**DBT**

* diferencias entre dbt core, dbt cloud free y dbt cloud de pago
* Paso a paso de un proyecto dbt + postgre
* Qué es dbt y para qué nos sirve (demostrado con el ejemplo anterior)?

**INTRODUCCIÓN**

Data Build Tool(dbt) es una herramienta de análisis de datos open source que permite a los equipos de análisis y datos transformar, documentar y orquestar datos de una manera eficiente.

Se enfoca principalmente en la transformación de datos, toma datos en crudo de diferentes fuentes, los limpia y los transforma. Posteriormente los entrega en un formato estructurado para que los analistas o científicos de datos puedan trabajar con ellos.

DBT permite trabajar con una amplia variedad de siestemas de almacenamiento de datos, como bases de datos relacionales, cloud, data warehouse, etc.

DBT se escribe en SQL y permite a los equipos de análisis colaborar en el código, documentar y automatizar la transformación de datos.

**TARIFAS**

A continuación, se muestran las principales diferencias entre DBT core, DBT Cloud Free y DBT Cloud de pago:

* DBT Core: Versión gratuita, puede ser instalador localmente en un ordenador o servidor. Incluye las funciones esenciales de DBT como transformación de datos, testeo y orquestación de tareas. También dispone de una comunidad activa para proporcionar soporte.
* DBT Cloud Free: Versión gratuita de DBT Cloud, plataforma que ofrece herramientas de orquestación, programación, ejecución y monitoreo de flujos de trabajo de DBT. Al ser gratuita tiene una limitación en cuanto a la cantidad de proyectos y ejecuciones que pueden realizar. Es la ideal para probar y familiarizar.
* DBT Cloud de pago: Versión de pago, ofrece más funcionalidades en comparación a la gratuita y más capacidad de procesamiento de datos. Incluye funciones avanzadas como la integración con otras herramientas de análisis de datos, soporte a múltiples data warehouse, programación avanzada de flujos y el monitoreo y análisis en tiempo real. Es ideal para usuarios que necesitan trabajar con grandes volúmenes de datos y equipos de análisis de datos más grandes y complejos.

Como podemos ver, tenemos dos versiones gratuitas, DBT Core y DBT Cloud Free. La primera es una herramienta de línea de comandos, que permite transformar datos, mientras que la segunda es una plataforma que ademas de permitir transformar datos, también tiene un seguimiento de versiones (gracias a su compenetración con Git), programación de tareas, alertas, etc. Además DBT Cloud Free no hay que tenerlo instalado en nuestra máquina, a diferencia de DBT Core que si que es necesario tenerlo instalado. El hecho de ser una plataforma en línea, nos permite ejecutar DBT Cloud Free desde cualquier dispositivo, mientras que DBT Core solo podremos hacerlo desde la máquina local.

**CREAR PROYECTO DBT**

Para crear un proyecto dbt tenemos que tener instalado Python en una versión superior a la 3.5, y también nuestra BBDD Postgre SQL, la cual instalaremos desde su propia página web: <https://www.postgresql.org/download/>

Una vez hemos descargado el instalador de Postgre SQL para nuestro SSOO, durante el proceso de instalación nos va a pedir que asignemos diferentes nombres de usuario y contraseñas para poder acceder a la base de datos.

A partir de aquí, deberemos instalar dbt. Para nuestro sistema operativo, Windows, lo vamos a hacer desde el cmd o línea de comandos, ejecutando pip install dbt. Al hacer esto nos indicará que debemos especificar un determinado paquete para instalar dbt, por lo que haremos pip install dbt-postgres.

Para la creación de un proyecto deberemos ejecutar dbt init "ProjectName", y nos creará una carpeta del proyecto en el directorio donde nos encontrábamos.

Alli deberemos configurar el archivo profiles.yml para adaptarlo a nuestra base de datos PostgreSQL con el siguiente codigo:

company-name:

target: dev

outputs:

dev:

type: postgres

host: [hostname]

user: [username]

password: [password]

port: [port]

dbname: [database name] # or database instead of dbname

schema: [dbt schema]

threads: [optional, 1 or more]

[keepalives\_idle](#keepalives\_idle): 0 # default 0, indicating the system default. See below

connect\_timeout: 10 # default 10 seconds

[retries](#retries): 1 # default 1 retry on error/timeout when opening connections

[search\_path](#search\_path): [optional, override the default postgres search\_path]

[role](#role): [optional, set the role dbt assumes when executing queries]

[sslmode](#sslmode): [optional, set the sslmode used to connect to the database]

<https://docs.getdbt.com/reference/warehouse-setups/postgres-setup>

A la hora de crear un proyecto, o utilizar cualquier comando "dbt" puede dar alguna clase de problemas relacionados con no poder ejecutar dbt. Para comprobar que este paquete se ha instalado correctamente ejecutaremos:  
pip list -> Para ver todos los paquetes  
pip show dbt-postgres -> Para ver informacion sobre ese paquete  
pip show dbt-core ->Para ver informacion sobre ese paquete

Si estos se encuentran instalados, deberemos añadir a nuestras variables del sistema la ruta de la carpeta Scripts o bin de nuestra version de Python instalada. Para concoer esta ruta ejecutaremos: "python -m site --user-base."

Ahora para añadir esta ruta buscaremos en el buscador de Windows:  
- Editar las variables de entorno del sistema  
 - Variables de entorno  
 - "Editar", en Variables del sistema, variable Path  
 - Nueva (Añadir Ruta) y aceptar ventanas

Al crear un proyecto, deberemos tener una carpeta dentro de nuestro usuario (C:\Users\<Nombre\_De\_Usuario>\.dbt), donde estará el archivo de perfil, "profiles.yml". También se nos crea una carpeta con el nombre del proyecto que le hemos dado en el repositorio desde que lanzasemos el comando de "dbt init nombre\_de\_proyecto".

Esta carpeta la abriremos con un editor de código como puede ser Visual Studio Code, y abriremos el fichero "<Nombre\_Del\_Proyecto>.yml". Aqui se encontrará toda la configuración del proyecto, en nuestro caso introduciremos el codigo anterior ya que es el adaptado al perfil de postgres.

La linea primera del "profiles.yml" donde dice company-name, debe coincidir con la linea del .yml del proyecto que dice profile, en este caso en el archivo de configuración del proyecto deberemos ver:  
profile: 'company-name'

Una vez introducidos todos los datos correctamente (host, usuario, nombre de base de datos, contraseña, etc.), abriremos un terminal desde VS Code, dandole en la opción del menú superior Terminal > Nuevo Terminal, y escribiremos "dbt debug". De esta forma podremos comprobar que dbt y postgre se han conectado correctamente o por el contrario no se ha podido establecer la conexión.

Si nos aparecen las siguientes líneas será que todo ha ido de forma exitosa:

"Connection test: [OK connection ok]

All checks passed! "

**AIRBYTE**

Plataforma de integración de datos open source, que permite a las empresas mover datos de diversas fuentes a una amplia variedad de destinos de datos en tiempo real. Se enfoca en que la integración de datos sea fácil, rápida y confiable.

La instalación de Airbyte se puede realizar ya sea en máquinas locales, virtuales como en Docker o Kubernetes, etc. Airbyte ofrece instrucciones detalladas para mucha variedad de instalaciones.

A diferencia de otras herramientas de integración de datos, Airbyte ofrece:

* Código abierto
* Facilidad de uso
* Conectores preconstruidos
* Escalabilidad
* Comunidad

La diferencias de los planes esta