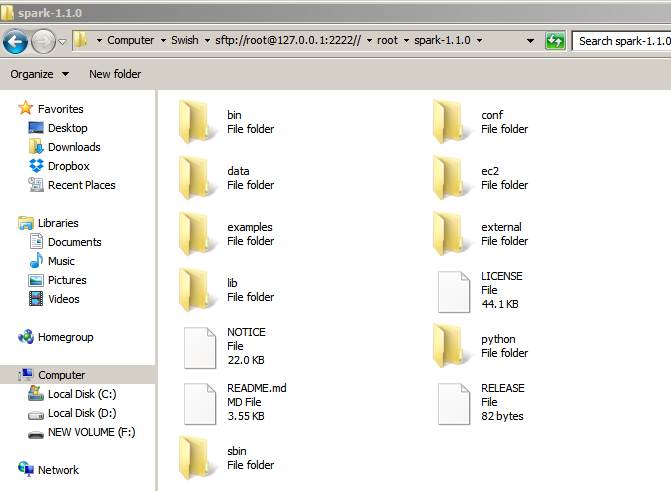
**Learning Objective 3**

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

1, Start the vm and copy the examples directory to a development directory on your box.



2, Create a java project with your favourite IDE you will need a build tool like maven or gradle

we will look at two options for the java code

1. maven
2. gradle

a,1 create empty maven project

copy code correctly

create build file

Build the sample source code into a ‘uber jar’

scala mvn scala:compile -DdisplayCmd=true

java mvn package

Build code on windows

When running the source if you get hadoop or winutil errors at runtime

download the code from, no need to build this code, use as is

<https://github.com/srccodes/hadoop-common-2.2.0-bin>

then set Environment Variable

HADOOP\_HOME

C:\hadoop-common-2.2.0-bin

submit to yarn

1, build application to run in ‘yarn-cluster mode’

2, set up flags for spark submission

3, run on cluster

2.

**./bin/spark-submit --class org.apache.spark.examples.JavaSparkPi \**

**--master yarn-cluster --num-executors 3 \**

**--driver-memory 512m \**

**--executor-memory 512m \**

**--executor-cores 1 \**

**../eduonix\_spark-deploy.jar 10**