



INGENIERÍA INFORMÁTICA
FACULTAD: GEOLOGÍA Y MINAS

TRABAJO DE DIPLOMA

PARA OPTAR POR EL TÍTULO DE

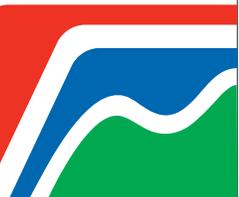
INGENIERO INFORMÁTICO

**TÍTULO: BASE DE DATOS DE LOS INDICADORES
DE LOS GRUPOS ELECTRÓGENOS DIESEL DE MOA**

AUTOR: ELBER BOSQUEZ ALMAGUER

**TUTORES: DRA. YIEZENIA ROSARIO FERRER
MSc. ALINA NOA RODRÍGUEZ
ING. ALEXEI CALA HINOJOSA**

MOA, 2013
“AÑO 55 DE LA REVOLUCIÓN”



Declaración de Autoría

Declaro que soy el único autor de este trabajo y autorizo al Instituto Superior Minero Metalúrgico de Moa “Dr. Antonio Núñez Jiménez” y al Departamento de Informática para que hagan el uso que estimen pertinente con este trabajo.

Para que así conste firmamos la presente a los ____ días del mes de _____ de 2013.

Elber Bosquez Almaguer

Firma Autor

Dra. Yiezenia Rosario Ferrer

Firma Tutor

MsC. Aliniuska Noa Rodriguez

Firma Tutor

Ing. Alexei Cala Hinojosa

Firma Tutor

Pensamiento



“Todo lo que es importante para la Revolución hay que chequearlo todos los días, la Revolución no pide disciplina, tiene derecho a exigirla, con la Revolución no se firman contratos, se firman compromisos. “

Fidel Castro Ruz.

Dedicatoria

Este trabajo va dedicado especialmente a mi madre, Flora, por haberme apoyado incondicionalmente en los momentos difíciles y ser la luz de mi vida, a mi padre, Bosquez, porque me enseñó a luchar, a no rendirme antes las dificultades y por ser mi ídolo.

A mi hermana tita que es la gordita que más amo en la vida.

A mi hermana pomba por soportarme las malacrianzas y quererme tanto.

A mi hermano tito por ser mi segundo padre y una de las personas en quien más confío.

A mis sobrinas Leyané y Leidis por ser las niñas de mis ojos.

A mis tías y tíos que me demuestran mucho cariño.

A mi amigo Osmani a quien quiero como mi otro hermano.

Agradecimientos

Primeramente quisiera agradecer a las razones de mi vida, a mis padres, que sin ellos este sueño no se hubiera hecho realidad.

A mis hermanos por todo el apoyo incondicional que me brindaron.

A mi familia en general por quererme tanto

A mis amigos de la infancia Osmani y Paco por demostrarme que los verdaderos amigos se prueban en los momentos difíciles y en esos momentos ellos me supieron extender su mano

A mis hermanos Guanchi y Rolandito por soportarme toda su vida.

A los amigos del barrio que compartimos juntos en las buenas y malas.

A mi grupo que me han ayudado en estos cinco años, sin duda son los mejores.

A mi tutora Yiezenia por su dedicación y entrega.

A mi tutora Alina por darme la oportunidad de participar en este proyecto.

A mi tutor Cala por su ayuda incondicional.

A Tony y Susi por su contribución en mi tesis.

Resumen

La Revolución Energética ha sido de gran significado para nuestro país. Los grupos electrógenos forman parte de esta, contribuyendo al ahorro energético del territorio nacional. Estos son controlados diariamente y en este proceso generan una gran cantidad de información, la cual es guardada en una hoja electrónica de cálculos Excel que presenta deficiencias.

Debido a la importancia que tiene esta información es necesario construir una base de datos para la explotación eficaz de los grupos electrógenos ya que ayudará en la toma de decisiones que se realicen con respecto a estos, apoyándose en datos confiables y seguros que serán almacenados en dicha base de datos.

Abstract

The energy revolution has been of great significance to our country by the importance. The generators are part of this, contributing to energy savings in the country. These are monitored daily, and in the process generates a large amount of information which is stored in a calculation Excel spreadsheet presents certain problems.

Due to the importance of this information is necessary to build a database for efficient operation of generators as it will help in making decisions that are made regarding these, based on reliable and secure data that will be stored in that database.

Índice

Declaración de Autoría	I
Pensamiento	II
Dedicatoria.....	III
Agradecimientos	IV
Introducción.....	1
CAPÍTULO 1. MARCO TEÓRICO-CONTEXTUAL DE LA INVESTIGACIÓN.....	8
1.1 Introducción.....	8
1.2 Antecedentes de la investigación.	8
1.3 Descripción del sistema de información de los indicadores de la Batería de Grupos Electrógenos Diésel.	12
1.3.1 Sistema de información para la gestión de los indicadores de la Batería de Grupos Electrógenos Diésel.....	12
1.3.2 Sistema de gestión de los indicadores de la Batería de Grupos Electrógenos Diésel.	14
1.3.3- Flujo de información del sistema de gestión de los indicadores de la Batería de Grupos Electrógenos Diésel.	18
1.4 Herramientas.....	20
1.4.1. Sistemas Gestores de Base de Datos (SGBD)	20
1.4.1.1 PostgreSQL.....	21
1.4.1.2 MySQL	22
1.4.1.3 SQL Server.....	24
1.4.2 Embarcadero ER/Studio.	25
1.4.3 Pentaho	26
1.4.4 Kettle de Pentaho (Spoon).....	26
1.5 Metodologías de desarrollo del software	27
1.5.1 Metodología MetdisBD.....	27
1.5.2 Metodología Hefesto.....	28
1.6 Metodologías y Herramientas a utilizar.....	30
1.7 Conclusiones del Capítulo.....	31
CAPÍTULO 2. DISEÑO Y POBLACIÓN DE LA BASE DE DATOS	32

2.1	Introducción.....	32
2.2	Descripción de los datos de los Grupos Eléctrogenos Diésel en Batería.....	32
2.3	Diseño de la Base de Datos.....	36
2.3.1	Aplicación de la <i>Metodología MetdisBD</i>	37
2.3.1.1	Determinación de entidades y atributos.....	37
2.3.1.2	Normalización de entidades.....	38
2.3.1.3	Determinación de relaciones (DER).....	42
2.3.1.4	Obtención del modelo lógico global de los datos.....	43
2.3.1.5	Diseño Físico de la Base de Datos.....	46
2.4	Poblado de la Base de Datos.....	47
2.4.1	Arquitectura de la solución propuesta.....	47
2.4.2	Procesos de extracción, transformación y carga.....	48
2.4.2.1	Proceso de extracción.....	48
2.4.2.2	Proceso de transformación.....	49
2.4.2.3	Transformaciones y cargas realizadas con SPOON.....	55
2.4.3	Trabajos.....	58
2.5	Estructura física de la base de datos.....	59
2.6	Conclusiones del Capítulo.....	61
	CAPÍTULO 3. ESTUDIO DE FACTIBILIDAD.....	62
3.1	Introducción.....	62
3.2	Evaluación Costo-Beneficio.....	62
3.3	Efectos Económicos.....	63
3.4	Elementos para identificar los Costos y Beneficios del Proyecto.....	64
3.5	Factibilidad Económica.....	65
3.6	Evaluación Económica.....	65
3.6.1	Beneficios tangibles.....	68
3.6.2	Beneficios y Costos Intangibles en el proyecto.....	69
3.7	Conclusiones del Capítulo.....	72
	Conclusiones Generales.....	73
	Recomendaciones.....	74
	Bibliografía.....	75

Anexos 1.X

Índice de Tablas

Tabla 1.1 Datos de la capacidad instalada de los grupos electrógenos	8
Tabla 2.1 Informe de Indicadores principales de Moa	33
Tabla 2.2 Informe de Indicadores principales de Moa	35
Tabla 2.3 Transformación para cargar la tabla TOTALES1.....	49
Tabla 2.4 Transformación para cargar la tabla TOTALES 2.....	51
Tabla 2.5 Transformación para cargar la tabla Tipo_emplazamiento	52
Tabla 2.6 Transformación para cargar la tabla Grupo_ electrógeno	52
Tabla 2.7 Transformación para cargar la tabla Consumo_aceite_refrigerante	54
Tabla 2.8 Transformación para cargar la tabla Combustible.....	54
Tabla 2.9 Transformación para cargar la tabla Localidad.....	55
Tabla 3.1 Datos de los costos en moneda libremente convertible	66
Tabla 3.2 Datos de los costos en moneda nacional	67

Índice de Figuras

Figura. 1.1 Diagrama en bloque de operación de la Batería de GE	13
Figura. 1.2 Diagrama del flujo de información de los GED.....	18
Figura 2.1 Diagrama de la determinación de relaciones.....	43
Figura 2.2 Diagrama del diseño físico de la base de datos	47
Figura 2.3 Diagrama de la arquitectura para poblar la base de datos	48
Figura 2.4 Diagrama de transformación para cargar la tabla2.3 TOTALES 1.....	56
Figura 2.5 Ventana de entrada de Excel para la obtención de los datos de entrada	57
Figura 2.6 Ventana pre-visualización de las filas de la HECE	57
Figura 2.7 Ventana de mapeo de campos	58
Figura 2.8 Diagrama de trabajo	59
Figura 2.9 Estructura física de la base de datos en SQLServer (2005-2008)	60
Figura 3.1 Gráfico de barra de las soluciones con la HECE y con la BD	70
Figura 1. Esquema del protocolo MODBUS del SCADA "Eros"	X

Introducción

Las Tecnologías de la Informática y las Comunicaciones (TIC) estudian el tratamiento automatizado de la información de una manera útil y oportuna. Estas son empleadas en los diferentes sectores existentes en la economía y la sociedad, porque nos permite almacenar y procesar una gran cantidad de información.

Nuestro país ha recibido los beneficios que proporcionan estas ciencias y las ha utilizado a su favor en la economía, la salud, el deporte, en la generación de energía eléctrica, etc.

En Cuba, el déficit de reserva de generación en las centrales eléctricas (5%) [Arrastía, 2008] en el 2005, condicionó la toma de decisiones con respecto a cómo suplir la demanda de energía en el sector residencial. La propuesta para solucionar este problema fue invertir en las tecnologías de Generación Distribuida (GD) con el fin de garantizar el suministro de energía eléctrica a la población. Estrategia que se implementó con la Revolución Energética en el 2006.

La Revolución Energética significó un acelerado despegue en la aplicación de las tecnologías energéticas renovables y no renovables. El Ministerio de Energía y Minas en coordinación con el Ministerio de la Industria Sideromecánica (SIME) controlaron e implementaron 28 programas para el desarrollo de la generación de electricidad basada en: la energía eólica, el empleo de la energía solar para el calentamiento de agua de uso doméstico, social y con fines industriales, el desarrollo de la capacidad hidroeléctrica y el aprovechamiento de los desechos sólidos con fines de valorización energética, así como la investigación sobre la energía geotérmica, oceánica y los grupos electrógenos diésel y fuel oil, constataron los avances del país en la inclusión de tecnologías energéticas renovables y no renovables concebidas como tecnologías usadas para la GD. El modelo de Generación Distribuida que se refiere a: "La generación o el almacenamiento de energía eléctrica a pequeña escala, lo más cercana al centro de carga, con la opción de interactuar con la red eléctrica, y en algunos casos, considerando la máxima eficiencia energética " [Bisset, 2010].

La tecnología introducida de GD con mayor capacidad en el país la constituyen los Grupos Electrógenos (GE) que queman combustibles fósiles (diésel o fuel oil; respectivamente GED y GEFO) [ONE, 2012]. Estos sistemas actualmente solucionan el problema de déficit de generación, pero encarecen la economía del país y el medioambiente. Los grupos electrógenos alcanzan una potencia instalada de 2 395,6 GW; el 1 272,5 GW en grupos de diésel y el 1 123,1 GW en grupos de fuel oil.

En Cuba en el período enero 2011-diciembre 2012 se ha incrementado el número de grupos de fuel oil en 250, en tanto disminuyen los grupos de diésel en 30. Las provincias que registran mayor capacidad instalada son: La Habana con 269,7 GW, Holguín con 261,5 GW y Artemisa 248,2 GW. En Holguín se han incrementado el número de grupos de fuel oil a 24, en tanto disminuyen los grupos de diésel en 3, en igual período. Se debe destacar que esta provincia es la segunda con mayor cantidad de GE instalados, después de La Habana [ONE, 2009 y 2012].

Los Grupos Electrógenos son: "Equipos formados por un motor primario de combustión interna y un generador sincrónico de corriente alterna acoplado mecánicamente para producir energía eléctrica. Para su funcionamiento puede consumir fuel oil, diésel o gas natural. Estos dispositivos pueden estar sincronizados al Sistema Electroenergético Nacional (SEN) para solucionar el déficit de potencia y contingencias aisladas (pertenecientes a la Unión Eléctrica [UNE]), para suministrar energía en lugares donde no llegue la red eléctrica nacional o de emergencia, ubicados en una entidad para operar en caso de fallo, desconexión o insuficiencias del fluido eléctrico en la red nacional" [ONE 2009].

Los grupos electrógenos desde que empezaron a explotarse usaron sistemas informáticos que han contribuido con la operación eficaz y eficiente de las unidades. Para ello se han realizado varias aplicaciones informáticas como software, páginas web, bases de datos y eventos internacionales donde se promueven estos. Uno de los más significativos, ha sido el Simposio Internacional de Generación Distribuida, dirigido y financiado por la CUJAE (Instituto Superior Politécnico José Antonio Hechavarría), en La Habana.

En el Simposio Internacional de Generación Distribuida en el 2010 se publicaron varios artículos que analizan los problemas de los sistemas de adquisición y monitoreo de los datos para la explotación de los GEFO y en la toma de decisiones para la operación de los GE ante desastres ambientales. Más Guerra, Peña y Cobos identificaron las deficiencias del SCADA (*Supervision Control and Data Acquisition*) usado en GEFO en la tecnología HYUNDAI e incorporaron información cuantitativa de los parámetros que determinan la eficiencia energética de los GE [Más; Guerra, 2010] [Peña; Cobos, 2012]. Pérez, Rodríguez y Vázquez plantearon la necesidad de crear un gestor de base de datos (BD) para facilitar a los órganos de dirección la adopción de medidas relacionadas con el acomodo de tecnologías ante el impacto u ocurrencia de desastres naturales a través de un Sistema de Información Geográfica (SIG). Estas investigaciones aunque tratan las deficiencias del SCADA, caracterizan la eficiencia energética de estos sistemas y abordan la explotación de los GE ante desastres naturales respectivamente [Pérez; Rodríguez; Vázquez, 2010]. No analizan los GED, unidades que presentan problemas energéticos más severos que los GEFO. En el 2012, solo se publicó un artículo de Pérez, en el mismo se analiza la eficiencia de los GED en dos emplazamientos de Santiago de Cuba, aportando resultados relevantes con respecto a las horas de operación de los GED frente a eventos meteorológicos en régimen paralelo y plantea un estrategia del área de control de calidad en la documentación del Manual de GED en Cuba [Pérez, 2012]. En estas investigaciones no se estudian todos los indicadores para la explotación eficaz de los GED.

En el contexto moense se han realizados varios estudios por investigadores del ISMMM (Instituto Superior Minero Metalúrgico de Moa): Noa en el 2009 y Rosales en el 2010 usaron el Matlab para estudios de la operación de la Batería de Grupos Electrónicos Diésel (GEDB) en ISLA¹, respectivamente [Noa; Marrero; Legra, 2009] [Rosales; Guerrero, 2010]. Por otro lado, Guerrero en el 2011 utilizó el mismo software, pero para analizar el comportamiento de un sistema híbrido de GD [Guerrero; Labore; Ignacio, 2011]. Bisset en el 2010, propone mejoras en la explotación de los GE

¹ Régimen ISLA. Las unidades de GE están diseñadas para funcionar de modo aislado. Las unidades trabajan de forma independiente o paralelas entre sí, pero separadas del SEN de forma tal que los parámetros de frecuencia, voltaje son fijados por ellas (Manual de Calidad para la operación, 2006).

aplicando una metodología de Minería de Datos y Cutiño en el 2012 desarrolló un sistema informático (aplicación de escritorio) para mejorar la toma de decisiones en la explotación de los GE Diésel de Moa [Bisett; Sánchez; Noa, 2010] [Cutiño; Rivera, 2012].

Las aplicaciones informáticas se enfocan en analizar la explotación de los GE ante situaciones excepcionales². Proponen el empleo de técnicas inteligentes de minería de datos y el uso de una aplicación informática de escritorio para la gestión de los indicadores técnico-productivos de operación y de eficiencia energética. En ambos casos los datos usados no están recogidos en un BD normalizada, o sea, se utiliza una hoja electrónica de cálculos de Excel³ que fue diseñada por la Unión Nacional Eléctrica para la Generación Distribuida (UNE-GD) de la provincia Holguín.

Los especialistas que la elaboraron no tuvieron en cuenta los aspectos teóricos y de contenido de los sistemas de BD, por lo que la misma presenta problemas de solapamiento y redundancia de información principalmente, elementos que afectan cualquier resultado de los estudios mencionados porque se simuló y estudió la explotación de los GE a partir de datos no validados, siendo esta **la situación problemática** que origina o perfila la investigación.

En los GED se hace indispensable, la gestión de los sistemas de indicadores como principal argumento para la explotación sistémica y dinámica de estas entidades [Blanco; González; Rodríguez; Miranda; Ortiz, 1989]. Estos indicadores permiten a la institución la interpretación de la información generada por un SCADA en función de la producción de energía. Para incluir estos indicadores en un sistema informático que garantice la explotación eficaz de los GE, se debe tener en cuenta que el diseño de la base de datos debe estar en función de los mismos, garantizando, que se pueda incluir

² Situación excepcional: constituyen estados de ese carácter que se establecen, de forma temporal, en todo el territorio nacional o en una parte de él, en interés de garantizar la defensa nacional o proteger a la población y la economía en caso o ante la inminencia de una agresión militar, de desastres naturales, otros tipos de catástrofes u otras circunstancias que por su naturaleza, proporción o entidad afecten el orden interior, la seguridad del país o la estabilidad del Estado (Gaceta Oficial de la República de Cuba: Ley 75 de la Defensa Civil, Artículo 9, 1998)

³ Hoja electrónica de cálculos de Excel: son las tablas en donde se guarda la información de los los indicadores técnicos-productivos de los GED la cual fue diseñada por la Unión Nacional Eléctrica para la Generación Distribuida de Holguín.

un nuevo indicador, con los respectivos campos o atributos que conformen la fórmula para su posterior cálculo.

En el preámbulo de la presente investigación se han observado ciertas deficiencias respecto a la construcción de la base de datos de los indicadores técnicos-productivos para la explotación eficaz de los Grupos Electrónicos Diésel en Batería de Moa, identificándose como **problema de investigación**.

Objeto de la investigación.

Sistema de información de la base de datos de los indicadores técnicos-productivos de los Grupos Electrónicos Diésel en Batería.

Campo de Acción.

Base de datos de los indicadores técnicos-productivos de los Grupos Electrónicos Diésel en Batería de Moa.

Objetivo General de la investigación.

Desarrollar una base de datos de los indicadores técnicos-productivos para la explotación eficaz de los Grupos Electrónicos Diésel en Batería de Moa.

Idea a defender.

Con el desarrollo de una base de datos de los indicadores técnicos-productivos de los Grupos Electrónicos Diésel en Batería de Moa, se facilita la toma de decisiones para explotación eficaz de los mismos.

