

ISW-911 Minería de Datos

Proyecto I

Supuestos de Negocio:

- Como consultor de la empresa, Grupo Innova S.A, usted fue contratado para definir los requerimientos para la construcción de un datawarehouse que permita conocer información de temas relevantes como ventas mensuales, rendimiento de las sucursales, rendimientos de los empleados, márgenes de utilidad neta, entre otros, con la finalidad de basar sus decisiones estratégicas en datos.

Supuestos de Proyecto:

- El tiempo máximo, por grupo, para la defensa del proyecto es de 20 minutos.
- Los grupos ya fueron creados con anterioridad, no se permitirán cambios durante la elaboración de este proyecto.
- Debe utilizar GIT (gitlab) como repositorio de control, la participación de los integrantes debe ser equitativa.
- Aunque el proyecto es grupal, la evaluación será de forma individual, predominando la participación y el conocimiento de los conceptos de forma equitativa por parte de los integrantes.
- La base de datos para el staging y datawarehouse sera SQL Server
- La herramienta para ETL sera SSDT
- La o las base de datos transaccionales pueden ser variables (Oracle,Web, ect) deben suponer una cantidad de datos apreciable para la visualización de los resultados, mayor a 5.000.00 transacciones por año y al menos de los últimos 5 años, por ejemplo.
- Se permite el uso de los scripts de clase como referencia, pero las estructuras (objetos de base de datos) debe ser completamente diferentes.

Objetivo del proyecto:

Comprender los conceptos aprendidos en el curso de minería de datos sobre datawarehouse, aplicando un ejercicio de índole investigativo y práctico en la resolución de los enunciados presentados a continuación.

Enunciado I:

- a. Realice una investigación acerca de la metodología para satisfacer necesidad de un marco de referencia para el diseño y la construcción de un Data Warehouse y un Data Lake.
- b. Evalué, comparativamente, los modelos bajo los enfoques de William H. Inmon y Ralph Kimball.

Requerimientos: Documento .pdf formato APA

Enunciado II:

- a. Utilizando la metodología estudiada en el enunciado I, realice el diseño de un datawarehouse.

Debe contar, al menos, con las dimensiones de:

- Clientes
- Empleados
- Agencias o Sucursales
- Tiempo
- Productos
- Proveedores
- Direcciones

Adicionalmente las tablas de hecho que requiera, Si requiere más dimensiones puede incluirlas.

Enunciado III:

- a. Realice el proceso ETL para cargar los datos de una base de datos transaccional a la base de datos staging área y posteriormente, se cargarán los datos al datawarehouse.
- b. Crear los procesos de llenado de datos tanto a las tablas de dimensión como a las tablas de hechos.

Requerimiento: utilizar al menos 10 componentes diferentes de data tools en la construcción del flujo de carga de datos.

Enunciado IX:

Crear las estructuras OLAP:

- a. Crear una estructura multidimensional que permita acceder a la información definida en los requerimientos de la empresa.
- b. Crear una estructura tabular que permita acceder a la información definida en los requerimientos de la empresa.
- c. Realizar un comparativo entre las ventajas y desventajas de la utilización de un modelo tabular y uno multidimensional. (PDF como anexo al documento del enunciado I)

- d. Conectar las estructuras a PowerBI de forma que se pueda demostrar el correcto funcionamiento de dichas estructuras (no es necesario profundizar en este punto solo demostrar el funcionamiento OLAP)