ASSIGNMENT # 2

**1.** In your own words: (Please do not copy and paste)

**A)** What is *Hadoop Ecosystem*? (3 or maximum 4 lines)

**B)** What is *HDFS*? (2 or maximum 3 lines)

**C)** What is the difference between Master Nodes and Slave Nodes? (4 lines)

**D)** What is the difference between Hadoop and HDFS? (4 lines)

**2.** Go to Omnivox (Section Documents)

Make the reading *How to set properties for Secondary NameNode in Hadoop?* And answer the following questions using your own words:

**A)** What is the meaning of EditLog and FsImage?

**B)** Is the Secondary NameNode a back up of NameNode server?

**C)** What is the main function of Secondary NameNode?

**D)** What is the property that defines the number of uncheckpointed transaction on the Name Node?

**3.** Explain using your own words.

**A)** Why do we use the expression *Data ingestion*? (3 lines)

**B)** What is the difference between *Data ingestion* and *ETL*? (4 lines)

**C)** Why *Data ingestion* and *ETL* are important in Big Data?

**D)** What is Sqoop and what are its features? (6 lines)

**4.** Go to:

<https://www.projectpro.io/hadoop-tutorial/hadoop-sqoop-tutorial-data-aggregation>

**A)** Make the reading and explain:

1. What is the problem statement? (3 lines)

2. What is the proposed solution? (2 lines)

**5.** Practice: Intro to *HDFS* and *Sqoop*.

- Attach the screenshots of the practice in the assignment.