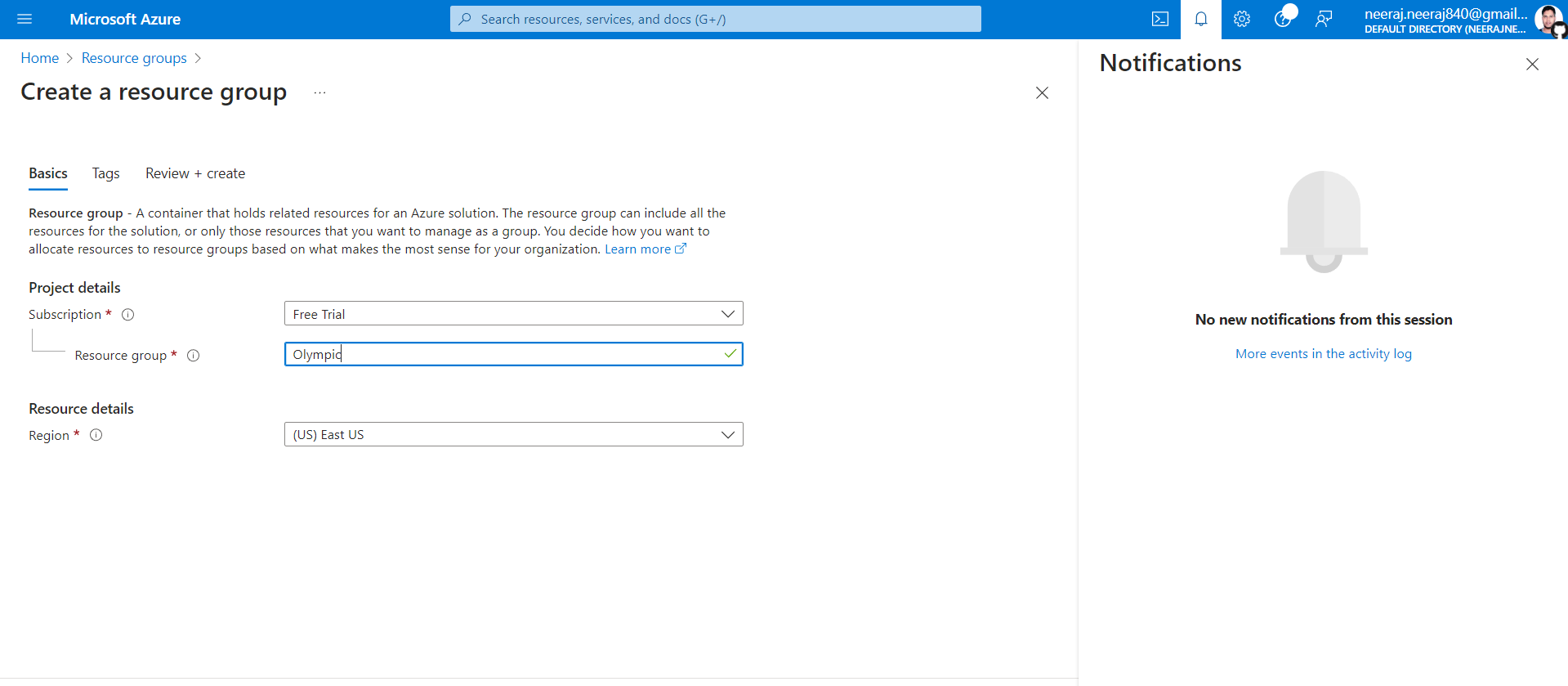
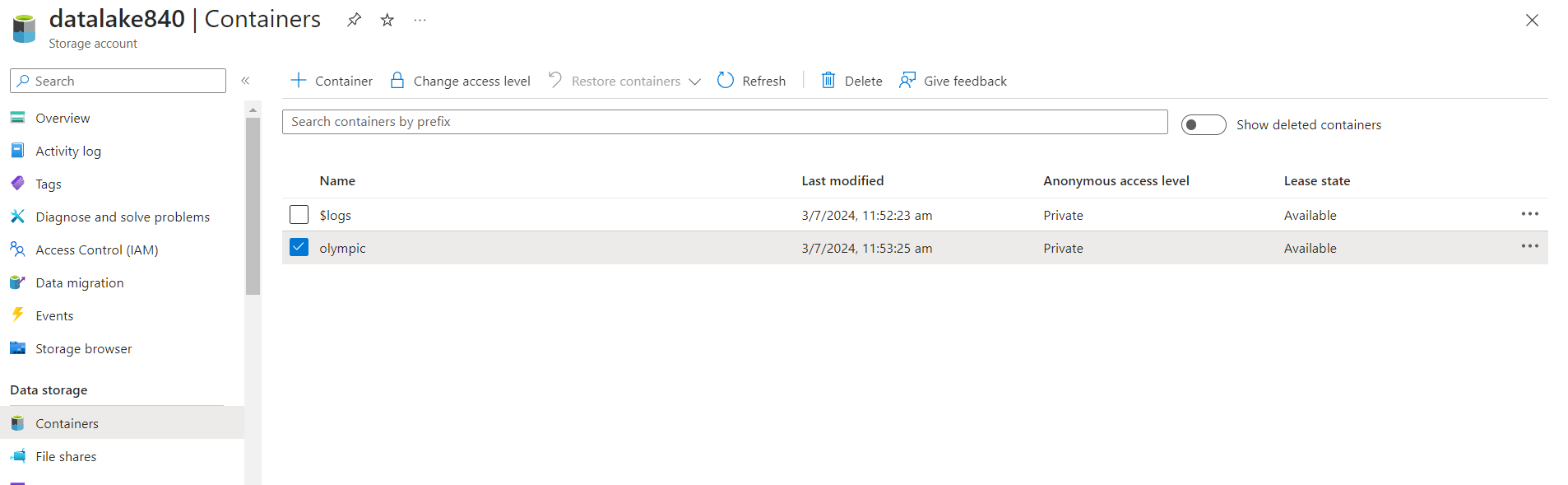