Objetivos específicos

- Establecer el marco teórico-contextual de los indicadores técnico-productivos de los Grupos Electrónicos Diésel en Batería de Moa.
- Construir la Base de Datos de los indicadores técnico-productivos de los Grupos Electrónicos Diésel en Batería de Moa.

- Establecer un algoritmo para poblar la Base de Datos de los indicadores técnico-productivos de los Grupos Electrógenos Diésel en Batería de Moa.
- Determinar la factibilidad económica del proyecto

Tareas de la Investigación.

- Estudiar detalladamente el proceso de planificación y control de los principales indicadores de la producción de energía eléctrica a los Grupos Electrógenos Diésel de Moa.
- Estudiar los diferentes sistemas de gestión de datos de los indicadores técnico-productivos de los Grupos Electrógenos Diésel en Batería.
- Estudiar las tecnologías existentes para el desarrollo de bases de datos y seleccionar las más apropiadas para resolver la problemática planteada.
- Diseñar la base de datos de los indicadores técnicos-productivos de los Grupos Electrógenos Diésel en Batería
- Validar la base de datos para que satisfaga las necesidades de los clientes.
- Poblar la base de datos.
- Aplicar la metodología costo-beneficio para evaluar la factibilidad del proyecto.

Estructura de la Tesis:

El presente trabajo está formado por la **introducción, 3 capítulos, conclusiones, recomendaciones y referencias bibliográficas.**

En el capítulo 1: se lleva a cabo el estudio del marco teórico-contextual de la investigación, se realiza un análisis de las aplicaciones informáticas en los Grupos Electrógenos Diésel en Batería. También se efectúa un estudio de las herramientas y metodologías para el desarrollo de bases de datos y se seleccionan las adecuadas para resolver la problemática.

En el capítulo 2: se desarrolla la fase de diseño de la base de datos y se establecen los procedimientos, algoritmos y herramientas que se van a emplear en la construcción

y poblado de la base de datos para para la planificación y control de los principales indicadores de la producción de energía eléctrica a los Grupos Electrógenos Diésel de Moa.

En el capítulo 3: se realiza un análisis sobre los costos y beneficios del desarrollo del proyecto y se identifican los efectos económicos de este.

Capítulo 1: Marco Teórico-Contextual de la Investigación

CAPÍTULO 1. MARCO TEÓRICO-CONTEXTUAL DE LA INVESTIGACIÓN

1.1 Introducción.

En este capítulo se expondrán los antecedentes de la investigación para el establecimiento de marco teórico-contextual de la explotación de la GEDB, citaremos los principales resultados de las investigaciones desarrolladas en el ISMMM en esta área del conocimiento. Además abordaremos los elementos del sistema de información, la gestión y el flujo de información de los indicadores técnico-productivos de los GED.

También se realiza un análisis de las herramientas y las metodologías a emplear en el desarrollo del campo de investigación.

1.2 Antecedentes de la investigación.

Desde el mes de diciembre 2008 hasta diciembre 2012 la instalación y utilización de las tecnologías de grupos electrógenos en Cuba aumentó en 524 GW de capacidad instalada⁴, la provincia de Holguín creció en 56,2 GW (*Gigawatt*). La generación bruta⁵ de electricidad en la provincia varió de 3003,1 a 343,9 GWh (*Gigawatt hora*), para ello se instalaron 24 GEFO y de 118 GED, se desmontaron 3 quedando actualmente 115 GED [ONE, 2009 y 2012], como se muestra en la tabla 1.1.

Tabla 1.1 Datos de la capacidad instalada de los grupos electrógenos

Capacidad instalada (GW)	
Cuba	
2008	2012

⁴ Capacidad instalada: Se refiere a la potencia activa máxima continua, que es capaz de entregar un bloque o unidad generadora medida en los terminales del generador. La capacidad instalada de una central eléctrica o de una empresa, está determinada por la sumatoria de las capacidades instaladas en todas sus unidades generadoras.

⁵ Generación bruta: Se refiere a la generación de energía eléctrica, incluyendo el insumo, de todas las plantas eléctricas de servicio público o de instalaciones generadoras de otros productores.

Capítulo 1: Marco Teórico-Contextual de la Investigación

1871,6	2395,6
Holguín	
205,3	261,5
Diésel	
205,3	201,5
Fuel oil	
0	60,0

Las facilidades que trajo consigo la nueva capacidad de generación provocaron que el consumo de electricidad⁶ en el sector residencial aumentase de 498,4 a 609,2 GWh, donde se encuentran los principales consumidores de la potencia generada por los GE.

En la provincia Holguín, en estos momentos, se cuenta con 20 Grupos Electrógenos Diésel Aislados (GEA) y 7 Baterías Grupos Electrógenos Diésel, el sistema de GD con GED más grande del país y de América Latina en su tipo [Video UEB, 2012].

El municipio Moa cuenta con una Batería de Grupos Electrógenos Diésel (capacidad instalada de 15,104 MW) y se está terminando el montaje de una Batería de Grupos Electrógenos Fuel oil (GEFOB) (capacidad instalada de 184 MW), es el segundo municipio con mayor potencia instalada en GE de la provincia [OBE municipal, 2012; Noa, 2009]. Esto se debe a que en Moa la población es de 73 879, en la zona urbana hay 63 027 (representa 85,3%) habitantes para un consumo de electricidad de 3,3675 GWh [Despacho del Gobierno Municipal, 2012].

Teniendo en cuenta el impacto que tiene en la provincia la instalación de los GE y fundamentalmente en el territorio moense, en el ISMMM se han desarrollado varias

⁶ Consumo de electricidad: Se refiere al consumo de electricidad registrado por todos los sectores de la economía (incluyendo el privado) y con independencia de la fuente de origen (plantas de servicio público o auto productores). (ONE,2012)

Capítulo 1: Marco Teórico-Contextual de la Investigación

investigaciones que abordan diversas temáticas acerca de la explotación de la GEDB. Aproximadamente 18 Tesis de Grado utilizaron la información de la hoja electrónica de cálculos de Excel de la GEDB. Entre las carreras de Ingeniería Eléctrica (12), Mecánica (3) e Informática (3), los temas que se abordaron fueron principalmente:

- La operación en régimen ISLA y paralelo con el SEN; el estudio de los parámetros efectivos y el impacto del medio ambiente en el motor de combustión interna.
- Sistemas informáticos para el análisis de los indicadores que determina el rango del Índice de Consumo de Combustible (ICC) de la GEDB.

Las aplicaciones informáticas que se emplearon para estos estudios fueron:

- Excel
- Power Systems explorer
- Radial
- Matlab
- Win SOS
- Mapinfo
- Winder
- AutoCAD
- Mat Cad
- Weka
- MySQL

De estas investigaciones se debe resaltar que solamente tres de los resultados que se obtienen tratan nuestro objeto de investigación:

Rosales (2010), establece por medio de las simulaciones en Matlab los límites de operación del microsistema GEDB y de las subestaciones de distribución ante una desconexión del SEN, para ello utiliza la hoja electrónica de cálculos de Excel (HECE).

Capítulo 1: Marco Teórico-Contextual de la Investigación

En el 2010 teniendo en cuenta los problemas con los altos costo de generación de la GEDB, se realizó una investigación que utilizó la técnicas de Minería de Datos (MD), en la misma se describe el problema y se establecen los objetivos principales junto con los objetivos más específicos, los costos y beneficios así como la planificación inicial de un proyecto de MD. Se utiliza la metodología CRISP– DM (*Cross – Industry Standard Process for Data Mining*) y las herramientas de análisis de datos WEKA (*Waikato Environment Knowledge Analysis*), MATLAB y Excel. Con estas herramientas y guiada por la metodología se demuestra con un estudio estadístico y las técnicas seleccionadas que las variables Índice de Consumo de Combustible Neto (ICCN) e Índice de Consumo de Combustible Bruto (ICCB) en la data (una hoja electrónica de cálculos de Excel) presentan incoherencias, se eliminan estas variables del estudio y el insumo eléctrico. Se elabora un modelo basado en las reglas inteligentes y el procedimiento por el cual se estableció la dinámica de operación de la Batería de Grupos Electrónicos Diésel de Moa para la toma de decisiones [Bisset, 2010]

A continuación y tomando como referencia el estudios de Bisset, en el 2012 se desarrolla un Sistema Informático (aplicación de escritorio) para mejorar la toma de decisiones en la explotación de los Grupos Electrónicos Diésel de Moa, se construyó una base de datos en MySQL, la cual presenta problemas de redundancia de datos y de seguridad [Cutiño, 2012].

También se han defendido dos Tesis de Maestría: Noa (2009) propone un procedimiento para la operación de los GEDB en ISLA y utiliza la hoja electrónica de cálculos de Excel y el Matlab, estos resultados no se pudieron validar adecuadamente por los problemas de los datos que se han mencionado. Teniendo en cuenta estas deficiencias y en otras condiciones de explotación; Guerrero (2011) simula con el Matlab la operación de un sistema híbrido de GD, con la que estima el trabajo en conjunto de la GEDB, una central hidroeléctrica y un parque eólico. Independientemente de estos resultados se debe destacar que todavía falta por estudiar el diseño que debe tener la base de datos de los Grupos Electrónicos Diésel de manera que admita la inclusión de nuevos indicadores, con sus respectivos campos, apoyados en la implementación de técnicas de base de datos que garantizan el

Capítulo 1: Marco Teórico-Contextual de la Investigación

proceso de cálculo y la conformación de los datos para que los reportes representen la información necesaria para la explotación eficaz. Para ello en la investigación se deben conocer de las técnicas de Data Mining, Base de Datos Activa y la técnica de Base de Datos Relacional con Atributos, las cuales forman parte de la solución del problema de investigación.

1.3 Descripción del sistema de información de los indicadores de la Batería de Grupos Electrónicos Diésel

1.3.1 Sistema de información para la gestión de los indicadores de la Batería de Grupos Electrónicos Diésel

Todo sistema de información (SI)⁷ está constituido por varios subsistemas, existiendo entre ellos interacción y relación para cumplir con el objetivo fundamental por el que fue creado el sistema atendiendo a todo esto se puede declarar que un sistema de información es: "(...) un proceso clave que responde a un macroproceso, en este caso identificado por la actividad principal u objeto social de una organización" [Rodríguez, 2013].

Para establecer el SI en la GEDB se debe conocer la interrelación de los subsistemas de operación, como se muestra en la figura 1.1.

⁷ Sistema de información: Los sistemas de información responden a la satisfacción de necesidades de una organización o de un individuo o grupo. Un sistema de información es un conjunto de elementos o componentes relacionados con la información que interaccionan entre sí para lograr un objetivo (Ponjuan-Dante, Mugia, 2004)

Capítulo 1: Marco Teórico-Contextual de la Investigación

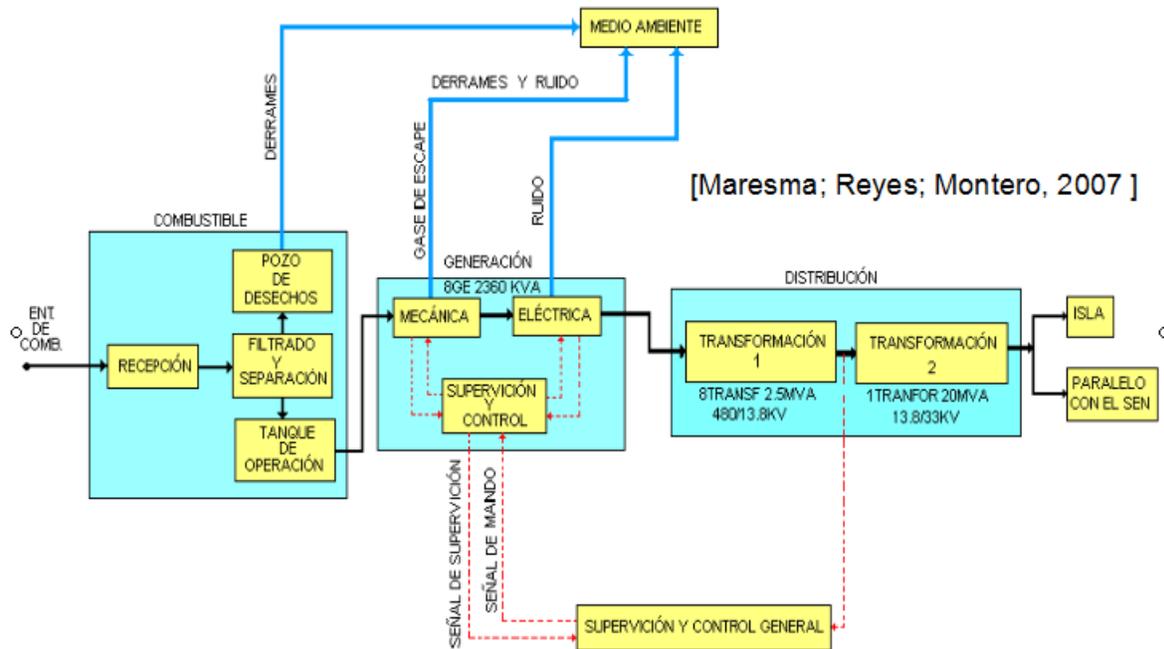


Figura. 1.1 Diagrama en bloque de operación de la Batería de GE

Cada uno de estos subsistemas de operación generan indicadores (datos e información) que se envían a través del protocolo MODBUS al SCADA "Eros" [Manual de Usuario Eros, 2006] que está en la computadora de la GEDB (vea la Figura 1, del Anexo 1). También se miden los datos de consumo de combustible en la Estación de Combustible pero estos se entran por teclado al SCADA. Este SCADA envía por medio de *frame relay* (sistema de transmisión de datos) hacia UNE-GD provincial. Los indicadores que se almacenan en la hoja electrónica de cálculo de Excel y se mandan luego a la UNE-GD nacional, con el objetivo de garantizar la generación conjunta de los GE del país para satisfacer las necesidades energéticas del sector residencial principalmente.

Los indicadores generados permiten en los diferentes niveles de la UNE-GD la toma de decisiones y para ello son necesarios: documentos, registros, ficheros o archivos, equipos, elementos de apoyo a los sistemas, personas. Según Rodríguez, estos son los componentes básicos de un sistema de información [Rodríguez, 2013].

Capítulo 1: Marco Teórico-Contextual de la Investigación

1.3.2 Sistema de gestión de los indicadores de la Batería de Grupos Electrógenos Diésel

La gestión del sistema de indicadores de los GED los estableció la UNE-GD nacional [Estructura Funcional, 2006] con los siguientes objetivos:

1. Disponer oportunamente (**frecuencia diaria**, antes de las 07:00) de la información necesaria para la evaluación de las condiciones existentes para garantizar el servicio y poder adoptar las **decisiones operativas** que se requieran.
2. Disponer oportunamente (**frecuencia mensual**, antes del día 7 del mes siguiente) de la información validada que permita realizar análisis de cualquier tipo, la evaluación del desempeño de la gestión y la **certificación oficial** de los indicadores a cualquier efecto.

Niveles del sistema de gestión:

- Nivel de emplazamiento: captación, registro, procesamiento y reporte.
(Emplazamientos con Baterías y Emplazamientos con Grupos Aislados)
- Nivel provincial: recepción, validación, procesamiento y reporte.
(Despacho Provincial, Despacho Territorial y OBE Provincial)
- Nivel de la organización: recepción y procesamiento.
(Despacho Nacional, Dirección de Generación UNE)

Gestión de la información a nivel de emplazamiento:

Las funciones del sistema de gestión establece a nivel de emplazamiento son las siguientes:

- Captación de la Información
- Registro de la Información
- Procesamiento de la información
- Reporte al Despacho (OBE) Provincial

Capítulo 1: Marco Teórico-Contextual de la Investigación

La captación de la información primaria se realiza por lecturas directas del Operador en los instrumentos del equipamiento instalado. En particular se establecen los siguientes parámetros a captar:

- Eventos Primarios de operación:
 - Hora de inicio del evento.
 - Potencia activa del evento iniciado.
 - Tipo de evento.
 - Incidencias de la operación.

Registro: Los eventos primarios se asientan por el operador en el Registro de Incidencias. Las horas que se registran son las que valida el Despacho Territorial (a través del Despacho Provincial).

- Recepción del Combustible
 - Fecha y hora de entrada del carro cisterna.
 - Chapa del carro cisterna.
 - Número de la factura comercial o conduce.
 - Volumen de combustible declarado en la factura comercial o conduce.
 - Volumen de combustible realmente recibido.
 - Nombre y firma del operador.

Registro: Los datos sobre la recepción del combustible se asientan por el operador en el Registro de Recepción de Combustible.

En cada emplazamiento donde se realiza la descarga parcial del carro cisterna se recogerá un conduce firmado por el chofer, consignando los datos de la descarga efectuada. En el emplazamiento donde se realiza una descarga total o se termina de descargar el carro cisterna en una serie de descargas parciales, se recibe adicionalmente la factura comercial. Las desviaciones (reclamaciones) se registran al dorso del conduce o factura. Se firman por el chofer y el operador.

Capítulo 1: Marco Teórico-Contextual de la Investigación

Cálculos: - Existencia de combustible antes de la recepción.

- Capacidad de recepción de combustible antes de la descarga.
- Volumen de combustible recibido en cada recepción.
- Volumen de combustible recibido en 24 horas (22:00 – 22:00).

Balance Producción-Consumo - Potencia activa de cada GE, lectura horaria

- Potencia reactiva de cada GE, lectura horaria
- Lectura del metro contador de generación bruta de cada GE, en los horarios: 06:00; 18:00 y 22:00
- Lectura de todos los metro-contadores de insumo eléctrico, en los horarios: 06:00; 18:00 y 22:00.
- Nivel de cada uno de los tanques de combustible, en el horario 22:00.

Registro: Los datos sobre el balance producción-consumo se asientan por el operador en las Hojas de Lectura. El diseño de las Hojas de Lectura será específico para cada emplazamiento, garantizando la posibilidad de calcular los datos del reporte.

Cálculos: - Potencia horaria activa de cada Batería de GE o del emplazamiento en el caso de los GE aislados.

- Potencia horaria reactiva de cada Batería de GE o del emplazamiento en el caso de los GE aislados.
- Generación Bruta del Emplazamiento (últimas 24 horas, con cierre 22.00) así como su desglose en los horarios de la madrugada, el día y el pico.
- Insumo eléctrico del emplazamiento (últimas 24 horas, con cierre 22.00) así como su desglose en los horarios de la madrugada, el día y el pico.

Capítulo 1: Marco Teórico-Contextual de la Investigación

- Existencia de combustible en el Emplazamiento a las 22:00.
- Cobertura de combustible.
- Reportes: - Como parte del proceso de operación, se reportan al Despacho Provincial:
 - Los datos que caracterizan los eventos primarios.
- Cada segunda hora se reportan al Despacho Provincial:
 - La Potencia horaria activa y reactiva de cada Batería de GE o del emplazamiento en el caso de los GE aislados.
- Antes de las 00:00 se reportan al Despacho Provincial:
 - Generación bruta del emplazamiento (24 horas, con cierre 22.00), así como su desglose en los horarios de la madrugada, el día y el pico.
 - Insumo eléctrico del emplazamiento (24 horas, con cierre 22.00), así como su desglose en los horarios de la madrugada, el día y el pico.
 - Existencia de combustible en el emplazamiento a las 22:00.
 - Suministro de combustible al emplazamiento (24 horas, con cierre 22.00).
- Antes de las 08:00 se reportan a la OBE Provincial:
 - Las reclamaciones reportadas por el suministro de combustible.
 - Existencia de combustible en el emplazamiento a las 06:00 (en el caso en que la cobertura sea menor que la magnitud crítica).
- En el término de 48 horas se recogen por la OBE Provincial:
 - Las facturas comerciales o conduces de combustible recibidas.
- El día primero del mes se recogen por la OBE Provincial:
 - Los registros de recepción de combustible del mes vencido.
 - Las Hojas de Lecturas del mes vencido.
- Luego de agotados (libros llenos) se recogen por la OBE Provincial:

Capítulo 1: Marco Teórico-Contextual de la Investigación

- Los Registros de Operación.

