

# Proyecto de Business Intelligence para caso de estudio de parques eólicos

**Jesús Martín Muriel**

Master Universitario en Ingeniería Informática  
Inteligencia de negocio (Business Intelligence)

**David Amorós Alcaraz**

**María Isabel Guitart Hormigo**

9 de Enero de 2017

© (Jesús Martín Muriel)

Reservados todos los derechos. Está prohibido la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la impresión, la reprografía, el microfilme, el tratamiento informático o cualquier otro sistema, así como la distribución de ejemplares mediante alquiler y préstamo, sin la autorización escrita del autor o de los límites que autorice la Ley de Propiedad Intelectual.

## FICHA DEL TRABAJO FINAL

<b>Título del trabajo:</b>	<i>Proyecto de Business Intelligence para caso de parque eólicos</i>
<b>Nombre del autor:</b>	<i>Jesús Martín Muriel</i>
<b>Nombre del consultor/a:</b>	<i>David Amorós Alcaraz</i>
<b>Nombre del PRA:</b>	<i>María Isabel Guitart Hormigo</i>
<b>Fecha de entrega (mm/aaaa):</b>	01/2017
<b>Titulación::</b>	<i>Master Oficial en Ingeniería Informática</i>
<b>Área del Trabajo Final:</b>	<i>Business Intelligence</i>
<b>Idioma del trabajo:</b>	<i>Castellano</i>
<b>Palabras clave</b>	<i>Data warehouse, ETL, OLAP, Pentaho.</i>
<p><b>Resumen del Trabajo (máximo 250 palabras):</b> <i>Con la finalidad, contexto de aplicación, metodología, resultados y conclusiones del trabajo.</i></p>	
<p>La finalidad del presente proyecto ha sido identificar una plataforma de inteligencia de negocio de uso libre y utilizarla en un caso práctico, en concreto, en el negocio de parques eólicos.</p> <p>En un contexto donde una gran parte de la actividad humana está asistida por sistemas informáticos, se cuenta con grandes volúmenes de datos, pero resulta complejo poder tomar información relevante de los mismos, que nos permita explicar que ha sucedido o que nos ayude a tomar decisiones.</p> <p>El camino para poder afrontar esto, es una mezcla de metodologías de trabajo y herramientas. Dichas herramientas, responden a diferentes aplicaciones informáticas, unas pueden ser de uso general y son aprovechables en esta actividad, otras son específicas o exclusivas de la propia inteligencia de negocio.</p> <p>A su vez, diferentes fabricantes de software, ofrecen suites comerciales, que suelen encapsular en exceso los diferentes aspectos de un proyecto de inteligencia de negocio, obligando al uso exclusivo de las mismas en todas las fases y etapas de un proyecto de estas características.</p> <p>El mayor problema que se plantea en general en el uso del software libre y en particular en contextos complejos como este, es que el mismo, constituye un conjunto de aplicaciones diferentes, poco integradas, deslavazadas, y que en cierto modo carecen del hilo conductor como base metodológica para alcanzar los objetivos de negocio que amparan estos proyectos.</p> <p>El resultado de este proyecto, es un framework (documentación, ejemplos y aplicaciones), que puede servir como patrón para proyectos de esta tipología, además de satisfacer las necesidades de información del caso práctico tratado.</p>	

**Abstract (in English, 250 words or less):**

The target for this project was find a open source solution for intelligence business environment, study that solution and use it in a real case. This case has been a wind energy business.

Today, large portion of human activity is supported by computer systems. There are lot of datum but is very complex recover information from these datum for to know what happened or for help to decision makers.

The way to deal with this is a mix of work methodologies and tools. These tools respond to different computer applications, some can be of general use but profitable in this activity, others are specific or exclusive to the business intelligence itself.

Different software manufacturers offer commercial suites, which often encapsulate the different aspects of a business intelligence project, forcing the exclusive use of the same in all phases and stages of a project of these characteristics.

The biggest problem in general in the use of free software, and particularly in complex contexts such as this one, is that it is a set of different applications that are not well integrated. It really is not a software suite, so a methodological approach is even more important.

The result of this project is a framework (documentation, examples and applications), which can serve as a template for projects of this typology, in addition to meeting the information needs of the case in study (wind energy case).

# Índice

1.	Introducción .....	1
1.1.	Contexto y justificación del Trabajo .....	1
1.2.	Objetivos del Trabajo.....	2
1.3.	Enfoque y método seguido .....	2
1.4.	Planificación del Trabajo.....	5
1.5.	Breve resumen de productos obtenidos .....	7
1.6.	Breve descripción de los otros capítulos de la memoria.....	10
2.	Resultado del proyecto .....	12
2.1.	Identificación y estudio del software necesario.....	12
2.1.1.	Selección de software de Business Intelligence.....	12
2.1.2.	Visión global de Pentaho.....	16
2.1.3.	Pautas generales de adquisición, instalación y uso .....	17
2.1.4.	Software necesario adicional.....	20
2.2.	Construcción del data warehouse.....	21
2.2.1.	Diseño conceptual del data warehouse (planteamiento inicial) .....	21
2.2.2.	Diseño conceptual del data warehouse (Enfoque 2) .....	22
2.2.3.	Diseño lógico del data warehouse.....	23
2.2.4.	Alternativa finalmente adoptada de diseño lógico del data warehouse 24	
2.2.5.	Diseño físico inicial del data warehouse.....	25
2.2.6.	Diseño físico finalmente adoptado para el data warehouse .....	26
2.3.	Construcción de artefactos ETL.....	27
2.3.1.	Diseño ETL para la carga de datos maestros .....	27
2.3.2.	Implementación ETL para la carga de datos maestros .....	31
2.3.3.	Diseño ETL para la carga de lecturas de todos los parques .....	32
2.3.4.	Implementación ETL de carga de lecturas de todos los parques ....	38
2.4.	Construcción de artefactos OLAP.....	39
2.4.1.	Diseño de artefactos OLAP .....	39
2.4.2.	Schema Workbench .....	40
2.4.3.	(PUC) Pentaho User Console .....	42
2.5.	Obtención de informes.....	44
2.6.	Resolución de los problemas que originaron el proyecto .....	47
2.6.1.	Parques eólicos más productivos.....	47
2.6.2.	Incidencia de los hechos meteorológicos sobre la producción.....	49
2.6.3.	Zonas con mejor relación viento potencia .....	52
2.6.4.	Incidencia de los hechos meteorológicos sobre las alarmas técnicas 55	
2.6.5.	Análisis de las empresas de mantenimiento .....	72
2.6.6.	Relación entre las variables productivas y meteorológicas .....	73
3.	Conclusiones .....	78
4.	Glosario .....	80
5.	Bibliografía .....	82
6.	Anexos.....	83

## Lista de figuras

1 Imagen del diagrama de Gantt de la planificación del proyecto	6
2 Cuadrante Mágico de Gartner de Febrero de 2016 sobre BI	14
3 Diagrama de diseño conceptual inicial (No es el enfoque ejecutado)	21
4 Diagrama de diseño conceptual ejecutado	22
5 Diagrama del modelo físico inicialmente planteado elaborado con herramientas de MySQL	25
6 Diagrama del modelo físico ejecutado elaborado con herramientas de MySQL	26
7 Imagen del diseño del proceso de carga de datos maestro	27
8 Imagen del detalle de configuración de la tarea de lectura del Excel de entrada de datos	28
9 Imagen del detalle de configuración de los campos de entrada	28
10 Imagen del detalle de configuración de tarea de búsqueda y actualización	29
11 Imagen del detalle de configuración de tarea de búsqueda y actualización	30
12 Imagen del diseño del proceso de carga de las lecturas	32
13 Imagen con detalle de configuración de mapeo entre Excel de entrada y modelo	32
14 Imagen detallada de limpieza de variables, quedando solo las válidas	33
15 Imagen de detalle de tarea de formateo de fechas, estableciendo nuevas variables del flujo	34
16 Detalle de configuración de la carga de dimensión fecha mediante búsqueda y actualización	35
17 Imagen con detalle de configuración de la tarea	36
18 Imagen de tarea de carga de datos en tabla de hechos	37
19 Imagen con configuración de conexión al data warehouse del proyecto	40
20 Imagen de un momento del diseño del cubo Análisis_Productividad_Meteorología	40
21 Opción de publicación y detalle de cuenta en (PUC) Pentaho User Console	41
22 Mensaje de publicación realizada	41
23 Imagen de login en PUC	42
24 Menú principal PUC	42
25 Visualización de cubo	43
26 Visualización de cubo Productividad	43
27 <i>Visualización de cubo</i>	44
28 <i>Visualización de cubo</i>	44
29 Visualización de cubo	45
30 Visualización de cubo	45
31 Visualización de cubo	46
27 Imagen de informe sobre productividad elaborado con Pentaho Report	48
28 Imagen de confección de cubo mediante Schema Workbench	49
29 Imagen de uso de cubo y gráfico con JPivot	50
30 Imagen de análisis gráfico mediante Pentaho Report sobre la relación entre Meteorología y Productividad	51
31 Imagen de creación de origen de datos basado en una vista	53
32 Imagen de modelado de cubo OLAP desde PUC	53

33 Visualización de cubo desde JPivot	54
34 Visualización de análisis gráfico desde Pentaho Report	56
35 Visualización de análisis gráfico desde Pentaho Report	57
36 Visualización de análisis gráfico desde Pentaho Report	58
37 Visualización de análisis gráfico desde Pentaho Report	59
38 Visualización de análisis gráfico desde Pentaho Report	60
39 Visualización de análisis gráfico desde Pentaho Report	61
40 Visualización de análisis gráfico desde Pentaho Report	62
41 Visualización de análisis gráfico desde Pentaho Report	63
42 Visualización de análisis gráfico desde Pentaho Report	64
43 Visualización de análisis gráfico desde Pentaho Report	65
44 Visualización de análisis gráfico desde Pentaho Report	66
45 Visualización de análisis gráfico desde Pentaho Report	67
46 Visualización de análisis gráfico desde Pentaho Report	68
47 Visualización de análisis gráfico desde Pentaho Report	69
48 Visualización de análisis gráfico desde Pentaho Report	70
49 Visualización de análisis gráfico desde Pentaho Report	71
50 Visualización de informe desde Pentaho Report	72
51 Visualización de análisis gráfico desde Pentaho Report	74
52 Visualización de análisis gráfico desde Pentaho Report	74
53 Visualización de análisis gráfico desde Pentaho Report	75
54 Visualización de análisis gráfico desde Pentaho Report	75
55 Visualización de análisis gráfico desde Pentaho Report	76
56 Visualización de análisis gráfico desde Pentaho Report	76
57 Visualización de análisis gráfico desde Pentaho Report	77
58 Visualización de análisis gráfico desde Pentaho Report	77

# 1. Introducción

## 1.1. Contexto y justificación del Trabajo

Los parques eólicos se caracterizan por ubicarse en localizaciones de difícil acceso, especialmente los denominados offshore que se encuentran en el mar.

Un parque eólico está formado por un conjunto de molinos o aerogeneradores, que de forma permanente están generando eventos, que informan sobre el estado del conjunto o de las partes y que con una frecuencia establecida (en términos de minutos) pueden comunicar esos datos a un centro de control donde se reciben.

A los eventos que informan del estado de los aerogeneradores, se suman los de las estaciones meteorológicas instaladas en los parques, en tanto la relación entre producción y averías y condiciones meteorológicas, pueden guardar un relación directa.

Ante la inversión económica que representa cada uno de estos molinos, es fundamental detectar cuando sufren averías y dicha producción se ha parado, pero también poder entender las circunstancias que pudieron proporcionar la parada, encontrando relaciones entre los diferentes eventos proporcionados por la variedad de sensores instalados.

El problema final se centra en ser capaz de manejar un conjunto extenso de datos unitarios, unos valores de estado, otros indicadores cuantitativos, que responden a diferentes protocolos según los fabricantes de los equipos que los envían, y en concreto la solución del problema está en la capacidad de que los expertos en el negocio, sean capaces de encontrar patrones de comportamiento, así como de poder dar respuesta a las diferentes cuestiones que en ese camino se plantean.



## 1.2. Objetivos del Trabajo

**Objetivo 1:** Obtener un repositorio central de información que aglutine en un modelo dimensional la información procedente de los diferentes generadores y parques, disponiendo de las prestaciones y capacidades básicas de un motor de base de datos.

**Objetivo 2:** Contar con los mecanismos automáticos o semiautomáticos necesarios, para que la carga en la base de datos antes referida, de la información diversa (diferentes estructuras según fabricantes involucrados en la ingeniería de cada parque) se realice de forma adecuada, presentando en todo momento una información histórica que llegue hasta el momento actual.

**Objetivo 3:** Contar con los mecanismos que reestructuren la información almacenada en el repositorio, de forma que esta se pueda orientar a las necesidades de análisis de información, como base para un tratamiento inteligente de las mismas.

**Objetivo 4:** Disponer las opciones de consulta o informes, que reflejen los datos fundamentales que la dirección del negocio demanda, haciendo uso de medios que permitan a usuarios avanzados y especialistas en el negocio de la aerogeneración, la modificación o ampliación de los informes iniciales, sin necesidad de trabajos complejos de programación informática.

## 1.3. Enfoque y método seguido

Para satisfacer los objetivos del proyecto, se buscará en el mercado de software libre tanto un sistema gestor de base de datos relacional, como una plataforma de análisis de información que albergue herramientas ETL, que gestione metadatos para la implementación de modelos de datos multidimensionales y que cuente con un motor de informes, así como herramientas de diseño de informes y cuadros de mando.

A su vez el siguiente cuadro recoge en mayor detalle las tareas asociadas a cada objetivo.

Objetivo	Tareas
<b>Objetivo 1</b>	<p>Análisis de ficheros enviados por los parques/aerogeneradores o documentos descriptivos de los protocolos de los mismos.</p> <p>Obtener ficheros de ejemplo, bien reales o contruidos a partir de la descripción de los protocolos.</p> <p>Diseño de modelo conceptual que albergue la realidad recogida en los ficheros, orientado a un uso de data warehouse, tomando en consideración los procesos y métricas manifestados en la demanda del proyecto.</p> <p>Diseño de modelo lógico</p> <p>Elección de sistema gestor de base de datos relacional (SGBDR), dentro de la oferta existente en el ámbito de tecnologías open source.</p> <p>Diseño de modelo físico e implementación en el gestor de base de datos elegido del modelo establecido.</p>
<b>Objetivo 2</b>	<p>Análisis de productos ETL y plataformas de Business Inteligente. Si bien para atender este objetivo, existen soluciones individuales, desde la programación directa de aplicativos que lean los ficheros de entrada, practiquen las modificaciones y adendas necesarios y los inserten en base de datos, hasta productos individuales existentes en el mercado de la integración de datos, es habitual que los paquetes o plataformas de Business Intelligence incluyan una solución al respecto, y dado que hay que elegir una de estas, dispondremos en este punto los trabajos de búsqueda y selección de la misma.</p> <p>Diseño de mapeo y transformación de datos</p> <p>Implementación sobre el producto elegido del diseño obtenido</p>
<b>Objetivo 3</b>	<p>Diseño de modelos de datos multidimensionales que partiendo de</p>

los datos del data warehouse se orienten a satisfacer las necesidades de información y análisis expuestas.

Implementación de los modelos diseñados en la herramienta incluida en la plataforma elegida (OLAP, ROLAP, ...)

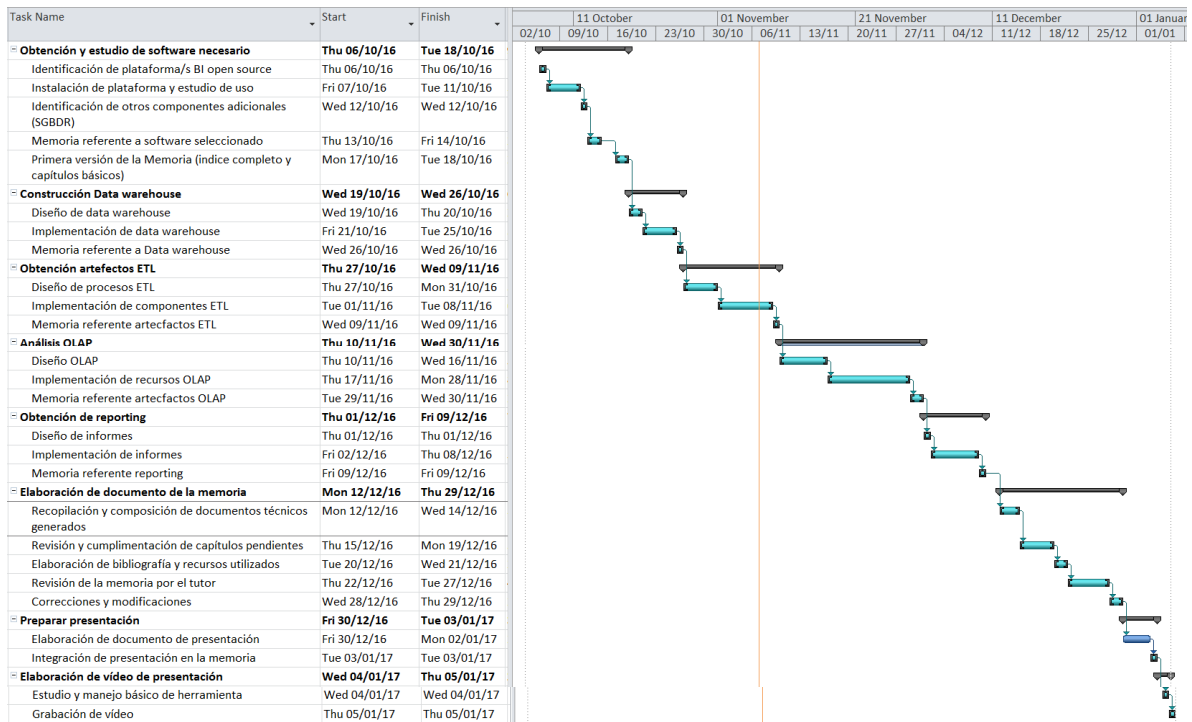
**Objetivo 4** Diseño de las consultas e informes tipo que el sistema a entregar incluirá, para satisfacer los requisitos solicitados.

## 1.4. Planificación del Trabajo

Tomando como referencia las fases fundamentales de un proyecto de Business Intelligence, los cuatro hitos fundamentales (obtener el Data warehouse, los elementos ETL, los elementos OLAP y el Reporting), así como otras actividades propias de la ejecución de un proyecto, los hitos fundamentales de esta planificación y su correlación con las entregas parciales del mismo (PEC) se distribuyen de la siguiente forma:

Entrega	Fecha de entrega	Descripción
<b>PEC1</b>	3-10-2016	Planificación
<b>PEC2</b>	7-11-2016	Construcción del data warehouse Construcción de artefactos ETL Memoria actualizada a este punto
<b>PEC3</b>	12-12-2016	Construcción OLAP Construcción de informes Memoria actualizada a este punto
<b>Entrega final</b>	9-01-2017	Confección de presentación Realización de vídeo de presentación Subsanación de errores en la memoria Introducción/completar de bibliografía

El diagrama de Gantt posterior, recoge el detalle necesario obtenido para atender al anterior resumen.



1 Imagen del diagrama de Gantt de la planificación del proyecto

## 1.5. Breve resumen de productos obtenidos

La siguiente tabla identifica todos los productos obtenidos con la ejecución del proyecto, indicando en cada caso la motivación o finalidad específica del mismo y su naturaleza.

Producto	Tipo de producto	Finalidad
<b>Selección/identificación de plataforma BI adecuada</b>	Documento	Infraestructura base requerida
<b>Diseño del data warehouse</b>	Documento	Infraestructura base requerida
<b>Diseño del data warehouse optimizado para los fines específicos del proyecto</b>	Documento	Infraestructura base requerida
<b>Implementación en MySQL del data warehouse</b>	Metadatos de SGBDR	Infraestructura base requerida
<b>Diseño de procesos ETL para el data warehouse</b>	Documento	Infraestructura base requerida
<b>Implementación de procesos ETL para el data warehouse:</b> <b>Carga_Datos_Maestros_v10.ktr</b> <b>Carga_Lecturas.ktr</b>	Metadatos en formato XML/PDI	Infraestructura base requerida
<b>Diseño de procesos ETL para el data warehouse optimizado</b>	Documento	Infraestructura base requerida
<b>Implementación de procesos ETL para el data warehouse optimizado:</b>	Metadatos en formato	Infraestructura base requerida

<b>Carga_Maestro_Enfoque_2.ktr</b>	XML/PDI	
<b>Carga_Lecturas_Enfoque2.ktr</b>		
<b>Base de datos resultante de ejecución procesos ETL</b>	Base de datos	Infraestructura base requerida
<b>Analisis_Productividad.mondrian.xml</b>		
<b>Analisis_Productividad.xml</b>		
<b>Productividad_x_Molino.zip</b>	Metadatos OLAP	
<b>Vision_global.zip</b>		
<b>Vista_Produccion_Ampliada.zip</b>		Parques eólicos más productivos
<b>2_6_1_Productividad_Molinos.prpt</b>	Metadatos reporting	
<b>Ejecucion_2_6_1.pdf</b>	Informe	
<b>Analisis_Productividad_Meteorologia.xml</b>	Metadatos OLAP	
<b>2_6_2_Meteo_Produccion.prpt</b>	Metadatos reporting	Incidencia de los hechos meteorológicos sobre la producción
<b>Ejecucion_2_6_1.pdf</b>	Informe	
<b>Análisis_Zona_Viento_Potencia.xjpivot.zip</b>	Metadatos en formato JPivot	Zonas con mejor relación viento potencia
<b>2_6_4_1_Meteorologia_alarmas_NAMPER.prpt</b>		
<b>2_6_4_2_Meteorologia_alarmas_RIAS_BAIXAS.prpt</b>		
<b>2_6_4_3_Meteorologia_alarmas_GUARACHICO.prpt</b>		
<b>2_6_4_4_Meteorologia_alarmas_KIRSKEN.prpt</b>	Metadatos reporting	Incidencia de los hechos meteorológicos sobre las alarmas técnicas
<b>2_6_4_5_Meteorologia_alarmas_POLVARS.prpt</b>		
<b>2_6_4_6_Meteorologia_alarmas_COUNSCOT.prpt</b>		
<b>2_6_4_7_Meteorologia_alarmas_GREENBLU</b>		

E.prpt 2_6_4_6_Meteorologia_alarmas_NORTHENC AP.prpt		
Resultados_2_6_4_1.pdf Resultados_2_6_4_2.pdf Resultados_2_6_4_3.pdf Resultados_2_6_4_4.pdf Resultados_2_6_4_5.pdf Resultados_2_6_4_6.pdf Resultados_2_6_4_7.pdf Resultados_2_6_4_8.pdf	Informe	
Análisis_Mantenedoras.prpt	Metadatos reporting	Análisis de las empresas de mantenimiento
Resultados_2_6_5.pdf	Informe	
2_6_6_1_Produccion_Meteorologia.prpt 2_6_6_2_Produccion_Meteorologia.prpt 2_6_6_3_Produccion_Meteorologia.prpt 2_6_6_4_Produccion_Meteorologia.prpt 2_6_6_5_Produccion_Meteorologia.prpt 2_6_6_6_Produccion_Meteorologia.prpt 2_6_6_7_Produccion_Meteorologia.prpt 2_6_6_8_Produccion_Meteorologia.prpt	Metadatos reporting	Relación entre las variables productivas y meteorológicas
Resultados_2_6_6_1.pdf Resultados_2_6_6_2.pdf Resultados_2_6_6_3.pdf Resultados_2_6_6_4.pdf Resultados_2_6_6_5.pdf Resultados_2_6_6_6.pdf Resultados_2_6_6_7.pdf Resultados_2_6_6_8.pdf	Informe	



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

El resultado del proyecto y las conclusiones generales que ha permitido alcanzar, dan contenido a los siguientes capítulos de la presente memoria.

La estructura del capítulo de Resultado del Proyecto, sigue la generación de cada uno de los productos que integran el resultado, y que en unos casos responden a análisis o informes que se integran en la memoria, y en otros a la descripción del artefacto resultante, por su naturaleza, complementario a la misma.

La secuencia de acciones, y con ello de subcapítulos, representa la metodología de trabajo aplicada para la ejecución de un proyecto de inteligencia de negocio, que partiendo de unos requisitos iniciales, los analiza, diseña el data warehouse más adecuado (ejemplo de resultado del proyecto que por su naturaleza queda integrado en la propia memoria), llegando a la propia implementación de la base de datos (ejemplo de artefacto complementario a esta memoria, pero a cuya descripción y referencia, responde uno de los subcapítulos posteriormente recogidos).

Uno de los productos del proyecto, es la elección de un software open source que de soporte al mismo.

Tras una breve fundamentación de dicha decisión, se proporciona una visión global de la plataforma, entrando en detalles en cuanto al uso de las funcionalidades de la misma, que serán utilizadas en el proyecto.

Volviendo a la metodología de proyecto de inteligencia de negocio seguida, ya con la asistencia de la plataforma, se diseñan e implementan los artefactos propios de este tipo de proyecto (cubos OLAP e informes concluyentes).

Con toda la infraestructura necesaria obtenida y reflejada en la memoria (análisis, diseño, elección de la tecnología, explicación e instalación general de la misma), se procede a su uso para terminar satisfaciendo las cuestiones generales planteadas sobre el caso práctico de negocio, entorno al cual se planteó el proyecto. Se incluyen en la memoria, las conclusiones obtenidas

vinculadas con la identificación y descripción de los artefactos implementados para alcanzar dichas conclusiones.

Finalmente el capítulo de conclusiones, recoge las propias del proyecto de Business Intelligence ejecutado, pues las conclusiones técnicas a la problemática del caso de negocio analizado, forman parte del Resultado del Proyecto.

## 2. Resultado del proyecto

El resultado del proyecto es la combinación de una plataforma de Business Intelligence perteneciente al existentes del tipo open source junto con los documentos y metadatos diseñados para la plataforma elegida para dar una solución de inteligencia de negocio a la problemática de los parque eólicos offshore.

### 2.1. Identificación y estudio del software necesario

#### 2.1.1. Selección de software de Business Intelligence

Los requisitos funcionales establecidos, demandan la necesidad de una infraestructura tipo para el análisis de información, que asista a una toma de decisiones inteligente en base a un adecuado diagnóstico del negocio y de su organización.

Se manifiesta la necesidad de extraer información de diversas fuentes (ficheros de texto enviados desde los parques por los dos subsistemas identificados en los mismos), almacenarla atendiendo tanto a un enfoque que permita la gestión y análisis histórico como a unos análisis de la información que acaben proporcionando el acceso a los diferentes indicadores de los procesos de negocio en estudio, contando con unos recursos mínimos para la generación de informes y de cuadros de mando.

Todo lo anterior nos lleva a demandar un conjunto de aplicaciones informáticas, de uso integrado entre las mismas, que en la industria del software se engloba en el segmento de mercado denominado Business Intelligence.

Se establece también como requisito, la necesidad de que el software de infraestructura que se utilice para implementar la solución demandada, debe responder al colectivo de software libre (Open Source).

Son muchas las aplicaciones disponibles que funcionalmente se posicionan como plataforma de Business Intelligence dentro del contexto Open Source. Un

estudio exhaustivo de dicho conjunto de aplicaciones, excedería el alcance de este proyecto, por lo que, debemos centrar esta parte del presente trabajo académico, en una consideración de carácter general sobre una selección que más o menos arbitrariamente, recoja los productos más populares en este sentido, y que son los siguientes.

Producto	Consideraciones
<b>BIRT</b>	Se enfoca como un conjunto de recursos a ser integrados en el desarrollo de aplicaciones, por lo que no constituye la mejor opción para este proyecto que debe focalizarse en la resolución de un problema concreto de Business Intelligence, ajeno a la construcción de una aplicación al efecto.
<b>Jaspersoft</b>	Es una solución completa bajo el paraguas de TIBCO, lo que plantea dudas sobre su carácter de software libre, si bien sería necesario un testeado completo de la aplicación para valorar este extremo.
<b>Pentaho</b>	Es posiblemente la solución más popular y resulta fácil encontrar documentación de ayuda en la comunidad Internet, inclusive en castellano. Funcionalmente se presenta como una solución completa.
<b>RapidMiner</b>	Software libre focalizado en minería de datos.  Una revisión de la información de arquitectura que publica su sitio web, pone de manifiesto que cuenta con todas las piezas de una plataforma de inteligencia de negocio (ETL, OLAP, Reporting).
<b>SpagoBI</b>	Funcionalmente se presenta como una solución completa de Business Intelligence, si bien su requerimiento de un servidor de aplicaciones/servidor web (Tomcat) puede plantear dificultades en el

despliegue y uso y resultar más recomendable una aplicación de cliente pesado.

Con todas las anteriores consideraciones, se ha realizado una consulta a uno de los analistas de mayor reconocimiento y repercusión en la industria del software como es Gartner, en concreto se ha consultado su conocido producto divulgativo Cuadrante Mágico de Gartner para el sector de Business Intelligence del año 2016.



