**Assignment -1 (Weblogs dataset)**

Creating the folder in local machine(Linux) the upload the all logs file.

Creating the folder on console using [itv003722@g02 ~]$ hdfs dfs -mkdir ./weblogs

**Creating the table schema**

-To set metadata in itversity lab

set hive.metastore.warehouse.dir=/user/itv003722/weblogs;

On Hive Console

Create database in hdfs

hive> create database weblog;

to check in Hdfs [itv003722@g02 ~]$ hdfs dfs -ls ./weblogs

create table logs35(data string);

to check in Hdfs [itv003722@g02 ~]$ hdfs dfs -ls ./weblogs/weblog.db

-To load the logs file directly from local to logs table

hive> load data local inpath '/home/itv003722/weblogs' into table logs;

To check in Hdfs

[itv003722@g02 ~]$ hdfs dfs -ls ./weblogs/weblog.db/logs

To check in Hdfs

[itv003722@g02 ~]$ hdfs dfs -cat ./weblogs/weblog.db/logs/2014-03-15.log

To check  
22.252.151.164 - 23251 [15/Mar/2014:00:00:30 +0100] "GET /KBDOC-00150.html HTTP/1.0" 200 19203 "http://www.loudacre.com" "Loudacre Mobile Browser Sorrento F11L"

CREATE TABLE logs\_complete(IP\_ADDR STRING, USER\_ID STRING, DATES STRING,TIMES STRING,ZONE STRING, REQUEST\_URL STRING, STATUS\_CODE STRING, OBJECT\_SIZE STRING);

insert overwrite table logs\_complete

select substr(data,1,locate('-',data,1)-1),

substr(data,(locate('-',data,1)+2),(locate('[',data,1)-1)-(locate('-',data,1)+2)),

substr(data,(locate('[',data,1)+1),(locate(':',data,1)-1)-locate('[',data,1)),

substr(data,(locate(':',data,1)+1),(locate('+',data,1)-1)-locate(':',data,1)),

substr(data,(locate('+',data,1)+1),(locate(']',data,1)-1)-locate('+',data,1)),

substr(data,(locate('"',data,1)+1),(locate('1.0',data)+2) - (locate('"',data))),

substr(data,(instr(data,'1.0')+5),(locate('1.0',data)+8)-(locate('1.0',data)+5)),

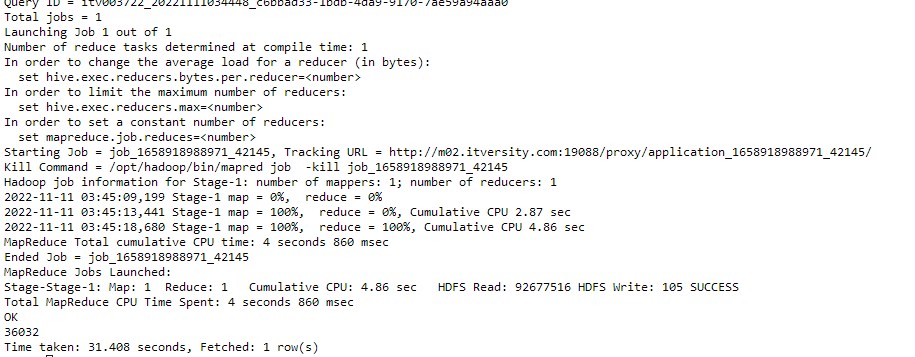
substr(data,(locate('1.0',data)+8),(locate('http:',data)-2) - (locate('1.0',data)+8))

from logs ;

drop table logs;

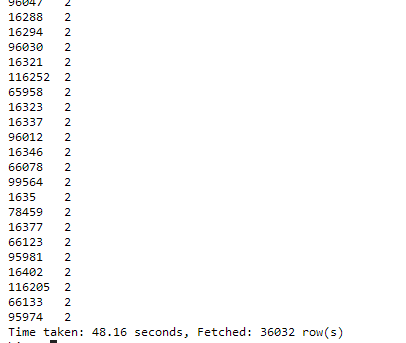
**How many users are there?**

hive> select count(distinct user\_id) from logs\_complete;



**Every user has made how many requests.**

hive> select user\_id,count(request\_url) as number\_of\_requests FROM logs\_complete group by user\_id order by number\_of\_requests desc;



**Display total number of successful request.**

select count(status\_code) as total\_number\_of\_successful\_request from logs\_complete where status\_code like '2%';

