B2.341– Trabajo final de postgrado
Sistema de inteligencia de negocio entorno a los parques eólicos offshore (BI.2)

Robert Berchtold Palacios
Postgrado - Sistemas de Información de Inteligencia de Negocio y Big data
Área de Sistemas de Información de Inteligencia de Negocio

David Amorós Alcaraz
María Isabel Guitart Hormigo

02/07/2017

FICHA DEL TRABAJO FINAL

Título del trabajo:	*B2.341 - Sistema de inteligencia de negocio entorno a los parques eólicos*
Nombre del autor:	*Robert Berchtold Palacios*
Nombre del consultor/a:	*David Amorós Alcaraz*
Nombre del PRA:	*María Isabel Guitart Hormigo*
Fecha de entrega (mm/aaaa):	07/2017
Titulación:	Postgrado en Sistemas de Información de Inteligencia de Negocio y Big data
Área del Trabajo Final:	Área de Sistemas de Información de Inteligencia de Negocio
Idioma del trabajo:	*Castellano*
Palabras clave	*Procesos ETL* *Almacén de datos* *Cubos multidimensionales*

Resumen del Trabajo (máximo 250 palabras): *Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.*

El objetivo general de este trabajo es realizar el diseño e implementación de un sistema de Business Intelligence que facilite la adquisición, el almacenamiento y la explotación de datos provenientes de diferentes parques eólicos donde hay instaladas boyas meteorológicas. Esto permitirá poder disponer de información adicional sobre las condiciones meteorológicas y tener más capacidad de análisis.

Los objetivos específicos del trabajo son:

- Realizar el almacenamiento de datos históricos en un almacén de datos.
- Extraer y manipular los datos disponibles en las fuentes de datos.
- Generar cubos multidimensionales.
- Visualizar y analizar los datos almacenados.

Este proyecto se ha dividido en seis fases donde se ha:

1) Identificado el hardware y software necesario.

2) Instalado y configurado las herramientas del sistema.

3) Realizado la modelización e implementación del almacén de datos (data warehouse) a partir del análisis de requerimientos y de los datos de las fuentes de datos.

4) Generado las transformaciones y jobs necesarios para la carga de datos en el almacén de datos.

5) Creado un cubo de análisis multidimensional.

6) Analizado los datos mediante Pentaho Analytics y respondido las preguntas analíticas del proyecto.

A nivel de implantación del proyecto, se ha logrado implementar un sistema de inteligencia de negocio para parques eólicos offshore donde se han podido satisfacer varias de las preguntas analíticas que se deseaban resolver a partir de los datos disponibles para este proyecto.

A nivel de planificación de proyecto, se han podido alcanzar los objetivos en el tiempo estipulado, introduciendo nuevos apartados y realizando breves cambios en la memoria a lo largo de este.

Abstract (in English, 250 words or less):

The general objetive of this work plan is perform the design and implementation of a Bussiness Intelligence system that to ease the adquisition, storage and explotation the data coming of the different Eolic parks where it has been installed meteorologic buoys. This will allow to provide available information about the meteorologic conditions and we will have more analysis capacity.

The specific objectives of this working plan are:

• Perform the storage of historical data in a data warehouse.

• Extract and manipulate data available in data sources.

• Generate multidimensional cubes.

• Display and analyze the stored data.

This project has been divided into six phases where it has:

1) Identified the necessary hardware and software.

2) Installed and configured the system tools.

3) Made the modeling and implementation of the data warehouse (data warehouse) from the analysis of requirements and data of data sources.

4) Generated the transformations and necessary jobs for loading data into the data warehouse.

5) Created a multidimensional analysis cube.

6) Analyzed the data using Pentaho Analytics and answered the analytical questions of the project.

At the level of implementation of the project, it has been implemented a business intelligence system for offshore wind parks where it has been asnwered some of the analytical questions that were desired to answer from the data available for this project.

At level of project planning It has been achieved to reach the objectives in the time stipulated, introducing new sections and making brief changes in the memory throughout this.

Índice

Lista de figuras

1. Introducción

1.1. Contexto y justificación del Trabajo

Hoy en día, la energía eólica es la más avanzada técnicamente dentro de las energías renovables. Además, de todas las energías renovables, es también la más rentable. Según algunos expertos, sólo esta energía podría llegar algún día a abastecer Europa sin problemas, siendo su potencial, para el 2020, tres veces superior a la demanda prevista.

Un aerogenerador es un dispositivo capaz de transformar la energía eólica en energía eléctrica. Las partes del aerogenerador que permiten realizar esta transformación son las palas, la multiplicadora (caja de engranajes) y el generador, encargado de producir la energía eléctrica.

Los aerogeneradores que actualmente se construyen, están altamente sensorizados. El elevado número de datos que se recogen, puede dar mucha información sobre el estado actual de los diferentes sistemas que lo componen. Indicadores como la potencia, la temperatura, las vibraciones o el viento nos pueden ayudar en gran medida a anticipar paradas y evitar fallos que pueden llegar a ser muy costosas.

El rendimiento de un aerogenerador depende, entre otros, de su localización geográfica. No obstante, un mantenimiento adecuado permite sacar más provecho cuando las condiciones atmosféricas son las adecuadas.

La energía eólica tiene un futuro prometedor, aunque los mejores terrenos han sido ya ocupados o están en trámites de autorización y cada vez es más difícil encontrar zonas con altas velocidades de viento sin explotar; como solución a este problema aparecen una nueva posibilidad, la energía eólica offshore.

En cualquier parque eólico el factor decisivo es la velocidad del viento, en el mar se calcula una velocidad en 1 m/s por encima de las zonas costeras cercanas a causa de que en el mar no existen obstáculos y la rugosidad del suelo es muchísimo menor, esto significa que en un parque eólico offshore la producción de electricidad a lo largo del año es de la orden de un 20% más que en tierra.

Debido al hecho que los parques eólicos offshore se sitúan mar adentro y su mantenimiento es costoso (monitorización, actuaciones, recambios, etc.), han empezado a apareces soluciones que pueden mejorar estas actividades. Una de ellas es la colocación de una Boya meteorológica dentro del parque que permita obtener información sobre las condiciones del parque en todo

momento. Esto permitirá conocer el origen de determinadas incidencias en el parque y prever posibles alarmas o averías.

1.2. Objetivos del Trabajo

El objetivo general es el ***diseño e implementación de un sistema de Business Intelligence*** que facilite la adquisición, el almacenamiento y la explotación de datos provenientes de diferentes parques eólicos donde tenemos instaladas boyas meteorológicas. Esto permitirá poder disponer de información adicional sobre las condiciones meteorológicas y tener más capacidad de análisis.

Los objetivos específicos del trabajo son:

1. Diseñar un almacén de datos (Data Warehouse).

2. Implementar este almacén de datos.

3. Utilizar la suite de Pentaho en su versión Community para la realización de los siguientes trabajos:

    a. Programar los procesos ETL (extracción, transformación y carga) que permitan alimentar el DW a partir de los ficheros base facilitados. Para ellos se podrá utilizar la herramienta Data Integration de Pentaho o crear procesos ad-hoc manualmente.

    b. Utilizar la herramienta Schema Workbench de Pentaho para la creación de cubos.

    c. Utilizar el Pentaho Business Analytics (plugin Sayku) como herramienta Front End de Análisis.

1.3. Enfoque y método seguido

La estrategia a seguir para el correcto desarrollo e implementación de este sistema de inteligencia de inteligencia de negocio, orientado al análisis de los datos de los diferentes parques eólicos, estará basada en la realización y ejecución secuencial de una serie de fases y puntos clave del proyecto.

Las fases del proyecto a realizar son:

1. Hardware y Software
2. Instalación y configuración de las herramientas.

2.1. Pentaho Data Integración (ETL).
2.2. Pentaho Schema Workbench (Cubos de análisis).
2.3. Pentaho Business Analytics (Frontal web de análisis).
2.4. PostgreSQL (Data Warehouse).

3. Modelización e implementación del almacén de datos.

   3.1. Análisis de requerimientos.
   3.2. Análisis de fuente de datos.
   3.3. Análisis funcional.
   3.4. Diseño conceptual.
   3.5. Diseño lógico.
   3.6. Diseño físico.

4. Carga de datos mediante procesos ETL.

   4.1. Identificar los procesos ETL necesarios.
   4.2. Descripción de las acciones a realizar.
   4.3. Implementación de procesos ETL.

5. Diseño multidimensional.

   5.1. Identificación de las necesidades analíticas.
   5.2. Creación e implementación de cubos analíticos.

6. Análisis multidimensional.

   6.1. Interpretación de resultados.

1.4. Planificación del Trabajo

- Hardware y Software.

  Inicio de fase: 27/03
  Fin de fase: 09/04

  Estimación: 2 semanas

- Instalación y configuración de las herramientas.

  Inicio de fase: 27/03
  Fin de fase: 09/04

  Estimación: 2 semanas

- Modelización e implementación del almacén de datos.

  Inicio de fase: 10/04

Fin de fase: 30/04

Estimación: 3 semanas

- Carga de datos mediante procesos ETL.

  Inicio de fase: 01/05
  Fin de fase: 21/05

  Estimación: 3 semanas

- Diseño multidimensional.

  Inicio de fase: 22/05
  Fin de fase: 11/06

  Estimación: 3 semanas

- Análisis multidimensional.

  Inicio de fase: 12/06
  Fin de fase: 25/06

  Estimación: 2 semanas

- Finalizar documentación y preparar exposición.

  Inicio de fase: 26/06
  Fin de fase: 07/07

  Estimación: 2 semanas

Se adjunta Excel Anexo 1: Documento de planificación del proyecto con los detalles de la planificación de tareas y fechas de entrega del proyecto.

1.5. Breve sumario de productos obtenidos

Los productos obtenidos para la realización y ejecución de este Trabajo son:

- Para los procesos Extracción, transformación y carga de datos se ha seleccionado el producto ***Pentaho Data Integration***.

- Para la creación de cubos de análisis se ha seleccionado el producto ***Pentaho Schema Workbench***.

- Para la capa frontal de análisis se ha seleccionado el producto ***Pentaho Bussiness Analytics (plugin Saiku)***.

- Para la capa de almacenaje de datos (data warehouse) se ha seleccionado el producto ***PostgreSQL***.

1.6. Breve descripción de los otros capítulos de la memoria.

En los siguientes capítulos de esta memoria serán descritas todas las fases de este proyecto además, de detallar en profundidad todos los puntos de análisis, diseño, identificación, configuración e implementación de cada uno de los procesos de este sistema de Business Intelligence.

En el primer capítulo o fase del proyecto con nomenclatura ***"Hardware y Software"*** se realizará una breve descripción de los elementos de hardware y software de la infraestructura.

En el segundo capítulo o fase del proyecto con nomenclatura ***"Instalación y configuración de las herramientas",*** encontraremos los detalles referentes a la descarga, requisitos, instalación y configuración de cada uno de los productos y herramientas descritos en el ***apartado 1.5 de esta memoria***.

En el tercer capítulo o fase del proyecto con nomenclatura ***"Modelización e implementación del almacén de datos"*** se detallaran todos los puntos de análisis de los datos previos y de diseño requeridos para la implementación del almacén de datos Data Warehouse.

