



Bases en el diseño e implantación de un proyecto de Análisis de Datos y de Inteligencia de Negocio. (borrador)

Análisis de datos e inteligencia de negocio (Ad y In)
Data Analysis and Business Intelligence (DA and BI)

Sobre el diseño de un modelo de datos. Cómo comenzar un proyecto de Business Intelligence

Jose Ignacio González Gómez Ver. 2.1

Departamento de Economía, Contabilidad y Finanzas - Universidad de La Laguna

www.jggomez.eu

INDICE

1	Introducción al proceso de diseño de un proyecto de BI	1
2	Un modelo basado en preguntas.....	1
3	Análisis de los datos necesarios y disponibles. Diseño del modelo estrella	1
4	El proceso ETL Power Query (PQ)	2
5	Presentación de la dimensión calendario y su incorporación al modelo.....	2
6	Fases y conceptos a tener en cuenta en un proyecto de análisis de datos y BI	3
6.1	Presentación del caso. Estudio del caso	3
6.1.1	Análisis de la empresa y en concreto del proceso de negocio objeto de estudio	3
6.1.2	Recursos necesarios y disponibles	3
6.2	El proceso ETL con Power Query. Preparando los datos y modelado.....	3
6.2.1	Conectar y/o agregar fuentes de datos (tablas, ficheros planos, etc)	3
6.2.2	Transformar y limpiar los datos con Power Query	3
6.2.3	Modelado. Esquema Estrella	3
6.3	Creación de Medidas. Introducción a DAX	3
6.3.1	Medidas y métricas básicas.	3
6.3.2	Funciones basicas de inteligencia de tiempo	3
6.3.3	Otras funciones.....	3
6.4	Visualización de datos	3
6.4.1	Listas, tarjetas, matriz y gráficas	3
6.4.2	Visualizaciones de KPI y Medidor	3
6.4.3	Aplicando filtros de tarjeta slicers	3
6.4.4	Navegación de páginas	3

1 Introducción al proceso de diseño de un proyecto de BI

Presentamos a continuación una serie de consideraciones a tener en cuenta en la fase inicial de desarrollo de un proyecto de Business Intelligence (BI) y donde la base crítica está en el adecuado diseño del modelo de datos que sustenta el proyecto BI.

Estas recomendaciones se derivan de nuestra experiencia en el diseño e implantación de cuadros de mando en diferentes sectores y áreas empresariales, así como de las recomendaciones de otros expertos en inteligencia de negocios.

2 Un modelo basado en preguntas

Siempre un modelo de datos dirigido a la inteligencia de negocios debe ser el resultado de analizar y reflexionar sobre preguntas vinculadas con la empresa, área de negocio o sector y a la que debe dar respuesta los informes propuestos. Es fundamental desde el punto de vista de la consultoría el hablar con el cliente, conocer los objetivos de las personas y sus necesidades, los KPI (la rentabilidad de los productos o gama, análisis de mercados, de clientes, etc.). Es decir, entender las distintas áreas de negocios, los departamentos, las decisiones de negocio, los KPI que utilizan, que indicadores son importantes, todo ello es la clave para el éxito en el desarrollo un proyecto de inteligencia de negocio.

Así, antes de diseñar un cuadro de mando o dashboard es fundamental tener un conocimiento de la organización más allá de lo técnico para dar soluciones y respuestas al personal.

En resumen, si vamos a diseñar un modelo de datos lo fundamental es conocer y entender cuál es el objetivo, cuáles son las preguntas del negocio, es decir que debe contener el informe que debemos diseñar, indicadores KPIs, etc.

Muchos consultores con amplia experiencia comienzan esta primera etapa con una hoja en blanco sin abrir Excel ni Power BI e intentan pensar en cómo será el modelo de datos intentando identificar y conocer el proceso de negocio que se quiere analizar e identificar todos los filtros o dimensiones asociados.

Generar esta estructura en papel y después plasmarla es lo que nos va a condicionar posteriormente que el código DAX para crear nuestras formulas y medidas sea mucho más fácil de implementar.

3 Análisis de los datos necesarios y disponibles. Diseño del modelo estrella

Además, en este análisis previo va a determinar los datos que necesitamos y confirmar si existen para dar respuestas a las preguntas de negocio y lo que supone un conocimiento de la infraestructura técnica con que cuenta la empresa (CRM, ERP, etc.) y si estos orígenes de datos son fiables y de que información disponemos y tenemos acceso para incorporar a nuestro modelo de datos a través de una importación o vinculación y en caso necesario de un proceso una depuración y adaptación.

Señalar que es recomendable conocer los aspectos técnicos básicos relacionados con las bases de datos, como son las consultas, tipo de campos, estructuras, etc.. que están íntimamente relacionados con el proceso de diseño y construcción de los modelos de datos.

