Hive Certification Project Report –Edureka

Manoj K V

**Index:**

I. Problem statement

II. Technology/Software Used

III. Solution Flow Diagram

IV. Solution

**Problem Statement**

1. Find the list of people with grade “B” who have taken loan.

2. Find the list of people having interest more than 1000.

3. Find the list of people having loan amount more than 1000.

4. Get the highest loan amount given to grade users (A-G).

5. Highest loan amount given in that year with that Employee id and Employees annual income.

6. Get the total number of loans with loan id and load amount which are having loan status as Late.

7. Average loan interest rate with 60-month term and 36-month term.

**Technology/Software Used:**

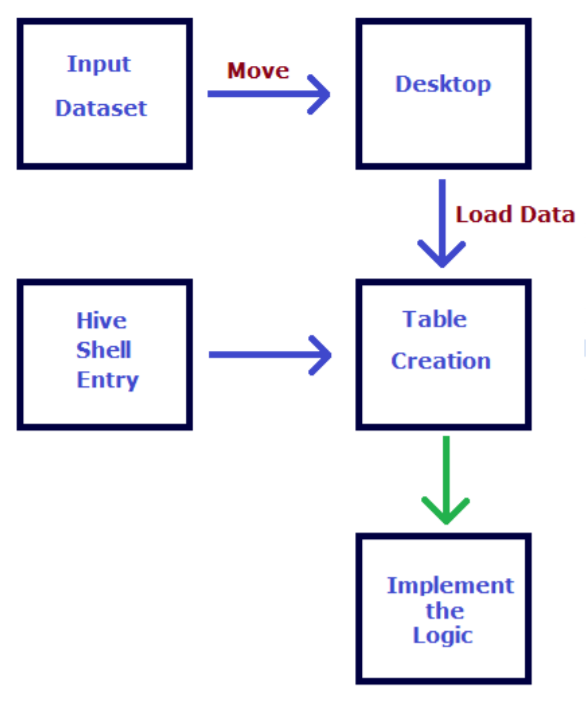
* Hadoop environment

(HDP sandbox)

* Apache Hive

Hive 1.2.1000.2.4.2.57-1

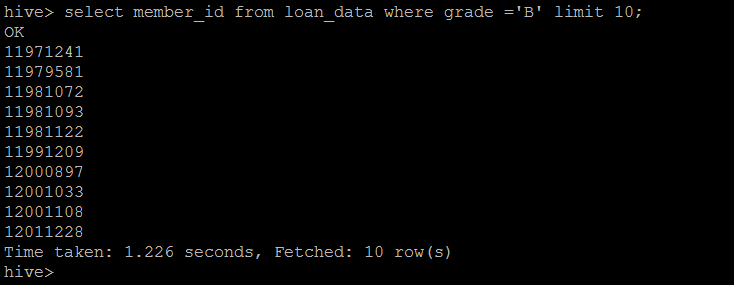
**Solution Flow Diagram**



**Solution:**

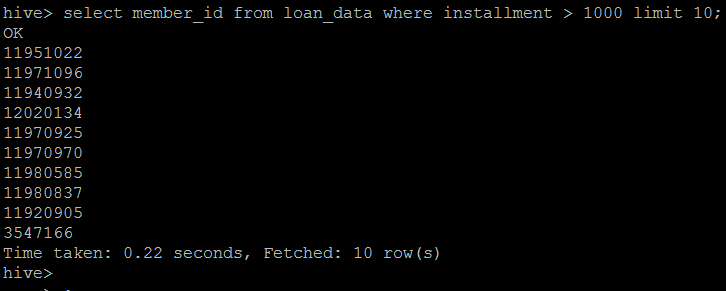
*--1. Find the list of people with grade "B" who have taken loan.*

