establecieron para el prototipo:

Requerimientos de negocios

- Se desea poder hacer análisis del crecimiento de las redes de distribución y sus equipos y estructuras con los siguientes análisis:
 - Totalización de km de redes de Media Tensión clasificados por Provincia y Cantón
 - Totalización de km de redes de Media Tensión clasificados por Subestación y Alimentador
 - Indicar la cantidad de transformadores y la suma de la potencia de los mismos, clasificados por Provincia y Cantón y totalizados por Fase (Monofásicos, Bifásico y Trifásicos)

- Indicar la cantidad de transformadores y suma de potencia de los mismos clasificados por Subestación y Alimentador y Totalizados por Fase (Monofásicos, Bifásico y Trifásicos)
- Totalización de la cantidad y suma de la energía por Provincia y Cantón de los clientes de la empresa.
- Totalización de la cantidad y suma de la energía por Subestación y Alimentador de los clientes de la empresa.
- Totalización de la cantidad y suma de la energía por provincia y Cantón del alumbrado público de la empresa.
- Totalización de la cantidad y suma de la energía por Subestación y Alimentador del alumbrado público de la empresa.
- Totalización de la cantidad por provincia y Cantón de los postes de la empresa.
- Totalización de la cantidad por Subestación y Alimentador de los postes de la empresa.
- Se desea poder obtener información totalizada de la inversión que la empresa ha hecho con sus proyectos clasificados por provincias, cantón, subestaciones y alimentadores.
 - Inversión en transformadores totalizada por Provincia y Cantón y por fases (Monofásicos, Bifásico y Trifásicos)
 - Inversión en transformadores totalizada por Alimentador y Subestaciones y por fases (Monofásicos, Bifásico y Trifásicos)
 - Inversión en km de tramos de media tensión clasificado por Provincia y Cantón y totalizado por Fases

- Inversión en km de tramos de media tensión clasificado por Subestación y Alimentador.

Requerimientos técnicos

Se especificó que se requiere que se trabaje con la marca Oracle, ya que es la misma base de datos utilizada para el SIG, se puede instalar en el servidor Windows de que se dispone para este prototipo y podemos bajar el software para evaluaciones debido a que estamos al día en el pago del mantenimiento con este proveedor.

Que disponga de una herramienta que realice los geoprosos de extracción y transformación de la información espacial para ser utilizada como fuente para el Data Mart.

Que disponga de una interfaz en ambiente web, donde los usuarios de negocios pueden armar sus reportes y análisis. Basado en mínimo tres cubos de información y pueda generar de igual manera tres tipos de reportes o análisis multidimensionales.

Que disponga de un tablero de control en ambiente web, donde el Director puede ver dos indicadores del estado actual de su área de negocio.

Es importante anotar que un listado detallado de los requisitos de los clientes y proveedores de los procesos están documentados en cada uno de los mismos en la sección 5 de este documento.

4. Análisis y Diseño Lógico del prototipo

Misión del Área de Negocios: Departamento SIGADE

Recopilar, procesar, actualizar, analizar y difundir la información técnica, administrativa, económica y espacial, con calidad y oportunidad, para facilitar la toma de decisiones de la dirección y de la administración, a fin de optimizar la gestión técnica y administrativa que beneficie al cliente externo e interno.

Estrategia

La estrategia propuesta tiene el enfoque Ralph Kimball, conocido autor en el tema de los data Warehouse, quien determinó que un data warehouse no era más que: "la unión de todos los Data marts de una entidad", porque nos permite la creación de un Data Mart independiente de forma iterativa y aplicando el ciclo de vida dimensional del negocio.

Macro procesos beneficiados con la arquitectura.

Los dos macro procesos beneficiados con esta solución son:

- El proceso de evaluación y aprobación económica de proyectos.
- El proceso de registro del inventario de operación (SIG).

Usuario con Acceso

Los usuarios con acceso al sistema están clasificados en aquellos llamados

Operativos.- donde están los ingenieros eléctricos de la Dirección, cuyas actividades son repetitivas y a los cuales les interesa la información proporcionada por el ODS en sus ámbitos más específicos y con mayor nivel de detalle.

Tácticos.- donde están los Jefes Departamentales, que son los que utilizan la información del Data Mart para ver si sus objetivos propuestos para el plan operativo anual se están cumpliendo y si se encuentra dentro del rango de sus indicadores.

Estratégicos: donde está el Director de Distribución, el cual utiliza la información del Data Mart para tomar decisiones gerenciales orientadas a la visión global de la empresa y de sus requerimientos, soportados por datos históricos.



Gráfico 11. Usuario con acceso al Data Mart.

Tipo de información para cada usuario

La información debe estar orientada a las preguntas estratégicas del negocio para cada nivel de usuario como se indica a continuación:

Para usuario no frecuente (nivel 1)

La información para este tipo de usuario le sirve para realizar consultas esporádicas más vinculadas al día a día del negocio, y en ellos existe una baja demanda de herramientas de análisis y reporte del Data Mart.

Por ejemplo: información que le permita saber cuáles fueron los proyectos aprobados en el mes de julio en un determinado Cantón.

Para usuario que requiere consultas ad-hoc y capacidades de análisis (nivel 2)

Toda la información del Data Mart le puede ser útil en algún momento a este tipo de usuario, éste necesita capacidades de disponibilidad, desempeño con cortes verticales (slice) y horizontales (dice) desde diferentes perspectivas, en cualquier momento, donde las consultas buscan responder a una pregunta de negocio.

Por ejemplo: ¿Cuáles son los alimentadores en los que se ha invertido más dinero en los últimos tres años?

Para usuarios analista de negocio (avanzado, nivel 3)

La información para este tipo de usuario le permite contestar a preguntas más estratégicas asociadas con pronósticos o análisis de tendencias, al cual todos los datos del Data Mart le son potencialmente importantes y generalmente lo utilizan para realizar análisis especializados sobre datos preprocesados.

Por ejemplo: ¿Cuál fue la tendencia de crecimiento de los proyectos de electrificación en la provincia de Cañar en el último trimestre?

Modelo Lógico del Data Mart

El diseño de este prototipo de Data Mart está basada en la metodología sugerida por Oracle en su documento “Oracle Business Intelligence Estándar Edition One Tutorial” [8] en su anexo “Design the Data Mart” específicamente en la sección B.3.1 “Creating a Logical Design”. Como resultado de la aplicación de esta metodología a continuación se presentan los modelos lógicos de los seis cubos de información que conforman el prototipo de Data Mart:

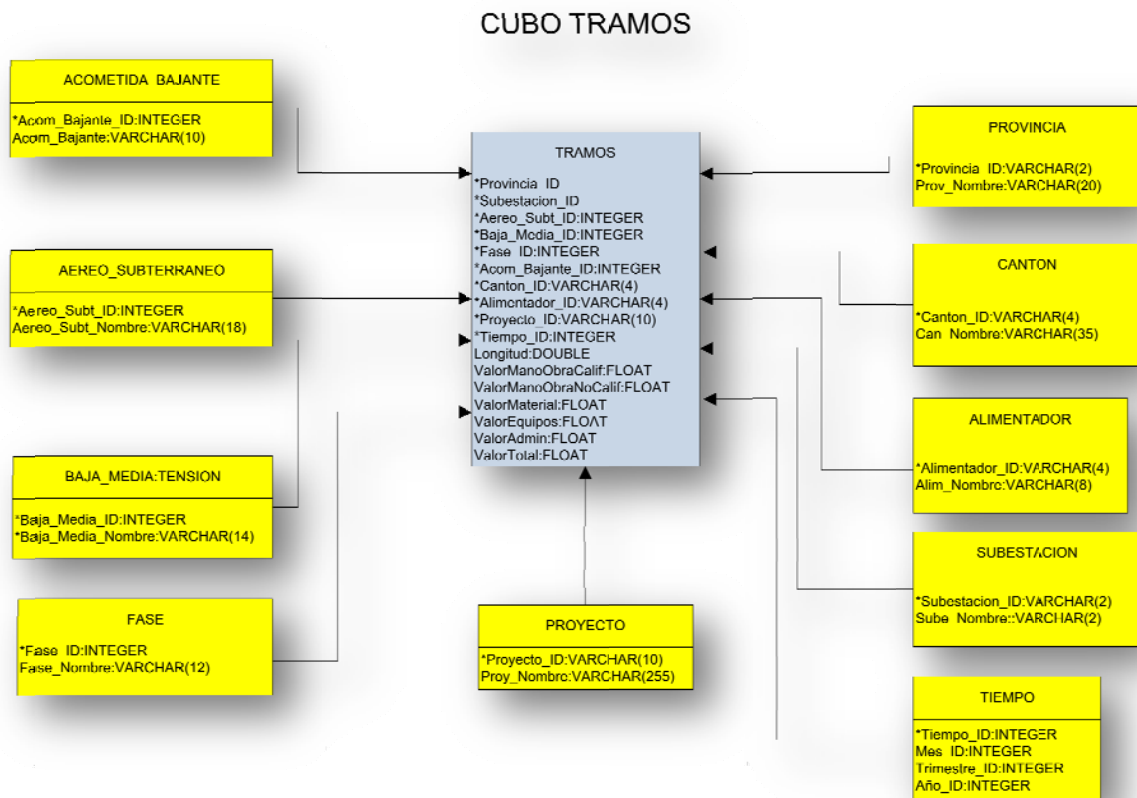


Gráfico 12. Modelo Lógico del cubo Tramos.

CUBO TRANSFORMADORES

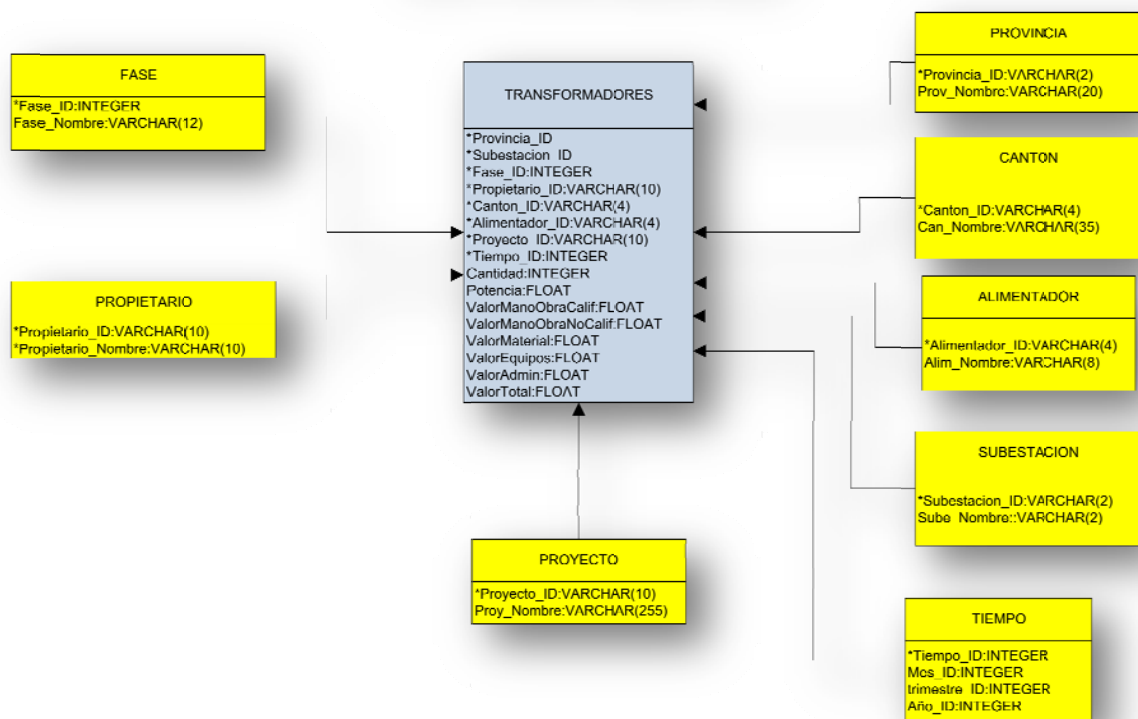


Gráfico 13. Modelo Lógico del cubo Transformadores.

CUBO LUMINARIAS

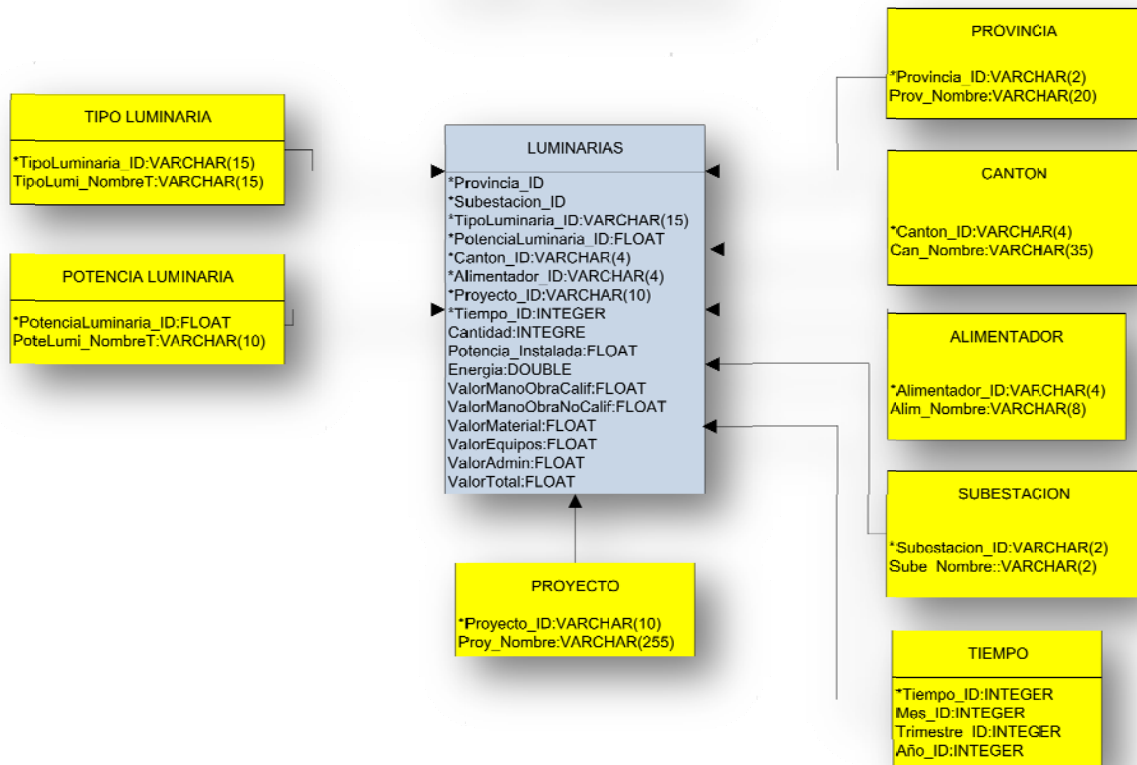


Gráfico 14. Modelo Lógico del cubo Luminarias.

CUBO CLIENTES

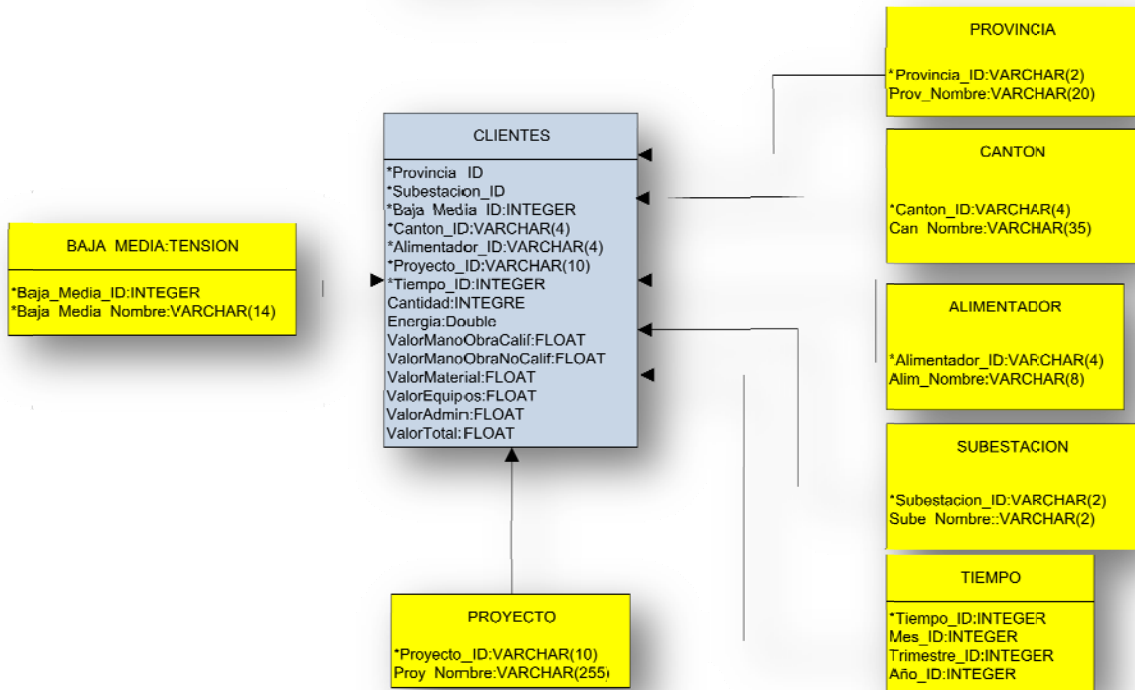


Gráfico 15. Modelo Lógico del cubo Clientes.

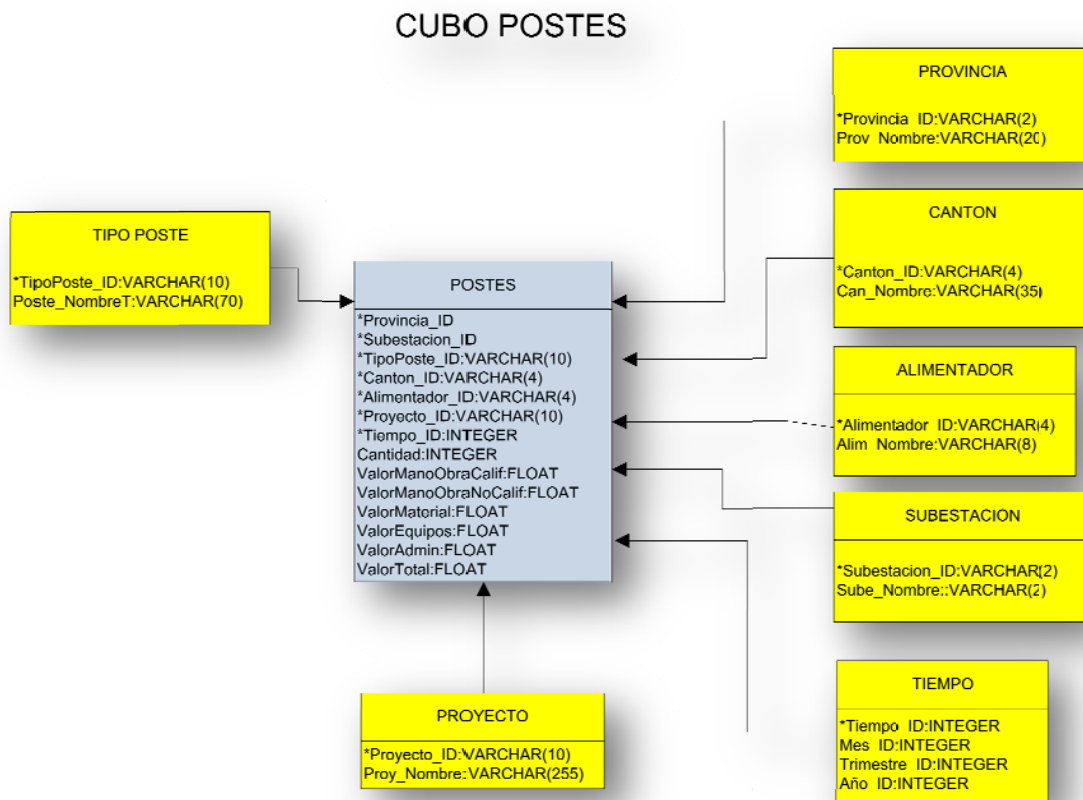


Gráfico 16. Modelo Lógico del cubo Postes.

CUBO ESTRUCTURAS EN POSTE

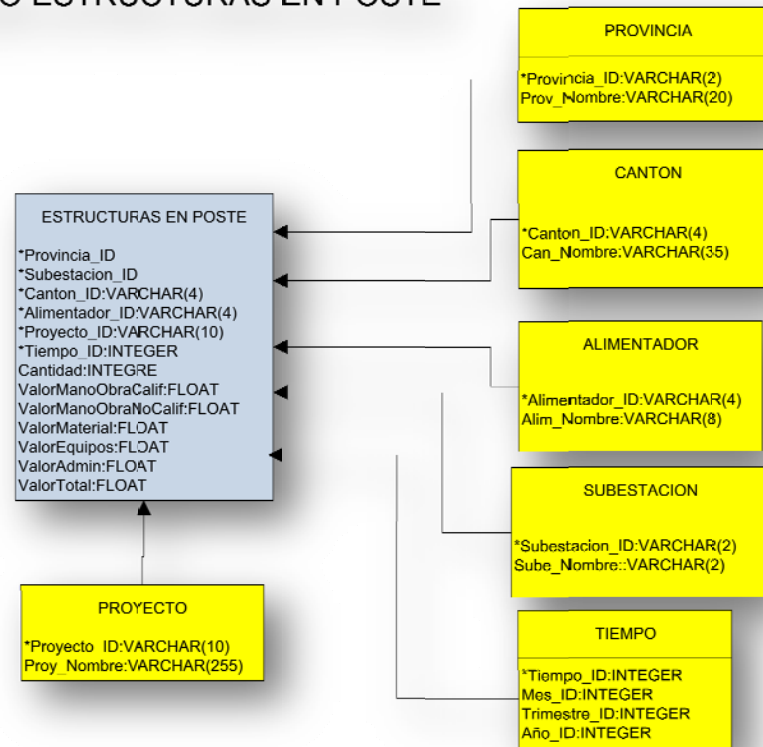


Gráfico 17. Modelo Lógico del cubo Estructuras en Poste.

Diagramas de Mapeo

A continuación se presentan los diagramas de Mapeo desde las tablas fuentes del SIG y del SGP, a las dimensiones y tablas de hechos del Data Mart:

MAPEO PARA LAS TABLAS DE HECHOS (Parte 2/2)

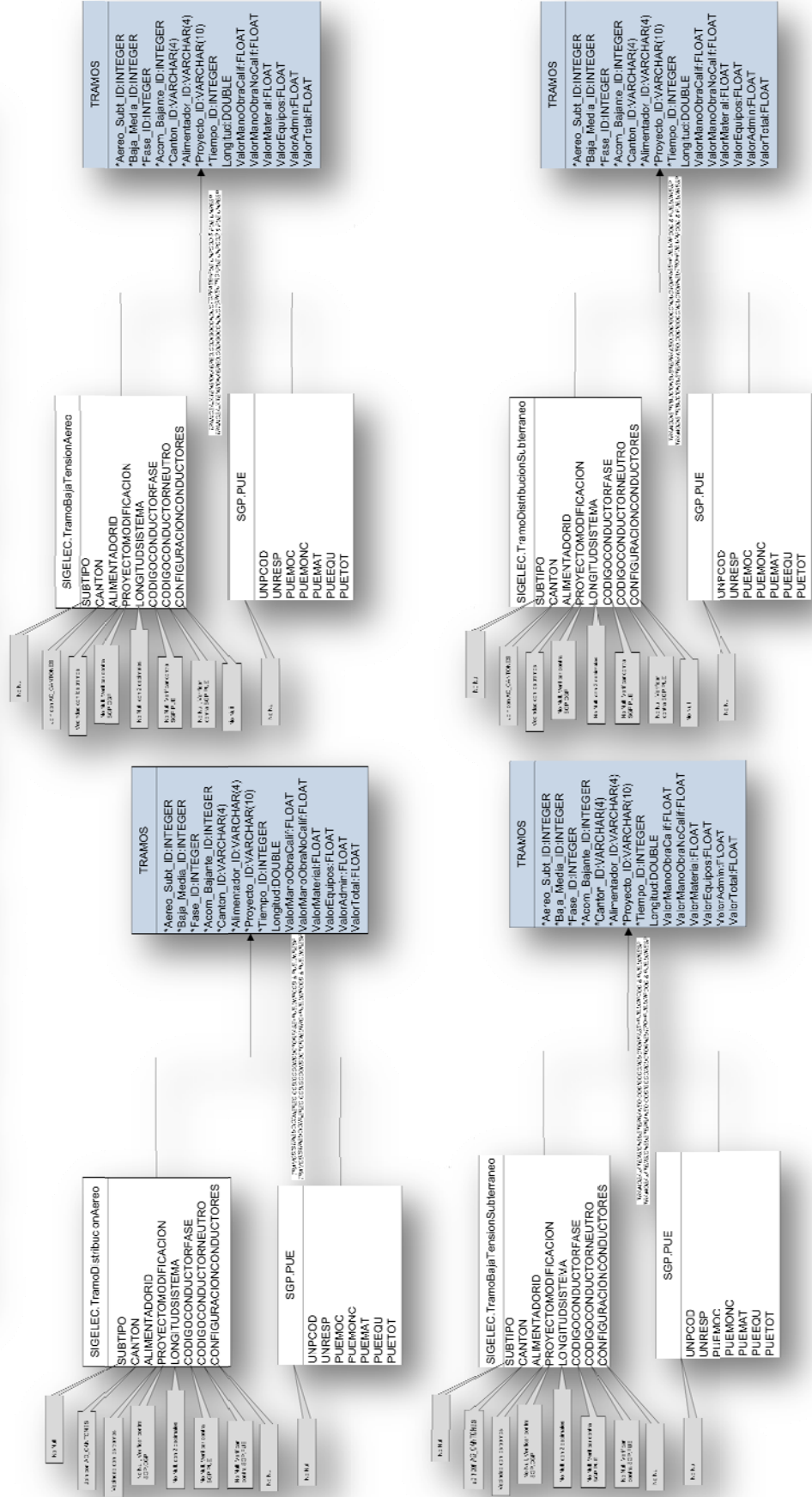


Gráfico 19. Mapeo de las tablas de hechos del Data Mart (2/2)

MAPEO DE LAS DIMENSIONES DEL DATA MART

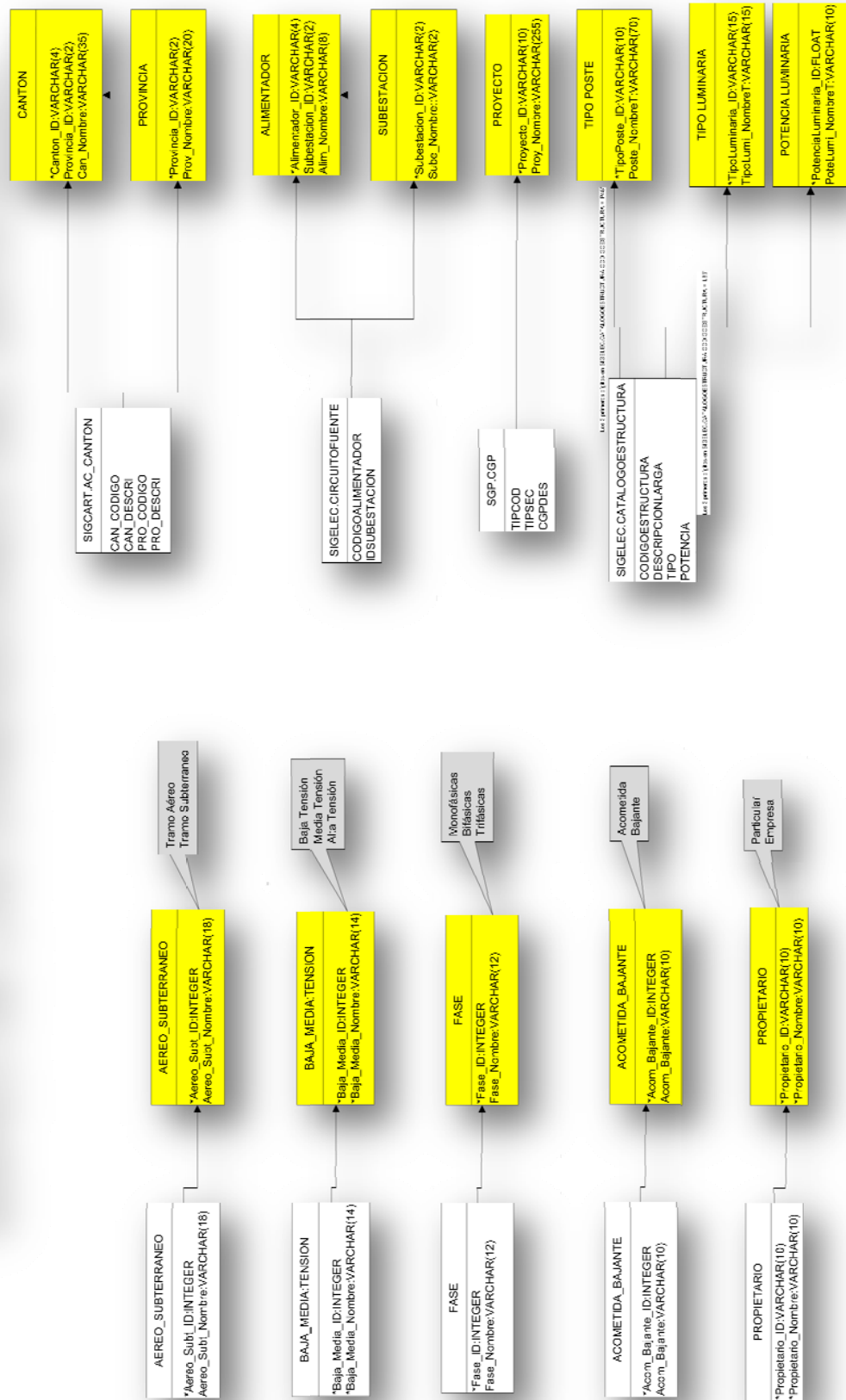


Gráfico 20. Mapeo de las dimensiones del Data Mart.

