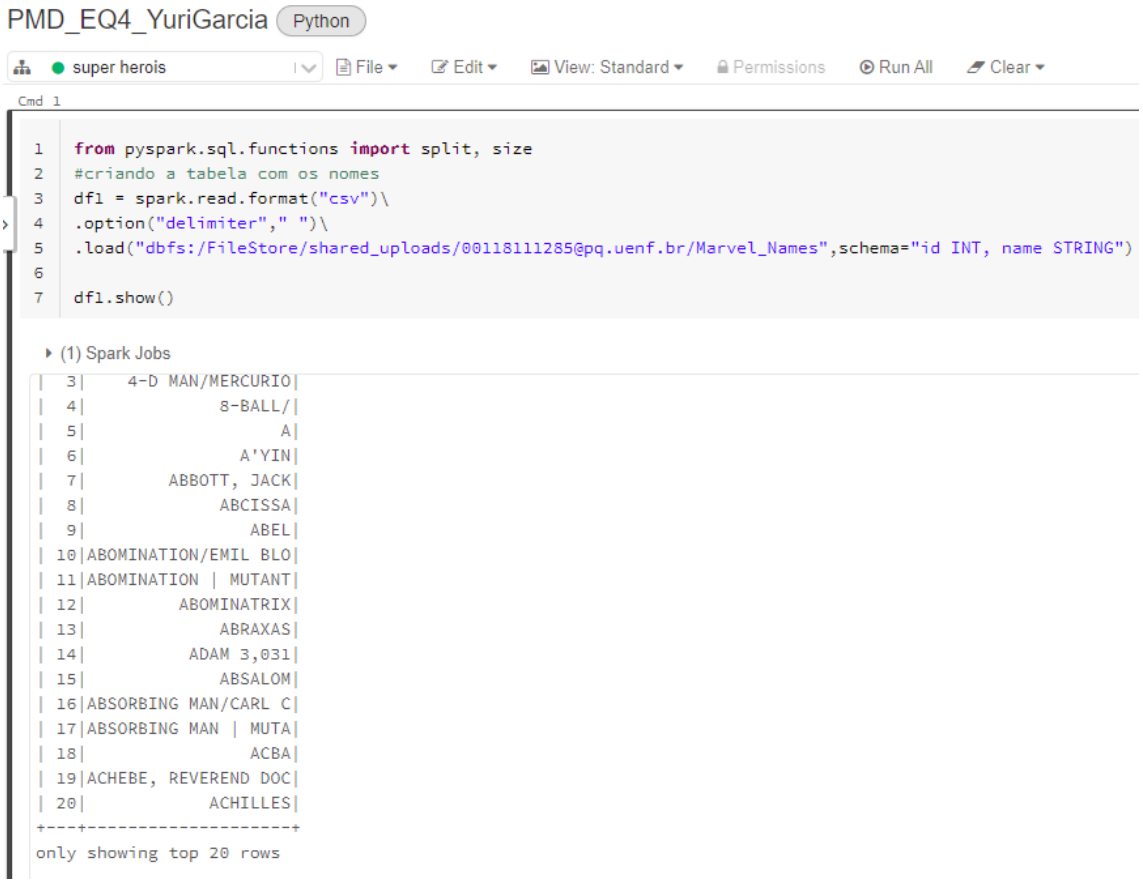
Problema 1 - REDE SOCIAL DE SUPER-HERÓIS

Dado os arquivos de super-heróis a seguir, crie um *dataframe* usando PySpark com o *ranking* dos super-heróis usando o critério de quem tem mais conexões fica no topo da lista.



```
#criando a tabela com as conexões
df2 = spark.read.format("csv")\
.load("dbfs:/FileStore/shared_uploads/00118111285@pq.uenf.br/Marvel_Graph")

df2.show()
```

▸ (2) Spark Jobs

```
+--------------------+
|                 _c0|
+--------------------+
|5988 748 1722 375...|
|5989 4080 4264 44...|
|5982 217 595 1194...|
|5983 1165 3836 43...|
|5980 2731 3712 15...|
|5981 3569 5353 40...|
|5986 2658 3712 26...|
|5987 2614 5716 17...|
|5984 590 4898 745...|
|5985 3233 2254 21...|
|6294 4898 1127 32...|
|270 2658 3003 380...|
|271 4935 5716 430...|
|272 2717 4363 408...|
|273 1165 5013 511...|
|274 3920 5310 402...|
|275 4366 3373 158...|
|276 2277 5251 480...|
```

```
#separando o id do heroi que possui as conexões
df3 = df2.select(split(df2["_c0"], ' ').getItem(0).alias("id"),
                 split(df2["_c0"], ' ').alias("conections"))
display(df3)

df3 = df3.withColumn("num_conections", size("conections") - 2)#contando o numero de conexões de cada heroi

df3 = df3.drop("conections")

display(df3)
```