Así partiendo del papel y pensando el modelo de datos correspondiente al área de negocio objeto de estudio y siguiendo la estructura del modelo estrella lo primero que se suele hacer es identificar lo que en el modelo de estrella se llama la tabla de hechos que viene a contener el procesos de negocio (por ejemplos las ventas, atenciones telefónicas, producción, etc..) es decir cualquier hecho que queramos analizar y que normalmente suceden en un momento determinado y que contienen valores que son los que deseamos analizar.

La dimensión viene a ser como queremos analizar ese proceso, por ejemplo el proceso de ventas, la quiero analizar por producto, por cliente, por fecha, empleado, sucursal, etc.. esas serían las dimensiones, como vemos contiene las preguntas que queremos estudiar sobre el proceso, es decir las dimensiones vienen a ser como las preguntas que queremos hacer sobre ese proceso, quien hizo ese proceso, cuando hizo ese proceso, como lo hizo, donde, porque, etc.. esas serían las tablas de dimensión.

Por tanto la estructura de datos no va a ser la del origen de datos sino que vamos a tener el modelo planificado como origen de datos logrando alcanzar normalmente un modelo de datos tipo estrella (con una tabla de hechos y una o varias de dimensión que estarán relacionadas) y que luego cualquier pregunta de negocio relacionados con esos procesos se pueda responder.

De esta forma todos los informes partan de ese modelo final tipo estrella.

En resumen, lo primero que debemos hacer es identificar el proceso de negocio y en segundo lugar todos los filtros que nos van a permitir analizar y dar respuestas a las preguntas detectadas en el modelo de datos.

4 El proceso ETL Power Query (PQ)

Como hemos señalado los datos que nos interesan pueden estar en una base de datos, en una hoja de cálculo, en una página web, en un fichero de datos, pero en cualquier caso una vez localizados necesitamos importarlos y tratarlos o adaptarlos para incorporar al modelo de datos, es decir conectar esos orígenes (importación y/o vinculación), vamos a transformarlos y a modelar y a este proceso es el que se le denomina ETL que significa Extraer, Transformar y Cargar (Extract, Transform, Load). Es el proceso de compilación de datos a partir de un número ilimitado de fuentes, su posterior organización y centralización en un único repositorio.

Debemos evitar comenzar un proyecto conectando los datos y cargando las distintas fuentes de información tal y como están, es indispensable en la mayoría de los casos un proceso de transformación de los datos, pero no solo de limpieza (quitar columnas y filas, ajustar encabezados, remplazar un dato por otro, formateo de los campos, etc..) sino además cambiar la estructura de las tablas con procesos de consulta y consolidación de tablas, estableciendo relaciones, etc.. y para este proceso de depuración y transformación contamos con Power Query que es la herramienta fundamental del proceso ETL.

Es decir si tenemos 2 tablas o 20 tablas no la vamos a cargar directamente al modelo sino con PQ las vamos a limpiar y transformar e incluso en algún debemos consolidar algunas tablas y relacionar y así finalmente carga el resultado de una tabla combinada al modelo de datos para construir posteriormente las métricas con DAX.

5 Presentación de la dimensión calendario y su incorporación al modelo

Por último en cualquier modelo de datos debe existir la dimensión calendario y para ello debemos considerar el disponer como mínimo de esta tabla en nuestro modelo de datos, es fundamental, ya que nos va a permitir analizar por cualquier dimensión de tiempo, como semestre, año, semana, día de la semana, etc.. aplicar funciones de inteligencia de tiempo con DAX (time intelligence dax), como son comparativas de tiempo, mes actual con respecto al mes anterior, semana actual con respecto a la misma semana del años pasado, etc...

6 Fases y conceptos a tener en cuenta en un proyecto de análisis de datos y BI

6.1 Presentación del caso. Estudio del caso

- 6.1.1 Análisis de la empresa y en concreto del proceso de negocio objeto de estudio
- 6.1.2 Recursos necesarios y disponibles

6.2 El proceso ETL con Power Query. Preparando los datos y modelado

- 6.2.1 Conectar y/o agregar fuentes de datos (tablas, ficheros planos, etc)
- 6.2.2 Transformar y limpiar los datos con Power Query
 - Depuración y adaptación
 - Preparando datos, eliminando columnas y filas.
 - Jerarquías
- 6.2.3 Modelado. Esquema Estrella
 - Identificación de Tabla de Hechos y de Dimensiones.
 - Tabla de calendario
 - Conexión y relaciones de tablas de datos

6.3 Creación de Medidas. Introducción a DAX

- 6.3.1 Medidas y métricas básicas.
- 6.3.2 Funciones básicas de inteligencia de tiempo
- 6.3.3 Otras funciones.

6.4 Visualización de datos

- 6.4.1 Listas, tarjetas, matriz y gráficas
- 6.4.2 Visualizaciones de KPI y Medidor
- 6.4.3 Aplicando filtros de tarjeta slicers
 - Interacciones y sincronizar segmentadores.
- 6.4.4 Navegación de páginas