5. Procesos para el manejo de la información

Luego del diseño del Data Mart, es importante el desarrollo de los procesos para el manejo de la información tanto espacial como la del SGP. Estos procesos deben ser enfocados para el manejo tanto de la información gráfica como de la información transacción y además en las dos etapas previas al ingreso de estos al Data Mart, como son:

- Los procesos ETL espaciales para el paso de la información con característica espacial a la base de datos transitoria.
- Los procesos ETL para el llenado de las tablas de hechos en el Data Mart.

El grafico a continuación ilustra la aplicación de los dos procesos mencionados:

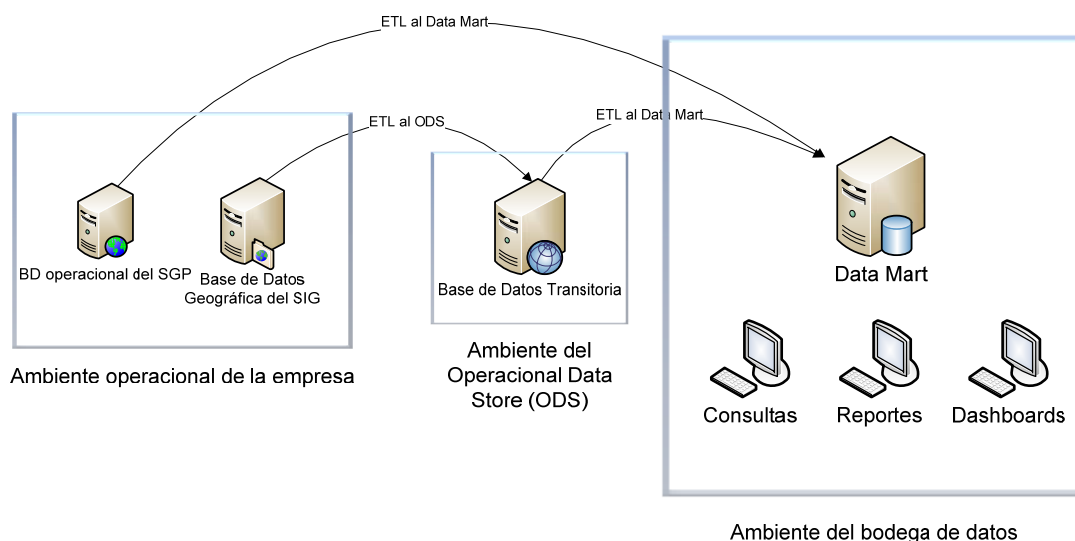


Gráfico 21. Procesos ETL para el Data Mart

(ETL al ODS y ETL al Data Mart).

Se han desarrollado y documentado los procesos más importantes para las tareas de extracción, transformación y carga de la información en estas dos etapas. La metodología de documentación utilizada es la adoptada por la Empresa Eléctrica Regional Centro Sur, con la ampliación de algunos ítems adicionales que se han tomado de otras metodologías de procesos como RUP y OSSAD, estos procesos se detallan a continuación:

5.1 Proceso de manejo de la información tabular y espacial en la etapa ETL para el Operacional Data Store (ODS).

Existen cinco procesos en la etapa ETL hacia la base de datos transitoria la cual, por sus características, puede ser utilizada como un Operacional Data Store (ODS) y es así cómo la llamaremos de ahora en adelante. Estos procesos se detallan a continuación:

5.1.1 Proceso ETL para los TRAMOS del ODS

5.1.1.1 Breve descripción del proceso:

Este proceso sirve para la extracción, transformación, depuración y carga de la información de las cuatro capas de tramos desde SIG al ODS en una sola capa llamada TRAMOS, además de la integración de la información de las tablas CGP y PUE del Sistema de Gestión de Proyectos.

Las cuatro capas espaciales mencionadas son feature class de líneas y se denominan:

- TramoBajaTensionAereo
- TramoBajaTensionSubteraneo
- TramoDistribucionAereo y
- TramoDistribucionSubteraneo

Estas capas están almacenadas en un feature dataset en la geodatabase del SIG de la EERCS cuyo DATUM es WGS84 y que se denomina “SIGELEC.Electrico”

La capa “TRAMOS” esta almacenada en el ODS y mantiene la misma proyección espacial que los datos fuentes.

5.1.1.2 Responsable:

El responsable de este proceso es un rol llamado Superintendente Informático, el cual es el administrador del SIG y, solo para el caso del prototipo, también será el administrador del Data Mart.

5.1.1.3 Objetivo y Alcance:

El objetivo general de este proceso es crear o actualizar los datos espaciales de la capa “TRAMOS” en el ODS. Los objetivos específicos de este proceso son:

- Crear la capa TRAMOS, que es un feature class de líneas, donde se almacena la información depurada de todos los tramos de la red eléctrica.
- Crear la capa TRAMOS_SinProyectoModificacion, que es un feature class de líneas, donde se almacenan los tramos que no tienen lleno el campo “ProyectoModificacion”. Para su respectiva corrección por parte de los Auxiliares SIG.
- Crear la capa TRAMOS_SinCodigoConductor, que es un feature class de líneas, donde se almacenan los tramos que no tienen lleno los campos “CodigoConductorFase” y “CodigoConductorNeutro”. Para su respectiva corrección por parte de los Auxiliares SIG.
- Crear la capa TRAMOS_SinCanton, que es un feature class de líneas, donde se almacenan los tramos que no tienen lleno los campos “Provincia” o “Canton”. Para su respectiva corrección por parte de los Auxiliares SIG.
- Crear la capa TRAMOS_SinAlimentador, que es un feature class de líneas, donde se almacenan los tramos que no tienen lleno el campo Alimentador. Para su respectiva corrección por parte de los Auxiliares SIG.

5.1.1.4 Clientes (del proceso) y sus Requisitos:

Los clientes del proceso actual son:

- Los procesos ETL para los Trafos del ODS.- el único requisito de este procesos es que la capa “Tramos” tenga lleno el campo Alimentador, el cual lo utilizará en una geoprocesos espacial para especificar a qué alimentador pertenecen los Transformadores.
- Los procesos ETL para los Clientes del ODS.- el único requisito de este procesos es que la capa “Tramos” tenga lleno el campo Alimentador, el cual lo utilizará en un geoproceso espacial para especificar a qué alimentador pertenecen los Clientes.
- Los procesos ETL para las Luminarias del ODS.- el único requisito de este procesos es que la capa “Tramos” tenga lleno el campo Alimentador, el cual lo utilizará en una geoproceso espacial para especificar a qué alimentador pertenecen las Luminarias.
- Los procesos ETL para los Postes con Estructuras en Poste del ODS.- el único requisito de este procesos es que la capa “Tramos” tenga lleno el campo Alimentador, el cual lo utilizará en un geoproceso espacial para especificar a qué alimentador pertenecen los Postes y sus estructuras.
- Además del proceso ETL para el Data Mart, el requerimiento de este proceso es que la capa Tramos tenga llenos todos los campos especificados en la plantilla para la capa “TRAMOS” indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1.
- Los Auxiliares SIG también son clientes de este proceso para hacer las correcciones a la información utilizando el software SIG, el requerimiento de

este rol es que las tablas de las capas que se nombran a continuación tengan llenos todos los campos especificados en las plantillas respectivas indicadas en la sección “Documentos y/o plantillas relacionadas” del Anexo 1 para las capas:

- TRAMOS_SinAlimentador
- TRAMOS_SinProyectoModificacion
- TRAMOS_SinCodigoConductor
- TRAMOS_SinCanton.

5.1.1.5 Proveedores (del proceso) y sus Requisitos

Los proveedores de este proceso son dos:

5.1.1.5.1 *El Sistema de Información Geográfico (SIG)*

Los requisitos de este proveedor están relacionados con las capas que se utilizan de este sistema para la depuración de la información de las redes y carga en la capa “TRAMOS” del ODS. Estas capas son cinco y todas deben estar en la misma proyección espacial y unidades de medida:

- a) Las primeras cuatro capas, que a su vez contienen sus respectivas tablas, son:
- TramoBajaTensionAereo
 - TramoBajaTensionSubterraneo
 - TramoDistribucionAereo
 - TramoDistribucionSubterraneo

Las mismas que están almacenadas en el SIG, contienen la información de las redes de distribución de la empresa eléctrica y los valores en sus campos deben estar acorde con las plantillas respectivas indicadas en la sección “Documentos

y/o plantillas relacionadas” del Anexo 1, además de cumplir los siguientes requisitos:

- El campo SUBTIPO, no puede ser nulo ya que este sirve para clasificar las redes en:
 - Tramos, Bajantes y Acometidas
 - Además de indicar si son Monofásicos, Bifásicos o Trifásicos.
- El campo Alimentador debe contener el código correspondiente al respectivo alimentador y subestación, no puede ser nulo.
- El campo ProyectoModificacion debe contener el código correspondiente al último proyecto, en la tabla CGP del Sistema de Gestión de Proyectos, que modificó los elementos de la respectiva red y no puede ser nulo.
- El campo CodigoConductorFase debe contener el código correspondiente al catálogo de estructuras, en la tabla PUE del Sistema de Gestión de Proyectos, para el conductor Fase del elemento de la respectiva red y no puede ser nulo.
- El campo CodigoConductorNeutro debe contener el código correspondiente al código del catálogo de estructuras, en la tabla PUE del Sistema de Gestión de Proyectos, para el conductor Neutro del elemento de la respectiva red y no puede ser nulo.
- El campo LongitudSistema, que contiene la longitud del respectivo tramo, no puede tener valores nulos.
- El campo “ConfiguracionConductores”, que contiene el dato de cómo está configurada la red que está representada por una sola línea en el SIG. Este campo no debe tener valores nulos.

- b) La última capa necesaria para este proceso es la capa “AC_CANTONES” la cual es una capa de polígonos que contiene la división política de los cantones del país y será utilizada en los geoprocursos para la asignación de la información de la provincia y cantón en la capa Tramos resultante. Los valores en los campos de esta tabla deben estar acorde con la plantilla para la capa “AC_CANTONES” indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1.

5.1.1.5.2 *El Sistema de Gestión de Proyectos (SGP)*

Los requisitos para este proveedor están relacionados con las tablas que se utilizan de este sistema para la depuración y carga en la capa “TRAMOS” del ODS de la información de las redes. Estas tablas están almacenadas en el servidor AS400 en producción y son las siguientes:

- a) La tabla CGP, que contiene la información de los proyectos de electrificación y donde los valores de sus campos no pueden ser nulos y deben estar acuerdo al formato especificado para la tabla CGP indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1, además de cumplir los siguientes requisitos:
- TipCod: que corresponde al tipo de proyecto, debe estar lleno con el respectivo tipo de proyecto, Urbano o Rural.
 - CGPSeq: que es un número secuencia, que junto con el campo TipCod conforman el código del proyecto, debe estar lleno con el correspondiente valor del proyecto.
 - CGPDes: que corresponde con el nombre del proyecto, esté lleno con el respectivo nombre del proyecto.
 - CGPFEC: que corresponde a la fecha de construcción del proyecto debe estar lleno con la respectiva fecha de construcción del proyecto.

b) La tabla PUE, que contiene la información de las unidades constructivas y sus costos y donde los valores de sus campos no pueden ser nulos y deben estar acuerdo al formato especificado para la tabla PUE, indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1, además de cumplir los siguientes requisitos:

- UNPCOD que contiene el código de las unidades de propiedad.
- UNRESP que es un secuencial que junto con UNPCOD conforman el código de estructura.
- PUEMOC el cual contiene el costo de mano de obra calificada para su respectiva estructura.
- PUEMONC el cual contiene el costo de mano de obra no calificada para su respectiva estructura.
- PUEMAT el cual contiene el costo del material para su respectiva estructura.
- PUEEQU el cual contiene el costo del equipo utilizado para su respectiva estructura.
- PUEADMT el cual contiene el costo de administración para su respectiva estructura.

5.1.1.6 Descripción detallada del proceso:

El proceso de ETL para los TRAMOS del ODS se lo detalla a continuación como actividades:

1. Leer las tablas de las capas TramoBajaTensionAereo, TramoBajaTensionSubterraneo, TramoDistribucionAereo,

TramoDistribucionSubterraneo desde el Sistema de Información Geográfico y mediante un “merge” espacial unificarlas para manejarlas como una sola capa.

2. Llena los campos Aereo_Subt_ID y Baja_Media_ID para cada registro de las tablas fuente de acuerdo al nombre de la capa de la que proviene, cuyos valores se especifican en la Tabla siguiente:

| Tabla Fuente | Aereo_Subt_ID | Baja_Media_ID |
|-------------------------------------|----------------------|----------------------|
| TramoBajaTensionAereo | 1 | 1 |
| TramoBajaTensionSubterraneo | 2 | 1 |
| TramoDistribucionAereo | 1 | 2 |
| TramoDistribucionSubterraneo | 2 | 2 |

Tabla 2. Valores asignados a los campos Aereo_Subt_ID y BajaMedia_ID de acuerdo al nombre de las tablas fuente.

3. Llena el campo Fase_ID para cada registro de las tablas fuente de acuerdo al nombre al campo "Subtipo", cuyos valores se especifican en la Tabla siguiente:

| Subtipo | Campo Subtipo | Fase_ID |
|----------------|--------------------------|----------------|
| 1 | Tramo BTA Monofásico | 1 |
| 4 | Bajante BTA Monofásico | 1 |
| 7 | Acometida BTA Monofásica | 1 |
| 2 | Tramo BTA Bifásico | 2 |
| 5 | Bajante BTA Bifásica | 2 |
| 8 | Acometida BTA Bifásica | 2 |
| 3 | Tramo BTA Trifásico | 3 |
| 6 | Bajante BTA Trifásica | 3 |
| 9 | Acometida BTA Trifásica | 3 |

Tabla 3. Valor asignado a lo campo Fase_ID de acuerdo al campo "Subtipo" de la tabla fuente.

4. Llena el campo Acom_Bajante_ID para cada registro de las tablas fuente de acuerdo al nombre al campo "Subtipo"

| Subtipo | Campo Subtipo | Acom_Bajante_ID |
|---------|--------------------------|-----------------|
| 1 | Tramo BTA Monofásico | 1 |
| 2 | Tramo BTA Bifásico | 1 |
| 3 | Tramo BTA Trifásico | 1 |
| 4 | Bajante BTA Monofásica | 2 |
| 5 | Bajante BTA Bifásico | 2 |
| 6 | Bajante BTA Trifásico | 2 |
| 7 | Acometida BTA Monofásico | 3 |
| 8 | Acometida BTA Bifásico | 3 |
| 9 | Acometida BTA Trifásico | 3 |

Tabla 4 . Valor asignado a lo campo Acom_Bajante_ID de acuerdo al campo "Subtipo" de la tabla fuente.

5. Se verifica si el campo "ProyectoModificacion", es nulo.
6. El campo "ProyectoModificacion" almacena el código de la carpeta del último proyecto con el que se realizaron modificaciones en una red, si este campo está vacío, quiere decir que no se han hecho mejoras o modificaciones a la red y se debe almacenar el código que está en el campo "Proyecto Construcción", que es el de la construcción original de la red; por lo tanto, si el campo "ProyectoModificacion" es nulo, se llena el campo "ProyectoModificacion" con el valor respectivo del campo "Proyecto Construcción".
7. Verifica que los valores en el campo "ProyectoModificacion" de la tabla TRAMOS tengan correspondencia con los campos TIPCOD+CGPSEC de la tabla CGP del Sistema de Gestión de Proyectos (SGP) y que son los que nos

servirá más adelante para tomar de la tabla CGP, la fecha de construcción de la red.

8. Si no existe correspondencia entre algunos registros de las tablas CGP y TRAMOS o si tienen valor nulo, el proceso los filtra y almacena en una capa llamada "TRAMOS_SinProyectoModificacion", que como se trata de una capa espacial, sirve para registrar dónde está ubicado el elemento que no tiene un valor adecuado en su campo "ProyectoModificacion" para que se puedan tomar los correctivos necesarios.
9. Utiliza el campo "ProyectoModificacion" para hacer un Join con la tabla CGP del SGP, para llenar los campos de "Fecha_Construccion" con el valor correspondiente del campo CGP.CGPFEC de cada tramo en la red.
10. Verifica que los valores en los campos "CodigoConductorFase" y "CodigoConductorNeutro" tengan correspondencia con el código de la estructura (UNPCOD+UNRESP) respectiva almacenada en la tabla PUE del sistema de gestión de proyectos.
11. Si no existe correspondencia entre algunos registros de las tablas TRAMOS y PUE o si tienen valor nulo, el proceso los filtra y almacena en una capa llamada "TRAMOS_SinCodigoConductor" la cual, al ser una capa espacial, sirve para registrar dónde está ubicado el elemento que no tiene un valor adecuado en sus campos "CodigoConductorFase" y "CodigoConductorNeutro" para que se puedan tomar los correctivos necesarios.
12. Utiliza el campo CodigoConductorFase para hacer un Join con la tabla PUE del SGP, para llenar los campos de costos del conductor Fase de cada tramo en la red.

13. Utiliza el campo `CodigoConductoNeutro` para hacer un Join con la tabla PUE del SGP, para llenar los campos de costos del conductor Neutro de cada tramo en la red.
14. Se realiza un geoproceso de sobre posición espacial contra la capa `AC_Cantones` para ubicar espacialmente en qué provincia y cantón está ubicada la respectiva red. Este geoproceso consiste en hacer un Join espacial de la capa `Tramos` contra la capa `AC_Cantones`, lo que da como resultado que a la capa `Tramos` se le añadan los campos de la capa `AC_Cantones` que coinciden espacialmente (los tramos que están dentro del respectivo polígono que representa los límites de un cantón). En el caso de que un tramo atraviese más de un cantón, éste es cortado por el límite de los polígonos que representan los cantones. Al final del proceso se toma los dos primeros dígitos de la capa `AC_Cantones`, para además llenar el campo provincia y de esta manera se evita tener que realizar otro geoproceso espacial para llenar el campo de la provincia en la capa `Tramos`.
15. Verifica que los campos "Provincia" y "Canton" no sean nulos. Esto podría darse por ejemplo en el caso de que por error se grafique un tramo fuera de los límites del Ecuador para donde no tenemos información en nuestra capa "AC_Cantones".
16. Si los campos "Provincia" y "Canton" son nulos en algunos registros, estos son filtrados y almacenados en una capa llamada "TRAMOS_SinCanton", que sirve para ubicar dónde están estos tramos que no caen dentro de la capa de cantones y provincias.

17. Verifica los registros de la capa Tramos donde el campo "Alimentador" es nulo, esto puede darse por algún error en la edición que no permitió actualizar automáticamente este campo.
18. Almacena en una capa temporal los tramos que no tienen lleno el campo "Alimentador"
19. Almacena en una capa temporal los tramos que sí tienen lleno el campo "Alimentador"
20. Realiza un geoproceto de proximidad espacial, entre las dos capas temporales mencionadas en los puntos 18 y 19 de esta sección, para determinar cuál es el valor del campo "Alimentador" y asignarlo al tramo vecino que no tuviere valor en este campo.
21. Luego se realiza una última verificación del campo "Alimentador" de la capa resultante del geoproceto anterior.
22. Los tramos que contienen el campo alimentador con "Nulo" son filtrados y almacenados en una nueva capa llamada "TRAMOS_SinAlimentador", Lo que puede darse en los casos de que existan tramos aislados, que se graficaron por error o realmente sean tramos desconectados.
23. Como resultado de cada una de estas actividades en este proceso, se obtiene una sola capa llamada Tramos, la cual almacena como líneas, toda la información depurada de los tramos tomados del SIG, en la cual se debe garantizar que se cumplan los criterios de aceptación mencionados en la siguiente sección.

Esta capa queda almacenada en el ODS, la cual a su vez sirve de suministro para los otros procesos en esta etapa y como fuente depurada de datos para el proceso ETL al Data Mart.

5.1.1.7 Criterios de Aceptación:

Se debe cumplir los siguientes criterios para ser considerada una carga exitosa de la capa TRAMOS en el ODS:

1. El campo Alimentador debe contener el código correspondiente al respectivo alimentador y subestación y no puede ser nulo.
2. El campo Subestación debe contener el número de subestación correspondiente a la respectiva subestación y no puede ser nulo.
3. El campo ProyectoModificacion debe contener el código correspondiente al último código de proyecto, en la tabla CGP del Sistema de Gestión de Proyectos, que modificó los elementos de la respectiva red y no puede ser nulo.
4. El campoCodigoConductoFase debe contener el código correspondiente al catálogo de estructuras, en la tabla PUE del Sistema de Gestión de Proyectos, para el conductor Fase del elemento de la respectiva red y no puede ser nulo.
5. El campoCodigoConductoNeutro debe contener el código correspondiente al catálogo de código de estructuras, en la tabla PUE del Sistema de Gestión de Proyectos, para el conductor Neutro del elemento de la respectiva red y no puede ser nulo.
6. El campo Provincia debe contener el código correspondiente a la provincia en la que está ubicado espacialmente y no puede ser nulo.
7. El campo Cantón debe contener el código correspondiente al cantón en la que está ubicado espacialmente y no puede ser nulo.
8. El campo Aereo_Subt_ID debe estar lleno solo con valores entre 1 y 2. Que indican si la red es Aérea o Subterránea.
9. El campo Baja_Media_ID debe estar lleno solo con valores entre 1 y 3. Que indica si la red es de: Baja Tensión, Media Tensión o Alta Tensión

10. El campo Fase_ID debe estar lleno solo con valores entre 1 y 3. Que indica si la red es Monofásica, Bifásica o Trifásica.
11. El campo Acom_Bajante_ID debe estar lleno solo con valores entre 1 y 3. Que indica si la red es Tramo, Bajante o Acometida.

5.1.1.8 Indicador de Excepciones

Una parte importante del proceso es determinar las excepciones del mismo, es decir los datos que no fueron transferidos por qué no cumplieron los criterios de aceptación (ver arriba), como resultado de este proceso obtenemos las capas

- TRAMOS_SinAlimentador
- TRAMOS_SinProyectoModificacion
- TRAMOS_SinCodigoConductor
- TRAMOS_SinCanton

Donde no solo podemos determinar cuáles son los tramos que no cumplieron los criterios de aceptación, sino también dónde están ubicados espacialmente, cuándo fueron graficados o modificados y quién los graficó, para poder identificar tendencias que se pueden mostrar en el transcurso del tiempo, y si se analizan de manera integral permiten mantener un diagnóstico de la situación y tomar acciones preventivas o correctivas sobre los datos transacciones y verificar si éstas fueron o no acertadas.

5.1.1.9 Glosario o Definición de Términos

SGP.- Sistema de Gestión de Proyectos.

SIG.- Sistema de Información Geográfico.

Alimentador.- El alimentador pertenece a una subestación y es la fuente de la energía a ser distribuida.

Proyecto Modificación.- Es el código de la carpeta a la que pertenecen los elementos de un proyecto de electrificación.

Código Conductor Fase.- Es el código del conductor eléctrico de la línea Fase.

Código Conductor Neutro.- Es el código del conductor eléctrico de la línea Neutro.

ODS.- Operational Data Store, Base de Datos Operacional, es una base de datos diseñada para integrar datos de múltiples fuentes para hacer más fácil el análisis y la presentación de informes. Debido a que los datos origen vienen de diferentes fuentes, la integración implica a menudo la limpieza, la solución de redundancia y comprobación de reglas de negocio para integridad. El ODS es generalmente diseñado para contener datos de bajo nivel o indivisible (como las transacciones y los precios) con una ventana de actualización muy pequeña en contraposición con el mayor volumen de datos almacenados en el Data Warehouse que generalmente es con menor frecuencia. En otras palabras, es un contenedor de datos activos, es decir operacionales que ayudan al soporte de decisiones y a la operación y está entre un OLAP y un OLTP.

Auxiliar SIG.- Es un rol dentro del departamento SIGADE encargado de la actualización de la información espacial en el SIG.

5.1.1.10 Documentos y/o plantillas relacionadas:

En el anexo 1 se presentan las estructuras de datos (plantilla) para cada una de las tablas fuente y resultantes del presente proceso. En cada plantilla se puede obtener datos como: Nombre del campo, tipo, tamaño y una pequeña descripción.

5.1.1.11 Diagramas de actividades (workflows):

A continuación se adjunta el diagrama de actividades del proceso ETL para los tramos del ODS:

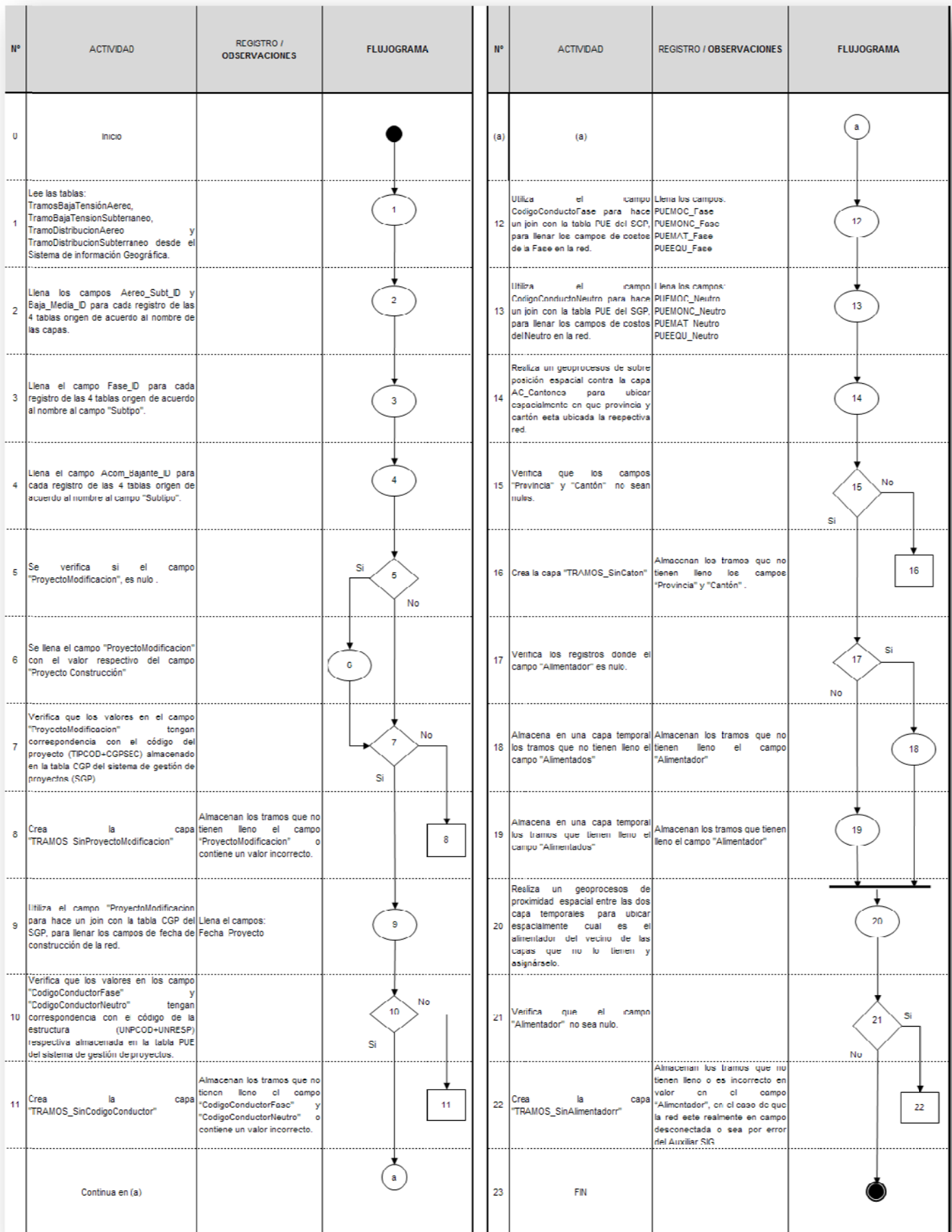


Gráfico 22. Diagrama de actividades del proceso ETL para los tramos del ODS.