En el cuarto capítulo o fase del proyecto con nomenclatura ***"Carga de datos mediante procesos ETL"*** se pretende identificar, describir e implementar cada uno de los procesos ETL requeridos para alimentar el Data Warehouse con la información de los parques eólicos.

En el quinto capítulo o fase del proyecto con nomenclatura ***"Diseño multidimensional"*** se realizará un estudio de las preguntas analíticas a responder, se generarán los cubos de análisis necesarios para satisfacer estas cuestiones y finalmente se cargaran los cubos multidimensionales en la aplicación frontal web Pentaho Analytics.

La última fase del proyecto con nomenclatura ***"Análisis multidimensional"***, se generará a partir de los cubos analíticos los datos de análisis necesarios para responder a las preguntas analíticas a satisfacer con este proyecto.

2. Hardware y Software

Uno de los puntos críticos para un correcto desarrollo de un proyecto de implantación de un sistema de inteligencia de negocio es realizar una correcta elección del **software** y del **hardware** a implementar en la infraestructura, ya que de estos dos activos dependerá el correcto funcionamiento, rendimiento y desarrollo de los procesos y servicios que ofrecerá el entorno.

De cara al **software**, es necesario una correcta selección de los siguientes productos:

- Software de almacenamiento de datos históricos.
- Software de extracción y manipulación de los datos.
- Software de generación de cubos multidimensionales.
- Herramientas Middleware para la visualización y análisis de los datos.

En el caso que nos ocupa, se han seleccionado diversos productos de **Pentaho BI Suite** debido a que son herramientas Business Intelligence libres, con diversos frameworks para cada tipo de requerimiento analítico, fáciles de utilizar, con grandes posibilidades y opciones de configuración muy por encima de lo que nos ofrecen otros productos no libres.

Para la parte de almacenamiento de datos históricos se ha seleccionado el producto **PostgreSQL** debido a que las características son muy similares a la suite de Pentaho. PosgreSQL es un programa libre de alta calidad, con mucha documentación, con soporte tanto profesional como de la comunidad, fiable, estable, con un rendimiento excelente para volúmenes altos de tráfico y transacciones muy por encima de las posibilidades que nos ofrecen otros productos no libres.

Estas características descritas hacen ideales los productos de Pentaho y PostgreSQL para satisfacer los requerimientos de software de nuestro sistema.

De cara al **hardware**, se utilizará para este proyecto una sola máquina donde se simulará y se instalarán todos los productos necesarios para este sistema de inteligencia de negocio. La elección de una sola maquina es debido a que a día de hoy desconocemos la envergadura real del proyecto y es muy difícil dimensionar la capacidad real que tendrá y requerirá el sistema productivo.

La capacidad de la máquina y los requerimientos mínimos son:

SO: Windows 7 64 bits
Procesador: Intel(R) Core (TM) i7-3610QM CPU @ 2.30GHz
RAM: 6 GB
Almacenamiento: 500 GB

Para el caso que nos ocupa y como **entorno de desarrollo**, esta capacidad a nivel de hardware debería ser suficiente para desarrollar y mantener los procesos y servicios del sistema a implementar.

En un **entorno productivo** final, donde se realice la implantación de este sistema, la infraestructura mínima, con la finalidad no solo de dar servicio, sino de ser un sistema con un buen rendimiento y de alta disponibilidad, debería ser similar a la ilustración siguiente:

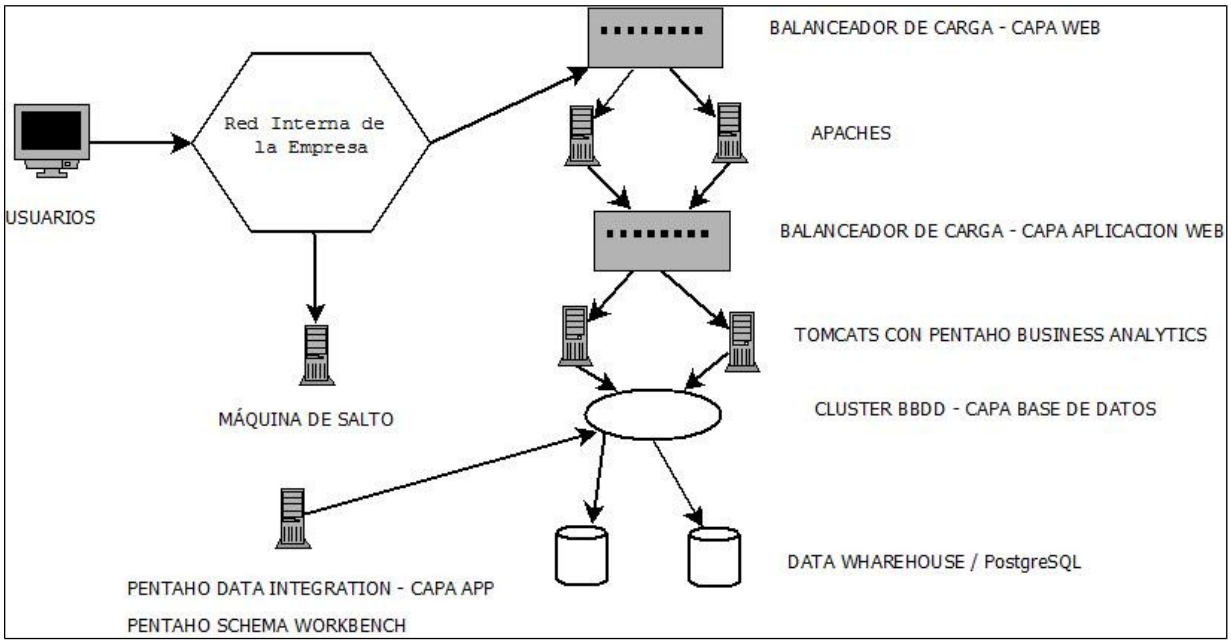

Ilustración 1 - Infraestructura entorno de producción

Este sistema de producción deberá contener un mínimo de 8 servidores con siguientes capacidades y distribuidos en:

Capa Web:

Esta capa contará con un balanceador de carga y dos servidores con producto Apache que se utilizarán únicamente para generar la URL de la aplicación Pentaho Business Analytics y redirigir a los usuarios a la capa de aplicaciones web.

SO: Linux Red Hat
Procesador: Intel(R) Core (TM) i3-6100U CPU @ 2.30GHz
RAM: 2 GB
Almacenamiento: 40 GB

Capa Aplicaciones Web:

Esta capa contará con un balanceador de carga y dos servidores con el producto Tomcat donde se ha desplegado el archivo .war de Pentaho Business Analytics. Estos servidores contarán con conexión al Data Warehouse para realizar las consultas analíticas requeridas.

SO: Linux Red Hat
Procesador: Intel(R) Core (TM) i3-6100U CPU @ 2.30GHz
RAM: 8 GB
Almacenamiento: 60 GB

Capa base de datos:

Esta capa contará con un clúster de base de datos y dos servidores con el producto de almacenamiento de datos PostgreSQL.

SO: Linux Red Hat
Procesador: Intel(R) Core (TM) i3-6100U CPU @ 2.30GHz
RAM: 4 GB
Almacenamiento: 100 GB

Capa de aplicaciones:

Esta capa contará con un solo servidor que tendrá instalados los productos Pentaho Data Integration y Pentaho Schema Workbench.

SO: Microsoft Windows Server
Procesador: Intel(R) Core (TM) i3-6100U CPU @ 2.30GHz
RAM: 8 GB
Almacenamiento: 60 GB

Capa de administración:

Máquina de salto con acceso a todos los servidores de la infraestructura mediante putty (o un programa similar) y remote desktop.

SO: Microsoft Windows Server
Procesador: Intel(R) Core (TM) i3-6100U CPU @ 2.30GHz
RAM: 2 GB
Almacenamiento: 40 GB

Todos los servidores del sistema, además de los productos mencionados anteriormente, deberán tener incluido un sistema de copias de seguridad, antivirus y un sistema de monitorización que permita la generación de alertas en el caso que se detecte algún problema en alguno de los servicios o procesos.

3. Instalación y configuración de las herramientas.

En este apartado se detallan y resumen todos los pasos requeridos para la correcta instalación de cada uno de los productos de nuestro sistema de inteligencia de negocio.

3.1. Pentaho Data Integration (ETL).

El software seleccionado para la realización de extracciones, manipulaciones y carga de datos es el framework **Data Integration** de suite de Pentaho.

Los detalles del proceso de instalación y configuración se pueden encontrar en el Anexo 2: Manual de instalación - Pentaho Data Integration.

3.2. Pentaho Schema Workbench (Cubos de análisis).

El software seleccionado para la generación de cubos multidimensionales es el framework **Schema Workbench** de suite de Pentaho.

Los detalles del proceso de instalación y configuración se pueden encontrar en el Anexo 3: Manual de instalación - Pentaho Schema Workbench.

3.3. Pentaho Business Analytics (Frontal web de análisis).

El software seleccionado para la para la visualización y análisis de los datos es el framework **Business Analytics** de la suite de Pentaho. Este framework utiliza por defecto para la visualización de cubos multidimensionales el plugin **JPivot**. En nuestro caso, para la visualización y gestión que cubos multidimensionales utilizaremos el plugin **Saiku**, ya que dispone de una interfaz más amigable, intuitiva y sencilla de utilizar.

Los detalles del proceso de instalación y configuración se pueden encontrar en el Anexo 4: Manual de instalación - Pentaho Business Analytics.

3.4. PostgreSQL (Data Warehouse).

El software seleccionado para la implantación del datawarehouse es PostgreSQL.

Los detalles del proceso de instalación y configuración se pueden encontrar en el Anexo 5: Manual de instalación - PostgreSQL.

4. Modelización e implementación del almacén de datos.

A partir del análisis del contexto del caso y de las fuentes de datos disponibles, se diseñará un almacén de datos que ofrezca soporte a los requerimientos analíticos a partir de los datos aportados de cada uno de los parques eólicos offshore.

4.1. Análisis de requerimientos.

Este análisis de requerimientos se basa en identificar las necesidades específicas requeridas en el sistema de inteligencia de negocio para parques eólicos offshore.

A nivel analítico, las preguntas que el sistema deberá ser capaz de resolver son:

1) Cuáles son los parques eólicos más productivos.
    - No necesariamente las zonas con constantes de viento más altas son las mejores, a veces la colocación de los aerogeneradores o la tecnología (marca, modelo) pueden hacer variar la potencia generada en cada caso.

2) Hacer un análisis profundo sobre los datos meteorológicos de cada parque y las incidencias sobre la producción.

3) Zonas con mejor relación viento/potencia.
    - Dado el caso que se quiera ampliar un parque o crear uno de nuevo, este dato nos puede ser muy útil para valorar la ubicación definitiva.

4) Análisis de alarmas.
    - Ver si hay alguna relación entre las alarmas que se producen (que acostumbran a parar el aerogenerador) y las variaciones meteorológicas. Esto puede ayudar a analizar cuáles son los problemas de los parques.

5) Efectividad de las empresas de mantenimiento a partir del dato de disponibilidad.

    - Es deseable que una empresa de mantenimiento minimice las alarmas y maximice la disponibilidad. Esta información es vital para la establecer los criterios contractuales o renovarlos.

6) Relaciones entre variables de producción y meteorológicas.