Los indicadores recogidos en estos registros y los configurados en el SCADA permiten establecer la Planificación del sistema (indicadores que se gestionan):

- *Balance de disponibilidad*
- *Balance producción – consumo*
- *Disponibilidad materias primas*

Los indicadores a partir del cual se elabora los balances se guardan en una hoja electrónica de cálculos de Excel conocida como "Informe de Indicadores Principales".

1.3.3- Flujo de información del sistema de gestión de los indicadores de la Batería de Grupos Electrónicos Diésel

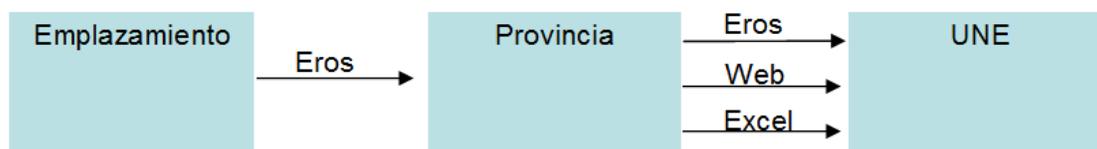


Figura. 1.2 Diagrama del flujo de información de los GED

La actualización del Sitio Web de Generación Distribuida, está diseñado para concentrar la información proveniente del Eros y de la hoja electrónica de cálculos de Excel y mostrar estos datos de la manera más cómoda y factible para los usuarios.

El SCADA Eros, es la herramienta más eficaz, porque es capaz de recoger las variables de cada emplazamiento de forma automática y enviarlas hasta la base de datos nacional. La intervención del hombre es mínima y la actualización de los datos es instantánea.

A partir de los datos recogidos en la hoja electrónica de cálculos de Excel (Informe de Indicadores principales) se llenan las variables, estas se dividen en funcionales o principales y se facilita su captación ya sea por lecturas a nivel de cada grupo o valores a nivel de emplazamiento. Esta herramienta se mantendrá a pesar del EROS, para casos de contingencia del SCADA en algún emplazamiento o provincia.

Capítulo 1: Marco Teórico-Contextual de la Investigación

Por dificultades en la forma de llenado de la Web (on line). Se implementa un reporte semanal en una hoja electrónica de cálculo de Excel (con la idea de desarrollar una variante que logre importar automáticamente datos de la hoja electrónica de cálculos de Excel (enviados por correo electrónico) hacia las bases de datos nacionales. Esto es una tercera vía de actualización de información.

Se puede constatar a partir de la información que se presenta en los acápite que el núcleo del sistema de gestión de los indicadores de los GED es la hoja electrónica de cálculos de Excel "Informe de Indicadores Principales", porque garantiza primero la elaboración de los balances y segundo el flujo de información hasta el nivel nacional.

Si se tiene en cuenta el número de investigaciones que se han desarrollado en el ISMMM y la cantidad de GED instalados en Holguín (20 GEA y 7 GEDB), se puede afirmar que la solución del problema de investigación repercutirá no solo a nivel provincial, sino a nivel nacional, porque la hoja electrónica de cálculo de Excel: Informe de Indicadores Principales, presenta problemas de solapamiento y redundancia de información.

El desarrollo de una base de datos que solucione los problemas identificados debe estar analizado de manera lógica que no force el uso de las técnicas de Data Mining, Base de Datos Activa y la técnica de Base de Datos Relacional con Atributos, sino que responda al problema analizado. Es por ello que el estudio de las herramientas y las metodologías a usar debe satisfacer en primer lugar al usuario, o sea, a la UNE-GD provincial y en este caso, solo es necesario diseñar una Base de Datos Relacional.

A continuación se expone los sistemas de gestores de base de datos y las metodologías. Se abordan las características de cada uno y se seleccionan las herramientas y las metodologías a usar teniendo en cuenta que es necesario diseñar una Base de Datos Relacional.

Capítulo 1: Marco Teórico-Contextual de la Investigación

1.4 Herramientas.

1.4.1. Sistemas Gestores de Base de Datos (SGBD)

Un Sistema Gestor de Bases de Datos (SGBD) o DBMA es una colección de programas cuyo objetivo es servir de interfaz entre la base de datos, el usuario y las aplicaciones. Se compone de un lenguaje de definición de datos, de un lenguaje de manipulación de datos y de un lenguaje de consulta. Un SGBD permite definir los datos a distintos niveles de abstracción y manipular dichos datos, garantizando la seguridad e integridad de los mismos.

Algunos ejemplos de SGBD son: PostgreSQL, MySQL, SQL Server, etc.

Un SGBD debe permitir: definir una base de datos: especificar tipos, estructuras y restricciones de datos. Construir la base de datos: guardar los datos en algún medio controlado por el mismo SGBD. Manipular la base de datos: realizar consultas, actualizarla, generar informes.

Características de un Sistema Gestor de Base de Datos: [CAVSI, 2004]

- a) Abstracción de la información. Los SGBD ahorran a los usuarios detalles acerca del almacenamiento físico de los datos. Da lo mismo si una base de datos ocupa uno o cientos de archivos, este hecho se hace transparente al usuario. Así, se definen varios niveles de abstracción.
- b) Independencia. La independencia de los datos consiste en la capacidad de modificar el esquema (físico o lógico) de una base de datos sin tener que realizar cambios en las aplicaciones que se sirven de ella. Redundancia mínima. Un buen diseño de una base de datos logrará evitar la aparición de información repetida o redundante. De entrada, lo ideal es lograr una redundancia nula; no obstante, en algunos casos la complejidad de los cálculos hace necesaria la aparición de redundancias.
- c) Consistencia. En aquellos casos en los que no se ha logrado esta redundancia nula, será necesario vigilar que aquella información que aparece repetida se actualice de forma coherente, es decir, que todos los datos repetidos se actualicen de forma simultánea. Seguridad. La información almacenada en una base de

Capítulo 1: Marco Teórico-Contextual de la Investigación

datos puede llegar a tener un gran valor. Los SGBD deben garantizar que esta información se encuentre protegida frente a usuarios malintencionados, que intenten leer información privilegiada; frente a ataques que deseen manipularla o destruirla; o simplemente ante las torpezas de algún usuario autorizado pero despistado. Normalmente, los SGBD disponen de un complejo sistema de permisos a usuarios y grupos de usuarios, que permiten otorgar diversas categorías de permisos. Integridad. Se trata de adoptar las medidas necesarias para garantizar la validez de los datos almacenados. Es decir, se trata de proteger los datos ante fallos de hardware, datos introducidos por usuarios descuidados, o cualquier otra circunstancia capaz de dañar la información almacenada.

- d) Respaldo y recuperación. Los SGBD deben proporcionar una forma eficiente de realizar copias de respaldo de la información almacenada en ellos, y de restaurar a partir de estas copias los datos que se hayan podido perder.
- e) Control de la concurrencia. En la mayoría de entornos (excepto quizás el doméstico), lo más habitual es que sean muchas las personas que acceden a una base de datos, bien para recuperar información, bien para almacenarla. Y es también frecuente que dichos accesos se realicen de forma simultánea. Así pues, un SGBD debe controlar este acceso concurrente a la información, que podría derivar en inconsistencias.

1.4.1.1 PostgreSQL

PostgreSQL: es un servidor de Base de Datos Relacional orientado a objetos de software libre, publicado bajo la licencia BSD40. Como muchos otros proyectos open source, el desarrollo de PostgreSQL no es manejado por una sola compañía sino que es dirigido por una comunidad de desarrolladores y organizaciones comerciales las cuales trabajan en su desarrollo. Dicha comunidad es denominada el PGDG (*PostgreSQL Global Development Group*).

Principales características:

Alta concurrencia: Mediante un sistema denominado MVCC (*Acceso concurrente Multiversión, por sus siglas en inglés*) PostgreSQL permite que mientras un proceso escribe en una tabla, otros accedan a la misma tabla sin necesidad de bloqueos. Cada

Capítulo 1: Marco Teórico-Contextual de la Investigación

usuario obtiene una visión consistente de lo último a lo que se le hizo (*commit*). Esta estrategia es superior al uso de bloqueos por tabla o por filas común en otras bases, eliminando la necesidad del uso de bloqueos explícitos. Amplia variedad de tipos nativos. PostgreSQL provee soporte para:

- Números de precisión arbitraria.
- Texto de largo ilimitado.
- Figuras geométricas (con una variedad de funciones asociadas)
- Direcciones IP (IPv4 e IPv6).
- Bloques de direcciones estilo CIDR.
- Direcciones MAC.
- Adicionalmente los usuarios pueden crear sus propios tipos de datos, los que pueden ser por completo indexables gracias a la infraestructura GiST de PostgreSQL. Algunos ejemplos son los tipos de datos GIS creados por el proyecto PostGIS [Eaprende.com, 2001].

Gracias a su licencia BSD, se permite la utilización del código para ser comercializado. Uno de los casos ejemplo es la de Enterprise DB (*Postgresql Plus*), la cual incluye varios agregados y una interfaz de desarrollo basada en Java. Entre otras empresas que utilizan PostgreSQL para comercializar se encuentra *CyberTech* (Alemania), con su producto *CyberCluster*.

1.4.1.2 MySQL

MySQL: es un sistema de gestión de Bases de Datos Relacionales. Fue creada por la empresa sueca MySQL AB, la cual tiene el copyright del código fuente del servidor SQL, así como también de la marca.

MySQL es un software de código abierto, licenciado bajo la GPL de la GNU, aunque MySQL AB distribuye una versión comercial, en lo único que se diferencia de la versión libre, es en el soporte técnico que se ofrece y la posibilidad de integrar este gestor en un software propietario, ya que de otra manera, se vulneraría la licencia GPL. El lenguaje de programación que utiliza MySQL es *Structured Query Language* (SQL) que

Capítulo 1: Marco Teórico-Contextual de la Investigación

fue desarrollado por IBM en 1981 y desde entonces es utilizado de forma generalizada en las Bases de Datos Relacionales [Nieves, 2007]. Inicialmente, MySQL carecía de algunos elementos esenciales en las Bases de Datos Relacionales, tales como integridad referencial y transacciones. A pesar de esto, atrajo a los desarrolladores de páginas web con contenido dinámico, debido a su simplicidad, de tal manera que los elementos faltantes fueron complementados por la vía de las aplicaciones que la utilizan. Poco a poco estos elementos faltantes, están siendo incorporados tanto por desarrolladores internos, como por desarrolladores de software libre.

En las últimas versiones se pueden destacar las siguientes características principales:

- El principal objetivo de MySQL es velocidad y robustez.
- Soporta gran cantidad de tipos de datos para las columnas.
- Gran portabilidad entre sistemas, puede trabajar en distintas plataformas y sistemas operativos.
- Cada base de datos cuenta con 3 archivos: Uno de estructura, uno de datos y uno de índice y soporta hasta 32 índices por tabla.
- Aprovecha la potencia de sistemas multiproceso, gracias a su implementación multihilo.
- Flexible sistema de contraseñas (passwords) y gestión de usuarios, con un muy buen nivel de seguridad en los datos.
- El servidor soporta mensajes de error en distintas lenguas.

Ventajas: [Eaprende.com, 2001]

- Velocidad al realizar las operaciones, lo que le hace uno de los gestores con mejor rendimiento.
- Bajo costo en requerimientos para la elaboración de bases de datos, ya que debido a su bajo consumo puede ser ejecutado en una máquina con escasos recursos sin ningún problema.
- Facilidad de configuración e instalación.
- Soporta gran variedad de Sistemas Operativos

Capítulo 1: Marco Teórico-Contextual de la Investigación

- Baja probabilidad de corromper datos, incluso si los errores no se producen en el propio gestor, sino en el sistema en el que está.
- Conectividad y seguridad

Desventajas:

- Un gran porcentaje de las utilidades de MySQL no están documentadas.
- No es intuitivo, como otros programas (ACCESS).

Los SGBD mencionados anteriormente, aunque se pueden emplear para diseñar la base de datos, no responde a los requisitos del cliente (UNE-GD). Además, la propuesta del SGBD está fundamentada en que el sistema deber tener un módulo que permita aplicar Minería de Datos, porque la provincia de Holguín es la segunda con mayor cantidad de GED en explotación y por tanto precisan que la base de datos del "Informe de Indicadores Principales", aporte la información necesaria para la explotación eficaz de los GED.

Teniendo en cuenta estos aspectos, el usuario impone como requisito usar el SGBD: **SQL Server.**

1.4.1.3 SQL Server

SQL Server es un conjunto de objetos eficientemente almacenados. Los objetos donde se almacena la información se denominan tablas, y estas a su vez están compuestas de filas y columnas. En el centro de SQL Server está el motor de SQL Server, el cual procesa los comandos de la base de datos. Los procesos se ejecutan dentro del sistema operativo y entienden únicamente de conexiones y de sentencias SQL.

SQL Server incluye herramientas para la administración de los recursos que el ordenador proporciona y los gestiona para un mejor rendimiento de la base de datos.

Transact-SQL es el lenguaje que utiliza SQL Server para poder enviar peticiones tanto de consultas, inserciones, modificaciones, y de borrado a las tablas, así como otras peticiones que el usuario necesite sobre los datos. En definitiva, es un lenguaje que utiliza SQL Server para poder gestionar los datos que contienen las tablas.

Capítulo 1: Marco Teórico-Contextual de la Investigación

El lenguaje estándar SQL (*Structured Query Language*) se emplea para los sistemas de bases de datos relacionales RDBMS (*Relational Database Management System*), es estándar ANSI (*American National Standards Institute*). (*FormaSelect*)

Características de Microsoft SQL Server:

- Soporte de transacciones.
- Escalabilidad, estabilidad y seguridad.
- Soporta procedimientos almacenados.
- Incluye también un potente entorno gráfico de administración, que permite el uso de comandos DDL y DML gráficamente.
- Permite trabajar en modo cliente-servidor, donde la información y datos se alojan en el servidor y las terminales o clientes de la red solo acceden a la información.
- Además permite administrar información de otros servidores de datos. Este sistema incluye una versión reducida, llamada MSDE con el mismo motor de base de datos pero orientado a proyectos más pequeños, que en sus versiones 2005 y 2008 pasa a ser el SQL Express Edition, que se distribuye en forma gratuita. [Islasoft.com, 2006]

En el diseño y la construcción lógica y física de la base de datos se empleará el Embarcadero ER/Studio.

1.4.2 Embarcadero ER/Studio.

Es una herramienta de modelado de datos. Su ambiente es de gran alcance y multinivel. Simple y fácil al usuario, ayuda a las organizaciones para tomar decisiones en cómo resolver embotellamientos de los datos. Elimina redundancia y alcanza en última instancia usos de más alta calidad que entreguen datos más eficientes y exactos a la empresa [Bureaudeprensa.com, 2008].

Luego de establecido que SQL Server será SGBD, que con el Embarcadero ER/Studio se realizará diseño y la construcción de la base de datos, analizamos como paso a seguir , que es necesario para poblar la base de datos seleccionar la suite **Pentaho**

Capítulo 1: Marco Teórico-Contextual de la Investigación

que consta de la herramienta **Kettle de Pentaho (Spoon)**, encargada del proceso de extracción, transformación y carga de datos (ETL).

1.4.3 Pentaho

Pentaho es una suite completa de inteligencia de negocio que la ha convertido en la más popular del mundo por sus resultados obtenidos. Está orientada a la solución y centrada en procesos, contiene una serie de componentes para implementar soluciones basadas en estos. Fue desarrollada bajo la filosofía de código abierto para la gestión y toma de decisiones empresariales. Es una plataforma compuesta de diferentes programas que satisfacen los requisitos de inteligencia de negocio.

Ofrece soluciones para la gestión y análisis de la información, incluyendo el análisis multidimensional OLAP: presentación de informes, minería de datos y creación de cuadros de mando para el usuario, por lo que constituye una solución factible para cualquier empresa.

También, a pesar de ser una suite completa, permite elegir entre el uso completo o la utilización de algunas herramientas para resolver los problemas de negocio. A continuación se mencionan algunas herramientas que se encuentran dentro de la suite:

- a) Subsistema de integración: es donde se realiza el perfilado de los datos, los procesos de extracción, transformación y carga (*Extract, Transform and Load*, ETL, por sus siglas en inglés) de los datos, de la fuente de origen de la información hacia el mercado de datos.
- b) Subsistema de almacenamiento: Subsistema que contiene la base de datos diseñada multidimensionalmente a través de las tablas de hechos y dimensiones cargadas con el proceso de ETL.

1.4.4 Kettle de Pentaho (Spoon)

Kettle de Pentaho.: es el componente de Pentaho responsable de los procesos de ETL. Si bien el uso más frecuente de toda herramienta ETL, incluyendo esta, es la población de almacenes de datos (datawarehouses), se puede utilizar también para otros propósitos tales como:

Capítulo 1: Marco Teórico-Contextual de la Investigación

- Migración de datos entre bases de datos o aplicaciones
- Exportación de datos de bases de datos a archivos planos
- Carga masiva de información en bases de datos
- Limpieza de datos
- Integración de aplicaciones

Esta herramienta es fácil de usar. Todos los procesos se crean en un entorno gráfico donde uno especifica qué hacer sin necesidad de escribir código para indicar cómo hacerlo. Es por esto que se dice que la solución está orientada a meta-data. Se puede utilizar en forma independiente, o integrado con el resto de los componentes de la suite de Pentaho. Como herramienta ETL, es la más popular entre las de código abierto. Soporta amplia variedad de formatos de entradas y salidas, incluyendo archivos planos, planillas de cálculo y conexión con motores de bases de datos tanto comerciales como abiertas. Además, de las entradas y salidas, posee una amplia gama de transformaciones que permiten manipular datos sin limitaciones.

1.5 Metodologías de desarrollo del software

Las metodologías de desarrollo de software son un conjunto de procedimientos, técnicas y ayudas a la documentación para el desarrollo de productos software. Pueden ser comparadas con un plan de contingencias en el que se van indicando paso a paso, todas las actividades a realizar para lograr el producto informático deseado, indicando además quiénes deben participar en el desarrollo de las actividades y qué papel deben tener. Detallan, además, la información que se debe producir como resultado de una actividad y la información necesaria para iniciarla.

En la investigación se han utilizado dos metodologías para guiar los pasos del diseño y la población de la base de datos.

1.5.1 Metodología MetdisBD

La metodología MetdisBD es empleada en el diseño de la base de datos, esta consta de los siguientes pasos:

Capítulo 1: Marco Teórico-Contextual de la Investigación

- Determinación de entidades y atributos
- Normalización de entidades
- Determinación de relaciones (DER)
- Obtención del modelo lógico global de los datos
- Diseño físico de la BD

Cuando se va a realizar el diseño de la base de datos para un sistema, es imprescindible determinar los datos que son necesarios tomar en cuenta y las dependencias funcionales existentes entre ellos.

Rigurosamente, esto se obtiene luego de realizada la etapa de análisis del sistema y partiendo de lo obtenido en esta, pero se tratará de dar indicaciones que permitan lograr un buen diseño de la base de datos, a partir de lo que se conoce [Mato, 1999].

1.5.2 Metodología HEFESTO

HEFESTO es una metodología propia, cuya propuesta está fundamentada en procesos de confección de almacenes de datos.

La idea principal, es comprender cada paso que se realizará, para no caer en el tedio de tener que seguir un método al pie de la letra sin saber exactamente qué se está haciendo, ni por qué.

