This the first push :

Command to initiate mysql : mysql -u root -p, it indicates that the user is “root”

After I create a database called “ca2\_tweets”

Enabling loading local with the commandf : sudo nano /etc/mysql/mysql.conf.d/mysqld.cnf

Mysql command to download a table as csv:

SELECT \* INTO OUTFILE ' /home/hduser/Downloads/test\_mysql.csv’

FIELDS TERMINATED BY ',' OPTIONALLY ENCLOSED BY '"'

LINES TERMINATED BY '\n'

FROM ProjectTweets;

Check the status of the mysql database :

CREATE TABLE ProjectTweets (

`index` INT NOT NULL,

ids BIGINT NOT NULL,

`date` VARCHAR(50) NOT NULL,

flag VARCHAR(50) NOT NULL,

user VARCHAR(50) NOT NULL,

`text` TEXT NOT NULL,

PRIMARY KEY (`index`)

);

SELECT \* FROM ca2\_tweets.ProjectTweets LIMIT 10;

Install locally pandas: sudo apt-get install python3-pandas

sudo apt-get install python3-scikit-learn

Problem using : !pip install python-dotenv

“To install Python packages system-wide, try apt install python3-xyz, where xyz is the package you are trying to install.” So sudo apt-get install python3-dotenv

Check which libraries I have in my jupyter notebook : !pip list

Before using pyspark:  
  
**jps**

**$start-dfs.sh**

**$start-yarn.sh**

**To know the host name : SELECT @@hostname;**

**$stop-dfs.sh**

**$stop-yarn.sh**

Command to create a workload opn YCSB with 1000 units :

./bin/ycsb.sh load jdbc -P ./jdbc-binding/conf/db.properties -P workloads/workloada\

In the case that I want to save a txt file :  
  
./bin/ycsb.sh load jdbc -P ./jdbc-binding/conf/db.properties -P workloads/workloada > /home/hduser/outputMySQL\_WORKLOADA.txt

In the case to use YCSB with mongo into the file ycsb-0.17.0:  
  
./bin/ycsb.sh load mongodb -s -P workloads/workloada

START mongo database: sudo systemctl start mongod

Check status of mongo database: sudo systemctl status mongod

Mongosh to interact with the databases;

Command to show databases: show dbs;

To use ycsb database : use ycsb

To show collections: show collections

To count documents use : db.usertable.countDocuments()

To save a text file of mongo metrics in, use the command:

./bin/ycsb load mongodb -s -P workloads/workloada > /home/hduser/output-

Load.txt

**hadoop fs -put ./ProjectTweets.csv /ca2\_bd\_ad**

**sudo hadoop fs -put /var/lib/mysql-files/test\_mysql.csv /ca2\_bd\_ad**

**hadoop fs -put . /test\_mysql.csv /ca2\_bd\_ad**

**check if the file is in Hadoop: change to push**

A graph with a line

Description automatically generated