

# Desafío Técnico

Federico De Grazia

Junio 2022

## 1 Introducción

---

El objetivo del presente documento es explicar cómo está compuesto el entregable y inicialización del ambiente de trabajo para resolver los ejercicios planteados en el siguiente enunciado:



Desafío Ing de  
Datos.pdf

## 2 Puesta en marcha

El ambiente de trabajo fue montado sobre Docker con las siguientes imágenes:

1. jupyter/pyspark-notebook: <https://hub.docker.com/r/jupyter/pyspark-notebook>
2. mcr.microsoft.com/mssql/server: [https://hub.docker.com/\\_/microsoft-mssql-server](https://hub.docker.com/_/microsoft-mssql-server)

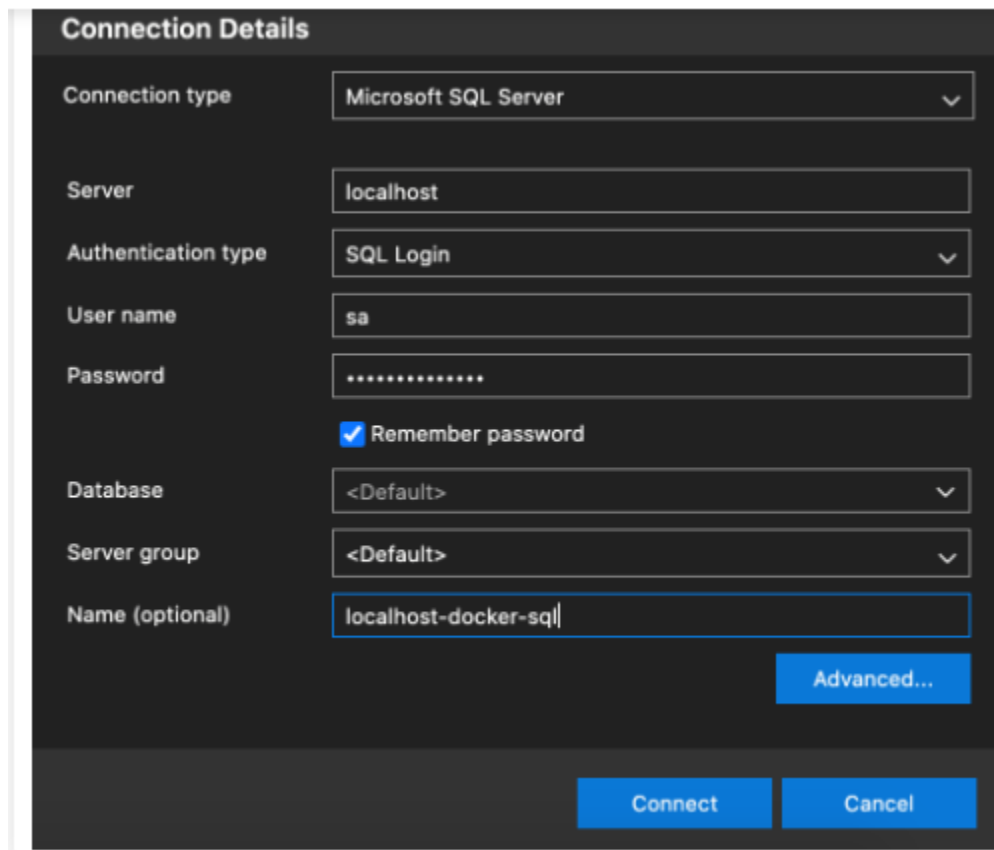
Se descargaron dichas imágenes con los comandos pull sugeridos en cada url.

### 2.1 SQL Server configuración:

1. Ejecución del container:

```
docker run -e "ACCEPT_EULA=Y" -e "SA_PASSWORD=<password>" -p 1433:1433 -d mcr.microsoft.com/mssql/server
```

2. Conexión a la base desde Azure data studio:



**Connection Details**

Connection type: Microsoft SQL Server

Server: localhost

Authentication type: SQL Login

User name: sa

Password: .....

☒ Remember password

Database: <Default>

Server group: <Default>

Name (optional): localhost-docker-sql

Advanced...

Connect Cancel

3. Habilitar preview features para realizar el restore:

<https://docs.microsoft.com/en-us/sql/azure-data-studio/preview-features?view=sql-server-ver16>

4. Disponibilizar el archivo bak en el filesystem de docker:

```
docker cp AdventureWorks2014.bak <CONTAINER ID>:/var/opt/mssql/data
```

## 2.2 Organización del entregable:

El directorio raíz es “Desafio\_tecnico\_nubimetrics”, el cual se divide en “pyspark” y “sql”.

La parte de spark se encuentra dividido en los siguientes directorios:

1. Input\_files: Contiene los archivos facilitados para resolver los ejercicios.
2. Out\_files: Destino de los archivos generados por la ejecución del script.
3. Src: Contiene el script de pyspark, un archivo de configuración con la metadata de cada archivo en input\_files.

El directorio de “sql” contiene un source con el código de las queries de los últimos desafíos.