

# Tarea 1

Luis Felipe Cubero Vargas

Guia de Instalación

## Resumen

- Instalar miniconda
- Instalar java
- Instalar java development kit
- Bajar Spark
- Crear ambiente
- Instalar pyspark, Jupiter notebook, pandas, numpy, matplotlib,
- Instalar FindSpark (invocar FindSpark)

# Instalar mini conda

<https://docs.conda.io/en/latest/miniconda.html>

## Miniconda

Miniconda is a free minimal installer for conda. It is a small, bootstrap version of Anaconda that includes only conda, Python, the packages they depend on, and a small number of other useful packages, including pip, zlib and a few others. Use the `conda install command` to install 720+ additional conda packages from the Anaconda repository.

[See if Miniconda is right for you.](#)

## Windows installers

### Windows

Python version	Name	Size	SHA256 hash
Python 3.7	<a href="#">Miniconda3 Windows 64-bit</a>	51.5 MiB	<code>f18060cc0bb50ae75e4d602b7ce35197c8e31e81288d069b758594f1bb46ab45</code>
	<a href="#">Miniconda3 Windows 32-bit</a>	54.0 MiB	<code>7c30778941d2bba03531ba269a78a108b01fa366530290376e7c3b467f3c66ba</code>
Python 2.7	<a href="#">Miniconda2 Windows 64-bit</a>	50.9 MiB	<code>8647c54058f11842c37854edeff4d20bc1fbdad8b88d9d34d76fda1630e64846</code>
	<a href="#">Miniconda2 Windows 32-bit</a>	48.7 MiB	<code>0d106228d6a4610b599df965dd6d9bb659329a17e3d693e3274b20291a7c6f94</code>

## MacOSX installers

### MacOSX

Python version	Name	Size	SHA256 hash
Python 3.7	<a href="#">Miniconda3 MacOSX 64-bit bash</a>	49.4 MiB	<code>5cf91dde8f6024061c8b9239a1b4c34380238297adbdb9ef2061eb9d1a7f69bc</code>
	<a href="#">Miniconda3 MacOSX 64-bit pkg</a>	59.8 MiB	<code>9927f1de5151a1a6431b02846fbca089e8b97a55a244f02ffc3207522092907b</code>
Python 2.7	<a href="#">Miniconda2 MacOSX 64-bit bash</a>	39.4 MiB	<code>0db8f4037e40e13eb1d2adc89e054dfb165470cc77be45ef2bf9cb31c8b72f39</code>
	<a href="#">Miniconda2 MacOSX 64-bit pkg</a>	47.8 MiB	<code>fcc30b2e18f7a292b34b2e24ad855786a66423f860157fa2b77e48b6392f0abb</code>

# Descargar e instalar java

<https://www.java.com/es/download/>

## Descargar Java para Mac OS X

**Recomendado Version 8 Update 231 (Tamaño de archivo: 79.91 MB)**

Fecha de versión: 15 de octubre de 2019



### Actualización importante de la licencia de Oracle Java

**La licencia de Oracle Java ha cambiado para las versiones publicadas a partir del 16 de abril de 2019.**

El nuevo [acuerdo de licencia de Oracle Technology Network para Oracle Java SE](#) es sustancialmente diferente a las licencias de Oracle Java anteriores. La nueva licencia permite ciertos usos, como el uso personal y de desarrollo, sin coste alguno (aunque podría haber otros usos autorizados en licencias de Oracle Java anteriores que ya no estén disponibles). Revise las condiciones con atención antes de descargar y utilizar este producto. Puede consultar las preguntas frecuentes [aquí](#).

La licencia comercial y el soporte están disponibles con una [suscripción de Java SE](#) de bajo coste.

Oracle también ofrece la última versión de OpenJDK con la [licencia pública general](#) de código abierto en [jdk.java.net](http://jdk.java.net).

