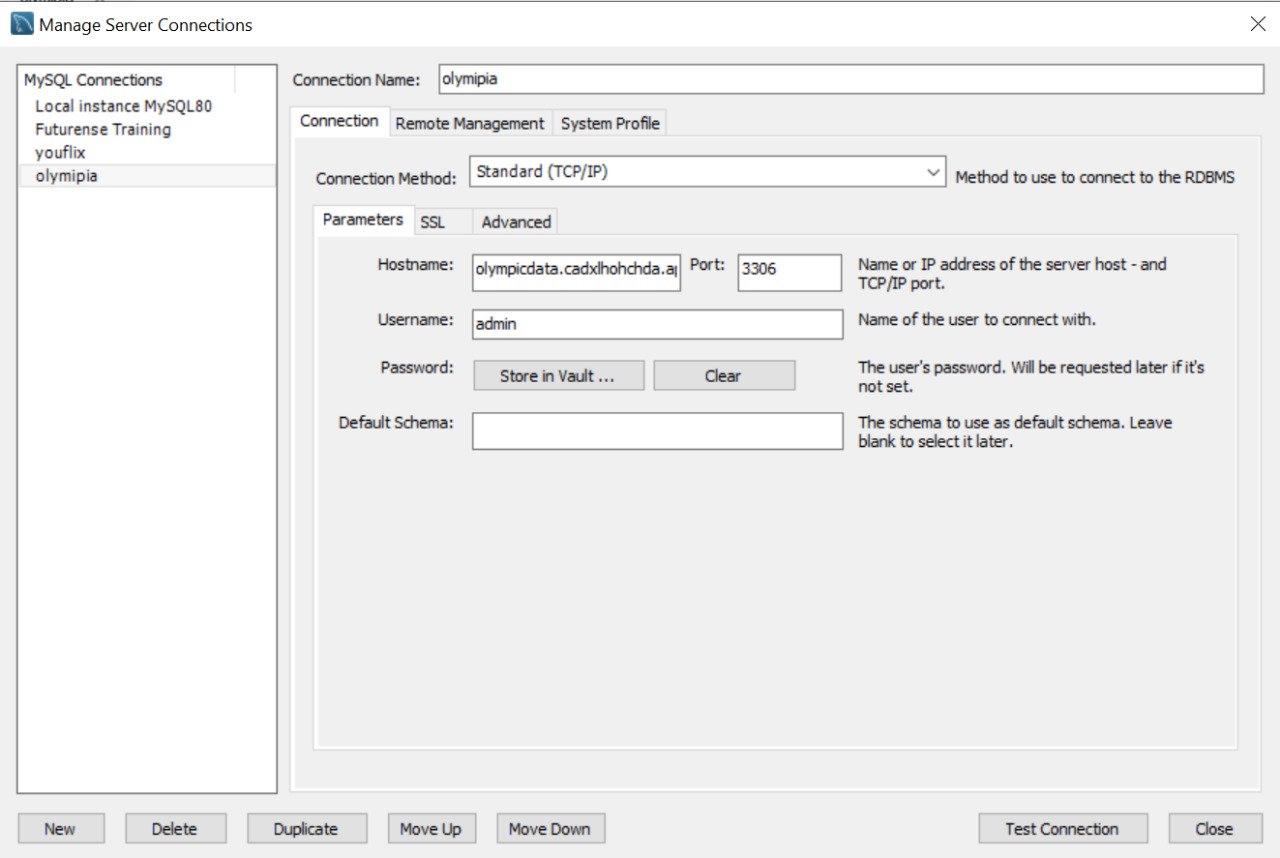
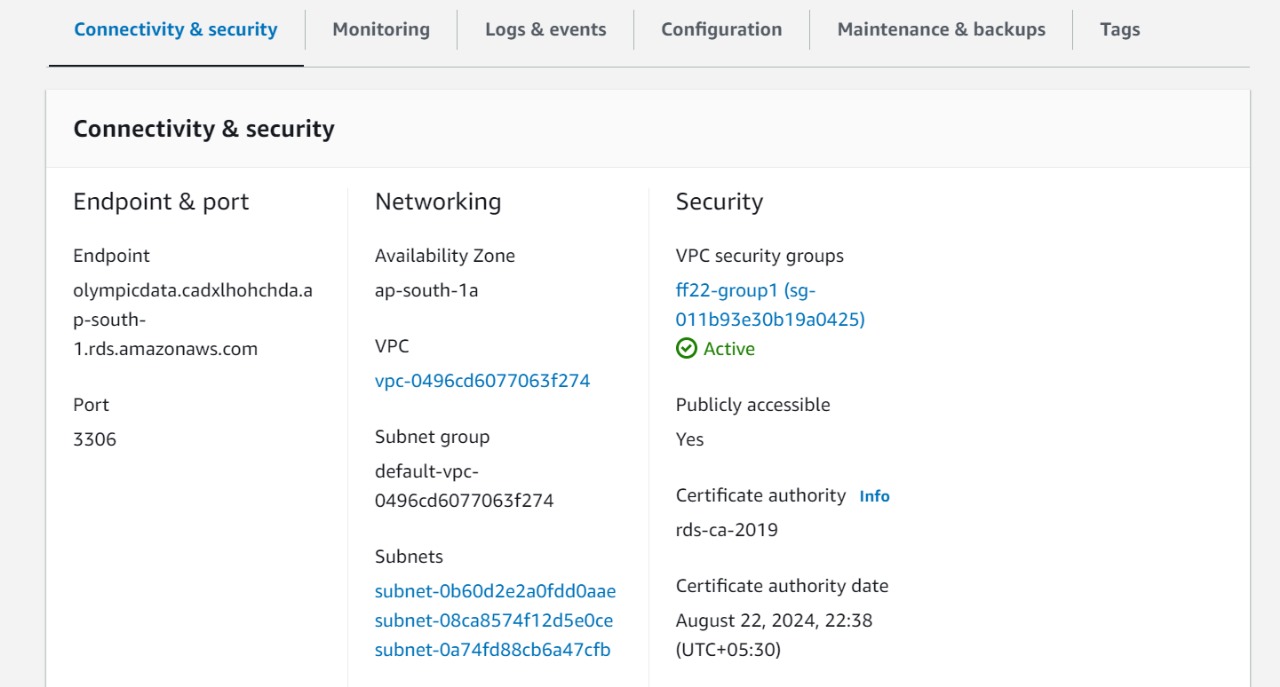
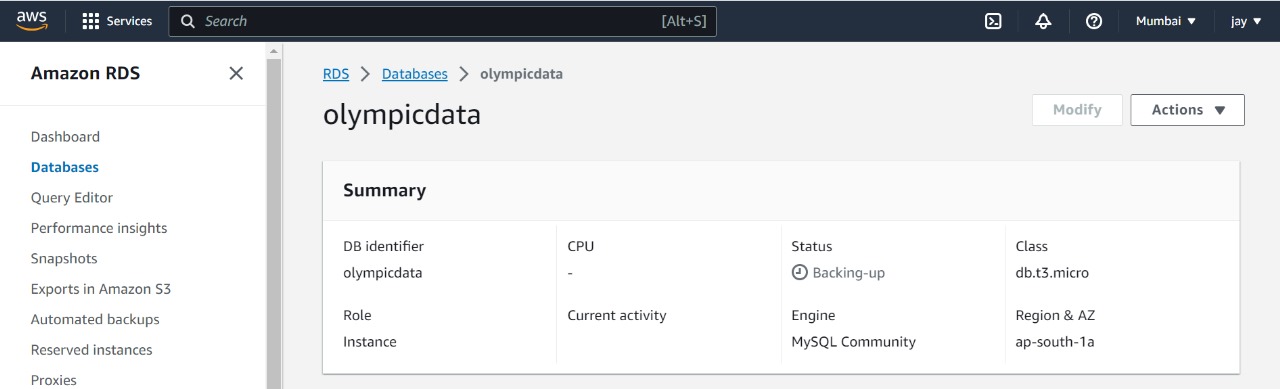
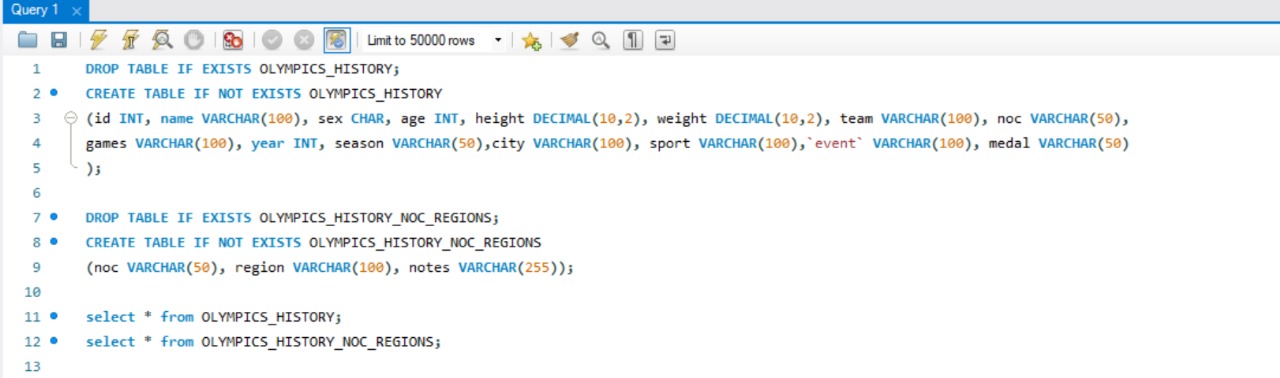
