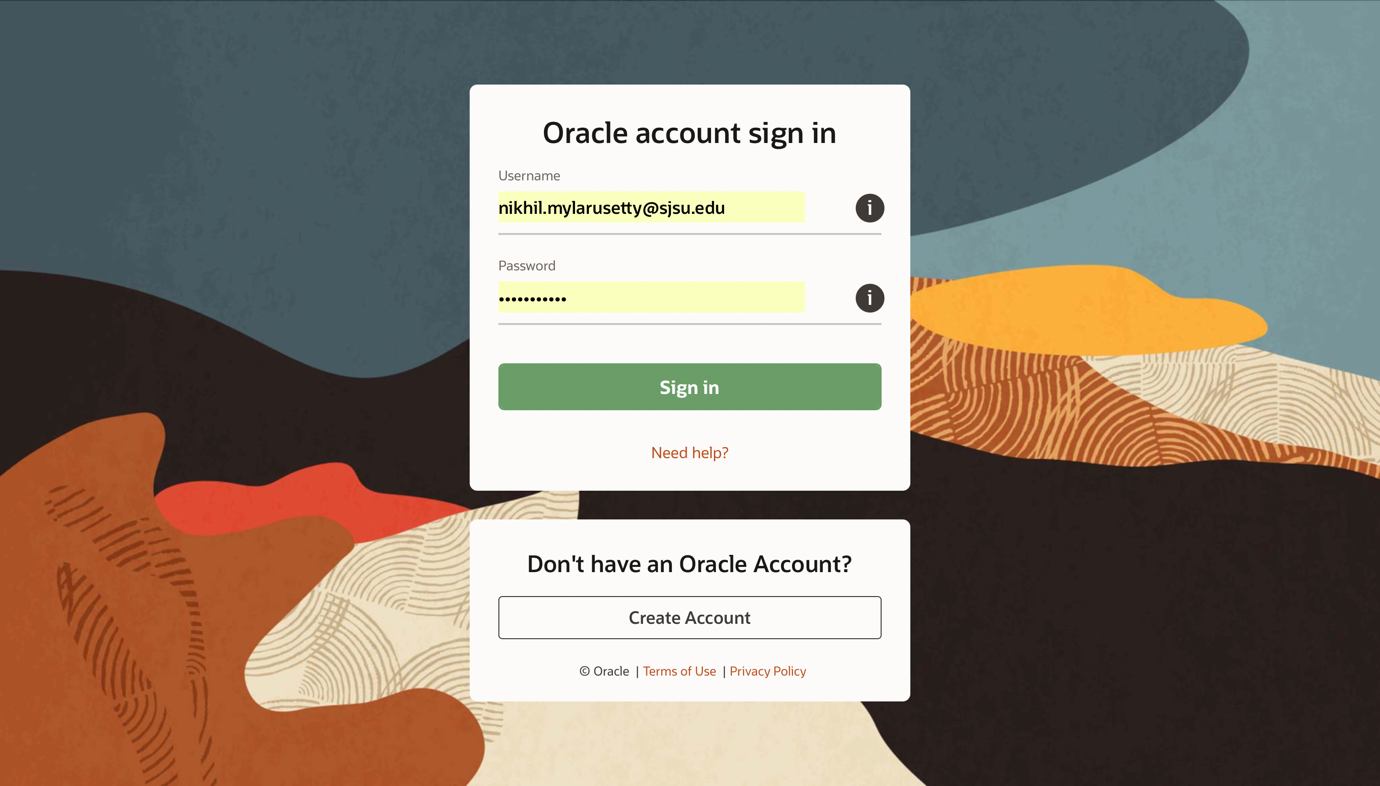
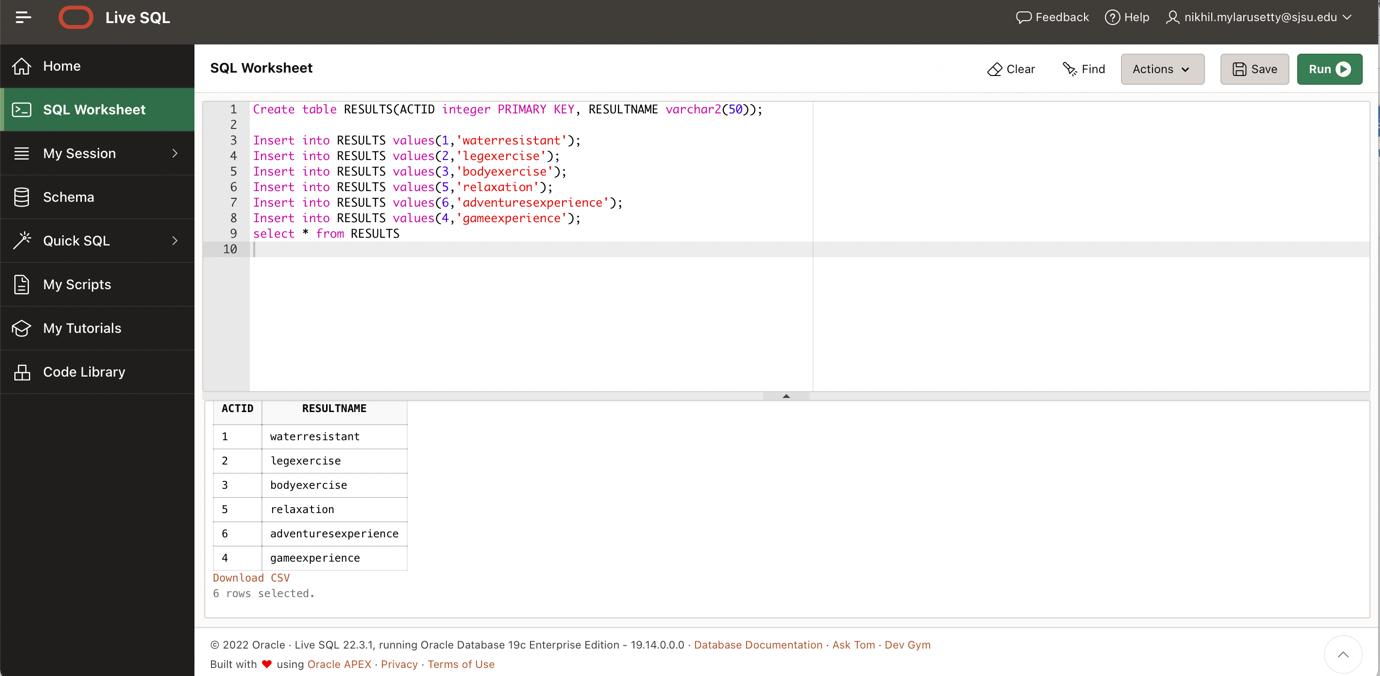
**Database Assignment(Homework 5)**

