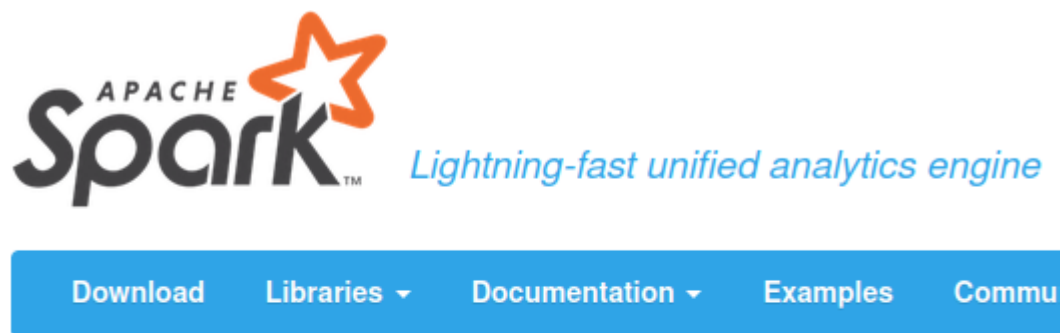


1) Instale a última versão do Apache Spark em Python via Pip



## Download Apache Spark™

1. Choose a Spark release: 3.1.2 (Jun 01 2021) ▾

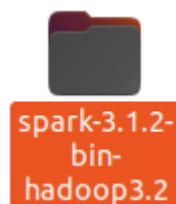
2. Choose a package type: Pre-built for Apache Hadoop 3.2 and later ▾

3. Download Spark: [spark-3.1.2-bin-hadoop3.2.tgz](#)

- Site do Apache Spark, baixando .tar.gz



- Arquivo comprimido do spark



- Descompactado



2) Usando o Spark Shell faça um programa que conte as palavras de um arquivo README.md da pasta home do Spark – Somente da Seção Basics do link

```
>>> textFile = spark.read.text("apps/spark-3.1.2-bin-hadoop3.2/README.md")
>>> textFile.count()
108
>>> textFile.first()
Row(value='# Apache Spark')
>>> linesWithSpark = textFile.filter(textFile.value.contains("Spark"))
>>> textFile.filter(textFile.value.contains("Spark")).count()
19
```

- Tutorial do Spark, parte Basics.