Como es de esperar en dicho informe, aparecen los productos de los grandes fabricantes generalistas y los más pequeños especializados en este segmento, todos ellos con soluciones comerciales fuera del alcance de este proyecto.

Pero como viene siendo habitual en los últimos años en estas publicaciones, aparece alguna solución de software libre, o que cuenta con versiones de software libre (versión community con una disposición funcional completa o casi completa) y estos sí forman parte del objeto de este trabajo,

Se trata en cualquier caso de validar el reconocimiento de la industria hacia alguna de las opciones de software libre consideradas, manifestándose que Pentaho es la única valorada en el informe, en una posición media-alta dentro del mercado, aspecto que viene a cumplimentar la percepción de ser la solución más popular.

A la vista de todas las consideraciones, se opta por la suite Pentaho para ejecutar el proyecto.

## 2.1.2. Visión global de Pentaho

Actualmente Pentaho es un producto del grupo empresarial japonés Hitachi, que en 2015 compró a la compañía estadounidense con sede en California, Pentaho Corporation.

Pentaho nació como un software de uso libre (de acuerdo a los principios Open Source) y basó su estrategia de negocio en los servicios profesionales entorno a dicho software y en una versión del mismo que incorporaba ciertas ventajas, ofrecida al mercado en modo de suscripción, pero siempre basada en un núcleo ajustado a Open Source.

Actualmente dan cobertura a tendencias como Big Data o Internet of Things (IoT).

Su presencia en Internet gira entorno a dos sites:

- [www.pentaho.com](http://www.pentaho.com)
- [community.pentaho.com](http://community.pentaho.com)

El primero es el site corporativo donde se presentan los productos de suscripción (de pago), que recogen las características más innovadoras.

En el segundo se da cabida a la comunidad de usuarios y es donde se encuentra el software de uso libre.

En uno y otro podemos comprobar como Pentaho responde a una suite de aplicaciones, donde cada una da cobertura a las diferentes necesidades que se manifiestan en la implementación de un proyecto de business intelligence como el presente, de forma que podemos distinguir la siguiente relación de necesidades y aplicaciones o módulos Pentaho que las satisfacen:

Necesidad	Aplicación
<b>Funciones ETL</b> <b>Capturar datos de diferentes fuentes y formatos, procesarlos y adaptarlos a los formatos adecuados para su análisis.</b>	PDI (Pentaho Data Integration)

<b>OLAP</b> <b>Aproximación a los hechos mediante la visualización de cubos de forma ajustable.</b>	Mondrian y Pentaho User Console
<b>Reporting</b> <b>Capacidad de construir informes y representaciones gráficas para dar respuesta a las necesidades de análisis, permitiendo presentaciones formales para clientes finales.</b>	Pentaho Report Designer

### 2.1.3. Pautas generales de adquisición, instalación y uso

Las versiones de uso libre se encuentran en <http://community.pentaho.com/>

Analizaremos estas pautas por aplicación o módulo.

<b>Módulo Pentaho</b>	<b>Descripción</b>
<b>PDI (Pentaho Data Integration)</b> <b>también llamado Kettle</b>	<p>Se obtiene de dicho site el fichero comprimido pdi-ce-6.1.0.1-196.zip, correspondiente a la versión 6.1. con la que se ha llevado a cabo este proyecto.</p> <p>Su descompresión despliega directamente la aplicación Java que requiere la correspondiente máquina virtual.</p> <p>La aplicación queda dispuesta en el directorio data-integration donde encontramos los siguientes ficheros .bat que lanzaran las siguientes aplicaciones disponibles:</p> <p>Spoon.bat lanza Spoon: Utilidad gráfica en (en la que centraremos la mayor parte de los trabajos ETL) para diseñar e implementar los procesos ETL, basado en el modelo drag and drop, genera un fichero de metadatos</p>



en formato XML (si bien los ficheros llevan la extensión .ktr de Kettle) que alberga los metadatos para describir las tareas a ser ejecutadas por el motor.

Spoon también permite el test de trabajo realizado, al instanciar al motor entregándole el XML para monitorizar el proceso y facilitar la información de log correspondiente al diseñador.

Además de Spoon, PDI también dispone en el mismo directorio las siguientes aplicaciones:

- Pan: Motor para la ejecución de transformaciones.
- Kitchen: Motor para la ejecución de trabajos.
- Carte: Permite actuar como servidor remoto que ejecuta transformaciones y trabajos.

### **Mondrian**

Mondrian es el motor de OLAP de Pentaho.

El trabajo de diseño e implementación OLAP del presente proyecto es llevado a cabo con la aplicación Schema Workbench, disponible en la Community de Pentaho en el paquete psw-ce-3.12.0.1-196.zip, cuyo despliegue dispone workbench.bat que lanza dicha aplicación.

### **Pentaho Console**

**User** Aplicación web para la publicación y visualización de cubos OLAP y otros recursos.

En las versiones utilizadas en este proyecto, esta aplicación web, es el punto de visualización de los cubos diseñados con Schema Workbench, así como de los propios informes, si bien Pentaho Report Designer, permite la visualización independiente de los mismos en diferentes formatos estándar, sin necesidad de publicar en este portal.

Pentaho User Console, representa la tendencia de Pentaho, en cuanto a integrar en una sola aplicación web, toda la funcionalidad hoy dispersa en diferentes aplicaciones de negocio. En la versión actual, desde este portal se puede definir nuevos cubos o en este servidor, se pueden, en forma de plug-in, ir incorporando las diferentes aplicaciones y utilidades que la comunidad Pentho va liberando.

Constituye una buena base para un uso corporativo, puesto que su arquitectura técnica, facilita su uso integrado en Intranet corporativas que den acceso a determinados análisis de la información.

La aplicación, cuenta con una gestión de usuarios y el correspondiente control de acceso.

Disponible en la Community de Pentaho en el paquete pentaho-server-ce-7.0.0.0-25.zip.

Mediante start-pentaho.bat, se arranca el servidor web Tomcat que alberga la propia aplicación. El propio paquete proporciona Tomcat con la aplicación (Pentaho) desplegada en el mismo.

Mediante el uso de un navegador, se accede a la misma con la URL <http://localhost:8080/pentaho>.

**Pentaho Report Designer** Aplicación Java de escritorio que mediante drag and drop permite el diseño de informes de diferentes tipologías.

Cuenta con una paleta de diferentes tipos de gráficos (barras, funcionales, de tarta, etc) con los que facilitar el análisis de datos.

Mediante la configuración de las fuentes de datos, se puede bien conectar directamente con el data warehouse o hacer uso de los artefactos OLAP

diseñados con Schema Workbench o con el propio Pentaho User Console.

Disponible en la Community de Pentaho en el paquete `prd-ce-6.1.0.1-196.zip`.

La aplicación se usa mediante la ejecución del lanzador `report-designer.bat`

#### 2.1.4. Software necesario adicional

Para albergar el data warehouse, resulta necesario un sistema gestor de base de datos relacional.

El software open source elegido para esa función es MySQL.

MySQL es posiblemente la aplicación open source de mayor penetración en el mundo. Creado en 1995 por la empresa del mismo nombre (MySQL AB), esta fue comprada por Sun Microsystems (año 2005) y esta a su vez y posteriormente (año 2010) por Oracle, que mantiene una doble línea de licenciamiento, la de open source y la comercial, dando así a un escenario similar al propio de Pentaho y el Grupo Hitachi.

MySQL cuenta con su site independiente (<http://www.mysql.com/>) donde bajo el epígrafe MySQL Community Edition (GPL) se accede a las versiones libres.

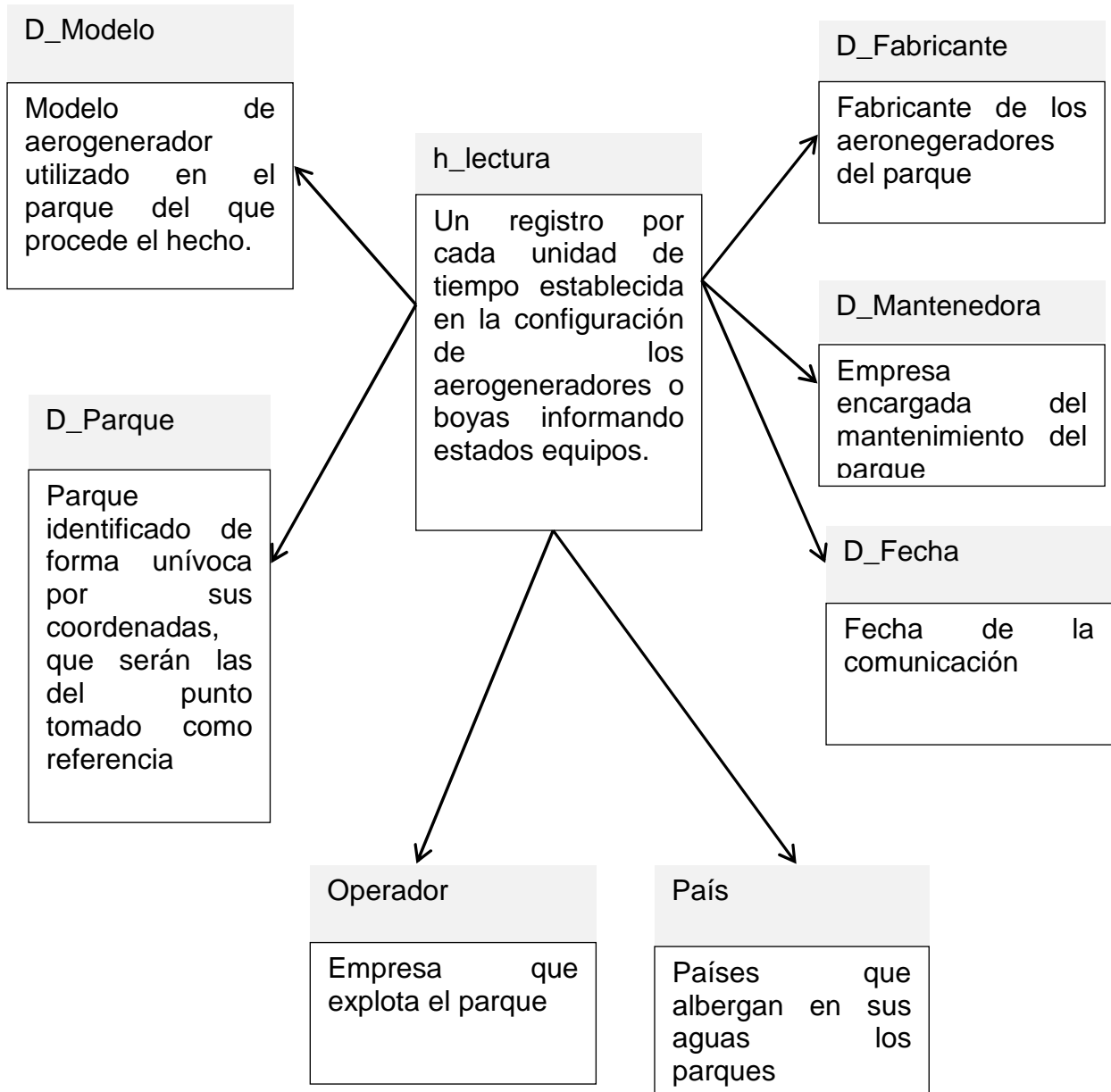
Para el ejecución del presente proyecto se ha hecho uso de la versión de MySQL 6.3.

Como complemento de esta parte de la infraestructura del proyecto, se ha hecho uso del driver de JDBC para MySQL, en concreto el componente `mysql-connector-java-5.1.40-bin.jar`, mediante el que las diferentes aplicaciones de Pentaho resuelven el acceso a la base de datos.

## 2.2. Construcción del data warehouse

### 2.2.1. Diseño conceptual del data warehouse (planteamiento inicial)

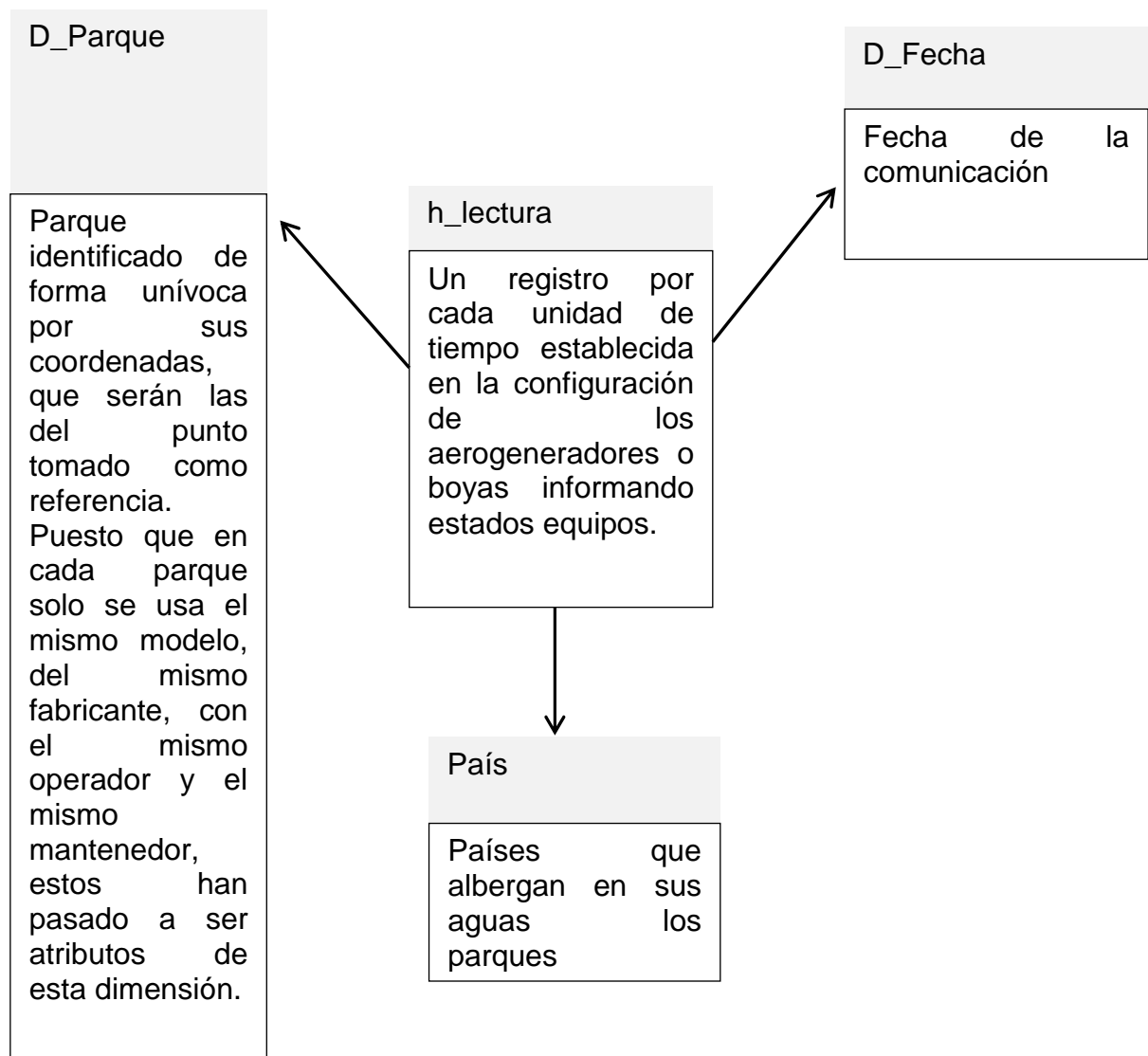
Para facilitar los indicadores planteados se plantea un modelo dimensional de copo de nieve.



3 Diagrama de diseño conceptual inicial (No es el enfoque ejecutado)

### 2.2.2. Diseño conceptual del data warehouse (Enfoque 2)

Partiendo del planteamiento inicial, como propuesta canónica, y analizados los datos, se propone un diseño alternativo que puede mejorar la productividad del trabajo en el proyecto, pasando al siguiente modelo, que ha dado lugar a que este Enfoque 2, se sume a los diseños e implementaciones iniciales tanto del data warehouse como de los artefactos ETL, optándose finalmente por este Enfoque 2:



4 Diagrama de diseño conceptual ejecutado

### 2.2.3. Diseño lógico del data warehouse

Las siguientes tablas recogen el detalle de las entidades definidas y sus relaciones en la opción de diseño inicialmente considerada.

Tabla de Hecho	Claves foráneas	Métricas
<b>h_lectura</b>	Id_Parque, Id_fecha, Id_Fabricante, Id_Modelo, Id_Mantenedora, Id_Operador, Id_País	Potencia_generada, velocidad_viento, temperatura, tasa_disponibilidad, altura_ola, tiempo_reparaciones, averías_multiplicadora, averías_generadora, averías_rotor, averías_otros

Dimensión	Clave primaria	Atributos
<b>d_fecha</b>	Id_fecha	Día, Mes, Año
<b>d_parque</b>	Id_Parque	CoordenadaX, CoordinadaY, nombre_parque, número_aerogeneradores
<b>d_pais</b>	Id_Pais	Nombre
<b>d_fabricante</b>	Id_Fabricante	Nombre
<b>d_modelo</b>	Id_Modelo	Nombre, fabricante
<b>d_mantenedora</b>	Id_Mantenedora	Nombre

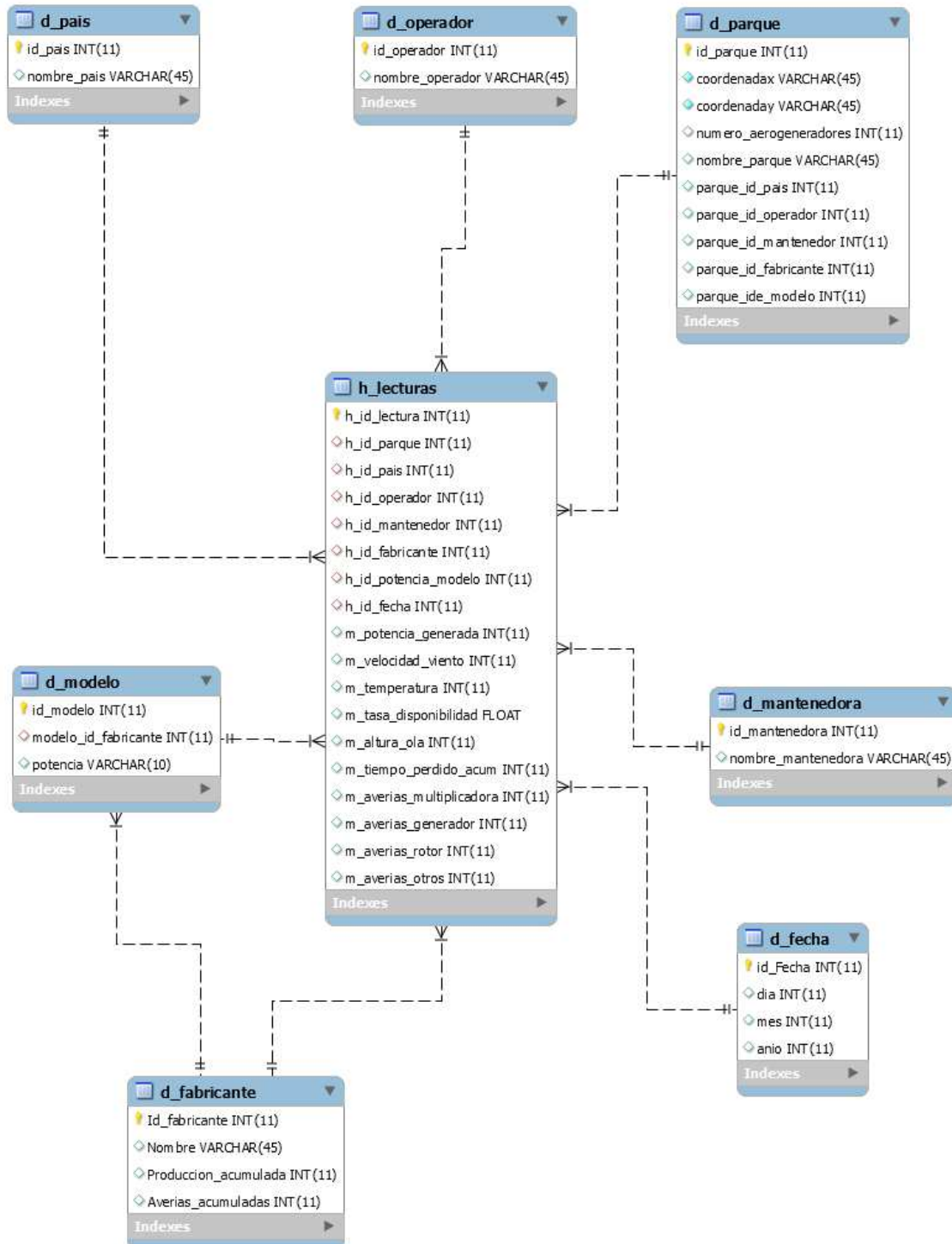
## 2.2.4. Alternativa finalmente adoptada de diseño lógico del data warehouse

El diseño por el que finalmente se ha optado (identificado como Enfoque 2) es el siguiente.

Tabla de Hecho	Claves foráneas	Métricas
<b>h_lectura</b>	Id_Parque, Id_Fecha, Id_País	Potencia_generada, velocidad_viento, temperatura, tasa_disponibilidad, altura_ola, tiempo_reparaciones, averias_multiplicadora, averias_generadora, averias_rotor, averias_otros

Dimensión	Clave primaria	Atributos
<b>d_fecha</b>	Id_Fecha	Día, Mes, Año
<b>d_parque</b>	Id_Parque	CoordenadaX, CoordenadaY, nombre_parque, número_aerogeneradores, mantenedor, operador, fabricante, modelo
<b>d_pais</b>	Id_Pais	Nombre

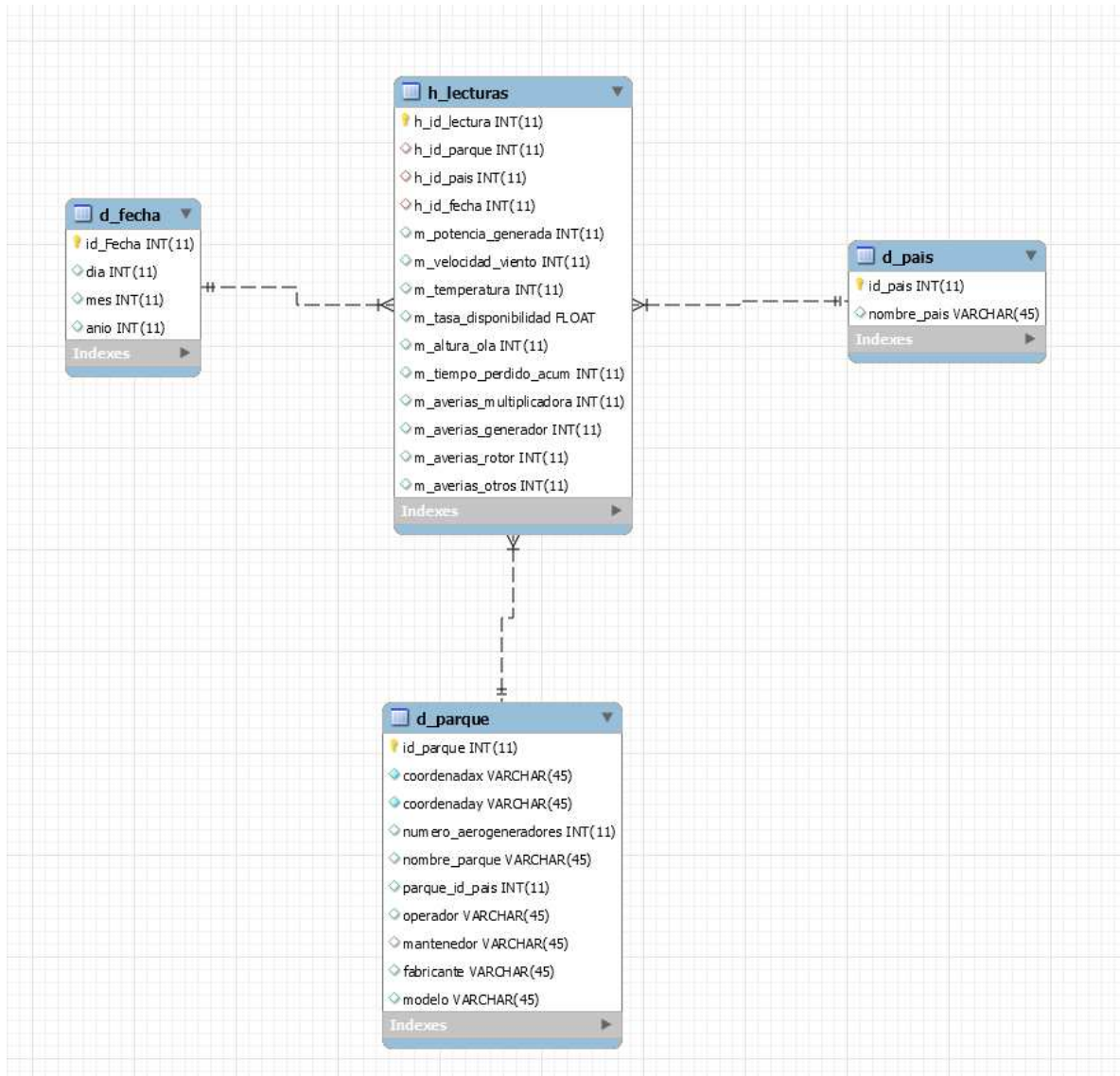
## 2.2.5. Diseño físico inicial del data warehouse



5 Diagrama del modelo físico inicialmente planteado elaborado con herramientas de MySQL



## 2.2.6. Diseño físico finalmente adoptado para el data warehouse



6 Diagrama del modelo físico ejecutado elaborado con herramientas de MySQL

## 2.3. Construcción de artefactos ETL

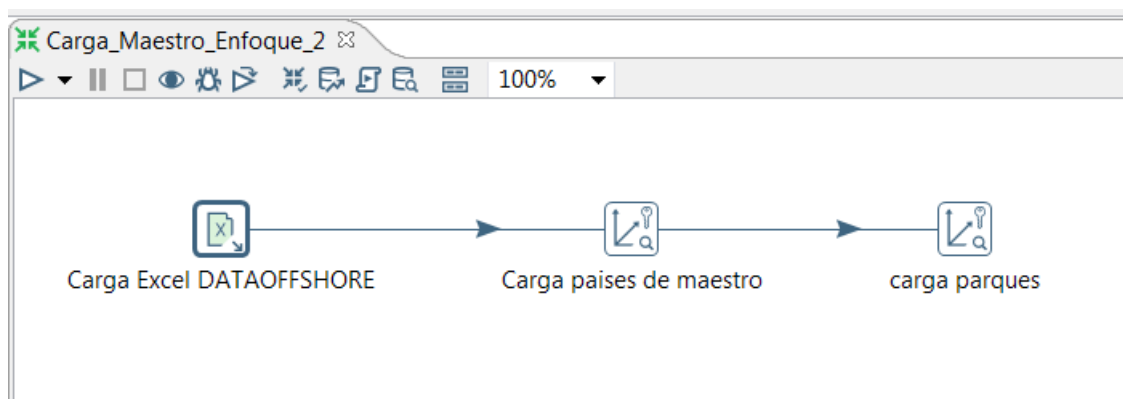
Se han diseñado dos transformaciones, una para la carga de los datos maestros, incluidos en la hoja OFFSHORE WIND FARM de DATAOFFSHORE.xls y otra para la carga de todas las lecturas, y que comprende todos los registros del resto de hojas del fichero Excel.

En el curso del proyecto, tras el planteamiento de un primer diseño del data warehouse, se optó por una segunda alternativa que facilitaría y simplificaría el resto del proyecto, como así ha sido. Dicha decisión se adoptó cuando ya se contaba con los procesos ETL adecuados a al modelado inicial del data warehouse.

Se han querido conservar los procesos antiguos (implementación y diseño), porque aunque desde la perspectiva del proyecto no han sido los que han generado el data warehouse con el que el mismo se ha continuado y finalizado, desde la perspectiva del conocimiento de la herramienta Pentaho PDI, responden a una casuística mayor, más compleja, y pueden resultar una ayuda de valor en el uso de la misma.

