

Ambiente de Trabajo

Watson Studio

Watson Studio democratiza la ciencia de datos y la IA para impulsar la innovación en su negocio. Con un conjunto de herramientas para todos los niveles de habilidad, todos pueden colaborar para preparar, analizar y modelar datos. Puede escribir código Python o R en cuadernos, codificar visualmente en un lienzo gráfico o crear modelos automáticamente.

<https://cloud.ibm.com/catalog/services/watson-studio>

1 usuario autorizado

Límite mensual de 10 horas de unidad de capacidad

Entorno = # de unidades de capacidad requeridas por hora

- 1 CPU virtual + 4 GB de RAM = 0,5
- 2 CPU virtuales + 8 GB de RAM = 1
- 4 CPU virtuales + 16 GB de RAM = 2
- Optimización de decisiones + Watson NLP = Medio ambiente + 5

Anaconda Distribution

La distribución de código abierto Anaconda es la forma más fácil de realizar ciencia de datos Python / R y aprendizaje automático en Linux, Windows y Mac OS X.



Instalación

<https://www.scribd.com/document/469169833/instalando-spark-con-Anaconda-Manualito-pdf>

<https://jarroba.com/instalar-apache-spark-en-cualquier-sistema-operativo-y-aprender-a-programarlo-con-python-scala-o-java/>

Anaconda Installers

Windows 

Python 3.8

64-Bit Graphical Installer (466 MB)

32-Bit Graphical Installer (397 MB)

MacOS 

Python 3.8

64-Bit Graphical Installer (462 MB)

64-Bit Command Line Installer (454 MB)

Linux 

Python 3.8

64-Bit (x86) Installer (550 MB)

64-Bit (Power8 and Power9) Installer (290 MB)

Instalación Spark

Anaconda

Aunque Spark fue diseñado para correr en clusters, es fácil de ejecutarlo localmente en una máquina; solo se necesita tener Java instalado y la variable de entorno `JAVA_HOME` que apunta a la instalación de Java.

Para usar en JupyterLab:

- Descargue e instale el JDK de Java versión 8
 - En ubuntu: `sudo apt install openjdk-8-jdk`
 - En ubuntu: `export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64`
- Ejecutar en consola anaconda e instalar Pyspark
 - `conda install -c conda-forge pyspark`
- Exportar variable de entorno `JAVA_HOME`

Externa

Si se desea utilizar la versión más nueva disponible en la pagina oficial de spark.

Para usar en JupyterLab:

- Descargue e instale el JDK de Java versión 8
 - En ubuntu: `sudo apt install openjdk-8-jdk`
 - En ubuntu: `export`
`JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64`
- Ejecutar en consola anaconda e instalar Py4J
 - `conda install -c conda-forge py4j`
- Descargar spark <https://spark.apache.org/downloads.html>
- Descomprimir spark en alguna carpeta
- Agregar ruta de pyspark
- Exportar variable de entorno JAVA_HOME

Rutas

Se mencionó que se debe crear una variable de entorno JAVA_HOME así como agregar la ruta a la instalación de spark (en caso que se desee conectar remotamente), dentro de python se puede hacer fácilmente:

```
import os
# en windows la diagonal es \\ en linux y mac es /
os.environ["JAVA_HOME"] = "C:\\Program Files\\Java\\jdk1.8.0_241"
```

```
import sys
# es la ruta donde se descomprime spark
sys.path.append("/home/octavio.renteria/spark/python")
```