    - La temperatura externa en un parque puede ser un indicador de futuras averías, de la misma forma que lo pueden ser las rachas de viento extremas muy continuadas o la altura de las olas.

A nivel de modelaje de datos debemos tener en cuenta los siguientes aspectos:

1) **Geografía y Parques Eólicos:**

    - Se quiere situar los parques eólicos offshore a nivel de mar u océano. Se tendrá información del país al que pertenecen las aguas. Cada parque estará geolocalizado para saber la ubicación exacta.

2) **Tiempo:**

    - Generalmente cada aerogenerador envía datos cada diez minutos, a pesar de que a veces se pueden obtener datos en frecuencias más pequeñas. Se quiere que el sistema almacene los datos con una granularidad diaria. Pudiéndose hacer agrupaciones a nivel mensual y anual.

    - No tendremos información detallada de los aerogeneradores de los parques eólicos. Sólo tendremos unos atributos por cada parque con la potencia de los aerogeneradores y el número de aerogeneradores (supondremos que dentro de un parque tenemos el mismo modelo).

3) **Alarmas, Mantenedores y operadores:**

    - Cuando se produce una alarma, esta viene informada con el sistema donde se ha producido la alarma: pitch (palas), multiplicadora, generador u otros (para el resto de partes del aerogenerador). Lo que recibiremos será información resumida para cada parque.

- Tendremos un indicador sobre la disponibilidad media del parque. Esto nos permitirá saber qué porcentaje del tiempo no se ha producido.

- Tendremos información de qué empresa es la encargada de llevar el mantenimiento de un parque. Aun así, también sabremos qué empresa explota un parque y por lo tanto lo gestiona.

4) Indicadores

Se prevé poder analizar datos numéricos diarios sobre:

- la potencia generada.
- la velocidad media del viento.
- la temperatura media externa.
- las alarmas producidas.
- la altura media de las olas.
- la disponibilidad media del parque (%).
- las actuaciones realizadas.

4.2. Análisis de fuente de datos.

En los proyectos de inteligencia de negocio resulta relevante analizar las fuentes de datos. Del análisis de las mismas puede desprenderse información clave para el éxito y la evolución de proyecto, así como la identificación de riesgos vinculados.

Para este proyecto de inteligencia de negocio se nos ha proporcionado un fichero de origen de fuente de datos con nomenclatura ***DATAOFFSHORE.xls.*** A continuación se detalla en profundad los principales aspectos a destacar sobre el fichero proporcionado:

DATAOFFSHORE.xls

- Un fichero Excel con un total de 9 hojas de datos (pestañas).

- Una de las hojas de datos contiene información general sobre cada uno de los parques eólicos offshore.
- Las ocho hojas de datos restantes contienen información relevante sobre varios factores meteorológicos y técnicos sobre cada uno de los parques eólicos offshore a analizar.

- Esta información relevante comentada en el punto anterior, a sido registrada en intervalos de 10 min.

- La información de los atributos está codificada de manera numérica, alfabéticamente y temporal.

- Parte de los campos y columnas están en ingles, por lo que debemos hacer una correspondencia al castellano.

- Las columnas que se incluyen en esta fuente de datos para la **primera hoja de datos** son:

Nombre atributo (En el fichero)	Nombre atributo (En Castellano)	Descripción	Tipo de dato	Nivel Información	Opcional
NAME	nombre	Nombre del parque eólico	Cadena	1 Nivel (nombre del parque)	NO
COUNTRY	país	País situado parque eólico	Cadena	1 Nivel (país)	NO
AEROGEN	aerogeneradores	Marca de los aerogeneradores instalados en el parque eólico	Cadena	1 Nivel (marca aerogeneradores)	NO
POWER	potencia	Potencia de los aerogeneradores del parque eólico (mw)	Cadena	1 Nivel (MW)	NO
NUMBER	número	Número de aerogeneradores del parque eólico (Unidad)	Numérico entero	1 Nivel (número de aerogeneradores)	NO
OPERATOR	operador	Empresa que opera en el parque eólico	Cadena	1 Nivel (empresa que opera)	NO
MAINTAINER	mantenedor	Empresa que mantiene el parque eólico	Cadena	1 Nivel (empresa que mantiene)	NO
X	X	Coordenada X (longitud) de situación cartográfica del parque	Numérico decimal	1 Nivel (coordenada X)	NO
Y	Y	Coordenada Y (latitud) de situación cartográfica del parque	Numérico decimal	1 Nivel (coordenada Y)	NO

- Las columnas que se incluyen en esta fuente de datos para las **ocho hojas de datos restantes** son:

Nombre atributo (En el fichero)	Nombre atributo (En Castellano)	Descripción	Tipo de dato	Nivel Información	Opcional
DATE	Fecha	Fecha del registro	Temporal	3 Niveles (año, mes y día)	NO
POWER (KW/H)	Potencia (KW/H)	Potencia generada (kw/h)	Numérico entero	1 Nivel (potencia)	NO
WIND SPEED	Velocidad del viento	Velocidad del viento (m/s)	Numérico decimal	1 Nivel (fuerza del viento)	NO
TEMP	Temperatura	Temperatura (grados)	Numérico entero	1 Nivel (Temperatura)	NO
AVAILABILITY	Disponibilidad	Disponibilidad (%)	Numérico decimal	1 Nivel (disponibilidad)	NO
WAVE HIGHT	Altura de las olas	Altura de las olas (metros)	Numérico entero	1 Nivel (altura de las olas)	NO
REPARATION TIME (H)	Tiempo de reparación (H)	Tiempo de reparación (horas)	Numérico entero	1 Nivel (tiempo total de reparación)	NO
ALARMS GEARBOX	Alarmas de la caja de cambios	Número de alarmas en la caja de cambios (unidad)	Numérico entero	2 Niveles (alarma / tipo de alarma)	NO
ALARM GENERATOR	Alarmas del generador	Número de alarmas en el generador (unidad)	Numérico entero		NO
ALARM ROTOR	Alarmas del rotor	Número de alarmas en el rotor (unidad)	Numérico entero		NO
ALARM OTHER	Otras alarmas	Número de otras alarmas (unidad)	Numérico entero		NO

Una primera estimación del volumen de datos a realizar en la **primera carga de datos** en el data wharehouse seria:

Fuente de datos	Hoja de datos	Valores a almacenar	Total de registros
DATAOFFSHORE.xls 1 fichero	primera hoja de datos 1 hoja de datos	8 parques eólicos 9 datos	1 fichero X 1 hoja de datos X 8 parques eólicos X 9 datos

			= 72 registros
DATAOFFSHORE.xls 1 fichero	de las segunda a la novena hoja de datos 8 hojas de datos	1 registro al dia 31 días aprox 11 datos	1 fichero X 8 hoja de datos X 1 registros / día X 31 días aprox X 3 meses X 11 datos 8184 = registros

Una vez realizada la carga inicial, podríamos plantear:

- Una **carga incremental planificada diariamente** después de medianoche de los datos relevantes sobre los factores meteorológicos y técnicos de cada parque eólico obtenidos ese día.

- Una **carga de actualización planificada trimestralmente** con la finalidad de cargar nuevos datos o actualizar los datos de las diferentes tablas de dimensiones.

4.3. Análisis funcional.

En este apartado se describen los requerimientos funcionales para el diseño del sistema de inteligencia de negocio que cubra las necesidades de análisis para los datos obtenidos de los diferentes parques eólicos offshore:

Número	Requerimiento	Prioridad	Exigible / Deseable
1	Se extraerá de forma adecuada la información de las fuentes de datos (considerando sólo la información relevante)	1	E
2	Se creará un almacén de datos para almacenar y guardar todos los datos consolidados de las diferentes fuentes de datos de origen.	1	E
3	Se consolidará y cargará la informa-	1	E

	ción en el almacén de datos corporativo.		
4	Se crearan diversos cubos multidimensionales.	2	E
5	Se creara un sistema Front End conectado a los cubos multidimensionales y al data warehouse para facilitar el acceso y las consultas analíticas a los usuarios.	3	E

Cabe comentar que en un caso genérico podemos encontrar otros requerimientos funcionales como:

- Creación de procesos de calidad de datos.
- Creación de una Staging Area.
- Automatizar cada proceso de carga en la Staging Area (según sus necesidades).
- Creación de procesos de cargas totales e incrementales.
- Se creará un soporte a los metadatos de gestión del almacén de datos así como de los procesos ETL.

La arquitectura de la factoría de información puede estar formada por varios elementos pero, en el caso que nos ocupa y como hemos comentado en apartados anteriores, en este entorno de desarrollo del proyecto, estarán alojados en una misma máquina los siguientes elementos:

- **Almacén de datos corporativo / data warehouse:** El almacén de datos servirá como punto de almacenaje de toda la información de interés analítico. Esta información será obtenida de las diferentes fuentes de datos proporcionadas para este caso y se almacenarán en el data warehouse de forma integrada, no volátil y variable en el tiempo con el objetivo de ayudar a responder las consultas analíticas requeridas.

- **Cubos de dimensión:** A partir de la información del almacén de datos se creerán diferentes cubos multidimensionales para que puedan consultarse los datos de manera agregada con el fin de satisfacer las necesidades analíticas que se puedan requerir.

- **Aplicación web front-end:** Se proporcionara a los una aplicación con acceso web la cual dispondrá de acceso al almacén de datos y a los cubos de dimensión creados. Esta aplicación web, será el punto de acceso a los usuarios a los datos según sus necesidades analíticas.

El siguiente gráfico resume los elementos que encontraremos en dicha máquina según los requerimientos funcionales:

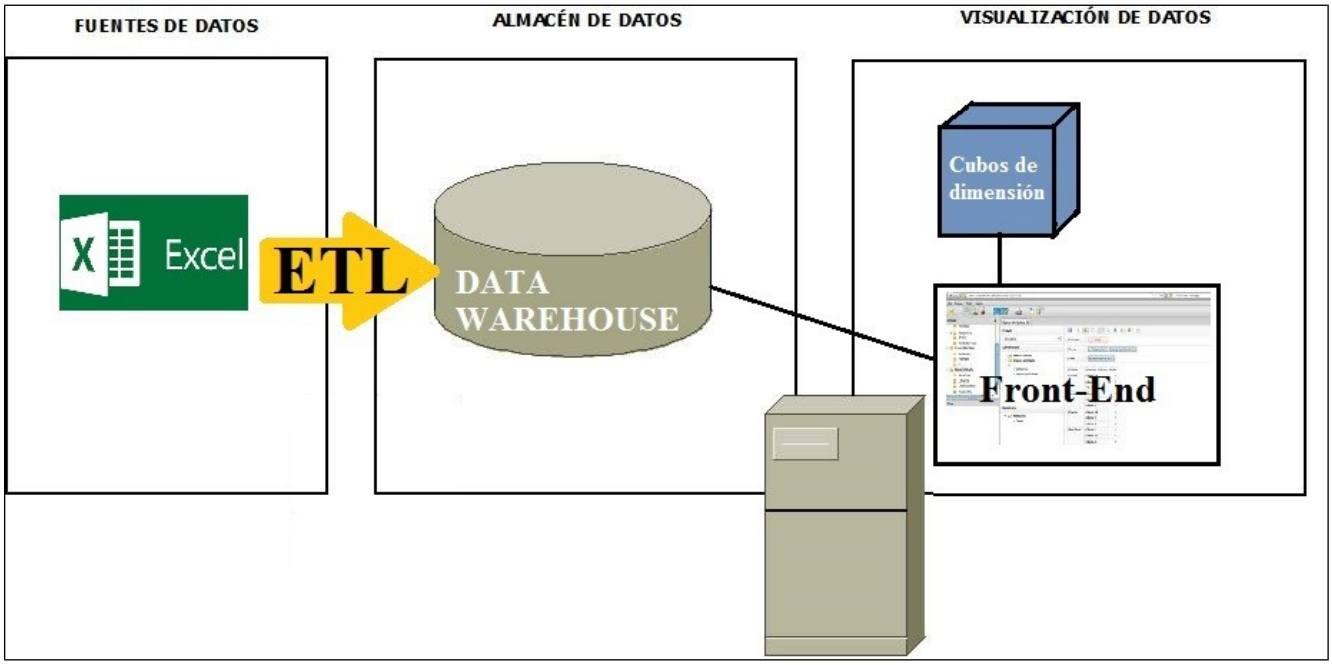

Ilustración 2 - Elementos según requerimientos funcionales

4.4. Diseño conceptual.

Para el correcto desarrollo del almacén de datos, es preciso definir las tablas de hechos, dimensiones de análisis, atributos y métricas que nos permitirán tener el nivel de agregado y información suficiente para cumplir los objetivos analíticos que se han definido anteriormente.