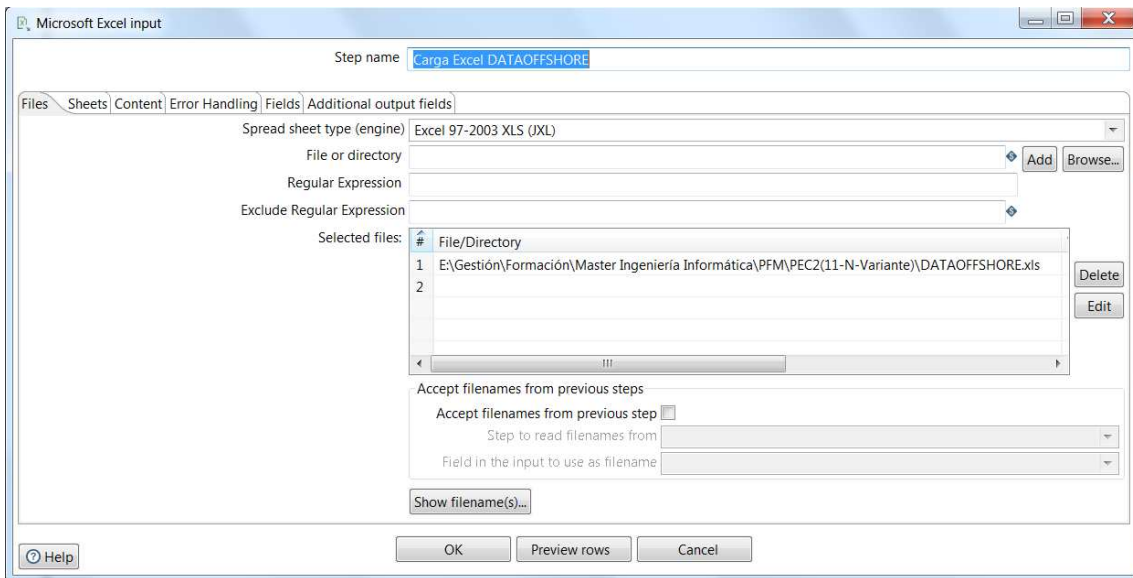
Los antiguos se han recogido en el Anexo I de la memoria, quedando aquí solo los posteriores.

### 1.1.1. Diseño ETL para la carga de datos maestros

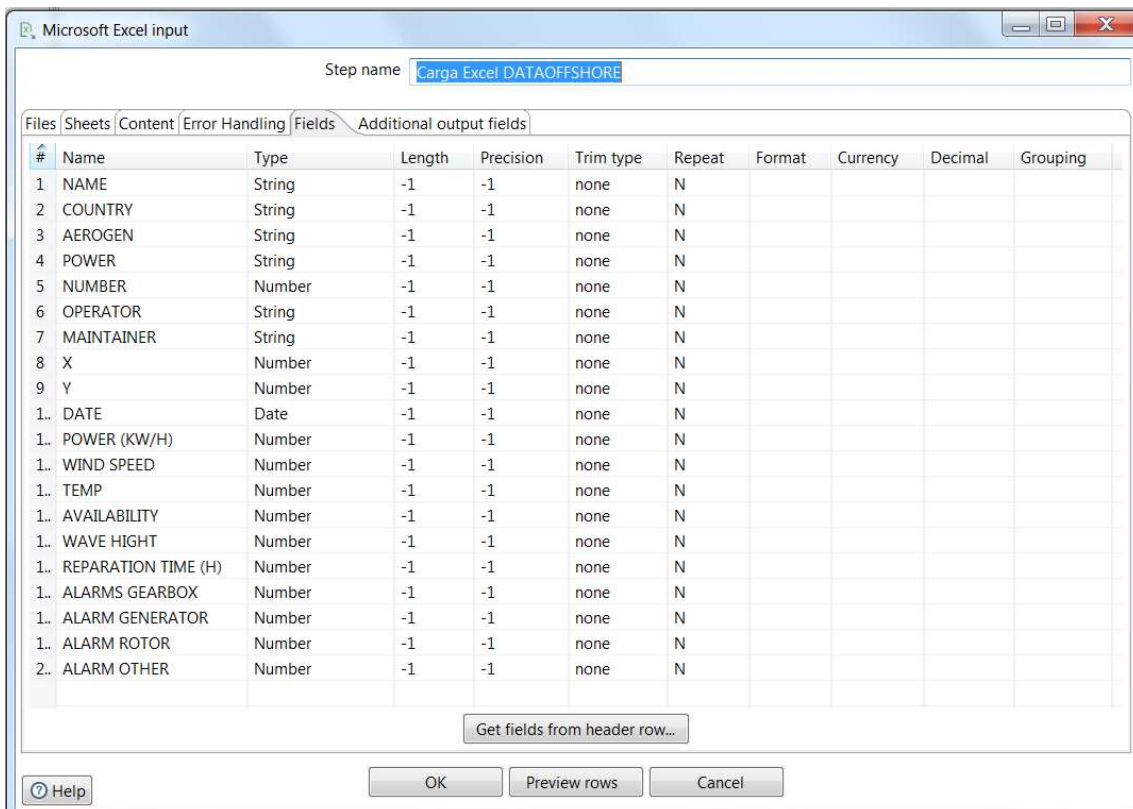


7 Imagen del diseño del proceso de carga de datos maestro

## a. Carga Excel DATAOFFSHORE



8 Imagen del detalle de configuración de la tarea de lectura del Excel de entrada de datos



9 Imagen del detalle de configuración de los campos de entrada

b. Carga países de maestro

Combination Lookup / Update

Step name: Carga países de maestro

Connection: conexion\_jmm\_db2

Target schema:

Target table: d\_pais

Commit size: 100

Cache size: 9999

Pre-load the cache?

Key fields (to look up row in table):

#	Dimension field	Field in stream
1	nombre_pais	COUNTRY

Technical key field: id\_pais

Creation of technical key:

Use table maximum + 1

Use sequence

Use auto increment field

Remove lookup fields?

Use hashcode?

Hashcode field in table:

Date of last update field (optional):

Buttons: Help, OK, Cancel, Get Fields, SQL

10 Imagen del detalle de configuración de tarea de búsqueda y actualización

### c. Carga parques

Step name

Connection

Target schema

Target table

Commit size  Cache size

Pre-load the cache?

Key fields (to look up row in table):

#	Dimension field	Field in stream
1	coordenadax	X
2	coordenaday	Y
3	nombre_parque	NAME
4	numero_aerogeneradores	NUMBER
5	parque_id_pais	id_pais
6	operador	OPERATOR
7	mantenedor	MAINTAINER
8	fabricante	AEROPEN
9	modelo	POWER

Technical key field

Creation of technical key

Use table maximum + 1

Use sequence

Use auto increment field

Remove lookup fields?

Use hashcode?

Hashcode field in table

Date of last update field (optional)

11 Imagen del detalle de configuración de tarea de búsqueda y actualización

### 1.1.2. Implementación ETL para la carga de datos maestros

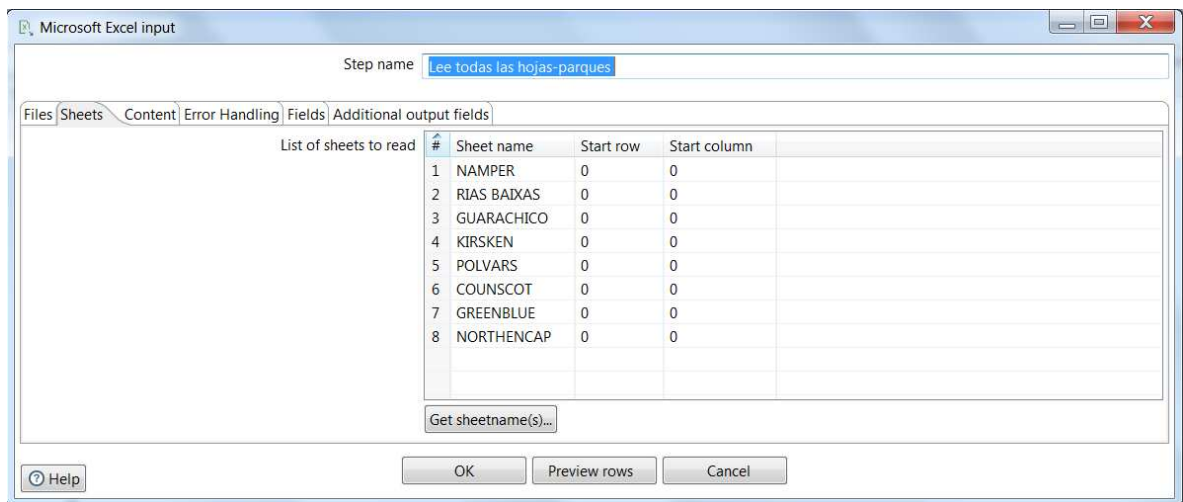
El fichero de formato XML, `Carga_datos_maestros_v10.ktr`, entregable del proyecto, alberga los metadatos resultantes de la implementación. Dicha información es interpretable por el motor PDI en la ejecución de las tareas, así como por la aplicación Spoon, capaz de construir el diagrama correspondiente con los objetos parametrizados, a partir de dicho fichero.

### 1.1.3. Diseño ETL para la carga de lecturas de todos los parques



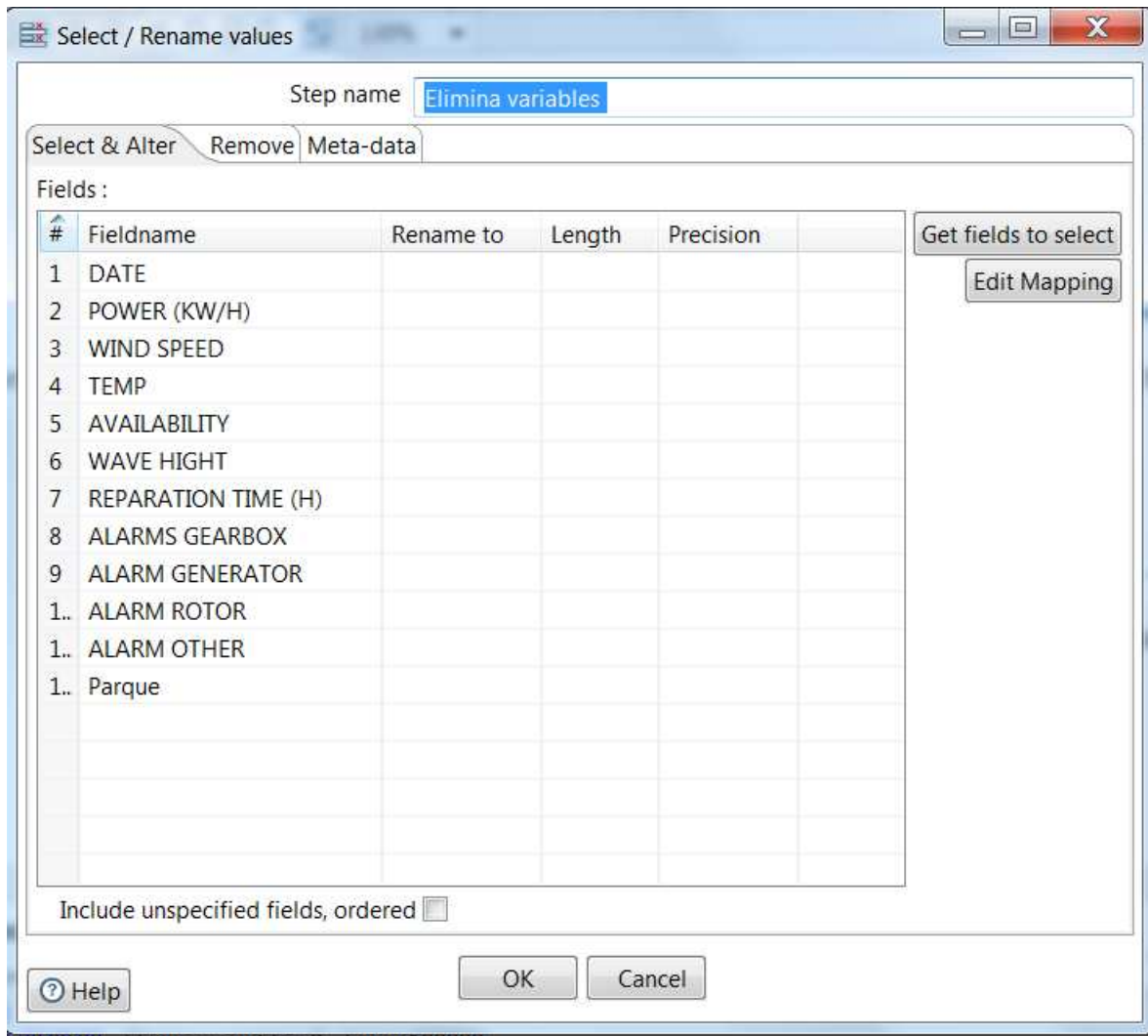
12 Imagen del diseño del proceso de carga de las lecturas

#### a. Lee las hojas-parques



13 Imagen con detalle de configuración de mapeo entre Excel de entrada y modelo

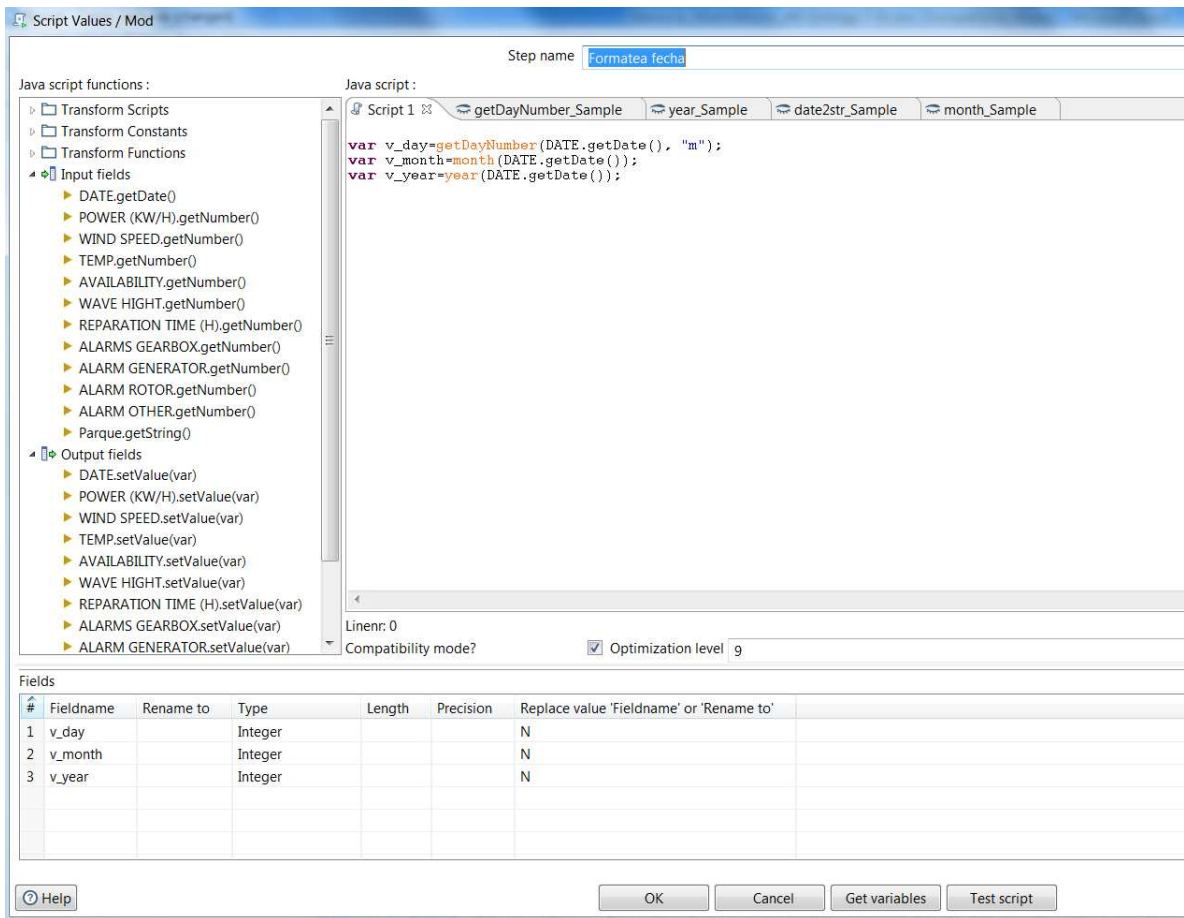
b. Elimina variables



14 Imagen detallada de limpieza de variables, quedando solo las válidas



### c. Formatea fecha



15 Imagen de detalle de tarea de formateo de fechas, estableciendo nuevas variables del flujo

#### d. Actualiza dimensión fecha

Step name: Actualiza dimensión fecha

Connection: Conexion

Target schema:

Target table: d\_fecha

Commit size: 100

Cache size: 999999

Pre-load the cache?

Key fields (to look up row in table):

#	Dimension field	Field in stream
1	Dia	v_day
2	Mes	v_month
3	Anio	v_year

Technical key field: id\_Fecha

Creation of technical key:

- Use table maximum + 1
- Use sequence
- Use auto increment field

Remove lookup fields?

Use hashcode?

Hashcode field in table:

Date of last update field (optional):

Buttons: Help, OK, Cancel, Get Fields, SQL

16 Detalle de configuración de la carga de dimensión fecha mediante búsqueda y actualización

e. Busca en dimensión Parque

Database Value Lookup

Step name: Busca en dimensión Parque

Connection: Conexion [Edit... New... Wizard...]

Lookup schema: [Browse...]

Lookup table: d\_parque [Browse...]

Enable cache?

Cache size in rows (0=cache): 0

Load all data from table

The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2
1	nombre_parque	=	Parque	

Values to return from the lookup table :

#	Field	New name	Default	Type
1	id_parque			Integer
2	coordenadax			String
3	coordenaday			String
4	numero_aerogeneradores			Integer
5	nombre_parque			String
6	parque_id_pais			Integer
7	parque_id_operador			Integer
8	parque_id_mantenedor			Integer
9	parque_id_fabricante			Integer
1..	parque_ide_modelo			Integer

Do not pass the row if the lookup fails

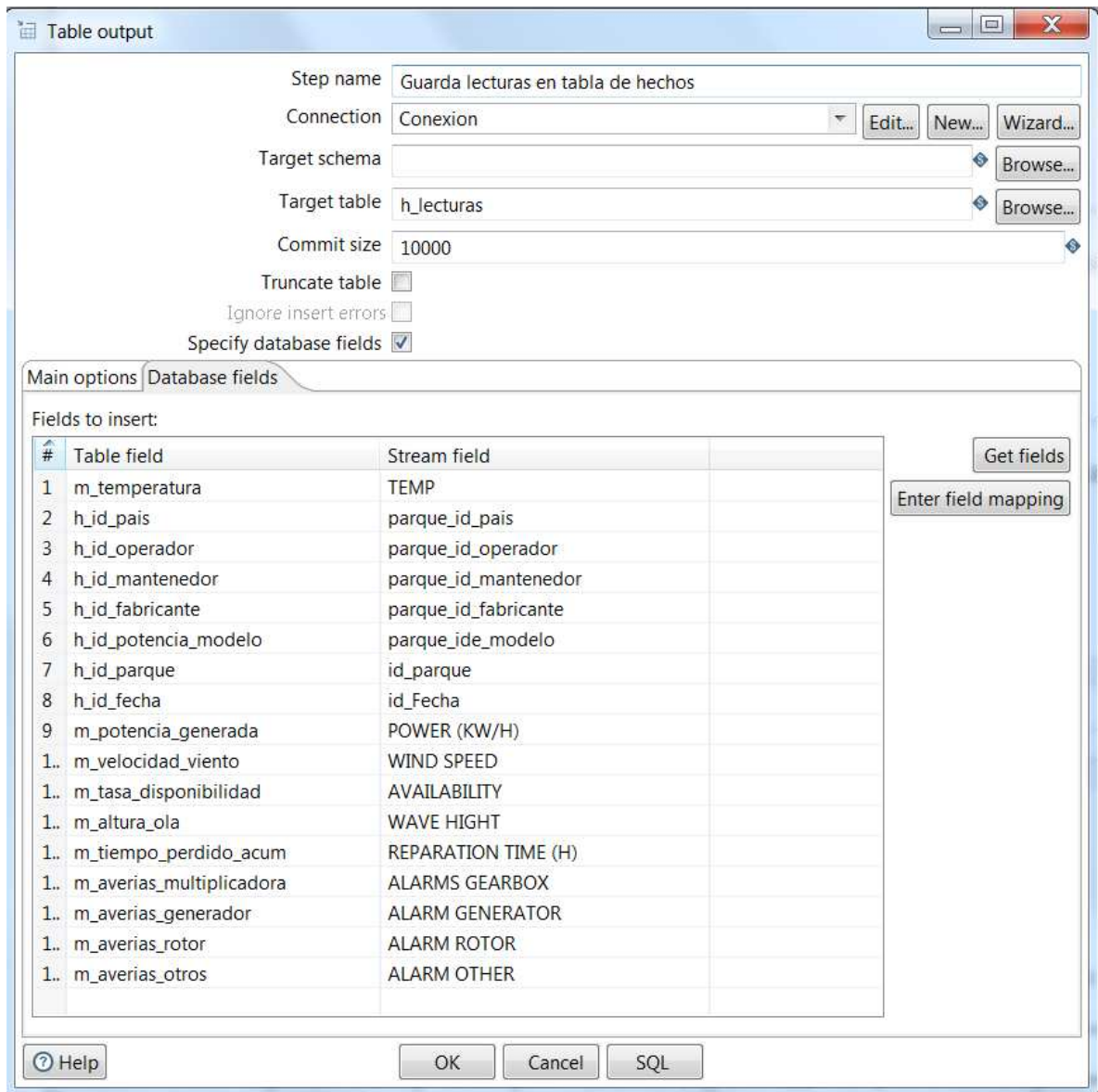
Fail on multiple results?

Order by: [ ]

[Help] [OK] [Cancel] [Get Fields] [Get lookup fields]

17 Imagen con detalle de configuración de la tarea

f. Guarda lecturas en tabla de hechos



18 Imagen de tarea de carga de datos en tabla de hechos

#### 1.1.4. Implementación ETL de carga de lecturas de todos los parques

El fichero de formato XML, `Carga_lecturas.ktr`, entregable del proyecto, alberga los metadatos resultantes de la implementación de los procesos ETL que cargan y procesan los registros de todos los parques.

## 2.4. Construcción de artefactos OLAP

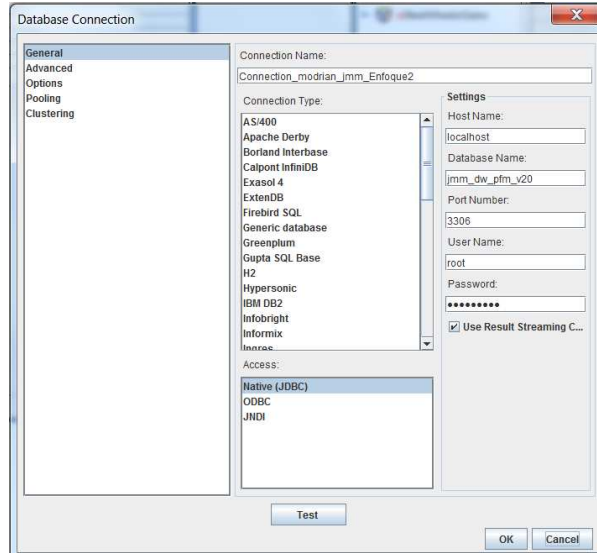
### 1.1.5. Diseño de artefactos OLAP

La versión Community de Pentho, ofrece dos herramientas para el diseño de cubos o recursos OLAP:

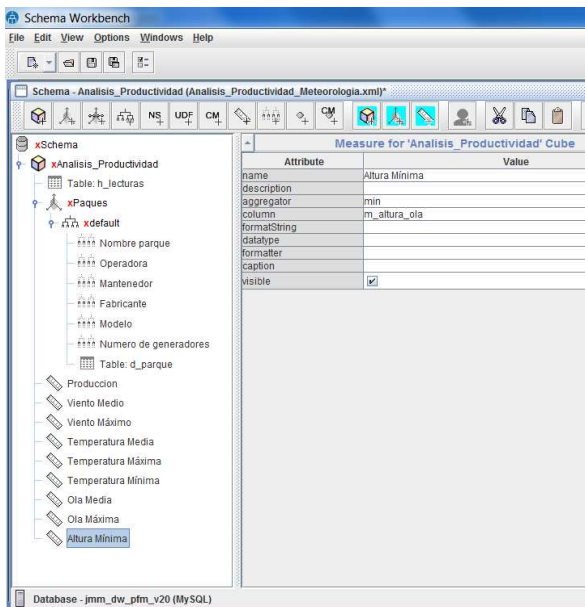
- **Schema Workbench:** Es una aplicación de cliente pesado (aplicación de escritorio), desarrollada en Java, y es la opción más completa para el diseño de cubos OLAP. Cuenta con un interfaz gráfico *drag and drop* y con un editor de MDX. Para una plena capacidad de diseño de cubos, es es la herramienta más adecuada dentro de la versión community.
- Desde el propio (PUC) Pentaho User Console, es posible también definir orígenes de datos y cubos, si bien presente más limitaciones. Es desde esta aplicación web donde se integra el visor OLAP JPivot.

### 1.1.6. Schema Workbench

La aplicación de escritorio requiere la configuración de una conexión a la base de datos del data warehouse, definida la cual, se está en disposición de crear un esquema sobre el que construir las correspondientes jerarquías en base a cubos y a sus dimensiones y métricas.



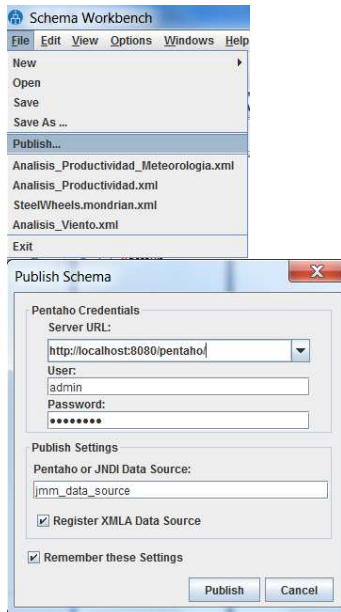
19 Imagen con configuración de conexión al data warehouse del proyecto



20 Imagen de un momento del diseño del cubo *Analisis\_Productividad\_Meteorologia*

Todo cubo requiere al menos de una tabla de hechos (en este caso *h\_lecturas*), de al menos una dimensión (Parque) y de las métricas a estudiar.

El trabajo realizado se alberga en un fichero xml (en este caso el artefacto generado es *Analisis\_Productividad\_Meteorologia.xml*) que puede ser trabajado de diversas formas.



La opción de publicación, crea en el portal Pentaho el origen de datos correspondiente, con oportunidad de practicar alguna modificación.

*21 Opción de publicación y detalle de cuenta en (PUC)  
Pentaho User Console*

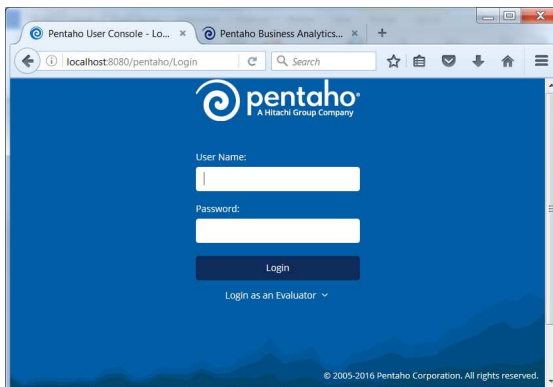
En caso de que el cubo sea correcto, se logra la publicación del mismo



*22 Mensaje de publicación realizada*



### 1.1.7. (PUC) Pentaho User Console

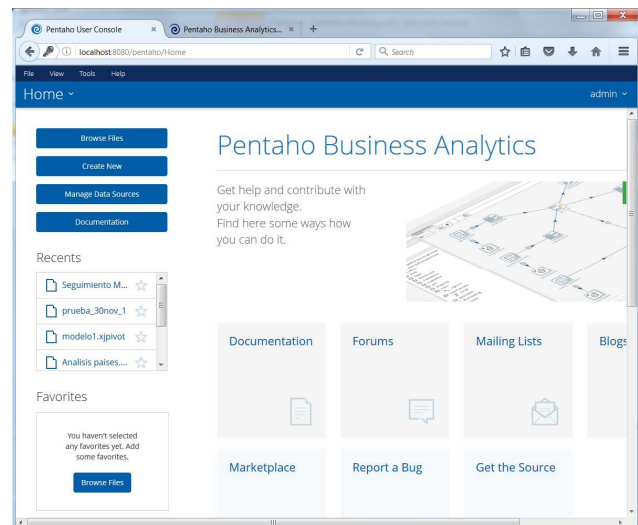


23 Imagen de login en PUC

Se trata de una aplicación Internet que requiere del uso de un servidor web, que es proporcionado en el propio paquete de instalación (se proporciona Tomcat), si bien en un entorno de producción se podría hacer uso de otro producto.

Cuenta con una gestión de usuarios que lo hace ideal para su uso en Intranet de corporaciones, como punto de acceso de los diferentes recursos y análisis en los que se vaya trabajando.