**select** member\_id **from** loan\_data **where** grade ='B'**;**



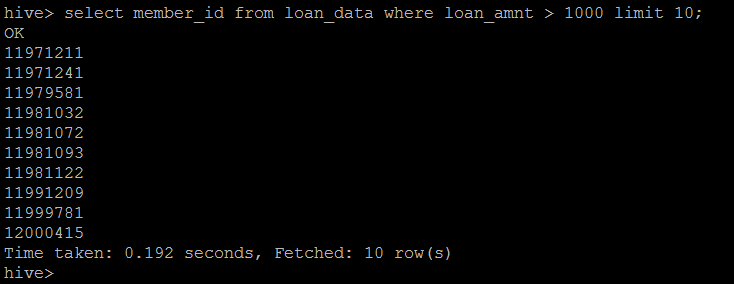
*--2. Find the list of people having interest more than 1000.*

**select** member\_id **from** loan\_data **where** installment > 1000**;**



*--3. Find the list of people having loan amount more than 1000.*

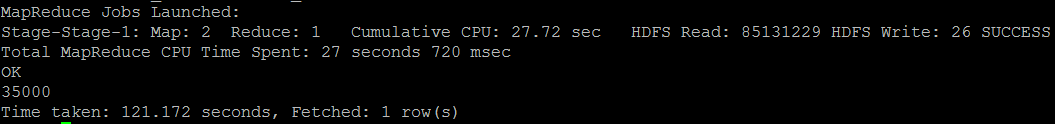
**select** member\_id **from** loan\_data **where** loan\_amnt > 1000**;**

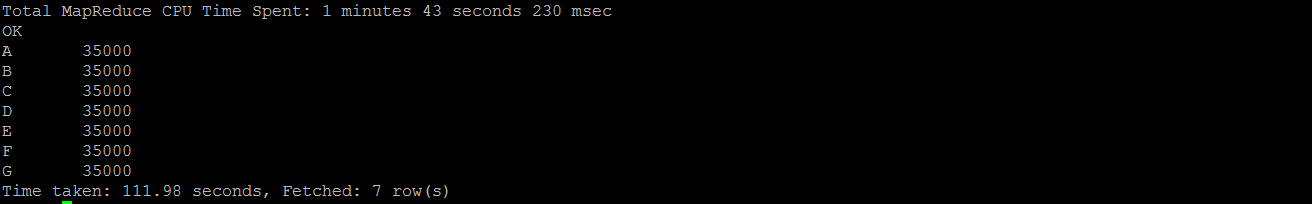


*--4. Get the highest loan amount given to grade users (A-G).*

**select** max(loan\_amnt) **from** loan\_data **where** grade between 'A' **and** 'G'**;**

**select** grade, max(loan\_amnt) **AS** max\_loan\_amnt **from** loan\_data **where** grade between 'A' **and** 'G' **group** **by** grade **order** **by** grade**;**





*--5. Highest loan amount given in that year with that Employee id and Employees annual income.*

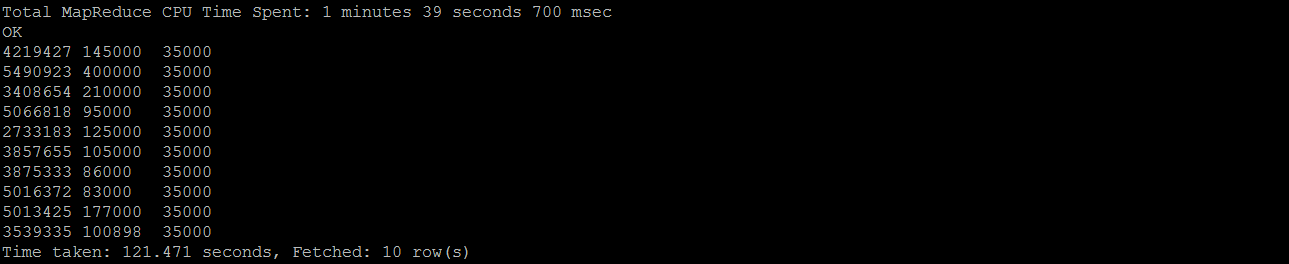
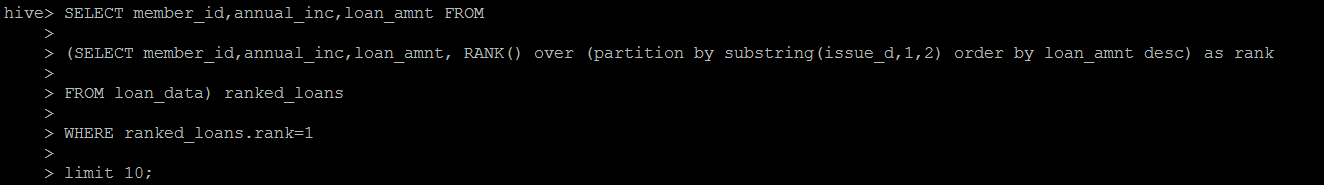
**SELECT** member\_id,annual\_inc,loan\_amnt **FROM**