**Solo es necesario  
descargar e ejecutar el  
instalador**



Hemos detectado que está utilizando Google Chrome y quizá no pueda utilizar el complemento Java desde este explorador. A partir de la versión 42 (abril 2015), Chrome ha desactivado la el método estándar por el que los exploradores soportan plugins. [Más información](#)

**Aceptar e iniciar descarga  
gratuita**

## Descargar e instalar java development kit

<https://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>

Java SE Development Kit 8u231		
You must accept the <a href="#">Oracle Technology Network License Agreement for Oracle Java SE</a> to download this software.		
<input type="radio"/> Accept License Agreement <input checked="" type="radio"/> Decline License Agreement		
Product / File Description	File Size	Download
Linux ARM 32 Hard Float ABI	72.9 MB	<a href="#">jdk-8u231-linux-arm32-vfp-hflt.tar.gz</a>
Linux ARM 64 Hard Float ABI	69.8 MB	<a href="#">jdk-8u231-linux-arm64-vfp-hflt.tar.gz</a>
Linux x86	170.93 MB	<a href="#">jdk-8u231-linux-i586.rpm</a>
Linux x86	185.75 MB	<a href="#">jdk-8u231-linux-i586.tar.gz</a>
Linux x64	170.32 MB	<a href="#">jdk-8u231-linux-x64.rpm</a>
Linux x64	185.16 MB	<a href="#">jdk-8u231-linux-x64.tar.gz</a>
Mac OS X x64	253.4 MB	<a href="#">jdk-8u231-macosx-x64.dmg</a>
Solaris SPARC 64-bit (SVR4 package)	132.98 MB	<a href="#">jdk-8u231-solaris-sparcv9.tar.Z</a>
Solaris SPARC 64-bit	94.16 MB	<a href="#">jdk-8u231-solaris-sparcv9.tar.gz</a>
Solaris x64 (SVR4 package)	133.73 MB	<a href="#">jdk-8u231-solaris-x64.tar.Z</a>
Solaris x64	91.96 MB	<a href="#">jdk-8u231-solaris-x64.tar.gz</a>
Windows x86	200.22 MB	<a href="#">jdk-8u231-windows-i586.exe</a>
Windows x64	210.18 MB	<a href="#">jdk-8u231-windows-x64.exe</a>

**Seleccionar el botón de:  
Accept License Agreement**

**Solo es necesario  
descargar e ejecutar el  
instalador**

# Descargar Apache Spark

<https://spark.apache.org/downloads.html>



Download Libraries ▾ Documentation ▾ Examples Community ▾ Developers ▾

## Download Apache Spark™

1. Choose a Spark release:  ←
2. Choose a package type:
3. Download Spark: [spark-2.4.4-bin-hadoop2.7.tgz](#) ←
4. Verify this release using the 2.4.4 [signatures](#), [checksums](#) and [project release KEYS](#).

Note that, Spark is pre-built with Scala 2.11 except version 2.4.4, which is pre-built with Scala 2.12.

## Latest Preview Release

Preview releases, as the name suggests, are releases for previewing upcoming features. Unlike nightly packages, preview releases have been audited by the project's management committee to satisfy the legal requirements of Apache Software Foundation's release policy. Preview releases are not meant to be functional, i.e. they can and highly likely will contain critical bugs or documentation errors. The latest preview release is Spark 3.0.0-preview, published on Nov 6, 2019. You can select and download it above.

