

Proyecto 3 (10%)

Fecha Entrega Miércoles 28 octubre de 2020

El objetivo de este proyecto es profundizar más en el manejo de herramientas que hoy en día son sumamente buscadas en el mercado laboral y que son importante que los estudiantes manejen. Se repasarán conceptos de bases de datos (DDL y DML) y programación en Python.

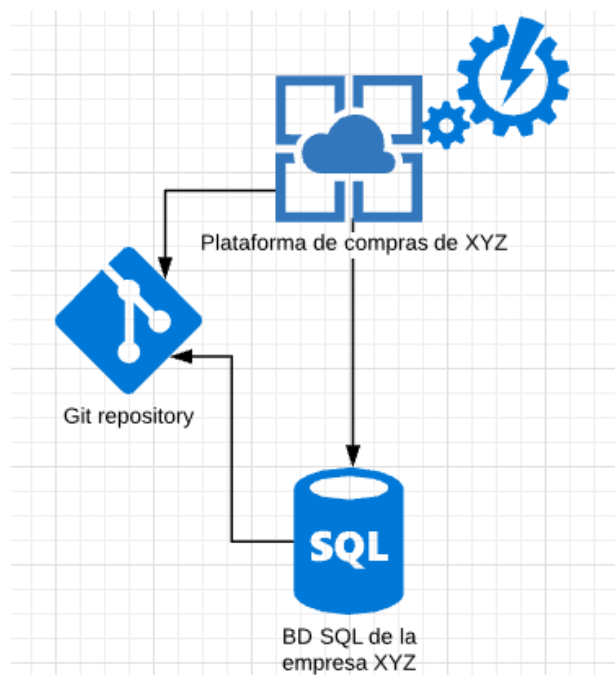
Problema

Para efectos de este proyecto se tiene la siguiente situación: Una empresa de nombre “XYZ” posee actualmente una Base de datos en SQL Server, encargada de guardar información de clientes, tarjetas de crédito, compras realizadas por los mismos, teléfonos de los clientes, direcciones de los clientes y correos electrónicos.

Para mantener la calidad de los datos y la disponibilidad al 100% se dispone de licencias Enterprise de SQL, que conllevan un alto costo mes a mes. El DevOps de la empresa ha propuesto realizar una migración de todos los datos almacenados en esta Base de Datos a una Base de Datos en PostgreSQL.

Con esto se planea eliminar las licencias de SQL Enterprise y ahorrarle a la empresa miles de dólares mes a mes que se gastan en las mismas.

Diagrama actual de la plataforma de “XYZ”