La construcción e implementación de un *Data Warehouse* puede adaptarse muy bien a cualquier ciclo de vida de desarrollo de software, con la salvedad de que para algunas fases en particular, las acciones que se han de realizar serán muy diferentes. Lo que se debe tener muy en cuenta, es no entrar en la utilización de metodologías que requieran fases extensas de reunión de requerimientos y análisis, fases de desarrollo monolítico que conlleve demasiado tiempo y fases de despliegue muy largas. Lo que se busca, es entregar una primera implementación que satisfaga una parte de las necesidades, para demostrar las ventajas del *Data Warehouse* y motivar a los usuarios.

Capítulo 1: Marco Teórico-Contextual de la Investigación

La metodología HEFESTO, puede ser embebida en cualquier ciclo de vida que cumpla con la condición antes declarada.

Con el fin de que se llegue a una total comprensión de cada paso o etapa, se ha realizado una detallada explicación de todo el ciclo de vida del DW, reflejándose todos los resultados obtenidos en cada paso de la metodología utilizada [Bernau, 2010].

Características de la Metodología HEFESTOS

- Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender.
- Se basa en los requerimientos del usuario, por lo cual su estructura es capaz de adaptarse con facilidad y rapidez, ante los cambios en el negocio.
- Reduce la resistencia al cambio, ya que involucra al usuario final en cada etapa para que tome decisiones respecto al comportamiento y funciones del *Data Warehouse*.
- Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- Es independiente del tipo de ciclo de vida que se emplee para contener la metodología.
- Es independiente de las herramientas que se utilicen para su implementación.
- Es independiente de las estructuras físicas que contengan el *Data Warehouse* y de su respectiva distribución.
- Cuando se culmina con una fase, los resultados obtenidos se convierten en el punto de partida para llevar a cabo el paso siguiente.
- Se aplica tanto para *Data Mart* como para *Data Warehouse* [Bernau,2010].

Estas características permiten que la metodología HEFESTO pueda ser usada también en el poblado de bases de datos, para ello se deben seguir los siguientes pasos:

- Transformar Hoja Electrónica de Cálculos de Excel para eliminar los errores.

Capítulo 1: Marco Teórico-Contextual de la Investigación

- Utilizar la herramienta *Spoon* de *Pentaho* para el proceso de ETL (extracción, transformación y carga).
- Carga en la base de datos en el servidor a usar.

1.6 Metodologías y Herramientas a utilizar

Luego del estudio realizado se arriba a las conclusiones de que el Sistema Gestor de Base de Datos a ser utilizada es el SQL Server 2008 por las características antes mencionadas, además tiene incorporado un módulo de minería de datos el cual será recomendado para aplicar técnicas de minería de datos sobre la base de datos construida. Se empleará la herramienta RSstudio para diseñar la Base de Datos, por las facilidades que brinda a la hora de diseñar y crear el modelo físico de la base de datos. Será usada la herramienta Spoon de Pentaho para poblar la base de datos, pues esta nos permite realizar todo el proceso ETL sobre los datos. El proceso será guiado por la unión de dos metodologías: MetdisBD y HEFESTOS. La primera guiará todo el proceso desde el análisis del negocio y de los requisitos hasta la fase de análisis y diseño de la base de datos. La segunda se utilizará para guiar el proceso ETL sobre los datos.

Capítulo 1: Marco Teórico-Contextual de la Investigación

1.7 Conclusiones del Capítulo.

- Se estableció el marco teórico-contextual de los indicadores técnico-productivos de los Grupos Electrónicos Diésel en Batería.
- En las referencias bibliográficas consultadas se determinó que era necesario profundizar el estudio del proceso de planificación y control para la explotación eficaz de los Grupos Electrónicos Diésel en Batería de Moa, usando un sistema de información y de gestión de los indicadores técnico-productivo de la entidad.
- Las tecnologías para el desarrollo de bases de datos seleccionadas son las más apropiadas para la solución del problema analizado.

Capítulo 2: Diseño y Población de la Base de Datos

CAPÍTULO 2. DISEÑO Y POBLACIÓN DE LA BASE DE DATOS

2.1 Introducción

En este capítulo se realiza la descripción de los datos de los Grupos Electrógenos Diésel aislados y en batería, se explica los campos de las tablas de la Hoja Electrónica de Cálculos de Excel (HECE). Luego se abordan los elementos teóricos para el diseño y poblado de la base de datos siguiendo las metodologías enunciadas en el capítulo 1. También se presenta la estructura física de la BD "Indicadores Técnicos-Productivos de la GEDB".

2.2 Descripción de los datos de los Grupos Electrógenos Diésel en Batería.

Mensualmente en la empresa UNE-GD se genera un reporte en una HECE con la información referente a los indicadores técnicos-productivos de los GEA y GEDB de la provincia Holguín, el cual contempla datos como: la generación bruta, insumo eléctrico, consumo de combustible, ICC, existencia y entrada de combustible. El tratamiento estadístico de la información para obtener el valor de los datos del reporte se realiza en dos tablas en dos hojas de la HECE "Informe de Indicadores Principales", vea las tablas 2.1 y 2.2.

Capítulo 2: Diseño y Población de la Base de Datos

Tabla 2.1 Informe de Indicadores principales de Moa

Tabla de Escritura																				Consumo de Aceite y líquido Refrig. por Grupos.				Lect.Inicial	
Día	Generación	Insumo	Índice Consumo	Consumo Comb.	Existencia Comb.	Entradas	Horas total G1	Horas total G2	Exist. Aceite	Cons. Relleno	Cons. Mito	Entrada aceite	Exist. Refrig	Cons. Refrig	Entrada refrigerante	Grupo 1		Grupo 2		62772481					
																Aceite	Refrig.	Aceite	Refrig.	BATERIA	GENERACION				
1	110554	228.6	OK	29569.0	227695	0	OK	5737	6215	455	115	0	0	OK	138	0	0	OK	15	0	15	0	62883035	110554	OK
2	164373	237.0	OK	45584.0	182111	0	OK	5754	6232	362	93	0	0	OK	130	8	0	OK	15	0	17	0	63047408	164373	OK
3	153404	227.2	OK	40784.0	260506	119179	OK	5771	6249	287	75	0	0	OK	115	15	0	OK	15	0	0	0	63200812	153404	OK
4	148502	228.4	OK	39699.0	220807	0	OK	5788	6263	453	0	260	416	OK	115	0	0	OK	0	0	0	0	63349314	148502	OK
5	202923	230.9	OK	54830.0	165977	0	OK	5800	6287	317	136	0	0	OK	115	0	0	OK	0	0	27	0	63552237	202923	OK
6	208777	233.2	OK	56971.0	175218	66212	OK	5824	6311	208	109	0	0	OK	115	0	0	OK	14	0	13	0	63761014	208777	OK
7	77191	236.2	OK	21339.0	180833	26954	OK	5833	6320	1499	165	0	1456	OK	115	0	0	OK	45	0	30	0	63838205	77191	OK
8	0			0.0	180833	0	OK	5833	6320	1499	0	0	0	OK	115	0	0	OK	0	0	0	0	63838205	0	OK
9	6680	244.6	Error	1912.0	212027	33106	OK	5836	6323	1239	0	260	0	OK	5	110	0	OK	0	0	0	0	63844885	6680	OK
10	63323	265.4	Error	19670.0	218686	26329	OK	5846	6338	1239	0	0	0	OK	5	0	0	OK	0	0	0	0	63908208	63323	OK
11	68505	261.5	Error	20966.0	257780.0	60060	OK	5846	6357	974	15	250	0	OK	5	0	0	OK	0	0	15	0	63976713	68505	OK
12	143541	225.6	OK	37889.0	219891.0	0	OK	5846	6379	751	15	0	-208	OK	5	0	0	OK	0	0	0	0	64120254	143541	OK
13	142455	236.5	OK	39429.0	180462	0	OK	5846	6394	461	30	260	0	OK	0	5	0	OK	0	0	0	0	64262709	142455	OK
14	219388	227.0	OK	58271.0	122191	0	OK	5846	6415	1121	120	260	1040	OK	208	0	208	OK	0	0	0	0	64482097	219388	OK
15	146527	231.5	OK	39701.0	206282	123792	OK	5846	6430	861	0	260	0	OK	208	0	0	OK	0	0	0	0	64628624	146527	OK
16	135601	226.0	OK	35858.0	203530	33106	OK	5846	6449	816	45	0	0	OK	198	10	0	OK	0	0	0	0	64764225	135601	OK
17	114236	225.3	OK	30113.0	173417	0	OK	5846	6460	720	96	0	0	OK	198	0	0	OK	0	0	0	0	64878461	114236	OK
18	156108	236.9	OK	43279.0	288891	158753	OK	5846	6479	694	26	0	0	OK	178	20	0	OK	0	0	0	0	65034569	156108	OK
19	177819	222.9	OK	46384.0	242507	0	OK	5846	6500	1644	90	0	1040	OK	178	0	0	OK	0	0	0	0	65212388	177819	OK
20	112948	232.3	OK	30706.0	211801	0	OK	5846	6514	1569	75	0	0	OK	516	0	338	OK	0	0	15	0	65325336	112948	OK
21	141905	226.0	OK	37534.0	174267	0	OK	5846	6531	1207	102	260	0	OK	506	10	0	OK	0	0	25	0	65467241	141905	OK
22	134957	235.3	OK	37167.0	137100	0	OK	5856	6546	1207	0	0	0	OK	360	146	0	OK	0	106	0	0	65602198	134957	OK
23	147470	231.3	OK	39911.0	97189	0	OK	5873	6563	1102	105	0	0	OK	360	0	0	OK	15	0	15	0	65749668	147470	OK
24	171979	228.4	OK	45958.0	51231	0	OK	5892	6577	982	120	0	0	OK	360	0	0	OK	45	0	30	0	65921647	171979	OK
25	156247	231.9	OK	42398.0	75045	66212	OK	5910	6577	832	150	0	0	OK	360	0	0	OK	25	0	0	0	66077894	156247	OK
26	129623	229.6	OK	34832.0	133379	93166	OK	5927	6577	708	124	0	0	OK	360	0	0	OK	31	0	0	0	66207517	129623	OK
27	112417	226.2	OK	29759.0	103620	0	OK	5945	6577	594	114	0	0	OK	360	0	0	OK	30	0	0	0	66319934	112417	OK
28	98041	229.6	OK	26340.0	77280	0	OK	5963	6577	594	0	0	0	OK	360	0	0	OK	0	0	0	0	66417975	98041	OK
29	93495	230.0	OK	25161.0	78448	26329	OK	5982	6577	516	78	0	0	OK	360	0	0	OK	50	0	0	0	66511470	93495	OK
30	69698	242.5	Error	19779.0	84998	26329	OK	5999	6577	1140	0	0	624	OK	360	0	0	OK	0	0	0	0	66581168	69698	OK

Descripción de los campos de la tabla 2.1:

GENERACIÓN (generación neta): Es la diferencia entre la generación bruta de la batería o emplazamiento y el insumo para generar. Corresponde a la energía eléctrica que se entrega al sistema de transmisión.

INSUMO (insumo eléctrico): Es la parte de la generación bruta que los productores consumen para su funcionamiento.

ÍNDICE DE CONSUMO (índice de consumo de combustible): Es la cantidad en gramos de combustible que se consume por cada kilowatts generado en un intervalo de tiempo multiplicado por la densidad del combustible.

CONSUMO COMB (consumo de combustible): Es la cantidad en gramos de combustible que se consume por cada kilowatts generado en un intervalo de tiempo.

EXIST. REFRIG (existencia de refrigerante): Es la cantidad de refrigerante en existencia en los tanques de almacenamiento de 55 galones de la batería o emplazamiento luego de explotados los GE o después de un mantenimiento.

Capítulo 2: Diseño y Población de la Base de Datos

CONS. REFRIG (consumo de refrigerante): Es el refrigerante que se consume en litros por cada kilowatts generado en un intervalo de tiempo.

ENTRADA REFRIGERANTE (entrada de refrigerante): Es la cantidad de refrigerante que se suministra a los GE y se almacena en los tanques de 55 galones de la batería o emplazamiento

CONSUMO DE ACEITE Y LÍQUIDO REFRIGERANTE POR GRUPO: Es la cantidad de aceite y refrigerante que consume cada GE durante su explotación.

LECTURA INICIAL: Es la lectura del metro contador del transformador de fuerza de la batería en la última hora del día anterior al de explotación que se muestra en el SCADA.

BATERÍA: Es la lectura del metro-contador del transformador de fuerza de la batería en la hora actual del día presente de explotación que se muestra en el SCADA.

COMBUSTIBLE RECIBIDO: Son los datos de la factura, la fecha, la cantidad de combustible en litros, la densidad en g/cm^3 , los kilogramos (kg) y la densidad diaria promedio de combustible que se suministró el último día del mes anterior y del día del mes presente en la batería o emplazamiento.

Capítulo 2: Diseño y Población de la Base de Datos

Tabla 2.2 Informe de Indicadores principales de Moa

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	Y	Z					
	MOA								MES: SEPTIEMBRE 2008								Generación		Generación Neta		Consumo		Generación		TECLEAR AQU	
																							Bruta x Grupo		Insumo	
	Día	Generación	Lectura G1	Lectura G2	Lectura G3	Lectura G4	Lectura G5	Lectura G6	Lectura G7	Lectura G8	Grupo a Grupo	Comp.	Leet. Totalizador	Por Totalizador	Comp.	Ventiladores										
	Inicial		7159389	8619122	6547204	8338889	5053919	8094927	8544230	8658672			62772481													
5	1	110168	7175429	8632217	6563219	8338889	5070409	8111443	8560238	8674676	110168	OK	62883035	110554	OK	5332	115500	4946	115500	4946						
6	2	163735	7198911	8655699	6586700	8338889	5093896	8134308	8583704	8698148	163735	OK	63047408	164373	OK	7440	171175	6802	171175	6802						
7	3	153260	7221138	8677928	6609493	8338889	5115648	8154148	8605916	8720355	153260	OK	63200812	153404	OK	6944	160204	6800	160204	6800						
8	4	148038	7241475	8698254	6631825	8338889	5137706	8176477	8626238	8740689	148038	OK	63349314	148502	OK	6448	154486	5984	154486	5984						
9	5	202653	7263081	8729808	6665968	8339013	5155709	8210628	8657778	8772221	202653	OK	63552237	202923	OK	8928	211581	8658	211581	8658						
10	6	208822	7300059	8762556	6702889	8339013	5155709	8247430	8690478	8804894	208822	OK	63761014	208777	OK	8928	217750	8973	217750	8973						
11	7	77217	7313027	8775368	6715852	8339013	5155709	8260287	8703287	8817702	77217	OK	63838205	77191	OK	3348	80565	3374	80565	3374						
12	8	0	7313027	8775368	6715852	8339013	5155709	8260287	8703287	8817702	0	OK	63838205	0	OK	0	0	0	0	0						
13	9	6999	7314405	8776733	6717223	8339013	5157400	8261086	8703452	8817932	6999	OK	63844885	6680	OK	2170	9169	2489	9169	2489						
14	10	64034	7319677	8785572	6726764	8339013	5170010	8271479	8711592	8827171	64034	OK	63908208	63323	OK	5828	69862	6539	69862	6539						
15	11	69148	7319677	8801641	6734729	8339013	5188158	8279479	8714635	8843094	69148	OK	63976713	68505	OK	5146	74294	5789	74294	5789						
16	12	143615	7319677	8826904	6749542	8353476	5213226	8301545	8732681	8866990	143615	OK	64120254	143541	OK	7192	150807	7266	150807	7266						
17	13	142186	7319677	8844336	6770179	8375400	5234972	8323542	8753021	8885100	142186	OK	64262709	142455	OK	7192	149378	6923	149378	6923						
18	14	218842	7319677	8871197	6804047	8409204	5267431	8357532	8785091	8910890	218842	OK	64482097	219388	OK	10106	228948	9560	228948	9560						
19	15	146463	7319677	8893678	6827410	8432567	5289899	8372996	8807535	8927770	146463	OK	64628624	146527	OK	6200	152663	6136	152663	6136						
20	16	135120	7319677	8918011	6852099	8438581	5315646	8379442	8831862	8951332	135120	OK	64764225	135601	OK	6200	141320	5719	141320	5719						
21	17	114301	7319677	8933485	6858013	8453978	5344009	8379442	8859223	8973126	114301	OK	64878461	114236	OK	5022	119323	5087	119323	5087						
22	18	155839	7319677	8959274	6886043	8480533	5370953	8379442	8885004	8995866	155839	OK	65034569	156108	OK	6882	162721	6613	162721	6613						
23	19	177438	7319677	8987795	6917404	8511973	5399960	8379442	8913516	9024303	177438	OK	65212388	177819	OK	7812	185250	7431	185250	7431						
24	20	112756	7319677	9006591	6936279	8530767	5418751	8379442	8932308	9043171	112756	OK	65325336	112948	OK	5146	117902	4954	117902	4954						
25	21	141601	7319677	9030611	6960892	8555229	5442403	8379442	8953151	9067182	141601	OK	65467241	141905	OK	6386	147987	6082	147987	6082						
26	22	134601	7334422	9050978	6982127	8576599	5462791	8379442	8969263	9087566	134601	OK	65602198	134957	OK	6014	140615	5658	140615	5658						
27	23	147259	7359781	9074271	6985352	8601952	5486267	8379442	8992540	9110842	147259	OK	65749668	147470	OK	6448	153707	6237	153707	6237						
28	24	171872	7387780	9093291	6993519	8635402	5518471	8379442	9018183	9136231	171872	OK	65921647	171979	OK	7688	179560	7581	179560	7581						
29	25	155120	7414742	9093291	7020480	8662362	5543924	8379442	9042572	9160626	155120	OK	66077894	156247	OK	6696	161816	5569	161816	5569						
30	26	130515	7439603	9093291	7045338	8666667	5567030	8386067	9065953	9184005	130515	OK	66207517	129623	OK	5766	136281	6658	136281	6658						
31	27	112657	7466035	9093291	7071767	8673460	5570500	8386067	9090724	9208767	112657	OK	66319934	112417	OK	4960	117617	5200	117617	5200						
32	28	98328	7492044	9093291	7097597	8673460	5570500	8386067	9113956	9232024	98328	OK	66417975	98041	OK	4278	102606	4565	102606	4565						
33	29	93944	7522131	9093291	7109578	8673460	5570500	8386067	9139891	9257965	93944	OK	66511470	93495	OK	4154	98098	4603	98098	4603						
34	30	70227	7545919	9093291	7109578	8673460	5570500	8386067	9163114	9281181	70227	OK	66581168	69698	OK	3162	73389	3691	73389	3691						

Descripción de los campos de la tabla 2.2:

GENERACIÓN Y GENERACIÓN GRUPO A GRUPO (generación neta): Es la diferencia entre la generación bruta de la batería o emplazamiento y el insumo para generar. Corresponde a la energía eléctrica que se entrega al sistema de transmisión.

LECTURA G1, G2,..., G8: Es la lectura del metro-contador del transformador de fuerza de la batería en la última hora del día anterior y la hora actual del día presente de explotación.

LECTURA TOTALIZADOR: Es la lectura del metro-contador del transformador de fuerza de la batería en la última hora del día anterior al de explotación.

Capítulo 2: Diseño y Población de la Base de Datos

GENERACIÓN NETA POR TOTALIZADOR: Es la diferencia entre la lectura del metro contador del transformador de fuerza de la batería en la última hora del día anterior y la primera hora del día presente de explotación.

CONSUMO DE VENTILADORES: Es la energía eléctrica que se consume de los kilowatts (62 kW) de lo generado por los GE durante la explotación de la batería o emplazamiento.