Para este sistema de inteligencia de negocio, se ha identificado un total de 4 dimensiones y una tabla de hechos con una métrica.

Tabla de hechos	Descripción
registro	Recoge la toma concreta de un dato o valor

Las dimensiones corresponderán a las perspectivas de negocio sobre las que queremos analizar los datos recogidos:

Tablas de dimensiones	Descripción
fecha	Fecha de registro del dato
parque eólico	Datos de cada parque eólico
variable	Variable concreta que se registrará con su respectiva categoría
contrato	Dimensión del tipo SCD con los detalles de cada empresa contratada.

El diagrama conceptual generado a partir del análisis realizado sería el siguiente:

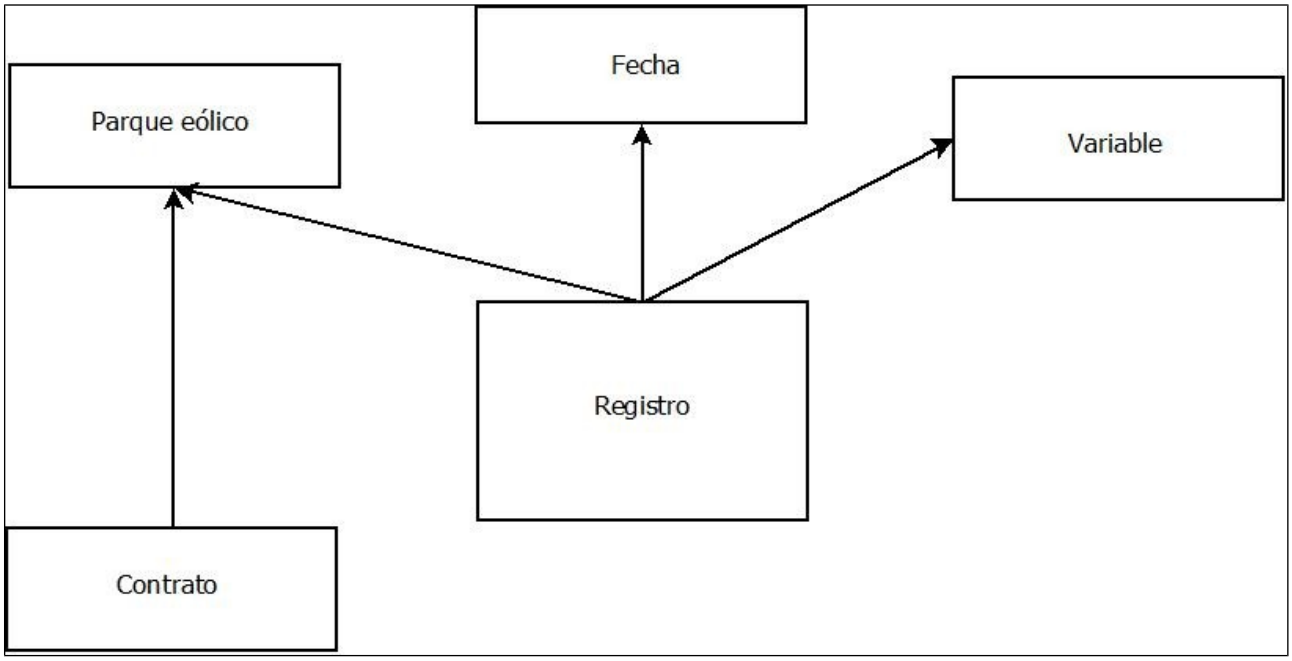

Ilustración 3 - Diagrama conceptual

4.5. Diseño lógico.

En el apartado de diseño lógico, para cada hecho y dimensión se han determinado sus métricas y atributos:

De cara a la tabla de hechos **f_registro,** se ha establecido la siguiente métrica:

Tabla de hecho	Métricas	Descripción
f_registro	valor	Recoge los valores numéricos obtenidos

		por las sondas de cada parque eólico. (Meteorología, alarmas, disponibilidad…)

Para las dimensiones detectadas anteriormente, se han establecido los siguientes atributos:

- **Dimensión d_fecha:**

d_fecha	
Atributos	**Descripción**
fecha	Recoge la fecha en la que se obtuvo el valor de la tabla de hechos (dd/mm/yyyy)
dia	Recoge el día en el que se obtuvo el valor de la tabla de hechos (dd)
mes	Recoge el mes en el que se obtuvo el valor de la tabla de hechos (mm)
anyo	Recoge el año en el que se obtuvo el valor de la tabla de hechos (yyyy)

- **Dimensión d_parque_eolico:**

d_parque_eolico	
Atributos	**Descripción**
parque_eolico	Nombre del parque eólico
pais	País donde se encuentra el parque eólico
marca_aerogen	Marca de los aerogeneradores del parque eólico
numero_aerogen	Número de aerogeneradores disponibles en el parque eólico
potencia_aerogen	Potencia máxima de cada generador del parque eólico
coordenada_x	Coordenada X (longitud) de situación cartográfica del parque eólico
coordenada_y	Coordenada Y (latitud) de situación cartográfica del parque eólico

- **Dimensión d_variable:**

d_variable	
Atributos	**Descripción**
categoría	Describe el tipo de valor registrado en la tabla de hechos (Meteorología, alarmas, disponibilidad…)
subcategoría	Este campo es un subnivel de la categoría (Caja de cambios, generador, rotor, otros…)
variable	Registra la variable del valor registrado en la tabla de hechos (mediana o sumatorio de valores del dia)
unidad	Registra la unidad de medida del valor registrado en la tabla de hechos (m/s, km/h, %...)

- **Dimensión d_contrato:**

d_contrato	
Atributos	**Descripción**
parque_eolico	Nombre del parque eólico
empresa	Empresa en contrato en el parque eólico
servicio	Servicio que ofrece la empresa dentro del parque eólico. (mantenimiento o operador)
fecha_inicio	Fecha de inicio de contrato con la empresa
fecha_fin	Fecha fin de contrato con la empresa
contrato_activo	Valor que indica si la empresa tiene un contrato en vigor actualmente o no

La dimensión **d_contrato** a diferencia del resto, se ha creado siguiendo un modelo **SCD (Slowly Changing Dimension) del tipo 2.** Esto se ha realizado así debido a que los registros de contratos de esta dimensión sufrirán cambios en el tiempo que deben permanecer en esta (cambios de empresa de mantenimiento o operadora de cada parque eólico), donde se informa en todo momento de la fecha de inicio de contrato, fecha fin de contrato con cada empresa y informando si es un contrato en vigor.

Este tipo de esquema con dimensiones extra de apoyo recibe el nombre de **esquema en copo de nieve**.

El diagrama lógico generado a partir del análisis realizado sería el siguiente:

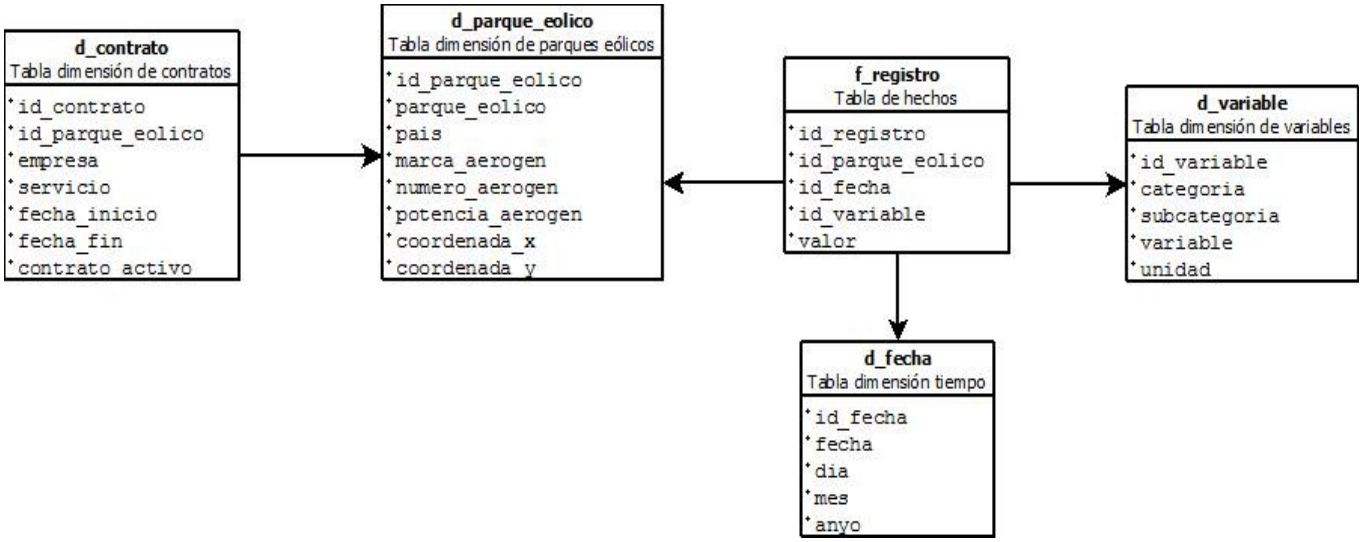

Ilustración 4 - Diagrama lógico

4.6. Diseño físico.

Si consideramos cada una de las tablas (hechos y dimensiones) así como el detalle de cada una de las métricas y atributos que las componen, el diseño físico materializado al tipo de dato y script a crear en base de datos.

La base de datos creada en **PostgreSQL** donde se almacenarán los datos para este sistema de inteligencia de negocio se denomina ***offshore.*** El esquema donde se crearan cada una de las tablas será en el esquema por defecto ***public*** de esta base de datos.