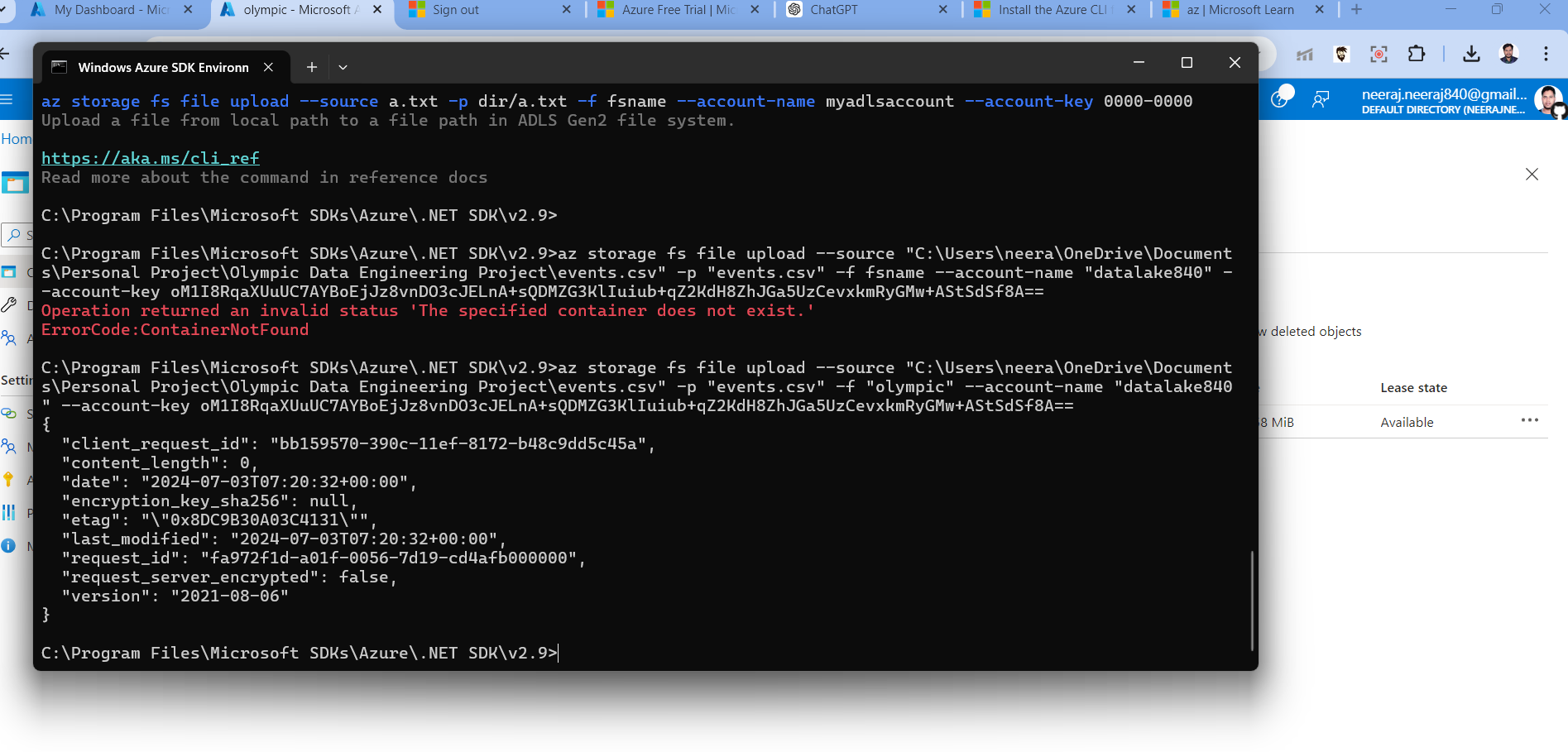
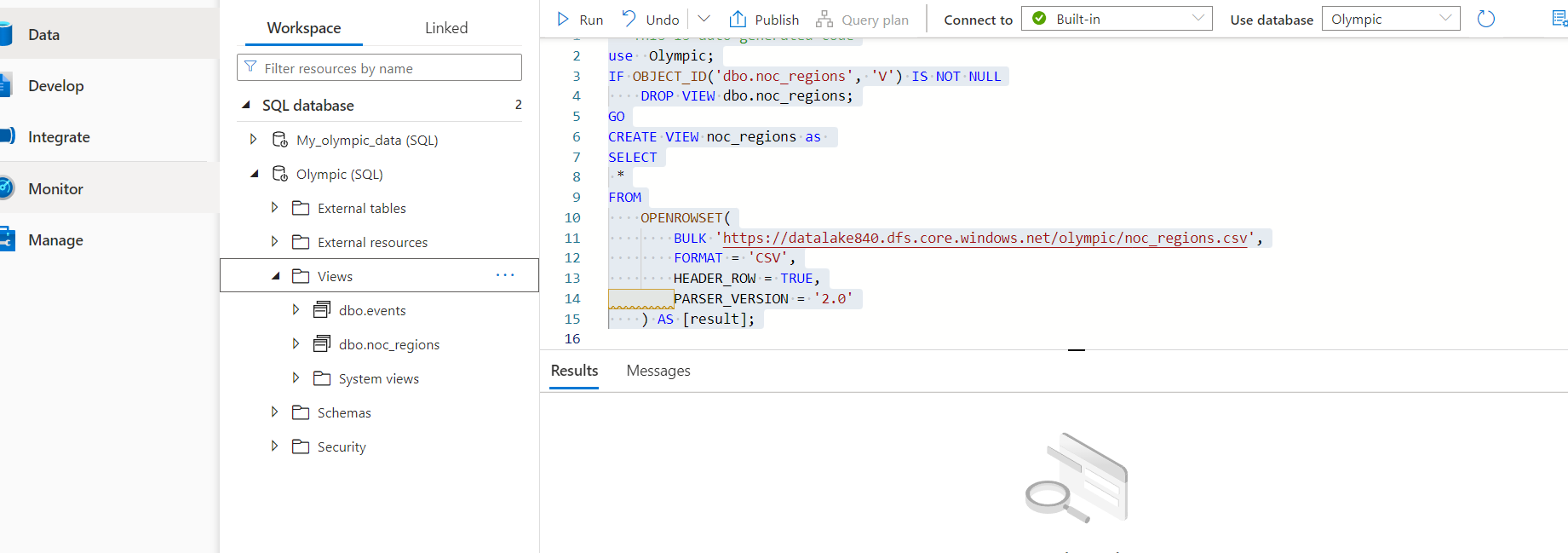