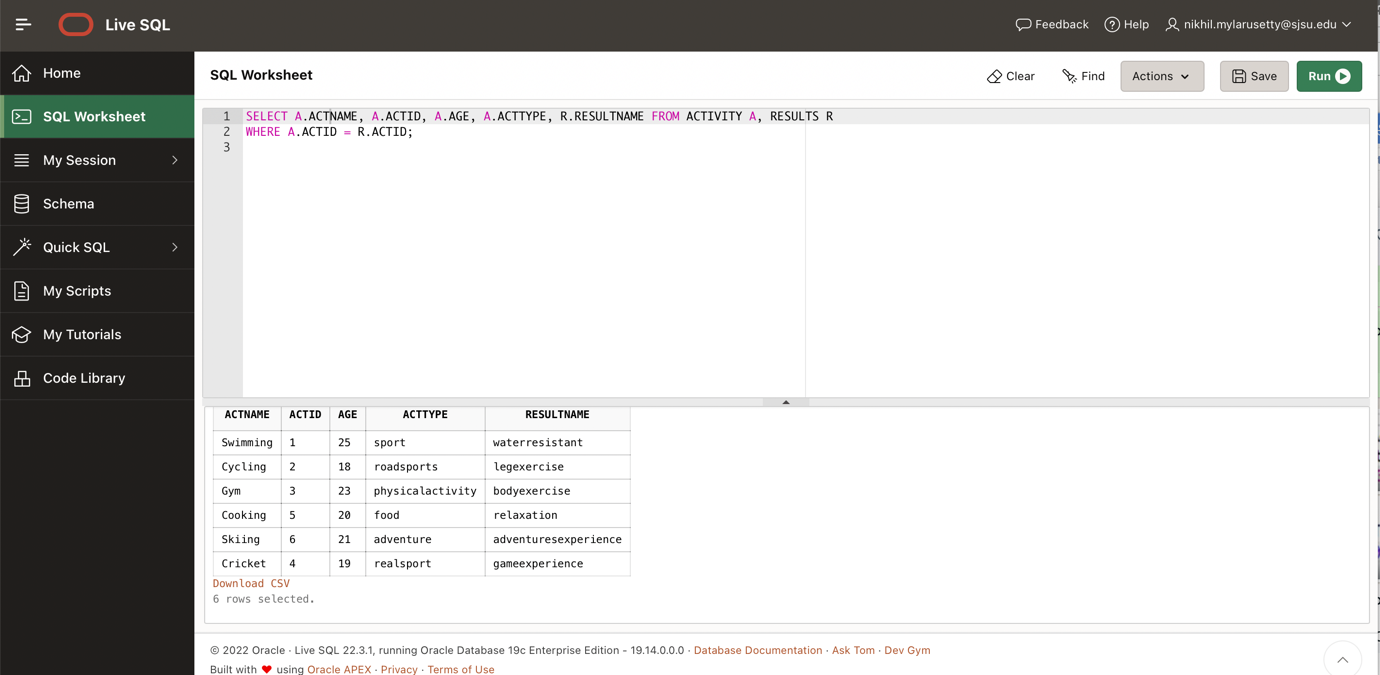
1. [Oracle Database Analytics in the Cloud]
2. Created account Successfully.