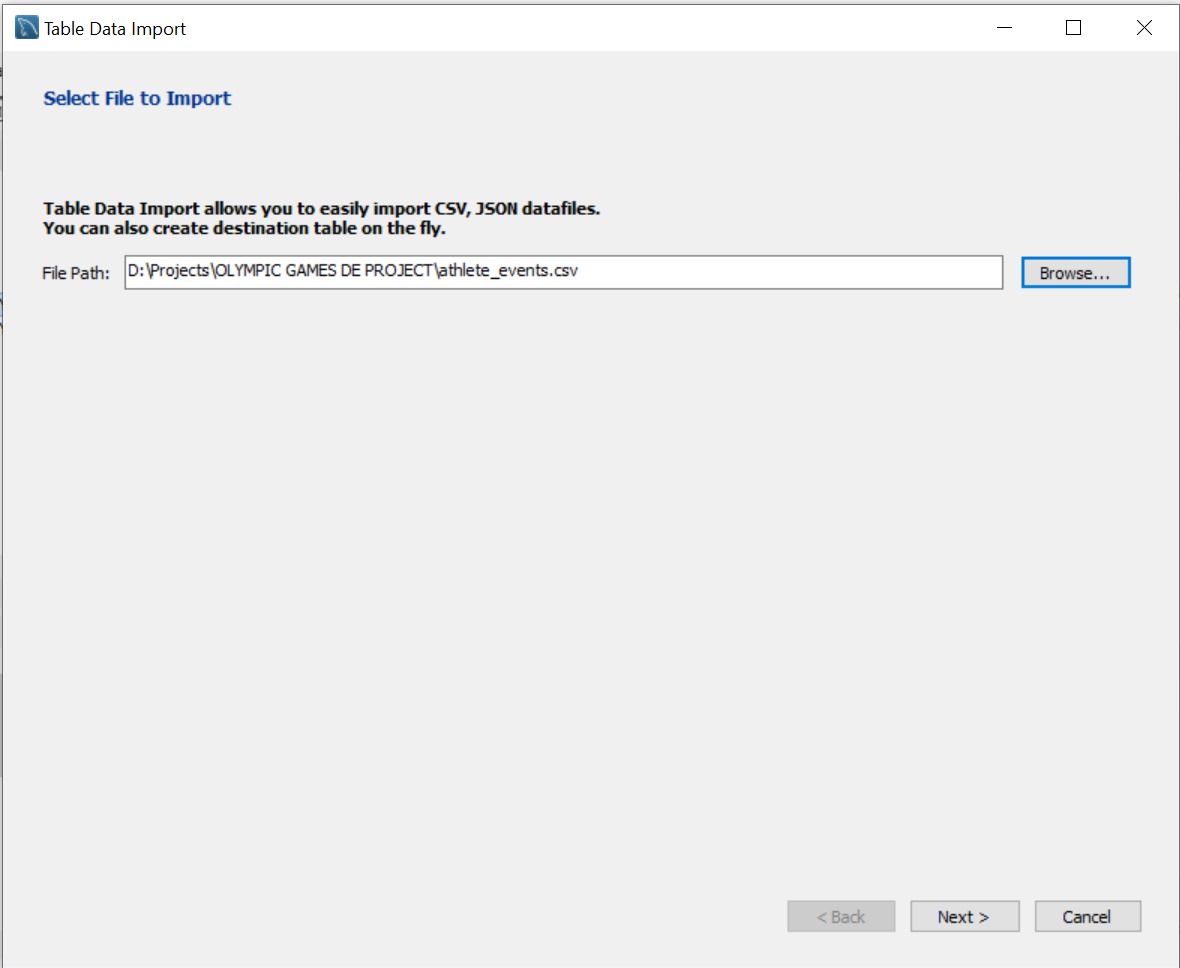
Step 1:- Creating Extraction pipelines from multiple client data sources to our data warehouse.