El foco fundamental de la aplicación es la gestión e los orígenes de datos, que se pueden definir desde este portal o importar procedentes de las herramientas de escritorio como Schema Workbench.



24 Menú principal PUC

El primer uso que haremos del PUC será visualizar el cubo antes creado con Schema Workbench, para lo que desde el menú principal, seleccionaremos File/New/JPIVOT view, dando lugar a la visualización del mismo.

Paques	Measures										
	· Produccion	· Viento Medio	· Viento Máximo	· Viento mínimo	· Temperatura Máxima	· Temperatura Mínima	· Temperatura Media	· Ola Máxima	· Altura Mínima	· Ola Media	· Tasa Productividad
▣ Todos los parques	255,187,114	13.931	33	2	20	3	9.36	18	2	7.373	98.475
▣ COUNSCOT	34,864,341	18.751	30	8	13	3	7.983	18	2	9.974	98.925
▣ EOLLAGER	78,197,557	13.544	30	2	20	3	11.153	18	2	7.01	98.658
▣ CUPRA	86,142,387	15.289	33	2	20	3	9.491	18	2	7.988	97.527
▣ GDEP	51,166,836	10.36	30	3	15	3	8.478	18	2	6.99	99.307
▣ MANWIN	117,877,891	17.251	30	4	18	3	10.487	12	2	7.023	98.649
▣ GREENBLUE	18,024,622	18.746	30	8	13	3	7.989	18	2	9.981	95.692
▣ ACTIONAK	27,011,299	11.53	30	3	15	3	8.993	18	2	7.486	97.523
▣ GUARACHICO	8,684,325	4.613	8	2	20	10	14.986	6	2	4.022	98.414
▣ EOLLAGER	78,197,557	13.544	30	2	20	3	11.153	18	2	7.01	98.658
▣ KIRSKEN	34,648,891	17.269	30	4	18	3	10.49	12	2	7.035	98.634
▣ EOLLAGER	78,197,557	13.544	30	2	20	3	11.153	18	2	7.01	98.658
▣ NAMPER	7,315,818	8.013	10	6	12	3	7.456	10	2	6.007	99.642
▣ EUROWIND	149,978,258	15.918	33	4	18	3	7.812	18	2	7.66	98.926
▣ NORTHCAP	59,433,440	22.508	33	12	8	3	5.496	18	2	9.961	98.473
▣ EUROWIND	149,978,258	15.918	33	4	18	3	7.812	18	2	7.66	98.926
▣ POLVARS	83,229,000	17.233	30	4	18	3	10.485	12	2	7.011	98.663
▣ EUROWIND	149,978,258	15.918	33	4	18	3	7.812	18	2	7.66	98.926
▣ RIAS BAIKAS	8,986,677	4.315	6	3	15	5	9.997	8	2	4.989	99.354
▣ ACTIONAK	27,011,299	11.53	30	3	15	3	8.993	18	2	7.486	97.523

25 Visualización de cubo

Otros cubos creados para obtener una visión global del modelo han sido, por ejemplo, el de Productividad Ampliada:

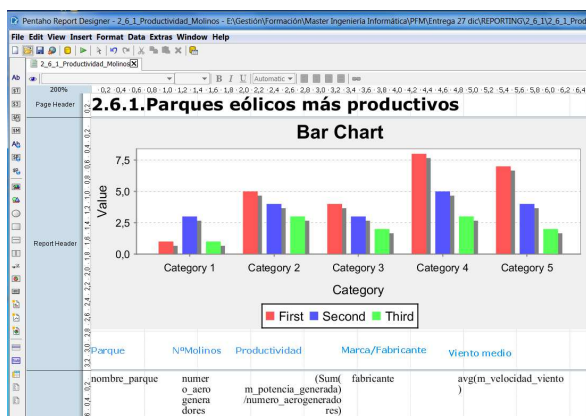
Parque	Fabricante	Modelo	Measures						
			· Productividad	· Disponibilidad	· Ola_Media	· Viento_Medio			
(All) Parque	(All) Fabricante	(All) Modelo							
▣ All Parques	▣ All Fabricantes	▣ All Modelos	9,409,040.609	795.64	58.982	111.448			
All Parques	COUNSCOT	▣ All Fabricantes	▣ All Modelos	1,202,218.655	99.483	9.974	18.751		
		All Fabricantes VESTAS	▣ All Modelos	1,202,218.655	99.483	9.974	18.751		
	GREENBLUE	All Fabricantes ALSTON	All Modelos	2.0 KW	1,202,218.655	99.483	9.974	18.751	
			▣ All Fabricantes	▣ All Modelos	1,802,462.2	99.312	9.981	18.746	
			All Modelos	3.0 KW	1,802,462.2	99.312	9.981	18.746	
	GUARACHICO	All Fabricantes GAMESA	▣ All Fabricantes	▣ All Modelos	289,477.5	97.4	4.022	4.613	
			All Fabricantes	All Modelos	2.0 KW	289,477.5	97.4	4.022	4.613
			All Modelos	2.0 KW	289,477.5	97.4	4.022	4.613	
	KIRSKEN	All Fabricantes ENERCON	▣ All Fabricantes	▣ All Modelos	1,385,955.64	100	7.035	17.269	
			All Fabricantes	All Modelos	2.5 KW	1,385,955.64	100	7.035	17.269
			All Modelos	2.5 KW	1,385,955.64	100	7.035	17.269	
	NAMPER	All Fabricantes VESTAS	▣ All Fabricantes	▣ All Modelos	487,721.2	99.76	6.007	8.013	
			All Fabricantes	All Modelos	2.0 KW	487,721.2	99.76	6.007	8.013
			All Modelos	2.0 KW	487,721.2	99.76	6.007	8.013	
	NORTHCAP	All Fabricantes ALSTON	▣ All Fabricantes	▣ All Modelos	1,801,013.333	99.925	9.961	22.508	
			All Fabricantes	All Modelos	2.5 KW	1,801,013.333	99.925	9.961	22.508
			All Modelos	2.5 KW	1,801,013.333	99.925	9.961	22.508	
	POLVARS	All Fabricantes ENERCON	▣ All Fabricantes	▣ All Modelos	2,080,725	100	7.011	17.233	
			All Fabricantes	All Modelos	2.5 KW	2,080,725	100	7.011	17.233
			All Modelos	2.5 KW	2,080,725	100	7.011	17.233	
RIAS BAIKAS	All Fabricantes ALSTON	▣ All Fabricantes	▣ All Modelos	359,467.08	99.76	4.989	4.315		
		All Fabricantes	All Modelos	2.5 KW	359,467.08	99.76	4.989	4.315	
		All Modelos	2.5 KW	359,467.08	99.76	4.989	4.315		

26 Visualización de cubo Productividad

## 2.5. Obtención de informes

Para la generación de informes se ha hecho uso de la aplicación de escritorio Pentaho Report Designer, que al igual que el resto de utilidades requiere de la correspondiente configuración para su conexión con la base de datos que implementa el modelo dimensional que implementa el data warehouse.

Adicionalmente puede trabajar con los metadatos resultantes de los diseños obtenidos tanto con Schema Workbench como como con Pentaho Metada Editor.

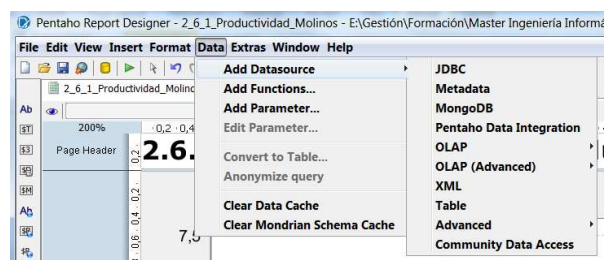


27 Pantalla principal con un documento en edición

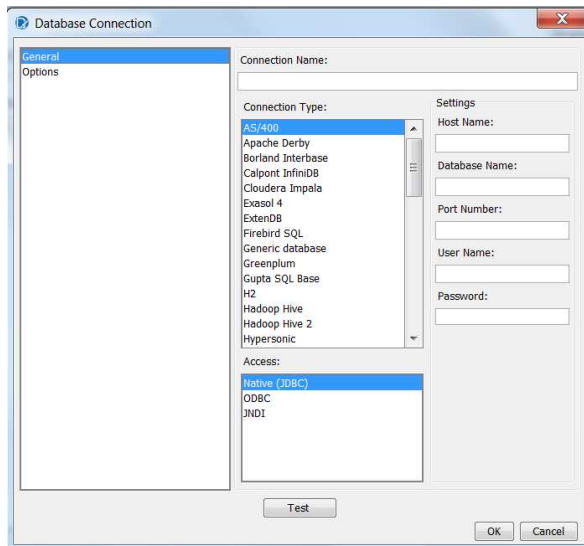
Un paso esencial es la conexión con una fuente de datos, permitiendo tanto el acceso de directo a la base de datos del data warehouse o bien restringiendo el mismo a los datos resultantes de un cubo OLAP, o bien utilizar como origen de datos ficheros XML o conexiones con otros repositorios.

Report Designer, cuenta con un canvas en el que construir los informes. Para ello se puede arrastrar bien objetos desde la paleta que por defecto aparece en la banda izquierda o elementos de datos.

Los objetos son desde cuadro de texto convencionales hasta un amplio abanico de gráficos que permiten la confección de informes emulando cuadros de mando.



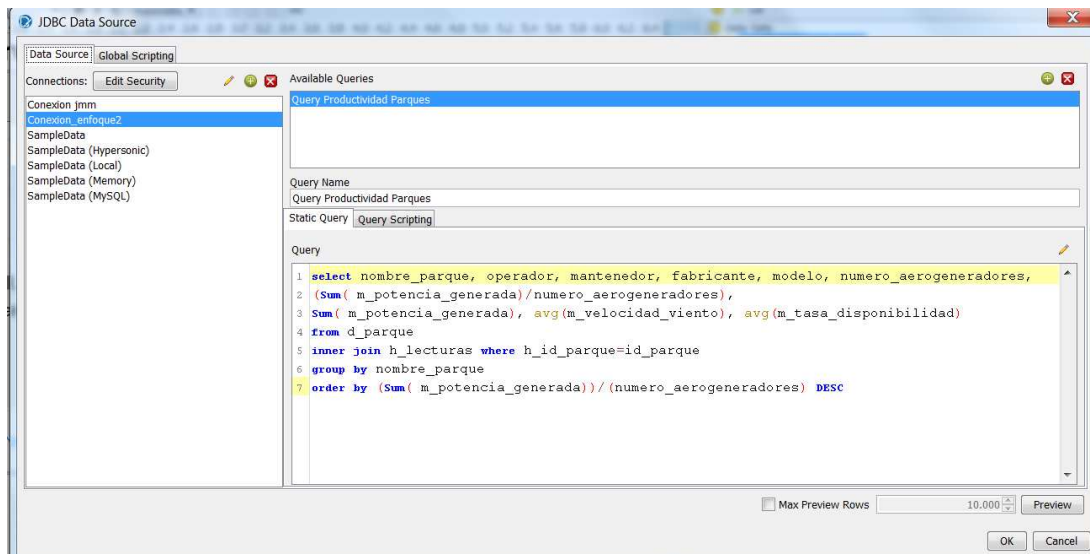
28 Jerarquía de menú para la indentificación de la fuente de datos



29 Configuración de conexión JDBC

Para poder definir fuentes de datos basadas en el propio sistema de gestión de base de datos, se debe crear previamente una conexión a base de datos, donde especificar el producto correspondiente (MySQL, Oracle, Postgre, SQL Server y un largo abanico de opciones) y los datos de la conexión y del esquema o base de datos.

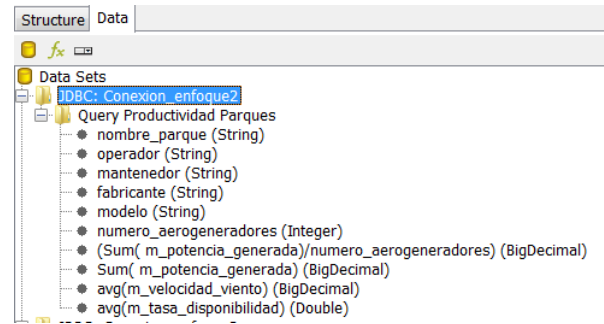
Es necesario proporcionar el driver del fabricante en el directorio correspondiente de la aplicación.



30 Ejemplo de configuración de fuente de datos basada en una conexión directa / data warehouse, en base a la consulta SQL que se puede observar

Una fuente de datos basada en una conexión JDBC a una base de datos, se define en base a una consulta concreta a la misma.

Los elementos resultantes de la consulta, son tratados visualmente como objetos a arrastrar al canvas o están disponibles, por ejemplo, al definir un gráfico. Definir los datos, es en cualquier caso una labor previa a la confección del informe.



31 Visualización de caja de objetos de datos para su arrastre (drag and drop) al documento en edición

## 2.6. Resolución de los problemas que originaron el proyecto

Los artefactos obtenidos en las etapas anteriores, nos permiten satisfacer las necesidades de información que originaron el proyecto, buscándose el mayor uso posible de los diferentes recursos que la herramienta dispone.

### 2.6.1. Parques eólicos más productivos

Para obtener la productividad, recurrimos al cociente de producción del parque entre los aerogeneradores instalados, accediendo así al término productividad por molino, como elemento representativo de la productividad del parque.

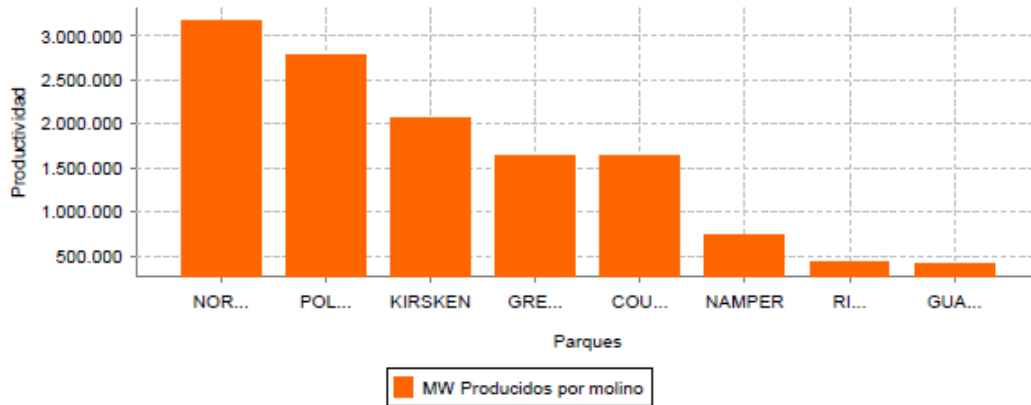
Mediante el uso de los cubos y su visualización con JPivot, logramos una aproximación a la realidad existente que nos adelanta las respuestas buscadas:

- De acuerdo al último fichero de lecturas, el parque más productivo es Northencap con una producción por molino de 3.171.090 MW.
- Dicha productividad guarda una relación directa con la media de viento, pues este parque tiene la mejor media de viento de todos los estudiados.
- Se observa como el fabricante de Northencap, que es Alston, cuenta con la mejor producción por molino, en el total de los parques. Podría plantearse que debe dicha posición a ser el producto de Northencap donde se dan las mejores condiciones de viento, pero su promedio productivo por molino en el conjunto de parques es aun mejor que el que obtiene de forma individual en Northencap.
- En cubos e informe, se incluye la tasa de disponibilidad, verificándose que nuevamente Alston, cuenta con el mejor ratio en el total de parques.

Para una formalización y entrega de todas las conclusiones anteriores, se ha construido el artefacto de reporting (2\_6\_1\_Productividad\_Molinos.prpt), fichero de metadatos Pentaho que forma parte de los entregables del proyecto y cuya información se incluye en la imagen posterior.

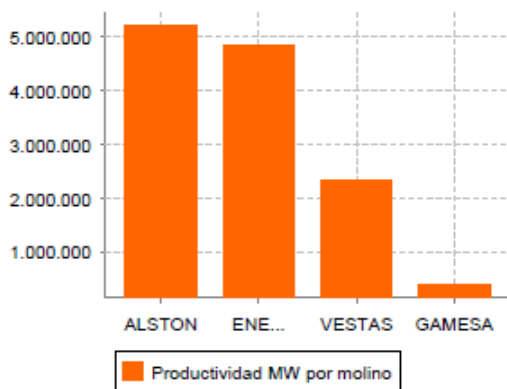
### 2.6.1. Parques eólicos más productivos

**Productividad por parques**

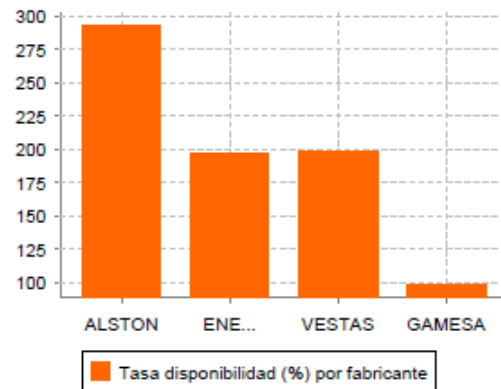


Parque	NºMolinos	Productividad	Marca/Fabricante	Viento medio
NORTHENCAP	33	3,171,090.7879	ALSTON	22.4806
POLVARS	40	2,780,110.65	ENERCON	17.292
KIRSKEN	25	2,058,578.44	ENERCON	17.1512
GREENBLUE	10	1,627,080.5	ALSTON	18.7634
COUNSCOT	29	1,626,333.5517	VESTAS	18.7918
NAMPER	15	728,415.2667	VESTAS	7.9933
RIAS BAIXAS	25	420,373.92	ALSTON	4.3524
GUARACHICO	30	403,084.8	GAMESA	4.5888

**Productividad por fabricante**



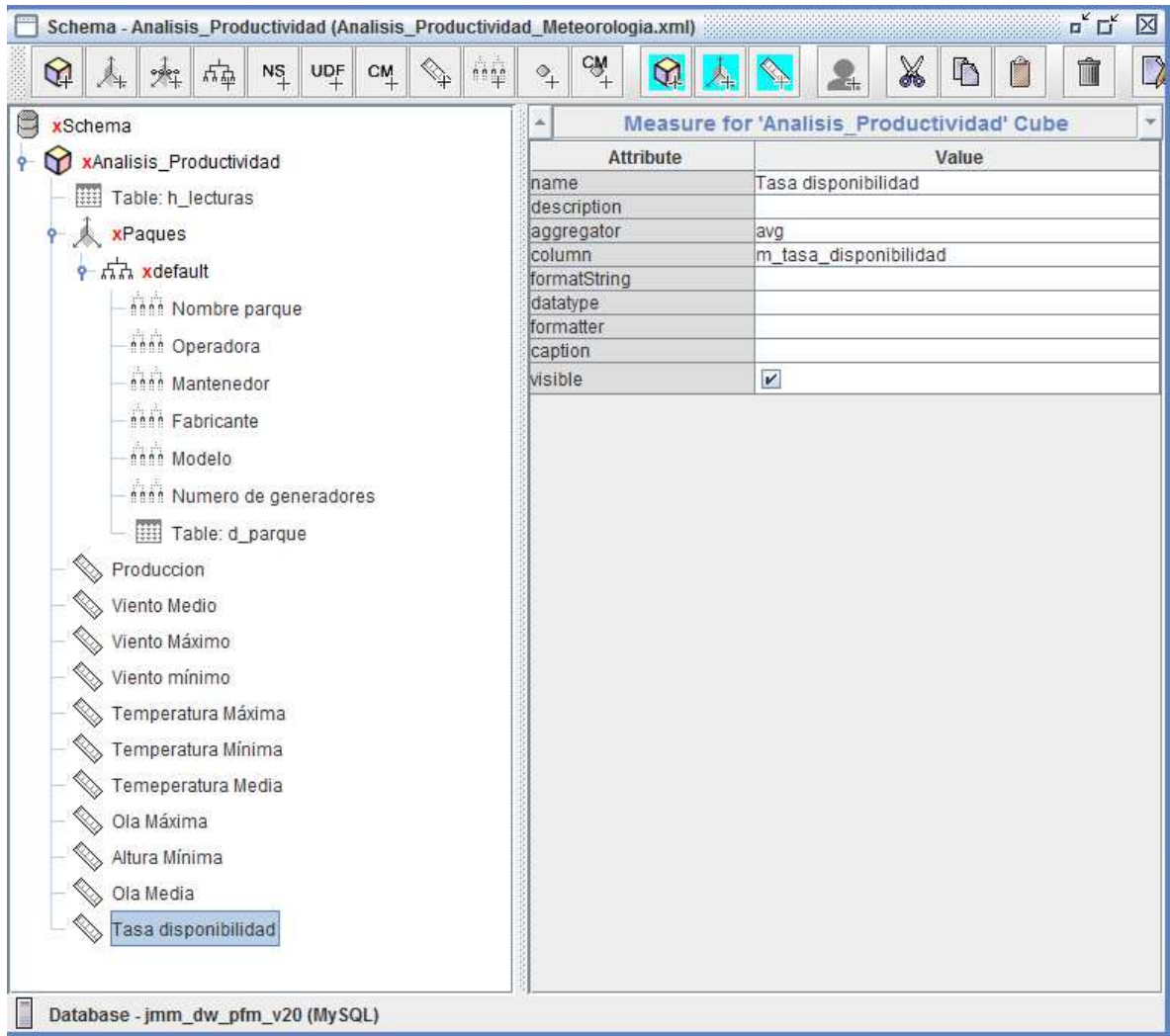
**Disponibilidad por fabricante**



32 Imagen de informe sobre productividad elaborado con Pentaho Report

## 2.6.2. Incidencia de los hechos meteorológicos sobre la producción

Para atender a esta petición, creamos en primer término con Workbench el cubo Analisis\_Productividad\_Meteorologia, tal y como refleja la siguiente figura:

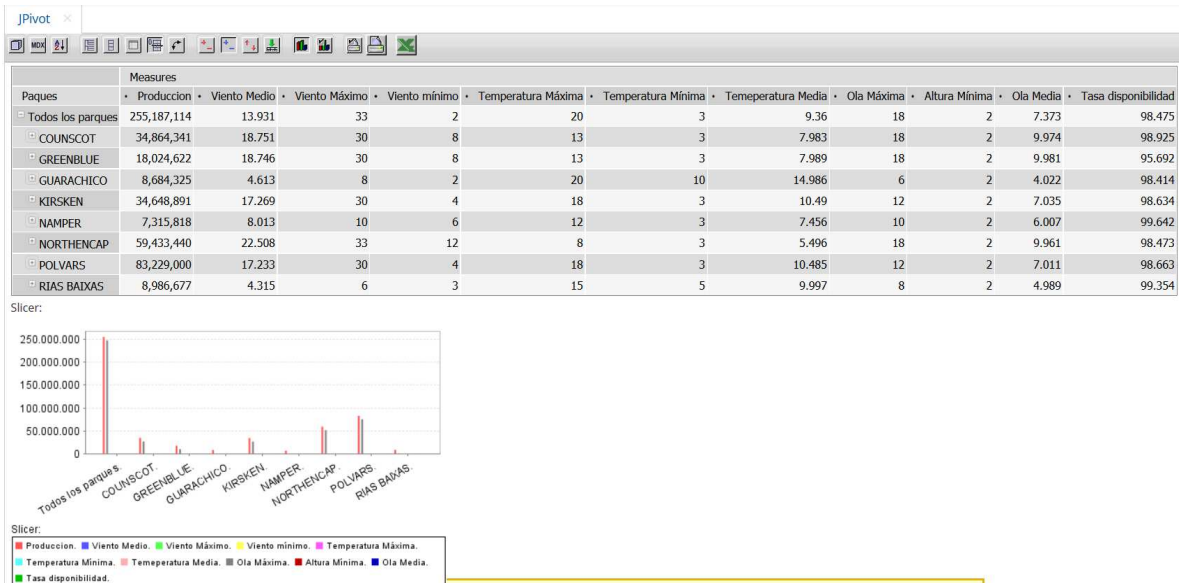


33 Imagen de confección de cubo mediante Schema Workbench

Para cada una de las variables meteorológicas con las que contamos, hemos tomado la mínima, la media y la máxima, además de otras características del parque, como norma general, que nos permita identificar hechos laterales.



La visualización de cubo OLAP con JPivot, nos ofrece la siguiente imagen, sobre las que orientamos las conclusiones posteriores y el diseño del reporting que proporciona datos concluyentes.

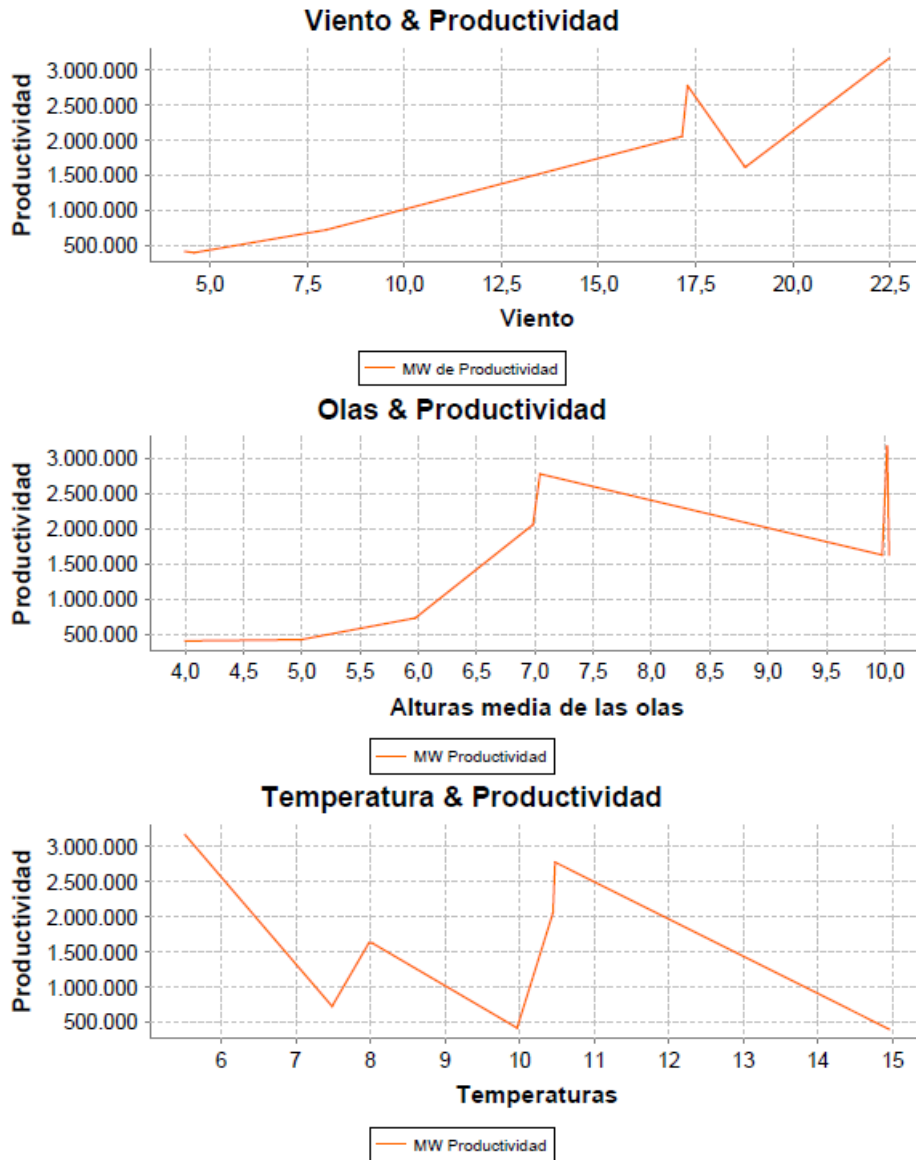


34 Imagen de uso de cubo y gráfico con JPivot

La respuesta se reduce a un problema de identificación de dependencia entre variables, cuya resolución más sencilla es obtener los gráficos correspondientes. El artefacto 2\_6\_2\_Meteo\_Produccion.prpt proporciona dichos gráficos, permitiendo las siguientes conclusiones:

- Más viento, más productividad.
- Hasta olas de 7 metros de altura media, la producción crece con las olas. Hay que considerar aquí que a su vez, las olas son una consecuencia del viento, por tanto más viento, más olas, más producción, pero a partir de cierta cantidad de olas, la producción cae, sin duda por la incidencia de estas sobre la disponibilidad.
- La temperatura vuelve a ser una variable afectada por el viento, y a su vez por tramos horarios. No es fácil interpretar sin la información técnica suficiente (termodinámica, meteorología, mecánica de fluidos) la información que se obtiene, sin embargo se observan tendencias que relacionan productividad y temperatura.

## 2.6.2. Incidencia de la meteorología sobre la producción



35 Imagen de análisis gráfico mediante Pentaho Report sobre la relación entre Meteorología y Productividad

### 2.6.3. Zonas con mejor relación viento potencia

Para analizar esta relación, debemos buscar una variable que iguale condiciones entre los diferentes parques, que responden a volúmenes de molinos instalados diferentes.

