Using partitioning as most of the queries require country wise grouping.

* Create olympicdata table. (base table)

create table if not exists olympicdata

(

athlete\_name string,

age int,

country string,

year int,

closing\_date string,

sport string,

gold int,

silver int,

bronze int,

total\_medals int

)

row format delimited

fields terminated by ',';

* Create partitioned table olympicdata\_part

create table if not exists olympicdata\_part

(

athlete\_name string,

age int,

year int,

closing\_date string,

sport string,

gold int,

silver int,

bronze int,

total\_medals int

)

PARTITIONED BY (country string)

row format delimited

fields terminated by ','

STORED AS TEXTFILE;

* Load data in base table

LOAD DATA LOCAL INPATH '/home/acadgild/nikidir/olympix\_data.csv' OVERWRITE INTO TABLE olympicdata;

* Load data in partitioned table from base table.

set hive.exec.dynamic.partition=true;

set hive.exec.max.dynamic.partitions.pernode=110;

insert overwrite table olympicdata\_part

partition (country)

select

athlete\_name ,

age ,

year ,

closing\_date ,

sport ,

gold ,

silver ,

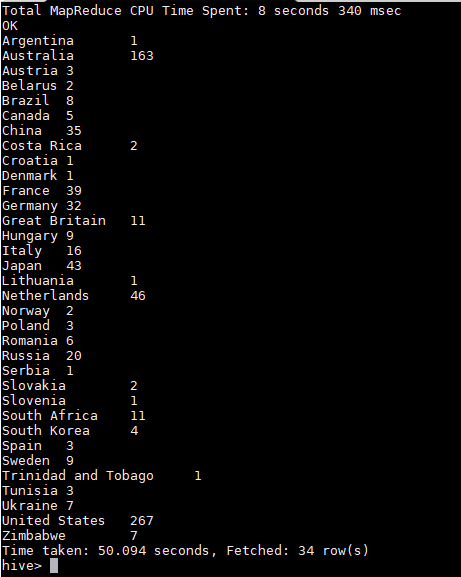
bronze ,

total\_medals,country from olympicdata;

**Questions**

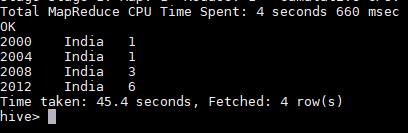
1. Write a Hive program to find the number of medals won by each country in swimming.

select country, SUM(total\_medals) from olympicdata\_part where sport='Swimming' group by country;



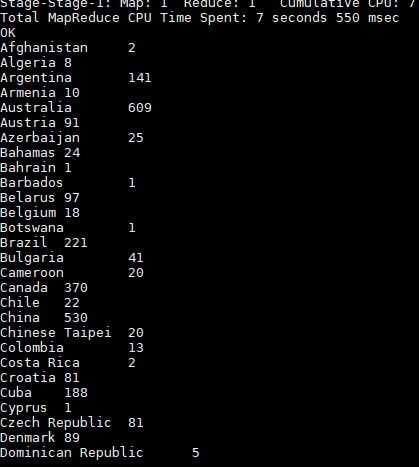
1. Write a Hive program to find the number of medals that India won year wise.

Select year, 'India',SUM(total\_medals) from olympicdata\_part where country='India' group by year;



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

Select country, SUM(total\_medals) from olympicdata\_part group by country;



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

Select country, SUM(gold) from olympicdata\_part group by country;

