Session 07: Hive Operations

Assignment 2

**Associated Data Files**

This Data set is about Olympics. You can download the data set from the below link:

<https://drive.google.com/open?id=0ByJLBTmJojjzV1czX3Nha0R3bTQ>

**Problem Statement**

1. Write a Hive program to find the number of medals won by each country in swimming.

2. Write a Hive program to find the number of medals that India won year wise.

3. Write a Hive Program to find the total number of medals each country won.

4. Write a Hive program to find the number of gold medals each country won.

**Solution :-**

Firstly creating table and loading data into the table:-

hive> create table if not exists athlete\_details

( athlete\_name string, age int, country string, year int, closing\_date string, sport string, gold\_medals int,

silver\_medals int, bronze\_medals int, total\_medals int)

row format delimited fields terminated by '\t';

hive> load data local inpath '/home/acadgild/Acadgild/Assignment7\_2/olympix\_data.csv' into table athlete\_details;



**Now finding solutions and saving the output in file :-**

1. Write a Hive program to find the number of medals won by each country in swimming.

**hive>** insert overwrite local directory '/home/acadgild/Acadgild/Assignment7\_2'

row format delimited fields terminated by '\t'

**select country, sum(gold\_medals), sum(silver\_medals), sum(bronze\_medals), sum(total\_medals) from athlete\_details group by country**;

Output file :-





1. Write a Hive program to find the number of medals that India won year wise.

Hive> insert overwrite local directory '/home/acadgild/Acadgild/Assignment7\_2'

row format delimited fields terminated by '\t'

**select year, country, sum(gold\_medals), sum(silver\_medals), sum(bronze\_medals), sum(total\_medals) from athlete\_details where country = "India" group by year, country;**

**Output file - **



1. Write a Hive Program to find the total number of medals each country won.

**Hive>** insert overwrite local directory '/home/acadgild/Acadgild/Assignment7\_2'

row format delimited fields terminated by '\t'

**select country, sum(total\_medals) from athlete\_details group by country;**

**Output file - **



1. Write a Hive program to find the number of gold medals each country won.

**Hive>** insert overwrite local directory '/home/acadgild/Acadgild/Assignment7\_2'

row format delimited fields terminated by '\t'

**select country, sum(gold\_medals) from athlete\_details group by country;**

**Output file –**

****

****