GENERACIÓN BRUTA X GRUPO Y GB X GRUPOS: Se refiere a la generación de energía eléctrica de los GE medida en los metros contadores de cada grupo al cual se le suma el consumo de los ventiladores multiplicado por las horas totales.

INSUMO (insumo eléctrico): Es la parte de la generación bruta que los productores consumen para su funcionamiento.

A partir del conocimiento de los datos y su significado se realiza el diseño de la base de datos de la GEDB de Moa como caso de estudio.

2.3 Diseño de la Base de Datos.

El diseño de una base de datos responde a características esenciales: redundancia, independencia de datos, estructura o esquema de la base de datos, etc. De acuerdo con estas características se definen una base de datos como una "colección o depósitos de datos integrados, almacenados en soporte secundario (no volátil) y con redundancia controlada. Los datos, que han de ser compartidos por diferentes usuarios y aplicaciones, debe mantenerse independientes de ellos, y su definición (estructura de la base de datos) única y almacenada junto con los datos, se ha de apoyar en un modelo de datos, el cual ha de permitir captar las interrelaciones y restricciones existentes en el mundo real. Los procedimientos de actualización y recuperación, comunes y bien determinados, facilitarán la seguridad del conjunto de los datos". [Castaño; Piattini, 2006]

Los aspectos que se nombran en la definición se desarrollan a continuación en la investigación a través de la aplicación de la metodología MetdisBD que se describe en el capítulo 1.

Capítulo 2: Diseño y Población de la Base de Datos

2.3.1 Aplicación de la *Metodología MetdisBD*

Esta metodología para el diseño de base de datos está formada por cinco pasos.

2.3.1.1 *Determinación de entidades y atributos* el cual expresa lo siguiente:

- Para cada salida debe: consultar su formato y determinar datos que se calculen u obtengan a partir de otros e ir sustituyéndolos hasta llegar a los primarios
- Determinar existencia de ficheros con información normativa y/o de consulta
- Cada salida y cada fichero, tomarlos como entidad y relacionar sus atributos

Para representar adecuadamente la información referente a la GEDB se identificaron las siguientes entidades con sus respectivos atributos.

Determinación de entidades y atributos de la GEDB

- **Localidad**(Id_loc,nom_loc)
- **Fecha** (Id_fec,dia,mes,año)
- **Grupo**(Id_gru,nom_gru)
- **Totales1**(id_loc,id_fec,nom_loc,dia,mes,año,gener,lec_tota,gener_neta,cons_ven, gener_bruta_PG,insu,índ_cons,exis_refr,exis_ace)
- **Totales2**(id_loc,id_fec,id_comb,exis_comb,entr_comb,nom_loc,dia,mes,año,nom_comb,id_tipo_comb,nom_tipo_comb)
- **Loc_fecha_gr**(id_loc,id_fec,id_gru,nom_loc,id_tipo_comb,nom_tipo_comb,nom_comb,can_comb,dia,mes,año,lec_act,lec_dia_ant,cons_dia,cons_ace,cant_refr,horas_acum,nom_grup,canti_aceite,horas_trab)
- **Tipo_combustible**(id_tipo_comb,nom_tipo_comb)
- **Combustible**(id_comb,nom_comb,id_tipo_comb,nom_tipo_comb)

Capítulo 2: Diseño y Población de la Base de Datos

2.3.1.2 Normalización de entidades

La normalización de entidades es el segundo paso dentro de la metodología de diseño de base de datos. Esta expresa lo siguiente:

- Determinar las dependencias funcionales (DF).
- Normalizar cada entidad.
- Fusionar, de ser lógico, las entidades normalizadas que tengan la misma llave.

A continuación se muestra el proceso de normalización de entidades:

Dependencias funcionales atributos de la GEDB:

- (Id_localidad)nom_loc
- (Id_fecha)día,mes,año
- (Id_grupo)nom_gru
- (id_localidad,Id_fecha)dia,mes,año,nom_loc,gener,lec_tota,gener_neta,con_ven,gener_bruta_PG,insu,índ_cons,comb,exis_comb,exis_refr,entra_comb,exis_ace,id_tipo_comb, nom_tipo_comb
- (id_localidad,fecha,Id_grupo)nom_loc,dia,mes,año,lec_act,lec_dia_ante,cons_dia,cons_ace, cons_refr,hor_acum
- (id_tipo_combustible)nom_tipo_comb
- (id_combustible)comb,id_tipo_comb,nom_tipo_comb

Luego de obtener las dependencias funcionales se obtuvieron las siguientes llaves (Id_localidad,Id_fecha,Id_grupo,id_localidad,Id_fecha,id_tipo_combustible)

Capítulo 2: Diseño y Población de la Base de Datos

Proceso de normalización

1ra Forma Normal

- **Localidad**(Id_loc,nom_loc)
- **Fecha** (Id_fec,dia,mes,año)
- **Grupos**(Id_gru,nom_gru),
- **Totales1**(id_loc,id_fec,nom_loc,dia,mes,año,gener,lec_tota,gener_neta,cons_ven, gener_bruta_PG,insu,índ_cons,exis_refr,exis_ace)
- **Totales2**(id_loc,id_fec,id_comb,exis_comb,entr_comb,nom_loc,dia,mes,año,nom_comb,id_tipo_comb,nom_tipo_comb)
- **Loc_fecha_gr**(id_loc,id_fec,id_gru,nom_loc,id_tipo_comb,nom_tipo_comb,nom_comb, cant_comb,dia,mes,año,lectura_act,lec_dia_ant,cons_dia,consu_ace,cant_refr,hor_acum, nom_grup,cant_ace,horas_trab)
- **Tipo_combustible**(id_tipo_comb,nom_tipo_comb)
- **Combustible**(id_comb,nom_comb,id_tipo_comb,nom_tipo_comb)

2da Forma Normal

- **Localidad**(Id_loc,nom_loc)
- **Fecha** (Id_fec,dia,mes,año)
- **Grupo**(Id_gru,nom_gru)
- **Totales1**(id_loc,id_fec,nom_loc,dia,mes,año,gener,lec_tota,gener_neta,cons_ven, gener_bruta_PG,insu,índ_cons,exis_refri,exis_ace)
- **Totales2**(id_loc,id_fec,id_com,exis_comb,entr_comb,nom_loc,dia,mes,año,nomb_comb,id_tipo_comb,nom_tipo_comb)
- **Fecha1** (Id_fec,dia,mes,año)

Capítulo 2: Diseño y Población de la Base de Datos

- **Localidad1** (id_loc, nom_loc)
- **Fecha2** (Id_fec, dia, mes, año)
- **Localidad2** (id_loc, nom_loc)
- **Combustible3**(id_comb, nom_comb, id_tipo_comb, nom_tipo_comb)
- **Loc_fecha_gr**(id_loc, id_fec, Id_gru, id_comb, id_tipo_comb, nomb_tipo_comb, cant_comb, lect_act, lec_dia_ant, cons_dia, cons_ace, cant_refr, horas_acum, cant_ace, horas_trab, nom_comb)
- **Localidad3** (id_loc, nom_loc)
- **Fecha3** (id_fec, dia, mes, año)
- **Grupo2** (Id_gru, nom_gru)

3ra Forma Normal

- **Localidad**(Id_loc, nom_loc)
- **Fecha** (Id_fec, dia, mes, año)
- **Grupo**(Id_gru, nom_gru)
- **Totales1**(id_loc, id_fec, gener, lec_tota, gener_neta, cons_ven, gener_bruta_pg, Insu, ind_cons, exis_refri, exis_ace)
- **Totales2**(id_loc, id_fec, id_comb, exis_comb, entr_comb, id_tipo_comb, nom_tipo_comb)
- **Fecha4** (Id_fec, dia, mes, año)
- **Localidad4** (id_loc, nom_loc)
- **Fecha4** (Id_fec, dia, mes, año)
- **Localidad4** (id_loc, nom_loc)
- **Combustible4**(id_comb, nom_comb, id_tipo_comb)
- **Tipo_combustible4** (id_tipo_comb, nom_tipo_comb)
- **Tipo_combustible2**(id_tipo_comb, nom_tipo_comb)

Capítulo 2: Diseño y Población de la Base de Datos

- **Loc_fecha_gr**(id_loc,id_fec,ld_gru,id_comb,id_tipo_comb,can_comb,lec_act,lec_dia_ante,cons_dia,cons_ace,can_refr,horas_acum,can_ace,horas_trab)
- **Tipo_combustible3**(id_tipo_comb,nom_tipo_comb)
- **Combustible2**(id_comb,nom_comb)

División de la entidad loc_fecha_gr

- **Loc_fecha_gr**(id_loc,id_fec,ld_gru,id_comb,id_tipo_comb,can_comb,lec_act,lec_dia_ante,cons_dia,cons_ace,cant_refr,horas_acum,can_ace,horas_trab)
- **Mantenimiento**(id_localidad,id_fecha,ld_grupo,cons_ace)
- **Lecturas**(id_loc,id_fec,ld_gru,horas_acum, lec_act, lec_dia_ant,cons_dia)
- **Consumo_aceite_refr**(id_loc,id_fec,ld_gru,cant_refr,can_ace)

Luego de realizar el proceso de normalización de las entidades iniciales para eliminar la redundancia de datos, se obtienen las siguientes entidades, las cuales se encuentran en tercera forma normal.

Entidades finales resultantes

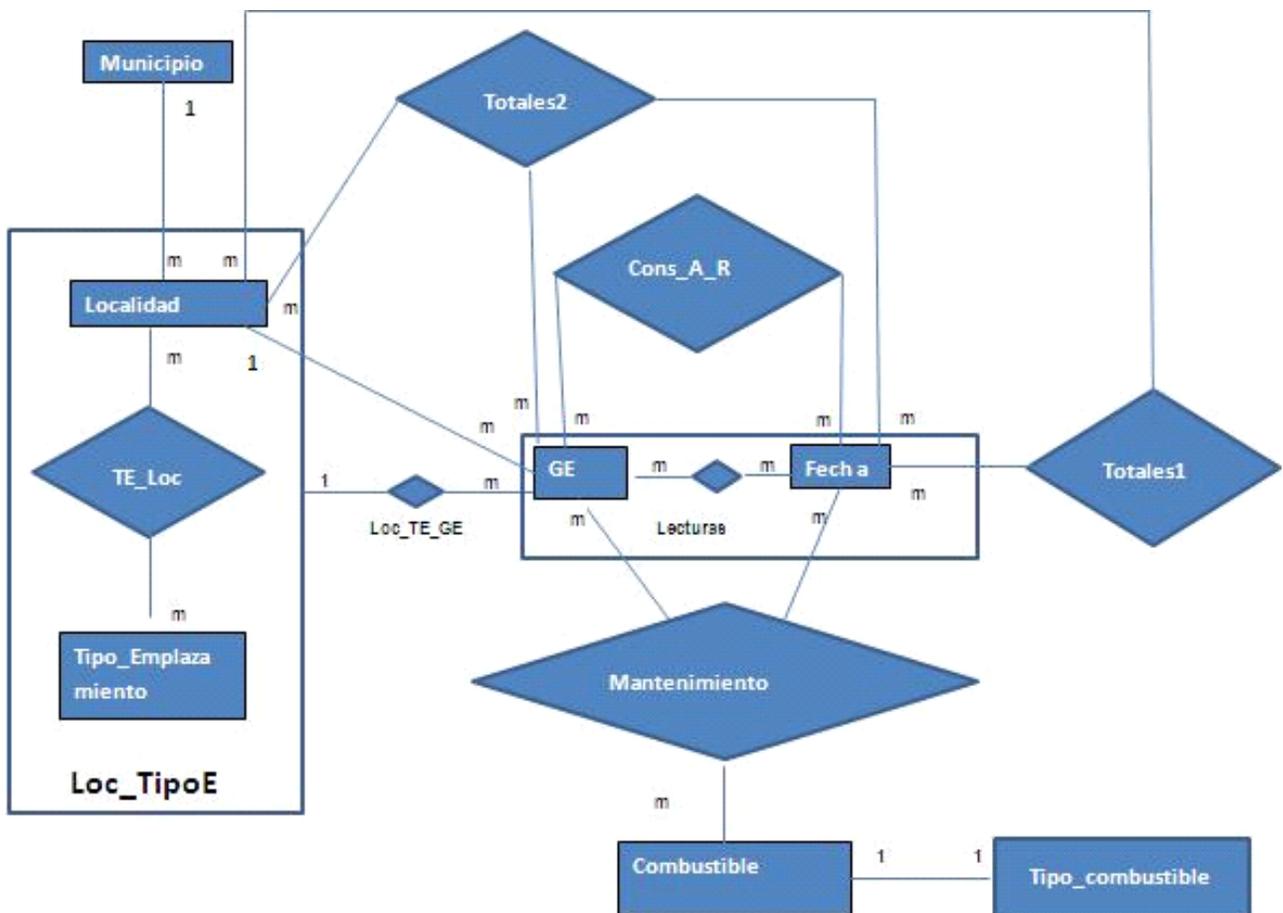
- **Localidad**(ld_loc,nom_loc)
- **Fecha** (ld_fec,dia,mes,año)
- **Grupo**(ld_gru,nom_gru)
- **Totales1**(id_loc,id_fec,gener,lec_tota,gener_neta,cons_ven,gener_bruta_PG,insu,índ_cons,exis_refr ,exis_ace)
- **Totales2**(id_loc,id_fec,id_comb,exis_comb,entr_comb)
- **Mantenimiento**(id_localidad,id_fecha,ld_grupo,cons_ace)
- **Lecturas**(id_loc,id_fec,ld_gru,horas_acum,lec_act,lec_dia_ant,cons_dia,horas_trabajadas_a_dia)
- **Consumo_aceite_refr**(id_loc,id_fec,ld_gru,can_refr,can_ace)
- **Combustible**(id_comb,nomb_comb,id_tipo_comb)
- **Tipo_combustible**(id_tipo_comb,nom_tipo_comb)

Capítulo 2: Diseño y Población de la Base de Datos

2.3.1.3 Determinación de Relaciones (DER)

La determinación de relaciones entre las entidades es el tercer paso dentro de la metodología de diseño de base de datos. Esta consiste en:

- Si hay alguna entidad cuya llave es combinación de llaves de otras entidades, generalmente, es porque representa una relación entre ellas (m: n)
- Analizar cada entidad con las restantes para determinar si existe relación, determinando tipo de proyección.
- Ir confeccionando el DER
- Analizar entidades agregadas, entidades débiles y sus relaciones con otras



Capítulo 2: Diseño y Población de la Base de Datos

Figura 2.1 Diagrama de la determinación de relaciones

En el DER anterior (*ver la figura 2.1*) se representa el funcionamiento de los Grupos electrógenos. Estos se encuentran en un municipio del que se conoce su código, nombre, cantidad de habitantes y km cuadrados. En un municipio puede haber varias localidades de las que se conocen sus identificadores, y el nombre. Estas tienen relación con el tipo de emplazamiento del cual se conoce su identificador y el nombre. La localidad y el tipo de emplazamiento forman una entidad agregada la cual se llama Loc_TipoE. En una Loc_TipoE pueden haber varios GE y estos pueden estar en una Loc_TipoE. También un GE puede encontrarse en una sola localidad pero en una localidad pueden existir varios GE. Del GE se conoce su identificador y el nombre. Un GE puede tener varios días, meses y años. Un GE en una fecha dada, junto a un combustible, genera un mantenimiento. Del mantenimiento se conoce sus identificadores, la cantidad de aceite y la cantidad de combustible. Del combustible se conoce, su identificador y el nombre del combustible. Un combustible tiene un tipo de combustible. De este se conoce su identificador y el nombre. Un GE y una fecha forman una entidad agregada llamada lectura, de esta se conoce sus identificadores, las horas acumuladas de mantenimiento, las horas trabajadas al día, la lectura actual, las horas acumuladas y la lectura del día anterior. Un GE y una fecha generan el consumo de aceite refrigerante, de esto se conoce sus identificadores, el aceite y el refrigerante. La entidad totales 1, es generada por una localidad y una fecha. De esta se conoce sus identificadores, la generación, la lectura totalizadora, la generación neta, el consumo ventiladores, la generación bruta por grupos, el insumo, el índice de consumo, la existencia de refrigerante, y la existencia de aceite. La entidad totales 2, es generada por una localidad, un GE y una fecha. De esta tabla se conocen: sus identificadores, la existencia de combustible, la entrada de combustible, el identificador de tipo de combustible y el nombre del tipo de combustible.

2.3.1.4 Obtención del modelo lógico global de los datos

El cuarto paso por la que está formada la metodología de diseño de base de datos es la obtención del modelo lógico global de los datos.

Capítulo 2: Diseño y Población de la Base de Datos

Se deben seguir los pasos para llevar del DER al modelo lógico global de los datos. Si en el paso anterior no se modifica el DER obtenido, entonces las tablas que se obtienen en este paso, en general, son las mismas que las obtenidas en el segundo paso (normalizar las entidades).

Al realizar el DER se introdujeron dos nuevos conceptos; municipio y tipo_emplazamiento, por tanto, el modelo global de los datos ahora contempla todas las entidades resultantes en el paso normalización de entidades, más las entidades municipio y tipo_emplazamiento.

- **Municipio** {id_mun, nom_mun, km_cua, can_hab, tip_emp}
- **Tipo Emplazamiento** {tip_emp, nom_te}
- **localidad**(ld_loc, nom_loc, id_mun)
- **fecha** (ld_fec, dia, mes, año)
- **grupo**(ld_gru, nom_gru)
- **totales1**(id_loc, ld_fec, gener, lec_tota, gener_neta, cons_ven, gener_bruta_PG, insu, ind_cons, exis_refr, exis_ace)
- **totales2**(id_loc, ld_fec, id_comb, exis_comb, entr_comb)
- **Mantenimiento**(id_localidad, id_fecha, ld_grupo, consumo_aceite,)
- **Lecturas**(id_loc, id_fec, ld_gru, horas_acum, lec_act, lec_dia_ant, cons_dia)
- **Consumo_aceite_refr**(id_loc, id_fec, ld_gru, can_refr, can_ace)
- **Combustible**(id_comb, nomb_comb, id_tipo_comb)
- **Tipo_combustible**(id_tipo_comb, nom_tipo_comb)

La definición de las entidades anteriores es:

- **Municipio:** Es la entidad donde se encuentran ubicadas las localidades.
- **Localidad:** Es el lugar donde se encuentran los grupos electrógenos.
- **Grupo Electrónico:** "Equipos formados por un motor primario de combustión interna y un generador sincrónico de corriente alterna acoplado mecánicamente para

Capítulo 2: Diseño y Población de la Base de Datos

producir energía eléctrica. Para su funcionamiento puede consumir fuel, diesel o gas natural. Estos dispositivos pueden estar sincronizados al Sistema Electroenergético Nacional (SEN) para solucionar los déficit de potencia y contingencias, aislados (pertenecientes a la Unión Eléctrica [UNE]) para suministrar energía eléctrica en lugares donde no llegue la red eléctrica nacional o de emergencia, ubicados en una entidad para operar en caso de fallo, desconexión o insuficiencias del fluído eléctrico en la red nacional” [ONE 2009].