Pipeline 1- Amazon RDS or client relational database to Hive/HBase.

Ingestion Tool used: Sqoop





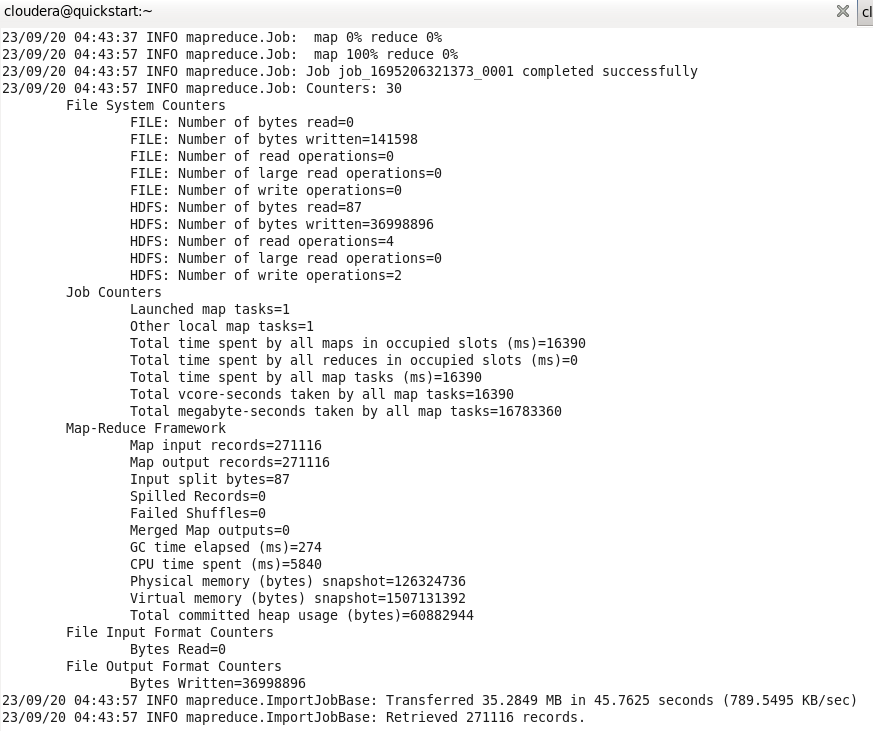




sqoop import-all-tables --connect jdbc:mysql://olympicdata.cadxlhohchda.ap-south-1.rds.amazonaws.com:3306/olympicdata --username admin --password olympia1 --hive-import --fields-terminated-by ',' -m 1;