5.1.1.12 Normas y políticas para el proceso:

a. Requerimientos de Información:

Los requerimientos de información fuente del SIG y SGP, deberían ser cumplidos y corregidos en caso de que no cumplan los criterios de aceptación.

b. Frecuencia:

Este proceso de actualización de la capa “Tramos” del ODS debería ser corrido todos los fines de semana para mantener la información actualizada con una ventana de tiempo pequeña.

5.1.2 Proceso ETL para los TRAFOS del ODS.

5.1.2.1 Breve descripción del proceso:

Este proceso sirve para la extracción, transformación, depuración y carga de la información espacial de la capa “PuestoTransfDistribucion” y la tabla “UnidadTransfDistribución” en el SIG, además de las tablas CGP y PUE, del sistema de gestión de proyectos, al ODS en la capa TRAFOS.

La capa “PuestoTransfDistribucion” es un feature class de puntos que está almacenada en un feature dataset cuyo DATUM es WGS84 y que se denomina “SIGELEC.Electrico”.

La tabla “UnidadTransfDistribución” esta almacenada en el SIG y existe una relación 1 a M entre la capa “PuestoTransfDistribucion” y “UnidadTransfDistribución”.

La capa “TRAFOS”, pertenece al ODS y mantiene la misma proyección espacial que sus datos fuente.

5.1.2.2 Responsable:

El responsable de este proceso es un rol llamado Superintendente Informático, el cual es el administrador del SIG y para solo para el caso del prototipo también será el administrador del Data Mart.

5.1.2.3 Objetivo y Alcance:

El objetivo general de este proceso es crear o actualizar los datos espaciales de la capa “TRAFOS” en el ODS. Los objetivos específicos de este proceso son:

- Crear la capa TRAFOS donde se almacena la información depurada de todos los transformadores de la red eléctrica.
- Crear la capa TRAFOS_SinProyectoModificacion, que es un feature class de puntos, donde se almacenan los transformadores que no tienen lleno el campo “ProyectoModificacion”, para su respectiva corrección por parte de los Auxiliares SIG.
- Crear la capa TRAFOS_SinCodigoEstructura, que es un feature class de puntos, donde se almacenan los transformadores que no tienen lleno los campos “CodigoEstructura” para su respectiva corrección por parte de los Auxiliares SIG.
- Crear la capa TRAFOS_SinCanton, que es un feature class de puntos, donde se almacenan los transformadores que no tienen lleno los campos “Provincia” o “Canton”, para su respectiva corrección por parte de los Auxiliares SIG.
- Crear la capa TRAFOS_SinAlimentador, que es un feature class de puntos, donde se almacenan los transformadores que no tienen lleno el campo Alimentador, para su respectiva corrección por parte de los Auxiliares SIG.

5.1.2.4 Clientes (del proceso) y sus Requisitos:

Los clientes del proceso actual son:

- El proceso ETL para el Data Mart, el requerimiento de este proceso es que la capa TRAFOS tenga llenos todos los campos especificados en la plantilla para la capa “TRAFOS” indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1.
- Los Auxiliares SIG también son clientes de este proceso para hacer las correcciones a la información utilizando el software SIG, el requerimiento de

este rol es que las tablas de las capas que se nombran a continuación tengan llenos todos los campos especificados en las plantillas respectivas indicadas en la sección “Documentos y/o plantillas relacionadas” del Anexo 1 para las capas:

- TRAFOS_SinAlimentador
- TRAFOS_SinProyectoModificacion
- TRAFOS_SinCodigoConductor
- TRAFOS_SinCanton.

5.1.2.5 Proveedores (del proceso) y sus Requisitos

Los proveedores de este proceso son dos:

5.1.2.5.1 *El Sistema de Información Geográfico (SIG)*

Los requisitos de este proveedor están relacionados con las capas que se utilizan de este sistema para la depuración de la información de los transformadores y carga en la capa “TRAFOS” del ODS. Estas capas son tres y deben estar en la misma proyección espacial, existe además una tabla donde se almacena la información de los transformadores. A continuación se describe éstas:

- a. La primera capa que a su vez contiene su tabla respectiva es PuestoTransfDistribucion. La misma que está almacenada en el SIG, contienen la información de la localización de los puestos de transformación de la EERCS y los valores en sus campos deben estar acorde con las plantillas respectivas indicadas en la sección “Documentos y/o plantillas relacionadas” del Anexo 1, además de cumplir los siguientes requisitos:
 - El campo SUBTIPO, no puede ser nulo ya que éste sirve para clasificar los transformadores en:
 - Monofásicos, Bifásicos o Trifásicos.

- El campo Alimentador debe contener el código correspondiente al respectivo alimentador y subestación, no debería ser nulo.
 - El campo ProyectoModificacion debe contener el código correspondiente al último proyecto, en la tabla CGP del Sistema de Gestión de Proyectos, que modificó un transformador en campo y no puede ser nulo.
 - El campo Propiedad debe informar si se trata de un transformador Particular o no, y no puede ser nulo.
- b. Tabla UnidadTransfDistribucion, la misma que está almacenada en el SIG y tiene una relación 1-M con la capa PuestoTransfDistribucion, contienen la información de la potencia y código de las unidades de transformación localizadas en el puesto de transformación de la EERCS y los valores en sus campos deben estar acorde con las plantillas respectivas indicadas en la sección “Documentos y/o plantillas relacionadas” del Anexo 1, además de cumplir los siguientes requisitos:
- El campoCodigoEstructura debe contener el código correspondiente al catálogo de estructuras, en la tabla PUE del Sistema de Gestión de Proyectos, para la unidad de transformación respectiva, y no puede ser nulo.
 - El campoCodigoUnidad debe contener el código de la unidad, que es el número pintado en la unidad de transformación en el campo, y no puede ser nulo.
 - El campoPotenciaNominal contiene el valor de la potencia nominal la unidad de transformación respectiva, y no puede ser nulo.
- c. Otra capa necesaria para este proceso en la capa “TRAMOS” la cual es una capa de líneas que contiene la información del Alimentador y será utilizada en los geoprocetos para la asignación de la información del Alimentador en la capa TRAFOS resultante.

Los valores en los campos de esta tabla deben estar acorde con la plantilla para la capa “TRAMOS” indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1.

- d. La última capa necesaria para este proceso es la capa “AC_CANTONES” la cual es una capa de polígonos que contiene la división política de los cantones del país y será utilizada en los geoprocetos para la asignación de la información de la provincia y cantón en la capa TRAFOS resultante. Los valores en los campos de esta tabla deben estar acorde con la plantilla para la capa “AC_CANTONES” indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1.

5.1.2.5.2 *El Sistema de Gestión de Proyectos (SGP)*

Los requisitos para este proveedor están relacionados con las tablas que se utilizan de este sistema para la depuración y carga en la capa “TRAFOS” del ODS de la información de las redes. Estas tablas están almacenadas en el servidor AS400 en producción y son las siguientes:

- a) La tabla CGP, que contiene la información de los proyectos de electrificación y donde los valores de sus campos no pueden ser nulos y deben estar acuerdo al formato especificado para la tabla CGP indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1, además de cumplir los siguientes requisitos:
- TipCod: que corresponde al tipo de proyecto, debe estar lleno con el respectivo tipo de proyecto, que puede ser Urbano o Rural.

- CGPSeq: que es un número secuencia, que junto con el campo TipCod conforman el código del proyecto, debe estar lleno con el correspondiente valor del proyecto.
 - CGPDes: que corresponde con el nombre del proyecto, esté lleno con el respectivo nombre del proyecto.
 - CGPFEC: que corresponde a la fecha de construcción del proyecto, debe estar lleno con la respectiva fecha de construcción del proyecto.
- b) La tabla PUE, que contiene la información de las unidades constructivas y sus costos y donde los valores de sus campos no pueden ser nulos y deben estar acuerdo al formato especificado para la tabla PUE, indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1, además de cumplir los siguientes requisitos:
- UNPCOD que contiene el código de las unidades de propiedad.
 - UNRESP que es un secuencial que junto con UNPCOD conforman el código de estructura.
 - PUEMOC el cual contiene el costo de mano de obra calificada para su respectiva estructura.
 - PUEMONC el cual contiene el costo de mano de obra no calificada para su respectiva estructura.
 - PUEMAT el cual contiene el costo del material para su respectiva estructura.
 - PUEEQU el cual contiene el costo del equipo utilizado para su respectiva estructura.
 - PUEADMT el cual contiene el costo de administración para su respectiva estructura.

5.1.2.6 Descripción detallada del proceso:

El proceso de ETL para los TRAFOS del ODS se lo detalla a continuación como actividades:

1. Leer la tabla de la capa PuestoTransfDistribucion desde el Sistema de Información Geográfico.
2. Llena el campo Fase_ID para cada registro de las tablas fuente de acuerdo al nombre al campo "Subtipo", cuyos valores se especifican en la Tabla siguiente:

| Subtipo | Campo Subtipo | Fase_ID |
|---------|---|---------|
| 1 | Puesto de Transformación Monofásico Aéreo | 1 |
| 2 | Puesto de Transformación Monofásico Subterráneo | 1 |
| 3 | Puesto de Transformación Bifásico Aéreo | 2 |
| 4 | Puesto de Transformación Trifásico Aéreo | 3 |
| 5 | Puesto de Transformación Trifásico Subterráneo | 3 |
| 6 | Puesto de Transformación Monofásico A Nivel | 1 |
| 7 | Puesto de Transformación Bifásico A Nivel | 2 |
| 8 | Puesto de Transformación Trifásico A Nivel | 3 |
| 9 | Pad Mount Monofásico | 1 |
| 10 | Pad Mount Trifásico | 3 |

Tabla 5. Valores asignados a los campos Fase_ID de acuerdo al campo Subtipo de la tabla fuente.

3. Se verifica si el campo "ProyectoModificacion" es nulo.
4. El campo "ProyectoModificacion" almacena el código de la carpeta del último proyecto con el que se realizaron modificaciones en un transformador, si este campo está vacío, quiere decir que no se han hecho mejoras o modificaciones y

se debe almacenar el código que está en el campo “Proyecto Construcción”, que es el de la construcción original de la red, por lo tanto si el campo "ProyectoModificacion" es nulo, se llena el campo "ProyectoModificacion" con el valor respectivo del campo "Proyecto Construcción".

5. Mediante un Join entre tablas se verifica que los valores en el campo "ProyectoModificacion" de la tabla TRAFOS tengan correspondencia con los campos TIPCOD+CGPSEC de la tabla CGP del Sistema de Gestión de Proyectos (SGP).
6. Los registros cuyo campo ProyectoModificacion es nulo o no tienen correspondencia en el Join del paso anterior, son almacenados en una capa llamada “Trafos_SinProyectoModificacion”, la cual es una capa espacial, que sirve para registrar dónde está ubicado el elemento que no tiene un valor adecuado en sus campos “ProyectoModificacion” para que se puedan tomar los correctivos necesarios.
7. En los registros que sí tienen correspondencia con el Join del paso 5 se actualiza el campo “FECHA_PROYECTO” con el valor correspondiente del campo CGP.CGPFEC, que es la fecha de puesta en operación del transformador.
8. Se hace un Join entre la tabla UnidadTransfDistribucion y la tabla de la capa PuestoTransfDistribucion. Para obtener el campoCodigoEstructura de la unidad.
9. Mediante un Join entre tablas se verifica los valores del campo "CodigoEstructura" de la tabla resultante del proceso anterior con el código de la estructura (UNPCOD+UNRESP) respectiva almacenada en la tabla PUE del sistema de gestión de proyectos.

10. Los registros cuyo campo `CodigoEstructura` es nulo o no tienen correspondencia en el Join del paso anterior, son almacenados en una capa llamada “`Trafos_SinCodigoEstructura`”, la cual es una capa espacial, que sirve para registrar dónde está ubicado el elemento que no tiene un valor adecuado en sus campos “`ProyectoModificacion`” para que se puedan tomar los correctivos necesarios.
11. En los registros que sí tienen correspondencia con el Join del paso 9 se actualiza los campos: `PUEMOC`, `PUEMONC`, `PUEMAT`, `PUEEQU` que son los costos de puesta en operación del transformador.
12. Se realiza un geoproceso de sobre posición espacial contra la capa `AC_Cantones` para ubicar espacialmente en qué provincia y cantón está ubicada el respectivo transformador. Este geoproceso consiste en hacer un Join espacial de la capa `PuestoTransfDistribucion` contra la capa `AC_Cantones`, lo que da como resultado que a la capa `PuestoTransfDistribucion` se le añadan los campos de la capa `AC_Cantones` que coinciden espacialmente (los transformadores que están dentro del respectivo polígono que representa los límites de un cantón). Se procede a actualizar el campo `Canton` de la tabla `TRAFOS` y al final del proceso se toma los dos primeros dígitos de la capa `AC_Cantones`, para además llenar el campo `provincia` y de esta manera se evita tener realizar otro geoproceso espacial para llenar el campo de la provincia en la capa `Trafos`.
13. Verifica que los campos “`Provincia`” y “`Canton`” no sean nulos. Esto podría darse por ejemplo en el caso de que por error se grafique un transformador fuera de los límites del Ecuador para donde no tenemos información en nuestra capa “`AC_Cantones`”.

14. Si los campos "Provincia" y "Canton" son nulos en algunos registro, estos son filtrados y almacenados en una capa llamada "TRAFOS_SinCanton", que sirve para ubicar donde están estos transformadores que no caen dentro de la capa de cantones y provincias.
15. Realiza un geoproceto de proximidad espacial, entre las capa TRAMOS y TRAFOS, para actualizar el valor del campo "Alimentador" basado en el valor de campo alimentador del tramo más cercano.
16. Luego se verifica el campo "Alimentador" de la capa resultante del geoproceto anterior, para saber cuáles son nulos.
17. Almacenan en la capa "TRAFOS_SinAlimentador" los registros donde el valor en el campo "Alimentador" es nulo o es incorrecto, en el caso de transformadores que estén realmente desconectados en campo o se ingresaron por error del Auxiliar SIG.
18. En los registros que sí tienen correspondencia con el Join espacial del paso 15 se actualiza los campos: "Alimentador".
19. Como resultado de cada una de estas actividades en este proceso se obtiene una capa llamada TRAFOS, la cual almacena como puntos, toda la información depurada de los transformadores tomados del SIG, en la cual se debe garantizar que se cumplan los criterios de aceptación mencionados en la siguiente sección. Esta capa queda almacenada en el ODS, la cual a su vez sirve de suministro y fuente depurada de datos para la el proceso ETL al Data Mart.

5.1.2.7 Criterios de Aceptación:

Se debe cumplir los siguientes criterios para ser considerada una carga exitosa de la capa TRAFOS en el ODS:

- 1 El campo Alimentador debe contener el código correspondiente al respectivo alimentador y no puede ser nulo.
- 2 El campo Subestación debe contener el número de subestación correspondiente a la respectiva subestación y no puede ser nulo.
- 3 El campo ProyectoModificacion debe contener el código correspondiente al último código de proyecto, en la tabla CGP del Sistema de Gestión de Proyectos, que modificó al elemento y no puede ser nulo.
- 4 El campo CodigoEstructura debe contener el código correspondiente al catálogo de estructuras, en la tabla PUE del Sistema de Gestión de Proyectos, para la unidad de transformación respectiva y no puede ser nulo.
- 5 El campo Provincia debe contener el código correspondiente a la provincia en la que está ubicado espacialmente y no puede ser nulo.
- 6 El campo Cantón debe contener el código correspondiente al cantón en la que está ubicado espacialmente y no puede ser nulo.
- 7 El campo Fase_ID debe estar lleno solo con valores entre 1 y 3. Que indica si la red es: Monofásica, Bifásica o Trifásica.
- 8 El campo Propiedad debe estar lleno con valores “EERCS” o “PARTICULAR”, Lo que indica si el transformador pertenece a la Empresa o no.

5.1.2.8 Indicador de Excepciones

Una parte importante del proceso es determinar las excepciones del mismo, es decir los datos que no fueron transferidos por qué no cumplieron los criterios de aceptación (ver arriba), como resultado de este proceso obtenemos las capas

- TRAFOS_SinAlimentador
- TRAFOS_SinProyectoModificacion

- TRAFOS_SinCodigoEstructura
- TRAFOS_SinCanton

Donde no solo podemos determinar cuáles son los transformadores que no cumplieron los criterios de aceptación, sino también dónde están ubicados espacialmente, cuándo fueron graficados o modificados y quien los graficó, para poder identificar tendencias que se pueden mostrar en el transcurso del tiempo, y si se analizan de manera integral permiten mantener un diagnóstico de la situación y tomar acciones preventivas o correctivas sobre los datos transacciones y verificar si éstas fueron o no acertadas.

5.1.2.9 Glosario o Definición de Términos

SGP.- Sistema de Gestión de Proyectos.

SIG.- Sistema de Información Geográfico.

Alimentador.- El alimentador pertenece a una subestación y es la fuente de la energía a ser distribuida.

Proyecto Modificación.- Es el código de la carpeta a la que pertenecen los elementos de un proyecto de electrificación.

Código Estructura.- Es el código de la estructura asignada como una unidad de Transformación.

ODS.- Operational Data Store, Base de Datos Operacional, es una base de datos diseñada para integrar datos de múltiples fuentes para hacer más fácil el análisis y la presentación de informes. Debido a que los datos origen vienen de diferentes fuentes, la integración implica a menudo la limpieza, la solución de redundancia y comprobación de reglas de negocio para integridad. El ODS es

generalmente diseñado para contener datos de bajo nivel o indivisible (como las transacciones y los precios) con una ventana de actualización muy pequeña en contraposición con el mayor volumen de datos almacenados en el Data Warehouse que generalmente es con menor frecuencia. En otras palabras, es un contenedor de datos activos, es decir operacionales que ayudan al soporte de decisiones y a la operación y está entre un OLAP y un OLTP.

Auxiliar SIG.-Es un rol dentro del departamento SIGADE encargado de la actualización de la información espacial en el SIG.

5.1.2.10 Documentos y/o plantillas relacionadas:

A el anexo 1 se presentan las estructuras de datos (plantilla) para cada una de las tablas fuente y resultantes del presente proceso. En cada plantilla se puede obtener datos como: Nombre del campo, tipo, tamaño y una pequeña descripción.

5.1.2.11 Diagramas de actividades (workflows):

A continuación se adjunta el diagrama de actividades del proceso ETL para los transformadores del ODS:

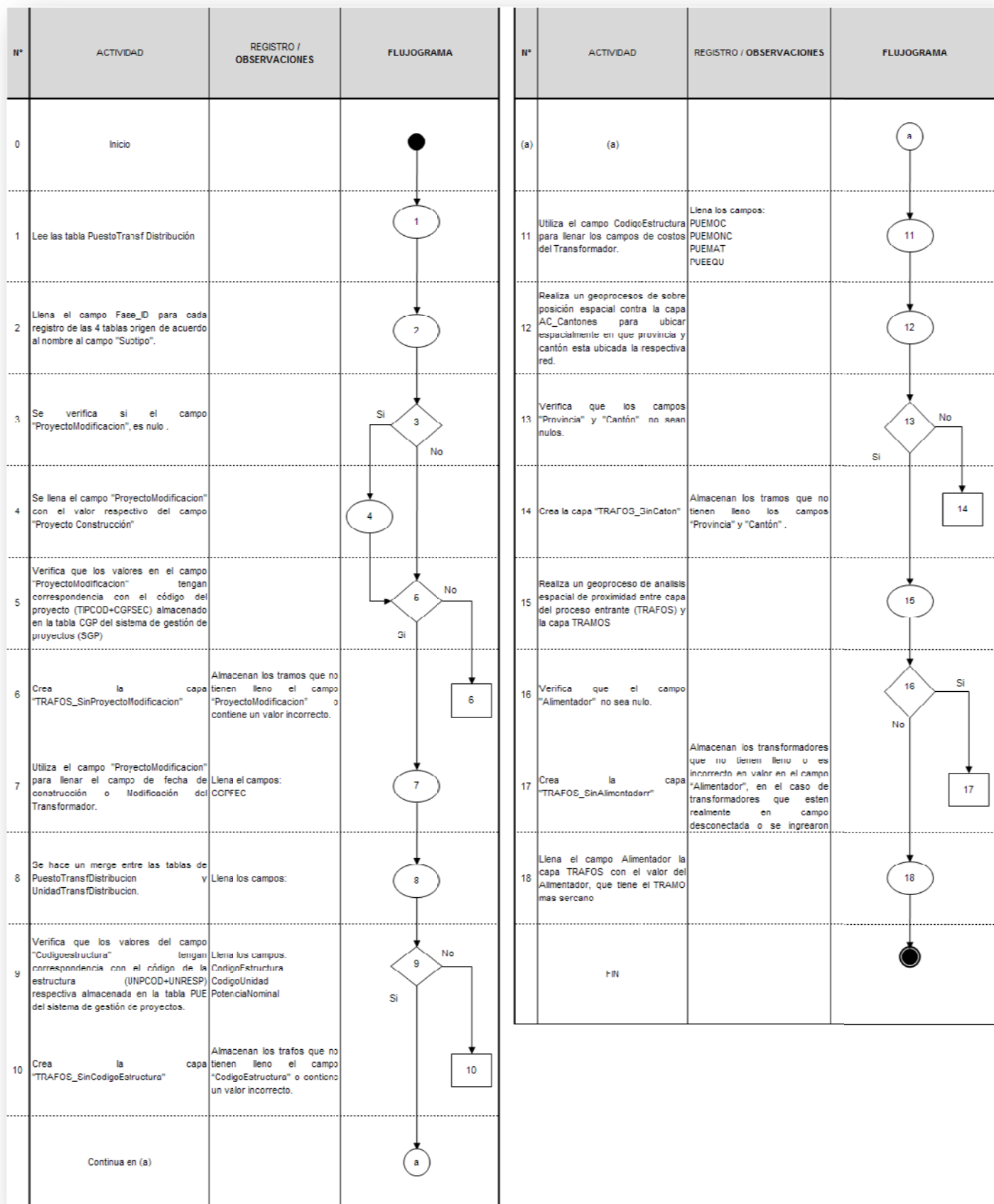


Gráfico 23. Diagrama de actividades del proceso ETL para los transformadores del ODS.

5.1.2.12 Normas y políticas para el proceso:

b. Requerimientos de Información:

Los requerimientos de información fuente del SIG y SGP, deberían ser cumplidos y corregidos en caso de que no cumplan los criterios de aceptación.

c. Frecuencia:

Este proceso de actualización de la capa “TRAFOS” del ODS, debería ser corrido todos los fines de semana, para mantener la información actualizada con una ventana de tiempo pequeña.

5.1.3 Proceso ETL para los CLIENTES del ODS.

5.1.3.1 Breve descripción del proceso:

Este proceso sirve para la extracción, transformación, depuración y carga de la información de las capas “PuntoCarga”, y las tablas “ConexionConsumidor” y “AtributosConsumidor”, desde SIG, y las tablas CGP y PUE, del sistema de gestión de proyectos, al ODS en la capa CLIENTES.

La capa “PuntoCarga” es un feature class de puntos que está almacenada en un feature dataset cuyo DATUM es WGS84 y que se denomina “SIGELEC.Electrico”

La tabla “ConexionConsumidor” esta almacenada en el SIG y existe una relación 1 a M entre la capa “PuntoCarga” y la tabla “ConexionConsumidor”.

La tabla “AtributosConsumidor” esta almacenada en el SIG y existe una relación 1 a 1 entre la tabla “ConexionConsumidor” y la tabla “AtributosConsumidor”.

La capa “CLIENTES”, pertenece al ODS y mantiene la misma proyección espacial que sus datos fuente.

5.1.3.2 Responsable:

El responsable de este proceso es un rol llamado Superintendente Informático, el cual es el administrador del SIG y solo para el caso del prototipo también será el administrador del Data Mart.

5.1.3.3 Objetivo y Alcance:

El objetivo general de este proceso es crear o actualizar los datos espaciales de la capa “CLIENTES” en el ODS. Los objetivos específicos de este proceso son:

- Crear la capa CLIENTES donde se almacena la información depurada de todos los clientes de la red eléctrica.
- Crear la capa CLIENTES_SinProyectoModificacion, que es un feature class de puntos, donde se almacena la ubicación de los clientes que no tienen lleno el campo “ProyectoModificacion”, para su respectiva corrección por parte de los Auxiliares SIG.
- Crear la capa CLIENTES_SinCodigoEstructura, que es un feature class de puntos, donde se almacenan los clientes que no tienen lleno los campos “CodigoEstructura” para su respectiva corrección por parte de los Auxiliares SIG.
- Crear la capa CLIENTES_SinCanton, que es un feature class de puntos, donde se almacenan los clientes que no tienen lleno los campos “Provincia” o “Canton”, para su respectiva corrección por parte de los Auxiliares SIG.
- Crear la capa CLIENTES_SinAlimentador, que es un feature class de puntos, donde se almacenan los clientes que no tienen lleno el campo Alimentador, para su respectiva corrección por parte de los Auxiliares SIG.

5.1.3.4 Clientes (del proceso) y sus Requisitos:

Los clientes del proceso actual son:

- El proceso ETL para el Data Mart, el requerimiento de este proceso es que la capa CLIENTES tenga llenos todos los campos especificados en la plantilla para la capa “CLIENTES” indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1.
- Los Auxiliares SIG también son clientes de este proceso para hacer las correcciones a la información utilizando el software SIG, el requerimiento de este rol es que las tablas de las capas que se nombran a continuación tengan llenos todos los campos especificados en las plantillas respectivas indicadas en la sección “Documentos y/o plantillas relacionadas” del Anexo 1 para las capas:
 - CLIENTES_SinAlimentador
 - CLIENTES_SinProyectoModificacion
 - CLIENTES_SinCodigoEstrucuta
 - CLIENTES_SinCanton.

5.1.3.5 Proveedores (del proceso) y sus Requisitos

Los proveedores de este proceso son dos:

5.1.3.5.1 *El Sistema de Información Geográfico (SIG)*

Los requisitos de este proveedor están relacionados con las capas que se utilizan de este sistema para la depuración de la información de los clientes y su carga en la capa “CLIENTES” del ODS. Estas capas son tres y deben estar en la misma proyección espacial, existen además dos tablas donde se relaciona y almacena la información de los clientes. A continuación se describen éstas:

- a. La primera capa que a su vez contiene su tabla respectiva es “PuntoCarga”. La misma que está almacenadas en el SIG, contienen la información de la localización de los clientes de la EERCS y los valores en sus campos deben estar acorde con las plantillas respectivas indicadas en la sección “Documentos y/o plantillas relacionadas” del Anexo 1, además de cumplir los siguientes requisitos:
- El campo SUBTIPO, no puede ser nulo ya que este sirve para clasificar los clientes en de:
 - Baja Tensión, o Media Tensión.
 - El campo Alimentador debe contener el código correspondiente al respectivo alimentador, no debería ser nulo.
 - El campo ProyectoModificacion debe contener el código correspondiente al último proyecto, en la tabla CGP del Sistema de Gestión de Proyectos, que modificó un cliente en campo y no puede ser nulo.
- b. Tabla “AtributosConsumidor”, la misma que está almacenada en el SIG, contiene la información del consumo mensual y código del cliente localizadas en el punto de carga y los valores en sus campos deben estar acorde con las plantillas respectivas indicadas en la sección “Documentos y/o plantillas relacionadas” del Anexo 1, además de cumplir los siguientes requisitos:
- El campo CodigoCliente debe contener el código comercial correspondiente del cliente y no puede ser nulo.
 - El campo CliUltConM contiene el valor del último consumo mensual de un cliente.
- c. Tabla “ConexionConsumidor”, la misma que está almacenada en el SIG, contiene la información necesaria para relacionar la información comercial del cliente que se

encuentra en el tabla “AtributosConsumidor” con la información de localización geográfica que se encuentra en la tabla “PuntoCarga” y los valores en sus campos deben estar acorde con las plantillas respectivas indicadas en la sección “Documentos y/o plantillas relacionadas” del Anexo 1, además de cumplir los siguientes requisitos:

- El campo PuntoCargaObjectID debe contener el id del punto de carga y no puede ser nulo.
 - El campo CodigoCliente contiene el código del cliente y no puede ser nulo.
- d. Otra capa necesaria para este proceso en la capa “TRAMOS” la cual es una capa de líneas que contiene la información del Alimentador y será utilizada en los geoprocetos para la asignación de la información del Alimentador en la capa TRAFOS resultante. Los valores en los campos de esta tabla deben estar acorde con la plantilla para la capa “TRAMOS” indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1.
- e. La última capa necesaria para este proceso es la capa “AC_CANTONES” la cual es una capa de polígonos que contiene la división política de los cantones del país y será utilizada en los geoprocetos para la asignación de la información de la provincia y cantón en la capa CLIENTES resultante. Los valores en los campos de esta tabla deben estar acorde con la plantilla para la capa “AC_CANTONES” indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1.

