**Learning Objective 2**

To be able to configure Apache Spark on a cluster with Yarn.

**Ports**

Some typical settings

**Spark ports**

spark.ui.port 4040

spark.history.provider 18080

**Yarn Ports**

yarn.nodemanager 8042

jobtracker 19888

Spark Configuration

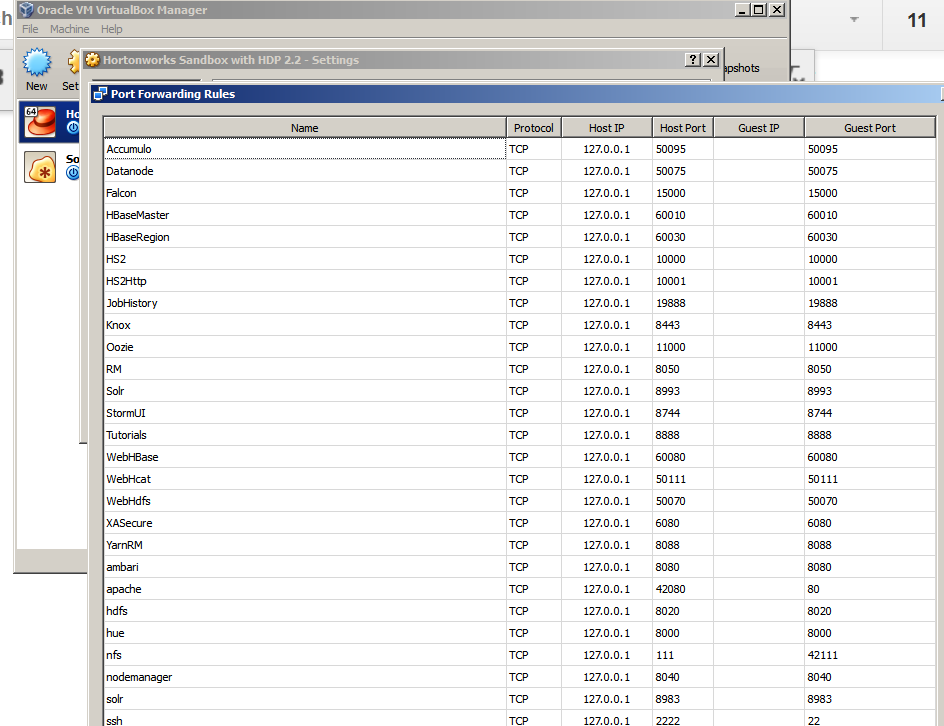
<https://spark.apache.org/docs/1.1.1/configuration.html>

YARN Configuration

<http://www.alexjf.net/blog/distributed-systems/hadoop-yarn-installation-definitive-guide/#yarn-configuration>

How you set them in the Hortonworks sandbox vm

Port Forwarding



Environment Variables

Spark

SPARK\_HOME/conf/spark-defaults.conf

spark.driver.extraJavaOptions -Dhdp.version=2.2.0.0-2041  
spark.yarn.am.extraJavaOptions -Dhdp.version=2.2.0.0-2041

Yarn

export YARN\_CONF\_DIR=/etc/hadoop/conf

Set the host file on the physical box

Windows

C:\Windows\System32\drivers\etc\hosts

Linux /etc/hosts

1. Ensure port forwarding (from host to guest) in the HDP Sandbox for ports 4040, 8042, 18080, and 19188.
2. /etc/hosts 127.0.0.1 localhost sandbox.hortonworks.com
3. export YARN\_CONF\_DIR=/etc/hadoop/conf
4. Create a file SPARK\_HOME/conf/spark-defaults.conf and add the following settings:

spark.driver.extraJavaOptions -Dhdp.version=2.2.0.0-2041  
spark.yarn.am.extraJavaOptions -Dhdp.version=2.2.0.0-2041