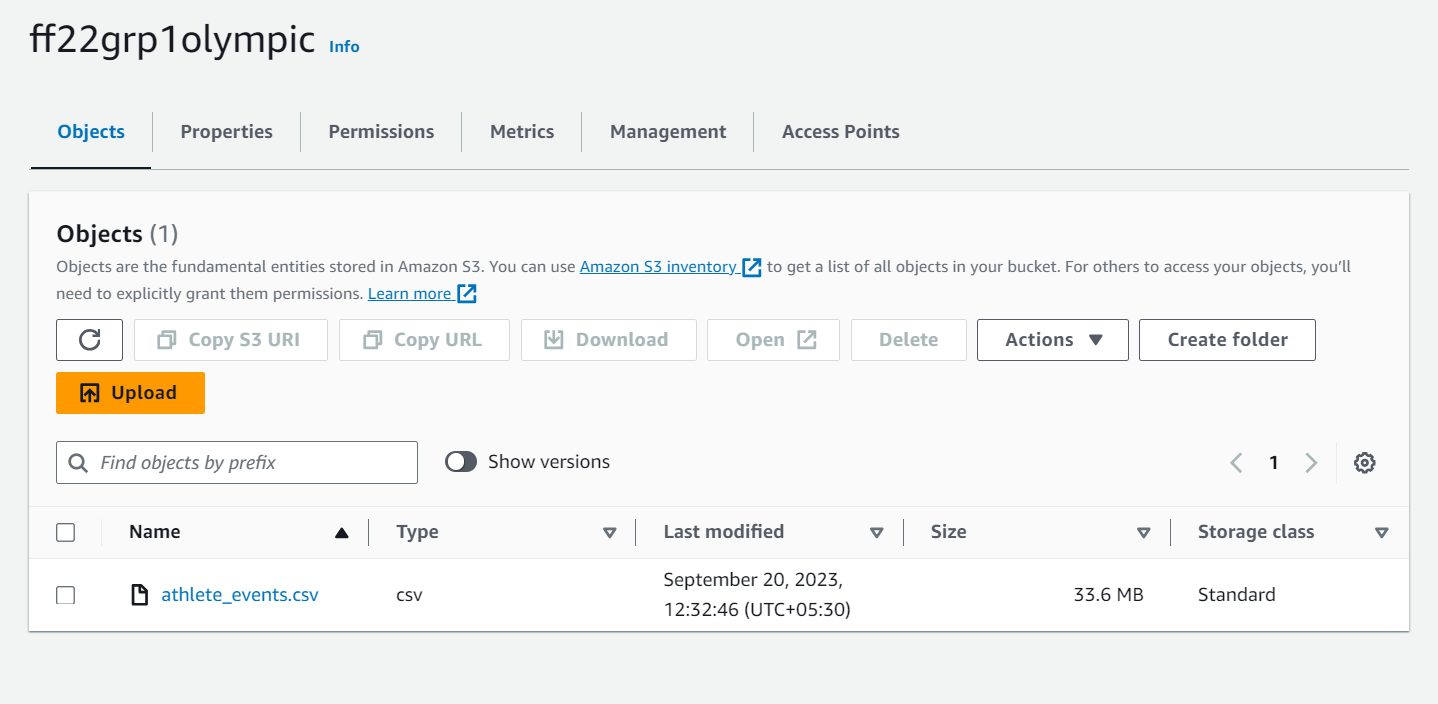


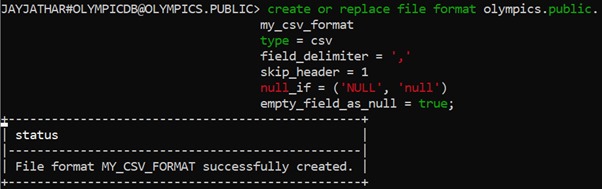
sqoop import-all-tables --connect jdbc:mysql://localhost:3306/olympicdb --username root --password cloudera –hive-import –fields-terminated-by ‘,’ -m 1;

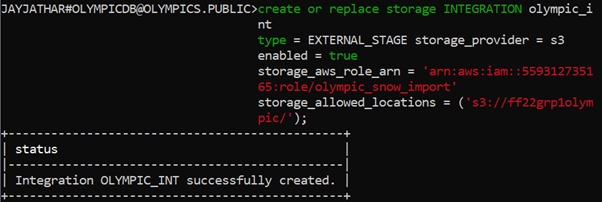


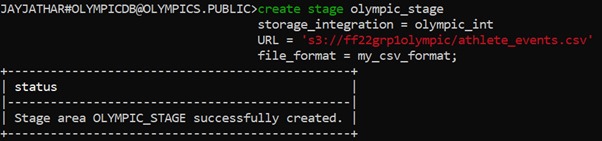
Pipeline 2- Amazon S3 Bucket to SnowFlake data warehouse.