A continuación se detallan las características físicas de cada tabla siguiendo el orden de creación en que se ha de ejecutar las sentencias DDL en base de datos:

- **Dimensión d_parque_eolico:**

d_parque_eolico		
Nombre del campo	**Tipo de dato**	**Descripción**
id_parque_eolico	INTEGER	Clave Primaria
parque_eolico	VARCHAR(30)	Atributo/ No nulo/ Único
pais	VARCHAR(30)	Atributo/ No nulo
marca_aerogen	VARCHAR(30)	Atributo/ No nulo
numero_aerogen	INTEGER	Atributo/ No nulo

potencia_aerogen	VARCHAR(10)	Atributo/ No nulo
coordenada_x	VARCHAR(15)	Atributo/ No nulo
coordenada_y	VARCHAR(15)	Atributo/ No nulo

Script

```
create table d_parque_eolico(
id_parque_eolico INTEGER NOT NULL,
parque_eolico VARCHAR(30) NOT NULL UNIQUE,
pais VARCHAR(30) NOT NULL,
marca_aerogen VARCHAR(30) NOT NULL,
numero_aerogen INTEGER NOT NULL,
potencia_aerogen VARCHAR(10) NOT NULL,
coordenada_x VARCHAR(15) NOT NULL,
coordenada_y VARCHAR(15) NOT NULL,

CONSTRAINT pk_parque_eolico PRIMARY KEY (id_parque_eolico)
);
```

- **Dimensión d_contrato:**

d_contrato		
Nombre del campo	**Tipo de dato**	**Descripción**
id_contrato	INTEGER	Clave Primaria
id_parque_eolico	INTEGER	Clave Foránea
empresa	VARCHAR(30)	Atributo / No nulo
servicio	VARCHAR(15)	Atributo / No nulo
fecha_inicio	DATE	Atributo/ No nulo Valor por defecto: fecha de hoy
fecha_fin	DATE	Atributo/ No nulo Valor por defecto: 01/01/9999
contrato_activo	BOOLEAN	Atributo/ No nulo Valor por defecto: true

Script

```
create table d_contrato(
id_contrato INTEGER NOT NULL,
id_parque_eolico INTEGER NOT NULL,
empresa VARCHAR(30) NOT NULL,
servicio VARCHAR(15) NOT NULL,
fecha_inicio DATE NOT NULL CHECK (fecha_inicio >= current_date) DE-
FAULT current_date,
fecha_fin DATE NOT NULL CHECK (fecha_inicio >= current_date) DEFAULT
'01/01/9999',
contrato_activo BOOLEAN NOT NULL DEFAULT true,

CONSTRAINT pk_contrato PRIMARY KEY (id_contrato),
FOREIGN KEY (id_parque_eolico) REFERENCES d_parque_eolico (id_par-
que_eolico)
```

```
);
```

Para esta dimensión se ha han añadido una cláusula check de seguridad para que siempre que se proceda a informar el campo ***fecha_inicio*** o ***fecha_fin***, el valor a insertar sea igual o superior a la fecha actual.

En esta dimensión también se ha creado un **trigger before insert** (antes de insertar) que comprueba si en la dimensión **d_contrato** existe un registro con **id_parque_eolico** similar al registro a insertar y con **contrato_activo = true**. De existir este registro, el trigger procede automáticamente a dar de baja ese contrato activo en base de datos y dar de alta el nuevo contrato.

```
---------------------------------------------------------------------------------------------
--
-- Creación o remplazamiento de la función sp_registro_activo
--
-----------------------------------------------------------------------------------------------

CREATE OR REPLACE FUNCTION sp_registro_activo()
RETURNS TRIGGER AS $$

/*
 *  Procedimiento: sp_registro_activo()
 *  Autor: Robert Berchtold Palacios
 *  Fecha creación: 16/04/2017
 *  Versión 1.0
 *  Parámetros: sin parámetros
  *    Descripción: Procedimiento creado para ser ejecutado en el trigger 'tg_i_registro_activo'.
 *  Este procedimiento se ejecuta cada vez que se inserta un nuevo registro en la dimensión d_contrato.
  *    Comprueba si existe en esta tabla un registro con el identificador id_parque_eolico que se está insertando, si existe algún registro comprueba si ese registro es un contrato activo
 *  de ser así, se actualiza el registro de base de datos actualizando el campo fecha_fin y estbaleciendo el campo contrato_activo a false.
 */

DECLARE
        v_count_parque_man INTEGER;   -- Variable para calcular el nº de lineas totales activas con servicio MANTENEDOR para un parque concreto.
        v_count_parque_oper INTEGER; -- Variable para calcular el nº de lineas totales activas con servicio OPERADOR para un parque concreto.
BEGIN
         -- Se comprueba si existe un contrato_activo en base de datos con el parque eolico que se desea insertar para un servicio MANTENEDOR.
```

```
        SELECT
            COUNT(*)
        INTO
            v_count_parque_man
        FROM
            d_contrato
        WHERE
            id_parque_eolico = NEW.id_parque_eolico
            AND servicio = 'MANTENEDOR'
            AND contrato_activo = true;

    -- Se comprueba si existe un contrato_activo en base de datos con el parque eolico que se desea insertar para un servicio OPERADOR.

        SELECT
            COUNT(*)
        INTO
            v_count_parque_oper
        FROM
            d_contrato
        WHERE
            id_parque_eolico = NEW.id_parque_eolico
            AND servicio = 'OPERADOR'
            AND contrato_activo = true;

    -- Si existe un contrato activo con servicio MANTENEDOR para el parque a insertar, se procede automaticamente a darlo de baja.

    IF (v_count_parque_man > 0 and NEW.servicio = 'MANTENEDOR') THEN

            UPDATE
                d_contrato
            SET
                contrato_activo = false,
                fecha_fin = current_date
            WHERE
                id_parque_eolico = NEW.id_parque_eolico
                AND servicio = 'MANTENEDOR'
                AND contrato_activo = true;
    END IF;

    -- Si existe un contrato activo con servicio MANTENEDOR para el parque a insertar, se procede automaticamente a darlo de baja.

    IF (v_count_parque_oper > 0 and NEW.servicio = 'OPERADOR') THEN

            UPDATE
                d_contrato
            SET
```

```
                    contrato_activo = false,
                    fecha_fin = current_date
                WHERE
                    id_parque_eolico = NEW.id_parque_eolico
                    AND servicio = 'OPERADOR'
                    AND contrato_activo = true;
        END IF;


        RETURN NEW;
END;
$$ LANGUAGE plpgsql;
-------------------------------------------------------------------------------------------
--
-- Creación o remplazamiento del trigger tg_i_registro_activo
--
-------------------------------------------------------------------------------------------

CREATE TRIGGER tg_i_registro_activo
BEFORE INSERT
ON d_contrato
FOR EACH ROW
EXECUTE PROCEDURE sp_registro_activo();
```

- **Dimensión d_fecha:**

d_fecha		
Nombre del campo	**Tipo de dato**	**Descripción**
id_fecha	INTEGER	Clave Primaria
fecha	DATE	Atributo/ No nulo
dia	INTEGER	Atributo/ No nulo
mes	INTEGER	Atributo/ No nulo
anyo	INTEGER	Atributo/ No nulo

Script

```
create table d_fecha(
id_fecha INTEGER NOT NULL,
fecha DATE NOT NULL,
dia INTEGER NOT NULL,
mes INTEGER NOT NULL,
anyo INTEGER NOT NULL,

CONSTRAINT pk_fecha PRIMARY KEY (id_fecha)
);
```

- **Dimensión d_variable:**

d_variable		
Nombre del campo	**Tipo de dato**	**Descripción**
id_variable	INTEGER	Clave Primaria
categoria	VARCHAR(30)	Atributo / No nulo
subcateogria	VARCHAR(30)	Atributo / No nulo
variable	VARCHAR(30)	Atributo / No nulo
unidad	VARCHAR(10)	Atributo / No nulo

Script

```
create table d_variable(
id_variable INTEGER NOT NULL,
categoria VARCHAR(30) NOT NULL,
subcategoria VARCHAR(30) NOT NULL,
variable VARCHAR(30) NOT NULL,
unidad VARCHAR(10) NOT NULL,

CONSTRAINT pk_variable PRIMARY KEY (id_variable)
);
```

- **Hecho f_registro:**

f_registro		
Nombre del campo	**Tipo de dato**	**Descripción**
id_registro	SERIAL	Clave Primaria / AutoIncremental
id_parque_eolico	INTEGER	Clave Foránea
id_fecha	INTEGER	Clave Foránea
id_variable	INTEGER	Clave Foránea
valor	NUMERIC(6,2)	Métrica

Script

```
create table f_registro(
id_registro SERIAL NOT NULL,
id_parque_eolico INTEGER NOT NULL,
id_fecha INTEGER NOT NULL,
id_variable INTEGER NOT NULL,
valor NUMERIC(6,2) NOT NULL,

CONSTRAINT pk_registro PRIMARY KEY (id_registro),
FOREIGN KEY (id_parque_eolico) REFERENCES d_parque_eolico (id_parque_eolico),
FOREIGN KEY (id_fecha) REFERENCES d_fecha (id_fecha),
FOREIGN KEY (id_variable) REFERENCES d_variable (id_variable)
);
```

El **script DDL** resultante para la creación en base de datos de todas las tablas y el trigger de la dimensión d_contrato se encuentra en el Anexo 6: Script DDL.

El diagrama físico generado a partir del análisis realizado sería el siguiente:

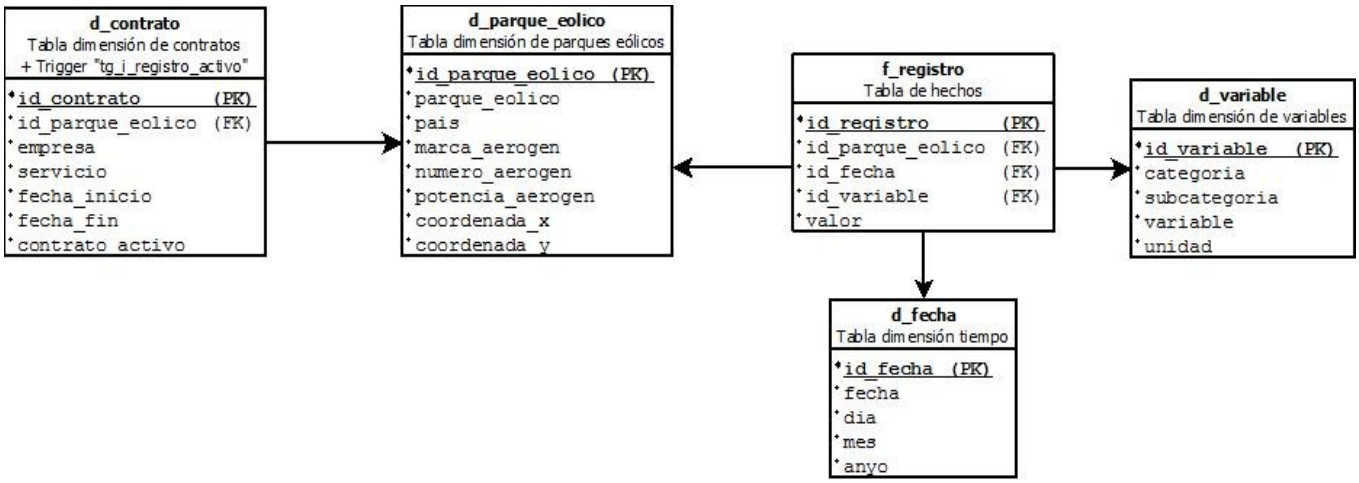

Ilustración 5 - Diagrama físico

5. Carga de datos mediante procesos ETL.

5.1. Identificar los procesos ETL necesarios.

Los procesos ETL deben conceptualizarse como manipulaciones de flujos de datos. Estos procesos deben diseñarse teniendo en cuenta distintos factores como los siguientes.

