

PEC 2.

Nombre y apellidos: Santiago Domínguez Collado

Usuario UOC: santiagodominguezc@uoc.edu

Estudios que cursa (Máster o Grado): Grado de Ciencia de Datos Aplicada (español)

Pregunta 1 (12,5%):

Desde mi perspectiva, utilizar fuentes de información no contrastadas es un riesgo y una irresponsabilidad. Si los datos finalmente son falsos, probablemente las decisiones tomadas a raíz de esos datos serían erróneas.

En cuanto a la presentación de conclusiones usando fuentes de dudosa fiabilidad, para mí sería equivalente a la divulgación de información falsa y/o manipulada, la cual nunca está justificada. Nadie tiene la potestad para mentir a la población en ningún contexto, esto incluye al gobierno, a la OMS o al órgano gubernamental u organización que se encargue de ello.

Pregunta 2 (12,5%):

- a) Un ejemplo que conozco es el de los grifos de cerveza en japon. Dichos grifos llevan un sensor que cuenta los litros de cerveza servidos, cuando estos se acercan a la cantidad que se ha comprado al proveedor, este vuelve a suministrar cerveza, sin necesidad de que el bar llame y asegurándose de que ningún negocio se queda sin suministro.
- b) Será necesario plantear un escenario realista, no todos los dispositivos tienen que formar parte de un ecosistema IOT si no se va a sacar beneficio de ello. Habrá que establecer qué hacer con esos datos, cómo almacenarlos, procesarlos y si es necesario, aplicarles inteligencia. Es importante asegurarse de que no se vulnera la ley de protección de datos al extraer información de los clientes. Está muy bien que Amazon te sugiera qué poner en la lista de la compra gracias a su nevera inteligente, pero no sería apropiado que Amazon les dijese a los contactos de un usuario que este compra mucho un producto.

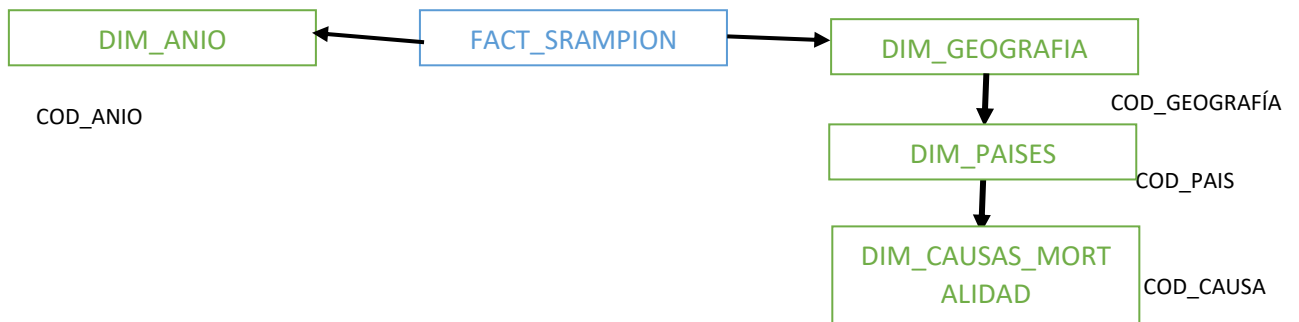
Pregunta 3 (7,5%):

La opción correcta es la b. El diseño conceptual es el de mayor nivel de abstracción, luego el diseño lógico y por último el físico.

Pregunta 4 (15%):

Es un diagrama de estrella. Los diagramas de copo de nieve, con el beneficio de reducir una cantidad de espacio mínimo, complican de forma exagerada el diagrama. Es obvio que este diagrama es más que entendible, por lo tanto, no es de copo de nieve.

Pregunta 5 (15%):



Pregunta 6 (7,5%):

La respuesta correcta es la c. Es cierto que una clave principal puede estar compuesta por múltiples columnas, sin importar que dichas columnas sean o no foreign keys de la tabla. Por otro lado, es seguro que solo hay un registro de cada país por año para cada causa, por lo tanto, en cada fila, la combinación de las tres columnas sería única.