- **Tipo Emplazamiento:** es la categoría que tiene el Grupo Electrónico, ya sea en paralelo o isla.
- **Fecha:** es la fecha donde queda registrado en el día, mes y año las lecturas, el mantenimiento, el consumo de aceite refrigerante y los totales que se le hacen a un grupo electrónico.
- **Combustible:** es el combustible que utilizan los grupos electrónicos.
- **Tipo Combustible:** representa los diferentes combustibles que se emplean en los grupos electrónicos.
- **Lectura:** Es la lectura que se le realizan a los Grupos Electrónicos midiendo diferentes parámetros.
- **Mantenimiento:** representa las reparaciones que se le realizan a los Grupos Electrónicos.
- **Consumo Aceite Refrigerante:** representa el aceite y el refrigerante que utilizan los Grupos Electrónicos.
- **Totales1:** es donde quedan almacenadas las generaciones, el insumo, la existencia de aceite y refrigerante, y la lectura totalizadora que se registran en una localidad en una fecha determinada.
- **Totales2:** es donde quedan almacenadas las entradas y la existencia de combustible que se registran en una localidad en una fecha determinada.

Capítulo 2: Diseño y Población de la Base de Datos

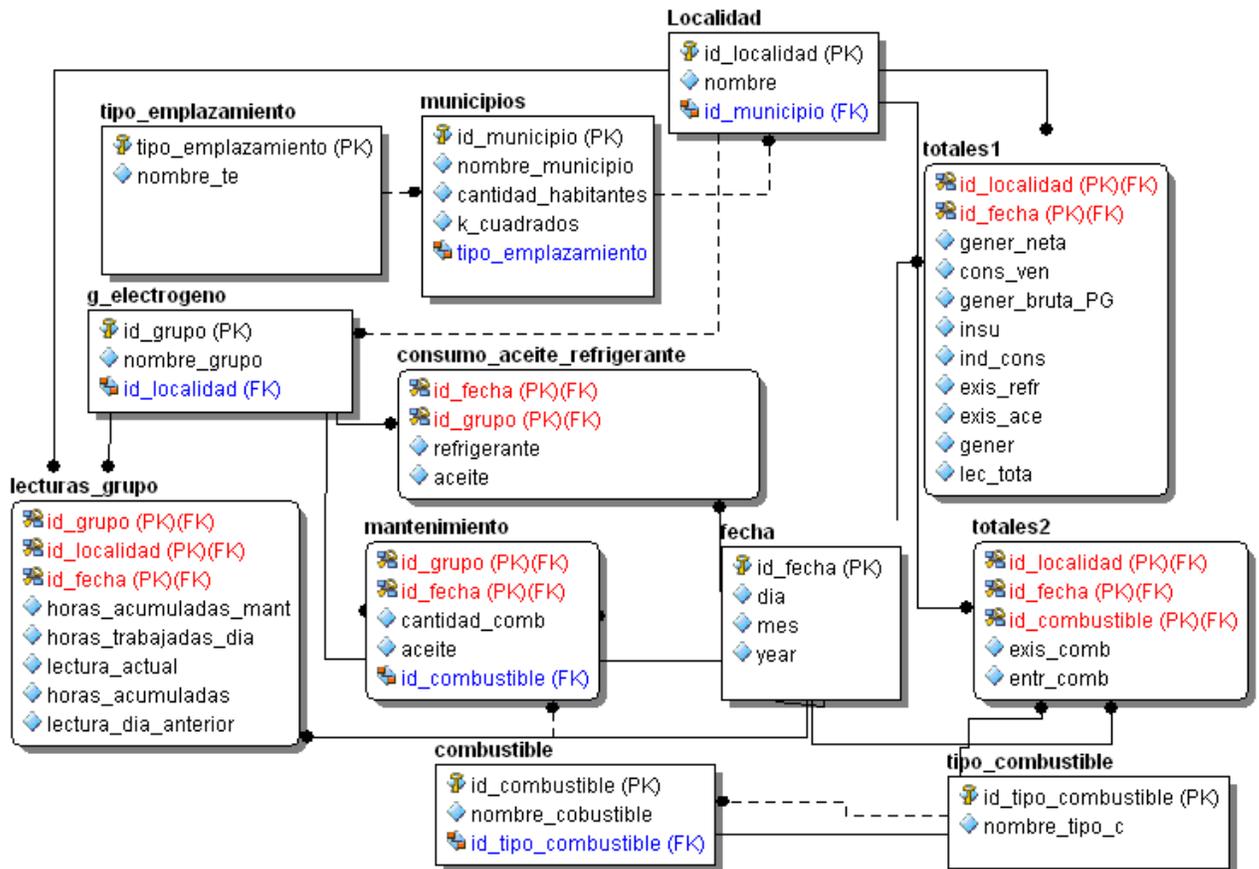
2.3.1.5 Diseño Físico de la Base de Datos

El diseño físico de la base de datos es el quinto y último paso dentro de la metodología para el diseño de la base de datos. En este paso hay que tener en cuenta:

- Aplicaciones a realizar sobre los datos.
- Características particulares del SGBD.

Ejemplos de estos dos aspectos a tener en cuenta:

- Ficheros índices
- Ficheros para reportes
- Considerar la inclusión de campos que sean secundarios debido a la cantidad de veces que habría que calcularlos en una determinada aplicación.
- Separar un fichero en varios, debido a la cantidad de campos y lo que se accesa cada vez



Capítulo 2: Diseño y Población de la Base de Datos

Figura 2.2 Diagrama del diseño físico de la base de datos

La figura 2.2 muestra el modelo físico de los datos, en el mismo la tabla GE es la que muestra el nombre del GE. La tabla totales es la que contiene toda la información que generan la batería de grupos electrógenos. La tabla mantenimiento es la encargada de controlar la cantidad de combustible y de aceite que utiliza un GE en un mantenimiento. A su vez, la Tabla Consumo Aceite Refrigerante, es la que controla el gasto de estos dos líquidos. La Tabla Lectura, tiene como función, llevar las horas de trabajo de los GE y las lecturas que se le realizan a este por parte de un especialista.

Después de culminado el análisis del diseño de la base de datos, se aplica la metodología HEFESTO descrita en el capítulo 1, para guiar el procedimiento del poblado de la base de datos.

2.4 Poblado de la Base de Datos

Para realizar el poblado de la base de datos, los datos que son generados por los GE son recogidos en una HECE. A estos datos es necesario exportarlos para la base de datos creada en SQL server (2005-2008). Para realizar este proceso primero es necesario transformar HECE para eliminar los errores que esta tiene.

Luego se procede a utilizar la herramienta *Spoon* de *Pentaho* para el proceso de ETL (extracción, transformación y carga) la cual tomará los datos antes mencionados y los carga en la base de datos en SQL server (versus 2005-2008).

Primeramente se debe realizar la carga inicial a la base de datos, poblando el modelo que hemos construido anteriormente. Para lo cual debemos llevar adelante una serie de tareas básicas, tales como limpieza de datos, calidad de datos, procesos ETL, etc.

Se debe evitar que la base de datos sea cargada con valores faltantes o anómalos, así como también, se deben establecer condiciones y restricciones para asegurar que solo se utilicen los datos de interés [Kimball; Reeves; Ross; Thornthwaite, 2009].

2.4.1 Arquitectura de la Solución Propuesta

En la arquitectura propuesta para el poblado de la Base de Datos de los Indicadores Técnico-Productivos de los GEDB se evidencia que la fuente de datos utilizada es la

Capítulo 2: Diseño y Población de la Base de Datos

Hoja Electrónica de Cálculos de Excel (HECE) donde se encuentra toda la información de los GED.

Con ayuda de la herramienta *Spoon* de *Pentaho*, estos datos son extraídos de sus fuentes originales y después de realizarles algunas transformaciones fundamentalmente de limpieza son cargados en la base de datos de SQLServer. Todo este proceso se realiza mediante el protocolo de comunicación TCP/IP, como se muestra en la figura 2.3.

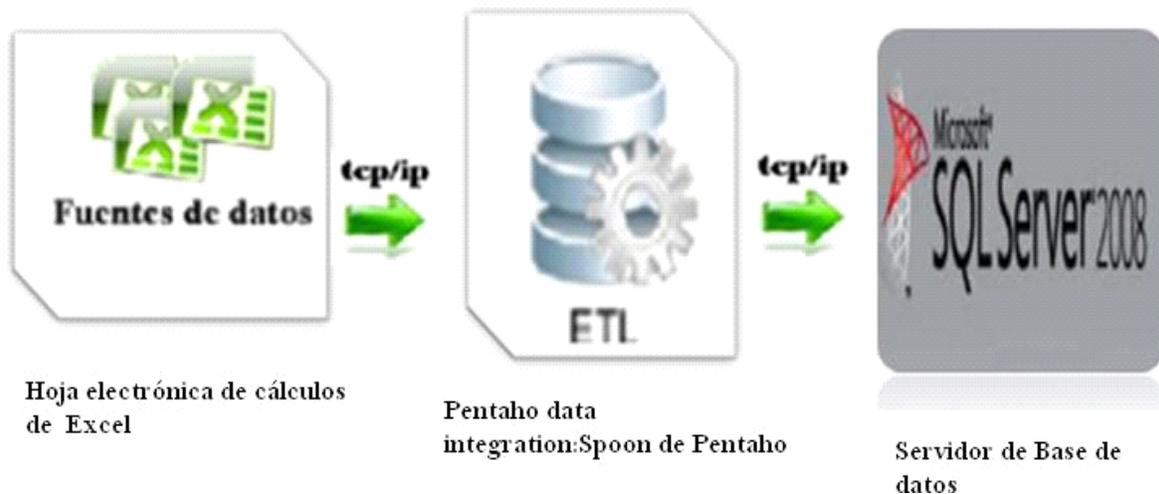


Figura 2.3 Diagrama de la arquitectura para poblar la base de datos

2.4.2 Procesos de Extracción, Transformación y Carga.

A la hora de extraer la información, un gran problema que se presenta es el de la diferencia de fuentes, sin embargo se ha visto a lo largo de la definición de las entidades que intervienen en la Base de Datos que esto no fue problema, dado que toda la información llega de una sola fuente en una HECE.

2.4.2.1 Proceso de Extracción.

Los patrones de diseño (*design patterns*) son la base para la búsqueda de soluciones a problemas comunes en el desarrollo de software y otros ámbitos referentes al diseño de interacción o interfaces. Puede encontrarse información sobre este tema en [Gamma, 1995]

Capítulo 2: Diseño y Población de la Base de Datos

Por otro lado, dado que la HECE es administrada a través de la misma herramienta y en forma centralizada, se puede decir que constituye la única fuente de datos para realizar la carga a la base de datos. En el algoritmo propuesto existe solo un tipo de extracción y carga; la extracción y carga inicial.

1. Extracción y carga inicial: En la que se extraen todos los valores a las tablas existentes en el momento que comienza a operar la base de datos, así como todo el histórico de movimientos existente en la Base de Datos de los Indicadores Técnico-Productivos de los GEDB.

Una vez definida las equivalencias de los atributos tanto en la HECE como en la base de datos se procede a programar una regla única de carga cuya función es la de cargar la información inicial a las tablas de la base de datos.

2.4.2.2 Proceso de Transformación

Las reglas a aplicar en este proceso de transformación son directas en su mayoría, es decir, los datos son tomados directamente del HECE y son cargados en la base de datos (BD), tal como se muestran las correspondencias de los atributos del sistema OLPT a la BD.

Seguidamente se mostrará el resumen de las fuentes de donde se extraerán los datos y se muestra el detalle de las decisiones de transformación en el diseño de la BD se muestran en las tablas 2.3 -2.9.

Tabla 2.3 Transformación para cargar la tabla TOTALES1

Atributo	Origen de datos	Tipo	Transformación	Observaciones
id_fecha	Excel para exportar .xls.Totales1	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro
id_localidad	Excel para	número	Convertirlo al tipo	Comprobar que pertenece a un

Capítulo 2: Diseño y Población de la Base de Datos

	exportar .xls.Totales1		integer(10)	municipio
gener	Excel para exportar .xls.Totales1	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro
lec_tota	Excel para exportar. xls.Totales1	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro
gener_neta	Excel para exportar .xls.Totales1	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro
cons_ven	Excel para exportar .xls.Totales1	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro
gener_bruta_pg	Excel para exportar .xls.Totales1	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro
Insu	Excel para exportar .xls.Totales1	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro

Capítulo 2: Diseño y Población de la Base de Datos

índ_cons	Excel para exportar .xls.Totales1	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro
exis_refri	Excel para exportar .xls.Totales1	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro
exis_ace	Excel para exportar .xls.Totales1	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro

Tabla 2.4 Transformación para cargar la tabla TOTALES 2

Atributo	Origen de datos	Tipo	Transformación	Observaciones
id_loc,	Excel para exportar .xls.Totales2	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro
Id_fec	Excel para exportar .xls.Totales2	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro
id_comb	Excel para exportar .xls.Totales2	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro
exis_comb	Excel para exportar .xls.Totales2	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro
entr_comb	Excel para exportar	número	Convertirlo al	Si es nulo, lo

Capítulo 2: Diseño y Población de la Base de Datos

id_tipo_comb	.xls.Totales2 Excel para exportar	número	tipo integer(10) Convertirlo al tipo integer(10)	ignoro Si es nulo, lo ignoro
nom_tipo_comb	.xls.Totales2 Excel para exportar	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro

Tabla 2.5 Transformación para cargar la tabla Tipo_emplazamiento

Atributo	Origen de datos	Tipo	Transformación	Observaciones
tipo_emplazamiento	Excel para exportar .xls.tipo_emplazamiento	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro
nombre_te	Excel para exportar .xls.tipo_emplazamiento	texto	Convertirlo al tipo varchar(10)	Si es nulo, lo ignoro

Tabla 2.6 Transformación para cargar la tabla Grupo_electrógeno

Atributo	Origen de datos	Tipo	Transformación	Observaciones
id_grupo	Excel para exportar .xls.grupo_electrógeno	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro
nombre_grupo	Excel para exportar .xls.grupo_	texto	Convertirlo al tipo varchar(10)	Si es nulo, lo ignoro

Capítulo 2: Diseño y Población de la Base de Datos

id_municipio	electrógeno Excel para exportar .xls.grupo_ electrógeno	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro
--------------	--	--------	------------------------------------	--------------------------

Capítulo 2: Diseño y Población de la Base de Datos

Tabla 2.7 Transformación para cargar la tabla Consumo_aceite_refrigerante

Atributo	Origen de datos	Tipo	Transformación	Observaciones
id_fecha	Excel para exportar .xls.cons_ace_refr	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro
id_grupo	Excel para exportar .xls.cons_ace_refr	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro
aceite	Excel para exportar .xls.cons_ace_refr	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro
refrigerante	Excel para exportar .xls.cons_ace_refr	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro

Tabla 2.8 Transformación para cargar la tabla Combustible

Atributo	Origen de datos	Tipo	Transformación	Observaciones
id_combustible	Excel para exportar .xls.combustible	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro
nombre_cobustible	Excel para exportar .xls.combustible	texto	Convertirlo al tipo varchar(10)	Si es nulo, lo ignoro

Capítulo 2: Diseño y Población de la Base de Datos

Tabla 2.9 Transformación para cargar la tabla Localidad

Atributo	Origen de datos	Tipo	Transformación	Observaciones
id_localidad	Excel para exportar .xls.localidad	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro
id_municipio	Excel para exportar .xls.localidad	número	Convertirlo al tipo integer(10)	Si es nulo, lo ignoro
nombre	Excel para exportar .xls.localidad	texto	Convertirlo al tipo varchar(10)	Si es nulo, lo ignoro

La información almacenada en una base de datos, por ser procesada, en principio no es actualizable, solo en casos excepcionales. Los casos de interés se presentan cuando ocurren cambios en atributos de alguna tabla que en definitiva son los que tienen una naturaleza textual [Kimball, 2005].

2.4.2.3 Transformaciones y Cargas Realizadas con SPOON

La transformación dentro de la herramienta *Spoon* de *Pentaho* no es más que el proceso de cambio al que es sometido cualquier tipo de dato, partiendo desde su origen o fuente hasta llegar a su destino final.

Capítulo 2: Diseño y Población de la Base de Datos

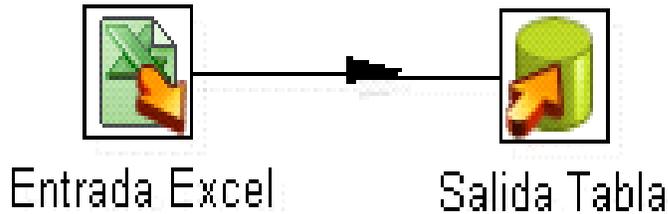
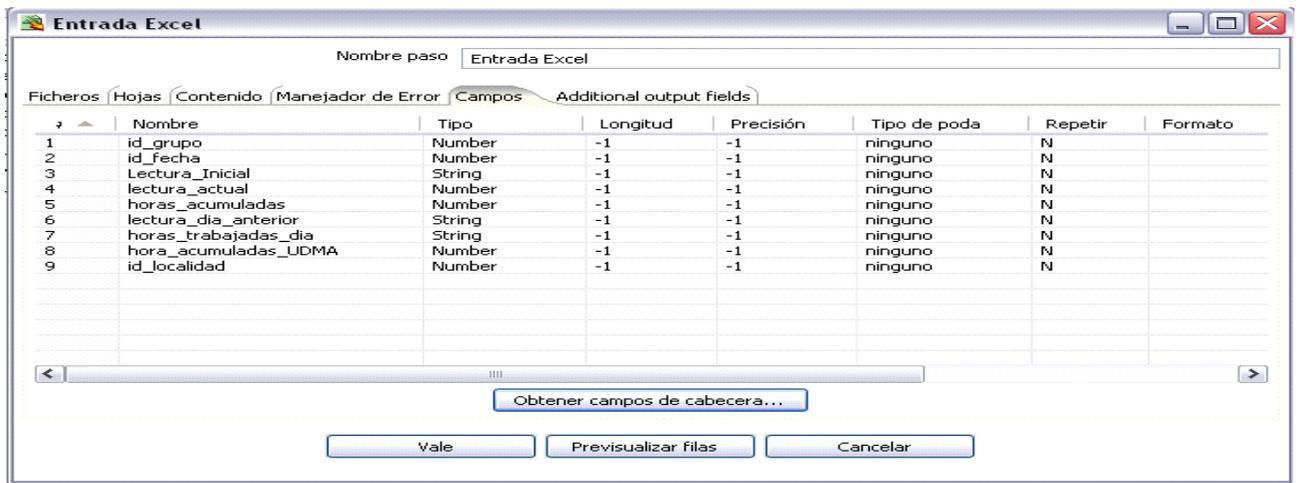


Figura 2.4 Diagrama de transformación para cargar la tabla2.3 TOTALES 1

En la figura 2.4 se representa una transformación. La entrada Excel simboliza los datos que se encuentran almacenados en la HECE la cual será transformada y sus datos serán cargados en la salida tabla, o sea en al servidor de base de datos SQL Server (2005-2008).