5.1.3.5.2 *El Sistema de Gestión de Proyectos (SGP)*

Los requisitos para este proveedor están relacionados con las tablas que se utilizan de este sistema para la depuración y carga en la capa “TRAFOS” del ODS de la

información de las redes. Estas tablas están almacenadas en el servidor en producción y son las siguientes:

- a. La tabla CGP, que contiene la información de los proyectos de electrificación y donde los valores de sus campos no pueden ser nulos y deben estar acuerdo al formato especificado para la tabla CGP indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1, además de cumplir los siguientes requisitos:
 - TipCod: que corresponde al tipo de proyecto, debe estar lleno con el respectivo tipo de proyecto, que puede ser Urbano o Rural.
 - CGPSeq: que es un número secuencia, que junto con el campo TipCod conforman el código del proyecto, debe estar lleno con el correspondiente valor del proyecto.
 - CGPDes: que corresponde con el nombre del proyecto, esté lleno con el respectivo nombre del proyecto.
 - CGPFEC: que corresponde a la fecha de construcción del proyecto debe estar lleno con la respectiva fecha de construcción del proyecto.

- b. La tabla PUE, que contiene la información de las unidades constructivas y sus costos y donde los valores de sus campos no pueden ser nulos y deben estar acuerdo al formato especificado para la tabla PUE, indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1, además de cumplir los siguientes requisitos:
 - UNPCOD que contiene el código de las unidades de propiedad.

- UNRESP que es un secuencial que junto con UNPCOD conforman el código de estructura.
- PUEMOC el cual contiene el costo de mano de obra calificada para su respectiva estructura.
- PUEMONC el cual contiene el costo de mano de obra no calificada para su respectiva estructura.
- PUEMAT el cual contiene el costo del material para su respectiva estructura.
- PUEEQU el cual contiene el costo del equipo utilizado para su respectiva estructura.
- PUEADMT el cual contiene el costo de administración para su respectiva estructura.

5.1.3.6 Descripción detallada del proceso:

El proceso de ETL para los CLIENTES del ODS se lo detalla a continuación como actividades:

1. Leer la tabla de la capa PuntoCarga desde el Sistema de Información Geográfico.
2. Llena el campo BajaMedia_ID para cada registro de las tablas fuente de acuerdo al nombre al campo "Subtipo", cuyos valores se especifican en la Tabla siguiente

| Subtipo | Campo Subtipo | Fase_ID |
|---------|------------------------------|---------|
| 1 | Punto de Carga Baja Tensión | 1 |
| 2 | Punto de Carga Media Tensión | 2 |

Tabla 6. Valor asignado a lo campo BajaMedia_ID de acuerdo al campo "Subtipo" de la tabla fuente.

3. Se verifica si el campo "ProyectoModificacion", es nulo.
4. El campo "ProyectoModificacion" almacena el código de la carpeta del último proyecto con el que se realizaron modificaciones en un cliente, si este campo está vacío, quiere decir que no se han hecho mejoras o modificaciones y se debe almacenar el código que está en el campo "Proyecto Construcción", que es el de la construcción original del elemento, por lo tanto si el campo "ProyectoModificacion" es nulo, se llena el campo "ProyectoModificacion" con el valor respectivo del campo "Proyecto Construcción".
5. Mediante un Join entre tablas verifica que los valores en el campo "ProyectoModificacion" de la tabla CLIENTES tengan correspondencia con los campos TIPCOD+CGPSEC de la tabla CGP del Sistema de Gestión de Proyectos (SGP).
6. Los registros cuyo campo ProyectoModificacion es nulo o no tienen correspondencia en el Join del paso anterior, son almacenados en una capa llamada "CLIENTES_SinProyectoModificacion", la cual es una capa espacial, que sirve para registrar dónde está ubicado el elemento que no tiene un valor adecuado en sus campos "ProyectoModificacion" para que se puedan tomar los correctivos necesarios.
7. En los registros que sí tienen correspondencia con el Join del paso 5 se actualiza el campo "FECHA_PROYECTO" con el valor correspondiente del campo CGP.CGPFEC, que es la fecha de energización del medidor del cliente.
8. Se hace un Join entre la tabla de la capa PuntoCarga con el campo ObjectID y las tablas ConexionConsumidor con el campo PuntoCargaObjectID y CodigoCliente con la tabla AtributosConsumidor con el campo también llamado CodigoCliente.

9. Con las tablas relacionadas se procede a actualizar los campos `CodigoCliente` y `CliUltConM` correspondientes en la tabla `CLIENTES`.
10. Mediante un `Join` entre tablas se verifica los valores del campo "`CodigoEstructura`" de la tabla resultante del proceso anterior con el código de la estructura (`UNPCOD+UNRESP`) respectiva almacenada en la tabla `PUE` del sistema de gestión de proyectos.
11. Los registros cuyo campo `CodigoEstructura` es nulo o no tienen correspondencia en el `Join` del paso anterior, son almacenados en una capa llamada "`CLIENTES_SinCodigoEstructura`", la cual es una capa espacial, que sirve para registrar dónde está ubicado el elemento que no tiene un valor adecuado en sus campos "`ProyectoModificacion`" para que se puedan tomar los correctivos necesarios.
12. En los registros que sí tienen correspondencia con el `Join` del paso 10 se actualiza los campos: `PUEMOC`, `PUEMONC`, `PUEMAT`, `PUEEQU` que son los costos de puesta en operación de los medidores.
13. Se realiza un geoproceso de sobre posición espacial contra la capa `AC_Cantones` para ubicar espacialmente en qué provincia y cantón está ubicado el respectivo cliente. Este geoproceso consiste en hacer un `Join` espacial de la capa `PuntoCarga` contra la capa `AC_Cantones`, lo que da como resultado que a la capa `PuntoCarga` se le añadan los campos de la capa `AC_Cantones` que coinciden espacialmente (los clientes que están dentro del respectivo polígono que representa los límites de un cantón). Se procede a actualizar el campo `Canton` de la tabla `CLIENTES` y al final del proceso se toma los dos primeros dígitos de la capa `AC_Cantones`, para además llenar el campo `Provincia` y de

esta manera se evita tener que realizar otro geoproceso espacial para llenar el campo de la provincia en la capa Clientes.

14. Verifica que los campos "Provincia" y "Canton" no sean nulos. Esto podría darse por ejemplo en el caso de que por error se grafique un cliente fuera de los límites del Ecuador para donde no tenemos información en nuestra capa "AC_Cantones".
15. Si los campos "Provincia" y "Canton" son nulos en algunos registro, estos son filtrados y almacenados en una capa llamada "CLIENTES_SinCanton", que sirve para ubicar dónde están estos clientes que no caen dentro de la capa de cantones y provincias.
16. Realiza un geoproceso de proximidad espacial, entre las capa TRAMOS y CLIENTES, para actualizar el valor del campo "Alimentador" basado en el valor del campo alimentador del tramo más cercano.
17. Luego se verifica el campo "Alimentador" de la capa resultante del geoproceso anterior, para saber cuáles son nulos.
18. Se almacena en la capa "CLIENTES_SinAlimentador" los registros donde el valor en el campo "Alimentador" es nulo o es incorrecto, en el caso de clientes que estén realmente desconectados en campo o se ingresaron por error del Auxiliar SIG.
19. En los registros que sí tienen correspondencia con el Join espacial del paso 16 se actualiza los campos: "Alimentador".
20. Como resultado de cada una de estas actividades en este proceso se obtiene una capa llamada CLIENTES, la cual almacena como puntos, toda la información depurada de los clientes tomados del SIG, en la cual se debe garantizar que se cumplan los criterios de aceptación mencionados en la siguiente sección.

Esta capa queda almacenada en el ODS, la cual a su vez sirve de suministro y fuente depurada de datos para la el proceso ETL al Data Mart.

5.1.3.7 Criterios de Aceptación:

Se debe cumplir los siguientes criterios para ser considerada una carga exitosa de la capa CLIENTES en el ODS:

- 1 El campo Alimentador debe contener el código correspondiente al respectivo alimentador y no puede ser nulo.
- 2 El campo Subestación debe contener el número de subestación correspondiente a la respectivo subestación y no puede ser nulo.
- 3 El campo ProyectoModificacion debe contener el código correspondiente al último código de proyecto, en la tabla CGP del Sistema de Gestión de Proyectos, que modificó al elemento y no puede ser nulo.
- 4 El campo CodigoEstructura debe contener el código correspondiente al catálogo de estructuras, en la tabla PUE del Sistema de Gestión de Proyectos, para el medidor del cliente respectivo y no puede ser nulo.
- 5 El campo Provincia debe contener el código correspondiente a la provincia en la que está ubicado espacialmente el cliente y no puede ser nulo.
- 6 El campo Cantón debe contener el código correspondiente al cantón en la que está ubicado espacialmente y no puede ser nulo.
- 7 El campo BajaMedia_Id solo debe contener valores entre 1 (baja tensión) y 2 (media tensión) y no puede ser nulo.

5.1.3.8 Indicador de Excepciones

Una parte importante del proceso es determinar las excepciones del mismo, es decir los datos que no fueron transferidos por qué no cumplieron los criterios de aceptación (ver arriba), como resultado de este proceso obtenemos las capas

- CLIENTES_SinAlimentador
- CLIENTES_SinProyectoModificacion
- CLIENTES_SinCodigoEstructura
- CLIENTES_SinCanton

Donde no solo podemos determinar cuáles son los clientes que no cumplieron los criterios de aceptación, sino también dónde están ubicados espacialmente, cuándo fueron graficados o modificados y quien los graficó, para poder identificar tendencias que se pueden mostrar en el transcurso del tiempo, y si se analizan de manera integral permiten mantener un diagnóstico de la situación y tomar acciones preventivas o correctivas sobre los datos transacciones y verificar si éstas fueron o no acertadas.

5.1.3.9 Glosario o Definición de Términos

SGP.- Sistema de Gestión de Proyectos.

SIG.- Sistema de Información Geográfico.

Alimentador.- El alimentar pertenece a una subestación y es la fuente de la energía a ser distribuida.

Proyecto Modificación.- Es el código de la carpeta a la que pertenecen los elementos de un proyecto de electrificación.

Código Estructura.- Es el código de la estructura asignada como una unidad de Transformación.

ODS.- Operational Data Store, Base de Datos Operacional, es una base de datos diseñada para integrar datos de múltiples fuentes para hacer más fácil el análisis y la presentación de informes. Debido a que los datos origen vienen de diferentes fuentes, la integración implica a menudo la limpieza, la solución de redundancia y comprobación de reglas de negocio para integridad. El ODS es generalmente diseñado para contener datos de bajo nivel o indivisible (como las transacciones y los precios) con una ventana de actualización muy pequeña en contraposición con el mayor volumen de datos almacenados en el Data Warehouse que generalmente es con menor frecuencia. En otras palabras, es un contenedor de datos activos, es decir operacionales que ayudan al soporte de decisiones y a la operación y está entre un OLAP y un OLTP.

Auxiliar SIG.- Es un rol dentro del departamento SIGADE encargado de la actualización de la información espacial en el SIG.

5.1.3.10 Documentos y/o plantillas relacionadas:

En el anexo 1 se presentan las estructuras de datos (plantilla) para cada una de las tablas fuente y resultantes del presente proceso. En cada plantilla se puede obtener datos como: Nombre del campo, tipo, tamaño y una pequeña descripción.

5.1.3.11 Diagramas de actividades (workflows):

A continuación se adjunta el diagrama de actividades del proceso ETL para los clientes del ODS:

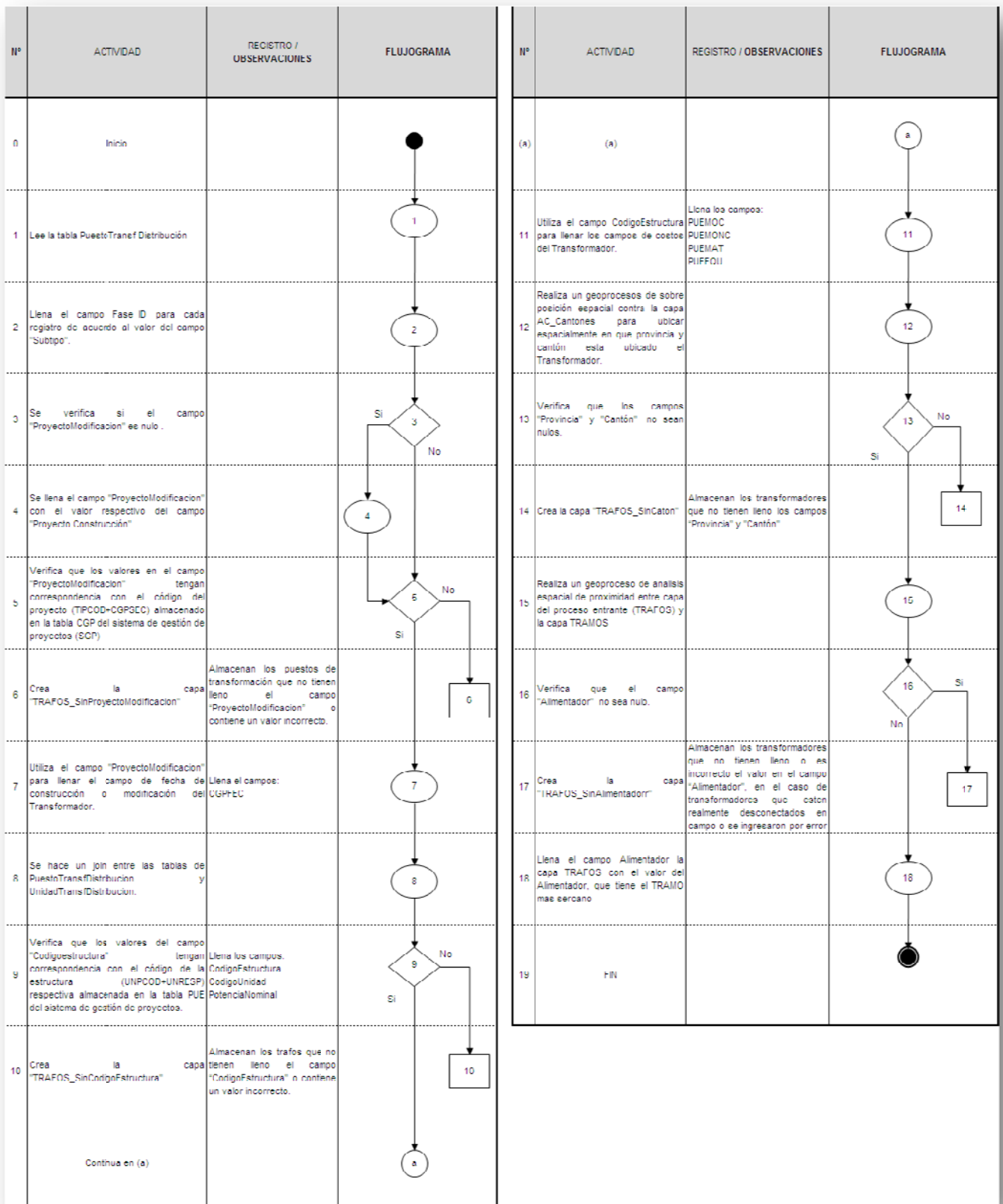


Gráfico 24. Diagrama de actividades del proceso ETL para los clientes del ODS.

Normas y políticas para el proceso:

a. Requerimientos de Información:

Los requerimientos de información fuente del SIG y SGP, deberían ser cumplidos y corregidos en caso de que no cumplan los criterios de aceptación.

b. Frecuencia:

Este proceso de actualización de la capa “CLIENTES” del ODS, debería ser corrido todos los fines de semana, para mantener la información actualizada con una ventana de tiempo pequeña.

5.1.4 Proceso ETL para las LUMINARIAS del ODS.

5.1.4.1 Breve descripción del proceso:

Este proceso sirve para la extracción, transformación, depuración y carga de la información de las capas “Luminarias” y la tabla “CatalogoEstructura”, desde SIG, y las tablas CGP y PUE, del sistema de gestión de proyectos, al ODS en la capa “LUMINARIAS_ODS”.

La capa “Luminarias” es un feature class de puntos que está almacenada en un feature dataset cuyo DATUM es WGS84 y que se denomina “SIGELEC.Electrico”

La tabla “CatalogoEstructura” esta almacenada en el SIG y mantiene una relación 1 a M con la capa “Luminarias”.

La capa “LUMINARIAS_ODS”, pertenece al ODS y mantiene la misma proyección espacial que sus datos fuente.

5.1.4.2 Responsable:

El responsable de este proceso es un rol llamado Superintendente Informático, el cual es el administrador del SIG y solo para el caso del prototipo también será el administrador del Data Mart.

5.1.4.3 Objetivo y Alcance:

El objetivo general de este proceso es crear o actualizar los datos espaciales de la capa “LUMINARIAS_ODS” en el ODS. Los objetivos específicos de este proceso son:

- Crear la capa LUMINARIAS_ODS donde se almacena la información depurada de todos las Luminarias de la red eléctrica.
- Crear la capa LUMINARIAS_SinProyectoModificacion, que es un feature class de punto, donde se almacena la ubicación de las luminarias que no tienen lleno el campo “ProyectoModificacion”, para su respectiva corrección por parte de los Auxiliares SIG.
- Crear la capa LUMINARIAS_SinCodigoEstructura, que es un feature class de punto, donde se almacenan las luminarias que no tienen lleno los campos “CodigoEstructura” para su respectiva corrección por parte de los Auxiliares SIG.
- Crear la capa LUMINARIAS_SinCanton, que es un feature class de punto, donde se almacenan las luminarias que no tienen lleno los campos “Provincia” o “Canton”, para su respectiva corrección por parte de los Auxiliares SIG.
- Crear la capa LUMINARIAS_SinAlimentador, que es un feature class de punto, donde se almacenan las luminarias que no tienen lleno el campo Alimentador, para su respectiva corrección por parte de los Auxiliares SIG.

5.1.4.4 Clientes (del proceso) y sus Requisitos:

Los clientes del proceso actual son:

- El proceso ETL para el Data Mart, el requerimiento de este proceso es que la capa LUMINARIAS_ODS tenga llenos todos los campos especificados en la plantilla para la capa “LUMINARIAS_ODS” indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1.
- Los Auxiliares SIG también son clientes de este proceso para hacer las correcciones a la información utilizando el software SIG, el requerimiento de

este rol es que las tablas de las capas que se nombran a continuación tengan llenos todos los campos especificados en las plantillas respectivas indicadas en la sección “Documentos y/o plantillas relacionadas” del Anexo 1 para las capas:

- LUMINARIAS_SinAlimentador
- LUMINARIAS_SinProyectoModificacion
- LUMINARIAS_SinCodigoEstructura
- LUMINARIAS_SinCanton.

5.1.4.5 Proveedores (del proceso) y sus Requisitos

Los proveedores de este proceso son dos:

5.1.4.5.1 *El Sistema de Información Geográfico (SIG)*

Los requisitos de este proveedor están relacionados con las capas que se utilizan de este sistema para la depuración de la información de las luminarias y su carga en la capa “LUMINARIAS_ODS”. Estas capas son tres y deben estar en la misma proyección espacial, existen además una tabla donde se almacena la información de técnica de las luminarias. A continuación se describen éstas:

- a. La primera capa que a su vez contiene su tabla respectiva es “Luminarias”. La misma que está almacenadas en el SIG, contienen la información de la localización de las Luminarias de la EERCS y los valores en sus campos deben estar acorde con las plantillas respectivas indicadas en la sección “Documentos y/o plantillas relacionadas” del Anexo 1, además de cumplir los siguientes requisitos:
 - El campo “Alimentador” debe contener el código correspondiente al respectivo alimentador, no debería ser nulo.

- El campo “ProyectoModificacion” debe contener el código correspondiente al último proyecto, en la tabla CGP del Sistema de Gestión de Proyectos, que modificó un cliente en campo y no puede ser nulo.
 - El campo “CodigoEstructura” debe contener el código correspondiente al catálogo de estructuras, en la tabla PUE del Sistema de Gestión de Proyectos, para cada Luminaria y no puede ser nulo.
 - Los campos “HoraFunc1” y “HoraFunc2” no deben ser nulos, ya que contienen el número de horas que permanece prendida a la noche cada luminaria, ya sea esta convencional o de doble nivel y son utilizados para calcular la energía que ocupa cada una de las luminarias al mes.
 - El campo “DiasFuncMes” no debe ser nulo, ya que indica el número de días al mes que permanece prendida una luminaria y es utilizado para el cálculo de la energía ocupada por cada una.
- b. La tabla “CatalogoEstructura” la misma que está almacenada en el SIG, contiene la información técnica de las Luminarias de la EERCS y los valores en sus campos deben estar acorde con las plantillas respectivas indicadas en la sección “Documentos y/o plantillas relacionadas” del Anexo 1, además de cumplir los siguientes requisitos:
- El campo “CodigoEstructura” debe contener el identificado de la estructura, por lo tanto no puede ser nulo.
 - El campo “Tipo” contiene una clasificación general de las luminarias y es utilizado en la carga de las dimensiones en el Data Mart, por lo tanto no puede ser nulo.
 - El campo “DobleNivel” este campo indica si una luminaria es de doble nivel, para de acuerdo a esto hacer el cálculo adecuado de la energía, por lo tanto este campo no puede ser nulo.

- Los campo “Potencia” y “Potencia2” contienen la información de la Potencia de la laminaras ya sea que estas son convencionales o doble nivel, el campo Potencia, no debería ser nulo.
 - El campo “Perdida_Porcentaje” contiene un valor de porcentaje de perdida producida en el balastro y que depende del tipo de luminaria, este campo no debería se nulo.
- c. Otra capa necesaria para este proceso en la capa “TRAMOS” la cual es una capa de líneas que contiene la información del Alimentador y será utilizada en los geoprosos para la asignación de la información del Alimentador en la capa LUMINARIAS_ODS resultante. Los valores en los campos de esta tabla deben estar acorde con la plantilla para la capa “TRAMOS” indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1.
- d. La última capa necesaria para este proceso es la capa “AC_CANTONES” la cual es una capa de polígonos que contiene la división política de los cantones del país y será utilizada en los geoprosos para la asignación de la información de la provincia y cantón en la capa LUMINARIAS_ODS resultante. Los valores en los campos de esta tabla deben estar acorde con la plantilla para la capa “AC_CANTONES” indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1.

5.1.4.5.2 *El Sistema de Gestión de Proyectos (SGP)*

Los requisitos para este proveedor están relacionados con las tablas que se utilizan de este sistema para la depuración y carga en la capa “LUMINARIAS” del ODS de la

información de las redes. Estas tablas están almacenadas en el servidor AS400 en producción y son las siguientes:

- a. La tabla CGP, que contiene la información de los proyectos de electrificación y donde los valores de sus campos no pueden ser nulos y deben estar acuerdo al formato especificado para la tabla CGP indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1, además de cumplir los siguientes requisitos:
 - TipCod: que corresponde al tipo de proyecto, debe estar lleno con el respectivo tipo de proyecto.
 - CGPSeq: que es un número secuencia, que junto con el campo TipCod conforman el código del proyecto, debe estar lleno con el correspondiente valor del proyecto.
 - CGPDes: que corresponde con el nombre del proyecto, esté lleno con el respectivo nombre del proyecto.
 - CGPFEC: que corresponde a la fecha de construcción del proyecto debe estar lleno con la respectiva fecha de construcción del proyecto.

- b. La tabla PUE, que contiene la información de las unidades constructivas y sus costos y donde los valores de sus campos no pueden ser nulos y deben estar acuerdo al formato especificado para la tabla PUE, indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1, además de cumplir los siguientes requisitos:
 - UNPCOD que contiene el código de las unidades de propiedad.

- UNRESP que es un secuencial que junto con UNPCOD conforman el código de estructura.
- PUEMOC el cual contiene el costo de mano de obra calificada para su respectiva estructura.
- PUEMONC el cual contiene el costo de mano de obra no calificada para su respectiva estructura.
- PUEMAT el cual contiene el costo del material para su respectiva estructura.
- PUEEQU el cual contiene el costo del equipo utilizado para su respectiva estructura.
- PUEADMT el cual contiene el costo de administración para su respectiva estructura.

5.1.4.6 Descripción detallada del proceso:

El proceso de ETL para las LUMINARIAS_ODS se lo detalla a continuación como actividades:

1. Leer la tabla de la capa Luminarias desde el Sistema de Información Geográfico.
2. Se verifica si el campo "ProyectoModificacion", es nulo.
3. El campo "ProyectoModificacion" almacena el código de la carpeta del último proyecto con el que se realizaron modificaciones en una luminaria, si este campo está vacío, quiere decir que no se han hecho mejoras o modificaciones y se debe almacenar el código que está en el campo "ProyectoConstrucción", que es el de la construcción original del elemento, por lo tanto si el campo "ProyectoModificacion" es nulo, se llena el campo "ProyectoModificacion" con el valor respectivo del campo "Proyecto Construcción".

4. Mediante un Join entre tablas verifica que los valores en el campo "ProyectoModificacion" de la tabla LUMINARAS tengan correspondencia con los campos TIPCOD+CGPSEC de la tabla CGP del Sistema de Gestión de Proyectos (SGP).
5. Los registros cuyo campo ProyectoModificacion es nulo o no tienen correspondencia en el Join del paso anterior, son almacenados en una capa llamada "LUMINARIAS_SinProyectoModificacion", la cual es una capa espacial, que sirve para registrar dónde está ubicado el elemento que no tiene un valor adecuado en sus campos "ProyectoModificacion" para que se puedan tomar los correctivos necesarios.
6. En los registros que sí tienen correspondencia con el Join del paso 4 se actualiza el campo "FECHA_PROYECTO" con el valor correspondiente del campo CGP.CGPFEC, que es la fecha de puesta en operación de la Luminaria.
7. Se hace un Join entre la tabla de la capa Luminarias con el campoCodigoEstructura y las tablas PUE con los campos UNPCOD+UNRESP.
8. Verifica los registros de la tabla LUMINARIAS_ODS que tienen correspondencia con los de la tabla PUE del sistema de gestión de proyectos.
9. Los registros cuyo campo CodigoEstructura es nulo o no tienen correspondencia en el Join del paso anterior, son almacenados en una capa llamada "LUMINARIAS_SinCodigoEstructura", la cual es una capa espacial, que sirve para registrar dónde está ubicado el elemento que no tiene un valor adecuado en sus campos "ProyectoModificacion" para que se puedan tomar los correctivos necesarios.

10. En los registros que sí tienen correspondencia con el Join del paso 7 se actualiza los campos: PUEMOC, PUEMONC, PUEMAT, PUEEQU que son los costos de puesta en operación de los medidores.
11. Se hace un Join entre la tabla de la capa Luminarias con el campo CodigoEstructura y las tablas CatalogoEstructura con el campo CodigoEstructura mismo y llena los campos TIPO, POTENCIA, POTENCIA2, DOBLENIVEL, PERDIDA_PORCENTAJE de la tabla LUMINARIAS_ODS.
12. Se realiza un geoproceso de sobre posición espacial contra la capa AC_Cantones para ubicar espacialmente en qué provincia y cantón está ubicada la respectiva luminaria. Este geoproceso consiste en hacer un Join espacial de la capa Luminarias contra la capa AC_Cantones, lo que da como resultado que a la capa Luminarias se le añadan los campos de la capa AC_Cantones que coinciden espacialmente (los clientes que están dentro del respectivo polígono que representa los límites de un cantón). Se procede a actualizar el campo Canton de la tabla LUMINARIAS_ODS y al final del proceso se toma los dos primeros dígitos de la capa AC_Cantones, para además llenar el campo Provincia y de esta manera se evita tener que realizar otro geoproceso espacial para llenar el campo de la provincia en la capa LUMINARIAS_ODS.
13. Verifica que los campos "Provincia" y "Canton" no sean nulos. Esto podría darse por ejemplo en el caso de que por error se grafique una luminaria fuera de los límites del Ecuador para donde no tenemos información en nuestra capa "AC_Cantones".
14. Si los campos "Provincia" y "Canton" son nulos en algunos registros, estos son filtrados y almacenados en una capa llamada "LUMINARIAS_SinCanton", que

sirve para ubicar donde están estas luminarias que no caen dentro de la capa de cantones y provincias.

15. Realiza un geoproceto de proximidad espacial, entre las capa TRAMOS y LUMINARIAS, para actualizar el valor del campo "Alimentador" basado en el valor del campo alimentador del tramo más cercano.
16. Luego se verifica el campo "Alimentador" de la capa resultante del geoproceto anterior, para saber cuáles son nulos.
17. Se almacena en la capa "LUMINARIAS_SinAlimentador" los registros donde el valor en el campo "Alimentador" es nulo o es incorrecto, en el caso de que existan Luminarias que estén realmente desconectados en campo o se ingresaron por error del Auxiliar SIG.
18. En los registros que sí tienen correspondencia con el Join espacial del paso 15 se actualiza los campos: "Alimentador".
19. Como resultado de cada una de estas actividades en este proceso se obtiene una capa llamada LUMINARIAS_ODS, la cual almacena como puntos, toda la información depurada de las luminarias tomada del SIG, en la cual se debe garantizar que se cumplan los criterios de aceptación mencionados en la siguiente sección.

Esta capa queda almacenada en el ODS, la cual a su vez sirve de suministro y fuente depurada de datos para la el proceso ETL al Data Mart.

