

Trabalho Final - Computação em Nuvem e Segurança de Redes

Aluna: Letícia Trein Medeiros

Este trabalho foi elaborado utilizando o WSL (Ubuntu) no Windows 11; também foi utilizado o docker para baixar imagens prontas de interesse.

Seguem principais passos seguidos e parte dos códigos utilizados:

- **Baixar arquivo via comando e desempacotamento de arquivo zip**

```
curl
http://www.labsom.ct.utfpr.edu.br/mrosa/computacao-nuvem-seguranca-redes/trabalho-final/dataset.zip -o dataset.zip
unzip dataset.zip
```

- **Rodar ambiente spark via docker**

```
docker run -v ./dataset:/dataset -it spark /opt/spark/bin/spark-shell
```

- **Ler arquivo de dados**

```
val df_raw = spark.read.option("delimiter", ";").csv("/dataset")
```

- **Organizar dos dados para melhor seleção**

```
val column_names = Seq( "copel_code", "contador unico da leitura no
arquivo", "data e hora da leitura", "mnemonico da empresa", "identificador
do smartmeter", "medida 1", "medida 2", "medida 3", "medida 4", "medida 5",
"medida 6", "medida 7", "medida 8", "medida 9", "medida 10", "medida 11",
"medida 12", "medida 13", "medida 14", "medida 15", "medida 16", "medida
17", "medida 18", "medida 19", "resumo criptografico da linha" )
```

```
val df_named = df_raw.toDF(column_names: _*)
```

Tarefas

1) Consolidar o consumo total, gerando a soma das medidas 14, 16 e 17 separadamente.

```
val df = df_named.select(col("mnemonico da empresa"), col("medida
14").cast("double"), col("medida 16").cast("double"), col("medida
17").cast("double"))
```

```
val resp_1 = df.select(sum("medida 14"), sum("medida 16"), sum("medida 17"))
```

resp_1.show

| | sum(medida 14) | sum(medida 16) | sum(medida 17) |
|--|----------------|--------------------|--------------------|
| | 240463.15 | 240001.93999999997 | 240256.71999999956 |

2) Consolidar o consumo POR CLIENTE, gerando a soma dos campos 14, 16 e 17 separadamente para cada uma das empresas/clientes.

```
val resp_2 = df.groupBy("mnemonico da empresa").agg(sum("medida 14").alias("sum_medida_14"), sum("medida 16").alias("sum_medida_16"), sum("medida 17").alias("sum_medida_17"))
resp_2.show
```

| mnemonico da empresa | sum_medida_14 | sum_medida_16 | sum_medida_17 |
|----------------------|---------------------|---------------------|---------------------|
| company_00018 | 5088.5700000000008 | 4875.3300000000005 | 4955.5400000000006 |
| company_00001 | 19985.2999999999996 | 20235.7599999999984 | 19629.3400000000015 |
| company_00017 | 4892.2499999999995 | 4992.28 | 4950.8899999999995 |
| company_00000 | 20302.9200000000006 | 19970.62 | 20076.0199999999986 |
| company_00012 | 9799.8899999999985 | 9878.9800000000007 | 10094.43 |
| company_00006 | 15033.619999999998 | 15118.649999999996 | 15045.039999999999 |
| company_00015 | 4966.5900000000002 | 5119.8299999999996 | 5020.8499999999996 |
| company_00014 | 10137.4900000000002 | 10052.1200000000003 | 10077.300000000001 |
| company_00019 | 4891.5899999999965 | 5007.1900000000003 | 4977.1299999999965 |
| company_00009 | 14995.149999999994 | 14710.9000000000014 | 15269.2000000000012 |
| company_00002 | 20101.2400000000023 | 19947.8300000000005 | 20006.5800000000013 |
| company_00016 | 5056.3099999999999 | 5196.1299999999999 | 4975.1999999999996 |
| company_00004 | 15045.4600000000005 | 14858.38 | 14748.0500000000007 |
| company_00011 | 9831.9399999999999 | 10001.27 | 10379.1100000000001 |
| company_00007 | 14980.1000000000002 | 15014.3900000000001 | 15094.84 |
| company_00005 | 14950.67 | 14943.479999999996 | 15025.7600000000006 |
| company_00003 | 14908.7100000000008 | 15082.6300000000006 | 15062.3500000000013 |
| company_00013 | 10166.3 | 10048.269999999997 | 9892.8999999999989 |
| company_00010 | 10051.25 | 10046.2300000000005 | 9984.7099999999994 |
| company_00008 | 15277.8 | 14901.669999999993 | 14991.4800000000009 |