Capítulo 2: Diseño y Población de la Base de Datos

Figura 2.5 Ventana de entrada de Excel para la obtención de los datos de entrada

Examine preview data									
Rows of step: Entrada Excel (240 rows)									
	id_...	id_f...	Lectura_...	lectura_a...	horas_...	lectura_...	horas_trabajadas_dia	hora_acumuladas_UDMA	id_lokali...
1	1,0	1,0		7159389,0	5737,0	7159389,0	12,0	5725,0	1,0
2	1,0	2,0	7159389,0	7175429,0	5754,0	7159389,0	17,0	5725,0	1,0
3	1,0	3,0	7159389,0	7198911,0	5771,0	7175429,0	17,0	5725,0	1,0
4	1,0	4,0	7159389,0	7221138,0	5786,0	7198911,0	15,0	5725,0	1,0
5	1,0	5,0	7159389,0	7241475,0	5800,0	7221138,0	14,0	5725,0	1,0
6	1,0	6,0	7159389,0	7263081,0	5824,0	7241475,0	24,0	5725,0	1,0
7	1,0	7,0	7159389,0	7300059,0	5833,0	7263081,0	9,0	5725,0	1,0
8	1,0	8,0	7159389,0	7313027,0	5833,0	7300059,0	0,0	5725,0	1,0
9	1,0	9,0	7159389,0	7313027,0	5836,0	7313027,0	3,0	5725,0	1,0
10	1,0	10,0	7159389,0	7314405,0	5846,0	7313027,0	10,0	5725,0	1,0
11	1,0	11,0	7159389,0	7319677,0	5846,0	7314405,0	0,0	5725,0	1,0
12	1,0	12,0	7159389,0	7319677,0	5846,0	7319677,0	0,0	5725,0	1,0
13	1,0	13,0	7159389,0	7319677,0	5846,0	7319677,0	0,0	5725,0	1,0
14	1,0	14,0	7159389,0	7319677,0	5846,0	7319677,0	0,0	5725,0	1,0
15	1,0	15,0	7159389,0	7319677,0	5846,0	7319677,0	0,0	5725,0	1,0
16	1,0	16,0	7159389,0	7319677,0	5846,0	7319677,0	0,0	5725,0	1,0
17	1,0	17,0	7159389,0	7319677,0	5846,0	7319677,0	0,0	5725,0	1,0
18	1,0	18,0	7159389,0	7319677,0	5846,0	7319677,0	0,0	5725,0	1,0
19	1,0	19,0	7159389,0	7319677,0	5846,0	7319677,0	0,0	5725,0	1,0
20	1,0	20,0	7159389,0	7319677,0	5846,0	7319677,0	0,0	5725,0	1,0
21	1,0	21,0	7159389,0	7319677,0	5846,0	7319677,0	0,0	5725,0	1,0
22	1,0	22,0	7159389,0	7319677,0	5856,0	7319677,0	10,0	5725,0	1,0
23	1,0	23,0	7159389,0	7334422,0	5873,0	7319677,0	17,0	5725,0	1,0
24	1,0	24,0	7159389,0	7359781,0	5892,0	7334422,0	19,0	5725,0	1,0
25	1,0	25,0	7159389,0	7387780,0	5910,0	7359781,0	18,0	5725,0	1,0
26	1,0	26,0	7159389,0	7414742,0	5927,0	7387780,0	17,0	5725,0	1,0
27	1,0	27,0	7159389,0	7439603,0	5945,0	7414742,0	18,0	5725,0	1,0
28	1,0	28,0	7159389,0	7466035,0	5963,0	7439603,0	18,0	5725,0	1,0
29	1,0	29,0	7159389,0	7492044,0	5982,0	7466035,0	19,0	5725,0	1,0
30	1,0	30,0	7159389,0	7522131,0	5999,0	7492044,0	17,0	5725,0	1,0
31	2,0	1,0	8619122,0	8632217,0	6215,0	8619122,0	10,0	6205,0	1,0
32	2,0	2,0	8619122,0	8655699,0	6232,0	8632217,0	17,0	6205,0	1,0
33	2,0	3,0	8619122,0	8677928,0	6249,0	8655699,0	17,0	6205,0	1,0
34	2,0	4,0	8619122,0	8698254,0	6263,0	8677928,0	14,0	6205,0	1,0
35	2,0	5,0	8619122,0	8729808,0	6287,0	8698254,0	24,0	6205,0	1,0
36	2,0	6,0	8619122,0	8762556,0	6311,0	8729808,0	24,0	6205,0	1,0

Figura 2.6 Ventana pre-visualización de las filas de la HECE

Las figuras 2.5 y 2.6 muestran la configuración necesaria para obtener los datos de entrada en el paso Entrada Excel además la pre-visualización de las filas leídas de la HECE.

Capítulo 2: Diseño y Población de la Base de Datos

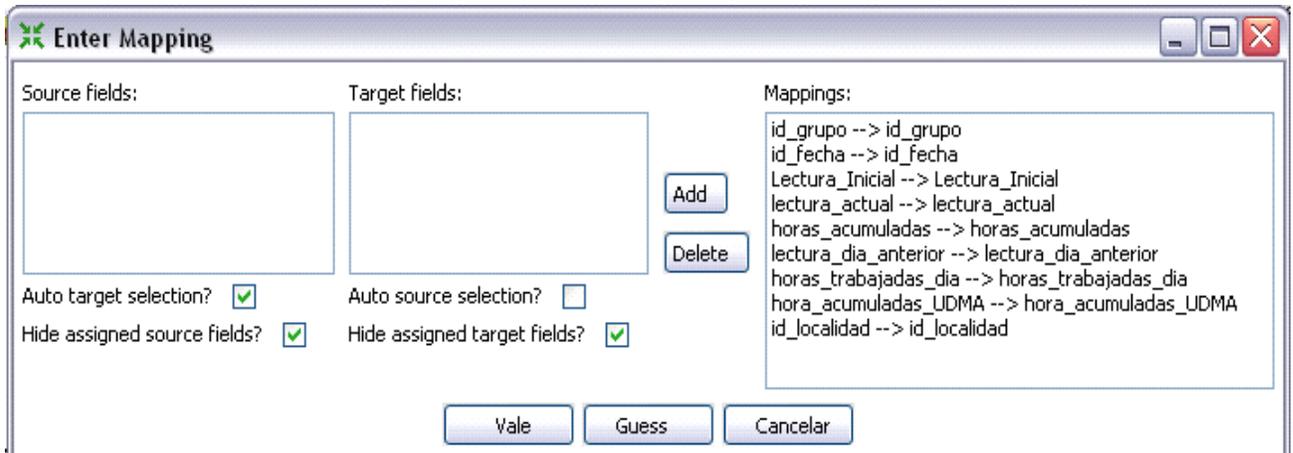
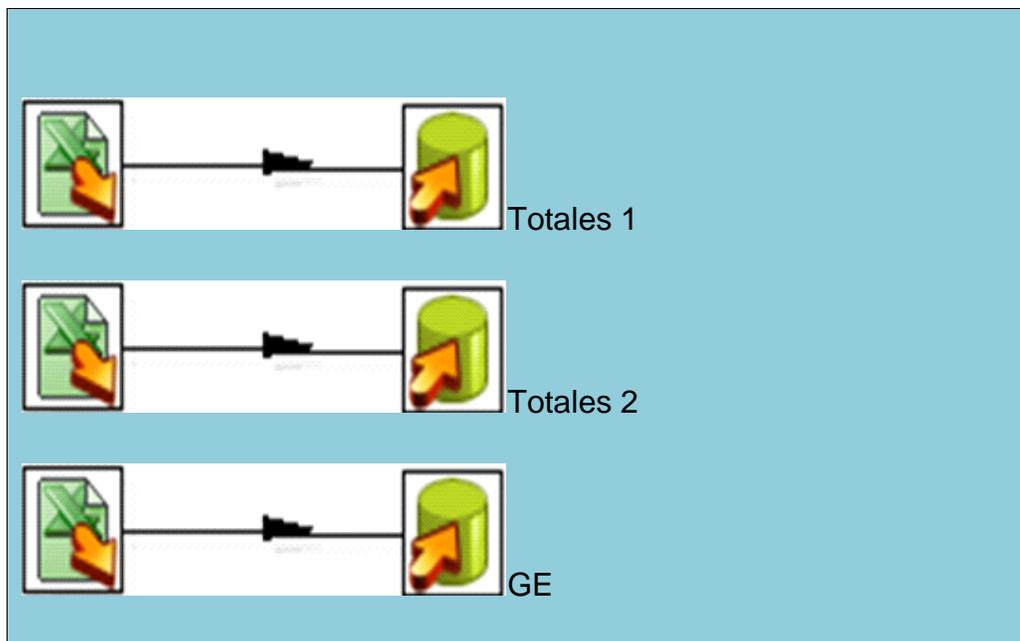


Figura 2.7 Ventana de mapeo de campos

En la figura 2.7 se muestra el proceso que se realiza con la herramienta Spoon de Pentaho para mapear los campos de la hoja electrónica de cálculos de Excel con los correspondientes campos creados en la tabla de SQLServer.

2.4.3 Trabajos

Los trabajos no son más que el proceso que nos permite agrupar y ejecutar un conjunto de transformaciones desde su origen hasta su destino.



Capítulo 2: Diseño y Población de la Base de Datos

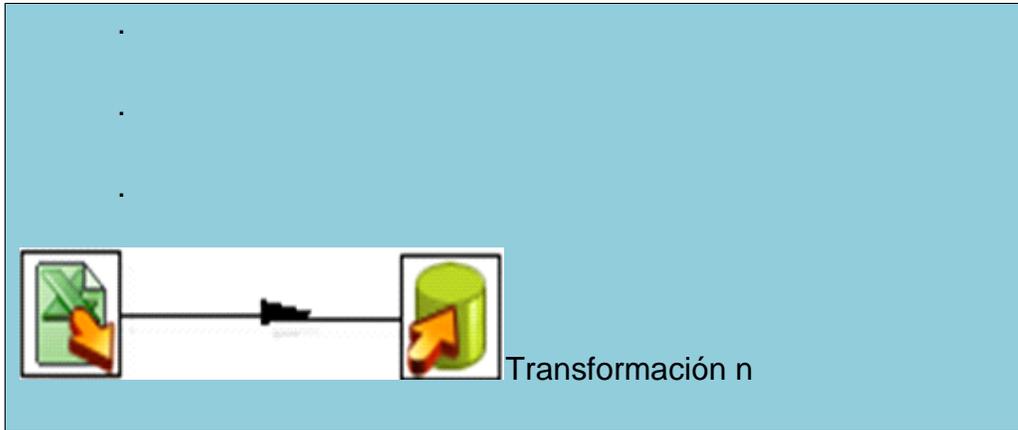


Figura 2.8 Diagrama de trabajo

La figura 2.8 ilustra el trabajo creado para poblar la base de datos, en el mismo se agrupan las transformaciones realizadas sobre los datos de los indicadores técnico-productivos de los GEDB, los cuales se encuentran en las HECE.

2.5 Estructura Física de la Base de Datos

La estructura física de la base de datos muestra la estructura final que tendrá esta dentro del SGBD SQLServer (2005-2008). Estará compuesta por once tablas las cuales guardarán toda la información referente a los indicadores técnico-productivos de los GEDB, almacenando todos los controles diarios individuales y colectivos que se le realizan. Vea la figura 2.9.

Capítulo 2: Diseño y Población de la Base de Datos

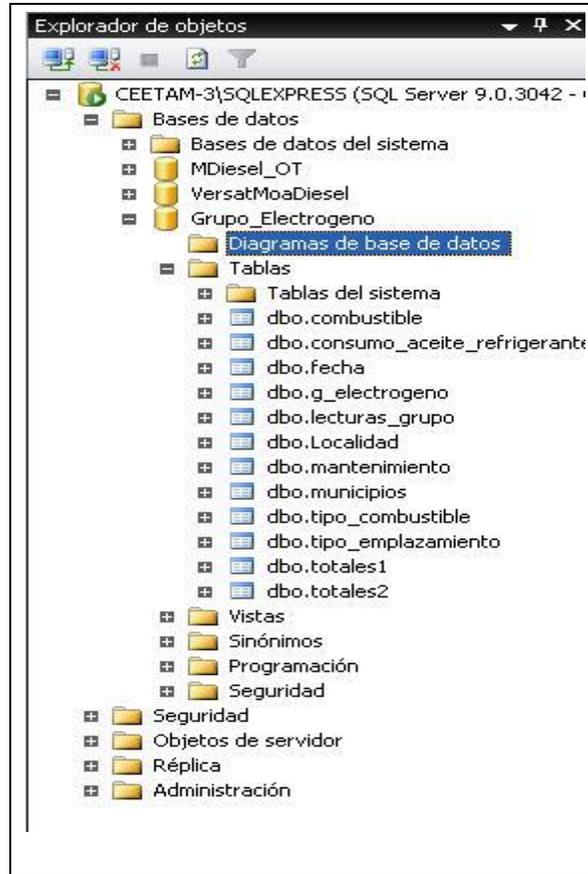


Figura 2.9 Estructura Física de la base de datos en SQLServer (2005-2008)

Capítulo 2: Diseño y Población de la Base de Datos

2.6 Conclusiones del Capítulo

- Se determinaron las entidades que permiten representar adecuadamente la información referente a los Grupos Electrógenos en batería así como las dependencias funcionales entre los campos representativos del negocio.
- Se obtuvo el diagrama de entidad de relación donde se muestra las relaciones entre las entidades.
- Se realizó el modelo lógico global de los datos siendo este la antesala del modelo físico.
- Se desarrolló el modelo físico por su gran importancia. Este representa el diseño final de la base de datos. Es en él donde quedarán definidas las tablas y sus atributos.
- Se obtuvo el algoritmo para poblar la base de datos
- Se desarrolló la Base de Datos de los Indicadores Técnicos Productivos de los GED de Moa obteniéndose un producto informático en el que se aplican los resultados de la investigación.

CAPÍTULO 3. ESTUDIO DE FACTIBILIDAD

3.1 Introducción

En los momentos actuales es muy importante tener en cuenta que a la hora de desarrollar un proyecto, el mismo cuente con eficacia y eficiencia a la hora de su implementación, y por supuesto, se hace preciso efectuar y evaluar la factibilidad antes de su elaboración. En general, los productos informáticos no están exentos de posibles riesgos en la concepción del proyecto, por lo que es válido minimizar de forma razonable recursos humanos, materiales y financieros. De ahí que sea de vital importancia estimar la relación costo-beneficio, así como el esfuerzo, capital humano y el tiempo de desarrollo que se demanda en la ejecución de los mismos.

En este capítulo se expone el estudio de factibilidad del proyecto, centrado en estimaciones de esfuerzo humano, tiempo de desarrollo para su ejecución y costo. Se estiman los beneficios tangibles e intangibles que representan para el sistema propuesto, un análisis de costos y beneficios.

3.2. Evaluación Costo-Beneficio

La mayoría, por no decir todos los proyectos de informática, son evaluados según el criterio de costo-beneficio. Esta metodología plantea que la conveniencia de la ejecución de un proyecto se determina por la observación conjunta de dos factores:

1. El costo, que involucra la implementación de la solución informática, adquisición y puesta en marcha del sistema hardware/software y los costos de operación asociados.
2. La efectividad, que se entiende como la capacidad del proyecto para satisfacer la necesidad, solucionar el problema o lograr el objetivo para el cual se ideó, es decir, un proyecto será más o menos efectivo con relación al mayor o menor cumplimiento que alcance en la finalidad para la cual fue ideado (costo por unidad de cumplimiento del objetivo). El desarrollo de un producto informático, siempre tiene un costo.

Capítulo 3: Estudio de Factibilidad

Este puede estar justificado por los beneficios tanto tangibles como intangibles que origina el mismo. En este proceso, se necesita de una selección adecuada de los elementos más convenientes para su evaluación.

3.3. Efectos Económicos.

1. Efectos directos.

POSITIVOS:

- Los usuarios con acceso a la BD tendrán la posibilidad de crear reportes confiables de información del consumo energético de los grupos electrógenos.
- Se mejora la eficiencia y calidad del proceso de control del consumo energético de los grupos electrógenos.
- Se obtendrá un resumen más legible de los consumos de materiales y combustibles que se emplean en el funcionamiento y la reparación de los grupos electrógenos.
- La información podrá ser guardada de una manera limpia y sin redundancia en una base de datos en SQL Server (2005-2008).

NEGATIVOS:

El algoritmo propuesto para el proceso ETL de los datos ha sido diseñado para el SGBD SQLServer (2005-2008) ya que la BD diseñada ha sido creada para este SGBD el cual fue especificado por el cliente.

2. Efectos indirectos:

- Los efectos económicos observados que pudiera repercutir sobre otros mercados no son perceptibles, aunque este proyecto no está construido con la finalidad de venta.
3. Efectos externos: Se obtuvo un producto disponible que facilita gran parte del trabajo de control de los indicadores técnico-productivos de los Grupos Electrógenos en Batería, del municipio Moa.

Capítulo 3: Estudio de Factibilidad

4. Intangibles: En la valoración económica siempre hay elementos como perjuicio o beneficio, pero al momento de ponderar en unidades monetarias esto resulta difícil o prácticamente imposible.

3.4. Elementos para identificar los Costos y Beneficios del Proyecto

Para la identificación de los costos y beneficios del proyecto que son pertinentes para su evaluación, es necesario definir una situación base o situación sin proyecto; la comparación de lo que sucede con proyecto, versus lo que hubiera sucedido sin proyecto, definirá los costos y beneficios pertinentes del mismo [María, 2009]. Estos escenarios, resultan ser una herramienta determinante, porque ayudan en gran medida, para la definición de los elementos necesarios en la evaluación.

Situación sin Proyecto:

Para llevar a cabo el control de los indicadores técnicos-productivos de los grupos electrógenos se realizan los siguientes pasos:

- Se almacenarán los datos relacionados con (generación, lecturas, consumo ventiladores) en la Hoja Electrónica de Cálculo Excel sin tener en cuenta la limpieza ni la redundancia de los datos.
- En otra Hoja Electrónica de Cálculos de Excel se controlan datos relacionados con (combustible, aceite, refrigerante, las horas trabajadas de los grupos,) importando los datos que sean necesarios de la primera hoja por lo que la información guardada presenta problemas de redundancia.

Situación con Proyecto:

- Se almacenarán los datos en la base de datos en SQL Server (2005-2008) después de haber pasado por un proceso ETL donde fueron limpiados.
- La información es guardada en una BD única de una manera legible y sin redundancia.

Capítulo 3: Estudio de Factibilidad

- Se controlarán las entradas de combustibles y las mediciones realizadas de una manera mejor organizada. Estas estarán dentro de las tablas totales 1 y totales 2.
- La información guardada podrá ser consultada de una manera más eficiente y segura.

Expuestas ambas situaciones, se procede a continuación con la identificación de los costos y beneficios del proyecto.

3.5 Factibilidad Económica

El análisis de factibilidad económica identifica los costos y beneficios asociados con el proyecto. El mismo incluye cuatro categorías [Torres, 2005]:

- Costo de desarrollo
- Costos operacionales
- Beneficios tangibles
- Beneficios intangibles

Esta tesis, por motivos de comprensión, englobará los costos de desarrollo y los costos operacionales en el epígrafe siguiente.

3.6. Evaluación Económica

Los principales factores a considerar para el cálculo de los costos son los relacionados: al personal, hardware y software, los que se pueden calcular de diversas maneras, que muchas veces se limitan al buen criterio y a la experiencia.

Para determinar el costo económico del proyecto, se desglosaron los costos en moneda libremente convertible y moneda nacional, vea la tabla 3.1 y la 3.2 respectivamente.

