

Lab 2.1: Primeiro contato com o Spark

- Vamos iniciar nosso primeiro contato acessando o spark-shell
- Estando no prompt do sistema operacional digite:
>spark-shell

```
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Spark context Web UI available at http://ip-172-31-45-99.us-east-2.compute.internal:4040
Spark context available as 'sc' (master = local[*], app id = local-1599678889208).
Spark session available as 'spark'.
Welcome to


  ____      _
 / ___|  _ \| | | |
 \___ \ |_) | | | |
  ___) ||  __| | | |
 /____||_| |_| |_| | |_|_|

version 2.4.6

Using Scala version 2.11.12 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_261)
Type in expressions to have them evaluated.
Type :help for more information.

scala> 
```

- Acesse <http://<sua>maquina>:4040>



JobsStagesStorageEnvironmentExecutors

Spark Jobs (?)

User: ec2-user
Total Uptime: 1,6 min
Scheduling Mode: FIFO

▼ [Event Timeline](#)
☐ Enable zooming

Executors						
<input type="checkbox"/> Added						
<input type="checkbox"/> Removed						
Jobs						
<input type="checkbox"/> Succeeded						
<input type="checkbox"/> Failed						
<input type="checkbox"/> Running						
		000	100	200	300	400
		19:14:48				

- Vá na aba EXECUTOR e veja qual é o Executor ID da tabela de Executors

- Veja quantos cores tem a sua máquina
- Na aba ENVIRONMENT e veja qual o conteúdo da seção Spark Properties
- Digite `sc.isLocal` e veja o resultado
- Digite `sc.<de um tab>` e veja o resultado

Lab 2.2: Primeiro contato com o Spark

- Entre no spark-shell
- Carregue o arquivo "README.txt"
`> val myfile = sc.textFile("README.txt")`
- Imprima o arquivo
`> myfile.foreach(println)`
- Olhe para o Spark UI e identifique quantas task foram usadas para executar a tarefa (coluna Tasks da aba Jobs)? Esse número lhe diz algo?
- Quantas linhas tem o arquivo?
`> myfile.count()`
- Pegue apenas as duas primeiras linhas e imprima
`> myfile.take(2).foreach(println)`
- Pegue todas as linhas do arquivo
`> myfile.collect()`
- Saia do spark
`> :quit`