También hay diferencias entre las potencias nominales de los molinos de unas y otras zonas, pero entiendo que las mismas no tienen un impacto tan determinante, por lo que este aspecto, aunque lo observaremos no lo integraremos en los cálculos.

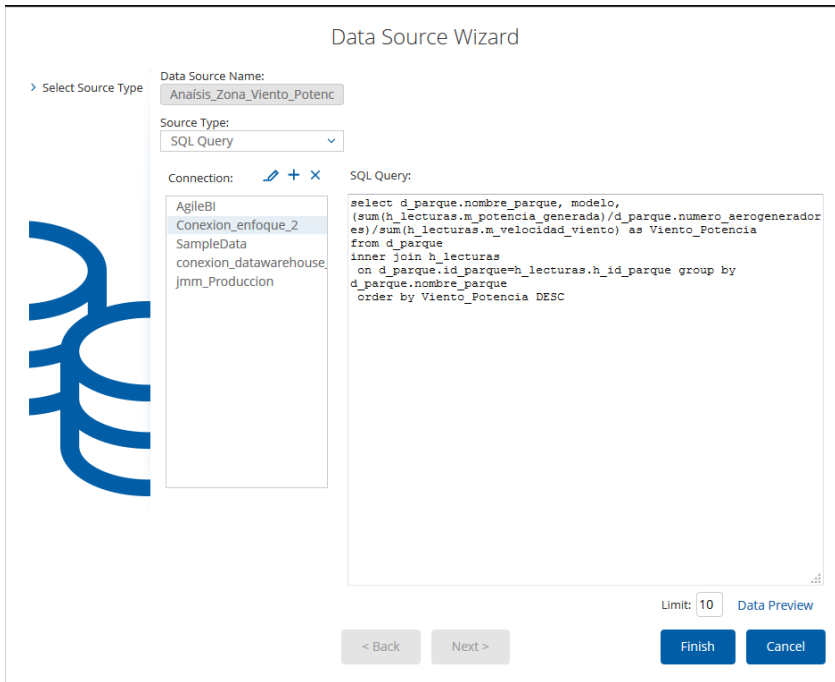
Para igualar las condiciones de todas las zonas o parques en estudio, haremos uso de la tasa de productividad antes ya utilizada y la relacionaremos con la suma de lecturas de viento obtenidas.

Es decir, Relación viento-potencia = Productividad por molino / suma de mediciones de viento.

Para este análisis, haremos uso del diseño de vistas desde el propio Pentaho User Console, como origen de datos del cuadro JPivot sobre el que realizaremos el análisis.

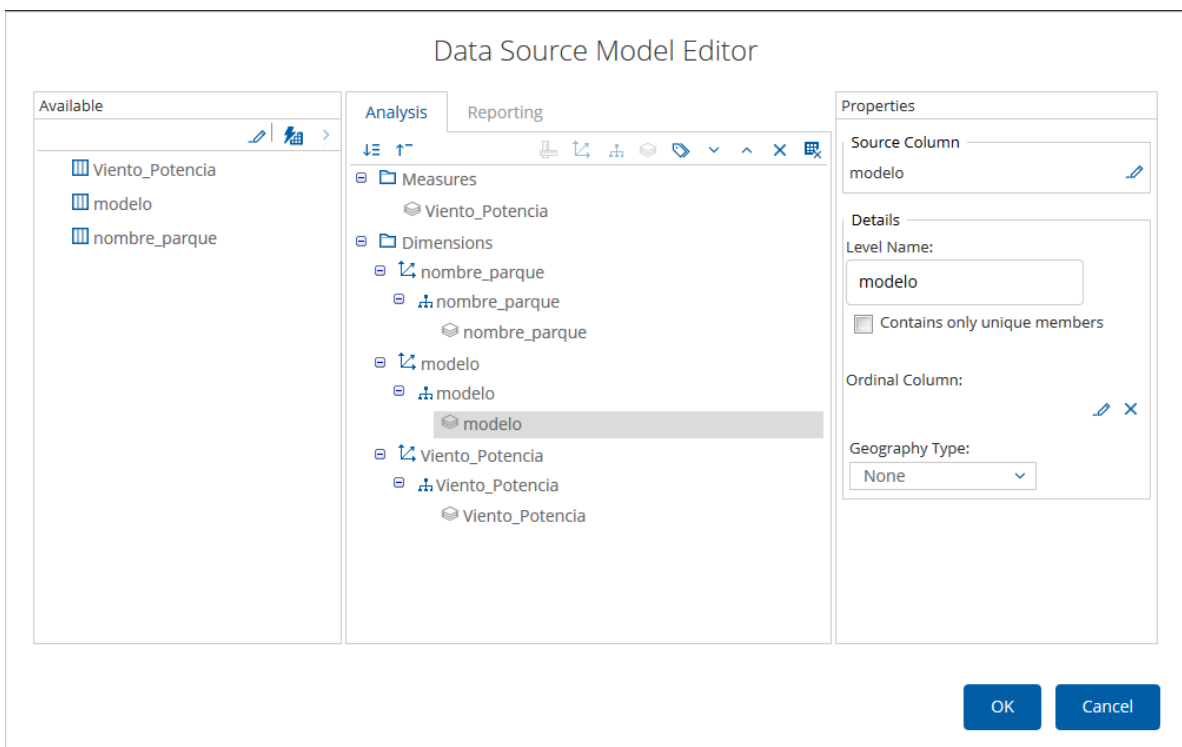
Para todo lo anterior seguiremos los siguientes pasos:

**Paso 1:** Implementación de la vista bajo la denominación `Análisis_Zona_Viento_Potencia` como origen de datos, mediante la sentencia SQL siguiente que ataca el data warehouse.



36 Imagen de creación de origen de datos basado en una vista

**Paso 2:** Definición del cubo, que esta vez realizaremos desde el interfaz al efecto de Pentaho User Console (como alternativa a Schema Workbench)



37 Imagen de modelado de cubo OLAP desde PUC

### Paso 3: Visualización mediante JPivot

		Measures
		Viento_Potencia
		Viento_Potencia
nombre_parque	modelo	• All Viento_Potencias
All nombre_parques	All modelos	67.91
	2.0 KW	20.709
	2.5 KW	40.437
	3.0 KW	6.764
COUNSCOT	All modelos	6.75
GREENBLUE	All modelos	6.764
GUARACHICO	All modelos	6.851
KIRSKEN	All modelos	9.362
NAMPER	All modelos	7.108
NORTHENCAP	All modelos	11.002
POLVARS	All modelos	12.54
RIAS BAIXAS	All modelos	7.533

38 Visualización de cubo desde JPivot

En esta ocasión la visualización del cubo nos permite concluir directamente:

- Northencap constituye la zona con mejor relación viento-potencia y por consiguiente es el candidato ideal para incrementar su parque instalado de aerogeneradores.
- Adicionalmente, los equipos de 2,5 MW son los que presentan mejor relación viento-potencia.

#### 2.6.4. Incidencia de los hechos meteorológicos sobre las alarmas técnicas

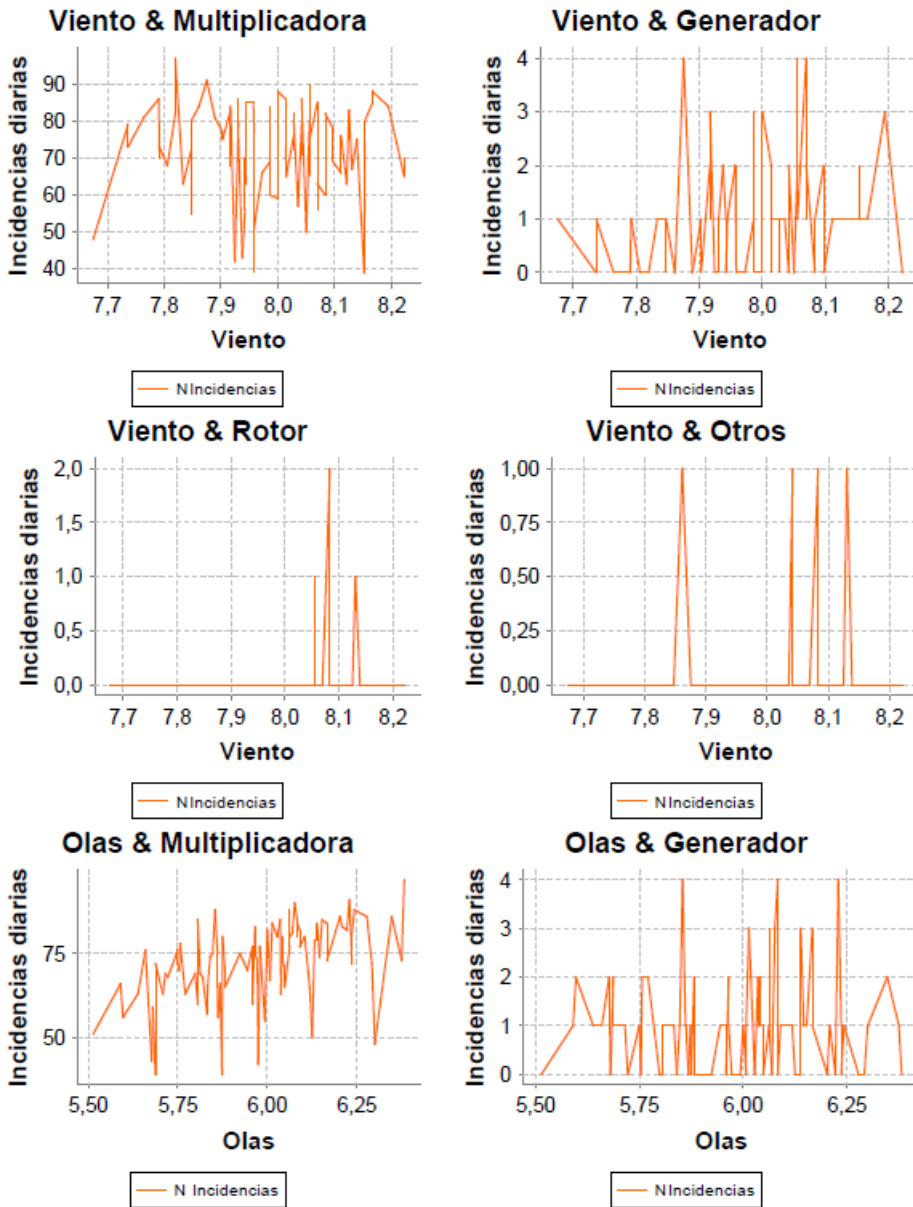
Se ha enfocado este análisis con una granularidad diaria, analizando por separado cada tipo de incidencia y hecho meteorológico, y estudiado a nivel de parque.

Se ha creado para ello un artefacto de reporting Pentaho por parque.

Las conclusiones generales que se pueden extraer al respecto son las siguientes:

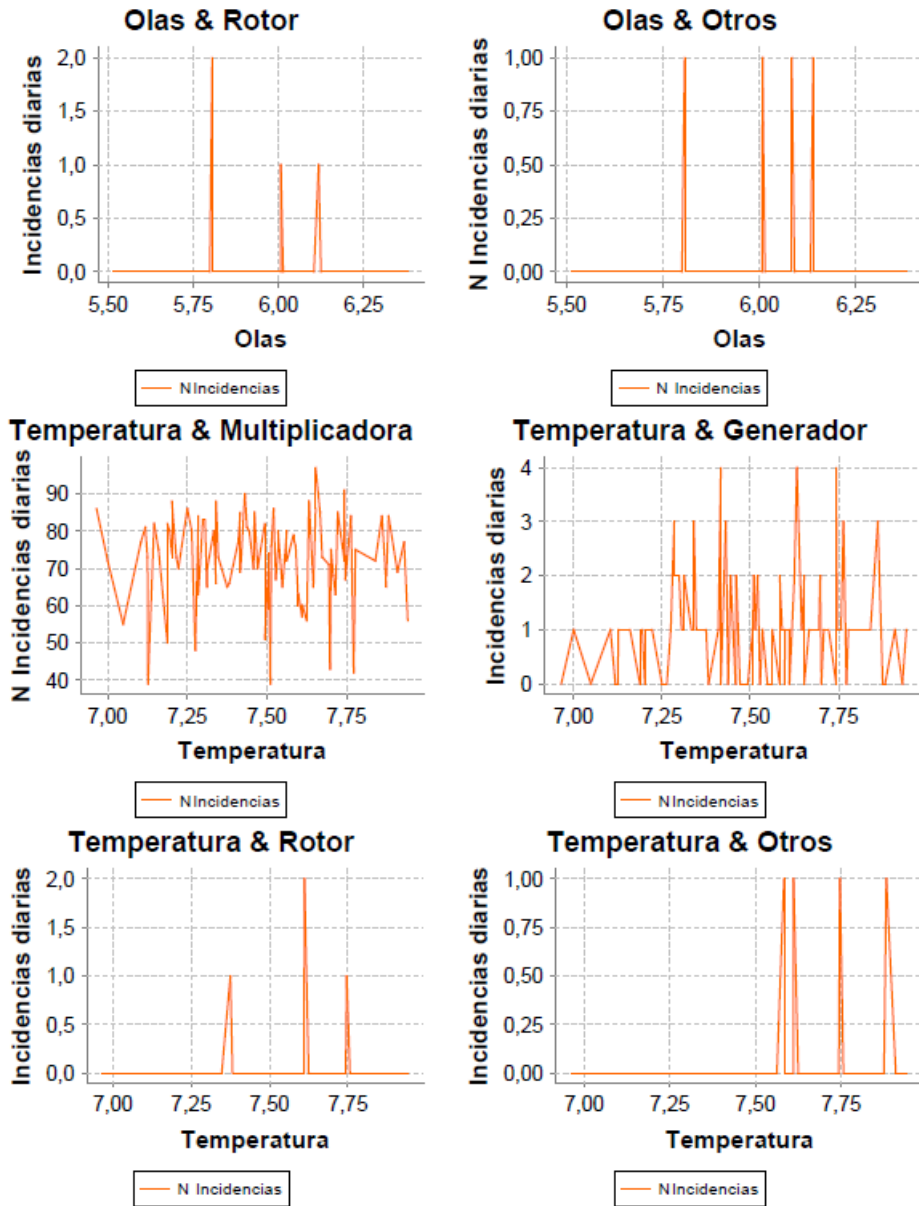
- Existen unos valores meteorológicos para las diferentes variables, a partir de los cuales las probabilidades de fallos en los diferentes componentes del sistema se consolidan, o lo que es lo mismo, más allá de ciertos valores hay incidencias seguras.
- Los patrones generales se mantienen entre los parques, si bien, los umbrales varían de unos a otros. Así, a modo de ejemplo, en Namper para entrar en la zona de riesgo viento-multiplicadora, hay que alcanzar los 7,7 kms/h y en Rias Baixas con 4,5 las incidencias son muy elevadas.
- La tendencia más constante de crecimiento-incidencia, la marcan las olas respecto de la multiplicadora.
- En general, se puede diseñar un sistema que anticipe incidencias partiendo de los valores meteorológicos, si bien, resultaría de interés extender este análisis en relación a los modelos y fabricantes, aunque se entiende que ese esfuerzo queda fuera del alcance del presente proyecto.

### 2.6.4.1. Análisis meteorología&alarmas NAMPER



39 Visualización de análisis gráfico desde Pentaho Report

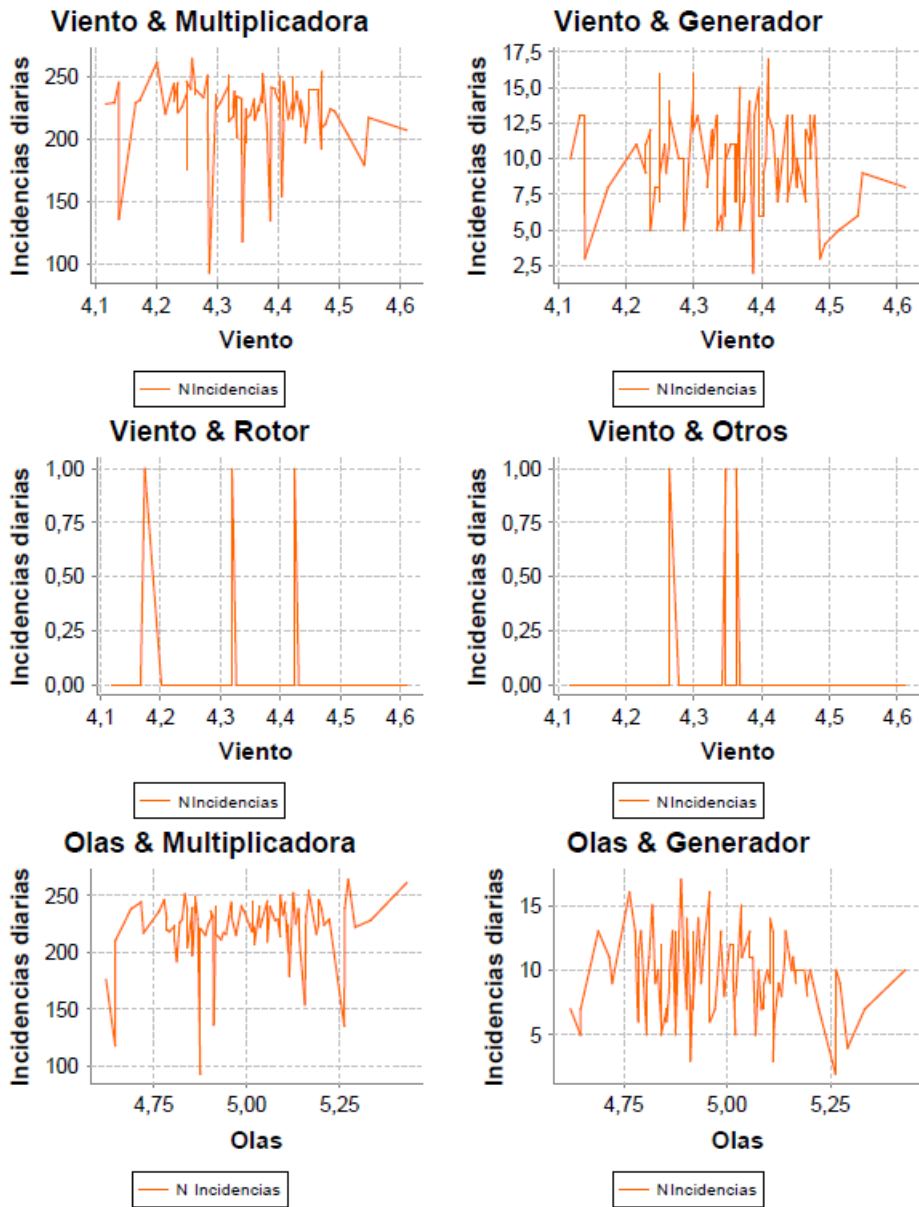
## 2.6.4.1. Análisis meteorología&alarmas NAMPER



40 Visualización de análisis gráfico desde Pentaho Report

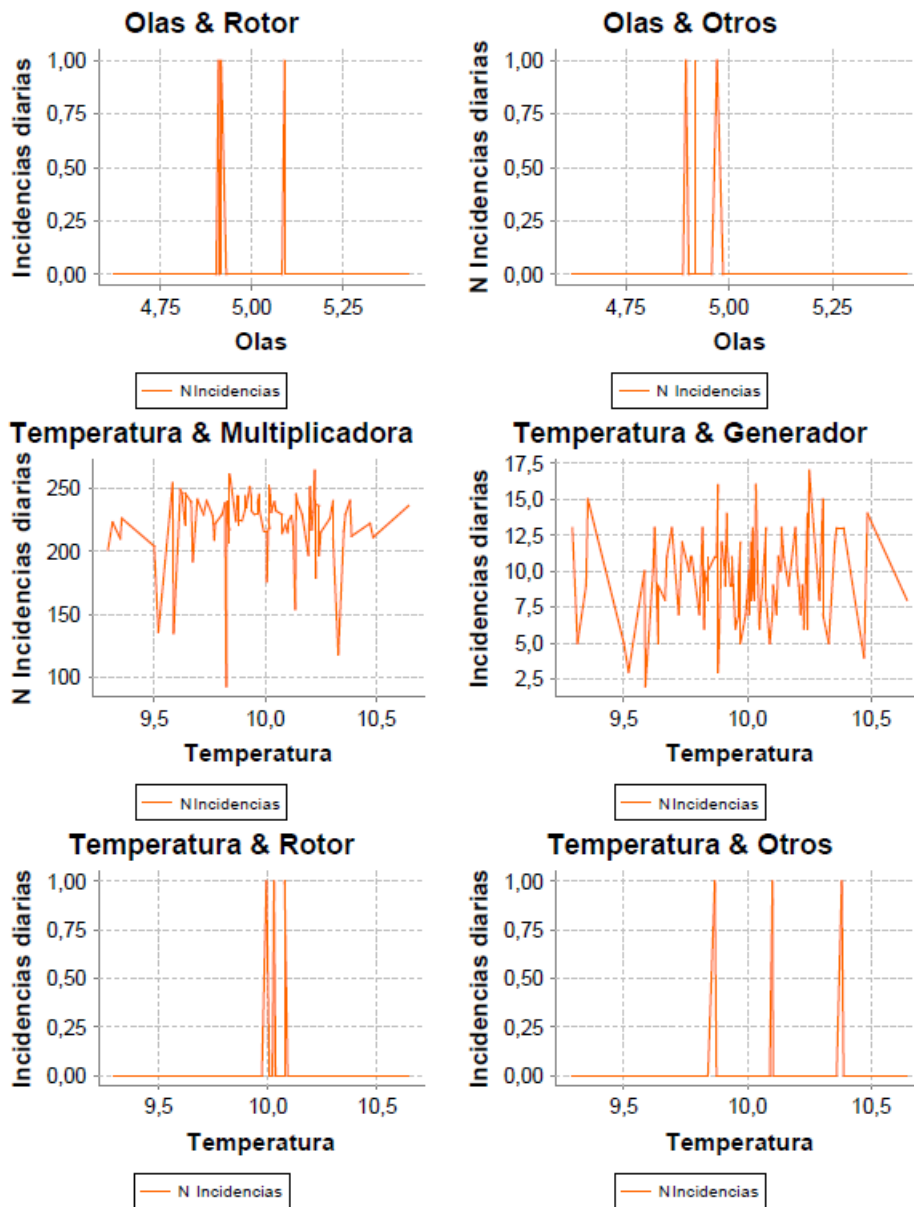


## 2.6.4.2. Análisis meteorología&alarmas RIAS BAIXAS



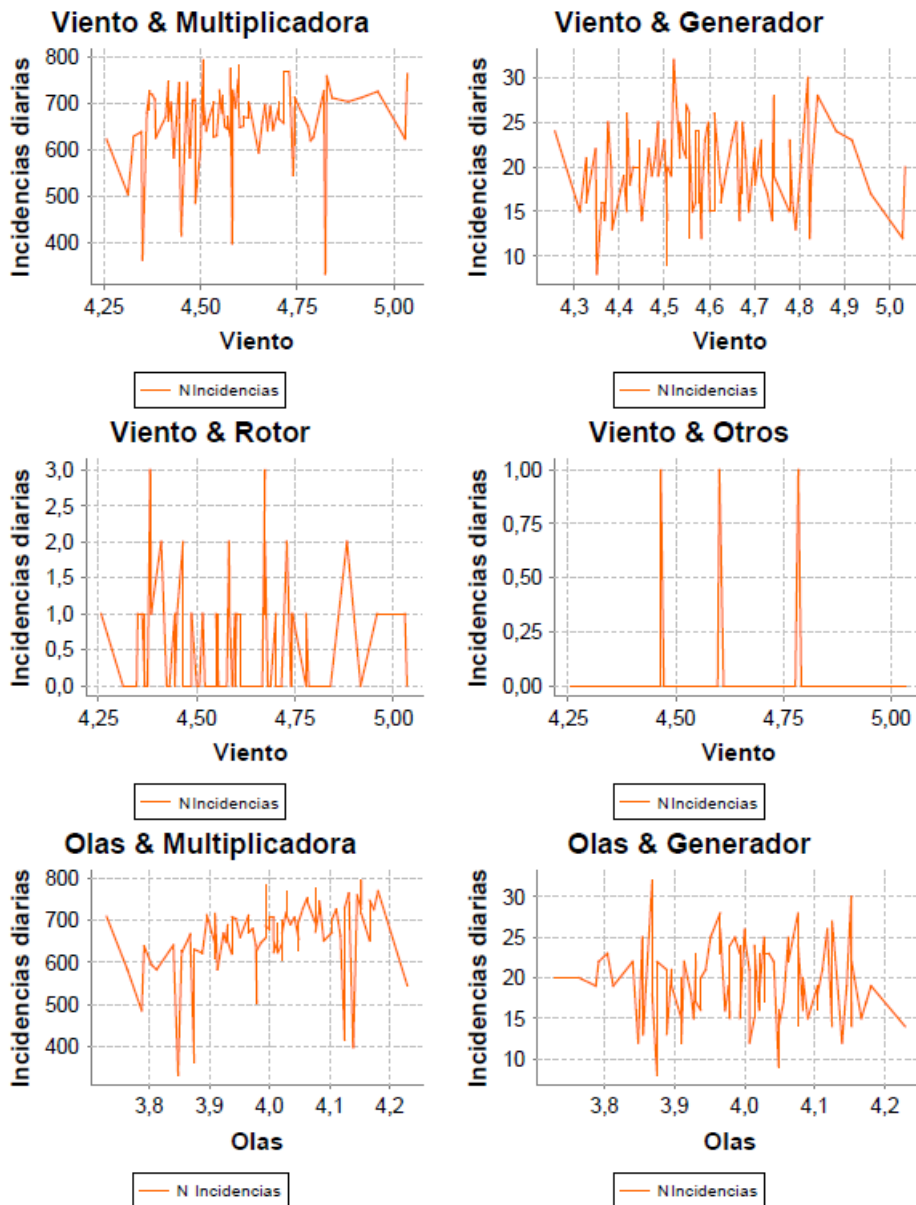
41 Visualización de análisis gráfico desde Pentaho Report

## 2.6.4.2. Análisis meteorología&alarmas RIAS BAIXAS



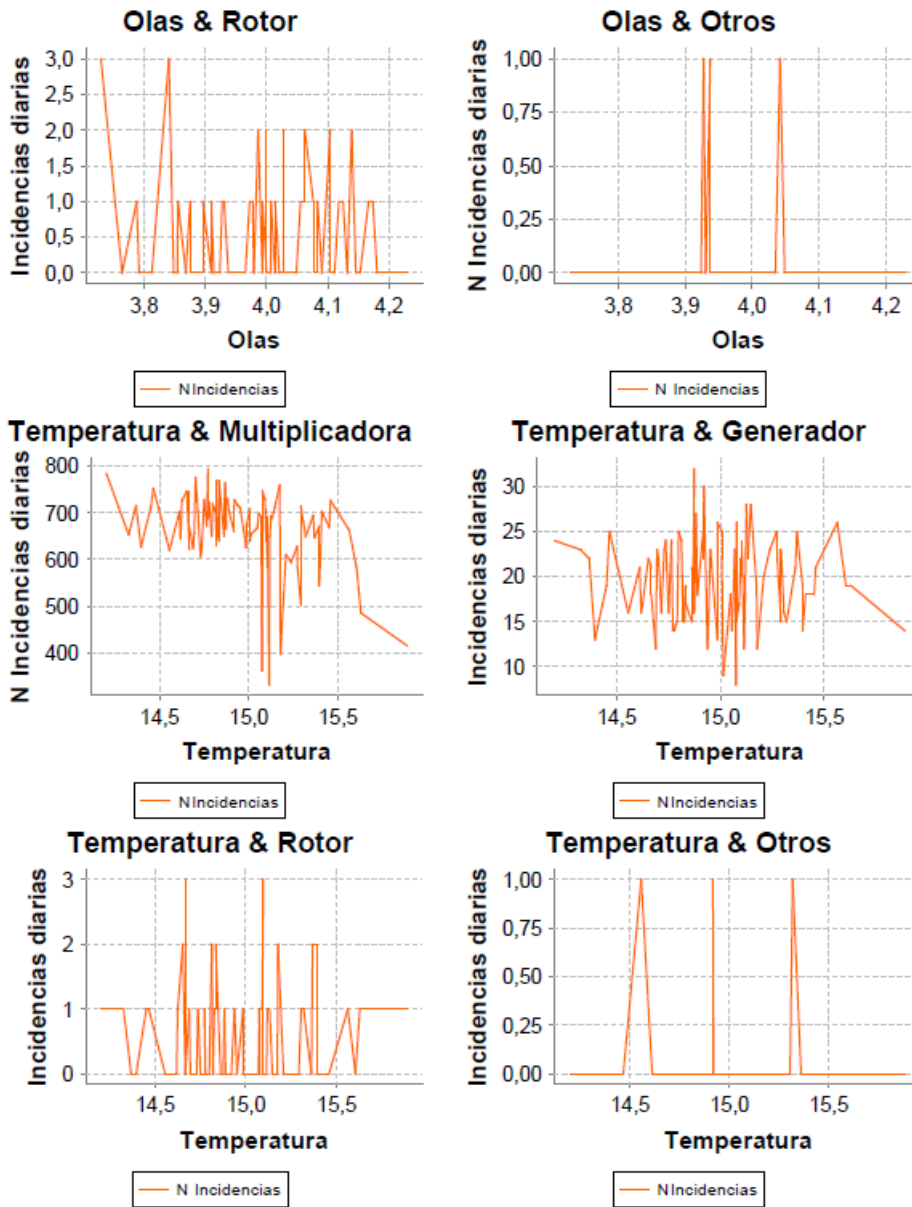
42 Visualización de análisis gráfico desde Pentaho Report

