**Data Testing Assessment**

A school 'Study & Play' has different branches in India. Students appeared in 10th & 12th board have scored good marks in different subjects. 'Study & Play' wants to recognize teachers across different streams to award them for their student's performance.

'Student & Play' are looking for a data warehouse solution to do analysis on student performance in a particular branch or subject.

**Exercise -**

As a part of this exercise, use the below Student data to import in database table.

Sample Data –

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Student Name | StudentRollNo | Class | Age | Father's Name |
| AJAY KUMAR | IETLINF01 | X | 16 | PREM KUMAR |
| RAJAT MISHRA | IETLINF02 | XI | 17 | PANKAJ MISHRA |
| SUDHEER SHARMA | IETLINF03 | IX | 15 | SUNDER SHARMA |
| NANCY KAUR | IETLINF04 | XII | 18 | AK SINGH |
| SUMITA SHARMA | IETLINF05 | IV | 14 | AJAY SHARMA |
| NANDINI GUPTA | IETLINF06 | VII | 18 | RAM GUPTA |
| RAM KUMAR | IETLINF07 | X | 16 | RAMAN KUMAR |

Tasks

1. Create above table in database and insert the data.

hive> SHOW DATABASES ;

hive> CREATE DATABASE students\_demo\_db ;

hive> SHOW DATABASES ;

hive> USE students\_demo\_db ;

hive> SHOW TABLES ;

hive> CREATE EXTERNAL TABLE Student(

StudentName STRING,

StudentRollNo INT,

Class STRING,

Age INT,

FatherName STRING

);

hive> SHOW TABLES ;

hive> DESCRIBE stocks ;

INSERT INTO TABLE Student VALUES

('AJAY KUMAR','IETLINF01','X','16','PREM KUMAR'),

('RAJAT MISHRA','IETLINF02','XI','17','PANKAJ MISHRA'),

('SUDHEER SHARMA','IETLINF03','IX','15','SUNDER SHARMA'),

('NANCY KAUR','IETLINF04','XII','18','AK SINGH'),

('SUMITA SHARMA','IETLINF05','IV','14','AJAY SHARMA'),

('NANDINI GUPTA','IETLINF06','VII','18','RAM GUPTA'),

('RAM KUMAR','IETLINF07','X','16','RAMAN KUMAR');

SudentName StudentRollNo Class Age FatherName

AJAY KUMAR IETLINF01 X 16 PREM KUMAR

RAJAT MISHRA IETLINF02 XI 17 PANKAJ MISHRA

SUDHEER SHARMA IETLINF03 IX 15 SUNDER SHARMA

NANCY KAUR IETLINF04 XII 18 AK SINGH

SUMITA SHARMA IETLINF05 IV 14 AJAY SHARMA

NANDINI GUPTA IETLINF06 VII 18 RAM GUPTA

RAM KUMAR IETLINF07 X 16 RAMAN KUMAR

1. Write sql query to find out total number of students in each class.

hive>select

COUNT(if(Class='IV',1,NULL)) as FOUR,

COUNT(if(class='VII',1,NULL)) as SEVEN,

COUNT(if(Class='IX',1,NULL)) as NINE ,

COUNT(if(Class='X',1,NULL)) as TEN,

COUNT(if(Class='XI',1,NULL)) as ELEVEN,

COUNT(if(Class='XII',1,NULL)) as TWELVE,

from Student;

NINE TEN ELEVEN TWELVE

1 1 1 2 1 1

1. Write sql query to find out average age of students in each class.

hive>SELECT Class, avg(Age) as AVERAGE\_AGE FROM Student GROUP BY Class;

Class AVERAGE\_AGE

X 16.0000

XI 17.0000

IX 15.0000

XII 18.0000

IV 14.0000

VII 18.0000

1. Write sql query to find out the number of students having same roll no.

hive>SELECT StudentRollNo, COUNT(StudentRollNo) FROM Student

GROUP BY StudentRollNo

HAVING COUNT(StudentRollNo) > 1;

The Query Doesn’t return anything as all Roll No are unique

1. Write sql query to view the duplicate name of a student from the table.

Hive>SELECT StudentName, COUNT(StudentName)FROM Student GROUP BY StudentName HAVING COUNT(StudentName) > 1;

The Query Doesn’t return anything as all Names are unique