

Cheat sheet for practice 2, Hadoop 30088

What are the addresses for the UIs?

- Hadoop jobs: **quickstart.cloudera:8088**
- Spark jobs: **quickstart.cloudera:8080**

What is my Rest URL for Spark?

- **spark://quickstart.cloudera:6066**
- if in doubt: open the Spark Master UI and find it listed on the upper left.

How do I run a spark job on the cluster?

Make sure your data is on the cluster (in HDFS)

```
$ spark-submit \  
--master spark://quickstart.cloudera:6066 \  
--deploy-mode cluster \  
--class <full classname> \  
<the name of your jar>.jar \  
<fully qualified input path> <fully qualified outputDir>
```

How do I run an MR2 job on the cluster?

Make sure your data is on the cluster (in HDFS)

```
$ hadoop jar <the name of your jar>.jar \  
<full classname> \  
<input on HDFS> <outputDir on HDFS>
```

How do I create a jar that I can run from the commandline?

- Open **Eclipse**.
- In the **Package Explorer**, right-click on the project you are using.
- If you have already built a jar, cleanup by selecting **Run-As** and then **Maven Clean**.
- Select **Run-As** and select **Maven-Install**.
 - This will create the jar file
 - You will see the process in the Eclipse console
- In the **Package Explorer**, open **target** folder
 - it should now contain a jar file called: **<projectname>-1.0.0-SNAPSHOT.jar**
 - this is the jar file you will use to run jobs from the commandline.

But where's my jar?

- *practice jars created by maven are in your workspace under:*
`/home/cloudera/Desktop/workspace/<project name>/target`

What's that command for mounting a share folder?

- **`sudo mount -t vboxsf share /home/cloudera/share`**

Where are hadoop configuration files?

- site.xml files for yarn, mapred, hdfs and core are in `/etc/hadoop/conf.pseudo`

Where are hadoop log files for jobs I ran on the cluster?

- In **HDFS** - use Hue to view.
- Look under:
`/var/log/hadoop-yarn/apps/cloudera/logs`
- click on the application of interest, it will open the log for the app in binary mode. Use "View file as Text".

Oh heck, how do I find anything?

- To find anything, try this: `$ find . -name 'anything'`
- This will search for anything, starting from your current directory
- You can add wildcards like `$ find . -name '*anything*'`
- learn more by typing `$man find`