5.1.4.7 Criterios de Aceptación:

Se debe cumplir los siguientes criterios para ser considerada una carga exitosa de la capa LUMINARIAS_ODS:

- 8 El campo Alimentador debe contener el código correspondiente al respectivo alimentador y no puede ser nulo.
- 9 El campo Subestación debe contener el número de subestación correspondiente a la respectiva subestación y no puede ser nulo.
- 10 El campo ProyectoModificacion debe contener el código correspondiente al último código de proyecto, en la tabla CGP del Sistema de Gestión de Proyectos, que modificó al elemento y no puede ser nulo.
- 11 El campoCodigoEstructura debe contener el código correspondiente al catálogo de estructuras, en la tabla PUE del Sistema de Gestión de Proyectos, para el medidor del cliente respectivo y no puede ser nulo.
- 12 El campo Provincia debe contener el código correspondiente a la provincia en la que está ubicado espacialmente el cliente y no puede ser nulo.
- 13 El campo Cantón debe contener el código correspondiente al cantón en la que está ubicado espacialmente y no puede ser nulo.
- 14 El campo Tipo no puede ser nulo.
- 15 El campo Potencia no puede ser nulo.

5.1.4.8 Indicador de Excepciones

Una parte importante del proceso es determinar las excepciones del mismo, es decir los datos que no fueron transferidos por qué no cumplieron los criterios de aceptación (ver arriba), como resultado de este proceso obtenemos las capas

- LUMINARIAS_SinAlimentador
- LUMINARIAS_SinProyectoModificacion
- LUMINARIAS_SinCodigoEstructura
- LUMINARIAS_SinCanton

Donde no solo podemos determinar cuáles son las luminarias que no cumplieron los criterios de aceptación, sino también dónde están ubicados espacialmente, cuándo fueron graficados o modificados y quien los graficó, para poder identificar tendencias que se pueden mostrar en el transcurso del tiempo, y si se analizan de manera integral permiten mantener un diagnóstico de la situación y tomar acciones preventivas o correctivas sobre los datos transacciones y verificar si éstas fueron o no acertadas.

5.1.4.9 Glosario o Definición de Términos

SGP.- Sistema de Gestión de Proyectos.

SIG.- Sistema de Información Geográfico.

Alimentador.- El alimentar pertenece a una subestación y es la fuente de la energía a ser distribuida.

Proyecto Modificación.- Es el código de la carpeta a la que pertenecen los elementos de un proyecto de electrificación.

Código Estructura.- Es el código de la estructura asignada como una unidad de Transformación.

Balastro.- El balastro, también conocido como balasto (no confundir con piedra para vía férrea) o balastra en España, es un equipo que sirve para mantener un flujo de corriente estable en lámparas, ya sea un tubo fluorescente, lámpara de vapor de sodio o lámpara de vapor de mercurio.

Doble Nivel.- Son luminarias que funcionan con dos niveles de potencia que pueden ser programados.

ODS.- Operational Data Store, Base de Datos Operacional, es una base de datos diseñada para integrar datos de múltiples fuentes para hacer más fácil el

análisis y la presentación de informes. Debido a que los datos origen vienen de diferentes fuentes, la integración implica a menudo la limpieza, la solución de redundancia y comprobación de reglas de negocio para integridad. El ODS es generalmente diseñado para contener datos de bajo nivel o indivisible (como las transacciones y los precios) con una ventana de actualización muy pequeña en contraposición con el mayor volumen de datos almacenados en el Data Warehouse que generalmente es con menor frecuencia. En otras palabras, es un contenedor de datos activos, es decir operacionales que ayudan al soporte de decisiones y a la operación y está entre un OLAP y un OLTP.

Auxiliar SIG.-Es un rol dentro del departamento SIGADE encargado de la actualización de la información espacial en el SIG.

5.1.4.10 Documentos y/o plantillas relacionadas:

En el anexo 1 se presentan las estructuras de datos (plantilla) para cada una de las tablas fuente y resultantes del presente proceso. En cada plantilla se puede obtener datos como: Nombre del campo, tipo, tamaño y una pequeña descripción.

5.1.4.11 Diagramas de actividades (workflows):

A continuación se adjunta el diagrama de actividades del proceso ETL para las Luminarias del ODS:

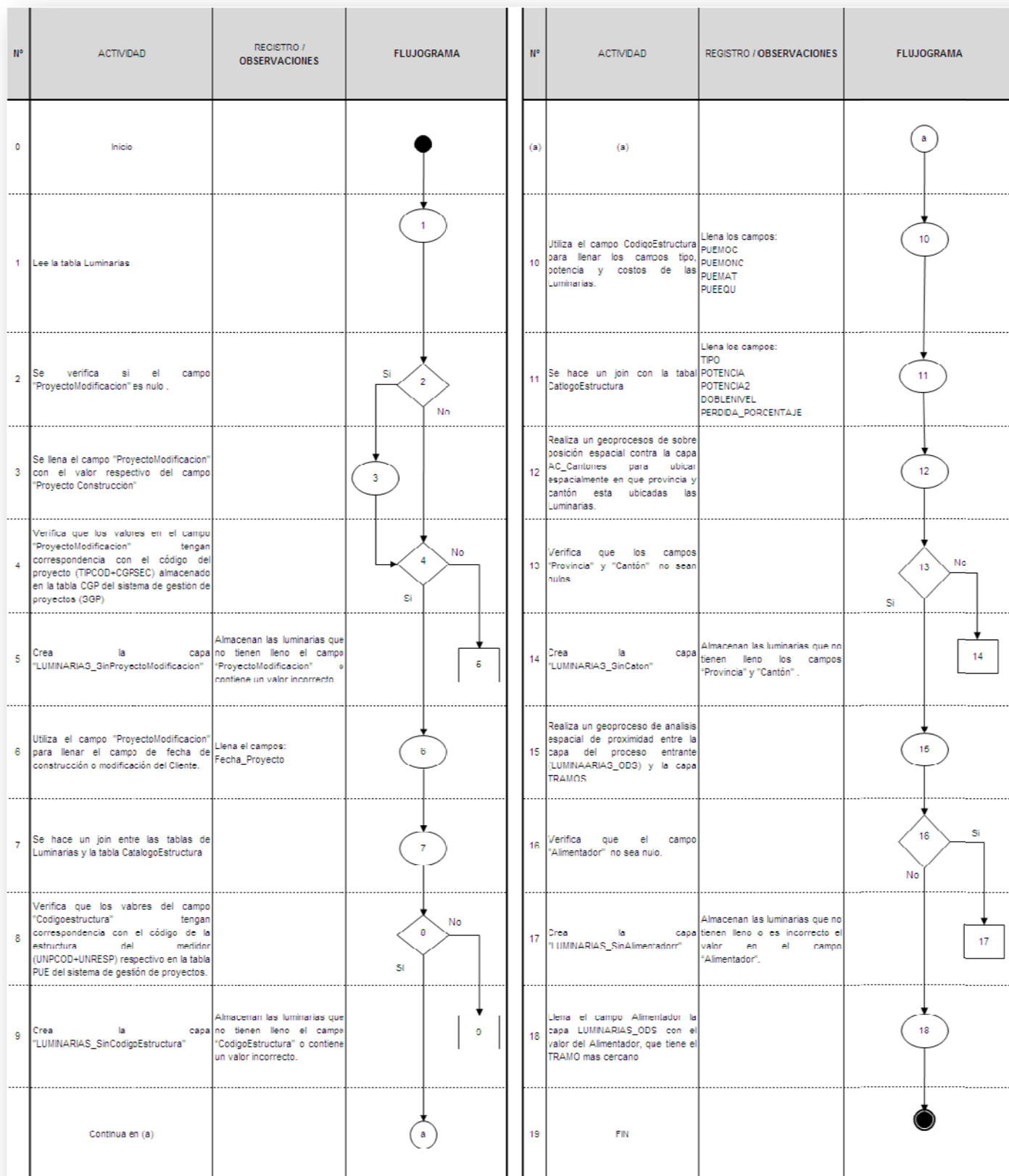


Gráfico 25. Diagrama de actividades del proceso ETL para las luminarias del ODS.

Normas y políticas para el proceso:

c. Requerimientos de Información:

Los requerimientos de información fuente del SIG y SGP, deberían ser cumplidos y corregidos en caso de que no cumplan los criterios de aceptación.

d. Frecuencia:

Este proceso de actualización de la capa “LUMINARIAS_ODS” del ODS, debería ser corrido todos los fines de semana, para mantener la información actualizada con una ventana de tiempo pequeña.

5.1.5 Proceso ETL para los POSTES y ESTRUCTURAS EN POSTE del ODS.

5.1.5.1 Breve descripción del proceso:

Este proceso sirve para la extracción, transformación, depuración y carga de la información de las capas “EstructuraSoporte” y la tabla “EstructuraEnPoste”, desde SIG, y las tablas CGP y PUE, del sistema de gestión de proyectos, al ODS en la capa POSTES y ESTRUCEnPOSTE.

La capa “EstructuraSoporte” es un feature class de puntos que está almacenada en un feature dataset cuyo DATUM es WGS84 y que se denomina “SIGELEC.Electrico”

La tabla “EstructuraEnPoste” esta almacenada en el SIG y existe una relación 1 a M con la capa “EstructuraSoporte” y la tabla “EstructuraEnPoste”.

La tabla “ESTRUCEnPOSTE” y capa “POSTES”, pertenece al ODS y mantiene la misma proyección espacial que sus datos fuente.

5.1.5.2 Responsable:

El responsable de este proceso es un rol llamado Superintendente Informático, el cual es el administrador del SIG y solo para el caso del prototipo también será el administrador del Data Mart.

5.1.5.3 Objetivo y Alcance:

El objetivo general de este proceso es crear o actualizar los datos espaciales de las capas “POSTES” y “ESTRUCEnPOSTE” en el ODS. Los objetivos específicos de este proceso son:

- Crear la capa POSTES donde se almacena la información depurada de todos los Postes en el sistema eléctrico.
- Crear la tabla ESTRUCEnPOSTE, que es un feature class de puntos, donde se almacena la información depurada de todas las estructuras montadas en los postes en el sistema eléctrico.
- Crear la capa POSTES_SinProyectoModificacion, que es un feature class de puntos, donde se almacena la ubicación de los postes que no tienen lleno el campo “ProyectoModificacion”, para su respectiva corrección por parte de los Auxiliares SIG.
- Crear la capa POSTES_SinCodigoEstructura, que es un feature class de puntos, donde se almacenan los postes que no tienen lleno los campos “CodigoEstructura” para su respectiva corrección por parte de los Auxiliares SIG.
- Crear la capa POSTES_SinCanton, que es un feature class de puntos, donde se almacenan los postes que no tienen lleno los campos “Provincia” o “Canton”, para su respectiva corrección por parte de los Auxiliares SIG.
- Crear la capa POSTES_SinAlimentador, que es un feature class de puntos, donde se almacenan los postes que no tienen lleno el campo Alimentador, para su respectiva corrección por parte de los Auxiliares SIG.

5.1.5.4 Clientes (del proceso) y sus Requisitos:

Los clientes del proceso actual son:

- El proceso ETL para el Data Mart, el requerimiento de este proceso es que la capa POSTES tenga llenos todos los campos especificados en la plantilla para la

capa “POSTES” indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1.

- Los Auxiliares SIG también son clientes de este proceso para hacer las correcciones a la información utilizando el software SIG, el requerimiento de este rol es que las tablas de las capas que se nombran a continuación tengan llenos todos los campos especificados en las plantillas respectivas indicadas en la sección “Documentos y/o plantillas relacionadas” del Anexo 1 para las capas:
 - POSTES_SinAlimentador
 - POSTES_SinProyectoModificacion
 - POSTES_SinCodigoEstructura
 - POSTES_SinCanton.

5.1.5.5 Proveedores (del proceso) y sus Requisitos

Los proveedores de este proceso son dos:

5.1.5.5.1 *El Sistema de Información Geográfico (SIG)*

Los requisitos de este proveedor están relacionados con las capas que se utilizan de este sistema para la depuración de la información de los postes y sus estructuras montadas y su carga en la capa “POSTES” y “ESTRUCEnPOSTE”. Estas capas son tres y deben estar en la misma proyección espacial, existen además de una tabla donde se almacena la relación entre los postes y sus estructuras montadas. A continuación se describen éstas:

- La primera capa que a su vez contiene su tabla respectiva es EstructuraSoporte, la misma que está almacenadas en el SIG, contienen la información de la localización de los Postes de la EERCS y los valores en sus campos deben estar acorde con las

plantillas respectivas indicadas en la sección “Documentos y/o plantillas relacionadas” del Anexo 1, además de cumplir los siguientes requisitos:

- El campo ProyectoModificacion debe contener el código correspondiente al último proyecto, en la tabla CGP del Sistema de Gestión de Proyectos, que modificó un poste en campo y no puede ser nulo.
 - El campo CodigoEstructura debe contener el código correspondiente al catálogo de estructuras, en la tabla PUE del Sistema de Gestión de Proyectos, para cada poste y no puede ser nulo.
- Otra capa necesaria para este proceso en la capa “TRAMOS” la cual es una capa de líneas que contiene la información del Alimentador y será utilizada en los geoprocetos para la asignación de la información del Alimentador en la capa POSTES resultante. Los valores en los campos de esta tabla deben estar acorde con la plantilla para la capa “TRAMOS” indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1.
 - La tabla EstructuraEnPoste, la misma que está almacenadas en el SIG, que es donde se almacena la relación entre el tipo de estructura montada en un poste y su relación y ubicación espacial con el poste mismo. Los valores en sus campos deben estar acorde con las plantillas respectivas indicadas en la sección “Documentos y/o plantillas relacionadas” del Anexo 1, además de cumplir los siguientes requisitos:
 - El campo CodigoEstructura debe contener el código correspondiente al catálogo de estructuras, en la tabla PUE del Sistema de Gestión de Proyectos, para cada estructura en poste y no puede ser nulo.

- EstructuraSoporteObjectId, que es el campo que relación las estructura montadas en un poste con el Id de los elementos gráficos en la capa EstructuraSoporte, este campo no debe ser nulo.
- La última capa necesaria para este proceso es la capa “AC_CANTONES” la cual es una capa de polígonos que contiene la división política de los cantones del país y será utilizada en los geoprosos para la asignación de la información de la provincia y cantón en la capa POSTES resultante. Los valores en los campos de esta tabla deben estar acorde con la plantilla para la capa “AC_CANTONES” indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1.

5.1.5.5.2 *El Sistema de Gestión de Proyectos (SGP)*

Los requisitos para este proveedor están relacionados con las tablas que se utilizan de este sistema para la depuración y carga en la capa “POSTES” y la tabla “ESTRUCEnPOSTE” del ODS de la información de las redes. Estas tablas están almacenadas en el servidor AS400 en producción y son las siguientes:

- a. La tabla CGP, que contiene la información de los proyectos de electrificación y donde los valores de sus campos no pueden ser nulos y deben estar acuerdo al formato especificado para la tabla CGP indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1, además de cumplir los siguientes requisitos:
 - TipCod: que corresponde al tipo de proyecto, debe estar lleno con el respectivo tipo de proyecto.

- CGPSeq: que es un número secuencia, que junto con el campo TipCod conforman el código del proyecto, debe estar lleno con el correspondiente valor del proyecto.
 - CGPDes: que corresponde con el nombre del proyecto, esté lleno con el respectivo nombre del proyecto.
 - CGPFEC: que corresponde a la fecha de construcción del proyecto debe estar lleno con la respectiva fecha de construcción del proyecto.
- b. La tabla PUE, que contiene la información de las unidades constructivas y sus costos y donde los valores de sus campos no pueden ser nulos y deben estar acuerdo al formato especificado para la tabla PUE, indicada en la sección “Documentos y/o plantillas relacionadas” del Anexo 1, además de cumplir los siguientes requisitos:
- UNPCOD que contiene el código de las unidades de propiedad.
 - UNRESP que es un secuencial que junto con UNPCOD conforman el código de estructura.
 - PUEMOC el cual contiene el costo de mano de obra calificada para su respectiva estructura.
 - PUEMONC el cual contiene el costo de mano de obra no calificada para su respectiva estructura.
 - PUEMAT el cual contiene el costo del material para su respectiva estructura.
 - PUEEQU el cual contiene el costo del equipo utilizado para su respectiva estructura.
 - PUEADMT el cual contiene el costo de administración para su respectiva estructura.

5.1.5.6 Descripción detallada del proceso:

El proceso de ETL para las POSTES y ESTRUCEnPOSTE se lo detalla a continuación como actividades:

1. Leer la tabla de la capa EstructuraEnPoste desde el Sistema de Información Geográfico.
2. Se verifica si el campo "ProyectoModificacion", es nulo.
3. El campo "ProyectoModificacion" almacena el código de la carpeta del último proyecto con el que se realizaron modificaciones en un poste, si este campo está vacío, quiere decir que no se han hecho mejoras o modificaciones y se debe almacenar el código que está en el campo "ProyectoConstrucción", que es el de la construcción original del elemento, por lo tanto si el campo "ProyectoModificacion" es nulo, se llena el campo "ProyectoModificacion" con el valor respectivo del campo "Proyecto Construcción".
4. Mediante un Join entre tablas verifica que los valores en el campo "ProyectoModificacion" de la tabla POSTES tengan correspondencia con los campos TIPCOD+CGPSEC de la tabla CGP del Sistema de Gestión de Proyectos (SGP).
5. Los registros cuyo campo ProyectoModificacion es nulo o no tienen correspondencia en el Join del paso anterior, son almacenados en una capa llamada "POSTES_SinProyectoModificacion", la cual es una capa espacial, que sirve para registrar dónde está ubicado el elemento que no tiene un valor adecuado en sus campos "ProyectoModificacion" para que se puedan tomar los correctivos necesarios.

6. En los registros que sí tienen correspondencia con el Join del paso 4 se actualiza el campo “FECHA_PROYECTO” con el valor correspondiente del campo CGP.CGPFEC, que es la fecha de energización del proyecto.
7. Se hace un Join entre la tabla de la capa POSTES con el campoCodigoEstructura y las tablas PUE con los campos UNPCOD+UNRESP, para verificar los registros de la tabla POSTES que tienen correspondencia con los de la tabla PUE del sistema de gestión de proyectos.
8. Los registros cuyo campo CodigoEstructura es nulo o no tienen correspondencia en el Join del paso anterior, son almacenados en una capa llamada “POSTES_SinCodigoEstructura”, la cual es una capa espacial, que sirve para registrar dónde está ubicado el elemento que no tiene un valor adecuado en sus campos “ProyectoModificacion” para que se puedan tomar los correctivos necesarios.
9. En los registros que sí tienen correspondencia con el Join del paso 7 se actualiza los campos: PUEMOC, PUEMONC, PUEMAT, PUEEQU que son los costos de puesta en operación de los medidores.
10. Se realiza un geoproceto de sobre posición espacial contra la capa AC_Cantones para ubicar espacialmente en qué provincia y cantón está ubicada los respectivos postes. Este geoproceto consiste en hacer un Join espacial de la capa POSTES contra la capa AC_Cantones, lo que da como resultado que a la capa POSTES se le añadan los campos de la capa AC_Cantones que coinciden espacialmente (los postes que están dentro del respectivo polígono que representa los límites de un cantón). Se procede a actualizar el campo Canton de la tabla POSTES y al final del proceso se toma los dos primeros dígitos de la capa AC_Cantones, para además llenar el campo Provincia y de esta manera se

evita tener que realizar otro geoproceso espacial para llenar el campo de la provincia en la capa POSTES.

11. Verifica que los campos "Provincia" y "Canton" no sean nulos. Esto podría darse por ejemplo en el caso de que por error se grafique un poste fuera de los límites del Ecuador para donde no tenemos información en nuestra capa "AC_Cantones".
12. Si los campos "Provincia" y "Canton" son nulos en algunos registro, estos son filtrados y almacenados en una capa llamada "POSTES_SinCanton", que sirve para ubicar donde están estos postes que no caen dentro de la capa de cantones y provincias.
13. Realiza un geoproceso de proximidad espacial, entre las capa TRAMOS y POSTES, para actualizar el valor del campo "Alimentador" basado en el valor del campo alimentador del tramo más cercano.
14. Luego se verifica el campo "Alimentador" de la capa resultante del geoproceso anterior, para saber cuáles son nulos.
15. Se almacena en la capa "POSTES_SinAlimentador" los registros donde el valor en el campo "Alimentador" es nulo o es incorrecto.
16. En los registros que sí tienen correspondencia con el Join espacial del paso 13 se actualiza el campo "Alimentador".
17. Una vez obtenida la capa POSTES con sus datos depurados hace un join con la tabla EstructuraEnPoste, para asignar a cada uno de los registros de esta tabla, los campos depurados de la tabla POSTES. El join de estas dos tablas se las realiza por los campos EstructuraSoporte.ObjectID y el campo EstructuraEnPoste.EstructuraSoporteObjectID.

18. Por último se hace un join con la tabla PUE del SGP, para actualizar los campos de costos, campos son: PUEMOC, PUEMONC, PUEMAT, PUEEQU. Y los campos por los cuales se relacionan son ESTRUCEnPOSTE.CodigoEstructura y PUE.UNPCOD+PUE.UNRESP
19. Como resultado de cada una de estas actividades en este proceso se obtiene una capa llamada POSTES la cual almacena como puntos, toda la información depurada de los postes tomados del SIG y una tabla llamada ESTRUCEnPOSTE donde se almacena toda la información depurada de las estructuras montadas en los postes. En ambos elementos se debe garantizar que se cumplan los criterios de aceptación mencionados en la siguiente sección.
- Estas capas quedan almacenadas en el ODS, las cuales a su vez sirven de suministro y fuente depurada de datos para la el proceso ETL al Data Mart.

5.1.5.7 Criterios de Aceptación:

Se debe cumplir los siguientes criterios para ser considerada una carga exitosa de la capa POSTES y ESTRUCEnPOSTE en el ODS:

- 1 El campo Alimentador debe contener el código correspondiente al respectivo alimentador y no puede ser nulo.
- 2 El campo Subestación debe contener el número de subestación correspondiente a la respectivo subestación y no puede ser nulo.
- 3 El campo ProyectoModificacion debe contener el código correspondiente al último código de proyecto, en la tabla CGP del Sistema de Gestión de Proyectos, que modificó al elemento y no puede ser nulo.

- 4 El campo `CodigoEstructura` debe contener el código correspondiente al catálogo de estructuras, en la tabla PUE del Sistema de Gestión de Proyectos, para el medidor del cliente respectivo y no puede ser nulo.
- 5 El campo `Provincia` debe contener el código correspondiente a la provincia en la que está ubicado espacialmente el cliente y no puede ser nulo.
- 6 El campo `Cantón` debe contener el código correspondiente al cantón en la que está ubicado espacialmente y no puede ser nulo.

5.1.5.8 Indicador de Excepciones

Una parte importante del proceso es determinar las excepciones del mismo, es decir los datos que no fueron transferidos por qué no cumplieron los criterios de aceptación (ver arriba), como resultado de este proceso obtenemos las capas

- `POSTES_SinAlimentador`
- `POSTES_SinProyectoModificacion`
- `POSTES_SinCodigoEstructura`
- `POSTES_SinCanton`

Donde no solo podemos determinar cuáles son los postes que no cumplieron los criterios de aceptación, sino también dónde están ubicados espacialmente, cuándo fueron graficados o modificados y quien los graficó, para poder identificar tendencias que se pueden mostrar en el transcurso del tiempo, y si se analizan de manera integral permiten mantener un diagnóstico de la situación y tomar acciones preventivas o correctivas sobre los datos transacciones y verificar si éstas fueron o no acertadas.

5.1.5.9 Glosario o Definición de Términos

SGP.- Sistema de Gestión de Proyectos.

SIG.- Sistema de Información Geográfico.

Alimentador.- El alimentar pertenece a una subestación y es la fuente de la energía a ser distribuida.

Proyecto Modificación.- Es el código de la carpeta a la que pertenecen los elementos de un proyecto de electrificación.

Código Estructura.- Es el código de la estructura asignada como una unidad de Transformación.

ODS.- Operational Data Store, Base de Datos Operacional, es una base de datos diseñada para integrar datos de múltiples fuentes para hacer más fácil el análisis y la presentación de informes. Debido a que los datos origen vienen de diferentes fuentes, la integración implica a menudo la limpieza, la solución de redundancia y comprobación de reglas de negocio para integridad. El ODS es generalmente diseñado para contener datos de bajo nivel o indivisible (como las transacciones y los precios) con una ventana de actualización muy pequeña en contraposición con el mayor volumen de datos almacenados en el Data Warehouse que generalmente es con menor frecuencia. En otras palabras, es un contenedor de datos activos, es decir operacionales que ayudan al soporte de decisiones y a la operación y está entre un OLAP y un OLTP.

Auxiliar SIG.- Es un rol dentro del departamento SIGADE encargado de la actualización de la información espacial en el SIG.

5.1.5.10 Documentos y/o plantillas relacionadas:

En el anexo 1 se presentan las estructuras de datos (plantilla) para cada una de las tablas fuente y resultantes del presente proceso. En cada plantilla se puede obtener datos como: Nombre del campo, tipo, tamaño y una pequeña descripción.

5.1.5.11 Diagramas de actividades (workflows):

A continuación se adjunta el diagrama de actividades del proceso ETL para los Postes del ODS:

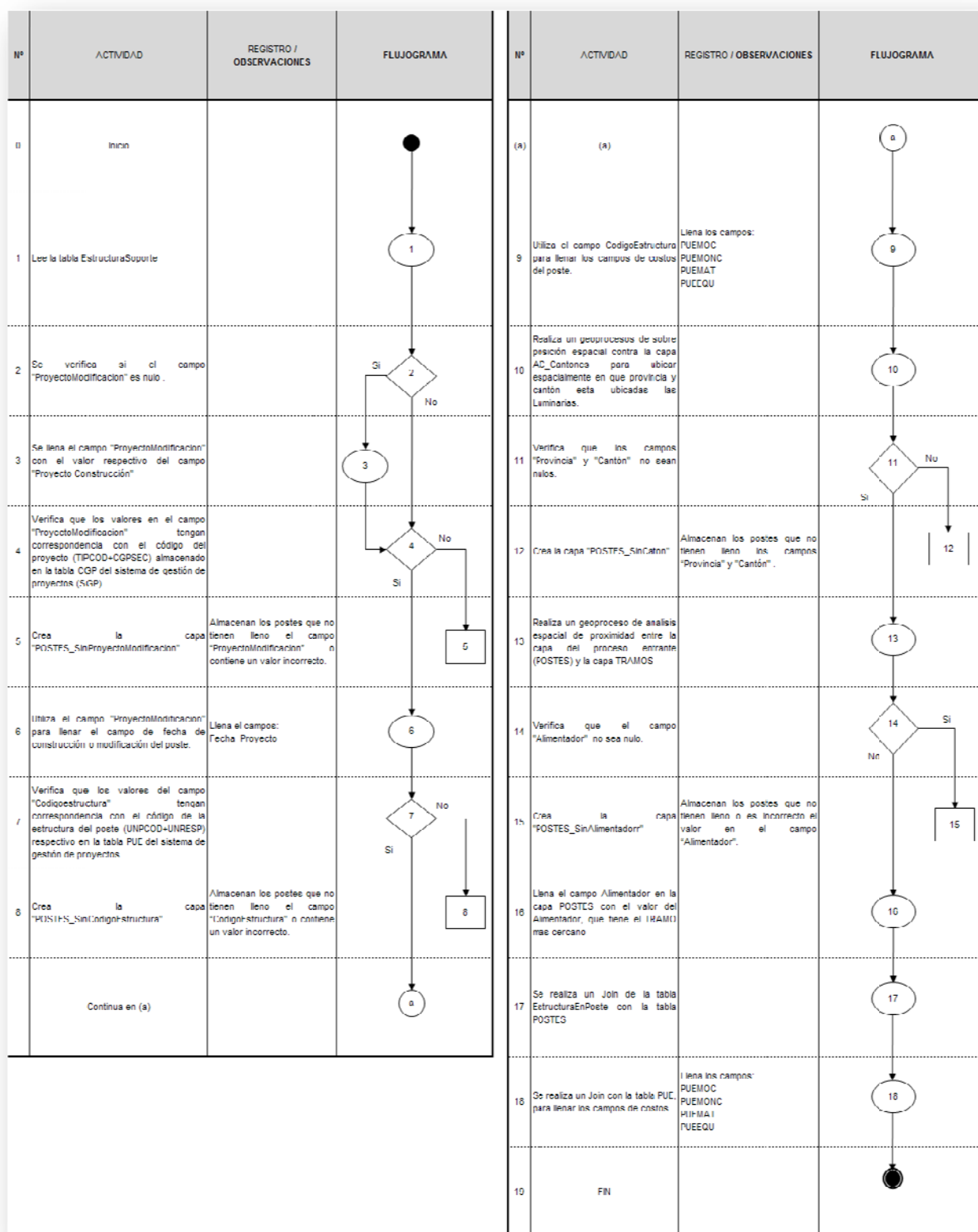


Gráfico 26. Diagrama de actividades del proceso ETL para los postes del ODS.

