

# Instalando o Spark

Faça o download de:

<https://spark.apache.org/downloads.html> - versão pre-built

<https://www.continuum.io/downloads> - Python 2 (instale como administrador no windows)

# Instalando o Spark

No prompt de comando digite:

```
pip install pyspark
```

# Instalando o Spark

Descompacte o spark em um diretório de sua preferência  
(no Windows tem que ser c:\opt\spark)

# Instalando o Spark

Crie as seguintes variáveis de ambiente:

SPARK\_HOME C:\opt\spark\spark-2.2.0-bin-hadoop2.7

HADOOP\_HOME C:\opt\spark\spark-2.2.0-bin-hadoop2.7

PYSPARK\_DRIVER\_PYTHON jupyter

PYSPARK\_DRIVER\_PYTHON\_OPTS notebook

# Instalando o Spark

Adicione:

`C:\opt\spark\spark-2.2.0-bin-hadoop2.7\bin`

Ao seu path e reinicie o computador.

# Instalando o Spark

Abra um terminal / prompt e digite:

```
Pyspark --master local[2]
```

Para abrir o Jupyter notebook com suporte ao Spark utilizando 2 cores da sua máquina.

# Instalando o Spark

Pode ser necessário a inserção e execução de uma célula com o seguinte comando:

```
sc = SparkContext.getOrCreate()
```

no início de cada notebook.