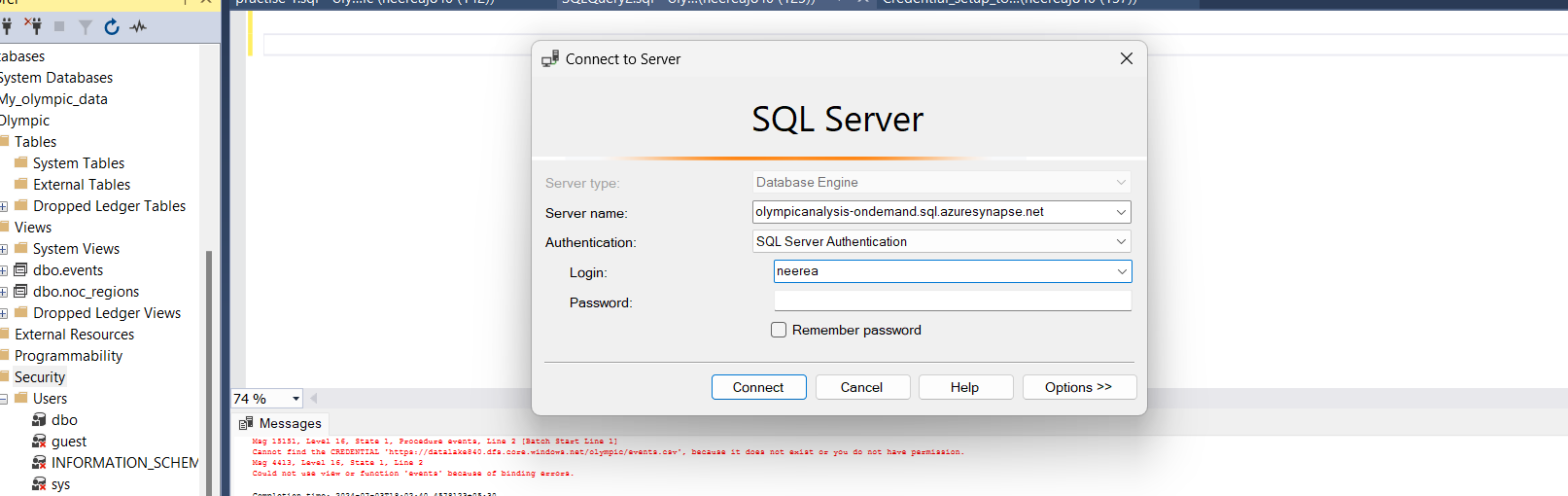
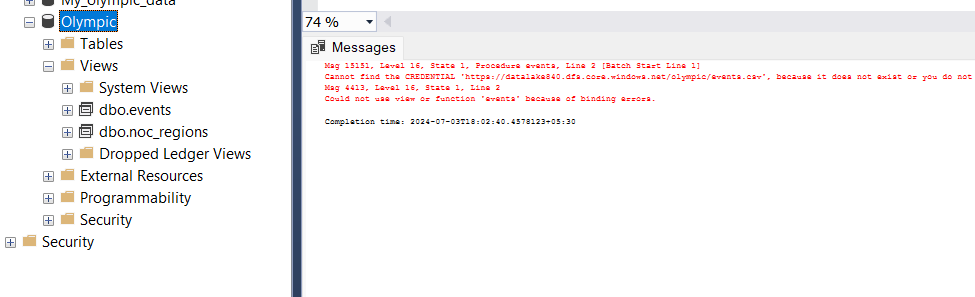