- Cómo debe cargarse de manera lógica la información, es decir, qué debe cargarse primero y qué después.

- La ventana de tiempo disponible, hecho que puede condicionar lo que debemos cargar.

- Identificar si los datos se deben cargar en un área intermedia. En nuestro caso, no es necesario.

- Tipo de carga: inicial, incremental o actualización.

Para nuestro caso, se trata de una carga inicial e incremental automatizada y periódica, por lo tanto, nuestros procesos deben estar preparados para estos objetivos.

Desconocemos la ventana de tiempo disponible, pero en nuestro caso sería irrelevante, ya que la carga se realiza mediante ficheros Excel o a través de sentencias INSERT ejecutadas directamente en el cliente del gestor de base de datos **pgAdminIII.**

Respecto a la realización del proceso de transformación, aparecen tres tipos de situaciones:

- **Dimensiones con valores fijos:** Con una variación en el tiempo mínima o directamente no varían en el tiempo. Estas dimensiones son:

  1) d_parque_eolico
  2) d_variable
  3) d_contrato

  Estas dimensiones se cargarán mediante ficheros Excel (fichero Excel proporcionado para el trabajo o ficheros Excel generados manualmente) o sentencias INSERT ejecutadas a través del cliente del PostgreSQL.

- **Dimensiones con valores no fijos:** Con una variación en el tiempo ocasional o constante. Estas dimensiones son:

  1) d_tiempo

  Para el caso de la dimensión tiempo, los registros y datos de esta dimensión se irán generando en paralelo a través de la carga de datos ETL de la tabla de hechos f_registro.

- **Tabla de hecho:** Se extraerán, transformarán y cargarán los datos para esta tabla mediante procesos ETL mediante el fichero Excel proporcionado para este trabajo. Esta tabla de hecho tiene la nomenclatura f_registro.

Hay que tener en cuenta que los datos pertenecientes a las dimensiones se cargaran antes que los datos pertenecientes a tabla de hecho, ya que la tabla de hecho depende directamente de los datos almacenados en las dimensiones.

5.2. Descripción de las acciones a realizar.

Se han identificado un total de 11 procesos de carga. Estos procesos de carga se dividen de la siguiente forma:

- 1 proceso de carga automática mediante Excel de carga personalizado.
- 10 procesos de carga automática mediante el Excel proporcionado para este trabajo.

Los ficheros Excel utilizados para el proceso de carga de datos en base de datos son:

Anexo 7: Fichero Excel DATAOFFSHORE.xlsx (Proporcionado por la UOC).

Anexo 8: Fichero Excel CARGA-DVARIABLE.xlsx (Personalizado).

En el siguiente cuadro se describen las acciones a realizar para cada uno de los procesos de carga identificados.

Proceso	Fichero	Descripción	Transformación	Subtareas
Inicial / Incremental	CARGA-DVARIABLE.xlsx	Carga de datos de la dimensión d_variable mediante proceso ETL.	TRA_D_VARIABLE	Extraer valores del Excel. Convertir valores a mayúsculas. Insertar valores en dimensión d_variable.
Inicial / Incremental	DATAOFFSHORE.xlsx	Carga de datos de la dimensión d_parque_eolico mediante proceso ETL.	TRA_D_PARQUE_EOLICO	Extraer valores del Excel. Filtrar campos no necesarios. Convertir valores a mayúsculas. Traducir valores del

				inglés al castellano Insertar valores en dimensión d_parque_eolico
Inicial / Incremental / Actualización	DATAOFFSHORE.xlsx	Carga de datos de la dimensión d_contrato mediante proceso ETL.	TRA_D_CONTRATO	Extraer valores del Excel. Filtrar campos no necesarios. Convertir valores a mayúsculas. Traducir valores del inglés al castellano. Obtener claves ID Insertar valores en dimensión d_contrato
Inicial / Incremental	DATAOFFSHORE.xlsx	Carga de datos de la dimensión d_fecha y de la tabla de hecho f_registro mediante proceso ETL para el parque eólico NAMPER	TRA_H_REGISTRO_NAMPER	Extraer valores del Excel. Realizar una agrupación y agregado de valores a partir del campo fecha. Obtener claves IDs. Obtener dia, mes y anyo a partir de la fecha Insertar valores fecha, dia, mes anyo en la dimensión d_fecha y obtener clave ID. Obtener el id_variable para cada variable a cargar. Insertar valores en la tabla de hecho f-

				registro.
Inicial / Incremental	DATAOFFSHORE.xlsx	Carga de datos de la dimensión d_fecha y de la tabla de hecho f_registro mediante proceso ETL para el parque eólico RIAS BAIXAS	TRA_H_REGISTRO_RIASBAIXAS	Extraer valores del Excel. Realizar una agrupación y agregado de valores a partir del campo fecha. Obtener claves IDs. Obtener dia, mes y anyo a partir de la fecha Insertar valores fecha, dia, mes anyo en la dimensión d_fecha y obtener clave ID. Obtener el id_variable para cada variable a cargar. Insertar valores en la tabla de hecho f-registro.

Inicial / Incremental	DATAOFFSHORE.xlsx	Carga de datos de la dimensión d_fecha y de la tabla de hecho f_registro mediante proceso ETL para el parque eólico GUARACHICO	TRA_H_REGISTRO_GUARACHICO	Extraer valores del Excel. Realizar una agrupación y agregado de valores a partir del campo fecha. Obtener claves IDs. Obtener dia, mes y anyo a partir de la fecha Insertar valores fecha, dia, mes anyo en la dimensión d_fecha y obtener clave ID. Obtener el id_variable para cada variable a cargar. Insertar valores en la tabla de hecho f-registro.

Inicial / Incremental	DATAOFFSHORE.xlsx	Carga de datos de la dimensión d_fecha y de la tabla de hecho f_registro mediante proceso ETL para el parque eólico KIRSKEN	TRA_H_REGISTRO_KIRSKEN	Extraer valores del Excel. Realizar una agrupación y agregado de valores a partir del campo fecha. Obtener claves IDs. Obtener dia, mes y anyo a partir de la fecha Insertar valores fecha, dia, mes anyo en la dimensión d_fecha y obtener clave ID. Obtener el id_variable para cada variable a cargar. Insertar valores en la tabla de hecho f-registro.
Inicial / Incremental	DATAOFFSHORE.xlsx	Carga de datos de la dimensión d_fecha y de la tabla de hecho f_registro mediante proceso ETL para el parque eólico POLVARS	TRA_H_REGISTRO_POLVARS	Extraer valores del Excel. Realizar una agrupación y agregado de valores a partir del campo fecha. Obtener claves IDs. Obtener dia, mes y anyo a partir de la fecha Insertar valores fecha, dia, mes anyo en la dimensión d_fecha

				y obtener clave ID. Obtener el id_variable para cada variable a cargar. Insertar valores en la tabla de hecho f-registro.
Inicial / Incremental	DATAOFFSHORE.xlsx	Carga de datos de la dimensión d_fecha y de la tabla de hecho f_registro mediante proceso ETL para el parque eólico COUNSCOT	TRA_H_REGISTRO_COUNSCOT	Extraer valores del Excel. Realizar una agrupación y agregado de valores a partir del campo fecha. Obtener claves IDs. Obtener dia, mes y anyo a partir de la fecha Insertar valores fecha, dia, mes anyo en la dimensión d_fecha y obtener clave ID. Obtener el id_variable para cada variable a cargar. Insertar valores en la tabla de hecho f-registro.

Inicial / Incremental	DATAOFFSHORE.xlsx	Carga de datos de la dimensión d_fecha y de la tabla de hecho f_registro mediante proceso ETL para el parque eólico GREENBLUE	TRA_H_REGISTRO_GREENBLUE	Extraer valores del Excel. Realizar una agrupación y agregado de valores a partir del campo fecha. Obtener claves IDs. Obtener dia, mes y anyo a partir de la fecha Insertar valores fecha, dia, mes anyo en la dimensión d_fecha y obtener clave ID. Obtener el id_variable para cada variable a cargar. Insertar valores en la tabla de hecho f-registro.
Inicial / Incremental	DATAOFFSHORE.xlsx	Carga de datos de la dimensión d_fecha y de la tabla de hecho f_registro mediante proceso ETL para el parque eólico NORTHENCAP	TRA_H_REGISTRO_NORTHENCAP	Extraer valores del Excel. Realizar una agrupación y agregado de valores a partir del campo fecha. Obtener claves IDs. Obtener dia, mes y anyo a partir de la fecha Insertar valores fecha, dia, mes anyo en la dimensión d_fecha

				y obtener clave ID. Obtener el id_variable para cada variable a cargar. Insertar valores en la tabla de hecho f-registro.

5.3. Implementación de procesos ETL.

Los pasos a seguir y las configuraciones a realizar para la generación de las transformaciones ETL correspondientes vistas en la tabla del apartado anterior se pueden encontrar en el siguiente documento:

Anexo 9: Manual de implantación - Creación de transformaciones

La ejecución de las transformaciones se ha de realizar a partir de jobs creados en **Pentaho**. Los detalles de cada Job se describen en la siguiente tabla:

Nombre del Job	Tipo de ejecución	Ejecución	Tablas afectadas
JOB_DIARIO	Inicial / Incremental	Diaria	d_fecha f_registro
JOB_TRIMESTRAL	Inicial / Incremental / Actualización	Trimestral	d_variable d_parque_eolico d_contrato

Los pasos a seguir y las configuraciones a realizar para la generación de los Jobs correspondientes se pueden encontrar en el siguiente documento:

Anexo 10: Manual de implantación - Creación de jobs

Todos los Jobs serán creados en **Pentaho Data Integration** y se programarán su ejecución mediante el planificador de tareas que incluya el sistema operativo. En nuestro caso al ser un sistema operativo Windows, ejecutaremos los correspondientes Jobs a partir de scripts **.bat** que contendrán la siguiente línea de ejecución en su programación.

```
Kitchen.(bat/sh) --file=path_to_file
```

Estos ficheros **.bat** a su vez serán ejecutados mediante el **planificador de tareas de Windows** en los periodos descritos en la tabla anterior.

6. Diseño multidimensional

6.1. Identificación de las necesidades analíticas.

Una vez efectuada la carga de datos en nuestro ***data warehouse***, es el momento de crear un modelo de análisis mediante la implantación de cubos analíticos multidimensionales (OLAP).

Los cubos multidimensionales permiten analizar los datos de gran volumen y variedad de nuestro ***data warehouse*** con una gran agilidad y rapidez, reduciendo enormemente el tiempo y los recursos empleados para el análisis.

En nuestro sistema de inteligencia de negocio para parques eólicos offshore dispondremos de un cubo de análisis multidimensional denominado ***offshore***.

Este cubo de análisis multidimensional nos permitirá responder las preguntas analíticas recogidas en el análisis de requerimientos. Las preguntas a resolver son:

1) Cuáles son los parques eólicos más productivos.

