**-- Changing from Master DB to User DB**

use riazalab;

--=============================================================================

**-- Q1. Create a table in Hive to access this data. Schema is mentioned above.**

**-- Dropping Table If its already exists**

drop table if exists logs\_test;

**-- Creating the Table**

CREATE external table logs\_test(

ip STRING,

dash string,

userid int,

reqdatetime string,

reqtime string,

pagereq string,

pagereq2 string,

code string,

statuscode int,

datasize int

)

row format delimited fields terminated by " ";

**-- Loading the data into table.**

load data inpath 'weblogs/' into table logs\_test;

**-- Creating a new table from the old table in**

**-- addition to cleaning and converting into required format.**

SELECT \* from logs\_test limit 5;

**--Creating the Final table Weblogs which holds the data along with**

**--appropriate format and data.**

create table weblogs as

select ip,

userid,

concat(replace(reqdatetime,"[",""),replace(reqtime,"]","")) ReqDateTime ,

replace(pagereq,'"',"") PageReq,

replace(pagereq2,"/","") PageName,

replace(code,'"',"") url,

statuscode,

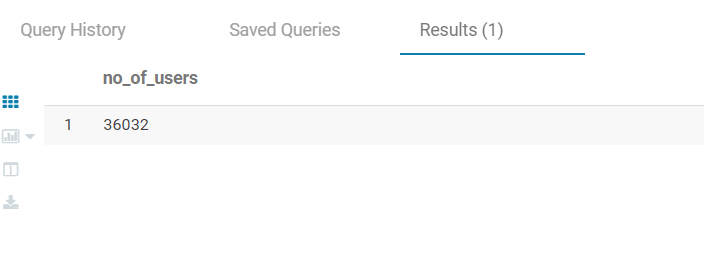
datasize

from logs\_test;

--=============================================================================

**-- Q2. How many users are there?**

select count(distinct userid) as No\_Of\_Users from weblogs;



--=============================================================================

**-- Q3. every user has made how many requests.**

select userid, count(Pagereq) as No\_Of\_Requests from weblogs

group by userid;

Query result exported to table 🡪 Q3\_table.csv

--=============================================================================

**-- Q4. Display total number of successful requests.**

select count(statuscode) as No\_Of\_Sucessful\_Req from weblogs

where statuscode like '2%';