5.1.5.12 Normas y políticas para el proceso:

a. Requerimientos de Información:

Los requerimientos de información fuente del SIG y SGP, deberían ser cumplidos y corregidos en caso de que no cumplan los criterios de aceptación.

b. Frecuencia:

Este proceso de actualización de la capa “POSTES” del ODS, debería ser corrido todos los fines de semana, para mantener la información actualizada con una ventana de tiempo pequeña.

5.2 Procesos ETL para la carga de las tablas de hechos en el Data Mart.

5.2.1 Breve descripción del proceso:

Este proceso sirve para la extracción, transformación, depuración y carga en el Data Mart de la información desde las capas del ODS, donde se ha realizado la limpieza y una primera transformación de la información tabular y geográfica, hacia el Data Mart.

5.2.2 Responsable:

El responsable de este proceso es un rol llamado Superintendente Informático, el cual es el administrador del SIG y solo para el caso del prototipo también será el administrador del Data Mart.

5.2.3 Objetivo y Alcance:

El objetivo general de este proceso es cargar las tablas de hechos del Data Mart.

Los objetivos específicos de este proceso son:

- Cargar la tabla de hechos WH_CUBO_TRAMOS, donde tenemos toda la información de las métricas para los tramos.
- Cargar la tabla de hechos WH_CUBO_TRANSFORMADORES, donde tenemos toda la información de las métricas para los transformadores.
- Cargar la tabla de hechos WH_CUBO_CLIENTES, donde tenemos toda la información de las métricas para los clientes.
- Cargar la tabla de hechos WH_CUBO_LUMINARIA, donde tenemos toda la información de las métricas para las luminarias.
- Cargar la tabla de hechos WH_CUBO_POSTES, donde tenemos toda la información de las métricas para los postes.

- Cargar la tabla de hechos WH_CUBO_ESTRUCTURA_POSTE, donde tenemos toda la información de las métricas para las estructuras en poste.

5.2.4 Clientes (del proceso) y sus Requisitos:

Los clientes del proceso actual son:

- El Data Mart en sí que es el que recibe los datos.
- Las herramientas de explotación de Data Mart, como son: el Oracle BI Answer y el Oracle Interactive Dashboards.

5.2.5 Proveedores (del proceso) y sus Requisitos

Los proveedores de este proceso son dos:

El Operational Data Store (ODS)

Los requisitos de este proveedor están relacionados con las capas que fueron depuradas y almacenadas en el ODS, las cuales deben tener llenos todos los campos especificados en la plantilla para las capas TRAMOS, TRAFOS, LUMINARIAS_ODS, CLIENTES, POSTES y ESTRUCEnPOSTE indicadas en la sección “Documentos y/o plantillas relacionadas” del Anexo 1 y cumplir los criterios de aceptación de cada uno de los procesos de la sección 5.1 de este documento.

5.2.6 Descripción del proceso:

El proceso de ETL para la carga de las tablas de hechos del Data Mart se lo detalla a continuación como actividades:

1. Lee la tabla de la capa TRAMOS desde el ODS que contiene toda la información depurada y producto de los geoprocursos espaciales.

2. Correr el proceso llamado "EXPRESSION" del Oracle Warehouse Builder, para realizar cálculos adicionales de los datos de la tabla TRAMOS. Las especificaciones de estos cálculos están detalladas como metadatos a continuación.

Metadatos para el proceso ETL de carga de la tabla de hechos WH_CUBO_TRAMOS en el Data Mart:

| DESTINO | FUENTE |
|-----------------------------|--|
| DW_ELECTRICO.WH_CUBO_TRAMOS | PROCESS.TRAMOS |
| *Aereo_Subt_ID:INTEGER | TRAMOS.Aereo_Subt_ID |
| *Baja_Media_ID:INTEGER | TRAMOS.Baja_Media_ID |
| *Fase_ID:INTEGER | TRAMOS.Fase_ID |
| *Acom_Bajante_ID:INTEGER | TRAMOS.Acom_Bajante_ID |
| *Canton_ID:VARCHAR(4) | TRAMOS.CANTON |
| *Alimentador_ID:VARCHAR(4) | TRAMOS.ALIMENTADORID |
| *Proyecto_ID:VARCHAR(10) | TRAMOS.PROYECTOMODIFICACION |
| *Mes_ID:INTEGER | FECHA(Mes) |
| Longitud:DOUBLE | Sum (TRAMOS.LONGITUDSISTEMA) |
| ValorManoObraCalif:FLOAT | Sum (TRAMOS.PUEMOC_FASE x (LONGITUDSIST_FASE/1000) + TRAMOS.PUEMOC_NEUTRO x (LONGITUDSIST_NEUTRO/1000)) |
| ValorManoObraNoCalif:FLOAT | Sum (TRAMOS.PUEMONC_FASE x (LONGITUDSIST_FASE/1000) + TRAMOS.PUEMONC_NEUTRO x (LONGITUDSIST_NEUTRO/1000)) |
| ValorMaterial:FLOAT | Sum (TRAMOS.PUEMAT_FASE x (LONGITUDSIST_FASE/1000) + TRAMOS.PUEMAT_NEUTRO x (LONGITUDSIST_NEUTRO/1000)) |
| ValorEquipos:FLOAT | Sum (TRAMOS.PUEEQU_FASE x (LONGITUDSIST_FASE/1000) + TRAMOS.PUEEQU_NEUTRO x (LONGITUDSIST_NEUTRO/1000)) |
| ValorAdmin:FLOAT | Sum (15% x (ValorManoObraCalif + ValorManoObraNoCalif + ValorMaterial + ValorEquipos)) |
| ValorTotal:FLOAT | Sum (ValorManoObraCalif + ValorManoObraNoCalif + ValorMaterial + ValorEquipos + ValorAdmin) |
| Observaciones: | <pre> if TRAMOS.N_CONDUCTORES > TRAMOS.N_FASES then LONGITUDSIST_FASE = TRAMOS.LONGITUDSISTEMA x (TRAMOS.N_FASES - 1) LONGITUDSIST_NEUTRO = TRAMOS.LONGITUDSISTEMA else LONGITUDSIST_FASE = TRAMOS.LONGITUDSISTEMA x (TRAMOS.N_FASES) LONGITUDSIST_NEUTRO = 0 end if </pre> |

Tabla 7. Metadatos para el proceso ETL de carga de la tabla de hechos WH_CUBO_TRAMOS en el Data Mart.

3. Realiza el proceso llamado "AGGREGATOR" para hacer los precálculos de las métricas de acuerdo a las dimensiones para los Tramos.
4. Realiza la carga de la información a la tabla de hechos "WH_CUBO_TRAMOS"

5. Lee la tabla de la capa TRAFOS desde el ODS que contiene toda la información depurada y producto de los geoprocesos espaciales.
6. Corre el proceso llamado "EXPRESSION" del Oracle Warehouse Builder, para realizar cálculos adicionales de los datos de la tabla TRAFOS. Las especificaciones de estos cálculos están detalladas como metadatos a continuación.

Metadatos para el proceso ETL de carga de la tabla de hechos

WH_CUBO_TRANSFORMADORES en el Data Mart:

| DESTINO | FUENTE |
|--------------------------------------|---|
| DW_ELECTRICO.WH_CUBO_TRANSFORMADORES | PROCESS.TRAFOS |
| *Fase_ID:INTEGER | TRAFOS.FASE_ID |
| *Propietario_ID:VARCHAR(10) | TRAFOS.PROPIEDAD |
| *Canton_ID:VARCHAR(4) | TRAFOS.CANTON |
| *Alimentador_ID:VARCHAR(4) | TRAFOS.ALIMENTADORID |
| *Proyecto_ID:VARCHAR(10) | TRAFOS.PROYECTOMODIFICACION |
| *Mes_ID:INTEGER | FECHA(Mes) |
| Cantidad:INTEGER | Count (TRAFOS.OBJECT_ID) |
| POTENCIA:FLOAT | Sum (TRAFOS.POTENCIANOMINAL) |
| ValorManoObraCalif:FLOAT | Sum (TRAFOS.PUEMOC) |
| ValorManoObraNoCalif:FLOAT | Sum (TRAFOS.PUEMONC) |
| ValorMaterial:FLOAT | Sum (TRAFOS.PUEMAT) |
| ValorEquipos:FLOAT | Sum (TRAFOS.PUEEQU) |
| ValorAdmin:FLOAT | Sum (15% x (TRAFOS.PUEMOC+TRAFOS.PUEMONC+TRAFOS.PUEMAT+TRAFOS.PUEEQU)) |
| ValorTotal:FLOAT | Sum (ValorManoObraCalif+ValorManoObraNoCalif+ValorMaterial+ValorEquipos+ValorAdmin) |

Tabla 8. Metadatos para el proceso ETL de carga de la tabla de hechos WH_CUBO_TRANSFORMADORES en el Data Mart.

7. Realiza el proceso llamado "AGGREGATOR" para hacer los precálculos de las métricas de acuerdo a las dimensiones para los Transformadores.
8. Realiza la carga de la información a la tabla de hechos "WH_CUBO_TRANSFORMADORES"

9. Lee la tabla de la capa LUMINARIAS_ODS desde el ODS que contiene toda la información depurada y producto de los geoprocursos espaciales.
10. Corre el proceso llamado "EXPRESSION" del Oracle Warehouse Builder, para realizar cálculos adicionales de los datos de la tabla LUMINARIAS_ODS. Las especificaciones de estos cálculos están detalladas como metadatos a continuación.

Metadatos para el proceso ETL de carga de la tabla de hechos WH_CUBO_LUMINARIA en el Data Mart:

| DESTINO | FUENTE |
|--------------------------------|---|
| DW_ELECTRICO.WH_CUBO_LUMINARIA | PROCESS.LUMINARIAS_ODS |
| *TipoLuminaria_ID:VARCHAR(15) | LUMINARIA.TIPO |
| *PotenciaLuminaria_ID:FLOAT | LUMINARIA.POTENCIA |
| *Canton_ID:INTEGER | LUMINARIA.CANTON |
| *Alimentador_ID:INTEGER | LUMINARIA.ALIMENTADORID |
| *Proyecto_ID:INTEGER | LUMINARIA.PROYECTOMODIFICACION |
| *Mes_ID:INTEGER | FECHA(Mes) |
| Cantidad:INTEGRE | Count (LUMINARIA.OBJECTID) |
| Potencia_Instalada:FLOAT | Sum (LUMINARIA.POTENCIA) |
| Energia:FLOAT | Sum (LUMINARIA.ENERGIA) |
| ValorManoObraCalif:FLOAT | Sum (LUMINARIA.PUEMOC) |
| ValorManoObraNoCalif:FLOAT | Sum (LUMINARIA.PUEMONC) |
| ValorMaterial:FLOAT | Sum (LUMINARIA.PUEMAT) |
| ValorEquipos:FLOAT | Sum (LUMINARIA.PUEEQU) |
| ValorAdmin:FLOAT | Sum (15% x (PUE.PUEMOC+PUE.PUEMONC+PUE.PUEMAT+PUE.PUEEQU)) |
| ValorTotal:FLOAT | Sum (ValorManoObraCalif + ValorManoObraNoCalif + ValorMaterial + ValorEquipos + ValorAdmin) |
| Observaciones | ENERGIA= <i>if</i> LUMINARIA.DOUBLENIVEL LIKE 'S' <i>THEN</i> (((LUMINARIA.POTENCIA * LUMINARIA.HORASFUNC1 * LUMINARIA.DIASFUNCMES) + (LUMINARIA.POTENCIA2 * LUMINARIA.HORASFUNC2 * LUMINARIA.DIASFUNCMES)) / 1000) * (1 + (LUMINARIA.PERDIDA_PORCENTAJE / 100)) ELSE (((LUMINARIA.POTENCIA * LUMINARIA.HORASFUNC1 * LUMINARIA.DIASFUNC MES) / 1000) * (1 + (LUMINARIA.PERDIDA_PORCENTAJE / 100))) |

Tabla 9. Metadatos para el proceso ETL de carga de la tabla de hechos WH_CUBO_LUMINARIA en el Data Mart.

11. Realiza el proceso llamado "AGGREGATOR" para hacer los precálculos de las métricas de acuerdo a las dimensiones para las Luminarias.
12. Realiza la carga de la información a la tabla de hechos "WH_CUBO_LUMINARIA".

13. Lee la tabla de la capa CLIENTES desde el ODS que contiene toda la información depurada y producto de los geoprocursos espaciales.
14. Corre el proceso llamado "EXPRESSION" del Oracle Warehouse Builder, para realizar cálculos adicionales de los datos de la tabla CLIENTES. Las especificaciones de estos cálculos están detalladas como metadatos a continuación.

Metadatos para el proceso ETL de carga de la tabla de hechos

WH_CUBO_CLIENTES en el Data Mart:

| DESTINO | FUENTE |
|-------------------------------|---|
| DW_ELECTRICO.WH_CUBO_CLIENTES | PROCESS.CLIENTES |
| *Baja_Media_ID:INTEGER | CLIENTES.Baja_Media_ID |
| *Canton_ID:INTEGER | CLIENTES.CANTON |
| *Alimentador_ID:INTEGER | CLIENTES.ALIMENTADORID |
| *Proyecto_ID:INTEGER | CLIENTES.PROYECTOMODIFICACION |
| *Mes_ID:INTEGER | FECHA(Mes) |
| Cantidad:INTEGRE | Count (CLIENTES.OBJECT_ID) |
| Energia:Double | Sum (CLIENTES.ENERGIA) |
| ValorManoObraCalif:FLOAT | Sum (CLIENTES.PUEMOC) |
| ValorManoObraNoCalif:FLOAT | Sum (CLIENTES.PUEMONC) |
| ValorMaterial:FLOAT | Sum (CLIENTES.PUEMAT) |
| ValorEquipos:FLOAT | Sum (CLIENTES.PUEEQU) |
| ValorAdmin:FLOAT | Sum (15% x (PUE.PUEMOC+PUE.PUEMONC+PUE.PUEMAT+PUE.PUEEQU)) |
| ValorTotal:FLOAT | Sum (ValorManoObraCalif + ValorManoObraNoCalif + ValorMaterial + ValorEquipos + ValorAdmin) |

Tabla 10. Metadatos para el proceso ETL de carga de la tabla de hechos WH_CUBO_CLIENTES en el Data Mart.

15. Realiza el proceso llamado "AGGREGATOR" para hacer los precálculos de las métricas de acuerdo a las dimensiones para los Clientes.
16. Realiza la carga de la información a la tabla de hechos "WH_CUBO_CLIENTES".

17. Lee la tabla de la capa POSTES desde el ODS que contiene toda la información depurada y producto de los geoprocursos espaciales.
18. Corre el proceso llamado "EXPRESSION" del Oracle Warehouse Builder, para realizar cálculos adicionales de los datos de la tabla POSTES. Las especificaciones de estos cálculos están detalladas como metadatos a continuación.

Metadatos para el proceso ETL de carga de la tabla de hechos WH_CUBO_POSTES en el Data Mart:

| DESTINO | FUENTE |
|-----------------------------|---|
| DW_ELECTRICO.WH_CUBO_POSTES | PROCESS.POSTES |
| *TipoPoste_ID:VARCHAR(10) | POSTES.TipoPoste_ID |
| *Canton_ID:INTEGER | POSTES.CANTON |
| *Alimentador_ID:INTEGER | POSTES.ALIMENTADORID |
| *Proyecto_ID:INTEGER | POSTES.PROYECTOMODIFICACION |
| *Mes_ID:INTEGER | FECHA(Mes) |
| Cantidad:INTEGER | Count (POSTES.OBJECT_ID) |
| ValorManoObraCalif:FLOAT | Sum (POSTES.PUEMOC) |
| ValorManoObraNoCalif:FLOAT | Sum (POSTES.PUEMONG) |
| ValorMaterial:FLOAT | Sum (POSTES.PUEMAT) |
| ValorEquipos:FLOAT | Sum (POSTES.PUEEQU) |
| ValorAdmin:FLOAT | Sum (15% x (POSTES.PUEMOC+POSTES.PUEMONG+POSTES.PUEMAT+POSTES.PUEEQU)) |
| ValorTotal:FLOAT | Sum (ValorManoObraCalif + ValorManoObraNoCalif + ValorMaterial + ValorEquipos + ValorAdmin) |

Tabla 11. Metadatos para el proceso ETL de carga de la tabla de hechos WH_CUBO_POSTES en el Data Mart.

19. Realiza el proceso llamado "AGGREGATOR" para hacer los precálculos de las métricas de acuerdo a las dimensiones para los Postes.
20. Realiza la carga de la información a la tabla de hechos "WH_CUBO_POSTES".
21. Lee la tabla de la capa ESTRUCENPOSTE desde el ODS que contiene toda la información depurada y producto de los geoprocursos espaciales.

22. Corre el proceso llamado "EXPRESSION" del Oracle Warehouse Builder, para realizar cálculos adicionales de los datos de la tabla ESTRUCENPOSTE. Las especificaciones de estos cálculos están detalladas como metadatos a continuación.

Metadatos para el proceso ETL de carga de la tabla de hechos WH_CUBO_ESTRUCTURA_POSTE en el Data Mart:

| DESTINO | FUENTE |
|---------------------------------------|--|
| DW_ELECTRICO.WH_CUBO_ESTRUCUTRA_POSTE | PROCESS.ESTRUCENPOSTE |
| *Canton_ID:INTEGER | <i>EstrucEnPoste.CANTON</i> |
| *Alimentador_ID:INTEGER | <i>EstrucEnPoste.ALIMENTADORID</i> |
| *Proyecto_ID:INTEGER | <i>EstrucEnPoste.PROYECTOMODIFICACION</i> |
| *Mes_ID:INTEGER | <i>FECHA(Mes)</i> |
| Cantidad:INTEGER | <i>Count (EstrucEnPoste.OBJECT_ID)</i> |
| ValorManoObraCalif:FLOAT | <i>Sum (EstrucEnPoste.PUEMOC)</i> |
| ValorManoObraNoCalif:FLOAT | <i>Sum (strucEnPoste.PUEMONC)</i> |
| ValorMaterial:FLOAT | <i>Sum (EstrucEnPoste.PUEMAT)</i> |
| ValorEquipos:FLOAT | <i>Sum (EstrucEnPoste.PUEEQU)</i> |
| ValorAdmin:FLOAT | <i>Sum (15% x (PUE.PUEMOC+PUE.PUEMONC+PUE.PUEMAT+PUE.PUEEQU))</i> |
| ValorTotal:FLOAT | <i>Sum (ValorManoObraCalif + ValorManoObraNoCalif + ValorMaterial + ValorEquipos + ValorAdmin)</i> |

Tabla 12. Metadatos para el proceso ETL de carga de la tabla de hechos WH_CUBO_ESTRUCTURA_POSTE en el Data Mart.

23. Realiza el proceso llamado "AGGREGATOR" para hacer los precálculos de las métricas de acuerdo a las dimensiones para los Clientes.

24. Realiza la carga de la información a la tabla de hechos "WH_CUBO_ESTRUCTURA_POSTE"

25. Como resultado de cada una de estas actividades en este proceso se obtiene la carga completa de todas las tablas de hechos del Data Mart. En ambos

elementos se debe garantizar que se cumplan los criterios de aceptación mencionados en la siguiente sección.

5.2.7 Criterios de Aceptación:

Se debe cumplir los siguientes criterios para ser considerada una carga exitosa de las tablas de hechos del Data Mart:

- 1 Ningún campo de las tablas de hechos puede contener valores nulos.
- 2 El indicador de carga completa de cada uno de las tablas de hechos debe ser satisfactorio
- 3 No pueden existir valores en los códigos de las tablas de hechos que no tengan correspondencia con sus dimensiones.

5.2.8 Glosario o Definición de Términos

ODS.- Operational Data Store, Base de Datos Operacional, es una base de datos diseñada para integrar datos de múltiples fuentes para hacer más fácil el análisis y la presentación de informes. Debido a que los datos origen vienen de diferentes fuentes, la integración implica a menudo la limpieza, la solución de redundancia y comprobación de reglas de negocio para integridad. El ODS es generalmente diseñado para contener datos de bajo nivel o indivisible (como las transacciones y los precios) con una ventana de actualización muy pequeña en contraposición con el mayor volumen de datos almacenados en el Data Warehouse que generalmente es con menor frecuencia. En otras palabras, es un contenedor de datos activos, es decir operacionales que ayudan al soporte de decisiones y a la operación y está entre un OLAP y un OLTP.

Data Mart.- Un Data Mart es una versión especial de almacén de datos (Data Warehouse). Son subconjuntos de datos con el propósito de ayudar a que un área

específica dentro del negocio pueda tomar mejores decisiones. Los datos existentes en este contexto pueden ser agrupados, explorados y propagados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los mismos de la forma más conveniente según sus necesidades.

5.2.9 Documentos y/o plantillas relacionadas:

El anexo 1 se presenta las estructuras de datos (plantilla) para cada una de las tablas resultantes del presente proceso. En cada plantilla se puede obtener datos como: Nombre del campo, tipo, tamaño y una pequeña descripción.

5.2.10 Diagramas de actividades (workflows):

A continuación se adjunta el diagrama de actividades del proceso ETL para carga de las tablas de hechos en el Data Mart:

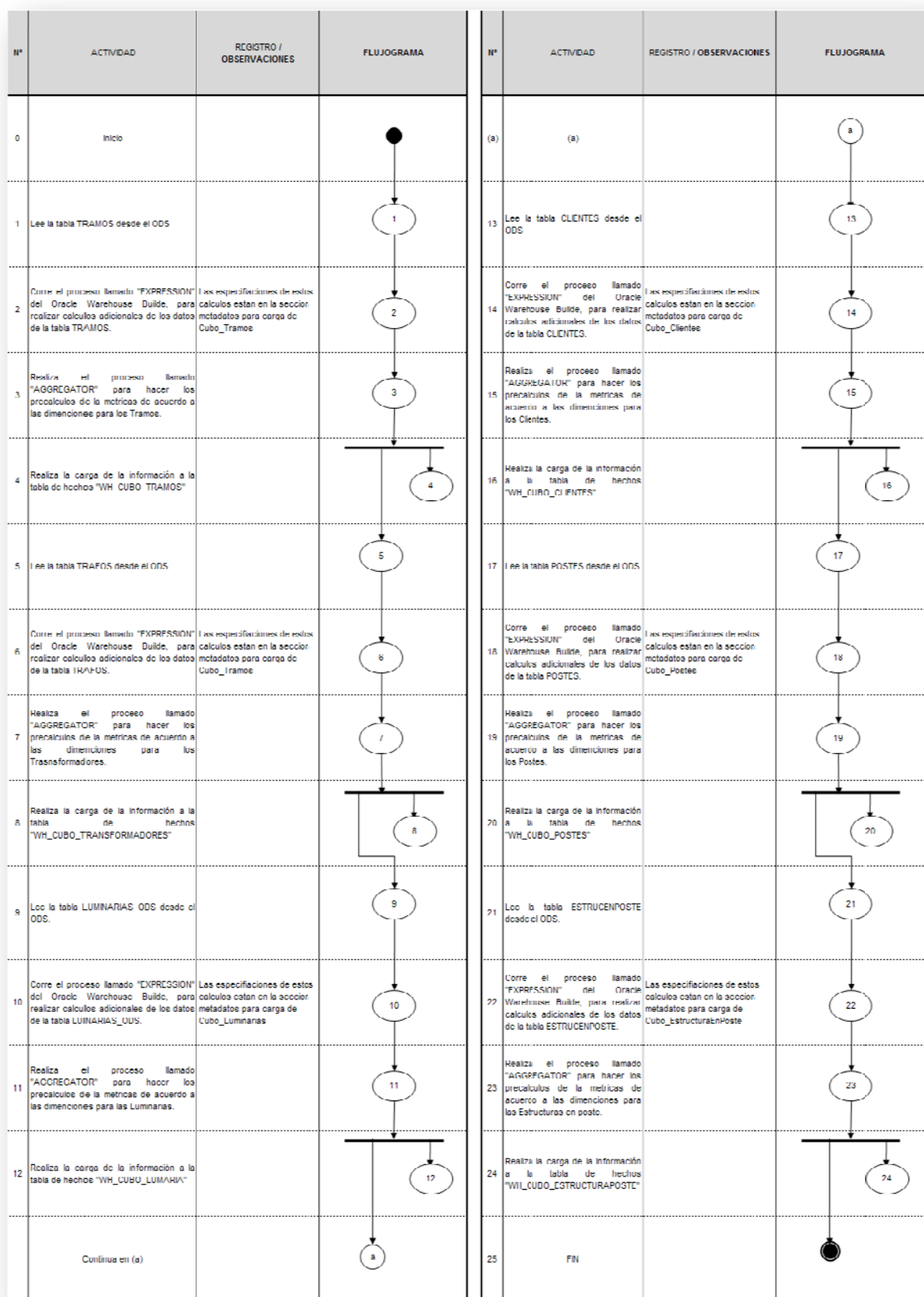


Gráfico 27. Diagrama de actividades del proceso ETL para carga de las tablas de hechos en el Data Mart.

5.2.11 Normas y políticas para el proceso:

a. Requerimientos de Información:

Los requerimientos de información fuente desde el ODS, deberían ser cumplidos y corregidos en caso de que no cumplan los criterios de aceptación.

b. Frecuencia:

Este proceso de carga de tablas de hechos en el Data Mart, debería ser corrido todos los fines de mes, para que sirva de información fuente para los informes mensuales y de soporte para estrategias a corto y largo plazo.

6. Descripción del prototipo

6.1 Origen y reseña de acontecimientos

La Empresa Eléctrica Regional Centro Sur cuenta con un Sistema de Información Geográfico desde principios de 2003, culminando el levantamiento de las redes en 2006, siendo la primera empresa de distribución eléctrica del país en contar con un Sistema de Información Geográfico corporativo. El SIG se convirtió en la empresa en una herramienta de apoyo muy importante para su personal técnico principalmente y se ha integrado con otros sistemas como son el SCADA, el sistema de calidad de producto, el sistema de análisis técnicos y el sistema comercial. Por otra parte, la Dirección de Distribución de la Empresa cuenta con el sistema de gestión de proyectos eléctricos, que es la herramienta principal para entregar la información administrativa como el estado actual, los avances físicos, los costos, fecha de construcción de los proyectos, para mencionar algunos.

6.2 Necesidad de la implementación.

Siempre ha existido la visión en la Dirección de Distribución de enfocar los sistemas no solo a la parte técnica, sino también a la parte del negocio en sí, es decir, que facilite información útil para la gestión en la Dirección de Distribución. Es así como se plantea el desarrollo de los procesos necesarios para la extracción y transformación de la información espacial del SIG y la información transaccional del SGP para ser cargada en un prototipo de Data Mart para demostrar al personal directivo, con datos propios de la empresa, las ventajas y potencialidad para la toma de decisiones que representa una integración entre el SIG y una herramienta de inteligencia de negocios como es el Data Warehouse.

6.3 Principales componentes

Los principales componentes del prototipo de Data Mart son tres y se enumera a continuación:

1. El Sistema de Gestión de Proyectos (SGP).
2. El Sistema de Información Geográfico con su herramienta ETL llamada “Data Interoperability”.
3. Herramientas de Inteligencia de negocios de Oracle.

6.3.1 El sistema de gestión de proyectos (SGP).

a. Descripción General

El sistema de gestión de proyectos es una aplicación que gestiona la información gerencial de los proyectos de electrificación en la Empresa Eléctrica Regional Centro Sur. En este sistema se mantiene la información histórica del estado de los proyectos, su fecha de construcción y sus costos para liquidación de los mismos.

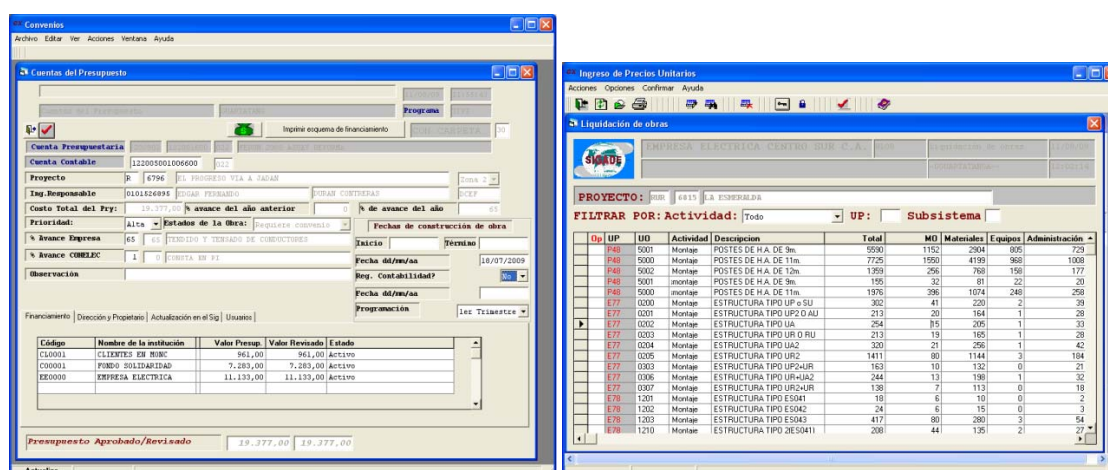


Gráfico 28. Pantallas del Sistema de Gestión de Proyectos.

b. Procedencia

El Sistema de Gestión de Proyectos (SGP) es un sistema desarrollado en la Empresa Eléctrica (in house). Fue desarrollado utilizando una herramienta de 4GL llamado GeneXus de la empresa Uruguaya llamada ArTech que fue fundada en 1988 y tiene su matriz en Montevideo.

c. Plataforma y arquitectura

El SGP es un sistema cliente servidor desarrollado en Visual Basic 6 con GeneXus 8 cuya información esta almacenada en un servidor AS400 con base de datos DB2.

d. Ámbito de servicio

Los módulos del SGP relacionados con el prototipo del Data Mart están instalados básicamente en la sección de gestión de proyectos del Departamento del Análisis y Sistemas Geográficos de Distribución (SIGADE) de la Dirección de Distribución de la Empresa.

e. Tiempo de operación

La primera versión del SGP funciona en la empresa desde el 1998 aproximadamente.