## Link with Spark

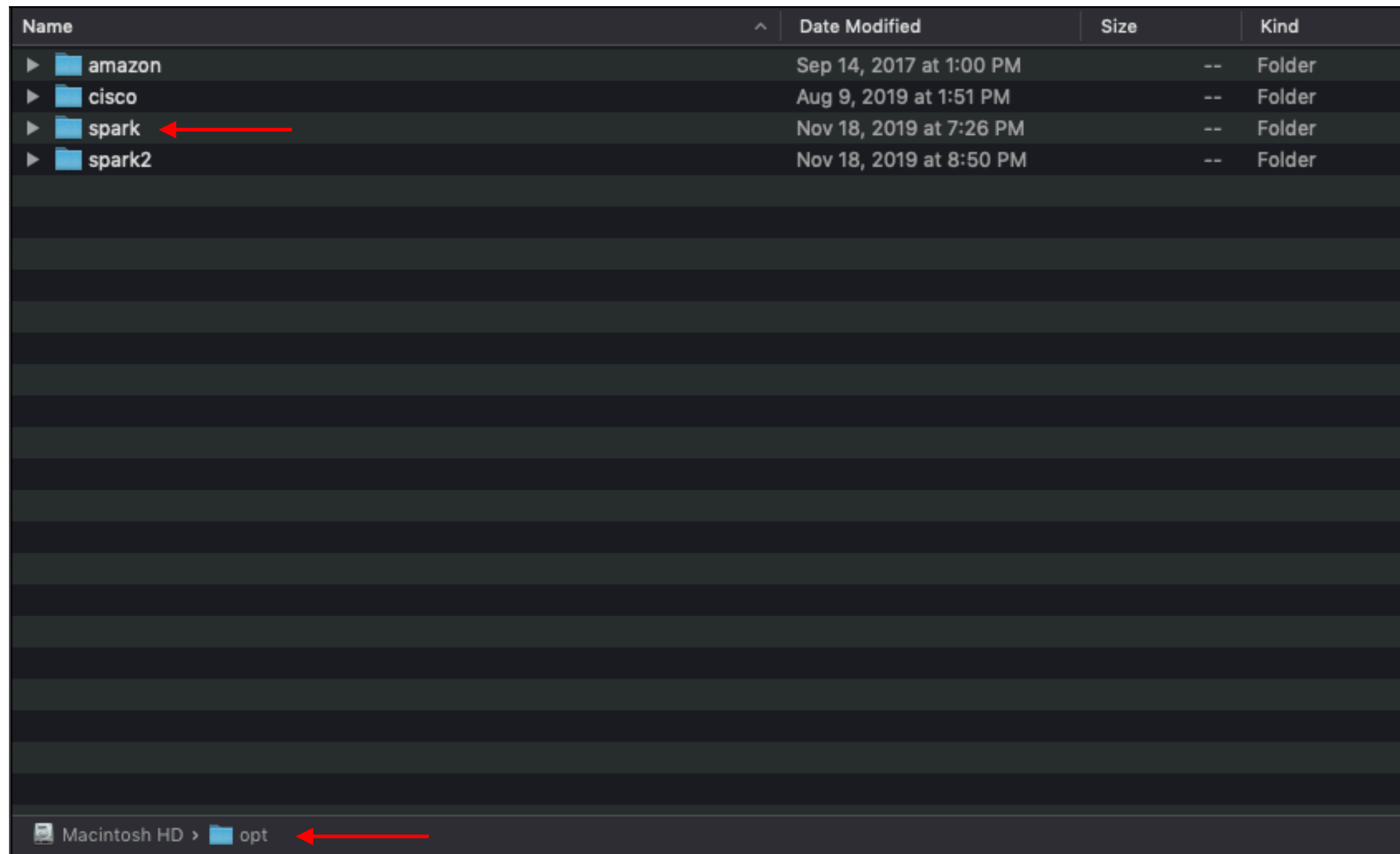
Spark artifacts are [hosted in Maven Central](#). You can add a Maven dependency with the following coordinates:

```
groupId: org.apache.spark
artifactId: spark-core_2.11
version: 2.4.4
```

Una vez descargado, se descomprime el archivo y se copia el interior del filtro en un folder con el nombre: apache

**En el caso de ser mac la carpeta se crea dentro del directorio opt**

**En el caso de Windows crear dentro del Disco local C**



Name	^	Date Modified	Size	Kind
▶ amazon		Sep 14, 2017 at 1:00 PM	--	Folder
▶ cisco		Aug 9, 2019 at 1:51 PM	--	Folder
▶ spark		Nov 18, 2019 at 7:26 PM	--	Folder
▶ spark2		Nov 18, 2019 at 8:50 PM	--	Folder

Macintosh HD > opt

**El siguiente paso es crear un ambiente virtual de miniconda**

En la terminal escribir:

conda create -n yourenvname **python**=x.x

Luego activamos el ambiente:

conda activate yourenvname

Y luego con pip install instalamos:

Jupyter notebook, pyspark, numpy, pandas, matplotlib

## Por último dentro de una notebook de jupyter realizamos la siguiente prueba

Ejecutando el código de a continuación en donde no haya error nos aseguramos que la instalación fue correcta

```
1 from pyspark.sql import SparkSession
2
3 spark=SparkSession.builder.appName('data_processing').getOrCreate()
4
5 import pyspark.sql.functions as F
6 from pyspark.sql.types import *
```

```
1 import findspark
2 findspark.init('/opt/spark') ← Aquí poner el path donde se guardo spark
3
4 from datetime import datetime
5 from pyspark.sql import SparkSession
6 from pyspark.sql.functions import col, date_format, udf
7 from pyspark.sql.types import DateType
8
```