1] RDD

- The dataset shared (`visits.txt`) captures the `visitors` visited the White House. Every line captures the details of a visitor. There are some visitors who visited White House to meet President Of the United States (POTUS)

- Please perform following activities

1. Load the data from visits.txt on HDFS into a Rdd
2. Count the number of elements in the above Rdd
3. Verify this count against the number of lines in the file
4. Select only those records from Rdd where the 19th field matches POTUS and then
   1. Capture the 0th field (This datapoint is lname in the data set)
   2. Capture the 1st field (This datapoint is fname in the data set)
   3. Capture the 6th field (This datapoint is time\_of\_arrival in the data set)
   4. Capture the 11th field (This datapoint is appt\_scheduled\_time in the data set)
   5. Capture the 21st field (This datapoint is location in the data set)
   6. Capture the 25th field (This datapoint is comment in the data set)
5. Count the number of records in a final Rdd
6. Save the result of above RDD (which has implemented point number 4 above) in a folder ‘wh\_visits’ on the local file system.

* Dataset
  + dataset/visits.txt
* TimeLine
  + 45 Mins

2] PairRdd

* Code/Dataset
  + dataset/Pair-RDD.zip
* TimeLine
  + Jan 2, 2024 EOD