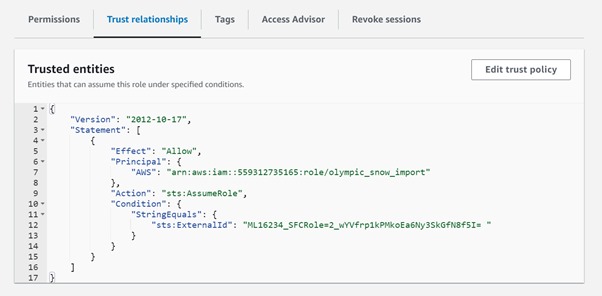
Ingestion Tool used: Copy into

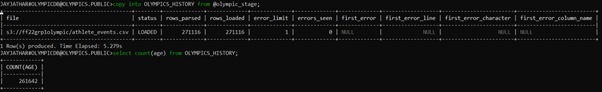






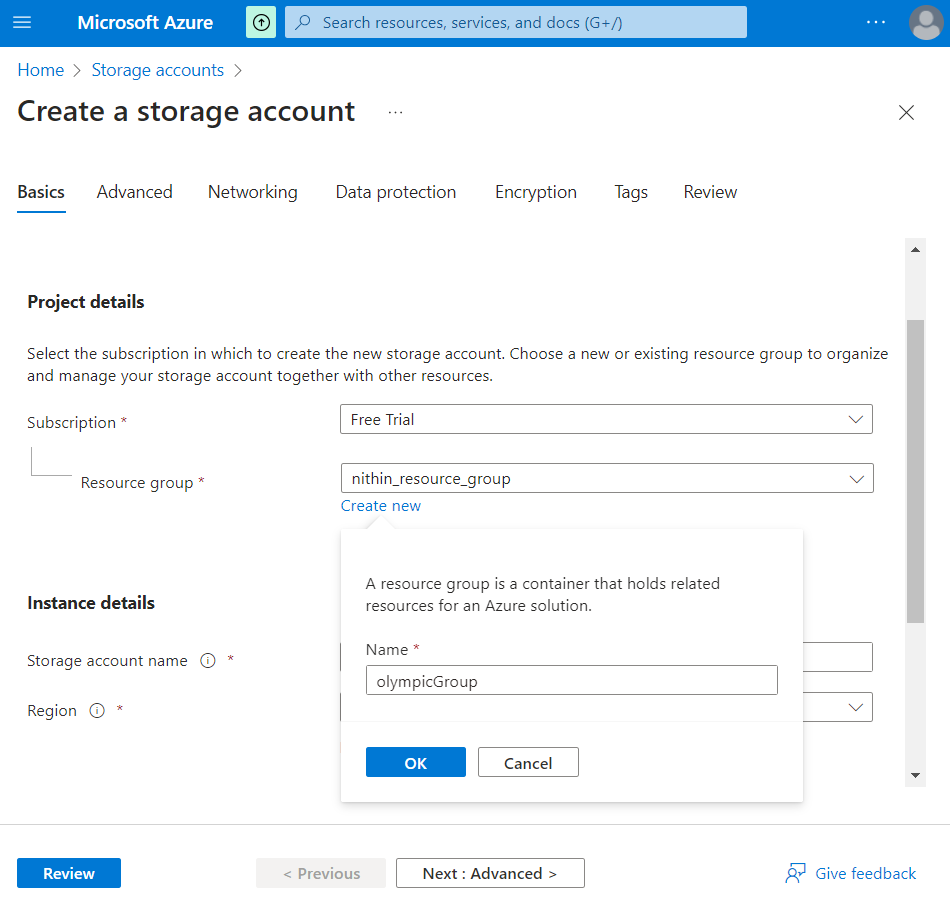