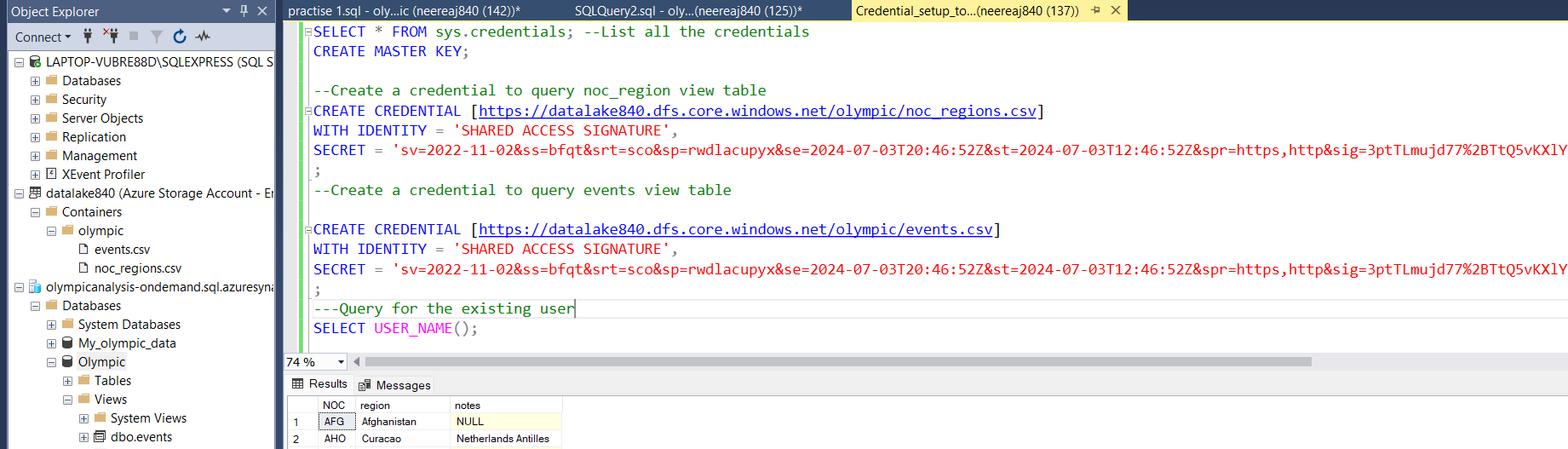
**Olympic Data Engineering Project**