Graphical user interface, application

Description automatically generated





Graphical user interface, text, table

Description automatically generated with medium confidence

Table

Description automatically generated with low confidence

Query 1.

Graphical user interface, text, application

Description automatically generated

Query 2

Graphical user interface

Description automatically generated with low confidence

Query 3

Graphical user interface

Description automatically generated with low confidence

Query 4

Text

Description automatically generated with medium confidence

Query 5.

Graphical user interface, application

Description automatically generated

ROLLUP

Table

Description automatically generated with low confidenceGraphical user interface, text, application

Description automatically generated

CUBETable

Description automatically generatedGraphical user interface, text, application

Description automatically generatedGraphical user interface, application, Word

Description automatically generated

1. [40 points] [Data warehousing]

The Data warehouse of ABC Theaters has 3 dimensions: movie, movie\_goer, and time and two summary metrics: number\_of\_movie\_goers and revenue.

2[a] [15 points]

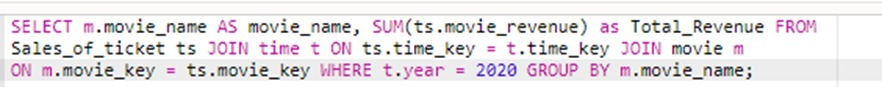
Draw a Star schema for this simple data warehouse, clearly showing the dimension and fact tables listing the columns that can help generating meaningful summaries such as the one in [b] below, from the data warehouse.

Graphical user interface

Description automatically generated

2[b] [5 points]

Assuming a relational model, write the SQL join query on the dimension and fact tables to output the total revenue collected through ticket sales by each movie in the pandemic year, 2020.



2[c] [5 points]

Suppose the transactional data was stored in an RDBMS table, ticket\_sales with 7 columns: theater, day, month, year, movie, movie\_goer, ticket\_price. Write the SQL query to generate the same output as [b] from this table.

