



Save to Evernote



# Running BigDL (Python) on Spark on your Local Machine

Updated Aug 31, 2017

This tutorial will teach you how to set up BigDL on your Mac and run the MNIST digit recognition BigDL Python application.

## Prerequisites:


- Mac/Linux OS
- Java latest version (if you don't have it run: `conda install java`)
- Jupyter Notebook (your pyspark driver)
- Python 2.7
- scikit-learn, scipy, numpy, pandas, matplotlib (pip install these in your environment)

## STEP 1: Set up your BigDL folder.


Create a new folder to store the necessary bigdl/spark requirements. We are going to put number of things in there:

- 1) BigDL 0.2.0 Package
- 2) Spark 2.1.1
- 3) `start_notebook.sh`
- 4) `cnn.ipynb`
- 5) `utils.py`

**#1:** Download the built BigDL Package under Release 0.2.0, for Spark 2.1.1 here: <https://bigdl-project.github.io/master/#release-download/>. There is a version for Mac and Linux. This has all BigDL dependencies and python files too.



master ↕



Fork on GitHub

Overview

▼ Releases

Download ▼

Nightly Build

Release 0.2.0

Release 0.1.1

Release 0.1.0

Documentation

Getting Started

► Scala User Guide

► Python User Guide

► Programming Guide

► API Guide

Powered by

Known Issues

Other Resource

Spark 1.6.2	<a href="#">download</a>	<a href="#">download</a>	<a href="#">download</a>
Spark 2.0.2	<a href="#">download</a>	<a href="#">download</a>	<a href="#">download</a>
Spark 2.1.1	<a href="#">download</a>	<a href="#">download</a>	<a href="#">download</a>

## Release 0.2.0

	Linux x64	Mac	Win64
Spark 1.5.2	<a href="#">download</a>	<a href="#">download</a>	<a href="#">download</a>
Spark 1.6.2	<a href="#">download</a>	<a href="#">download</a>	<a href="#">download</a>
Spark 2.0.2	<a href="#">download</a>	<a href="#">download</a>	<a href="#">download</a>
Spark 2.1.1	<a href="#">download</a>	<a href="#">download</a>	<a href="#">download</a>

## Release 0.1.1

	Linux x64	Mac
Spark 1.5.1	<a href="#">download</a>	<a href="#">download</a>

## Release 0.2.0

## Release 0.1.1

**#2:** Download Spark version 2.1.1 here: <https://spark.apache.org/downloads.html> This unzips into a folder with all spark related contents.

[Download](#)[Libraries ▾](#)[Documentation ▾](#)[Examples](#)[Community ▾](#)[Developers ▾](#)

## Download Apache Spark™

- Choose a Spark release:
- Choose a package type:
- Choose a download type:
- Download Spark: [spark-2.1.1-bin-hadoop2.7.tgz](#)
- Verify this release using the [2.1.1 signatures and checksums](#) and [project release KEYS](#).

*Note: Starting version 2.0, Spark is built with Scala 2.11 by default. Scala 2.10 users should download the Spark source package and build with Scala 2.10 support.*

1. Choose a Spark release: **2.1.1 (May 02 2017)**
2. Choose a package type: **Pre-built for Apache Hadoop 2.7 and later**
3. Choose a download type: **Direct Download**
4. Download Spark: [spark-2.1.1-bin-hadoop2.7.tgz](#)
5. Verify this release using the [2.1.1 signatures and checksums](#) and [project release KEYS](#).

### #3: Download this starter script.

**Note:** Remember to change **<path to folder you created>** after the **SPARK\_HOME** and **BigDL\_HOME** environment variables to the absolute path to your bigdl folder you just created. This is the script that you will run to start a spark session and use BigDL. It sets the correct parameters and environment for your spark session to run BigDL applications.

#4: Download cnn.ipynb here.

 **cnn.ipynb**  
132.7 KB

**Note:** There are more example notebooks in BigDL's tutorial repository. Feel free to clone and download these example notebooks here: [https://github.com/intel-analytics/BigDL-Tutorials/tree/branch-0.2/notebooks/neural\\_networks](https://github.com/intel-analytics/BigDL-Tutorials/tree/branch-0.2/notebooks/neural_networks)

#5: Download utils.py here.

 **utils.py**  
1.2 KB

#6 Verify that all items listed above are in your bigdl folder. You are ready to run your application now.

## STEP 2: Run your BigDL application

#1: On your terminal, make sure you are in the directory you created and run the command `./start_notebook.sh`. This starts Jupyter Notebook and creates your Spark Context (sc).

```
[I 12:32:23.101 NotebookApp] 0 active kernels
[I 12:32:23.101 NotebookApp] The Jupyter Notebook is running at: http://[all ip addresses on your system]:8889/?token=f84c615908a1ad189d717aa5ee31799d45fef2b4587ab252
[I 12:32:23.101 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 12:32:23.102 NotebookApp]

Copy/paste this URL into your browser when you connect for the first time,
to login with a token:
http://localhost:8889/?token=f84c615908a1ad189d717aa5ee31799d45fef2b4587ab252
```

#2: Copy and Paste the URL into your browser to start Jupyter Notebook.

**Note:** If you need to, change the access permission of start\_notebook.sh script using this command: `chmod +x <your path>/start_notebook.sh`

#3: Run cnn.ipynb file in Jupyter Notebooks **using a kernel that runs Python 2.7**. It should match the output on the BigDL tutorials.

**Note:** If your default kernel in Jupyter Notebook is Python 3, you have to change it to Python 2. You can create a virtual environment that uses Python 2.7 by following this guide <https://uoa-ereseearch.github.io/ereseearch-cookbook/recipe/2014/11/20/conda/>

**Note:** You can only run **one** .ipynb at a time because you have only one spark session, remember to shutdown the rest. Also make sure your kernel is running Python 2 (Python 3 is still being tested).

Thats it!



(

Add a save button to the internet

Get Web Clipper