2) Hacer un análisis profundo sobre los datos meteorológicos de cada parque y las incidencias sobre la producción.

3) Zonas con mejor relación viento/potencia.

4) Generación de alarmas respecto a las variaciones meteorológicas.

5) Efectividad de las empresas de mantenimiento.

6) Relaciones entre variables de producción y meteorológicas.

6.2. Creación e implementación de cubos analíticos.

Los pasos a seguir para la creación de nuestro cubo de análisis multidimensional denominado ***offshore*** y para la implementación de este en nuestro front-end web de análisis multidimensional se detallan en el siguiente documento:

Anexo 11: Manual de implantación - Cubo de análisis multidimensional

7. Análisis multidimensional.

7.1. Interpretación de resultados.

1) Parques eólicos más productivos.

Según el análisis realizado, los parques eólicos offshore más productivos a partir de los datos que disponemos del 2016 son:

- El primer parque eólico más productivo es NORTHENCAP en NORUEGA.
- El segundo parque eólico más productivo es POLVARS en POLONIA.
- El tercer parque eólico más productivo es KIRSKEN en ALEMANIA.

Para obtener esta clasificación se ha realizado una división entre la potencia total anual generada por cada parque entre el número de aerogeneradores del parque. Realizando este cálculo hemos visualizado aproximadamente que potencia se ha generado por cada aerogenerador.

El parque eólico de NORTHENCAP a pesar de tener un menor número de aerogeneradores ha generado más potencia por aerogenerador gracias a que los datos de viento medios del año han sido más favorables en este parque.

Respecto a los parques eólicos offshore de POLVARS y KIRSKEN, si analizamos la potencia media calculada por aerogenerador, vemos que a pesar de tener un viento menor que NORTHENCAP, los resultados de los valores de potencia por aerogenerador son muy elevados y prácticamente similares NORTHENCAP. Esto podría ser debido a que utilizan la misma marca de aerogeneradores (ENERCON).

anyo	parque eolic	pais	marca aerogeneradore	numero aerogeneradore	potencia aerogeneradore	DISPONIBILIDA	POTENC	VIENT	POTENCIA / AEROG
2016	NORTHENCAP	NORUEGA	ALSTON	33	2.5 MW	98,465	8.164,435	22,451	247,4071212
	POLVARS	POLONIA	ENERCON	40	2.5 MW	98,644	8.678,37	17,322	216,95925
	KIRSKEN	ALEMANIA	ENERCON	25	2.5 MW	98,662	4.013,217	17,178	160,52868
	GREENBLUE	IRLANDA	ALSTON	10	3.0 MW	95,706	1.269,609	18,776	126,9609
	COUNSCOT	REINO UNIDO	VESTAS	29	2.0 MW	98,928	3.681,402	18,792	126,9448966
	NAMPER	DINAMARCA	VESTAS	15	2.0 MW	99,643	852,272	7,588	56,81813333
	RIAS BAIXAS	ESPAÑA	ALSTON	25	2.5 MW	99,353	819,641	4,516	32,78564
	GUARACHICO	ESPAÑA	GAMESA	30	2.0 MW	98,408	942,533	4,489	31,41776667

Ilustración 6 - Análisis parques eólicos offshore más productivos

2) Zonas con mejor relación viento/potencia.

Según el análisis realizado, los parques eólicos offshore con mejor relación viento/potencia a partir de los datos que disponemos del 2016 son:

- El primer parque eólico con mejor relación viento/potencia es POLVARS en POLONIA.

- El segundo parque eólico con mejor relación viento/potencia es NORTHENCAP en NORUEGA.

- El tercer parque eólico más productivo con mejor relación viento/potencia RIAS BAIXAS en ESPAÑA.

Para obtener esta clasificación se han generado tres columnas en el Excel los cálculos necesarios.

La primera columna con nomenclatura “potencia por aerogenerador” realiza una división de la potencia anual media de cada parque eólico entre el número de generadores que tiene ese parque. Este cálculo nos dará como resultado la potencia media correspondiente que ha generado cada aerogenerador.

Una vez obtenida la “potencia por aerogenerador”, en la segunda columna con nomenclatura “comparativa con northencap” se realiza una regla de tres comparando cada parque eólico de la lista con el parque eólico con mas viento medio anual que es NORTHENCAP. El objetivo de este cálculo es calcular que potencia media anual se debería haber generado en cada parque a partir de relacionar el viento medio anual de ese parque, con la potencia media anual y el viento medio anual de NORTHENCAP.

Finalmente, la última columna con nomenclatura “diferencia” se realiza una resta entre el resultado de las columnas “potencia por aerogenerador” y “comparativa con northencap”. El resultado es la diferencia de potencia que existe en cada parque eólico comparando si en NORTHENCAP se hubiera tenido un viento y un numero de aerogeneradores similares.

anyo	parque eolico	pais	marca aerogeneradore	umero aerogeneradores	otencia aerogeneradore	DISPONIBILIDAD	POTENCIA	VIENTO	POTENCIA POR AEROGENERADOR	COMPARATIVA CON NOTHERCAP	DIFERENCIA
2016	NORTHENCAP	NORUEGA	ALSTON	33	2.5 MW	98,465	8.164,435	22,451	247,4071212	-	-
	POLVARS	POLONIA	ENERCON	40	2.5 MW	98,644	8.678,37	17,322	216,95925	190,8862034	26,0730466
	KIRSKEN	ALEMANIA	ENERCON	25	2.5 MW	98,662	4.013,217	17,178	160,52868	189,299342	-28,770662
	GREENBLUE	IRLANDA	ALSTON	10	3.0 MW	95,706	1.269,609	18,776	126,9609	206,9090957	-79,948196
	COUNSCOT	REINO UNIDO	VESTAS	29	2.0 MW	98,928	3.681,402	18,792	126,9448966	207,0854136	-80,140517
	NAMPER	DINAMARCA	VESTAS	15	2.0 MW	99,643	852,272	7,588	56,81813333	83,61878027	-26,800647
	RIAS BAIXAS	ESPAÑA	ALSTON	25	2.5 MW	99,353	819,641	4,516	32,78564	49,76573691	-16,980097
	GUARACHICO	ESPAÑA	GAMESA	30	2.0 MW	98,408	942,533	4,489	31,41776667	49,4682004	-18,050434

Ilustración 7 - Análisis zonas con mejor relación viento/potencia

3) Generación de alarmas respecto a las variaciones meteorológicas.

Según el análisis realizado sobre la generación de alarmas respecto a las variaciones meteorológicas a partir de los datos que disponemos del 2016 se ha obtenido como resultado que el viento es el factor meteorológico que mas influye en la generación de alarmas. También se ha detectado, aunque no es tan común como el viento, que la temperatura también puede ser un factor para la generación de algunas alarmas.

Las alarmas mas comunes son las alarmas de caja de cambios, a mayor viento y temperatura empiezan a aparecer alarmas de generador, rotor y otros tipos de alarmas no identificadas.

parque eolico	pais	ALARMAS	CAJA DE CAMBIOS	GENERADOR	OTROS	ROTOR	ALTURA OLAS	TEMPERATURA	VIENTO
POLVARS	POLONIA	2.062	1.958	86	0	18	8	10	18
POLVARS	POLONIA	1.994	1.852	120	0	22	7	10	17,4
POLVARS	POLONIA	1.970	1.840	102	0	28	7	10	17
POLVARS	POLONIA	1.960	1.832	106	0	22	7	11	17,9
POLVARS	POLONIA	1.936	1.802	122	0	12	7	11	17,1
POLVARS	POLONIA	1.924	1.810	98	0	16	7	11	16,4
POLVARS	POLONIA	1.922	1.812	94	0	16	8	10	18,1
POLVARS	POLONIA	1.920	1.818	86	0	16	7	10	16,3
POLVARS	POLONIA	1.916	1.798	104	0	14	7	10	17,6
POLVARS	POLONIA	1.900	1.774	98	0	28	7	11	17,8
POLVARS	POLONIA	1.896	1.796	88	0	12	7	10	16
POLVARS	POLONIA	1.884	1.768	100	0	16	7	11	17,2
NORTHENCAP	NORUEGA	1.880	1.794	86	0	0	11	6	22,9

Ilustración 8 - Análisis generación de alarmas respecto meteorología

4) Efectividad de las empresas de mantenimiento.

Según el análisis realizado sobre las empresas de mantenimiento mas efectivas a partir de los datos que disponemos del 2016, la empresa más efectiva realizando una comparación entre la suma de alertas totales, el tiempo de reparación total i el tiempo de reparación medio por alerta para cada parque eólico, nos aparece que la empresa MANWIN es la mas efectiva.

Cabe destacar, según los datos disponibles, que las alertas mas comunes son las alertas de caja de cambios.

Además, cabe decir, que las alertas de generador y rotor parecen más complicadas de solucionar y estas hacen aumentar el tiempo de reparación global para cada parque, por lo tanto, para realizar correctamente este análisis se requeriría incorporar a la ETL, los tiempos de SLA/SLO de origen obtenidos y permitidos para cada tipo de incidencia. Con estos datos sabríamos los incumplimientos de SLA/SLO por empresa en cada parque y el análisis de efectividad seria mucho mas correcto y efectivo.

anyo	parque eolico	pais	servicio	empresa	fecha inicio	fecha fin	contrato activo	ALARMA	CAJA DE CAMBIOS	GENERADOR	OTROS	ROTOR	TIEMPO REPARACION	Tiempo por Alerta
2016	NAMPER	DINAMARCA	MANTENEDOR	GDEP	2017-05-06	9999-01-01	true	6.951	6.852	91	4	4	4.569	1,521339462
	RIAS BAIXAS	ESPAÑA	MANTENEDOR	MANWIN	2017-05-06	9999-01-01	true	21.632	20.747	879	3	3	8.315	2,60156344
	COUNSCOT	REINO UNIDO	MANTENEDOR	CUPRA	2017-05-06	9999-01-01	true	43.283	39.814	3.417	4	48	13.281	3,25901664
	GREENBLUE	IRLANDA	MANTENEDOR	MANWIN	2017-05-06	9999-01-01	true	57.550	55.089	2.444	8	9	11.089	5,189827757
	NORTHENCAP	NORUEGA	MANTENEDOR	CUPRA	2017-05-06	9999-01-01	true	68.188	64.977	3.185	4	22	12.990	5,249268668
	KIRSKEN	ALEMANIA	MANTENEDOR	GDEP	2017-05-06	9999-01-01	true	45.535	42.879	2.648	4	4	8.594	5,298464045
	POLVARS	POLONIA	MANTENEDOR	GDEP	2017-05-06	9999-01-01	true	74.778	69.613	4.338	4	823	13.929	5,368511738
	GUARACHICO	ESPAÑA	MANTENEDOR	CUPRA	2017-05-06	9999-01-01	true	63.005	61.154	1.805	3	43	10.203	6,175144565

Ilustración 9 - Análisis efectividad de las empresas de mantenimiento

5) Relaciones entre variables de producción y meteorológicas.

Según el análisis realizado sobre la relación entre las variables de producción y meteorológicas a partir de los datos que disponemos del 2016, se ha obtenido como resultado que el viento es el principal factor meteorológico para la generación de potencia en los parques eólicos. A mayor velocidad del viento, mas potencia podremos generar en nuestro parque eólico.