6.3.2 El Sistema de Información Geográfico.

a. Descripción General

El SIG en su concepto básico, es una herramienta para la edición de información georeferenciada y su análisis espacial.

El SIG es utilizado principalmente para llevar la información de toda infraestructura de las redes de distribución de la empresa, esto incluye información de transformadores, redes de media y baja tensión, postes, estructuras en postes y clientes.

Para realizar automáticamente los geoprocursos y las tareas de extracción, transformación y carga de la información geográfica se utilizó una extensión del Arcgis llamada “Data Interoperability”, que es la que permite implementar las actividades de los “Proceso de manejo de la información tabular y espacial en la etapa ETL para el Operacional Data Store (ODS)” de la sección 5.1.

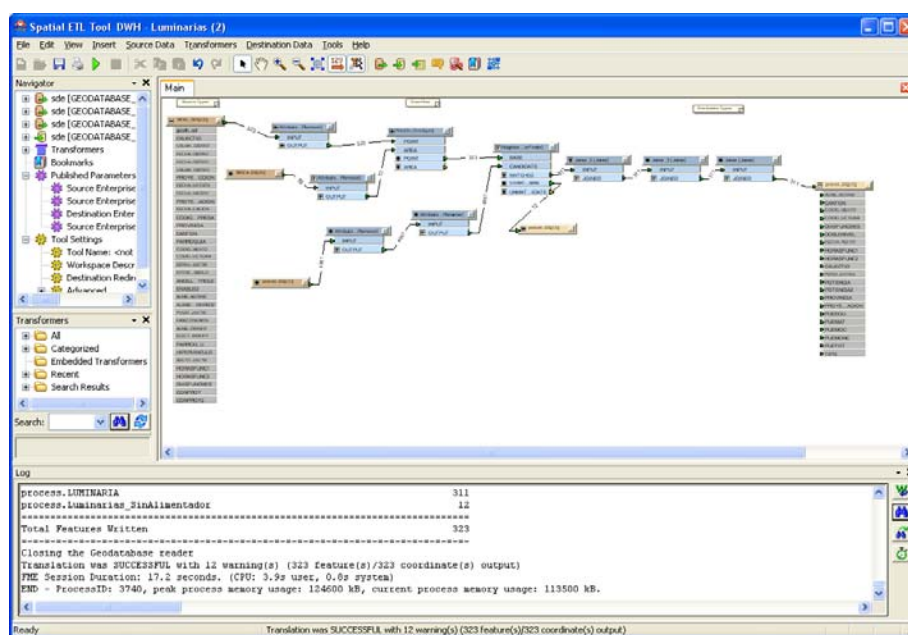


Gráfico 29. Pantalla del ArcGIS Data Interoperability. (Spatial ETL de Luminarias)

b. Procedencia

El Sistema de Información Geográfico de la Empresa Eléctrica es el ARCGIS, que al igual que “Data Interoperability” son de la marca ESRI (Environmental Systems Research Institute) fundada en 1969 en Redlands, California y es la empresa líder en el ámbito de los sistemas de información geográficos.

c. Plataforma y arquitectura

El SIG de la Empresa Eléctrica se encuentra en un servidor Risc, con sistema operativo AIX 5.2 en el cual está instalada la geodatabase que reside en una base de datos Oracle 8i, a la que acceden las estaciones de trabajo para la edición, consulta y análisis de la información georeferenciada. Para el caso del prototipo actual y como ambiente de pruebas se trabaja en un servidor Windows XP donde se encuentra instalado el ArcGis 9.3 y la base de Datos Oracle 11g.

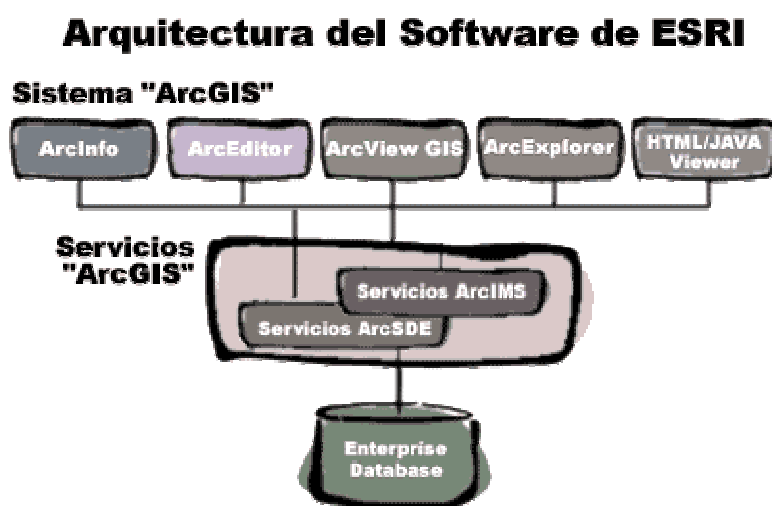


Gráfico 30. Arquitectura de software de ESRI.

Fuente: www.esri.com

d. Ámbito de servicio

Sus principales usuarios están en la Dirección de Distribución y corresponde a los operadores gráficos (encargados del mantenimiento del sistema); adicionalmente, en la Dirección de Comercialización existe una estación de trabajo para las actualizaciones de las nuevas instalaciones de medidores. En total el número de usuarios de este módulo son de 12 personas.

Adicionalmente para efectuar consultas de la información del SIG, el personal interno puede hacerlo a través de la aplicación llamada “SigCon”, mientras que personal externo a la empresa lo puede hacer a través de la página Web de la CENTROSUR: www.centrosur.com.ec.

La siguiente ilustración se muestra una pantalla del software SIG de la Empresa Eléctrica.

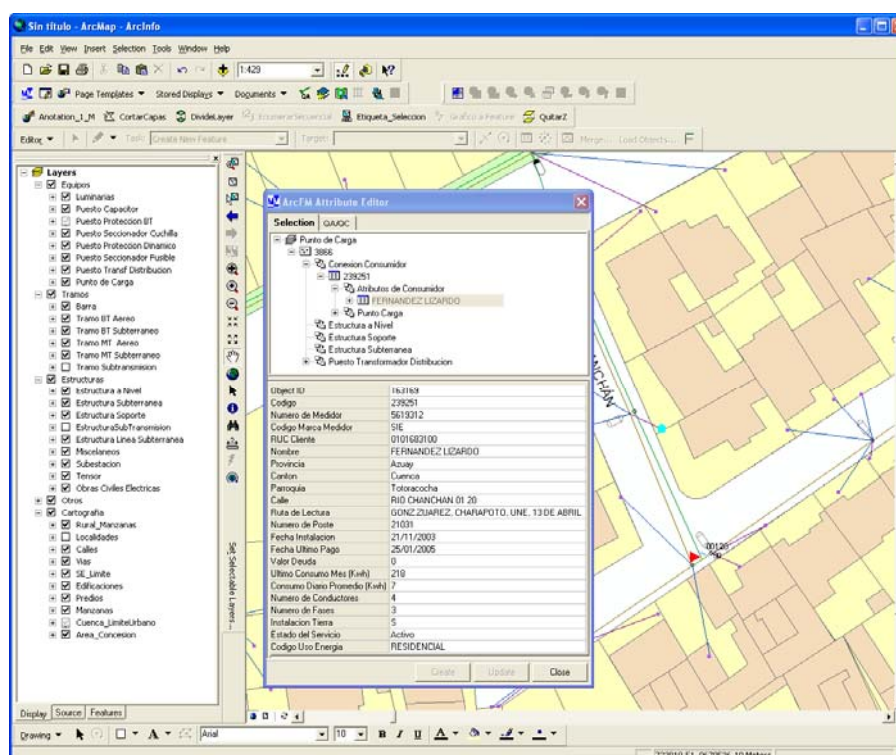


Gráfico 31. Pantalla de software SIG de EERCS.

e. *Tiempo de operación*

El SIG entró en funcionamiento desde principios de 2003, culminando el levantamiento de las redes en 2006.

6.3.3 Herramientas de Inteligencia de negocios de Oracle.

a. Descripción General

Básicamente se utilizan las siguientes herramientas que son propiedad de Oracle y que están orientadas a Inteligencia de Negocios:

- **Oracle Warehouse Builder**

Es una herramienta de Oracle para ETL (Extracción, Transformación y Carga), modelado de relaciones y dimensiones, manejo de datos y metadatos, que ofrece un ambiente gráfico para diseñar, construir, manejar y mantener los procesos de integración de datos en sistema de inteligencia de negocios.

Esta herramienta soporta una completa administración del ciclo de vida de la información del Data Mart o Data Warehouse:

- **Diseño:** Importación del modelo de datos empresarial, modelamiento gráfico del esquema multidimensional, y mapeo intuitivo de la fuente al destino.
- **Construcción:** Generación y poblado del Data Warehouse o Data Mart usando las características nativas de la base de datos Oracle.
- **Extracción, transformación, y carga de datos de diferentes fuentes** como base de datos relacionales, archivos planos, y fuentes ERP (enterprise resource planning), entre otros.
- **Integración:** Integración fácil con la Base de Datos de Oracle (Oracle Real Application Clusters, Oracle Online Analytical Processing, Oracle Spatial, Oracle Enterprise Manager, entre otros), Oracle E-Business Suite, y Oracle Business Intelligence

- **Mantenimiento:** Permite el mantenimiento de la bodega y la actualización de sus datos.

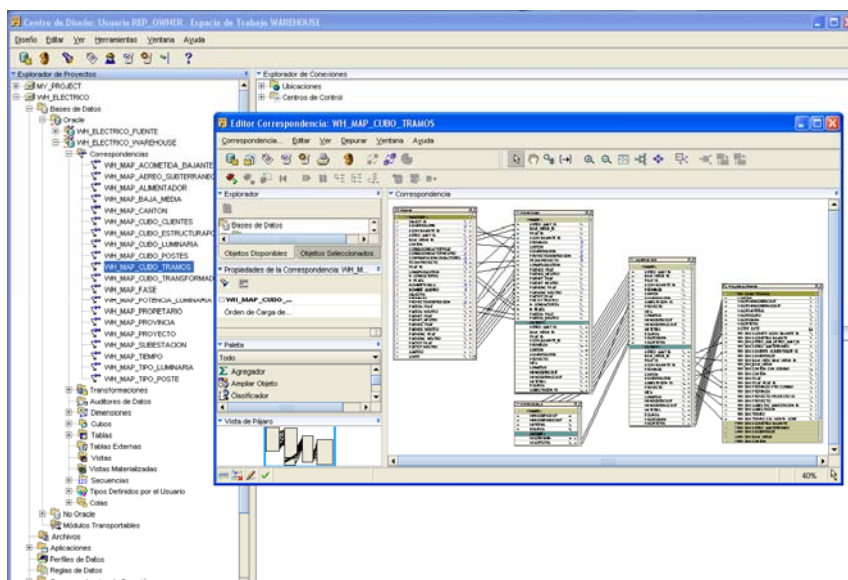


Gráfico 32. Diagrama de correspondencias de Oracle Warehouse Builder.

- **Oracle Business Intelligence Standard Edition One**

Son un conjunto de herramientas intuitivas para hacer consultas, reportes y análisis que dan al usuario acceso a la información en una base de datos relacional o multidimensional. Básicamente en el prototipo se utilizan tres de las herramientas que tiene Oracle Business Intelligence Standard Edition One y estas son:

- Oracle BI Administration Tool

Es donde se configura y mantiene la metada del Data Mart, aquí se controla el acceso a la información y se oculta la complejidad de los datos fuentes al usuario final. El Oracle BI Administrador permite:

- Importar los metadatos desde la base de datos y otras fuentes de datos.
- Simplifica y reorganiza los metadatos importados dentro del modelo de negocios
- Crea la presentación para el modelo de negocios, para que los usuarios puedan acceder a esta información como ellos la requieren, a través del Oracle BI Answer.

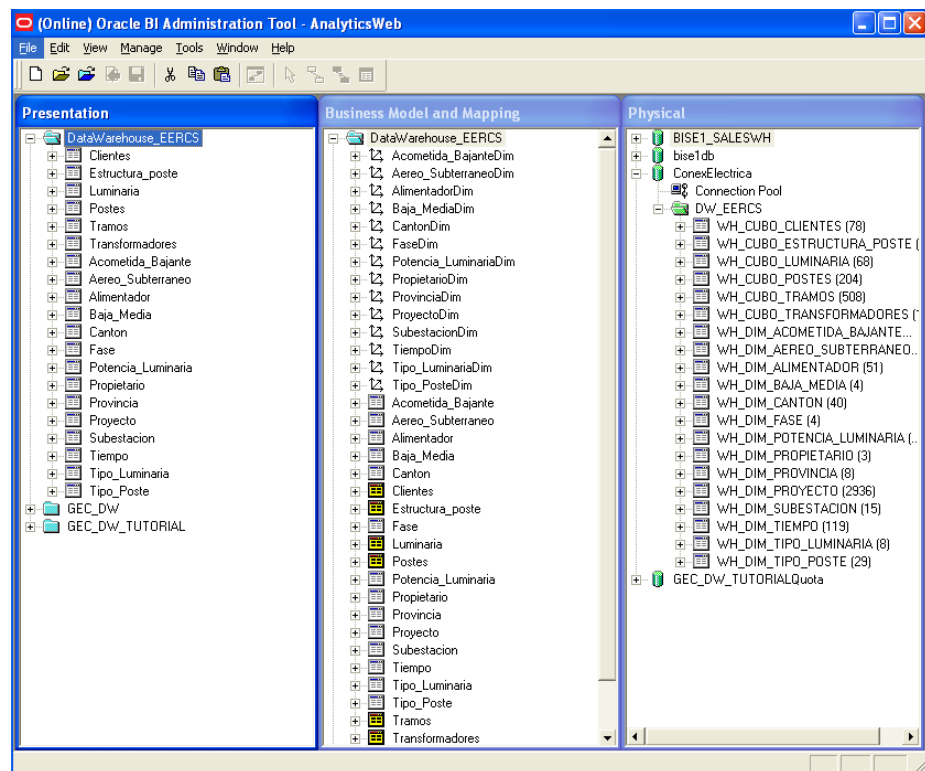


Gráfico 33. Oracle BI Administration.

- Oracle BI Answer

Es una aplicación en ambiente web que permite gracias a las funciones de consulta y análisis que los usuarios de negocio puedan crear nuevos análisis o modificar los existentes en los cuadros de mando sin tener que lidiar con estructuras de datos complejas.

Traduce los metadatos en conceptos comprensibles y los presenta en una vista lógica, de forma que los usuarios puedan combinar los datos procedentes de diversas fuentes de información sin necesidad de entender las relaciones complejas de las mismas.

Con esta aplicación los usuario pueden crear gráficos charts, reportes todo en un ambiente interactivo donde pueden ser grabados, compartidos, modificados, formateados, o empotrados en el Oracle BI Intelligence Dashboards personalizado por un usuario. Como resultado se tienen unos usuarios de negocios que pueden ser autosuficientes para obtener los reportes que necesita en un ambiente seguro y controlado.

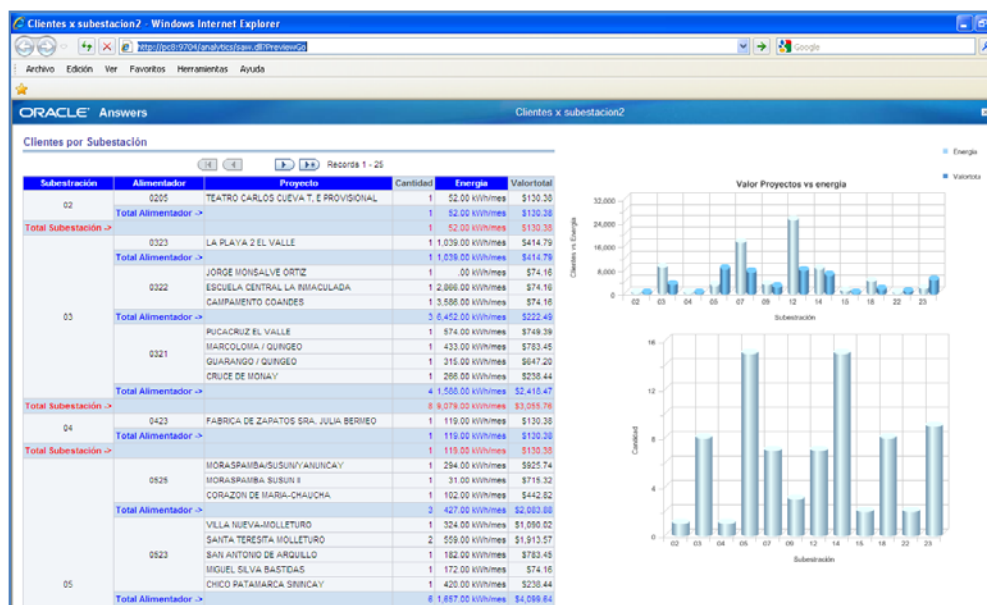


Gráfico 34. Pantalla de Oracle Answer.

Reporte de clientes por subestación.

- Oracle BI Interactive Dashboards

Esta aplicación, basada en una arquitectura de cliente web ligero, presenta un contenido analítico en diversos formatos interactivos. Los cuadros de mando se muestran en un navegador web y proporcionan a los usuarios información filtrada, personalizada y adaptada a su función o rol, tras aplicar una serie de reglas de seguridad predefinidas.

La interfaz de usuario es interactiva y presenta la información de forma intuitiva. Además la herramienta tiene sistemas de las alertas que permiten dar ágil respuesta y facilita que los usuarios tomen decisiones más acertadas y mejorar el rendimiento general de la Empresa.



Gráfico 35. Pantalla de Oracle Interactive Dashboard.

b. Procedencia

Con respecto a las herramientas de inteligencia de negocios de Oracle, estas no han sido adquiridas por la empresa, son herramientas que por contar con el mantenimiento de la base de datos de Oracle, nos fue fácil obtenerlas para su evaluación y el desarrollo del prototipo. Oracle Corporation es una de las mayores compañías de software del mundo. Sus productos van desde bases de datos (Oracle) hasta sistemas de gestión. Esta empresa surge a finales de 1970 con el nombre de Relational Software y es desde 1983 cuando cambia su nombre definitivo a Oracle Corporation.

c. Plataforma y arquitectura

El servidor de Data Mart y las herramientas de BI de Oracle utilizadas para este prototipo están instaladas en un servidor Intel con procesador core 2 duo de 2Ghz con 3GB de RAM, con sistema operativo Windows XP y en una base de Datos Oracle 10g.

d. Ámbito de servicio

A través del browser de internet el Director de Distribución y los jefes departamentales pueden acceder al Oracle BI Answer y al Oracle Interactive Dashboards.

Por otro lado, el Oracle Warehouse Builder y el Oracle BI Administration Tools estarán en el servidor de Data Mart que es administrado por el Superintendente Informático con el rol de Administrador del Data Warehouse.

e. Tiempo de operación

Al final el presente trabajo de tesis, la herramienta ha finalizado la fase de pruebas y entró en el periodo de operación hace unas pocas semanas.

6.4 Arquitectura del Sistema

A continuación se indican los diagramas de componentes y el diagrama de procesos del prototipo implementado.

6.4.1 Diagrama de componentes del Data Mart

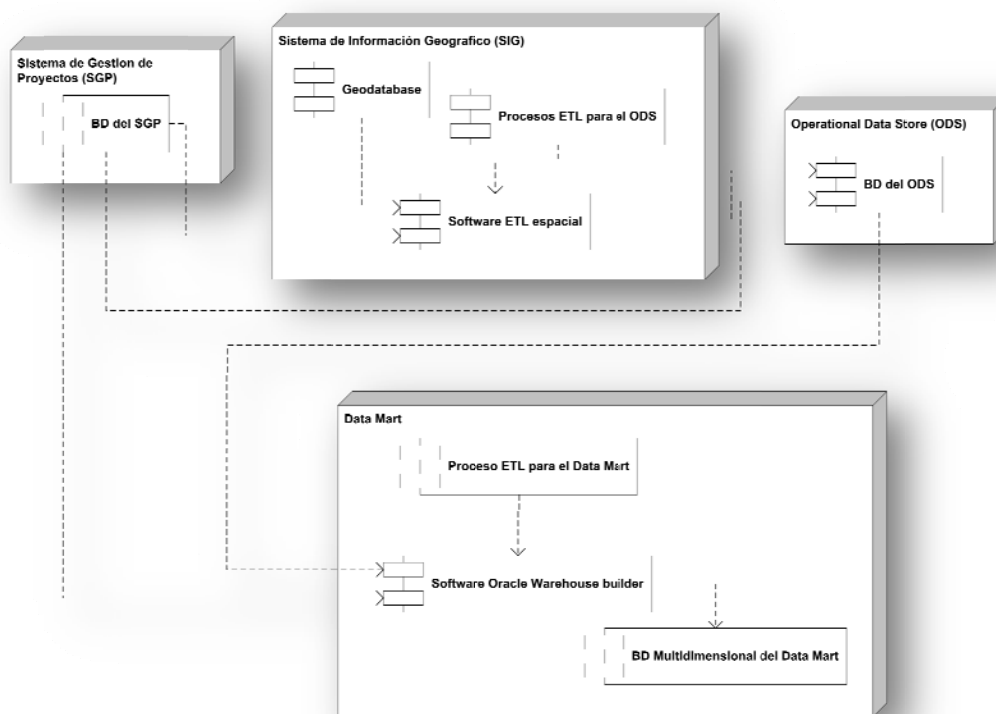


Gráfico 36. Diagrama de componentes del prototipo de Data Mart.

6.4.2 Diagrama de despliegue

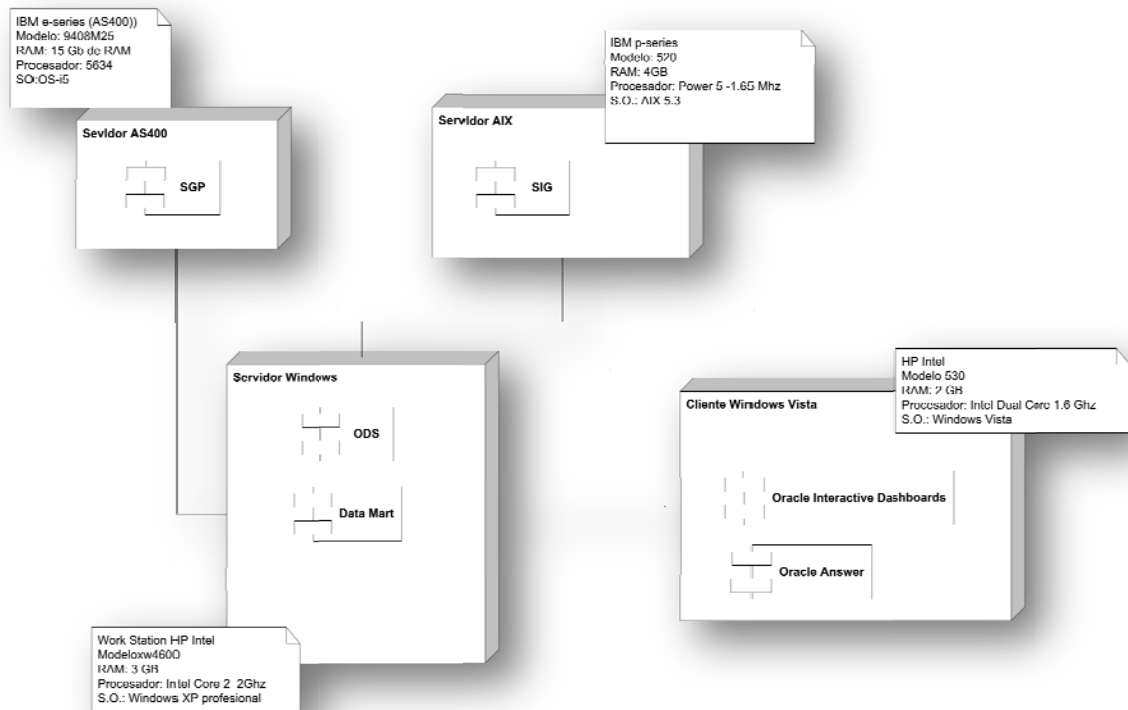


Gráfico 37. Diagrama de despliegue del prototipo de Data Mart.

6.4.3 Metadatos.

Los metadatos de este Data Mart los constituyen los diagramas de mapeo de las tablas de hecho, del diagrama de los mapeos de las dimensiones, así como la documentación de los procesos, que se encuentran en este documento en las secciones 4 y 5.

6.5 Implementación del Data Mart

La implementación de los Data Mart se puede realizar por áreas del negocio en una organización, los cuales se podrían consolidar en un solo Data Warehouse al final.

Lo que es importante tener en cuenta es que el análisis, el modelo y alcance deben ser generales para toda la organización, para poder definir dimensiones comunes entre los diferentes Data Marts, lo que evita que se realice un doble esfuerzo y se concientice en una unificación final, de otra manera cuando se intente implementar el Data Warehouse final que unifica los Data Mart no se podrá acoplar las diferentes áreas del negocio y el proyecto Data Warehouse final tendrá serias dificultades.

En la implementación de este Data Mart se definió procesos y fases basados en la aproximación metodológica sugerida por Bohorquez [6]. Con la puesta en práctica y conocimientos adquiridos se presenta aquí de forma sintetizada una sugerencia de los procesos y fases para la implementación de este sistema:

Procesos:

- **Definición de requisitos de la aplicación**, aquí se definen los aspectos iniciales del Data Mart como son el alcance, las áreas temáticas del mismo, los requisitos de información y se desarrolla un modelo conceptual que defina las estructuras multidimensionales y las fuentes de información espaciales y transaccionales.
- **Arquitectura tecnológica**, es donde se definen la arquitectura del Data Mart, los estándares a utilizar, las plataforma de desarrollo, pruebas y producción, se debe desarrollar las estrategias y planes de integración y calidad de datos, definir acuerdos sobre el nivel de servicio además de evaluar y seleccionar las herramientas GIS,ETL, ODS, Data Mart.

- **Operación de datos,** involucra la creación de los procesos y geoprosos de extracción, transformación y carga desde las fuentes de información, sean estas espaciales o no, establecer los criterios de aceptación e indicadores de excepciones de los datos. Establecer las políticas y normas de mantenimiento y limpieza de los datos geográficos y transaccionales en los sistemas OLTP.
- **Administración de los Metadatos,** este es un proceso que involucra todas las fases de implementación del Data Mart, donde se debería llevar un adecuado ingreso y actualización de los Metadatos técnicos, del negocio y los geográficos.
- **Diseño y construcción del Data Mart,** crear y validar los diseños lógicos y físicos, crear los diseños para el ODS, Data Mart y los metadatos, evaluar los diferentes tipos de roles y usuarios.
- **Documentación,** se debe disponer de manuales técnico y de usuario, claros y completos; así como documentación de procesos, referencias de los metadatos, entre otros.
- **Pruebas,** se debe probar los módulos que realizan los procesos ETL tanto a nivel geográfico (geoprosos) como transaccionales y realizar pruebas de requisitos funcionalidad y no funcionales del Data Mart.
- **Transición,** aquí se elabora el plan de instalación y capacitación, se adiestra al personal técnico en la administración y el mantenimiento del Data Mart, ODS y los geoprosos ETL, de igual manera, a los usuarios en el manejo de las herramientas de inteligencia de negocios para la creación y modificación de reportes y análisis, y se lanza el Data Mart a producción.
- **Soporte posterior a la implementación,** en este proceso se monitorea el uso del Data Mart y el rendimiento de los procesos y geoprosos, se corrigen problemas y se afinan las bases de datos para que mejoren el rendimiento.