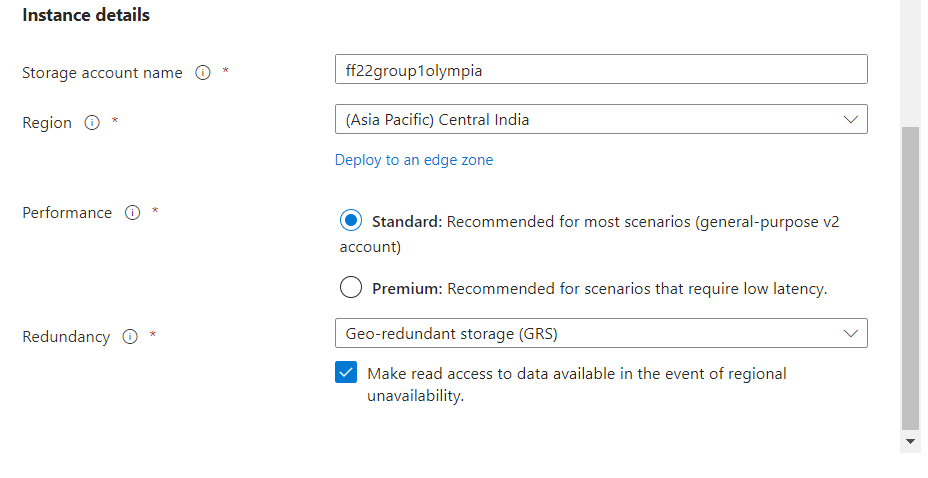


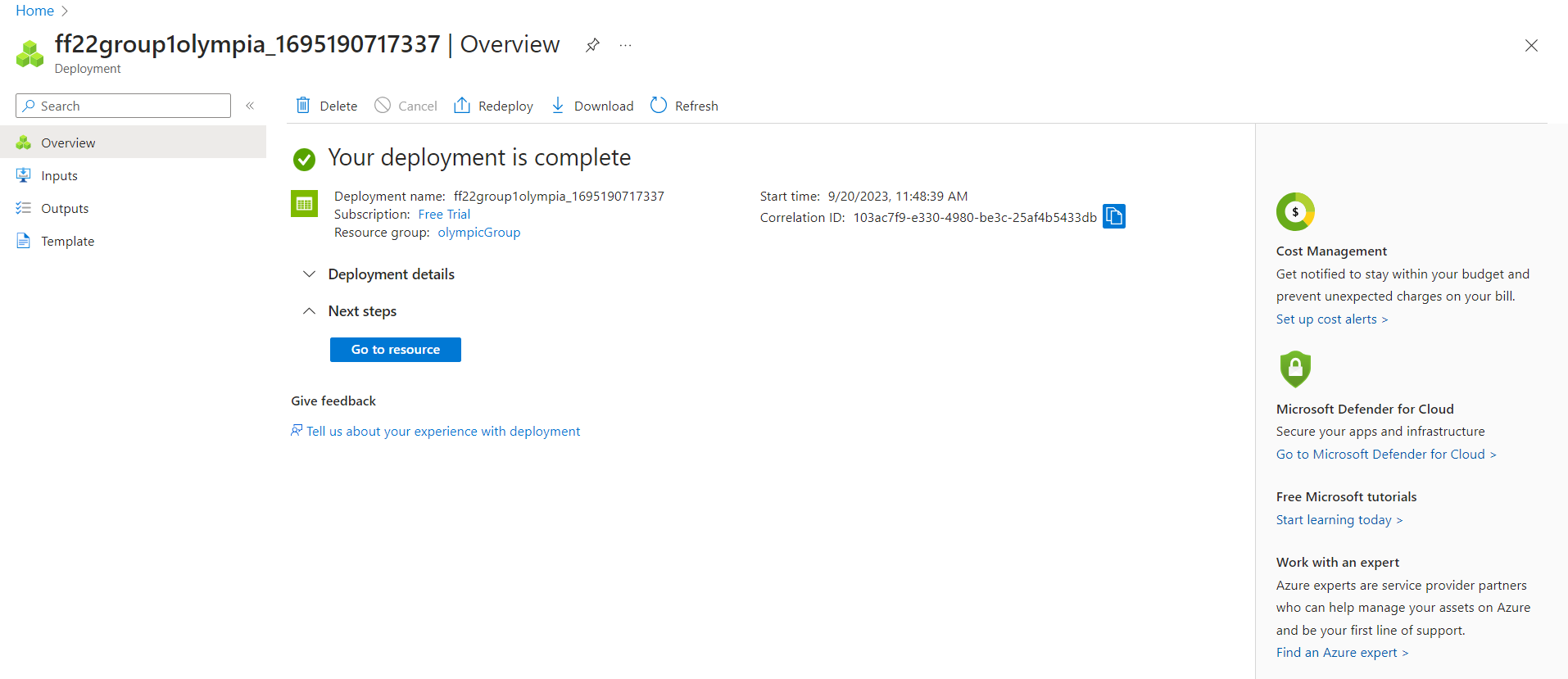


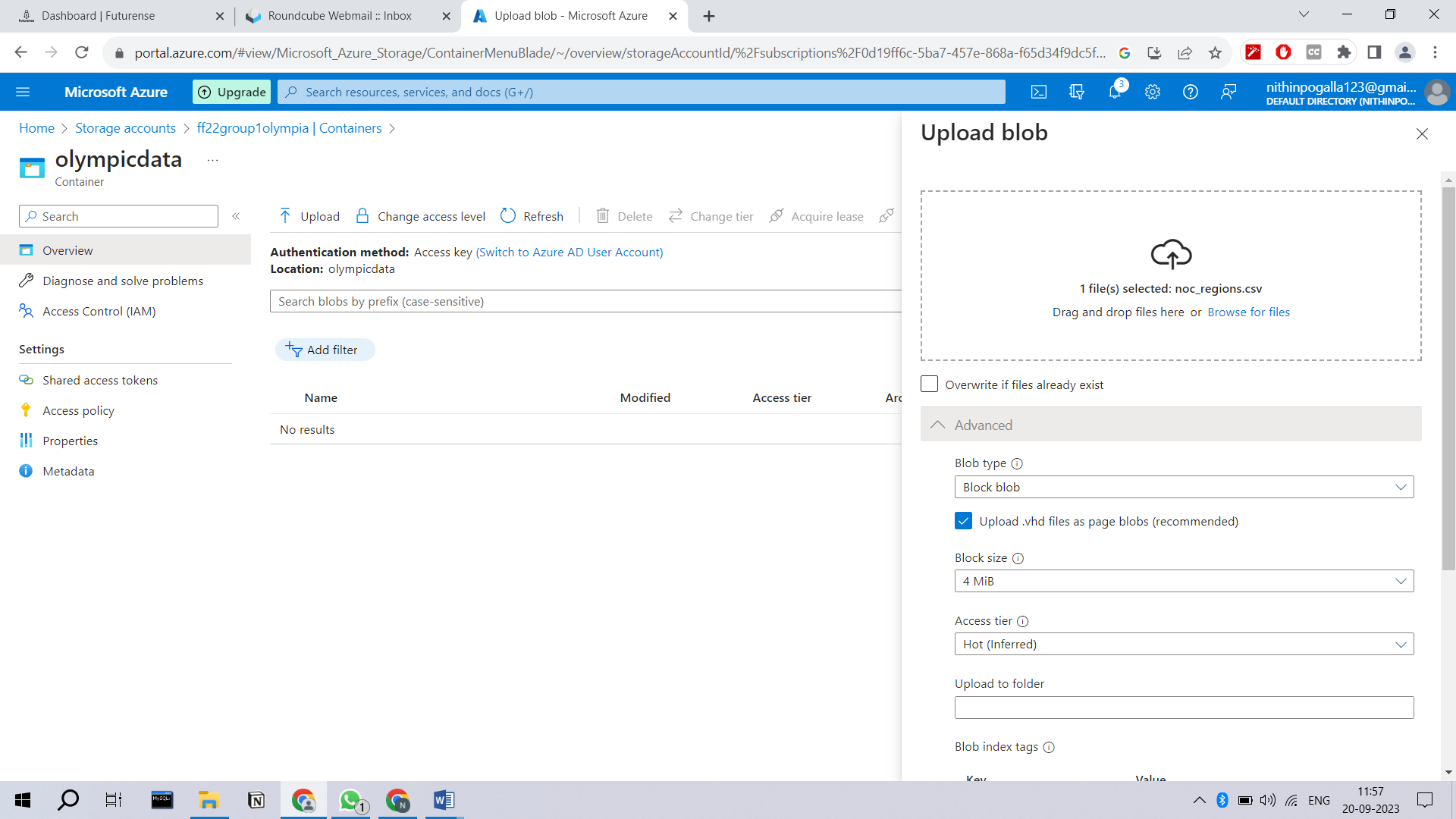
Pipeline 3- Azure Blob Container to SnowFlake data warehouse.

