



<https://spark.apache.org/docs/latest/submitting-applications.html>

Poner en marcha aplicaciones Spark en el cluster

programa.py (para Spark 2.4)

```
from pyspark.sql import SparkSession

spark = SparkSession.builder.appName("miPrograma").getOrCreate()

texto = "archivo.txt" #archivo de texto del sistema
datos = spark.read.text(texto).cache()

numAs = datos.where(logData.value.contains('a')).count()
numBs = datos.where(logData.value.contains('b')).count()

print("Líneas con a: %i, líneas con b: %i" % (numAs, numBs))

spark.stop()
```

programa.py (para Spark 1.6)

```
from pyspark import SparkContext
from pyspark.sql import SQLContext

sc = SparkContext("local", "miPrograma")
sqlContext = SQLContext(sc)

texto = "archivo.txt" #archivo de texto del sistema
datos = sqlContext.read.text(texto).cache()

numAs = datos.where(logData.value.contains('a')).count()
numBs = datos.where(logData.value.contains('b')).count()

print("Líneas con a: %i, líneas con b: %i" % (numAs, numBs))

spark.stop()
```

Con configuración por defecto

```
spark-submit programa.py
```

Configuración utilizando Yarn

```
spark-submit --master yarn-client programa.py
```