### 2.6.4.3. Análisis meteorología&alarmas GUARACHICO



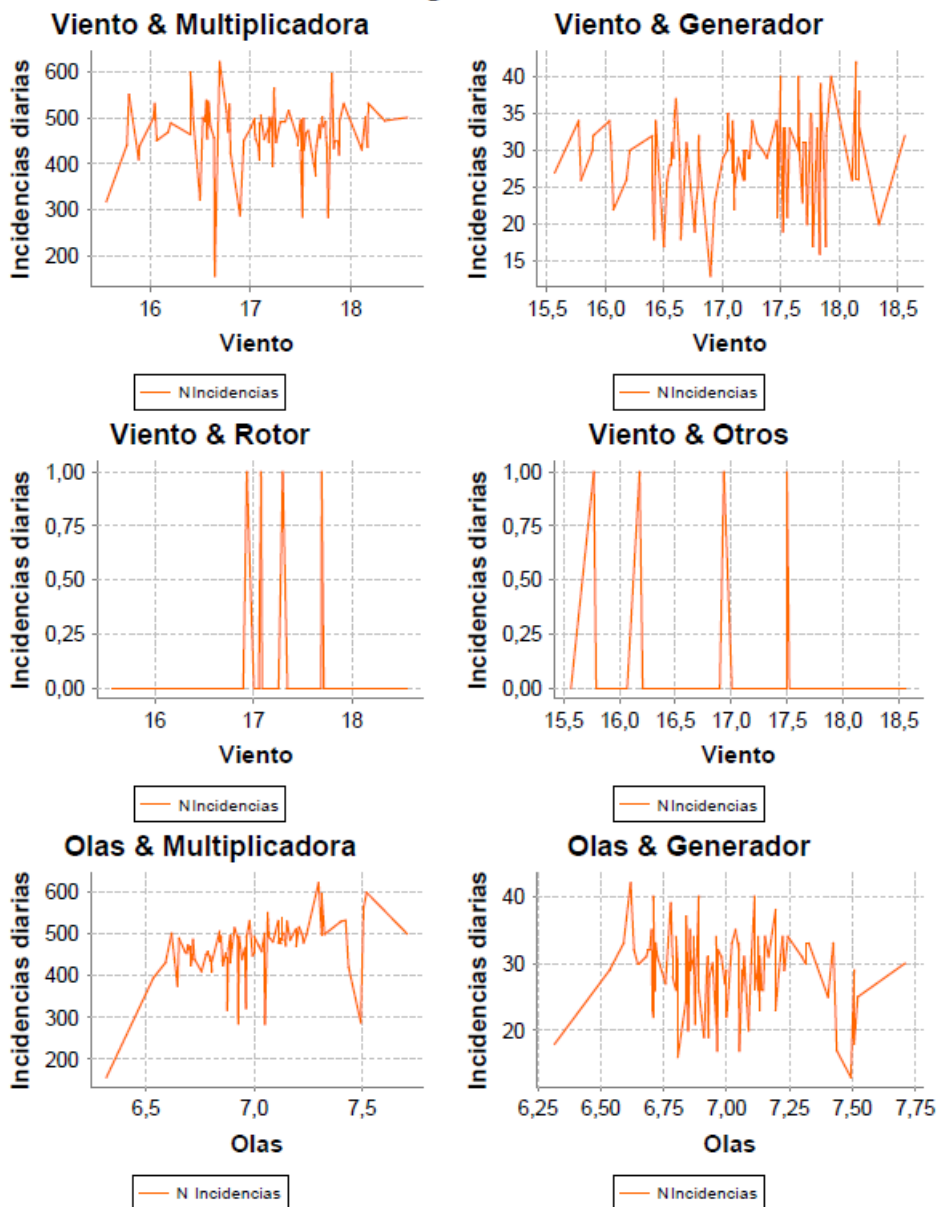
43 Visualización de análisis gráfico desde Pentaho Report

### 2.6.4.3. Análisis meteorología&alarmas GUARACHICO



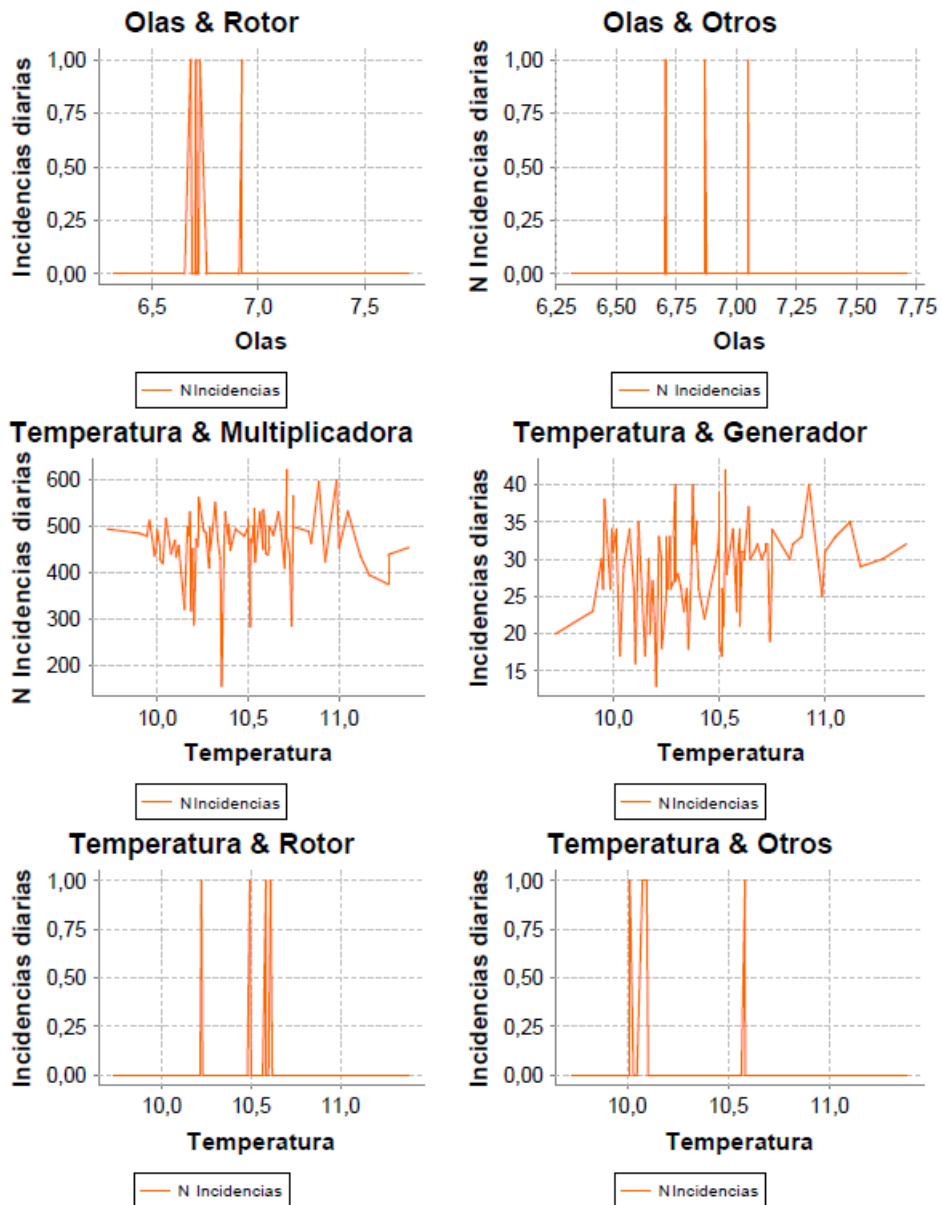
44 Visualización de análisis gráfico desde Pentaho Report

## 2.6.4.4. Análisis meteorología&alarmas KIRSKEN



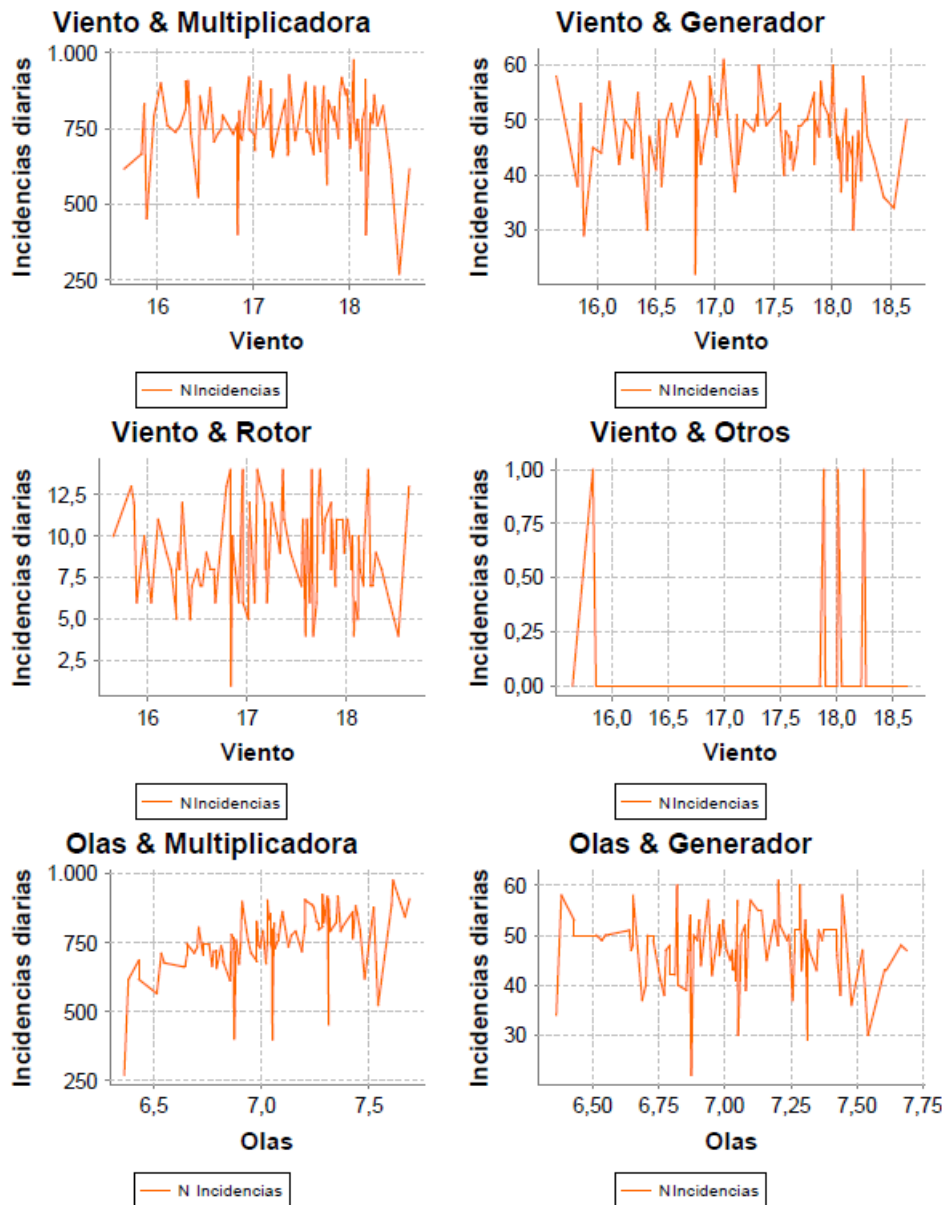
45 Visualización de análisis gráfico desde Pentaho Report

## 2.6.4.4. Análisis meteorología&alarmas KIRSKEN



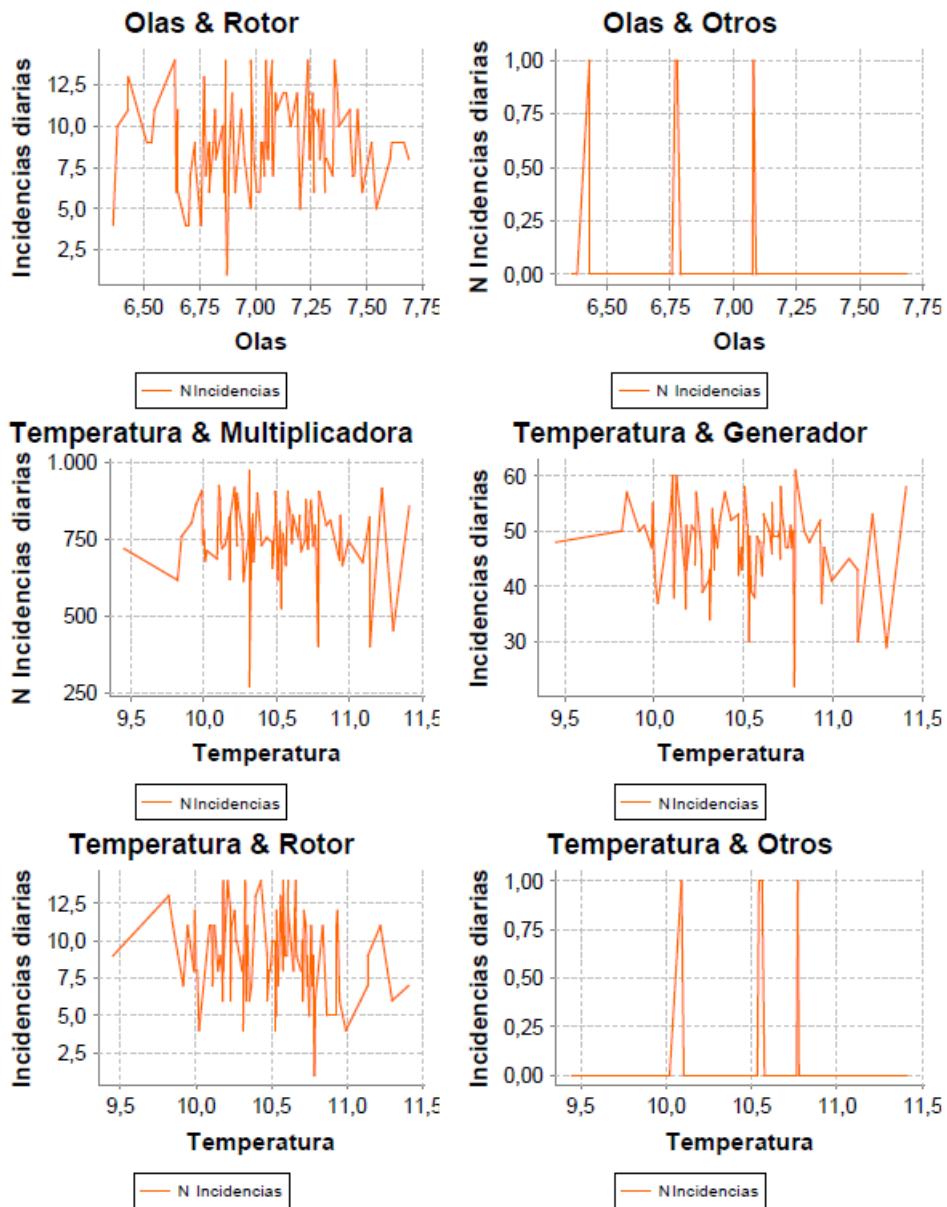
46 Visualización de análisis gráfico desde Pentaho Report

### 2.6.4.5. Análisis meteorología&alarmas POLVARS



47 Visualización de análisis gráfico desde Pentaho Report

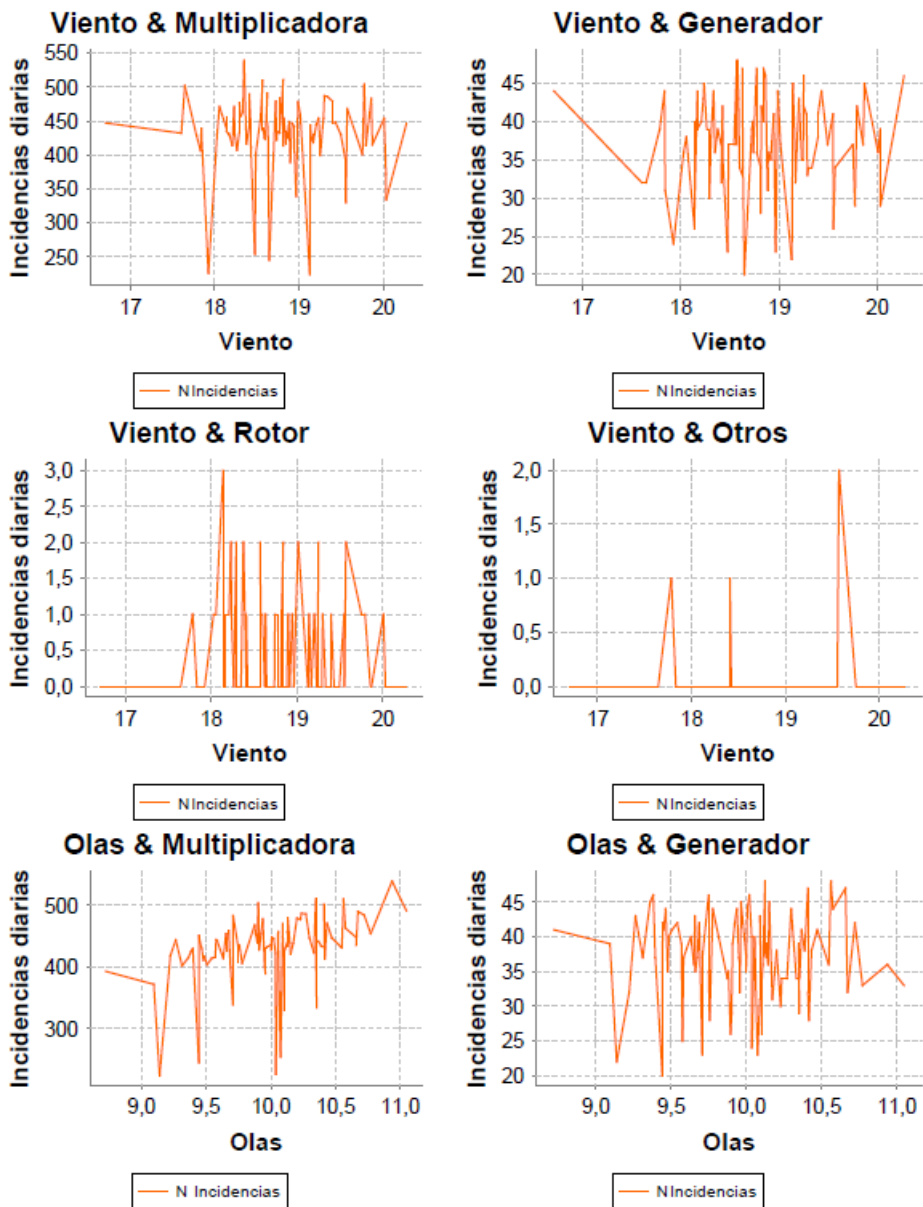
## 2.6.4.5. Análisis meteorología&alarmas POLVARS



48 Visualización de análisis gráfico desde Pentaho Report

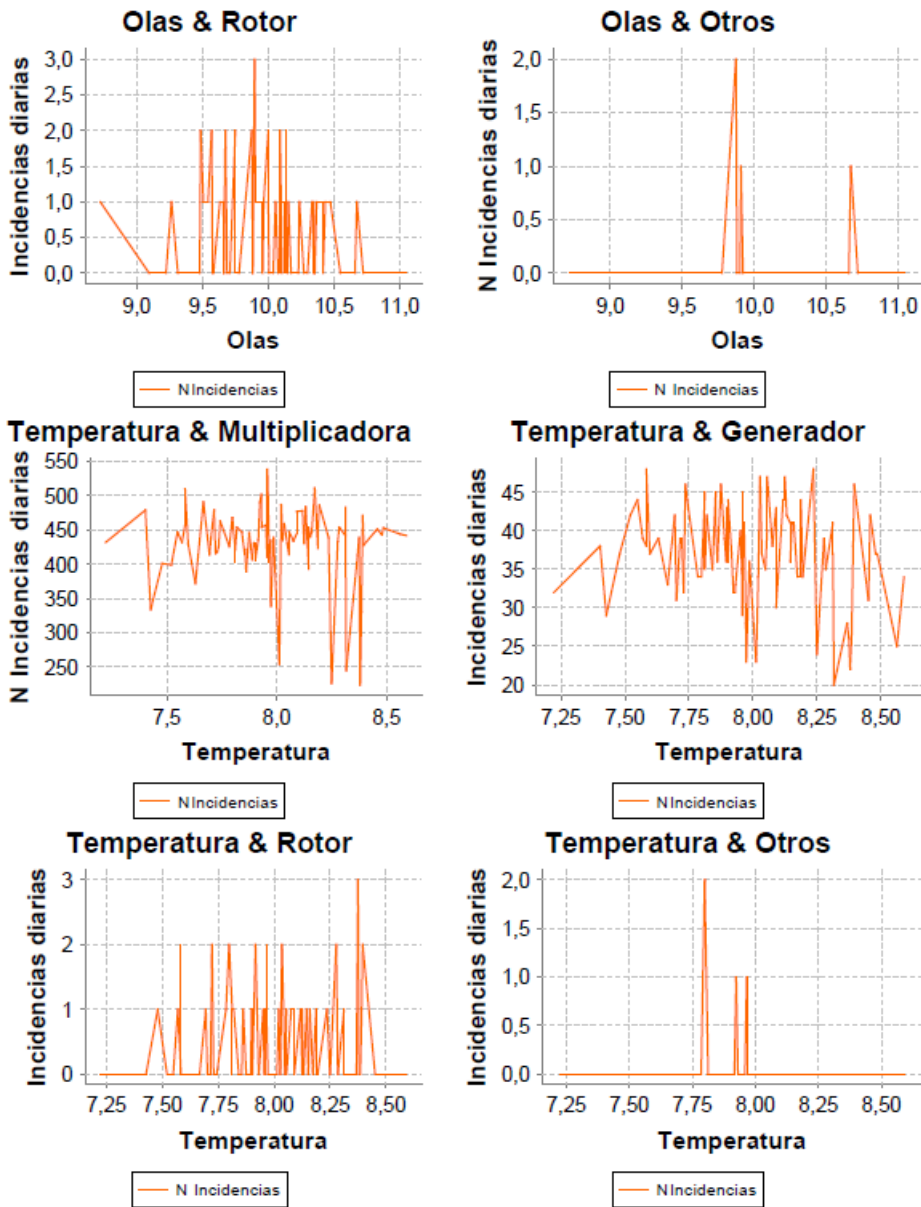


## 2.6.4.6. Análisis meteorología&alarmas COUNSCOT



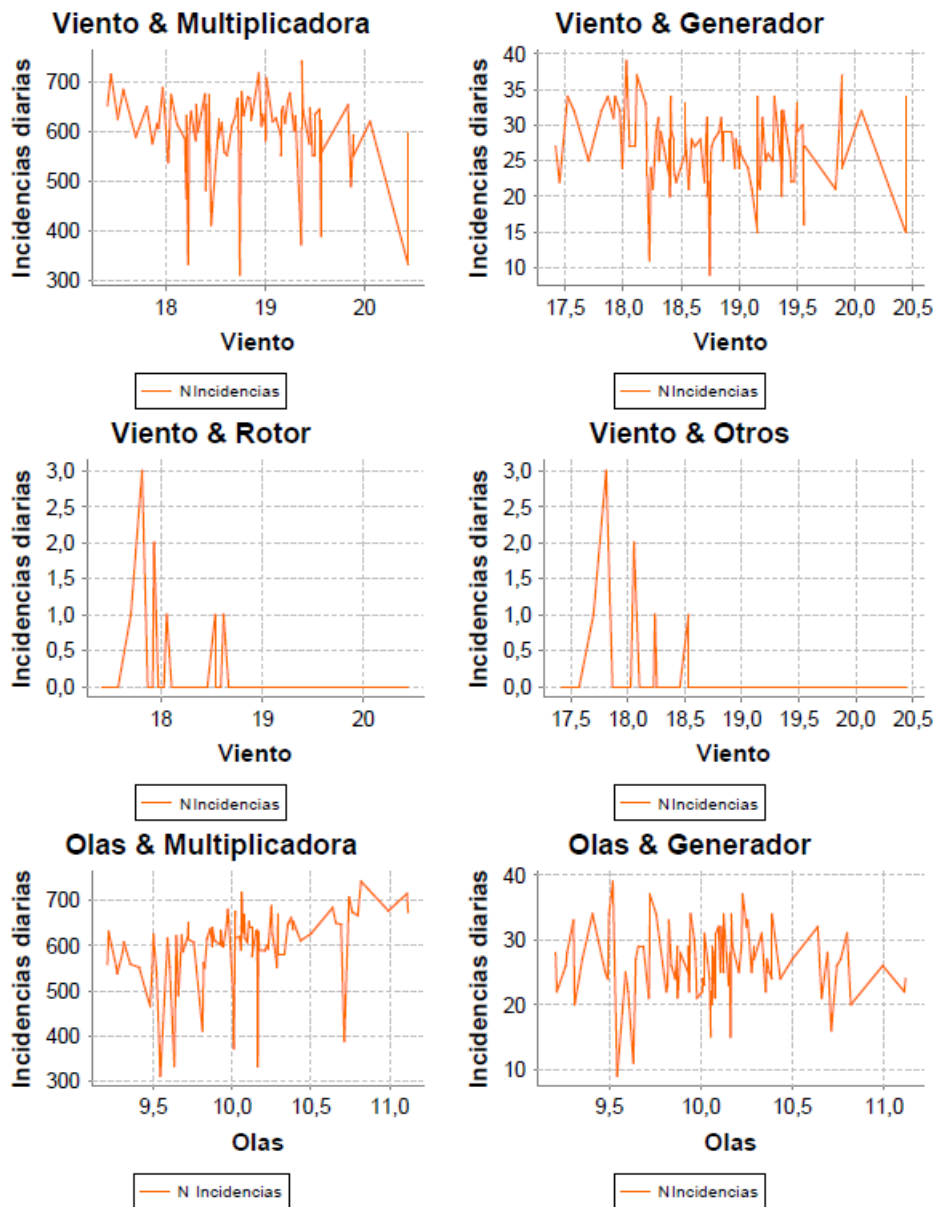
49 Visualización de análisis gráfico desde Pentaho Report

## 2.6.4.6. Análisis meteorología&alarmas COUNSCOT



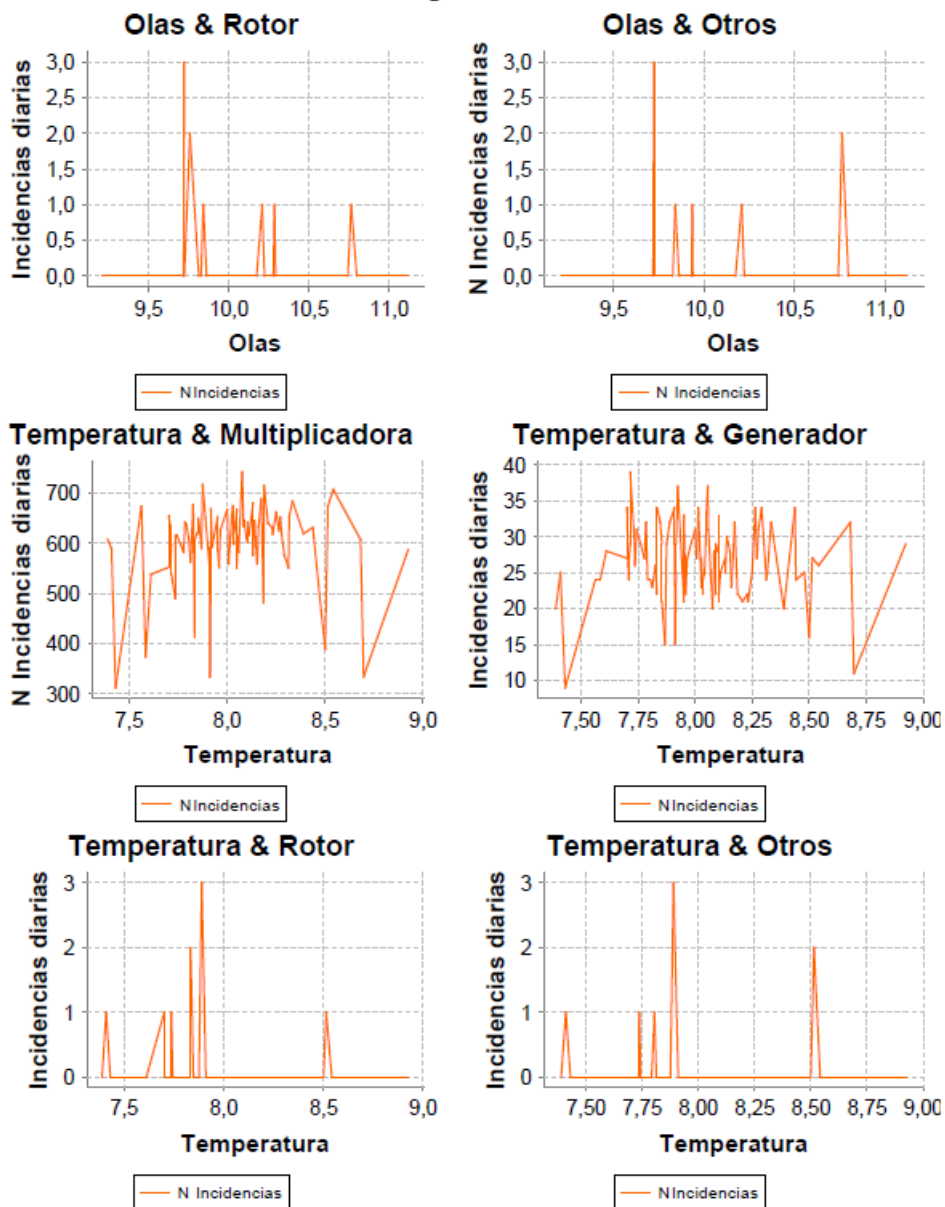
50 Visualización de análisis gráfico desde Pentaho Report