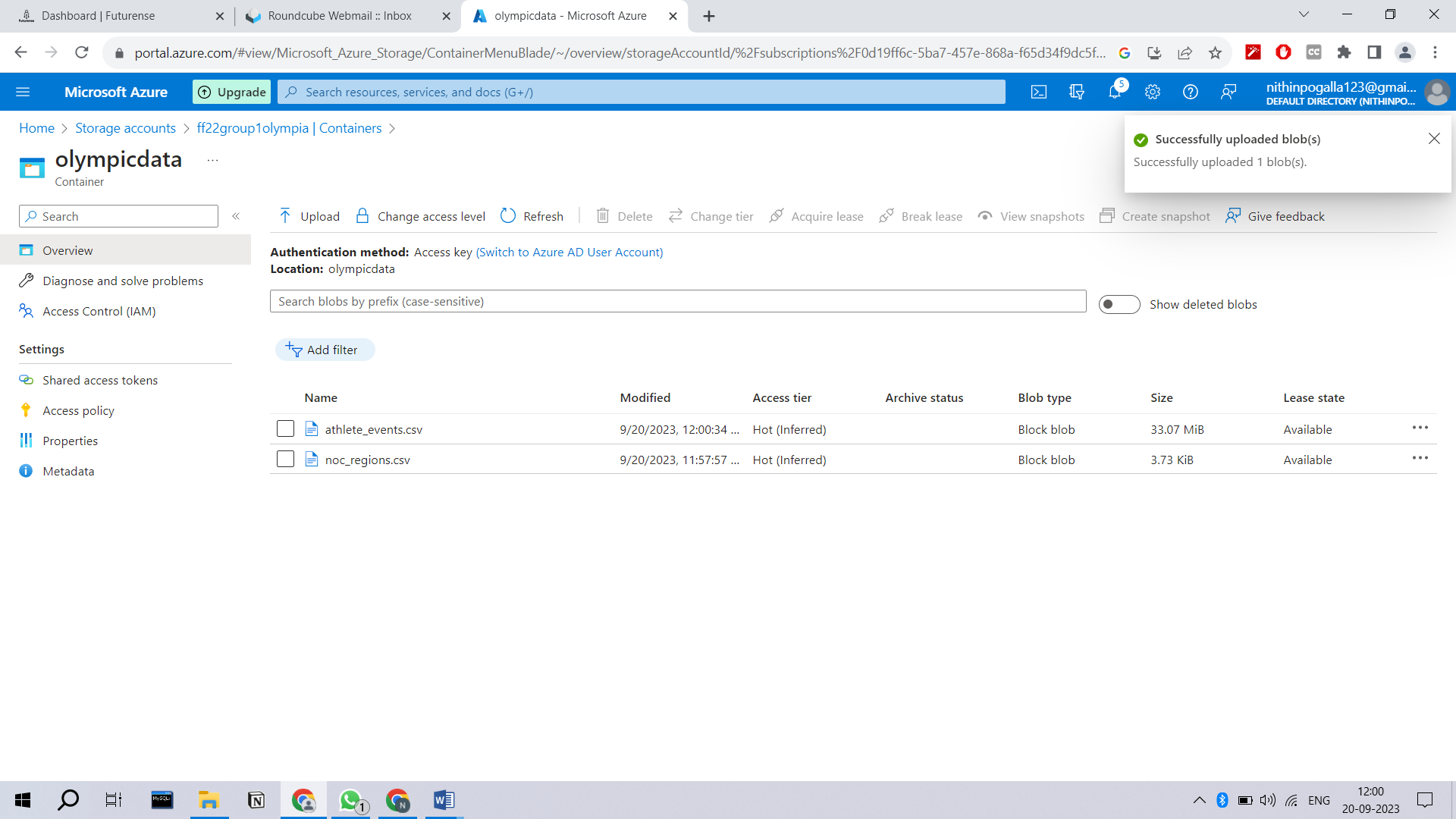
Ingestion Tool used: Snow Pipe

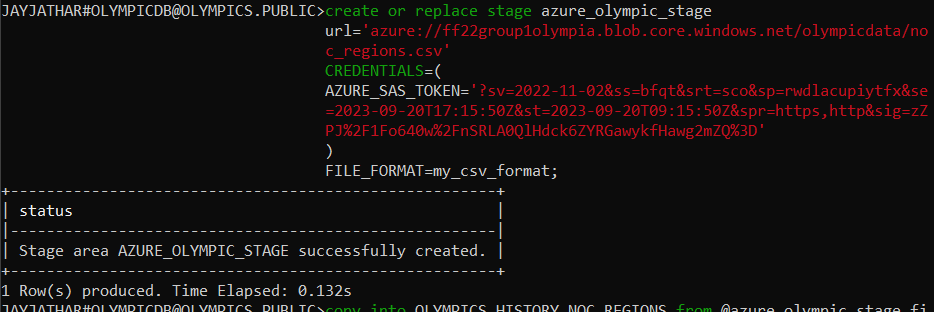


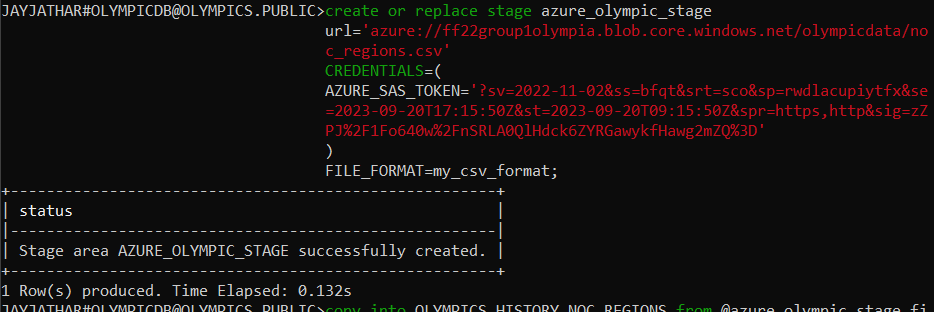












Step 2:- Transformation or Analysis of data in HIVE, HBASE, SNOWFLAKE according to the client requirements.

Analysis 1 - Snowfake

-- snowflake user stories

-- 1. How many olympics games have been held?

select count(distinct games) as Total\_Games from olympics\_history;

-- 2. List down all Olympics games held so far.

select distinct games from olympics\_history;

-- 3. Mention the total no of nations who participated in each olympics game?

select games , count(distinct n.region) as Total\_Nations from olympics\_history o join olympics\_history\_noc\_regions n on o.noc = n.noc group by games;

-- 4. Which year saw the highest and lowest no of countries participating in olympics?

with cte as (select year, count(distinct n.region) as Countries from olympics\_history o join olympics\_history\_noc\_regions n on o.noc = n.noc group by year)

select year, countries from (select year, countries, dense\_rank() over (order by countries) as l, dense\_rank() over (order by countries desc) as h from cte) where l = 1 or h = 1;

-- 5. Which nation has participated in all of the olympic games?

with cte as (select n.region as nations, count(distinct games) as game\_count from olympics\_history o join olympics\_history\_noc\_regions n on o.noc = n.noc group by nations)

select nations from cte where game\_count = (select count(distinct games) from olympics\_history);

-- 6. Identify the sport which was played in all summer olympics.

with cte as (select sport, count(distinct games) as game\_count from olympics\_history where season = 'Summer' group by sport)

select sport from cte where game\_count = (select count(distinct games) from olympics\_history where season = 'Summer');

-- 7. Which Sports were just played only once in the olympics?

select sport, count(distinct games) as game\_count from olympics\_history group by sport having game\_count = 1;

-- 8. Fetch the total no of sports played in each olympic games.

select games, count(distinct sport) as Total\_Sports from olympics\_history group by games;

-- 9. Fetch details of the oldest athletes to win a gold medal.

select name, age, games, sport, medal from olympics\_history where medal = 'Gold' and age = (select max(age) from olympics\_history where medal = 'Gold');

-- 10. Find the Ratio of male and female athletes participated in all olympic games

with cte as (select distinct name, sex from olympics\_history)

select sum(case when sex = 'M' then 1 else 0 end)/(select count(distinct name) from olympics\_history)\*100 as male\_ratio,

sum(case when sex = 'F' then 1 else 0 end)/(select count(distinct name) from olympics\_history)\*100 as female\_ratio from cte;

