



Proyecto No. 1
Caso de negocio
(Grupos de 3 o 4)

Modalidad y fecha de entrega

- El proyecto debe trabajarse en grupos de 3 o 4 integrantes, y ser entregado de acuerdo a la siguiente programación de entregas:
 - o Primera entrega: Jueves 16 de mayo a las 19:00
 - o Segunda entrega: Miércoles 5 de junio a las 16:00
- Luego de la fecha límite se restará 10 puntos por cada hora de atraso en la entrega

Descripción general del proyecto

El proyecto consiste en la implementación de conceptos de OLAP y Data warehousing para una base de datos relacional en PostgreSQL, simulando un entorno distribuido de puntos de venta. Contempla también el diseño y centralización de un Datawarehouse en Hive/Hadoop para los reportes de ventas. Además, incluye el uso de Hadoop como herramienta para el almacenamiento del tracking de los clientes.

El proyecto pretende simular el comportamiento de un conjunto de tiendas en donde cada una maneja una instancia independiente de un sistema de ventas. La premisa básica de la simulación es que cada tienda puede registrar empleados, clientes, productos y ventas independientes, y la información debe ser agregada hacia un *data warehouse* para procesamiento y presentación.

El proyecto debe contemplar las siguientes fases:

Fase 0: Pre-requisitos y setup de ambiente

- Descarga de herramientas que se utilizarán durante el proyecto
- Comunicación de todos los equipos mediante un switch para que cada uno se pueda comunicar con los demás en el ambiente

Fase 1: Diseño de una base de datos relacional (OLTP)

- Se debe diseñar una base de datos operativa que incluya datos de clientes, productos, categorías, marcas, facturas, líneas de facturas.
- Se deben agregar las restricciones de integridad referencial pertinentes
- Se debe almacenar la fecha y hora en la que ocurre cada venta.
- Se debe desarrollar una estructura que permita especificar y almacenar información adicional *custom* sobre los productos (sin utilizar ALTER TABLES).

Fase 2: Implementación de un sistema relacional operativo de ventas

- Se debe desarrollar una interfaz gráfica de usuario que permita consultar y registrar clientes, productos y ventas (facturas).
- Esta interfaz gráfica debe permitir también especificar la información adicional *custom* a asociar a los productos.
- Se debe ejecutar esta aplicación en al menos dos equipos simulando una sucursal de la tienda por equipo.
- El desarrollo de esta aplicación debe estar versionado de forma que se pueda revisar el historial de aportes
- El manejador de base de datos a utilizar es PostgreSQL.

Fase 3: Simulación de realización de ventas

- Desarrollar una funcionalidad en la interfaz que simule de manera aleatoria 1 día de operación dentro de una tienda: creación de clientes, productos, facturas, etc., dentro de la base de datos.
- Esta funcionalidad de simulación debe solicitar la fecha dentro de la cual se realizarán la creación de los registros

Fase 4: Instalación y configuración de Hadoop

- Se debe instalar Hadoop versión 3.1 sobre máquina virtual de Linux. Esto incluye la instalación y validación de pre requisitos
- La instalación se debe hacer en al menos 2 nodos (equipos) bajo la modalidad *cluster setup*.
- Diseño de instalación de nodos y servicios de Hadoop. Esto incluye master nodes y worker nodes.

Fase 5: Implementación de DWH en Hive (OLAP)

- Realizar el diseño de la base de datos analítica para el DWH
- La base de datos analítica debe poder responder consultas de ventas por cliente, producto, categoría, marca, tienda y fecha
- Instalación y configuración de Hive dentro del ambiente de Hadoop
- Centralizar en Hive la data de ventas generada por todas las tiendas de conveniencia
- Utilizar la herramienta *Talend Open Studio* para hacer los procesos de ETL (*extraction, transformation, loading*) entre las bases de datos operativas y el DWH
- Desarrollar consultas en HQL (*Hadoop query language*, lenguaje de consulta de datos basado en SQL) para mostrar la información de las ventas

Fase 6: Inteligencia de negocios

- Uso de Tableau en versión trial como herramienta de visualización y reportería de la información de la empresa
- Conexión de Tableau al DWH en Hive con ODBC o JDBC



- Mostrar todas las métricas posibles de ventas y sus combinaciones. Por cliente, por producto, categoría, marca, tienda y fecha

Fase 7: Interacción con DBs no relacionales

- Levantar una instancia de MongoDB y registrar allí los clientes y ventas consolidados de todas las tiendas
- Desarrollar un script en Javascript que analice el listado de clientes y publique en Twitter una oferta personalizada para cada uno de los cinco clientes con mayor volumen de facturación.

Temas a reforzar

- OLTP, SQL, PostgreSQL
- OLAP, DWH
- Hadoop y Hive
- Integración de programación con bases de datos
- Integración con herramientas para gestión de información
- Uso de herramientas de ETL
- Uso de herramientas de inteligencia de negocio

Referencias

- Entity-attribute-value model:
https://en.wikipedia.org/wiki/Entity%E2%80%93attribute%E2%80%93value_model
- How to install hadoop in standalone mode:
<https://www.digitalocean.com/community/tutorials/how-to-install-hadoop-in-stand-alone-mode-on-ubuntu-18-04>

VI. Evaluación

- Primera entrega (15 netos)
 - Fase 1: 30%
 - Fase 2: 30%
 - Fase 3: 20%
 - Fase 4: 20%
 - **Total: 100%**
- Segunda entrega (20 netos)
 - Fase 5: 40%
 - Fase 6: 30%
 - Fase 7: 30%
 - **Total: 100%**

Total: Total 35 puntos netos