



Formação Engenheiro de Dados

O que é Engenharia de Dados

Ingestão do DW para o Data Lake

- ◆ Carregar dados do Data Warehouse do JJBike para o HDFS

Preparar Postgres para Conexão Remota

- ♦ Alterar senha do user postgres:

```
sudo -u postgres psql  
alter user postgres PASSWORD '123456';
```

- ♦ Alterar arquivo de configuração do Postgres, para permitir conexão remota

```
sudo vi /var/lib/pgsql/11/data/pg_hba.conf
```

- ♦ Incluir linha

```
host all all 0.0.0.0/0 md5
```

- ♦ Deve estar antes de ident

- ♦ Parar e reiniciar Postgres

```
sudo service postgresql-11 stop  
sudo service postgresql-11 start
```

Baixar e Instalar Driver para Sqoop

❖ Baixar driver JDBC para o Sqoop

```
cd /home/cloudera/Downloads  
wget  
http://jdbc.postgresql.org/download/postgresql-9.2-1002.jdbc4.jar
```

❖ Copiar para var/lib/sqoop/

```
sudo cp postgresql-9.2-1002.jdbc4.jar  
/var/lib/sqoop/postgresql-9.2-  
1002.jdbc4.jar
```


Rodando Sqoop

```
sqoop list-databases --connect  
jdbc:postgresql://127.0.0.1/ --username  
postgres --password 123456
```

```
sqoop list-tables --connect  
jdbc:postgresql://127.0.0.1/ed --  
username postgres --password 123456 --  
--schema dimensional
```

```
sqoop import-all-tables --connect  
jdbc:postgresql://127.0.0.1/ed --  
username postgres --password 123456 --  
--schema dimensional
```

Resultados

- ◊ Verificar a importação (um diretório por tabela)

```
sudo hdfs dfs -ls /user/cloudera
```

- ◊ Abrir um Arquivo

```
sudo hdfs dfs -cat  
/user/cloudera/fatovendas/part-m-00000
```