



INCEPTEZ TECHNOLOGIES Installation guide

1. Install spark

```
cd /home/hduser/install  
tar xvzf spark-2.0.1-bin-hadoop2.7.tgz  
sudo mv spark-2.0.1-bin-hadoop2.7 /usr/local/spark
```

2. Goto the /usr/local/hadoop/etc/hadoop/, edit yarn-site.xml and add the below content

vi /usr/local/hadoop/etc/hadoop/yarn-site.xml

```
<property>  
  <name>yarn.log-aggregation-enable</name>  
  <value>>true</value>  
  <description>Whether virtual memory limits will be enforced for containers</description>  
</property>  
<property>  
  <name>yarn.nodemanager.vmem-check-enabled</name>  
  <value>>false</value>  
  <description>Whether virtual memory limits will be enforced for containers</description>  
</property>  
<property>  
  <name>yarn.nodemanager.vmem-pmem-ratio</name>  
  <value>4</value>  
  <description>Ratio between virtual memory to physical memory for containers</description>  
</property>
```

To Start Scala Spark cli

spark-shell

Other Commands to Run later:

3. To Start Spark Services

```
start-master.sh
```

4. Spark Workers/Executors

```
start-slaves.sh
```

Run the below configs if the above standalone jobs doesn't work

7. Configure Spark Copy spark-env.sh.template:

```
vi /usr/local/spark/conf/spark-env.sh
```

```
export SPARK_MASTER_IP=localhost  
export SPARK_WORKER_CORES=1 export  
SPARK_WORKER_MEMORY=800m export  
SPARK_WORKER_INSTANCES=1
```

8. Edit file /usr/local/spark/conf/spark-defaults.sh

```
vi /usr/local/spark/conf/spark-defaults.sh
```

```
export SPARK_MASTER_IP=localhost  
export SPARK_WORKER_CORES=1 export  
SPARK_WORKER_MEMORY=800m export  
SPARK_WORKER_INSTANCES=1
```

9. Web UI:

```
localhost:8080
```

10. Run Spark

To start Pyspark

pyspark

To Start spark shell with yarn-client

spark-shell --verbose --master yarn-client --driver-memory 512m --num-executors 1 --
executorcores 2 --executor-memory 512m

(or)

spark-shell --master yarn