▸ (2) Spark Jobs

	id	conections
1	5988	▸ ["5988", "748", "1722", "3752", "4655", "5743", "1872", "3413", "5527", "6368", "6085", "4319", "4728", "1636", "2397", "3364", "4001", "1614", "1819", "1585", "732", "2660", "3952", "2507", "3891", "2070", "2239", "2602", "612", "1352", "5447", "4548", "1596", "5488", "1605", "5517", "11", "479", "2554", "2043", "17", "865", "4292", "6312", "473", "534", "1479", "6375", "4456", ""]
2	5989	▸ ["5989", "4080", "4264", "4446", "3779", "2430", "2297", "6169", "3530", "3272", "4282", "6432", "2548", "4140", "185", "105", "3878", "2429", "1334", "4595", "2767", "3956", "3877", "4776", "4946", "3407", "128", "269", "5775", "5121", "481", "5516", "4758", "4053", "1044", "1602", "3889", "1535", "6038", "533", "3986", ""]
3	5982	▸ ["5982", "217", "595", "1194", "3308", "2940", "1815", "794", "1503", "5197", "859", "5096", "6039", "2664", "651", "2244", "528", "284", "1449", "1097", "1172", "1092", "108", "3405", "5204", "387", "4607", "4545", "3705", "4930", "1805", "4712", "4404", "247", "4754", "4427", "1845", "536", "5795", "5978", "533", "3984", "6056", ""]
4	5983	▸ ["5983", "1165", "3836", "4361", "1282", "716", "4289", "4646", "6300", "5084", "2397", "4454", "1913", "5861", "5485", ""]
5	5980	▸ ["5980", "2731", "3712", "1587", "6084", "2472", "2546", "6313", "875", "859", "323", "2664", "1469", "522", "2506", "2919", "2423", "3624", "5736", "5046", "1787", "5776", "3245", "3840", "2399", ""]
6	5981	▸ ["5981", "3569", "5353", "4087", "2653", "2058", "2218", "5354", "5306", "3135", "4088", "4869", "2958", "2959", "5732", "4076", "4155", "291", ""]
	5986	▸ ["5986", "2658", "3712", "2650", "1265", "133", "4024", "6313", "3120", "6066", "3546", "403", "545", "4860", "4337", "2295",

Truncated results, showing first 1000 rows

Truncated results, showing first 1000 rows.

Click to re-execute with maximum result limits.

	id	num_conections
1	5988	48
2	5989	40
3	5982	42
4	5983	14
5	5980	24
6	5981	17
7	5986	142

Truncated results, showing first 1000 rows.

Click to re-execute with maximum result limits.

```
#agrupando os id de herois que se repetem e somando as suas conexões
df_group = df3.groupBy("id").agg({"num_conections": 'sum'}).withColumnRenamed("sum(num_conections)", "num_conections")

display(df_group)

#juntando as informações numa unica tabela e ordenando
df4 = df1.join(df_group, df1["id"] == df_group["id"], "inner") \
             .select(df1["id"], df1["name"], df_group["num_conections"]) \
             .orderBy("num_conections", ascending=False)

display(df4)

```

▸ (4) Spark Jobs

	id	num_conections
1	691	6
2	1159	11
3	3959	142
4	1572	35
5	2294	14
6	1090	4
7	3606	171

Truncated results, showing first 1000 rows.

Click to re-execute with maximum result limits.

	id	name	num_conections
1	859	CAPTAIN AMERICA	1933
2	5306	SPIDER-MAN/PETER PAR	1741
3	2664	IRON MAN/TONY STARK	1528
4	5716	THING/BENJAMIN J. GR	1426
5	6306	WOLVERINE/LOGAN	1394
6	3805	MR. FANTASTIC/REED R	1386
7	2557	HUMAN TORCH/JOHNNY S	1371

Truncated results, showing first 1000 rows.

https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/2451291788623867/3119772944382309/5643440177508145/latest.html