## 2.6.4.7. Análisis meteorología&alarmas GREENBLUE



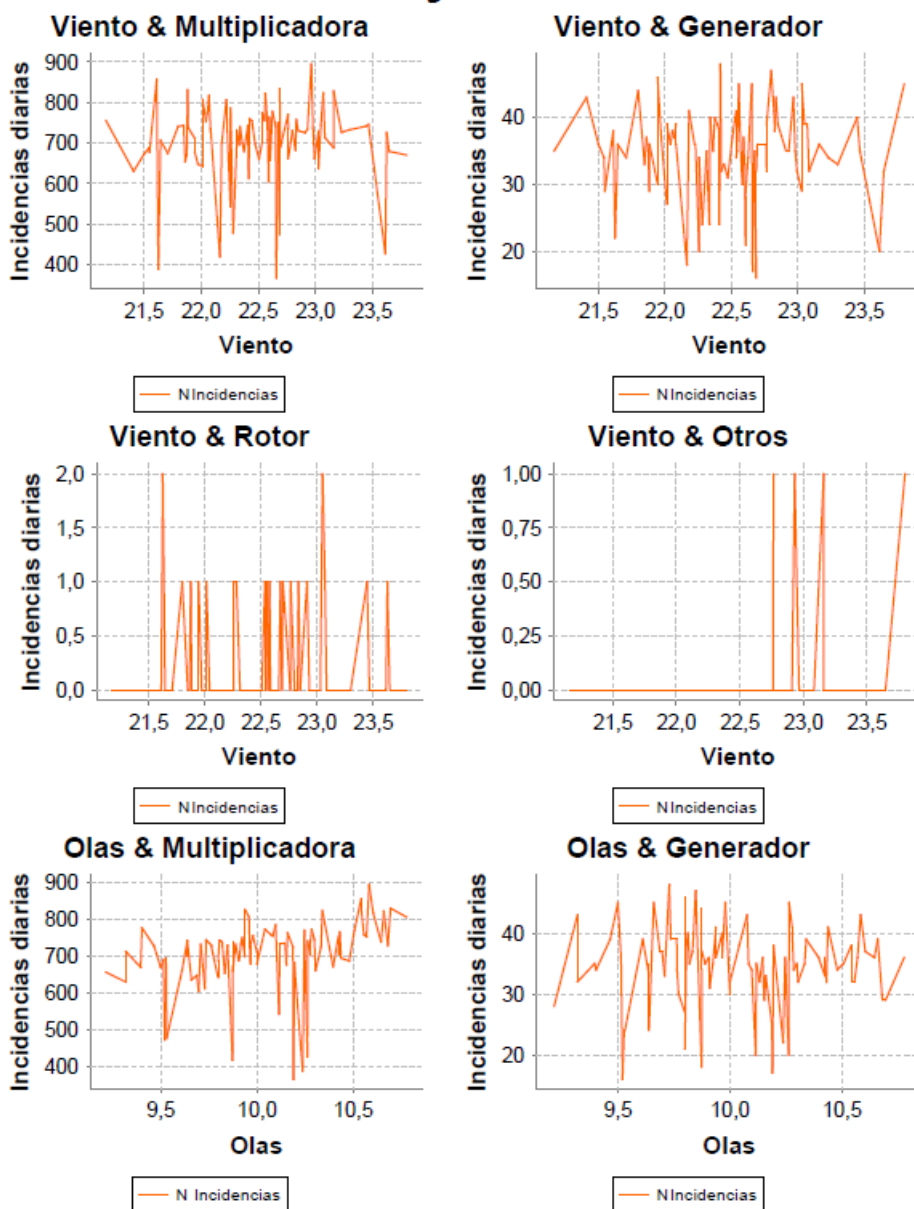
51 Visualización de análisis gráfico desde Pentaho Report

## 2.6.4.7. Análisis meteorología&alarmas GREENBLUE



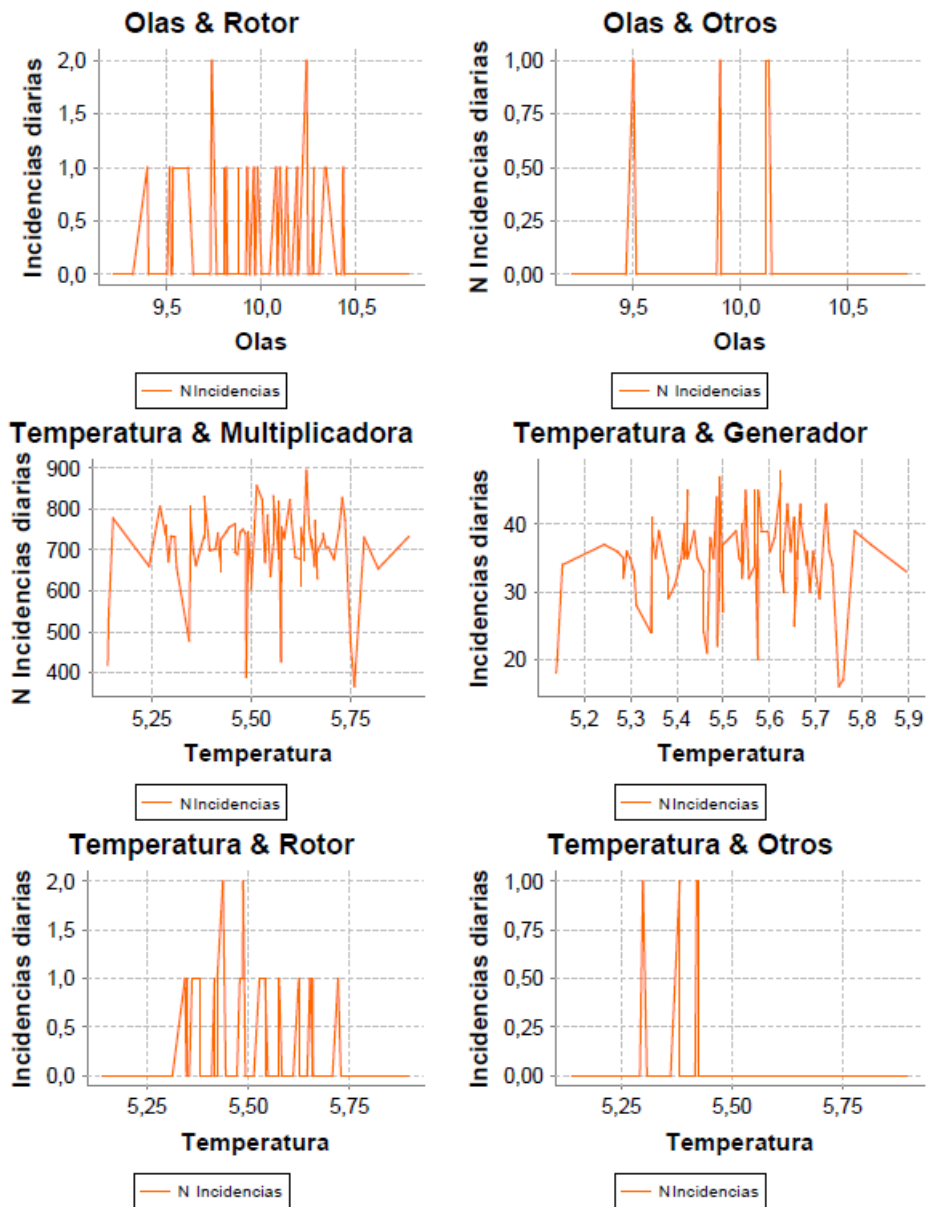
52 Visualización de análisis gráfico desde Pentaho Report

## 2.6.4.8. Análisis meteorología&alarmas NORTHENCAP



53 Visualización de análisis gráfico desde Pentaho Report

## 2.6.4.8. Análisis meteorología&alarmas NORTHENCAP



54 Visualización de análisis gráfico desde Pentaho Report

## 2.6.5. Análisis de las empresas de mantenimiento

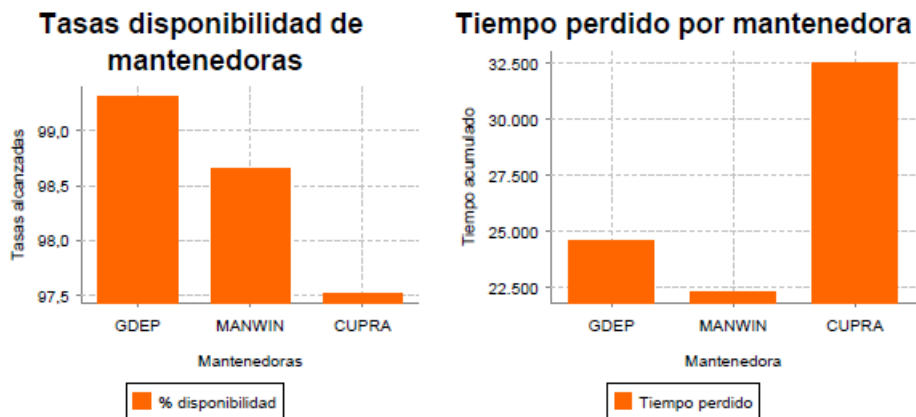
Mediante el artefacto de reporting Análisis\_Mantenedoras.prpt, se da respuesta a los resultados obtenidos por cada una de las empresas mantenedoras.

GDEP es la empresa mantenedora que tiene los ratios más satisfactorios.

### 2.6.5. Análisis de empresas mantenedoras

*El presente informe ordena de mejor a peor los resultados acumulados de las diferentes mantenedoras de todos los parques, tomando como elemento de ordenación la media de su tasa de disponibilidad e informando de forma complementaria la suma de tiempos de inproductividad alcanzado entre todos los parques a los que cada empresa da servicio.*

Empresa mantenedora	Tasa media de disponibilidad	Tiempo perdido acumulado
GDEP	99.30849202573137	24,635
MANWIN	98.65245650270393	22,340
CUPRA	97.52592438567717	32,546



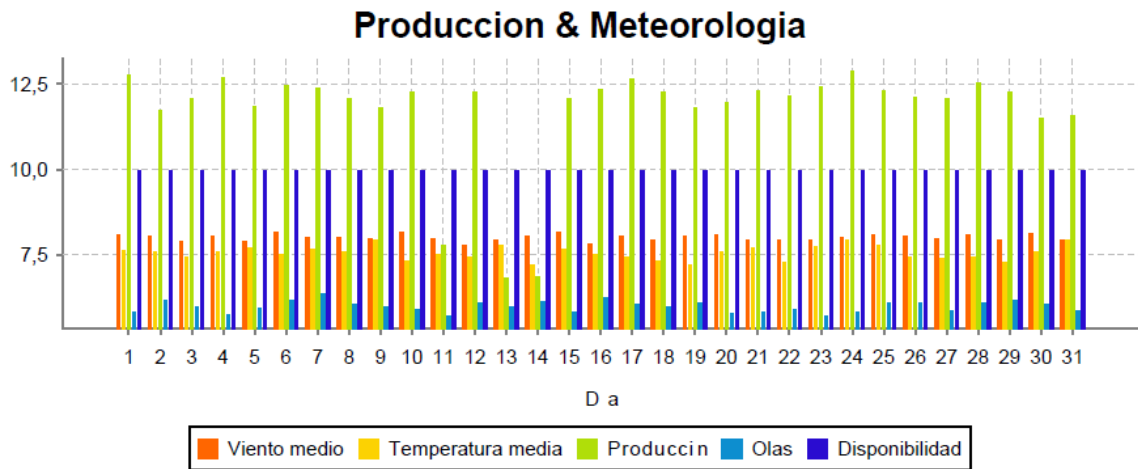
## 2.6.6. Relación entre las variables productivas y meteorológicas

Finalmente y obtenida en este punto una visión general de la información residente en el data warehouse mediante los recursos OLAP con los que se ha ido obteniendo una primera aproximación y los correspondientes informes con los que concretar las conclusiones, recurriremos aquí a un último informe Pentaho, que basado en diagramas de barras, proporciona las siguientes conclusiones sobre el conjunto de la actividad analizada, a nivel individualizado por parque y con una granularidad diaria, restringiendo al mes de Marzo el periodo de análisis:

- Hay una relación directa entre viento y producción. Si el viento cae, en igual proporción lo hace la producción.
- En la medida en la que existe una relación entre el viento y las restantes variable meteorológicas, existe entre estas y la producción, pero dicha relación no se puede considerar como una relación causa – efecto entre estas y la producción.
- En todo caso, la relación residual que puede existir entre las variables meteorológicas distintas al viento y la producción, se da por la relación que estas pueden tener con la generación de averías y por consiguiente con el impacto de ese hecho en la tasa de disponibilidad, si bien los en gráficos, no se manifiesta esa relación entre disponibilidad y esas variables. Para poner esta de manifiesto, habría que disponer gráficos de mayor escala centrados exclusivamente en esa relación.
- Adicionalmente, se observan días muy concretos, en los que la caída de la producción no se justifica con la caída del viento ni con la de la disponibilidad. Habría que profundizar en esos casos (días) la evolución de las incidencias, que en caso de confirmarse, pondría en discusión la corrección del dato de tasa de disponibilidad.

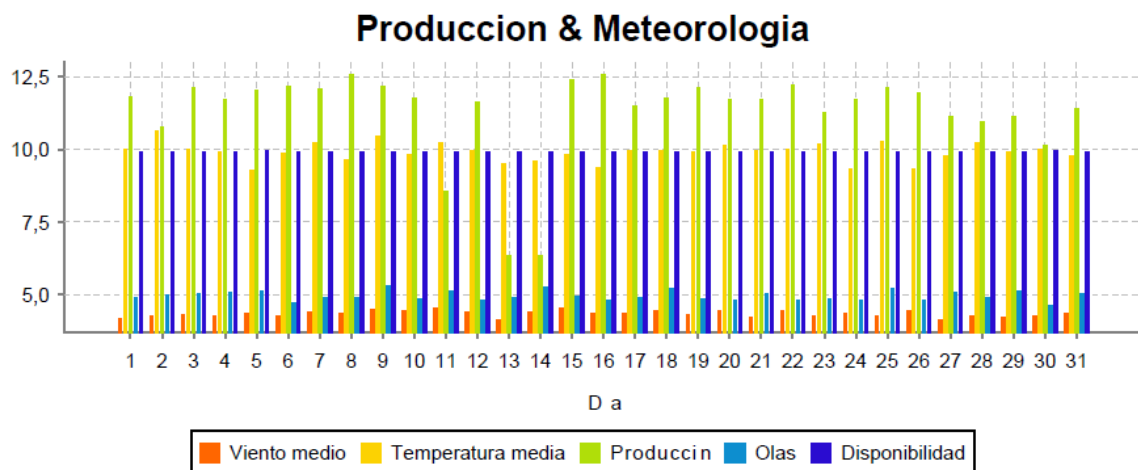


### 2.6.6.1. Relación entre las variables productivas y meteorológicas NAMPER



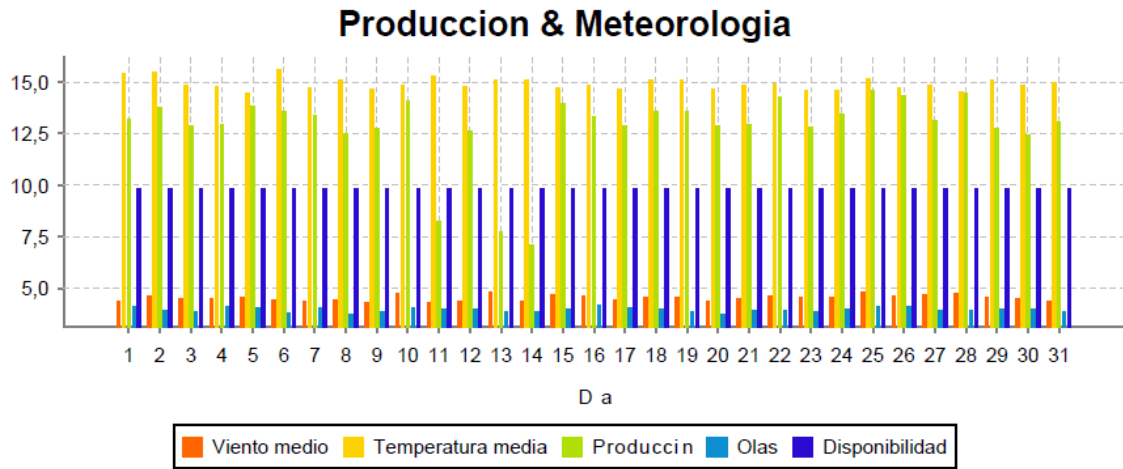
56 Visualización de análisis gráfico desde Pentaho Report

### 2.6.6.2. Relación entre las variables productivas y meteorológicas RIAS BAIXAS



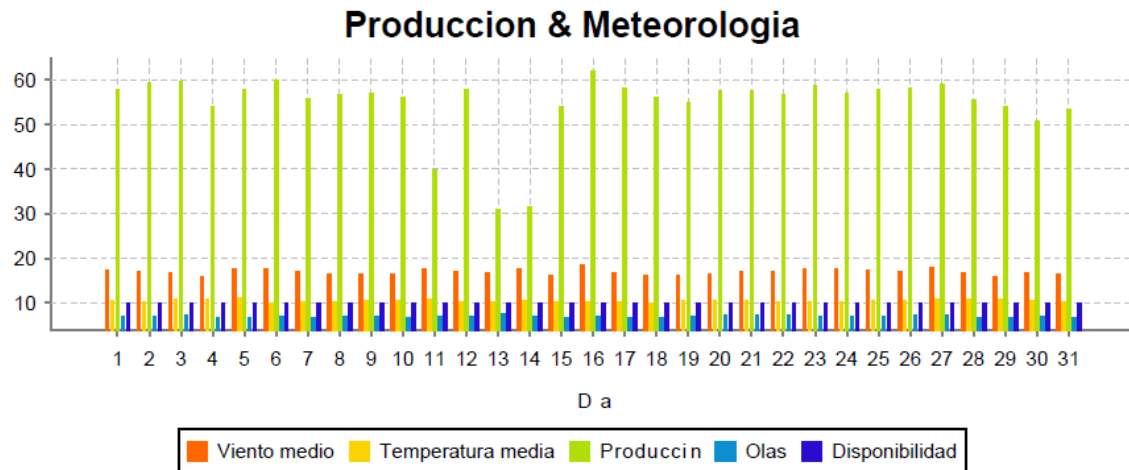
57 Visualización de análisis gráfico desde Pentaho Report

### 2.6.6.3. Relación entre las variables productivas y meteorológicas GUARACHICO



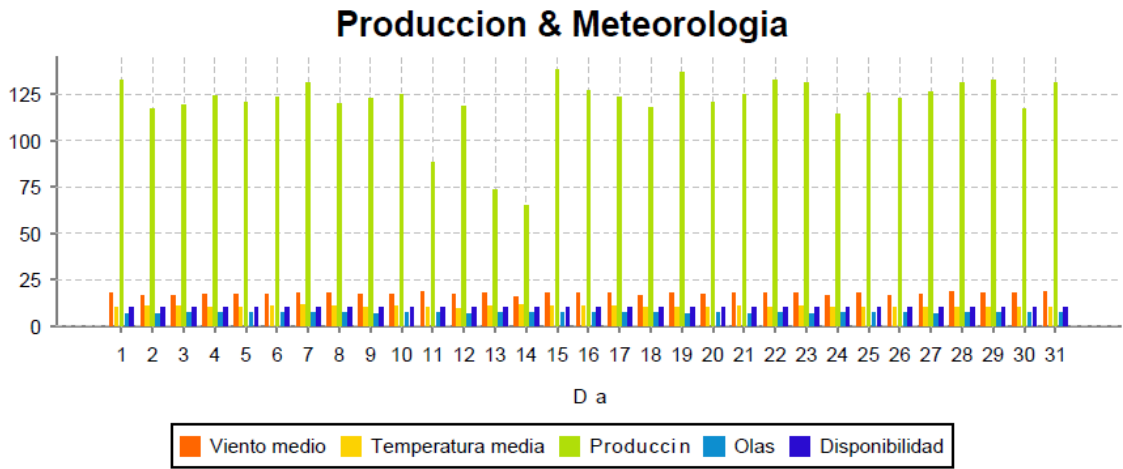
58 Visualización de análisis gráfico desde Pentaho Report

### 2.6.6.4. Relación entre las variables productivas y meteorológicas KIRSKEN



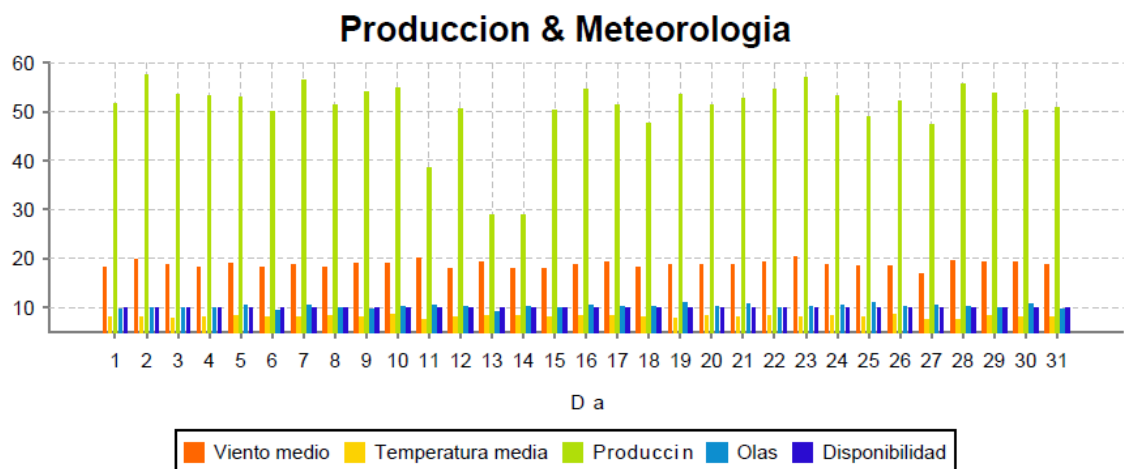
59 Visualización de análisis gráfico desde Pentaho Report

### 2.6.6.5. Relación entre las variables productivas y meteorológicas POLVARS



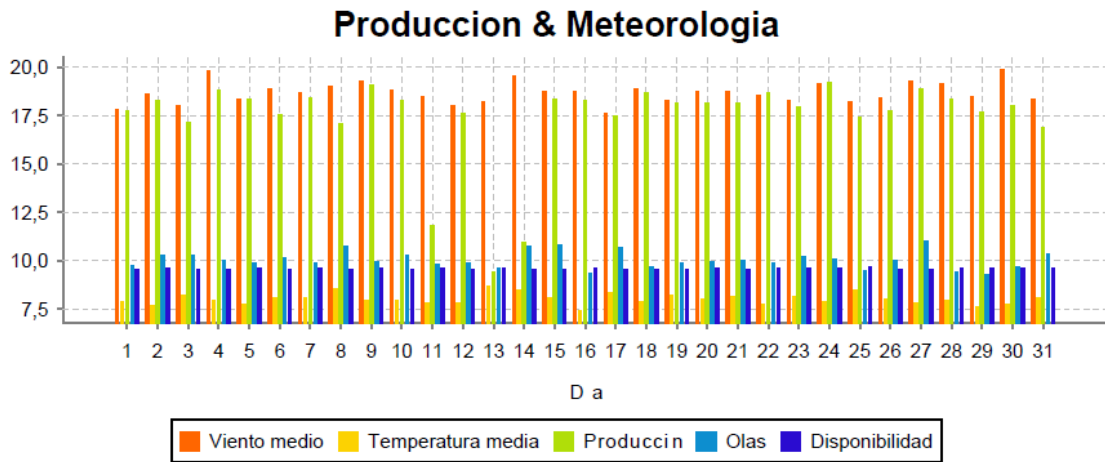
60 Visualización de análisis gráfico desde Pentaho Report

### 2.6.6.6. Relación entre las variables productivas y meteorológicas COUNSCOT



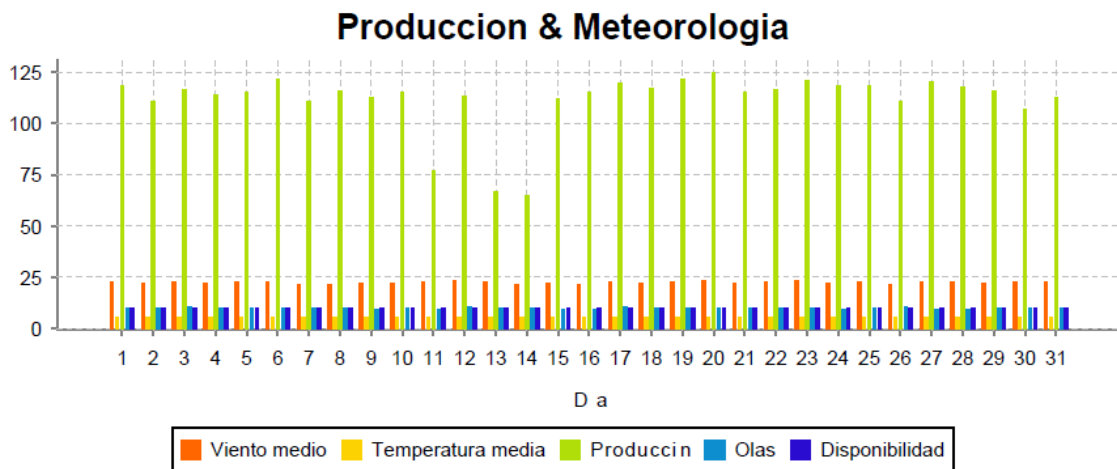
61 Visualización de análisis gráfico desde Pentaho Report

### 2.6.6.7. Relación entre las variables productivas y meteorológicas GREENBLUE



62 Visualización de análisis gráfico desde Pentaho Report

### 2.6.6.8. Relación entre las variables productivas y meteorológicas NORTHENCAP



63 Visualización de análisis gráfico desde Pentaho Report

### 3. Conclusiones

El resultado del proyecto, junto al software de código abierto utilizado, conforma una primera versión de una framework que puede resultar bastante útil para la ejecución de proyectos de business intelligent.

La propia memoria como parte del framework para proporcionar el enfoque metodológico, junto con los detalles sobre el uso general de las aplicaciones utilizadas, ausentes en la documentación de las mismas, sirve de elemento de unión del conjunto.

El software de Pentaho, ofrece en muchos momentos una imagen de conjunto de aplicaciones sin relación. Aunque al final la integración de la información acaba estando resuelta, en la medida que los metadatos de unos y otros son aprovechables, existiendo alguna opción de publicación desde aplicaciones como Pentaho Report o Workbench en el servidor.

Por otra parte y en algunos casos, la usabilidad de las aplicaciones no está integrada, debiéndose aprender formas distintas de hacer las cosas, según la aplicación.

Los procesos de instalación, recogen también problemas. Son menores, tales como ausencia de drivers básicos de JDBC, pero acaba consumiendo un tiempo adicional, no habiendo una documentación de conjunto que permita una visión global.