Fases:

- **Estrategia**, esta fase se enfoca en los aspectos globales de la solución de inteligencia de negocios a nivel de la organización completa, se establecen las estrategias de operación de los datos y administración de metadatos.
- **Definición**, esta fase consiste en enmarcar en la practico el alcance del proyecto, identificar las áreas de negocio involucradas y sus requisitos, conducir una visión inicial de la arquitectura, definir factores críticos de éxito y riesgos, establecer los procesos, usuarios y políticas involucradas e identificar los datos fuente.
- **Análisis**, en esta fase se crean los modelos de datos lógicos, se analiza las necesidades de adquisición de datos, cuáles serán los procesos y geoprocesos ETL necesarios y adecuar los modelos de datos de los sistemas fuente.
- **Diseño**, consiste en diseñar procesos y geoprocesos ETL, verificar la integridad de los datos espaciales y transaccionales, definir especificaciones de acceso a los datos e integrarlas con criterios de acceso geográfico a los mismos, seleccionar y comprar las herramientas, diseñar el modelo físico del Data Mart y configuración de la plataforma, además de diseñar los procedimientos integración y administración del Data Mart.
- **Construcción**, implementar y evaluar los procesos ETL, impactar el diseño físico en la base de datos, instalar y configurar las herramientas de modelamiento del Data Warehouse (repositorio de la metadatos), de ETL y de explotación, elaborar de reportes estándar y dashboards, crear planes de pruebas, crear procesos de administración del Data Mart y manuales de técnicos y usuario.
- **Transición**, se debe validar los requisitos no funcionales, afinar la base de datos, realizar los planes de instalación, capacitación y lanzamiento del Data Mart.

- **Revisión y evaluación**, se debe evaluar el cumplimiento del proyecto, identificar desviaciones, documentarlas y plantear requisitos de nuevas mejoras que surgirán con el involucramiento estricto de los usuarios estratégicos del Data Mart.

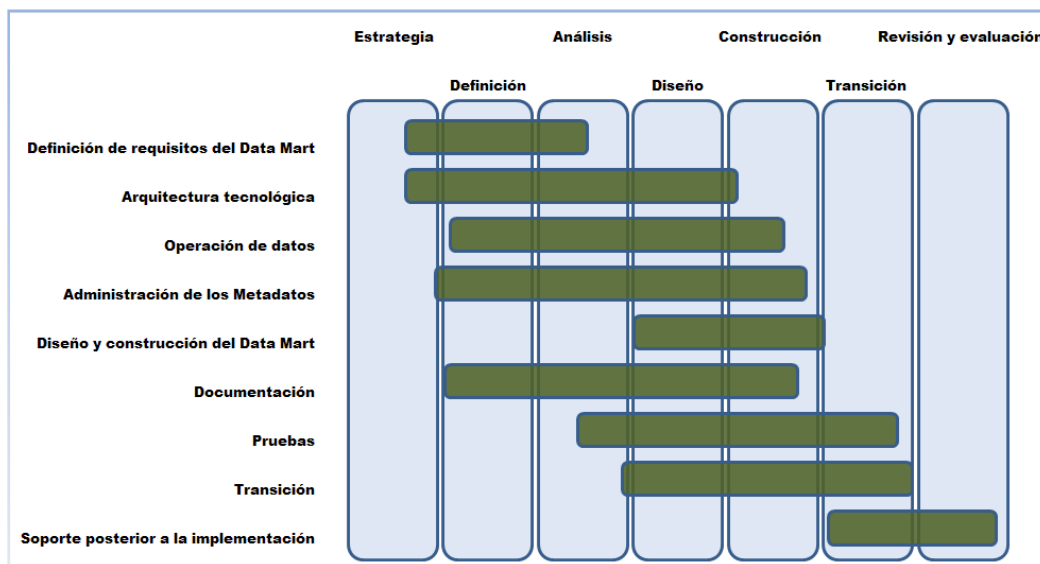


Gráfico 38. Fases y procesos de la implementación del prototipo de Data Mart.

Es importante mencionar que los procesos involucrados en las fases de Construcción, Transición y, Revisión y Evaluación no se aplicaron estrictamente a este proyecto, debido a que se trata de un prototipo, pero por otra lado los procesos de las fases de Estrategia, Definición, Análisis y Diseño sí fueron aplicados y su resultado está plasmado en el cuerpo del presente trabajo.

6.6 Beneficios

Los principales beneficio del proceso de manejo de la información aplicado al presente prototipo son:

- Se facilita al usuario el acceso ágil a la información resultante de geoprocesamiento, sin que esté obligado a conocer las herramientas SIG ni la forma en que se realizan estos análisis espaciales.
- Se facilita a través de las herramientas de explotación del prototipo, una información más ejecutiva, permitiendo a los usuarios hacer sus propias consultas.
- Ayuda a mejorar el funcionamiento de los sistemas operacionales (SIG y SGP) retroalimentándolos con los indicadores de excepciones y requerimientos de calidad de datos.
- La información compartida conduce a un lenguaje común y conocimiento común lo que mejora la comunicación en la empresa.

7. Factores críticos de éxito

Los principales factores críticos de éxito que se han encontrado del presente trabajo de tesis son los siguientes:

- El compromiso de todos los involucrados en el negocio, desde la alta gerencia hasta los operativos.
- Definir los requerimientos desde el punto de vista del negocio.
- Un completo conocimiento del área de negocio donde se aplicará el Data Mart.
- El adecuado y bien planificado establecimiento de procesos y geoprocesos espaciales, para la realización de la extracción, transformación y carga de los datos fuentes al Data Mart.
- Fomentar una cultura de entendimiento del alcance y complejidad de la información espacial.
- Tener la información del SIG y de los sistemas OLTP depurada y correcta.
- Una adecuada metodología de implementación.
- Utilización de estándares para levantamiento de requisitos, documentación de procesos y mejores prácticas para proyectos de software.

8. Hallazgos

Luego del presente trabajo de tesis se han encontrado los siguientes hallazgos:

1 Existen datos dentro del Sistema de Gestión de Proyectos (SGP) que deberían ser especificados como obligatorios, para evitar que los usuarios los dejen sin valores.

Los principales se mencionan continuación:

- El campo de fecha de energización de proyecto en la tabla CGP, donde solo el 12% de los registros de la tabla están llenos con este dato.
- En la tabla PUE del SGP no existen valoradas todas las estructuras que se manejan en el sistema, y para hacer valoraciones en ocasiones los usuarios toman códigos de estructuras similares. Aproximadamente un 3% de estructuras nuevas o poco utilizadas están sin valorar.

2 Existen datos dentro del Sistema de Información Geográfico que deberían ser especificados como obligatorios y validados contra el SGP, para evitar que los usuarios los dejen sin datos o cometan errores en el ingreso. Los principales se mencionan continuación:

- El código de estructura en las capas del SIG, tiene aproximadamente un 2% de datos nulos o no válidos.
- Los campos que almacenan el código del proyecto con el que se creó o modificó un elemento de la red en campo, tienen aproximadamente un 65% de datos nulos o no válidos.

3 La herramientas de inteligencia de negocios de Oracle son bastante poderosas, pero también complejas y costosas. En el caso del prototipo se realizó un análisis consiente pero no exhaustivo para la selección de la herramienta, la cual se

seleccionó con base en el hecho de que fue la más fácil de adquirir, más no necesariamente la más fácil de implementar.

- 4 ArcGis cuenta con una herramienta ETL llamada “Data Interoperability”, que permite, de forma ágil y fácil, la automatización de los geoprocesos ETL.

9. Conclusiones

A continuación se detallan algunas conclusiones del presente trabajo de tesis:

- Este trabajo aporta un beneficio en la operación estratégica de la Dirección de Distribución, por el hecho de que se simplifica la toma de decisiones ofreciendo a los usuarios estratégicos (Director y Jefes Departamentales) un acceso ágil a la información integrada de los proyectos de electrificación y sus costos con información espacial de las redes, en un solo ambiente de consultas y análisis que facilita el proceso de comparación, proyección o relación entre proyectos de electrificación.
- La correcta elaboración y documentación de los geoprocesos para la Base del ODS y los procesos ETL son importantes para una adecuada implementación del presente Data Mart.
- La calidad de la información espacial del SIG y la información en el sistema de gestión de proyectos (SGP), representa un aspecto que puede aportar una carga de tiempo tanto a los procesos ETL como a la culminación del proyecto como tal.
- El presente trabajo de tesis será un aporte importante al plan informático de la Empresa Eléctrica, donde uno de sus ejes es la implementación de un sistema de Inteligencia de Negocio que arrancará en 2010.
- Por último, y luego de los tropiezos y aciertos logrados a lo largo del presente trabajo de tesis, se puede concluir que un proyecto de integración de la información de un SIG y un sistema de Inteligencia de Negocios es un reto que involucra una correcta selección de los recursos informáticos a utilizar, al igual que un adecuado análisis, diseño y documentación de los procesos a implementar.

10.Recomendaciones

A continuación se detallan algunas recomendaciones luego de concluido el presente trabajo de tesis.

- Es importante realizar un plan de actualización de la información en el sistema de gestión de proyectos para mantener al día las fechas de energización de todos los proyectos y la valoración de las estructuras nuevas ingresadas al sistema.
- Es importante realizar un plan de actualización de la información en el Sistema de Información Geográfico para poner al día la información de los proyectos y las estructuras de cada elemento.
- Realizar un proceso de análisis, valoración y benchmarking de diferentes herramientas de administración y explotación de los datos del Data Mart.
- Si se pretende continuar operativamente con el Data Mart diseñado y pasar de un prototipo a una aplicación operación, se recomendable la selección y formalización de las licencias de software y hardware adecuadas.
- Se recomienda también en el campo de la investigación que:
 - Se realicen trabajos de comparación entre las dos herramientas de Inteligencia de Negocios que se visualizan como las más apropiadas para la empresa por el hecho de que se tiene experiencia en otros productos de la misma marca como son las de Oracle y Genexus.
 - Análisis de beneficios y productividad de las herramientas que han aparecido en el mercado para relacionar información de inteligencia de negocios con información espacial directamente en mapas en la aplicación del usuario.
 - Investigaciones de la creación de spatial OLAP (On Line Analytical Process)
 - Investigación en la factibilidad para nuestro medio de herramientas DSS-S (Decision Support System-Spatial).

- Es recomendable la incorporación progresiva de más indicadores en el Dashboard acordes con los nuevos requerimientos que surgirán conforme los usuarios de negocios vean sus ventajas y potencialidades.
- Como última recomendación respecto al próximo paso a dar en la implementación formal de un Data Warehouse para la Empresa Eléctrica es estrictamente necesario que se contrate una consultoría que se encargue del análisis y el modelo general de las dimensiones comunes y alcance de los Data Mart departamentales y enmarcarlos dentro de una estructura de datos estándar para toda la organización.

Bibliografía

[1] Francisco J. Moreno A. y Fernando Arango. (2007). *State of the Art of SpatioTemporal Multidimensional Models*.

[2] Diego Orlando Abril Fradel y José Nelson Pérez Castillo. (2007). *Current data warehousing and OLAP technologies. Status applied to spatial databases*.

[3] Carlos Andrés Herrera Parra. (2000). *Estructura de bodegas de datos para el almacenamiento de datos espaciales*.

[4] Ing. Carlos Andrés Herrera. (2001). *Spatial Hierarchy and OLAP Favored Search in Spatial Data Warehouse*.

[5] John William Cely y Yvan Bédard. (2005). *El paradigma multidimensional: Desarrollo de nuevas tecnologías para la gestión del territorio*.

[6] Juan Eulises Bohorquez. (2007). *Aproximación metodológica de un spatial Data Warehouse*.

[7] Universidad de Cuenca; Lucía Méndez Tapia. (2009). Maestría en Gerencia de Sistemas.

[8] Oracle. (2007). *Oracle Business Intelligence Estándar Edition One Tutorial*.

[9] Department of Computer Science; Dimitris Papadias, Panos Kalnis, Jun Zhang and Yufei Tao. Hong Kong University of Science and Technology. *Efficient OLAP Operations in Spatial Data Warehouses*.

Obtenido en línea el 10 de marzo de 2009. Disponible en:
<http://www3.ntu.edu.sg/home/jzhang/papers/sstd01.pdf>

[10] Faculty of Science; Taher Omran Ahmed. Aljabal Algharby University. *Continuous Spatial Data Warehousing*.

Obtenido en línea el 10 de marzo de 2009. Disponible en:
<http://eref.uqu.edu.sa/files/eref2/folder6/F180.pdf>

[11] Schlumbergersema, Graham Morgan, Gary Jurkowski, *Lessons In Geospatial Data Warehousing – A Proven Approach To Enterprise Data Integration*.

Obtenido en línea el 10 de marzo de 2009. Disponible en:

<http://www.google.com.ec/url?sa=t&source=web&cd=15&ved=0CDYQFjAEOAo&url=http%3A%2F%2Fciteeex.ist.psu.edu%2Fviewdoc%2Fdownload%3Fdoi%3D10.1.1.138.2476%26rep%3Drep1%26type%3Dpdf&rct=j&q=spatial%20data%20warehouse&ei=JEDETPacKcGp8Aajq7CsBg&usg=AFQjCNGS9pgVBRfUZRDjBk0q01B5LviiRA&cad=rja>

[12] Geer Services, Inc., William P. Geer, *Spatial Data Warehousing, GITA 2007 Annual Conference*.

Obtenido en línea el 11 de marzo de 2009. Disponible en:

http://www.mundogeo.com.br/revistas-interna.php?id_noticia=8306&lang_id=3

[13] Versailles-Saint-Quentin University, Lionel Savary, Karine Zeitouni, *Spatial Data Warehouse – A Prototype*.

Obtenido en línea el 11 de marzo de 2009. Disponible en:

<http://www.springerlink.com/content/hkygplb0nkt75v7d/>

[14] ESRI (March 1998), *White Paper, Spatial Data Warehousing*.

Obtenido en línea el 12 de marzo de 2009. Disponible en:

<http://spatialnews.geocomm.com/whitepapers/datawarehouse1.pdf>

[14] Ken Orr Institute; *Data Warehousing Technology*.

Obtenido en línea el 12 de marzo de 2009. Disponible en:

<http://www.kenorrinst.com/dwpaper.html>

[15] Karla Vanessa Barreto Stein. *Implementación incremental para data Warehouse*.

Obtenido en línea el 12 de marzo de 2009. Disponible en:

<http://www.gestiopolis.com/canales6/ger/data-warehouse.htm#mas-autor>

[16] Carmen Wolff. *Implementando un DataWarehouse*,

Obtenido en línea el 12 de marzo de 2009. Disponible en:

<http://www.inf.udec.cl/revista/edicion5/cwolff.htm>

Glosario

ACTUALIZACIÓN DE LA BASE DE DATOS.- Se trata de la fase final en la carga de datos. Es ahora cuando se trasladan los datos desde la tabla temporal (que se crea durante la fase de Carga de Datos) hasta su destino final en la base de datos. "

AIX.- (IBM) ADVANCE INTERACTIVE EXECUTIVE. Sistema operativo multiusuario y multitarea con posibilidad de trabajar en red y capacidad de gestionar procesos distribuidos.

ALIMENTACIÓN DE LA BASE DE DATOS.- Se trata de la información que el usuario proporciona para controlar la carga de datos a partir de uno o varios conjuntos de datos externos. La información queda almacenada y se puede volver a utilizar para actualizaciones de rutina.

CARGA DE LA BASE DE DATOS.- Se trata del módulo que se utiliza para añadir o actualizar registros en la base de clientes. Se puede añadir información desde un disco o cinta, o bien desde un equipo externo a través de una conexión en red. "

ANÁLISIS MULTI-DIMENSIONAL.- El objetivo del análisis multi-dimensional es que los usuarios finales lleguen a comprender lo que significan las bases de datos. El planteamiento multi-dimensional del análisis pone a un mismo nivel el contenido de los datos y el modelo concebido por el analista y, de este modo, evita confusiones y hace que disminuya el número de interpretaciones erróneas. También facilita la navegación por la base de datos, rastreando un subgrupo de datos concreto, solicitando datos con una orientación específica y definiendo cálculos analíticos. Además, puesto que los datos se encuentran físicamente almacenados en una estructura multi-dimensional, la velocidad en dichas operaciones es mucho más rápida y regular de lo que es posible con otras

estructuras de bases de datos. Una de las prestaciones más importantes del análisis multi-dimensional es dicha combinación de simplicidad y velocidad."

ANÁLISIS.- Operación que examina los datos con la intención de extraer o crear nuevos datos que cumplan ciertas condiciones o requerimientos

ASP (Active Server Pages).- Tecnología que puede utilizarse para crear aplicaciones web dinámicas e interactivas"

BASE DE DATOS.- conjunto de datos estructurado para permitir su almacenamiento, consulta y actualización en un sistema informático"

BD.- Base de datos.

CARGA DE LA BASE DE DATOS.- Se trata del módulo que se utiliza para añadir o actualizar registros en la base de clientes. Se puede añadir información desde un disco o cinta, o bien desde un equipo externo a través de una conexión en red.

CARTOGRAFÍA.- conjunto de técnicas utilizadas para la construcción de mapas.

CENTROSUR.- Nombre abreviado de la Empresa Eléctrica Regional Centro Sur.

CLASIFICACIÓN proceso de agrupamiento de un conjunto de elementos en clases en el sentido estadístico, una clasificación pretende agrupar los elementos en clases internamente homogéneas pero diferenciables entre ellas por los valores de una o varias variables"

CLIENTE OLAP.- Se trata de aplicaciones de usuario final que pueden solicitar fragmentos de un servidor OLAP y facilitar visualizaciones bi-dimensionales o multi-dimensionales, modificaciones de usuario, selecciones, ranking, cálculos, etc., para temas de visualización y navegación. Los clientes OLAP pueden ser tan simples como cuando un

usuario, conocedor de las hojas de cálculo, recupera un fragmento para ulteriores tareas a partir de un programa de hoja de cálculo, o bien tan sofisticado como en el caso de la aplicación de análisis de ventas o de un modelo financiero.

COORDENADA cantidad usada para definir una posición en un sistema de referencia las coordenadas pueden ser lineales (cartesianas) o angulares (esféricas), según el sistema de referencia.

DATA MART.- Un Data Mart es una versión especial de almacén de datos (Data Warehouse). Son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones. Los datos existentes en este contexto pueden ser agrupados, explorados y propagados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los mismos de la forma más conveniente según sus necesidades.

DATA MINING.- Ver ""Minería de Datos""."

DATA WAREHOUSE (DW).- Sistema de información que permite el almacenamiento en un único entorno de la información histórica e integrada proveniente de los distintos sistemas de la empresa"

DEFINICIÓN DE PROCESO.- Es la representación de un proceso de negocio de forma que admita manipulación y automatización. Está formado por un conjunto de actividades y relaciones entre ellas para indicar el comienzo y el fin del proceso."

DIGITALIZAR.- operación de codificar la información en cifras la digitalización se aplica habitualmente a la codificación de la información gráfica (mapas y planos convencionales) pero puede ser aplicada con propiedad a todo tipo de información para la construcción de bases de datos digitales.

DIMENSIONAL.- Se trata de un grupo de celdas de datos organizado según la dimensión de los datos. Por ejemplo, una hoja de cálculo ilustra una matriz bi-dimensional con sus celdas de datos organizadas en filas y columnas, cada una de las cuales representa una dimensión. Se puede visualizar una matriz tridimensional en forma de cubo y cada dimensión forma uno de los lados del cubo, inclusive cualquier fragmento paralelo ha dicho lado. Las matrices con mayor número de dimensiones no tienen traducción física, pero organizan los datos en la forma que los usuarios conciben su empresa. Las dimensiones corporativas más habituales son: tiempo, medidas, productos, zonas geográficas, canales de venta, etc.

DISOLUCIÓN (GEOMÉTRICA).- proceso de unión de dos o más polígonos mediante la eliminación de los lados comunes

DM.- Data Mart

DOBLE NIVEL.- Son luminarias que funcionan con dos niveles de potencia que pueden ser programados.

DRILL DOWN/UP.- Llevar a cabo un rastreo minucioso o superficial consiste en una técnica de análisis concreta con la cual el usuario rastrea los diferentes niveles de datos desde la forma más resumida (superficial) a la más detallada (minucioso). Se pueden definir rutas de rastreo a partir de las jerarquías que hay en las dimensiones u otras relaciones que puedan ser dinámicas con o entre dimensiones. Por ejemplo, al visualizar los datos sobre ventas en Norteamérica, si se solicita una operación de rastreo minucioso en la dimensión de Regiones, aparecerá Canadá, los Estados Unidos del Este y los Estados Unidos del Oeste. Si se solicita un rastreo aún más minucioso sobre Canadá, puede que muestre Toronto, Vancouver, Montreal, etc. "

DW.- Data Warehouse

EERCS.- Nombre abreviado de la Empresa Eléctrica Regional Centro Sur.

ERROR.- diferencia entre el valor medido o estimado y el valor real

GENERALIZACIÓN simplificación realizada al representar un objeto real mediante un modelo se aplica también al proceso de eliminación selectiva de vértices en una forma geométrica para simplificarla.

GEORREFERENCIAR.- asignar coordenadas geográficas a un objeto o estructura

GESTIÓN DEL CONOCIMIENTO.- Crear, adquirir, retener, mantener, utilizar y procesar el conocimiento antiguo y nuevo ante la complejidad de los cambios del entorno para poder poner al alcance de cada empleado la información que necesita en el momento preciso para que su actividad sea efectiva."

INFORMACIÓN.- Datos procesados u organizados.

INTEROPERABILIDAD.- Funcionamiento conjunto de dos o más programas"

LENGUAJE DE CONSULTA MULTI-DIMENSIONAL.- Se trata de un lenguaje informático que permite especificar cuáles son los datos a recuperar de un cubo. En general, al proceso de usuario para este tipo de consulta se le llama fragmentar en cortes y cubos (slicing and dicing). El resultado de una consulta multi-dimensional es una celda, un fragmento bi-dimensional o un sub-cubo multi-dimensional."

LÍNEA conjunto ordenado de vectores encadenados en el modelo de datos vectorial la línea se usa para representar objetos geográficos como carreteras, tendidos eléctricos, etc. En una estructura topológica, las líneas tienen un sentido y están definidos los lados izquierdo y derecho.

MAPA.- modelo gráfico de la superficie terrestre donde se representan objetos espaciales y sus propiedades métricas, topológicas y atributivas un mapa puede ser analógico (impreso sobre papel, por ejemplo) o digital (codificado en cifras, almacenado en un ordenador y presentado en una pantalla) existen mapas métricos, diseñados para representar distancias, superficies o ángulos y mapa topológicos, diseñados para representar vecindad, inclusión, conectividad y orden.

MEJORES PRACTICAS.- Ejemplos y casos que ilustran cuál es la manera más adecuada de llevar a cabo un proceso.

METADATOS.- información sobre las características de un conjunto de datos típicamente, los metadatos incluyen información anexa al cuerpo de datos principal (por ejemplo, un modelo digital de elevaciones) sobre extensión geográfica, estadísticas, autoría, metodología, calidad de la información, etc.

MINERÍA DE DATOS.- Tecnología cuyo objetivo es descubrir conocimiento oculto en un conjunto de datos y expresarlo de forma inteligible.

MODELO DE WORKFLOW.- Visión formal de un proceso de negocio, representado como una sucesión ordenada de actividades y procesos, que están conectados unos con otros con el fin de lograr un objetivo común.

MODELO representación simplificada de un objeto o proceso en la que se representan algunas de sus propiedades un modelo reproduce solamente algunas propiedades del objeto o sistema original que queda, por tanto, representado por otro objeto o sistema de menor complejidad; los modelos se construyen para conocer o predecir propiedades del objeto real.

NIVEL JERÁRQUICO.- Los miembros de una dimensión jerarquizada comparten nivel si, dentro de su jerarquía, tienen el mismo número máximo de descendientes en cualquiera de las rutas que se encuentran por debajo de ellos. Por ejemplo, dentro de la dimensión Cuentas que consta de libros de contabilidad general, todas las cuentas detalladas corresponden a miembros del Nivel 0. Las cuentas de un nivel superior son del Nivel 1, sus padres son del Nivel 2, etc. También puede suceder que unos padres tengan dos o más hijos con niveles diferentes, en cuyo caso el nivel de los padres queda definido en un nivel por encima del nivel del hijo con mayor nivel.

ODS.- Operational Data Store, Base de Datos Operacional, es una base de datos diseñada para integrar datos de múltiples fuentes para hacer más fácil el análisis y la presentación de informes. Debido a que los datos origen vienen de diferentes fuentes, la integración implica a menudo la limpieza, la solución de redundancia y comprobación de reglas de negocio para integridad. El ODS es generalmente diseñado para contener datos de bajo nivel o indivisible (como las transacciones y los precios) con una ventana de actualización muy pequeña en contraposición con el mayor volumen de datos almacenados en el Data Warehouse que generalmente es con menor frecuencia. En otras palabras, es un contenedor de datos activos, es decir operacionales que ayudan al soporte de decisiones y a la operación y está entre un OLAP y un OLTP.

POLÍGONO figura geométrica plana formada por, al menos, un anillo externo un polígono puede tener anillo(s) interno(s) en cuyo caso se habla de un polígono compuesto en vez de un polígono simple (sin "agujeros").

PROYECCIÓN conjunto de transformaciones métricas definidas para representar la superficie de la Tierra sobre un plano. Existe un gran número de proyecciones, cada una de las cuales posee propiedades diferentes en cuanto a las métricas del objeto real y de su

representación plana; por ejemplo, en una proyección conforme se conservan los ángulos (los paralelos y meridianos se cortan en ángulo recto) y en una equivalente se conservan las superficies.

PROYECTO MODIFICACIÓN.- Es el código de la carpeta a la que pertenecen los elementos de un proyecto de electrificación.

RASTER.- modelo de datos en el que la realidad se representa mediante celdas elementales que forman un mosaico regular cada celda del mosaico es una unidad de superficie que recoge el valor medio de la variable representada (altitud, reflectancia ...); las celdas pueden ser cuadradas (celdas) o no (triangulares, hexagonales...) un modelo de datos raster está basado en localizaciones.

RED modelo de datos formado por nodos y conexiones entre ellos tanto los nodos como las conexiones pueden tener atributos propios como, por ejemplo, longitud, resistencia, sinuosidad... El análisis de redes agrupa un conjunto de técnicas implicadas en la resolución de cuestiones que pueden ser modelizadas mediante una red, por ejemplo, determinación del camino de mínimo coste entre dos puntos.

RELACIONES JERÁRQUICAS.- Los miembros de cualquier dimensión se pueden organizar con base a relaciones de padres-hijos, siempre que los miembros padres representen la consolidación de los miembros que son sus hijos. Como resultado surge una jerarquía, y las relaciones padres/hijos son relaciones jerárquicas.

SGP.- Sistema de Gestión de Proyectos.

SIG.- Ver Sistema de Información Geográfico.

SISTEMA DE COORDENADAS.- marco de referencia espacial que permite la definición de localizaciones mediante coordenadas éstas pueden ser lineales (sistemas cartesianos, con ejes ortogonales) o esféricas (donde se utilizan como coordenadas el acimut y elevación angular).

SISTEMA DE GESTIÓN DE BASES DE DATOS (SGBD) sistema informático diseñado para la creación, modificación, corrección, actualización y consulta de bases de datos.

SISTEMA DE INFORMACIÓN GEOGRÁFICO.- sistema de gestión de bases de datos (SGBD) con herramientas específicas para el manejo de información espacial y sus propiedades los tipos de propiedades que un SIG debe poder analizar tanto independiente como conjuntamente son tres: métricas, topológicas y atributivas

SQL.- acrónimo de structured query language, un lenguaje estándar de gestión de bases de datos. SQL se ha convertido en un estándar por lo que es posible acceder a bases de datos de procedencia diversa mediante consultas en este lenguaje"

TABLA.- Se trata de una estructura en la cual se puede almacenar la información en la base de datos. Las tablas de datos se almacenan en filas y en columnas. Ver también code table; tabla de códigos.

TABLA DE CORRECCIÓN.- Se trata de una tabla de códigos a definir para luego utilizarla en la carga de datos. Esta tabla explica cómo cambiar valores concretos en un campo con registros del cliente y así corregir errores o bien estandarizar la información.

TIC.- Tecnologías de la Información y de las Comunicaciones."

TOPOLOGÍA.- referencia a las propiedades no métricas de un mapa en el contexto de los SIG, topología hace referencia a las propiedades de vecindad o adyacencia, inclusión,

conectividad y orden, es decir, propiedades no métricas y que permanecen invariables ante cambios morfológicos, de escala o de proyección se dice que una estructura de datos es ‘topológica’ cuando incluye información explícita sobre estas propiedades; en este caso, es posible realizar análisis y consultas “topológicas” sin necesidad de acudir a las tablas de coordenadas.

VALIDACIÓN.- proceso de comprobación de que datos y métodos responden a un estándar por ejemplo, la comprobación de que los códigos municipales de una base de datos se corresponden son coherentes con la codificación de referencia (por ejemplo, del

Instituto Nacional de Estadística

VECTOR.- entidad geométrica definida por una magnitud y un sentido un vector está formado por un par de puntos ordenados; el orden define el sentido del vector y la distancia entre origen y final su magnitud; si la magnitud es nula, el vector se reduce a un punto y el sentido queda indefinido.

VECTORIAL.- modelo de datos en el que la realidad se representa mediante vectores o estructuras de vectores una estructura vectorial puede ser compleja: una cadena de vectores forma un arco; una cadena de arcos forma un anillo; uno o varios anillos definen un polígono se trata de un modelo de datos basado en objetos (geométricos) frente al modelo raster, basado en localizaciones.

WORKFLOW.- Automatización de un proceso de negocio, en su totalidad o en parte, en el que intervienen diferentes participantes entre los que circulan tareas, información, documentos y datos, de acuerdo a unas reglas preestablecidas para ello.