ASSIGNMENT # 1

**1.** Explain **using your own words**, in 4 or 5 lines, the difference between *Big Data*, *Data Analytics* and *Distributed Systems*.

**2. Go to:**

<https://www.datamation.com/big-data/big-data-pros-and-cons.html>

Make the reading and answer the following questions:

1. What are the nine pros of Big Data? Explain each one of them in one or two lines.
2. What are the nine cons of Big Data? Explain each one of them in one or two lines.

**3.** Is Hadoop and Big Data the same? (Give your personal answer in 5 or maximum 6 lines).

**4.** Scenario:

* You have 100 terabytes of music: 44 terabytes of Indian music, 33 terabytes of rock music and 23 of electronic and hip-hop music. Then, create the partitioning and the replication in different nodes.

**5.** Practice: Intro to *Hadoop*.

A) Go to:

* The virtual machine in the computer.

B) Send me your *gmail* account. If you do not have one, you have to create it, because I will give you access to two *Cloudera* files.

C) Go to your *gmail* account, download and import the Cloudera files to the virtual machine.

D) Run Cloudera