

Installation Apache Spark 3.0.1

Spark can be used for processing batches of data, real-time streams, machine learning, and ad-hoc query.

Install Scala (master and slaves)

1. **sudo apt-get install scala**
2. **scala -version**

From here do only on master

3. **wget https://ftp.wayne.edu/apache/spark/spark-3.0.1/spark-3.0.1-bin-hadoop3.2.tgz**
4. **tar -xzf spark-3.0.1-bin-hadoop3.2.tgz**
5. **sudo mv spark-3.0.1-bin-hadoop3.2 /usr/local/spark**
- 6.
7. **sudo gedit ~/.bashrc**
 - export SPARK_HOME=/usr/local/spark/
 - export PATH=\$PATH:\$SPARK_HOME/bin
 - export LD_LIBRARY_PATH=\$HADOOP_HOME/lib/native:\$LD_LIBRARY_PATH
8. **source ~/.bashrc**

Integrate Spark with YARN

To communicate with the YARN Resource Manager, Spark needs to be aware of your Hadoop configuration. This is done via the HADOOP_CONF_DIR environment variable. The SPARK_HOME variable is not mandatory, but is useful when submitting Spark jobs from the command line.

9. **mv \$SPARK_HOME/conf/spark-defaults.conf.template \$SPARK_HOME/conf/spark-defaults.conf**
10. **sudo gedit \$SPARK_HOME/conf/spark-defaults.conf**
 - spark.master yarn
 - spark.driver.memory 512m
 - spark.yarn.am.memory 512m
 - spark.executor.memory 512m

Spark is now ready to interact with your YARN cluster.

Spark History server

11. **sudo gedit \$SPARK_HOME/conf/spark-defaults.conf**
 - spark.eventLog.enabled true
 - spark.eventLog.dir hdfs://master:9000/spark-logs
 - spark.history.provider
org.apache.spark.deploy.history.FsHistoryProvider
 - spark.history.fs.logDirectory hdfs://master:9000/spark-logs
 - spark.history.fs.update.interval 3s
 - spark.history.ui.port 18080

Create the log directory in HDFS:

12. **hadoop fs -mkdir /spark-logs**

Run the History Server:

13. **\$SPARK_HOME/sbin/start-history-server.sh**

Access the History Server by navigating to <http://master:18080> in a web browser:

Spark version:

```
spark-submit --version
```

Getting started with the classic Jupyter Notebook

```
sudo jupyter notebook --allow-root
```

```
sudo jupyter notebook list ---> to get the token
```