A picture containing company name

Description automatically generated

2[d] [15 points]

Expand the above data warehouse to justify the possibility of a Snowflake schema. What additional dimensional tables will you add and what additional summary information can you generate from it? Draw the snowflake schema and list one or more join queries you can perform to output the additional summary information.

It is common for snowflake schemas to have linked dimension tables. Thus, normalizing snowflake's dimensions reduces redundancy. Our ability to conserve storage space is improved because of these tables. However, we would need to perform more joins to perform the query.

Diagram

Description automatically generated

3a)[20 points] [Data warehouse practice]

[a] [10 points]

Explore the use of columnar databases for implementing data warehouse solutions in the industry and how it helps with analytic functions. Giving specific examples from the industry where possible, describe 5 salient observations from your study.

3a)Data warehousing is something that Amazon Redshift is used for by several companies, including Airbnb, Lyft, and Coursera. In order to make the transition to a cloud-based paradigm, many industries are using cloud-based data warehouses, such as Amazon Redshift, Snowflake, and Google Big Query.

Despite being built on PostgreSQL, Redshift differs from it in three key ways:

* Due to its cloud-based nature, **scalability** is easy (upgrades or downgrades).AIRBNB uses Redshift or columnar-based storage for extensive and intricate analytics, which we studied primarily (by browsing through various articles).
* A **columnar database** aggregates huge data sets to enhance analytical query performance.
* **Sharding** is the process of distributing a table over several servers for parallel processing. Data warehousing on Amazon Redshift is used by several companies, including Airbnb, Lyft, and Coursera. Cloud computing is becoming the new industry

**Salient Observations –**

* We can study vast amounts of data more effectively using columnar databases, since analytical queries on huge datasets respond much more quickly.
* It is possible to achieve excellent efficiency and performance by combining columnar storage with parallelism and data compression techniques.
* There is a significant reduction in disk space IO when using columns. The majority of NoSQL databases store data in columns.
* In certain columnar databases, columns can be self-indexed for improved processing and analysis.
* Columnar stores have the primary advantage of allowing query operators to handle compressed data directly (which takes up more CPU time but reduces I/O time greatly).

3b] [10 points]

Study the various OLAP tools used in the industry such as IBM Cognos, Microsoft SSAS, Oracle OLAP, etc. Giving specific examples from the industry where possible, describe 5 salient observations from your study.

3b) IBM COGNOS Analytics helps companies make better decisions by providing numerous reports and dashboards that can be easily generated in Cognos. IBM Cognos software allows us to analyze and model our data.

Reports and dashboards can be quickly extracted, analyzed, and generated with just a few clicks, making it beneficial for users with less technical skills. Using Cognos SQL, framework managers can interact with all relational data sources that are compliant with SQL standards.

ADP, the top provider of accounting services globally, uses IBM COGNOS to extract business insights from a never-ending stream of data.

One of ADP's Senior Analytics spoke about his experience using IBM COGNOS and mentioned how the ADP teams frequently utilize it for analysis and reporting.

According to him, ADP's team receives a lot of positive feedback from its numerous customers around the world. The process of iterating during testing was easy. As part of its recent upgrade to IBM COGNOS version 11, ADP used automated testing to quickly evaluate the insights gained from complicated analytical reports, saving time.

**Microsoft SSAS**

Here, SSAS and OLAP databases were used to create an OLAP cube format for data analysis. The data can be arranged into multidimensional structures to make it easier to analyze.

**Salient Observations –**

* By using Cognos, we can plan or model a data warehouse, analyze the data, and create forecast reports and dashboards.
* Microsoft Analysis Services (SAAS) provides support for both MOLAP and ROLAP solutions. Within a single model, they allow ROLAP, MOLAP, and HOLAP to be applied.
* For high performance, SAAS supports tabular storage mode, which keeps compressed data in memory.
* There are a variety of advantages associated with Cognos, such as lower expenses, less maintenance, quicker results, and better decision-making.
* In order to reduce hardware or disk access times, IBM uses in-memory databases or saves data in memory.