# SQOOP – Exercícios

## Parte 2

1. Dúvidas sobre sintaxe podem ser sanadas no site:

<https://sqoop.apache.org/docs/1.4.6/SqoopUserGuide.html>

Logs de erro podem ser obtidos com o comando:

**yarn logs -applicationId application\_<codigo\_da\_aplicação> > x.txt**

Exemplo:

**yarn logs -applicationId application\_1485423751090\_3566 > x.txt**

O arquivo X.TXT conterá o log dos erros.

1. Inicie a máquina virtual BigData Light. Usuário oracle. Senha welcome1.
2. Abra um terminal. Mude para o diretórios DOWNLOADS.

**cd Downloads**

1. Usando um editor de texto, crie um arquivo denominado c**sv\_generator.csv**.

**gedit csv\_generator.csv**

1. Insira o texto abaixo no arquivo **csv\_generator.csv**.

**#! /bin/bash**

**col\_three=('Rita' 'Humberto' 'Olivia' 'Barcelos' 'Patricia' 'Parducci');**

**col\_four=(20 21 23 25 53 67 68 69 80);**

**col\_five=('TCP' 'TCP' 'Telnet' 'SMTP' 'DNS' 'DHCP' 'DHCP' 'TFTP');**

**today=`date '+%Y\_%m\_%d\_\_%H\_%M\_%S'`**

**file\_name="log\_$today.csv"**

**for ((i=0; i<1000;i++))do echo "Registro $((i+1)) criado!" timestamp=`date +%s%3N`; ran\_one=$RANDOM%6; ran\_two=$RANDOM%8; printf "$((i+1)),$timestamp,${col\_three[ran\_one]},${col\_four[ran\_two]},${col\_five[ran\_two]}\n" >> $file\_namedone**

1. Altere as permissões do arquivo csv\_generator.csvpara que possa ser executado.

**chmod +x csv\_generator.csv**

Execute o *script*:

**./csv\_generator.csv**

Verifique o nome do arquivo gerado com o comando

**ls log\*.csv**

Examine o conteúdo do arquivo gerado com o comando de S.O. *cat*. Por exemplo:

**cat log\_2022\_03\_16\_\_21\_46\_55.csv**

O nome do seu arquivo deve ser o nome obtido pelo comando *ls*.

Para efeito do exercício considere que a estrutura do arquivo é a seguinte:

|  |  |  |  |
| --- | --- | --- | --- |
| Nome da Coluna | Tipo de Dados | Tamanho | Observações |
| CODIGO | Numérico | 8 | NOT NULL |
| IP | Numérico | 14 | NOT NULL |
| USUARIO | Texto | 16 | NOT NULL |
| PORTA | Numérico | 6 | NOT NULL |
| PROTOCOLO | Texto | 8 | NOT NULL |

1. Copie o arquivo gerado do *file system* (FS) local para o FS HADOOP. Durante a cópia, altere o nome do arquivo para **dados.csv**.
2. Altere os atributos do arquivo **dados.csv** para que o OWNER do arquivo tenha permissão de leitura e escrita. Todos os demais usuários deverão ter apenas permissão de leitura.
3. Conecte-se ao seu usuário no banco de dados da FIAP

sqlplus RM1234/senha@oracle.fiap.com.br:1521/orcl

1. e crie a tabela DADOS com a estrutura necessária para receber os dados do arquivo **dados.csv**.

CREATE TABLE dados (

col1 NUMBER(8),

col2 NUMBER(14),

col3 VARCHAR2(16),

col4 NUMBER(6),

col5 VARCHAR2(8)

);

1. Usando o SQOOP importe os dados do arquivo **dados.csv** para a tabela **dados** criada no banco de dados.
2. Crie a tabela DADOS\_ETL com a seguinte estrutura:

|  |  |  |  |
| --- | --- | --- | --- |
| Nome da Coluna | Tipo de Dados | Tamanho | Observações |
| USUARIO | Texto | 16 | NOT NULL |
| PORTA | Numérico | 6 | NOT NULL |
| PROTOCOLO | Texto | 8 | NOT NULL |
| ACESSOS | Numérico | 10 | NOT NULL |

CREATE TABLE dados\_etl (

col1 NUMBER(8),

col2 NUMBER(14),

col3 VARCHAR2(16),

col4 NUMBER(6),

col5 VARCHAR2(8),

col6 NUMBER(10)

);

1. Usando comandos SQL crie um relatório que mostre o nome do usuário, porta, protocolo e a quantidade de acessos que um mesmo usuário tenha feito usando a mesma porta e protocolo. Ordene os dados pelo nome do usuário em ordem crescente, protocolo em ordem decrescente, porta em ordem crescente e acessos por ordem crescente. Não liste os dados dos usuários que fizeram menos dos que 10 acessos usando a mesma porta e protocolo.

SELECT usuario, porta, protocolo, COUNT(\*) ACESSOS

FROM dados

GROUP BY usuário, porta, protocolo

HAVING acessos > 9

ORDER BY usuario ASC, port DESC, protocolo ASC, acessos DESC;

1. Usando comandos SQL, insira os dados obtidos no relatório na tabela DADOS\_ETL.

INSERT INTO dados\_etl

SELECT usuario, porta, protocolo, COUNT(\*) ACESSOS

FROM dados

GROUP BY usuário, porta, protocolo

HAVING acessos > 9

ORDER BY usuario ASC, port DESC, protocolo ASC, acessos DESC;

SELECT \* FROM empregado\_etl WHERE rownum <6;

1. Usando comandos SQOOP, verifique quais são as tabelas existentes no seu *schema* do banco de dados. Lembre-se de informar o seu RM e sua SENHA.
2. Usando comandos SQOOP importe os dados os dados da tabela DADOS\_ETL. Defina que os dados serão gravados em um diretório do HADOOP de nome DADOS. Importe apenas os dados dos empregados com ACESSOS > 15. Faça a importação usando 8 mappers.
3. Liste o conteúdo dos arquivos importados