On the Windows Powershell:

```
$ docker run --hostname=quickstart.cloudera --privileged=true -t -i -v
C:\Users\bina\Documents\dockerMR:/src --publish-all=true -p 8888 cloudera/quickstart
/usr/bin/docker-quickstart


Note: For windows this will prompt you to share a folder with docker and may ask for your credentials
```

- This command will initialize and configure the image and start various frameworks, and result will sequence of outputs. After it completes, it will return a # prompt. To confirm the configuration, do an `ls` , you should see the directory structure shown, where `src` is mapped to your `C:\Users\bina\Documents\dockerMR`

- `$ ls`

```
bin   dev  home  lib64       media  opt           proc  sbin     src  sys  usr
boot  etc  lib   lost+found  mnt    packer-files  root  selinux  srv  tmp  var
```

7. Provision Data

- Review the picture we discussed about HDFS.
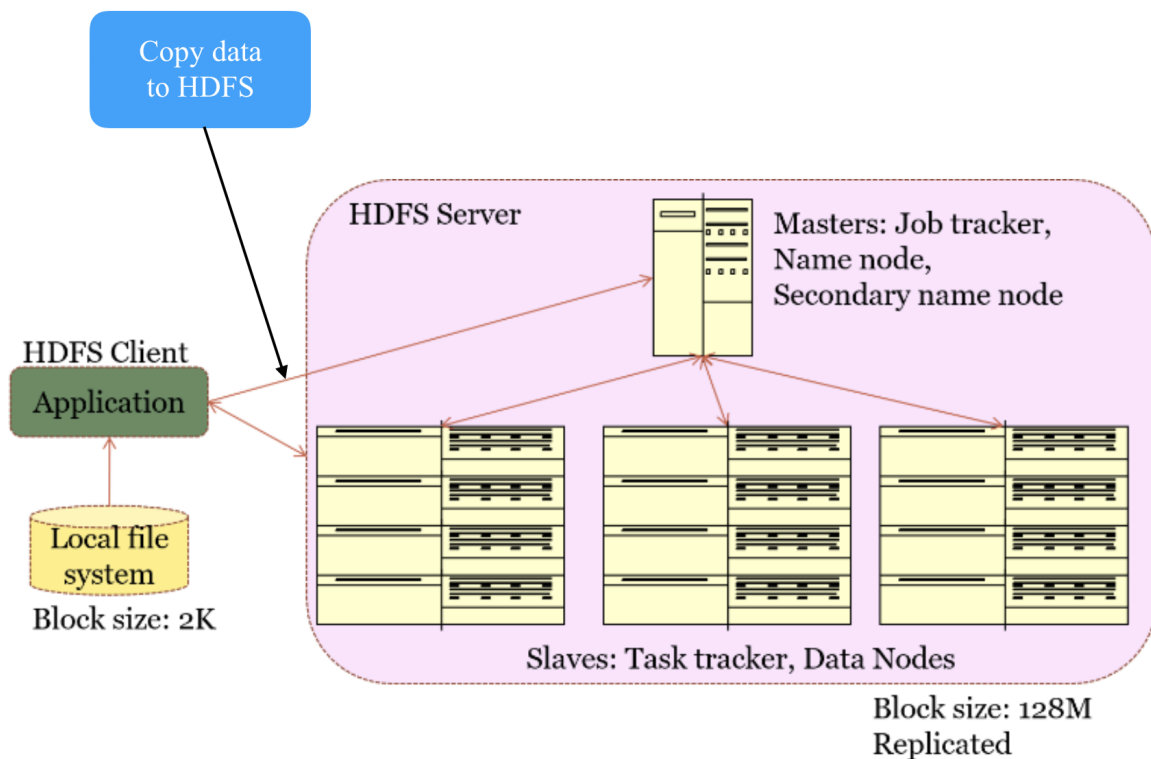
- To create input locations in your HDFS (name them in a suitable manner) run the following commands

```
$ hadoop fs -mkdir /user/yourName
$ hadoop fs -mkdir /user/yourName/MR
```

```
$ hadoop fs -mkdir /user/yourName/MR/input
```

- These commands create a directory structure for your files. Replace `yourName` with your first name or something like that. Figure above pictures the operations. Copy files from your local file system to Hadoop file system.

- Copy files from your shared folder i.e. `/src/data/` to the HDFS using the following commands, check the directory you created on hadoop fs, and copy data from local to hadoop fs

```
$ cd /src/data/
$ hadoop fs -put file1.txt /user/yourName/MR/input
```

  For multiple files

```
$ hadoop fs -put *.txt /user/yourName/MR/input
$ hadoop fs -ls /user/yourName/MR/input/
```

  Change back to the `/src` directory `$ cd ..`

8. Process the data using MR

- Now you are ready to run your MapReduce program using the following command. The hadoop streaming jar the docker provides is located at the following location in the command. If using a Virtual Machine or any other docker or your local machine with hadoop installed you will have to locate the streaming and jar and use that path and name for the jar. Make sure you replace "yourName" in the command below with the same name you used in step 7.

```
$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-streaming-2.6.0-cdh5.7.0.jar \
-file /src/mapper.py    -mapper /src/mapper.py \
-file /src/reducer.py   -reducer /src/reducer.py \
-input /user/yourName/MR/input/* -output /user/yourName/MR/output
```

  where `mapper.py` and `reducer.py` are your python programs to perform MapReduce

9. Observe the output

- Observe the output generated by MR and move it to your local file system for further processing.

```
$ hadoop fs -cat /user/bina/MR/output/part*
$ hadoop fs -get /user/bina/MR/output/ /src/
```

- Review the output, as we discussed earlier. Once transferred to the local system, you can use it for further processing and extracting information using other algorithms. (You can go back to your dockerMR directory and review the output directory and the MR output.)

10. Ready to Quit

- Quit the docker using `Ctrl+P` then `Ctrl+Q`

11. Further Explanation