-- 11. Fetch the top 5 athletes who have won the most gold medals.

select name, count(medal) as medalsWon

from olympics\_history

where medal = 'Gold'

group by name

order by medalsWon desc limit 5;

-- 12. Fetch the top 5 athletes who have won the most medals (gold/silver/bronze).

select name, count(medal) as medalsWon

from olympics\_history

group by name

order by medalsWon desc limit 5;

-- 13. Fetch the top 5 most successful countries in olympics. Success is defined by no of medals won.

select nr.region, count(oh.medal) as medalsWon

from olympics\_history oh

join olympics\_history\_noc\_regions nr on oh.noc = nr.noc

group by nr.region

order by medalsWon desc limit 5;

-- 14. List down total gold, silver and broze medals won by each country.

with MedalCounts as (

select noc,

sum(case when medal = 'Gold' then 1 else 0 end) as GoldCount,

sum(case when medal = 'Silver' then 1 else 0 end) as SilverCount,

sum(case when medal = 'Bronze' then 1 else 0 end) as BronzeCount

from olympics\_history

group by noc

)

select nr.region,

sum(mc.GoldCount) as GoldWon,

sum(mc.SilverCount) as SilverWon,

sum(mc.BronzeCount) as BronzeWon

from MedalCounts mc

join olympics\_history\_noc\_regions nr on mc.noc = nr.noc

group by nr.region

order by GoldWon desc, SilverWon desc, BronzeWon desc;

-- 15. List down total gold, silver and bronze medals won by each country corresponding to each olympic games.

select n.region, games, sum(case when medal='Gold' then 1 else 0 end) as gold, sum(case when medal='Silver' then 1 else 0 end ) as silver, sum(case when medal='Bronze' then 1 else 0 end ) as bronze from olympics\_history o join olympics\_history\_noc\_regions n on o.noc = n.noc group by n.region, games;

-- 16. Identify which country won the most gold, most silver and most bronze medals in each olympic games.

with gold\_cte as (

select noc,games,region,

sum(case when medal='Gold' then 1 else 0 end )as gold\_medals,

rank() over(partition by games order by gold\_medals desc) as gold\_rank

from olympics\_history

join olympics\_history\_noc\_regions using(noc)

group by noc,games,region

),

silver\_cte as (

select noc,games,region,

sum(case when medal='Silver' then 1 else 0 end )as silver\_medals,

rank() over(partition by games order by silver\_medals desc) as silver\_rank

from olympics\_history

join olympics\_history\_noc\_regions using(noc)

group by noc,games,region

),

bronze\_cte as (

select noc,games,region,

sum(case when medal='Bronze' then 1 else 0 end )as bronze\_medals,

rank() over(partition by games order by bronze\_medals desc) as bronze\_rank

from olympics\_history

join olympics\_history\_noc\_regions using(noc)

group by noc,games,region

)

select games,g.region as g\_region,s.region as s\_region,b.region as b\_region

from gold\_cte g

join silver\_cte s using(games)

join bronze\_cte b using(games)

where gold\_rank=1 and silver\_rank=1 and bronze\_rank=1

order by games;

-- 17. Identify which country won the most gold, most silver, most bronze medals and the most medals in each olympic games.

with gold\_cte as (

select noc,games,region,

sum(case when medal='Gold' then 1 else 0 end )as gold\_medals,

rank() over(partition by games order by gold\_medals desc) as gold\_rank

from olympics\_history

join olympics\_history\_noc\_regions using(noc)

group by noc,games,region

),

silver\_cte as (

select noc,games,region,

sum(case when medal='Silver' then 1 else 0 end )as silver\_medals,

rank() over(partition by games order by silver\_medals desc) as silver\_rank

from olympics\_history

join olympics\_history\_noc\_regions using(noc)

group by noc,games,region

),

bronze\_cte as (

select noc,games,region,

sum(case when medal='Bronze' then 1 else 0 end )as bronze\_medals,

rank() over(partition by games order by bronze\_medals desc) as bronze\_rank

from olympics\_history

join olympics\_history\_noc\_regions using(noc)

group by noc,games,region

),

total\_cte as (

select noc,games,region,

count(medal) as total\_medals,

rank() over(partition by games order by total\_medals desc) as total\_rank

from olympics\_history

join olympics\_history\_noc\_regions using(noc)

group by noc,games,region

)

select games,g.region as g\_region,s.region as s\_region,b.region as b\_region,t.region as total\_region

from gold\_cte g

join silver\_cte s using(games)

join bronze\_cte b using(games)

join total\_cte t using(games)

where gold\_rank=1 and silver\_rank=1 and bronze\_rank=1 and total\_rank=1

order by games;

-- 18. Which countries have never won gold medal but have won silver/bronze medals?

with cte as(

select noc,

sum(

case

when medal='Gold' then 1

else 0

end

) as gold\_medals,

sum(

case

when medal='Bronze' or medal='Silver' then 1

else 0

end

) as other\_medals

from olympics\_history

group by noc)

select region from cte

join olympics\_history\_noc\_regions using(noc)

where gold\_medals=0 and other\_medals<>0;

-- 19. In which Sport/event, India has won highest medals.

select sport,count(medal) as medals from olympics\_history

join olympics\_history\_noc\_regions using(noc)

where region='India'

group by sport

order by medals desc limit 1;

select event,count(medal) as medals from olympics\_history

join olympics\_history\_noc\_regions using(noc)

where region='India'

group by event

order by medals desc limit 1;

-- 20. Break down all olympic games where india won medal for Hockey and how many medals in each olympic games.

select games,count(medal) as medals from olympics\_history

join olympics\_history\_noc\_regions using(noc)

where region='India' and sport like '%Hockey%'

group by games

having medals <>0

order by medals desc;

Analysis 2 – Hive

Analysis 3 – Hbase

Step 3:- Loading Results to client database or snowflake data warehouse.