1. **Created a Azure resource group**

****

1. **Create a data lake storage and add Olympic container.** ****
2. **Use Azure CLI command prompt to migrate csv from local to ADL storage:** 
3. **Create Azure Synapse Analytics resource Group and lauch synapse studio**
   1. **Create a database Olympic**
   2. **Query tables stored in ADL using linked service and store it as views in Olympic database.** ****
   3. **Refer to regionView.sql and eventsView.sql script file.**
4. **Stage the olympic Database on the premises SSMS server by connecting using Database Engine and providing Azure SQL server credentials.** ****
   1. **Database has been mounted on SSMS** ****
5. **In order to Query view table on SSMS need to setup credentials.**
   1. **Generate SAS token to authenticate ADL.**
   2. **Use that token and tables file path stored in ADL to create credentials for all required tables.** ****

* 1. **Now views have been authenticated and could be read.**

1. **Now query views to derive information from tables**

**use olympic;**

**describe events;**

**-- STEP 1. Understand the dataxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx**

**SELECT \* FROM events;**

**select \* from Noc\_regions;**

**select count(\*) from Noc\_regions;**

**with t as (SELECT distinct(region) FROM NOC\_REGIONS)**

**select count(\*) from t;**

**select count(\*) from (SELECT distinct(noc) FROM NOC\_REGIONS) ;**

**select \* from events order by year;**

**-- Data Analysis xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx**

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

**-- Ans 51**

**with oly as**

**(Select year, season from events**

**group by year, season**

**)**

**select count(\*) from oly;**

**-- xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx**

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

**SELECT distinct YEAR, SEASON , CITY FROM events**

**order by year;**

**-- xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx**

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

**with all\_countries as**

**(select games, nr.region**

**from events e**

**join noc\_regions nr ON nr.noc = e.noc**

**group by games, nr.region)**

**select games, count(region)**

**from all\_countries**

**group by games**

**order by games;**

**-- xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx**

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

**with all\_countries as**

**(select games, nr.region**

**from events e**

**join noc\_regions nr ON nr.noc = e.noc**

**group by games, nr.region),**

**total\_countries as**

**(select games, count(region) total\_country**

**from all\_countries**

**group by games),**

**ans as (select concat(games,"\_",total\_country) max\_min\_country from total\_countries where**

**total\_country=(select max(total\_country) from total\_countries) or**

**total\_country=(select min(total\_country) from total\_countries))**

**select a.max\_min\_country lowest\_Country,b.max\_min\_country Highest\_Country**

**from ans a, ans b**

**where a.max\_min\_country != b.max\_min\_country**

**limit 1**

**;**

**-- xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxXXXXXXXXXX**

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

**with plays as**

**(select nr.region, games, count(games) over(partition by nr.region) games\_participated**

**from events e**

**join noc\_regions nr ON nr.noc = e.noc**

**group by nr.region,games)**

**Select distinct region, games\_participated from plays**

**where games\_participated= (select max(games\_participated) from plays) ;**

**-- xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxXXXXXXXXXX**

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

**select \* from events;**

**with summer as**

**(select year, season from events**

**where season="Summer"**

**group by year, season),**

**SPORTS AS (**

**select year,sport, count(sPORT)over(partition by sport) SUMMER\_COUNT from events**

**where season="Summer"**

**group by year,sport**

**order by sport)**

**select DISTINCT SPORT, SUMMER\_COUNT from SPORTS**

**WHERE SUMMER\_COUNT= (SELECT COUNT(\*) FROM SUMMER)**

**;**

**-- xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxXXXXXXXXXX**

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