Hay que tener presente que el viento también es uno de los factores principales para la generación de alertas y incidencias, por lo tanto, para obtener una buena producción en el parque, además de situar el parque en una zona con velocidades del viento elevadas, deberemos disponer de una empresa de mantenimiento que pueda solucionar incidencias de manera efectiva y en el menor tiempo posible para obtener la producción de potencia mas optima posible en cada parque eólico.

anyo	mes	dia	parque eolico	pais	POTENCIA	ALTURA OLAS	TEMPERATURA	VIENTO
2016	4	14	POLVARS	POLONIA	9.856	6	10	18,6
2016	3	15	POLVARS	POLONIA	9.599	7	11	18,2
2016	3	31	POLVARS	POLONIA	9.535	7	10	18,4
2016	3	19	POLVARS	POLONIA	9.528	7	10	18,1
2016	4	18	POLVARS	POLONIA	9.503	7	11	18,3
2016	5	4	POLVARS	POLONIA	9.501	6	10	18,7
2016	5	28	POLVARS	POLONIA	9.436	7	11	18,2
2016	4	26	POLVARS	POLONIA	9.375	7	11	18,3
2016	4	4	POLVARS	POLONIA	9.364	8	10	18,1
2016	3	13	POLVARS	POLONIA	9.292	7	11	18,2
2016	4	22	POLVARS	POLONIA	9.232	7	11	17,6
2016	3	1	POLVARS	POLONIA	9.225	6	10	18,1
2016	3	29	POLVARS	POLONIA	9.223	7	10	18
2016	3	22	POLVARS	POLONIA	9.208	8	10	18
2016	4	21	POLVARS	POLONIA	9.195	7	10	17,9
2016	3	11	POLVARS	POLONIA	9.165	7	10	18,5
2016	5	5	POLVARS	POLONIA	9.141	7	10	18,1
2016	5	2	POLVARS	POLONIA	9.115	7	11	17,7
2016	4	17	POLVARS	POLONIA	9.114	7	11	17,8

Ilustración 10 - Análisis relación variables de producción/meteorología

Todos los datos obtenidos sobre los análisis realizados pueden encontrarse en el siguiente documento:

Anexo 12: Resultado de análisis con saiku

8. Conclusiones

En este trabajo se ha aprendido a analizar, modelar, implementar y utilizar un sistema completo de inteligencia de negocio.

Inicialmente, se ha introducido el proyecto donde se han identificado los motivos i las justificaciones de la realización de dicho proyecto, los objetivos a alcanzar, el enfoque a seguir y la planificación del mismo.

En la segunda parte del proyecto, se ha identificado el hardware necesario a nivel de infraestructura para la correcta ejecución de este proyecto. A nivel de software, se ha realizado una selección, justificación e instalación de las herramientas necesarias para cumplir los objetivos de:

- Almacenamiento de datos históricos.
- Extracción y manipulación de los datos.
- Generación de cubos multidimensionales.
- Visualización y análisis de los datos.

En la tercera parte del proyecto, se ha realizado la modelización e implementación del data warehouse a partir del análisis de requerimientos i del origen de datos, además se han generado los diferentes esquemas de diseño conceptual, lógico i físico.

En la cuarta parte del proyecto, se han generado los procesos ETL necesarios para la carga de datos en el data warehouse a partir de una serie de transformaciones de procesado, modelización de datos y de jobs de ejecución de estas transformaciones. Estas transformaciones generadas contemplan en su configuración cargas de datos iniciales, incrementales y de actualización en su mismo flujo de ejecución.

En la quinta parte del proyecto, se ha generado un cubo multidimensional necesario para realizar el análisis de los datos. El cubo a sido diseñado para contemplar nuestro esquema en copo de nieve.

En la última parte del proyecto, este cubo multidimensional generado en la fase anterior se ha cargado en el frontal web de Pentaho Analytics. Una vez cargado, se ha realizado un primer análisis analítico a las preguntas analíticas que se deseaban resolver con este proyecto a partir de los datos disponibles en el data wharehouse.

Con este proyecto se ha logrado implementar un sistema de inteligencia de negocio para parques eólicos offshore además de poder profundizar en conceptos y usos de herramientas de BI de la suite Pentaho y bases de datos PostgreSQL.

Gracias a este sistema de inteligencia de negocio, se han podido resolver varias de las preguntas analíticas que se deseaban resolver. Para otras de estas preguntas analíticas, será necesario realizar una segunda fase del proyecto para obtener del origen de datos mas información para realizar un análisis más completo como pueden ser los tiempos de SLA/SLO por incidencia para identificar correctamente la efectividad de las empresas de mantenimiento contratadas.

A nivel de planificación de proyecto, se han podido alcanzar los objetivos en el tiempo estipulado, introduciendo nuevos apartados y realizando breves cambios en la memoria a lo largo de este.

Comentar que de cara a PostgreSQL y la aplicación Pentaho Data Integration no he tenido ningún problema en la implementación, configuración y uso.

De cara a Pentaho Schema Worbench, es una suite útil para la generación de cubos con esquemas en estrella. Para la generación de cubos con esquemas en copo de nieve puede resultar compleja su implementación o directamente no implementable dependiendo complejidad del nivel de relaciones que tenga nuestro data wharehouse si realizamos un diseño de esquema en copo de nieve.

De cara a Pentaho Bussiness Analytics en su versión community, comentar que personalmente, después de la realización del proyecto, los pluguin J-Pivot y Saiku me han parecido útiles para salir del paso pero, en el primer caso, J-Pivot me ha parecido un plugin complicado de utilizar y poco intuitivo. En el caso del plugin Saiku, es una buena herramienta para realizar agregados pero tiene un error de “caja” (error que me ha aparecido y he confirmado con el profesor y en los foros de meteorite) que no permite ejecutar filtros para ordenaciones o búsquedas. La solución encontrada a sido realizar los agregados a nivel de Saiku, exportar los agregados a Excel y realizar los filtros y búsquedas de datos concretos en el propio Excel. En una segunda fase del proyecto sería conveniente analizar si una versión Enterprise podría solucionar los problemas detectados con estos dos plugins.

9. Glosario

Aerogenerador: Generador de energía eléctrica que es accionado por la fuerza del viento.

Multiplicadora*:* Caja de engranajes del aerogenerador.

Generador: Es un dispositivo capaz de transformar energía mecánica en eléctrica.

Rotor: Es la parte giratoria de un aerogenerador compuesto por hélices.

Software: Conjunto de programas, instrucciones y reglas informáticas.

Hardware: Conjunto de componentes de una computadora.

Clúster: Conjuntos o conglomerados de ordenadores unidos entre sí normalmente por una red de alta velocidad y que se comportan como si fuesen una única computadora.

Balanceador de carga: Es un método para distribuir la carga de trabajo en varias computadoras separadas o agrupadas en un clúster.

Monitorización: Control de las constantes de un sistema a través de sensores.

Apache: Servidor web HTTP de código abierto que implementa el protocolo HTTP/1.1 y la noción de sitio virtual.

Tomcat: Es un contenedor web con soporte de servlets y JSPs.

PostgreSQL: Sistema de gestión de bases de datos relacional orientado a objetos y libre.

Pentaho BI Suite: Conjunto de programas libres para generar inteligencia empresarial (Business Intelligence). Incluye herramientas integradas para generar informes, minería de datos, ETL, etc.

Máquina de salto: Computadora informática de una infraestructura que se utiliza como único punto de conexión para la administración de los sistemas entre los usuarios y el resto de componentes de un sistema informático.

Staging Area: Es una zona intermedia entre los datos de origen y el almacen de datos utilizada para la realización de un primer procesamiento de los datos ETL.

ETL: Es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos, limpiarlos y cargarlos en otra base de datos.

Esquema en estrella: Modelo de datos que tiene una tabla de hechos (o tabla fact) que contiene los datos para el análisis, rodeada de las tablas de dimensiones.

Esquema copo de nieve*:* Modelo de datos algo más complejo que el esquema en estrella. Se da cuando alguna de las dimensiones se implementa con más de una tabla de datos.

DDL: (Data Definition languaje), lenguaje proporcionado por el sistema de gestión de base de datos que permite a los programadores realizar tareas de definición de estructuras que almacenarán datos.

Almacén de datos (Data warehouse): Es una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza.

Cubos multidimensionales (OLAP): Permiten procesar grandes volúmenes de información, en campos bien definidos, y con un acceso inmediato a los datos para su consulta y posterior análisis.

10. Bibliografía

- **Documentación PostgreSQL**

  URL: http://www.postgresql.org.es/documentacion

- **Documentación Pentaho**

  URL: https://help.pentaho.com/Documentation/7.0
  URL: http://mondrian.pentaho.com/documentation/workbench.php

- **Libro Introducción al Business Intelligence**

  Autores: Jordi Conesa Caralt / Josep Curto Díaz
  Editorial: Editorial UOC

- **Material teórico y práctico de las asignaturas de la UOC**

  B2.335 - Diseño y Construcción del Almacén de Datos
  B2.336 - Bases de Datos para Data Warehouse
  B2.337 - Explotación y Administración de Sistemas de Data Warehouse

- **Wikipedia**

  URL: https://es.wikipedia.org/wiki/Wikipedia:Portada

- **RAE**

  URL: http://www.rae.es/

- **Esquema en estrella y copo de nieve:**

  URL:http://josepcurto.com/2007/11/19/diseno-de-un-data-warehouse-estrella-y-copo-de-nieve/

- **Slowly Changing Dimension (SCD):**

  URL:
  http://www.guillesql.es/Articulos/Claves_Subrogadas_Slowly_Changing_Dimension_SCD_Tipo_2.aspx

- **Ejecución de Jobs mediante planificador de tareas:**

  URL: http://forums.pentaho.com/showthread.php?151873-How-to-schedule-the-Jobs-Transformations-in-Kettle

11. Anexos

Anexo 1: Documento de planificación del proyecto.

Planificación del proyecto.rar

Anexo 2: Manual de instalación - Pentaho Data Integration.

Manual de instalación - Pentaho Data Integration.rar

Anexo 3: Manual de instalación - Pentaho Schema Workbench.

Manual de instalación - Pentaho Schema Workbench.rar

Anexo 4: Manual de instalación - Pentaho Business Analytics.

Manual de instalación - Pentaho Business Analytics.rar

Anexo 5: Manual de instalación - PostgreSQL.

Manual de instalación - PostgreSQL.rar

Anexo 6: Script DDL

script_DDL_offshore.rar

Anexo 7: Fichero Excel DATAOFFSHORE.xlsx (Proporcionado por la UOC).

DATAOFFSHORE.rar

Anexo 8: Fichero Excel CARGA-DVARIABLE.xlsx (Personalizado).

CARGA-DVARIABLE.rar

Anexo 9: Manual de implantación - Creación de transformaciones

Manual de implantación - Creación de transformaciones.rar

Anexo 10: Manual de implantación - Creación de jobs

Manual de implantación - Creación de jobs.rar

Anexo 11: Manual de implantación - Cubo de análisis multidimensional

Manual de implantación - Cubo de analisis multidimensional.rar

Anexo 12: Resultado de análisis con saiku

Resultado de análisis con saiku.rar