(**SELECT** member\_id,annual\_inc,loan\_amnt, RANK() over (**partition** **by** substring(issue\_d,1,2) **order** **by** loan\_amnt **desc**) **as** rank

**FROM** loan\_data) ranked\_loans

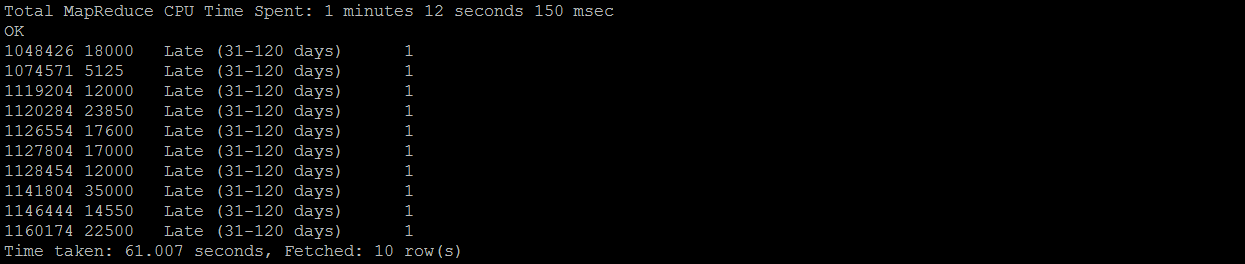
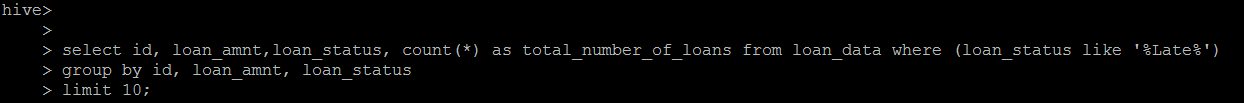
**WHERE** ranked\_loans.rank=1

**;**



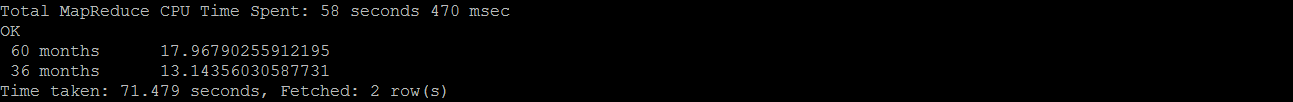
*--6. Get the total number of loans with loan id and loan amount which are having loan status as Late.*

**select** id, loan\_amnt,loan\_status, count(\*) **as** total\_number\_of\_loans **from** loan\_data **where** (loan\_status like '%Late%') **group** **by** id, loan\_amnt, loan\_status**;**



*--7. Average loan interest rate with 60-month term and 36-month term.*

**select** term,avg(regexp\_replace(int\_rate,'%','')) **from** loan\_data **where** trim(term) **in** ('60 months','36 months') **group** **by** term **;**



Thank you!