**with game\_no as**

**(select games,sport, count(sPORT)over(partition by sport) games\_count from events**

**group by games,sport**

**order by sport)**

**select games , sport, games\_count from game\_no**

**where games\_count=1;**

**-- xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxXXXXXXXXXX**

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

**with ngames as**

**(select games, sport from events**

**group by games,sport)**

**select distinct games, count(sport) over(partition by games order by games) No\_of\_sports**

**from ngames;**

**-- xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxXXXXXXXXXX**

**-- 9. Fetch oldest athletes to win a gold medal**

**describe events;**

**-- convert age, height, weight datatype from text to int**

**-- step 1. replace NA with null**

**SET SQL\_SAFE\_UPDATES = 0; -- unable safe update mode**

**update events**

**set age=null where age="NA";**

**update events**

**set height = null where height = 'NA';**

**update events**

**set weight=null where weight="NA";**

**-- -- step 2. Modify the data type**

**alter table events**

**modify column height int,**

**modify column weight int,**

**modify column age int;**

**--- Now solve**

**select \* from events;**

**select distinct medal from events;**

**select \* from events**

**where medal="Gold" and age=(select max(age) from events where Medal="gold")**

**order by age;**

**-- xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxXXXXXXXXXX**

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

**with t1a as**

**(select sex**

**from events**

**group by id,sex),**

**t1 as**

**(select sex, count(1) as cnt**

**from t1a**

**group by sex),**

**t2 as**

**(select \*, row\_number() over(order by cnt) as rn**

**from t1),**

**min\_cnt as**

**(select cnt from t2 where rn = 1),**

**max\_cnt as**

**(select cnt from t2 where rn = 2)**

**select concat('1 : ', round(max\_cnt.cnt/min\_cnt.cnt, 2)) as ratio**

**from min\_cnt, max\_cnt;**

**select sex, count(1) as cnt**

**from events**

**group by sex;**

**with t1a as**

**(select sex**

**from events**

**group by id,sex),**

**t1 as**

**(select sex, count(1) as cnt**

**from t1a**

**group by sex)**

**select concat('1 : ', round((select cnt from t1 where sex='M')/(select cnt from t1 where sex='F'), 2)) as ratio**

**from t1;**

**-- xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxXXXXXXXXXX**

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

**SELECT \* FROM EVENTS;**

**with player as**

**(select id, Name, medal, Count(medal) medal\_no from events**

**where medal="Gold"**

**group by id, Name)**

**select \* from player order by medal\_no desc**

**limit 5;**

**-- xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxXXXXXXXXXX**

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

**with player as**

**(select id, Name, Count(medal) medal\_no from events**

**where medal !="NA"**

**group by id, Name)**

**select \* from player order by medal\_no desc**

**limit 5;**

**-- xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxXXXXXXXXXX**

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

**with country as**

**( select team,count(medal) n\_medal from events**

**where medal!="NA"**

**group by team)**

**select \*, rank() over(order by n\_medal desc) RANKS from country;**

**-- xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx**

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

**SELECT \* FROM EVENTS;**

**select team ,medal,count(medal) counts from events where medal !="NA" group by team,medal order by team ;**

**select region ,medal,count(medal) counts from events e join noc\_regions r on e.noc=r.noc**

**where medal !="NA"**

**group by region ,medal**

**order by region;**

**with medal as**

**(select region ,medal,count(medal) counts from events e join noc\_regions r on e.noc=r.noc**

**where medal !="NA"**

**group by region ,medal**

**order by region),**

**medal\_1 as**

**(SELECT**

**region,**

**(CASE WHEN medal = "Gold" THEN counts ELSE 0 END) AS 'Gold',**

**(CASE WHEN medal = "Silver" THEN counts ELSE 0 END) AS 'Silver',**

**(CASE WHEN medal = "Bronze" THEN counts ELSE 0 END) AS 'Bronze'**

**FROM medal)**

**select region, sum(gold) gold, sum(Silver) as silver, sum(Bronze) bronze from medal\_1**

**group by region**

**order by region;**

**-- xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx**

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