Pregunta 7 (15%):

En la arquitectura Inmon se necesita transferir la información desde sistemas transaccionales a un lugar centralizado donde los datos puedan ser analizados. Esta arquitectura debe tener las siguientes características:

- Orientado a temas.
- Integrado: la base de datos contiene los datos consistentes de los sistemas operacionales de la organización.
- No volátil.
- Variante en el tiempo.

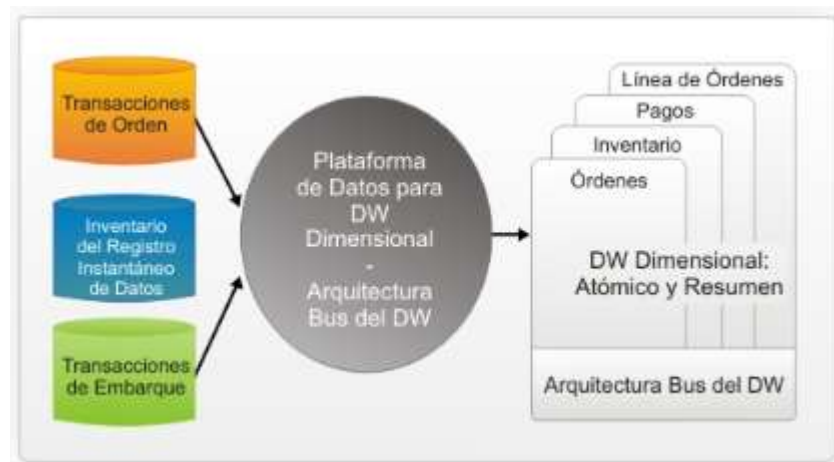
La información en este tipo de arquitectura tiene que estar a los máximos niveles de detalle posibles. Los datamarts son tratados como subconjuntos del almacén corporativo, y son contruidos para cubrir las necesidades individuales de análisis de cada departamento.



El enfoque Inmon ingesta los datos de forma Top-down. Los datos son extraídos de los sistemas operacionales mediante procesos ETL y cargados en la staging area para posteriormente cargarlos en el almacén.

En la arquitectura Kimball el almacén es un conglomerado de los almacenes departamentales, siendo una copia de los datos transaccionales estructurados de una forma especial para el análisis.

Este modelo no normalizado incluye las dimensiones de análisis y sus atributos y una organización jerárquica. Tenemos tablas tanto de dimensiones como de hechos. Los diferentes Data Marts están conectados entre sí (si es necesario) por una estructura de nominada *bus structure*.



Este enfoque ingesta los datos de forma Bottom-up, pues al final el almacén corporativo no es más que la unión de los almacenes departamentales.

A la hora de comparar ambos modelos, podemos concluir que el enfoque Inmon es más adecuado para sistemas complejos, con alta durabilidad y consistencia. Pero para pequeños proyectos, como ha sido el caso de la práctica 2, el modelo Kimball es más sencillo y rápido de desarrollar.

Pregunta 8 (15%):

Herramienta de ETL.

Pentaho Data Integration es un software muy capaz, sencillo e intuitivo que permite al usuario ahorrarse múltiples y engorrosas tareas, no obstante, tiene competidores que resaltan donde Pentaho no.

Oracle data integration es una potente herramienta que permite sin lugar a duda, realizar todo lo que permite Pentaho y más. Por ejemplo, la herramienta de Oracle permite una mayor integración con sistemas en la nube y trabaja mucho mejor con cantidades masivas de datos.

Aplicaciones BI.

Microsoft pone al servicio de los usuarios aplicaciones muy útiles e intuitivas. Por ejemplo, Power BI, ofrece por un bajo precio un potente software de análisis. A pesar de esto, Power BI no es el referente de aplicaciones BI en el mercado, ¿Por qué ocurre esto?

Alternativas como Tableau o Qlik permiten a un usuario experto, llegar más lejos en sus analíticas y exploraciones, así como una mayor red de partners por parte de las empresas y mayor flexibilidad en las implementaciones que se hacen con estos programas.

Conclusión:

A pesar de lo expuesto anteriormente, entiendo y apoyo porqué se nos han facilitado herramientas como Pentaho y Power BI en esta asignatura, pues dichas herramientas son más que competentes para las tareas que hemos tenido que abordar. Y teniendo en cuenta que ambas son más intuitivas que las de la competencia, me parecen la mejor elección posible.