# Laboratorio 1. (1° giorno - mattina)

## Obiettivi

* Familiarizzarsi con l'ambiente del laboratorio
  + Unix shell
  + Notebook Apache Zeppelin (<https://zeppelin.apache.org/>)
  + Internet browser (Google Chrome)
* Ripasso di alcuni comandi di HDFS
* Ripasso del linguaggio Scala

## Comandi HDFS

Possiamo usare una shell unix per analizzare il contenuto di HDFS

Per elencare i file:

$ hdfs dfs -ls

o anche:

$ hadoop fs -ls

È possibile specificare un path (relativo o assoluto):

$ hdfs dfs -ls spark

$ hdfs dfs -ls /user/cloudera/spark

$ hdfs dfs -ls hdfs:///user/cloudera/spark

Per esaminare il contenuto di un file:

$ hdfs dfs -cat spark/divina\_commedia.txt | less

## Linguaggio SCALA

Il riferimento principale è il sito <https://docs.scala-lang.org/>.

Per questo corso è stato preparato un "cheat sheet" (*Scala cheat sheet.pdf*)

## La shell di spark

La shell di spark si lancia con il comando:

$ spark-shell

La shell è un interprete che accetta ed esegue immediatamente comandi in Scala/Spark:

scala> spark.version

res0: String = 2.4.0

Per uscire:

scala> :quit

## Il notebook Apache Zeppelin

Useremo per la maggior parte degli esercizi il Notebook Apache Zeppelin (<https://zeppelin.apache.org/>).

Il notebook è una web-application, quindi prevede un "demone" che gira in background e un Internet Browser che visualizza l'interfaccia grafica. Il demone dovrebbe partire in automatico. In ogni caso lo si può controllare con i comandi:

e

### Obiettivi

### Strumenti

* Unix shell
* Notebook Apache Zeppelin (<https://zeppelin.apache.org/>)
* Internet browser (Google Chrome)

### Esercizio.1

ciccio