**with medal as**

**(select games,region ,medal,count(medal) counts from events e join noc\_regions r on e.noc=r.noc**

**where medal !="NA"**

**group by games, region ,medal**

**order by region),**

**medal\_1 as**

**(SELECT**

**games,region,**

**(CASE WHEN medal = "Gold" THEN counts ELSE 0 END) AS 'Gold',**

**(CASE WHEN medal = "Silver" THEN counts ELSE 0 END) AS 'Silver',**

**(CASE WHEN medal = "Bronze" THEN counts ELSE 0 END) AS 'Bronze'**

**FROM medal)**

**select games,region, sum(gold) gold, sum(Silver) as silver, sum(Bronze) bronze from medal\_1**

**group by games,region**

**order by region;**

**-- xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx**

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

**with region as**

**(select games, region ,medal,count(medal) counts from events e join noc\_regions r on e.noc=r.noc**

**where medal !="NA"**

**group by games,region ,medal**

**order by region),**

**ranks as**

**(select \*,concat(region,"-",counts) as ans, rank() over(partition by games,medal order by counts desc) ranks from region),**

**ans as**

**(select \* from ranks where ranks=1)**

**select s.games, g.max\_gold, s.max\_silver, b.max\_bronze from (select games,ans max\_gold from ans where medal="Gold") g join (select games,ans max\_silver from ans where medal="Silver") as s on g.games=s.games**

**join (select games, ans max\_bronze from ans where medal="Bronze") as b on s.games=b.games;**

**-- xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx**

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

**with total as**

**(select games, region ,count(medal) count, concat(region,"-",count(medal) ) ans from events e join noc\_regions r on e.noc=r.noc**

**where medal !="NA"**

**group by games,region**

**order by region),**

**total\_medal as**

**(select \*, rank()over(partition by games order by count desc) ranks from total) ,**

**total\_ans as**

**(select games, ans from total\_medal where ranks=1),**

**region as**

**(select games, region ,medal,count(medal) counts from events e join noc\_regions r on e.noc=r.noc**

**where medal !="NA"**

**group by games,region ,medal**

**order by region),**

**ranks as**

**(select \*,concat(region,"-",counts) as ans, rank() over(partition by games,medal order by counts desc) ranks from region),**

**ans as**

**(select \* from ranks where ranks=1)**

**select s.games, g.max\_gold, s.max\_silver, b.max\_bronze, ta.total\_medal from (select games,ans max\_gold from ans where medal="Gold") g join (select games,ans max\_silver from ans where medal="Silver") as s on g.games=s.games**

**join (select games, ans max\_bronze from ans where medal="Bronze") as b on s.games=b.games**

**join (select games, ans total\_medal from total\_ans) ta on ta.games=s.games;**

**-- xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx**

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

**with medal as**

**(select region ,medal,count(medal) counts from events e join noc\_regions r on e.noc=r.noc**

**where medal !="NA"**

**group by region ,medal**

**order by region),**

**medal\_1 as**

**(SELECT**

**region,**

**(CASE WHEN medal = "Gold" THEN counts ELSE 0 END) AS 'Gold',**

**(CASE WHEN medal = "Silver" THEN counts ELSE 0 END) AS 'Silver',**

**(CASE WHEN medal = "Bronze" THEN counts ELSE 0 END) AS 'Bronze'**

**FROM medal),**

**no\_gold as**

**(select region, sum(gold) gold, sum(Silver) as silver, sum(Bronze) bronze from medal\_1**

**group by region**

**order by region)**

**select \* from no\_gold**

**where gold=0 and (silver!=0 or bronze!=0) ;**

**-- xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx**

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

**--Ans Hockey**

**select \* from events;**

**select region,sport ,count(medal) counts from events e join noc\_regions r on e.noc=r.noc**

**where region='India' and medal is not NULL**

**group by region ,sport**

**order by counts desc**

**;**

**-- xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx**

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

**select region,sport ,games, count(medal) counts from events e join noc\_regions r on e.noc=r.noc**

**where region="India" and medal!="NA" and sport="Hockey"**

**group by region ,sport,games**

**order by games;**

**-- xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx**

**--**