El aspecto positivo, de cara al contexto metodológico de un proyecto de business intelligent, es que cada parte del mismo, queda claramente separada del resto, lo que proporciona una modularidad mayor a la que se acaba configurando con el uso de ciertas suites comerciales. Esto permite libertad, a la hora de elegir los diferentes componentes software con los que se puede acometer el proyecto. Incluso este proyecto, podría haber incorporado más de un software libre, si bien hubiera demandado indiscutiblemente un mayor número de horas de trabajo.

Dentro de las diferentes aplicaciones de Pentaho que se han utilizado, se invirtió tiempo tratando de usar la aplicación Saiku, para añadir un elemento

más en el análisis de información, como complemento a Report Designer. Los problemas surgidos en la configuración de esta aplicación y el hecho de que los objetivos del proyecto estaban satisfechos, hizo desistir de incorporar algún recurso de análisis con esta aplicación en lugar de hacerlo con Report, así como lo ajustado de la planificación, desaconsejaron esa opción.

En cualquier caso Pentaho, es una plataforma en evolución, en la que se advierte el paso hacia convertirse en una plataforma de publicación, combinando así sus capacidades funcionales originales con una potencia y facilidad nueva en cuanto a la gestión de los contenidos obtenidos en los análisis.

Por otra parte, han sido varios los bugs que en el curso del trabajo realizado se han localizado en las aplicaciones, especialmente en Workbench. Dichos fallos no han quedado documentados, pues finalmente ese capítulo del proyecto, el de la documentación, ha estado muy justo de recursos de tiempo. A pesar de que de acuerdo al plan del proyecto se estableció como metodología la confección de la memoria de forma totalmente paralela, resultaría un aspecto de mejorar, lograr incorporar a la misma, aspectos y reflexiones que finalmente han quedado fuera por falta de tiempo.

El enfoque metodológico de la planificación del proyecto en cascada, se ha encontrado en la realidad con diferentes procesos iterativos, planteando la idoneidad de las conocidas como metodologías ágiles, como una buena alternativa metodológica para proyectos BI.

Por otra parte, dichas iteraciones, han puesto a prueba, la capacidad del conjunto para maniobrar y rehacer diferentes partes del proyecto.

En el plano del análisis del negocio tratado, se ha obtenido una buena base, que podría proporcionar muy buenos resultados, con un trabajo de estudio y cruce de información entre las diferentes líneas trabajadas (marcadas por las cuestiones planteadas que sirvieron de motivación del proyecto). En este aspecto con unas pocas jornadas más de trabajo, se podrían definir las decisiones estratégicas del negocio con bastantes garantías.

## 4. Glosario

<b>Aerogenerador</b>	Término del negocio analizado. Equipo generador de electricidad a partir del viento existente.
<b>Artefacto</b>	Producto individual de un proyecto de software, que puede ser un programa o un fichero.
<b>Business Intelligence</b>	Anglicismo con el que es identificado la disciplina encargada del análisis de la información y su asistencia en la toma de decisiones.
<b>Data warehouse</b>	Anglicismo con el que se identifica el repositorio de datos histórico, generalmente diseñado bajo un enfoque dimensional, orientado a la consulta y el análisis.
<b>Diseño conceptual</b>	Primera representación de un modelo de datos de la realidad.
<b>Diseño físico</b>	Detalle de las entidades de datos y sus relaciones implementadas sobre un sistema gestor de base de datos específico.
<b>Diseño lógico</b>	Evolución de un diseño conceptual concretando detalles y aplicando criterios técnicos de diseño de bases de datos.
<b>ETL</b>	Acrónimo de Extract, Transform, Load o lectura , extracción y transformación de datos, usado en el contexto de Business Intelligence para las tareas de carga de datos del data warehouse con datos procedentes de diversas fuentes.
<b>Molino</b>	Término del negocio analizado. Término sinónimo de aerogenerador, utilizado en el curso de la memoria.

<b>Mondrian</b>	Motor de ejecución de cubos OLAP incluido en el conjunto de aplicaciones Pentaho.
<b>OLAP</b>	Acrónimo de On-Line Analytical Processing
<b>Parque</b>	Término del negocio analizado. Conjunto de aerogeradores o molinos en una ubicación geográfica concreta.
<b>PDI</b>	Acrónimo de Pentaho Data Integration
<b>PUC</b>	Acrónimo de Pentaho User Console
<b>Reporting</b>	Anglicismo utilizado para hacer referencia a la actividad con consulta y presentación de datos, generalmente albergados en una o varias bases de datos.
<b>Schema Workbench</b>	Aplicación incluido dentro de Pentaho, cuya funcionalidad es el diseño de cubos OLAP.



## 5. Bibliografía

- <http://community.pentaho.com/> (Octubre-Noviembre 2016)
- <https://www.spagobi.org/homepage/product/business-intelligence/> (Octubre-Noviembre 2016)
- <http://community.jaspersoft.com/> (Octubre-Noviembre 2016)
- <https://rapidminer.com/>
- <http://www.pentaho.org>
- <http://www.eclipse.org/birt/>
- <http://www.redbooks.ibm.com/redbooks/pdfs/sg242238.pdf>
- <http://web.mit.edu/profit/pdfs/slaughtera.pdf>
- Josep Curto Díaz y Jordi Conesa i Caralt , Introducción al Business Intelligence, publicado por Editorial UOC en Barcelona en Octubre de 2011.
- <http://culturacrm.com/business-intelligence/5-herramientas-open-source-de-business-intelligence/> (Octubre-Noviembre 2016)
- Josh Parenteau, Rita L. Sallam, Cindi Howson, Joao Tapadinhas, Kurt Schlegel, Thomas W. Oestreich , Magic Quadrant for Business Intelligence and Analytics Platforms, publicado el 4 Febrero de 2016 por Gartner

## 6. Anexos

- Anexo I: Diseño e implementación de procesos ETL construidos en el enfoque inicial

# ANEXO I: Diseño e implementación de procesos ETL construidos en el enfoque inicial

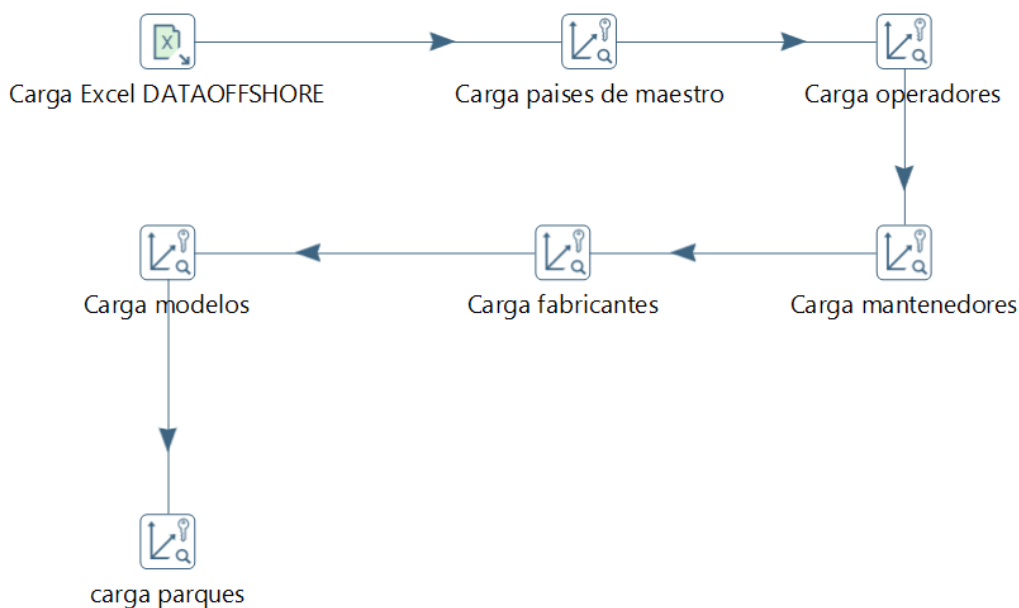
En el curso del proyecto, tras el planteamiento de un primer diseño del data warehouse, se optó por una segunda alternativa que facilitaría y simplificaría el resto del proyecto. Dicha decisión se adoptó cuando ya se contaba con los procesos ETL adecuados a ese modelo del data warehouse.

Se han querido conservar esos procesos, porque aunque desde la perspectiva del proyecto no han sido los que han generado el data warehouse con el que el mismo se ha continuado y finalizado, desde la perspectiva del conocimiento de la herramienta Pentaho PDI, responden a una casuística mayor, más compleja, y pueden resultar una ayuda de valor en el uso de la misma.

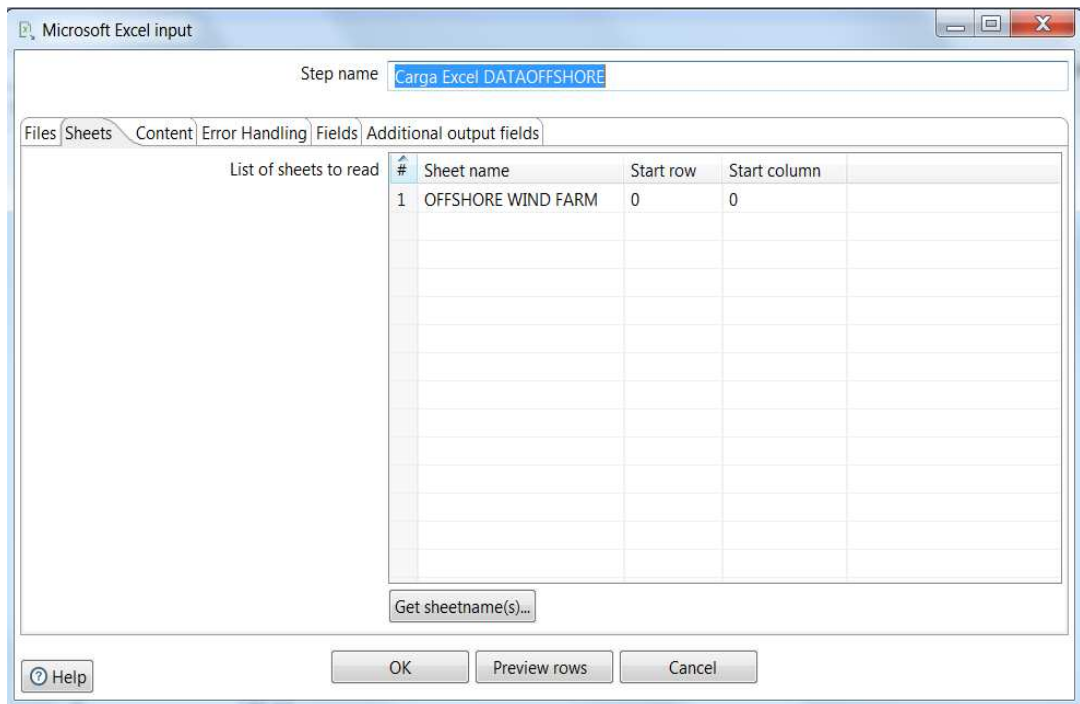
## 1.1. Construcción de artefactos ETL

**Se han diseñado dos transformaciones**, una para la carga de los datos maestros, incluidos en la hoja OFFSHORE WIND FARM de DATAOFFSHORE.xls y otra para la carga de todas las lecturas, y que comprende todos los registros del resto de hojas del fichero Excel.

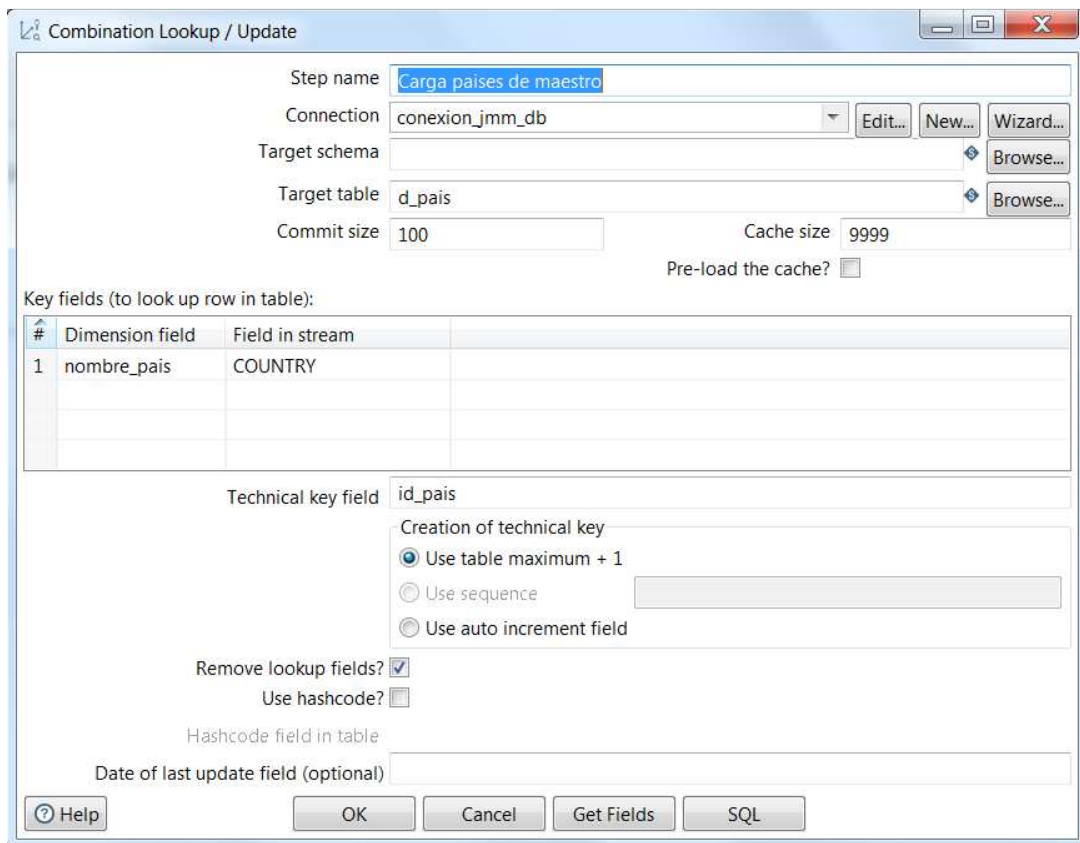
### 1.1.1. Diseño ETL para la carga de datos maestros



a. Carga Excel DATAOFFSHORE



b. Carga países de maestro



### c. Carga operadores

The dialog box 'Combination Lookup / Update' is configured for the step 'Carga operadores'. It is connected to 'conexion\_jmm\_db' and targets the 'd\_operador' table. The commit size is set to 100 and the cache size to 9999. The 'Pre-load the cache?' checkbox is unchecked. The 'Key fields (to look up row in table):' table contains one entry: '1' with 'nombre\_operador' as the dimension field and 'OPERATOR' as the field in stream. The 'Technical key field' is 'id\_operador', with 'Use table maximum + 1' selected for its creation. The 'Remove lookup fields?' checkbox is checked, and 'Use hashcode?' is unchecked. The 'Date of last update field (optional)' is empty.

#	Dimension field	Field in stream
1	nombre_operador	OPERATOR

### d. Carga modelos

The dialog box 'Combination Lookup / Update' is configured for the step 'Carga modelos'. It is connected to 'conexion2' and targets the 'd\_modelo' table. The commit size is set to 100 and the cache size to 9999. The 'Pre-load the cache?' checkbox is unchecked. The 'Key fields (to look up row in table):' table contains two entries: '1' with 'potencia' as the dimension field and 'POWER' as the field in stream, and '2' with 'modelo\_id\_fabricante' as the dimension field and 'Id\_fabricante' as the field in stream. The 'Technical key field' is 'id\_modelo', with 'Use table maximum + 1' selected for its creation. The 'Remove lookup fields?' checkbox is checked, and 'Use hashcode?' is unchecked. The 'Date of last update field (optional)' is empty.

#	Dimension field	Field in stream
1	potencia	POWER
2	modelo_id_fabricante	Id_fabricante

e. Carga fabricantes

The screenshot shows the 'Combination Lookup / Update' dialog box for the step 'Carga fabricantes'. The configuration is as follows:

- Step name: Carga fabricantes
- Connection: conexion2
- Target schema: (empty)
- Target table: d\_fabricante
- Commit size: 100
- Cache size: 9999
- Pre-load the cache?

Key fields (to look up row in table):

#	Dimension field	Field in stream
1	Nombre	AEROGEN

Technical key field: Id\_fabricante

Creation of technical key:

- Use table maximum + 1
- Use sequence
- Use auto increment field

Remove lookup fields?

Use hashcode?

Hashcode field in table: (empty)

Date of last update field (optional): (empty)

Buttons: Help, OK, Cancel, Get Fields, SQL

f. Carga mantenedores

The screenshot shows the 'Combination Lookup / Update' dialog box for the step 'Carga mantenedores'. The configuration is as follows:

- Step name: Carga mantenedores
- Connection: conexion\_jmm\_db
- Target schema: (empty)
- Target table: d\_mantenedora
- Commit size: 100
- Cache size: 9999
- Pre-load the cache?

Key fields (to look up row in table):

#	Dimension field	Field in stream
1	nombre_mantenedora	MAINTAINER

Technical key field: id\_mantenedora

Creation of technical key:

- Use table maximum + 1
- Use sequence
- Use auto increment field

Remove lookup fields?

Use hashcode?

Hashcode field in table: (empty)

Date of last update field (optional): (empty)

Buttons: Help, OK, Cancel, Get Fields, SQL

## g. Carga parques

Combination Lookup / Update

Step name:

Connection:

Target schema:

Target table:

Commit size:  Cache size:

Pre-load the cache?

Key fields (to look up row in table):

#	Dimension field	Field in stream
1	coordenadax	X
2	coordenaday	Y
3	nombre_parque	NAME
4	numero_aerogeneradores	NUMBER
5	parque_id_pais	id_pais
6	parque_id_operador	id_operador
7	parque_id_mantenedor	id_mantenedora
8	parque_ide_modelo	id_modelo

Technical key field:

Creation of technical key

Use table maximum + 1

Use sequence

Use auto increment field

Remove lookup fields?

Use hashcode?

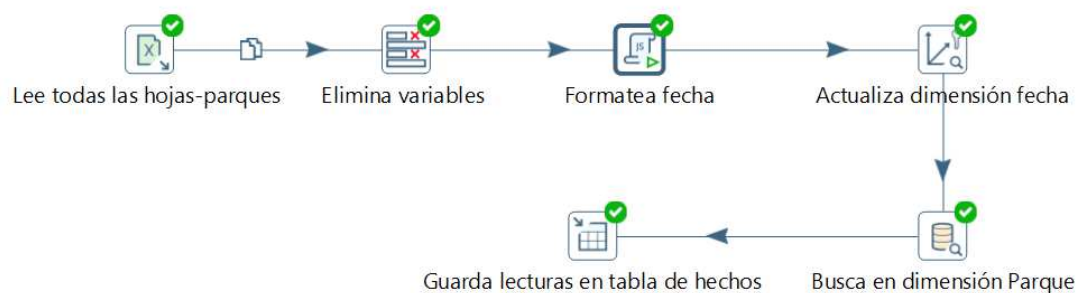
Hashcode field in table:

Date of last update field (optional):

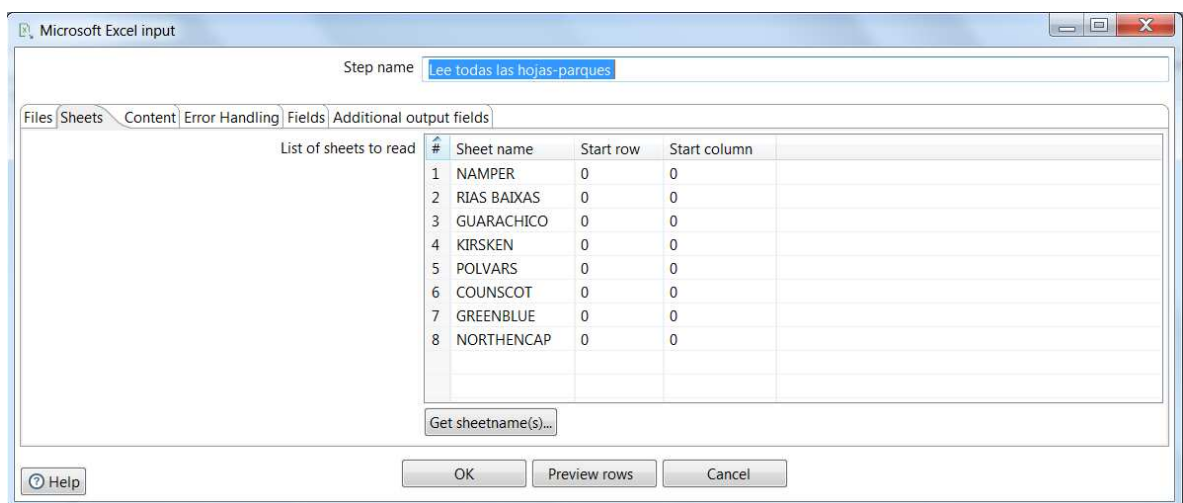
### 1.1.2. Implementación ETL para la carga de datos maestros

El fichero de formato XML, `Carga_datos_maestros_v10.ktr`, entregable del proyecto, alberga los metadatos resultantes de la implementación. Dicha información es interpretable por el motor PDI en la ejecución de las tareas, así como por la aplicación Spoon, capaz de construir el diagrama correspondiente con los objetos parametrizados, a partir de dicho fichero.

### 1.1.3. Diseño ETL para la carga de lecturas de todos los parques

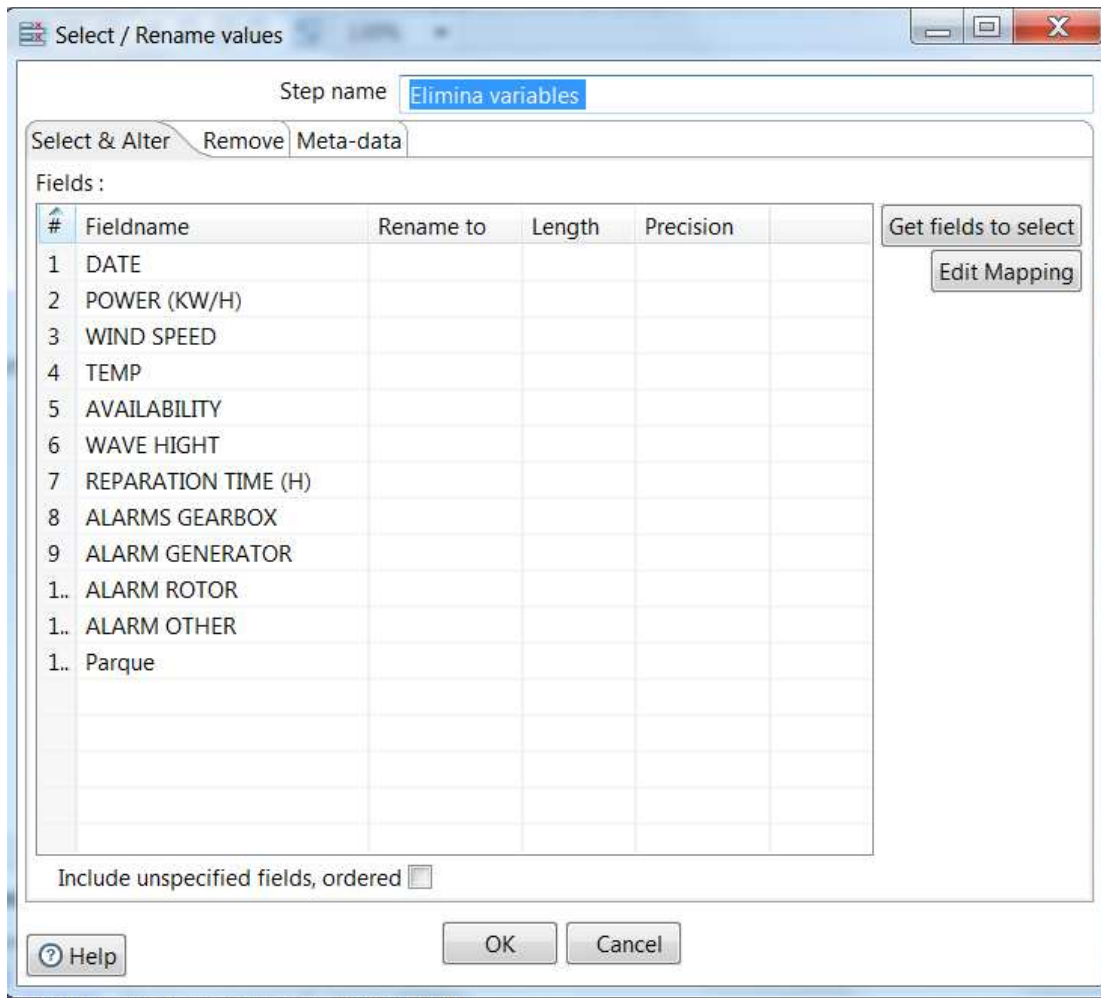


#### a. Lee las hojas-parques





b. Elimina variables



### c. Formatea fecha

Script Values / Mod

Step name: **Formatea fecha**

Java script functions:

- Transform Scripts
- Transform Constants
- Transform Functions
- Input fields
  - DATE.getDate()
  - POWER (KW/H).getNumber()
  - WIND SPEED.getNumber()
  - TEMP.getNumber()
  - AVAILABILITY.getNumber()
  - WAVE HIGHT.getNumber()
  - REPARATION TIME (H).getNumber()
  - ALARMS GEARBOX.getNumber()
  - ALARM GENERATOR.getNumber()
  - ALARM ROTOR.getNumber()
  - ALARM OTHER.getNumber()
  - Parque.getString()
- Output fields
  - DATE.setValue(var)
  - POWER (KW/H).setValue(var)
  - WIND SPEED.setValue(var)
  - TEMP.setValue(var)
  - AVAILABILITY.setValue(var)
  - WAVE HIGHT.setValue(var)
  - REPARATION TIME (H).setValue(var)
  - ALARMS GEARBOX.setValue(var)
  - ALARM GENERATOR.setValue(var)

Java script:

```
var v_day=getDayNumber (DATE.getDate(), "m");  
var v_month=month (DATE.getDate());  
var v_year=year (DATE.getDate());
```

Linens: 0  
Compatibility mode?  Optimization level 9

Fields

#	Fieldname	Rename to	Type	Length	Precision	Replace value 'Fieldname' or 'Rename to'
1	v_day		Integer			N
2	v_month		Integer			N
3	v_year		Integer			N

Help OK Cancel Get variables Test script



e. Busca en dimensión Parque

Database Value Lookup

Step name: Busca en dimensión Parque

Connection: Conexion [Edit... New... Wizard...]

Lookup schema: [Browse...]

Lookup table: d\_parque [Browse...]

Enable cache?

Cache size in rows (0=cache): 0

Load all data from table

The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2
1	nombre_parque	=	Parque	

Values to return from the lookup table :

#	Field	New name	Default	Type
1	id_parque			Integer
2	coordenadax			String
3	coordenaday			String
4	numero_aerogeneradores			Integer
5	nombre_parque			String
6	parque_id_pais			Integer
7	parque_id_operador			Integer
8	parque_id_mantenedor			Integer
9	parque_id_fabricante			Integer
1..	parque_ide_modelo			Integer

Do not pass the row if the lookup fails

Fail on multiple results?

Order by: [ ]

[Help] [OK] [Cancel] [Get Fields] [Get lookup fields]

f. Guarda lecturas en tabla de hechos

Table output

Step name: Guarda lecturas en tabla de hechos

Connection: Conexion [Edit... New... Wizard...]

Target schema: [Browse...]

Target table: h\_lecturas [Browse...]

Commit size: 10000

Truncate table:

Ignore insert errors:

Specify database fields:

Main options | Database fields

Fields to insert:

#	Table field	Stream field
1	m_temperatura	TEMP
2	h_id_pais	parque_id_pais
3	h_id_operador	parque_id_operador
4	h_id_mantenedor	parque_id_mantenedor
5	h_id_fabricante	parque_id_fabricante
6	h_id_potencia_modelo	parque_id_modelo
7	h_id_parque	id_parque
8	h_id_fecha	id_fecha
9	m_potencia_generada	POWER (KW/H)
1..	m_velocidad_viento	WIND SPEED
1..	m_tasa_disponibilidad	AVAILABILITY
1..	m_altura_ola	WAVE HIGHT
1..	m_tiempo_perdido_acum	REPARATION TIME (H)
1..	m_averias_multiplicadora	ALARMS GEARBOX
1..	m_averias_generador	ALARM GENERATOR
1..	m_averias_rotor	ALARM ROTOR
1..	m_averias_otros	ALARM OTHER

[Get fields] [Enter field mapping]

[Help] [OK] [Cancel] [SQL]

#### 1.1.4. Implementación ETL de carga de lecturas de todos los parques

El fichero de formato XML, `Carga_lecturas.ktr`, entregable del proyecto, alberga los metadatos resultantes de la implementación de los procesos ETL que cargan y procesan los registros de todos los parques.