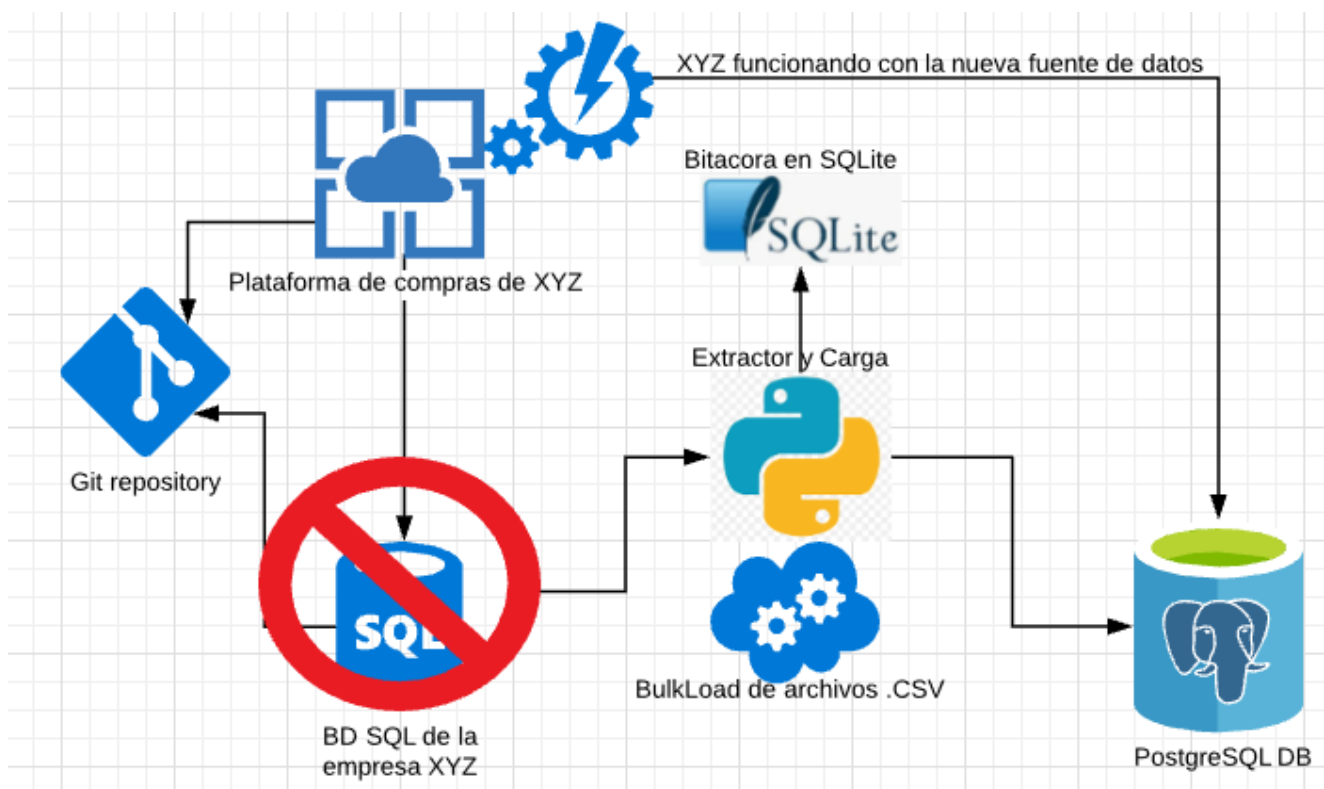


Propuesta de solución:

El DevOps de XYZ ha hecho solicitado una serie de tareas que deberán ser concluidas para dar por finalizado el entregable:

- 1) Se deberá crear un extractor de datos en Python el cuál estará encargado de extraer todos los datos de la Base de Datos en SQL Server.
- 2) Todos los datos extraídos de SQL deberán generar archivos .CSV por día de ejecución
- 3) Se deberá programar un segundo modulo en Python el cuál tomará cada .CSV y lo almacenará en PostgreSQL.
- 4) No se permite el uso de sentencias INSERT en PostgreSQL, si no que deberá hacer uso de BulkLoad de datos.
- 5) Se deberá crear una bitácora como paso intermedio la cuál lleve el control de que se ha extraído de SQL, que archivos se han generado, y que archivos se han cargado a PostgreSQL
- 6) Esta bitácora deberá ser creada en SQLite
- 7) Todo deberá ser respaldado en un repositorio (Scripts de Python, SP's de SQL, script de creación de BD en PostgreSQL
- 8) El extractor de Python deberá ser capaz de recibir parámetros de fechas, y así controlar que datos migrar y que datos no migrar.

Diagrama de propuesta



Requerimientos del proyecto:

- 1) Crear una base de datos en SQL Server la cual representa la BD de la empresa "XYZ". 5pts
- 2) La misma deberá estar normalizada, y contener una tabla para cada entidad definida al inicio del documento (clientes, órdenes de compra, teléfonos, direcciones, correos electrónicos y cualquier otra tabla que considere necesario). 5pts
- 3) Deberá encriptar teléfonos, direcciones, correos electrónicos, y tarjetas de crédito a nivel de columnas en cada tabla SQL (Usar certificados y encriptación por certificados). 10pts
- 4) Crear una Base de Datos gemela, pero en PostgreSQL que vendrá a ser la nueva Base de Datos a utilizar por "XYZ". 10pts
- 5) La BD en SQL deberá poseer registros en cada tabla (al menos 1000 rows por tabla, generados aleatoriamente, pero con fechas de inserción o actualización variados). 5pts
- 6) El script en Python deberá poseer 2 módulos separados, el extractor de datos y el que se encarga de almacenar datos en PostgreSQL 5pts
- 7) EL extractor deberá generar archivos .CSV por cada batch de ejecución que se solicite 5pts
- 8) Cada vez que se extrae datos y se genera un .CSV se guarda en una bitácora en SQLite. 5pts
- 9) Cada vez que se almacena datos en PostgreSQL se guarda en otra tabla de la misma bitácora. 5pts
- 10) Apenas se logra almacenar datos en "PostgreSQL" la misma deberá ser eliminada de SQL Server
- 11) Teléfonos, Direcciones y Correos electrónicos deberán ser eliminados mediante JOBS de SQL 5pts
- 12) Ordenes de compra y clientes, deberán ser eliminados mediante un Procedimiento almacenado, llamado desde Python. 5pts
- 13) Todo código que se programe deberá estar en el repositorio definido previamente, el día de revisión de proyecto, si el código de Python o los scripts de SQL no están en el repositorio, no se calificará. 5pts
- 14) Se da por concluido el proyecto con nota 100, si todos los datos que se prueban ese día se ven reflejados en la BD de PostgreSQL 30pts

Requerimientos del proyecto:

- 1) El proyecto se revisa el jueves 31 de octubre sin excepción:
- 2) La defensa del proyecto será publica y en horas lectivas.
- 3) El proyecto se desarrollará en los grupos asignados previamente durante el curso.
- 4) El proyecto deberá estar alojado en un repositorio (GitHub, GitLab o Bitbucket)
- 5) Los lenguajes de programación a utilizar son solo Python, SQL, PostgreSQL, SQLite

Importante:

La comprobación o sospecha fundada de que algún proyecto presenta interfaces, códigos, o trabajos escritos iguales o que se parezcan llevará aplicar las sanciones que contempla el reglamento de Régimen Académico Estudiantil, en el artículo 4, inciso C.

De comprobarse que existe dicho plagio el o los estudiantes de los grupos involucrados pierden el proyecto.