Capítulo 3: Estudio de Factibilidad

Tabla 3.1 Datos de los costos en moneda libremente convertible

Ficha de Costo	
	Precios (\$)
Costos Moneda Libremente Convertible	
Costos Directos	
Compra de equipos de cómputo	0,00
Alquiler de equipos de cómputo	0,00
Compra de licencia de Software	0,00
Depreciación de equipos	25,00
Materiales directos	7,50
Subtotal	32,50
Costos Indirectos	
Formación del personal que elabora el proyecto	0,00
Gastos en llamadas telefónicas	0,00
Gastos para el mantenimiento del centro	0,00
Know How	0,00
Gastos en representación	0,00
Subtotal	0,00
Gastos de Distribución y Venta	
Participación en ferias o exposiciones	0,00

Capítulo 3: Estudio de Factibilidad

Gastos en transportación	0,00
Compra de materiales de propagandas	0,00
Subtotal	0,00
Total	32,50

Tabla 3.2 Datos de los costos en moneda nacional

Ficha de Costo	
	Precios (\$)
Costos Moneda Nacional	
Costos Directos	
Salario del personal que laborará en el proyecto	100,00
12,5% del total de gastos por salarios se dedica a la seguridad socia	0,00
9.09% de salario total, por concepto de vacaciones a acumular	0,00
Gasto por consumo de energía eléctrica	15,00
Gastos en llamadas Telefónicas	0,00
Gastos administrativos	0,00

Capítulo 3: Estudio de Factibilidad

Subtotal		115,00
Costos Indirectos		
Know How		0,00
Subtotal		0,00
Gastos de Distribución y Venta		
	Total	115,00

La evaluación económica se efectúa conjuntamente con la que se puede llamar evaluación técnica del proyecto, que consiste en cerciorarse de la factibilidad técnica del mismo. En el análisis de la factibilidad técnica del proyecto, se pudo apreciar que:

- Se cuenta con hardware mayor a los requisitos mínimos
- Se cuenta con la disponibilidad de software

Por lo que se puede inferir que el proyecto es factible técnicamente.

Normalmente no se encuentran problemas en relación con el mercado o la tecnología disponible que se empleará en la fabricación del producto; por tanto la decisión de inversión casi siempre recae en la evaluación económica.

Los costos de los proyectos de informática son relativamente simples de cuantificar, no así los beneficios, que se presentan como ahorro de costos con respecto a la situación base [Rebaza, 2007]

3.6.1. Beneficios Tangibles.

Son los ingresos adicionales y/o reducción de costos que la nueva BD proveerá. Debido a que este proyecto no es construido con la finalidad de comercializarse, no posee ingresos monetarios perceptibles, no así con la reducción de costos, pues el sistema proveerá, mejoras en la gestión y la toma de decisiones.

Capítulo 3: Estudio de Factibilidad

Estos elementos son de muy difícil cuantificación, por lo que en esta tesis se consideran sólo como intangibles.

3.6.2. Beneficios y Costos Intangibles en el Proyecto

Costos

- Resistencia al cambio

Beneficios

- Mejor comodidad para los usuarios.
- Mejora la calidad del control de los indicadores técnico-productivos de los GEDB.
- Menor tiempo empleado en el proceso de control de los indicadores técnico-productivos de los GEDB.

Como se hizo referencia anteriormente, la técnica seleccionada para evaluar la factibilidad del proyecto es la Metodología Costo- Efectividad. Dentro de la misma los gráficos de barra (*vea la figura 3.1*) son aplicable a proyectos donde los beneficios tangibles no son evidentes, el análisis se basa exclusivamente en los costos. Para esta técnica es imprescindible definir una variable discreta (tipo de variables) que haga variar los costos. Teniendo en cuenta que el costo para este proyecto es despreciable, tomaremos como costo el tiempo en minutos (tiempo (min)) empleado para resolver el control de los indicadores técnico-productivos de los GEDB y las variables serían la complejidad de las pruebas que se realizan durante este proceso, las mismas se enuncian a continuación:

1. Valores de la variable 1: Solución con la HECE.

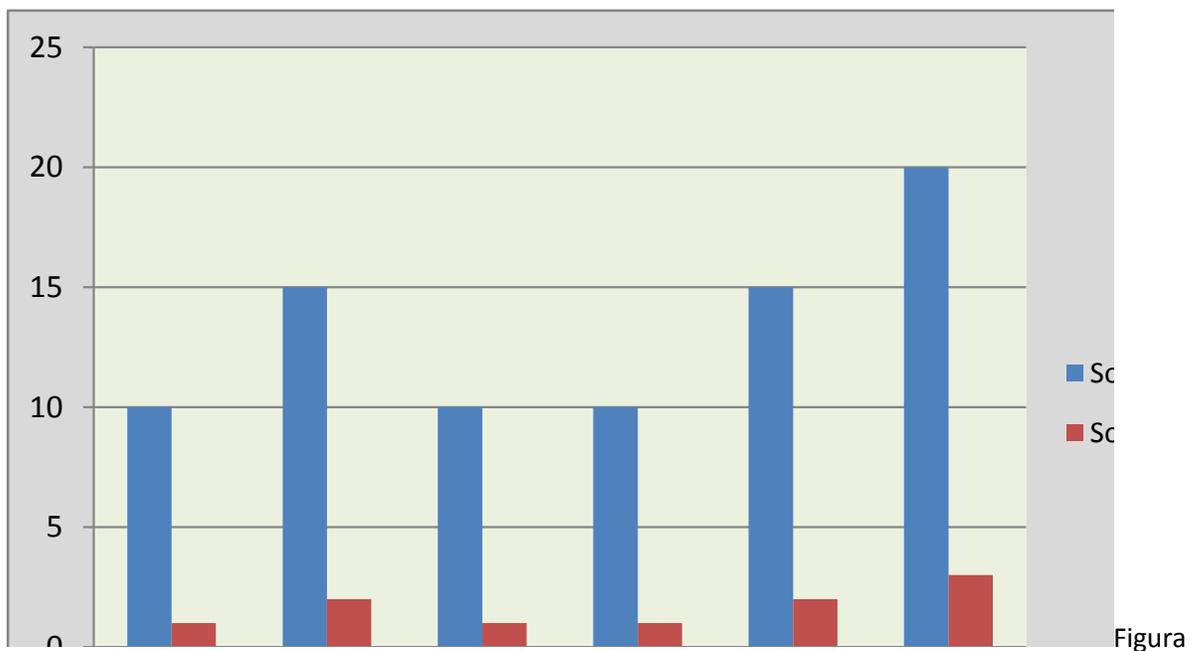
- Gestionar la lectura diaria de cada Grupo Electrónico. Cuatro variables (10 min)
- Ubicar los GEDB y efectuar el control. Tres variables(15 min)
- Elaborar un resumen del consumo de combustible y aceite diario por cada Grupo Electrónico. Dos variables (10 min)
- Elaborar un resumen del consumo de combustible y aceite del mes por cada Grupo Electrónico. Dos variables (10 min).

Capítulo 3: Estudio de Factibilidad

- Elaborar un resumen del mantenimiento que se le realiza a los Grupos Electrógenos. Cuatro variables(15 min)
- Almacenar toda la información referente a la batería. Cuatro variables(20 min)

2. Valores de la variable 2: Solución con la BD

- Gestionar la lectura diaria de cada Grupo Electrónico. Cuatro variables (1 min)
- Ubicar los GEDB y efectuar el control. Tres variables (2 min)
- Elaborar un resumen del consumo de combustible y aceite diario por cada Grupo Electrónico. Dos variables (1 min)
- Elaborar un resumen del consumo de combustible y aceite del mes por cada Grupo Electrónico. Dos variables (1 min).
- Elaborar un resumen del mantenimiento que se le realiza a los Grupos Electrónicos. Cuatro variables(2 min)
- Almacenar toda la información referente a la batería. Cuatro variables(3 min)



3.1 Gráfico de barra de las soluciones con la HECE y con la BD

Capítulo 3: Estudio de Factibilidad

El comportamiento de ambas variables se muestra en la figura 3.1. Como se observa, con la Solución con la HECE el estado de los seis campos es mayor a los de la Solución con la BD y el de máxima incidencia es el sexto, porque el sistema que se usa (SQL Server) es un gestor de base de datos a diferencia de la HECE. En conclusión el tiempo en minutos es 9 min para la variable 2, o sea, el costo intangible de la solución con la BD es factible.

3.7 Conclusiones del Capítulo.

- Se aplicó la metodología costo-beneficio en la cual se analizaron los efectos económicos, los beneficios y costos intangibles del proyecto.
- Se estimó el costo de ejecución y como resultado la utilización es de \$ 32,50 en CUC y \$ 115,00 en MN, demostrándose la factibilidad del proyecto.

Conclusiones Generales

Con el desarrollo de la Base de Datos de los Indicadores Técnicos- Productivos se dio cumplimiento a los objetivos propuestos en este trabajo. Pues se obtuvo como resultado un producto informático capaz de incidir de forma eficiente en la toma de decisiones para la explotación eficaz de los Grupos Electrónicos Diésel en Batería, arribándose a las siguientes conclusiones:

- Se realizó el estudio del marco teórico-contextual de los indicadores técnico-productivos de los Grupos Electrónicos Diésel en Batería.
- Se determinaron las metodologías y herramientas más factibles para realizar la construcción de la base de datos.
- Se diseñó la Base de Datos de los Indicadores Técnicos-Productivos
- Se pobló la base de datos quedando definidas las tablas y sus atributos.
- Se realizó un estudio de factibilidad en el que se demuestra lo rentable que es el proyecto apoyándose en el tiempo de estimación y en el costo de este.

Recomendaciones

En el presente trabajo recomendamos:

1. La Base de Datos de los Indicadores Técnicos-Productivos de los GED implementada sea usada por los especialistas de la UNE-GD de la provincia Holguín.
2. A la Base de Datos de los Indicadores Técnicos-Productivos de los GED se la añade el indicador "Demanda Eléctrica" para futuros estudio de la explotación de la GEDB ante situaciones excepcionales.
3. Desarrollar una aplicación informática para el trabajo con la base de datos de los GED.
4. Utilizar la Base de Datos de los Indicadores Técnicos-Productivos de los GED para un estudio de Minería de Datos.

Bibliografía

ARRASTRÍA AVILA, M, A., "Distributed generation in Cuba: parto f a transition towards a new energy paradigm". Revista Cogeneration ona On-Site Power Production. Cuba, Noviembre-diciembre 2008. p 61-65. [Disponible [http: www.cosp.com](http://www.cosp.com)]

Bernau Ricardo, D. (2010). HEFESTO Metodología para la construcción de un Data Warehouse. Córdoba.

Bisset, K; Sánchez, D; Noa, A. Obtención de Mejoras en la Explotación de Grupos Electrógenos Diésel Aplicando Minería de Datos. Tesis de Grado. ISMMM. Holguín. Moa. Julio 2010.pp104.

Blanco, H; González, B; Rodríguez, E; Miranda, R; Ortiz, R. Separata de planificación de empresas. Universidad de La Habana. Editorial: André Voisin, 1989. Cuba. pp 58-62.

Bureaudeprensa.com. 2008. Embarcadero ER/Studio. [En línea] 2008.

http://bureaudeprensa.com/es/view.php?bn=bureaudeprensa_software&key=1153755975

CAVSI [En línea] // ¿Qué es un Sistema Gestor de Bases de Datos o SGBD? - 2004. – diciembre de 2010. -[http://www.cavsi.com/preguntasrespuestas/que-es-un-sistema-gestor-de-bases-de-datos-o-sgbd/...](http://www.cavsi.com/preguntasrespuestas/que-es-un-sistema-gestor-de-bases-de-datos-o-sgbd/)

Cutiño, A; Rivera, Y. Sistema Informático para Mejorar la Toma de Decisiones en la Explotación de los Grupos Electrógenos Diésel de Moa. Tesis de Grado. ISMMM. Holguín. Moa. Julio 2012.pp60.

Eaprende.com [En línea] // Gestor de Base de Datos: MySQL, PostgreSQL, SQLite. - 2001. - enero de 2011. <http://www.eaprende.com/gestor-de-basededatos-mysql-postgresql-sqlite.html>....

GACETA OFICIAL DE LA REPÚBLICA DE CUBA. "Ley 75 de la Defensa Nacional:

Artículo 9°. La Habana, Cuba. 14 de Septiembre 1998.

GAMMA, E. (1995) Design Patterns. Elements of Reusable Object Oriented Software.

Guerrero, Y; Laborde, R; Romero, I. Sincronización de Sistema Híbrido Diésel- Eólico-Hídrico en Régimen Autónomo en el Municipio Moa .Tesis de Maestría ISMMM. Holguín Moa.2011.pp93

Islasoft.com Islasoft.com [En línea] // Diferencias entre Microsoft Access y Microsoft SQL Server/MSDE.. -2006. -enerode2011. -

http://www.islasoft.com/Producto/Diferencias_SQL_Server_Office.pdf..

KIMBALL, M. R. R. (2005) slowly changing dimensions are not always as easy as 1, 2, 3. InformationWeek.

Kimball, R., & Ross, M. (s.f.). The Data Warehouse Toolkid (Segunda edición ed.). Canada.

Manual de operación. Grupos de Desarrollo EROS. Nícaro. Cuba 2006.

Maresma, D; Reyes, A; Montero, R. Características operacionales y compartimientos de las principales variables de los Grupos Electrógenos para la Generación Distribuida en Moa. Tesis de Grado. ISMMM de Holguín. Moa. Julio 2007.pp40.

María F. S. Rincondelvago.com [En línea]// "Evaluación de proyectos.".- 2009. - noviembre de 2010. - <http://html.rincondelvago.com/evaluacion-de-proyectos.html...>

Mas, D. O; Guerra, S. Desarrollo y Aplicación de funcionalidades del proyecto de SCADA EN Grupos Feul Oil con Tecnología HYUNDAI. 15 Convención Científica de Ingeniería y Arquitectura. Simposio Internacional de Generación Distribuida. CUJAE. La Habana. Cuba. 29Noviembre-3Diciembre. ISBN.978-959-261-317-1.

Mato, R, M. Diseño de Bases de Datos. Octubre 1999. pp 42.

Nieves Borrero Martha y Góngora Rodríguez Asnier Herramienta Informática para automatizar los procesos en el Laboratorio de Calidad de Software: Módulo Gestión de las No Conformidades [Libro]. - Ciudad de La Habana: Universidad de las Ciencias Informáticas, 2007.

Noa, A; Marrero, S; Legrá, A.A. Operación del Sistema de Generación Distribuida de Moa para Situaciones Excepcionales. Tesis de Maestría ISMMM. Holguín. Moa. Julio 2009.pp76.

Oficina Nacional de Estadísticas (ONE). Producción y Consumo de Electricidad. Dirección de Industria. Ciudad de la Habana .Cuba. Febrero 2009. pp15.<http://www.one.cu>.

Oficina Nacional de Estadísticas (ONE). Producción y Consumo de Electricidad. Dirección de Industria. Ciudad de la Habana .Cuba. Agosto2012. pp10.<http://www.one.cu>.

Pérez, B. Análisis de la Eficiencia de los Grupos Electrónicos Diésel Santiago Industrial y Santiago Este. Segundo Congreso Cubano de Ingeniería Eléctrica. Simposio Internacional de Generación Distribuida. Cuba.26-30 Noviembre 2012ISBN.978-959-261-405-5.

Pérez, I; Rodríguez, M; Vázquez, A. Gestión de la Información de la Generación Distribuida en Cuba con un Sistema de Información Geográfica. 15 Convención Científica de Ingeniería y Arquitectura. Simposio Internacional de Generación Distribuida. CUJAE. La Habana. Cuba. 29Noviembre-3Diciembre. ISBN.978-959-261-317-1.

Peña, D; Cobos, A. Rediseño del Sistema de Supervisión y Adquisición de Datos de los Grupos Electrónicos Feul Oil de Tecnología HYUNDAI de 2,5MW. 15 Convención Científica de Ingeniería y Arquitectura. Simposio Internacional de Generación Distribuida. CUJAE. La Habana. Cuba. 29Noviembre-3Diciembre. ISBN.978-959-261-317-1.

Pon Juan-Dante, G; Mugía, M. M. Sistemas de Información: principios y aplicaciones. Editorial Félix Varela. La Habana. Cuba.

Rebaza J. C. V. Sociedad de Estudiantes de Ciencia de la Computación SECC [En línea]// Metodología para la Priorización de Proyectos Informáticos. - 2007. - enero de 2011. - <http://www.seccperu.org/node/302>..

Rodríguez, G; López, M.J. La Toma de Decisiones Multicriterio en la Organización del Conocimiento: aplicación del caso del Centro de Estudio de la Energía y Tecnología de Avanzada en Moa. Tesis de Doctorado. Universidad de Granada. España. Octubre 2012. pp20-23.

Rozales, A; Guerrero, Y. Simulación en Matlab de la Demanda de la Batería de Grupos Electrónomos Moa en Isla. Tesis de Grado ISMMM. Holguín. Moa .2010.pp67.

Torres D.[En línea]// Proyectos Informáticos. - 2005. - enero de 2011. - <http://notasprisma.tripod.com/Proyectos.htm...>

Unión Eléctrica .Dirección de Generación Distribuida Manual de calidad para la operación .La Habana.Cuba.2007.

Unión Nacional Eléctrica para la Generación Distribuida (UNE-GD). Estructura Funcional del Sistema de Gestión de la Información de la Generación Distribuida. La Habana. Cuba. 2006. pp10.

UEB Generación Distribuida. Generando Logro. Video. 2012

. 2004. UEB Generación Distribuida. Generando logros. Video. 2012

Anexos 1.

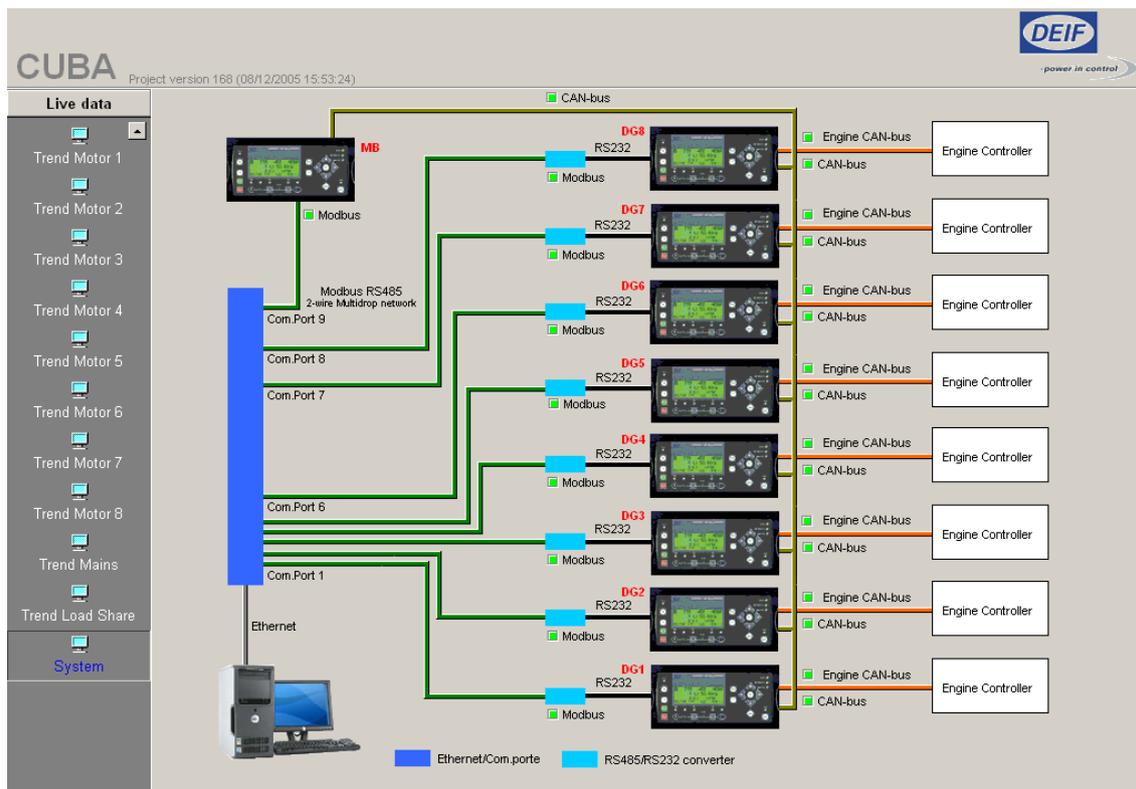


Figura 1. Esquema del protocolo MODBUS del SCADA "Eros"