**Midterm Part II: Practical**

**Question 7: Cloud Spanner**

**Objective:**

* To create an instance, database, add tables.
* Insert data and retrieve the data from console.

**Task 1: Instance creation**

**Step 1: Enable cloud spanner API**

* Login to the GCP console, select your project and type “Cloud Spanner API” in the search bar.

**A screenshot of a computer

Description automatically generated**

* Select “Cloud Spanner API” from the API section.
* In my case, my API is enabled.

A screenshot of a computer

Description automatically generated

* If it is not enabled for you, click on “enable”.

**Step 2: Create an Instance**

* Activate cloud shell from the top right corner of the screen.

**A screenshot of a computer

Description automatically generated**

* Type the below command to create the instance.
  + gcloud spanner instances create sonali-midterm-instance --config=regional-us-central1 --description="sonali instance" --nodes=2

**A screenshot of a computer

Description automatically generated**

* Cloud spanner instance is created.
* To set the default instance, run the below command.
  + gcloud config set spanner/instance sonali-midterm-instance

A black and white screen

Description automatically generated

* The default instance is set.

**Step 3: Create database and add tables**

* Execute the below command to create the database.
  + gcloud spanner databases create midterm-db

**A black screen with white text

Description automatically generated**

* The database “midterm-db” is now created.

**Step 4: Load Data**

* Execute the below two queries to create the tables.
  + gcloud spanner databases ddl update midterm-db --ddl='CREATE TABLE Student ( StudentId INT64 NOT NULL, FirstName STRING(1024), LastName STRING(1024), DepartmentID STRING(1024) ) PRIMARY KEY (StudentId)'
  + gcloud spanner databases ddl update midterm-db --ddl='CREATE TABLE Department ( DepartmentID INT64 NOT NULL, SubjectID INT64 NOT NULL, Faculty STRING(MAX), Score INT64, AssesmenType STRING(1024)) PRIMARY KEY (DepartmentID)'

A black screen with white text

Description automatically generated

* The table “Student” and “Department” are now created.
* Execute the below queries to insert data into the student table
  + gcloud spanner rows insert --instance=sonali-midterm-instance --database=midterm-db --table=Student --data=StudentId=1,FirstName='Lebron',LastName='James',DepartmentID='1',Semester='1'
  + gcloud spanner rows insert --instance=sonali-midterm-instance --database=midterm-db --table=Student --data=StudentId=2,FirstName=’Stephen’,LastName=‘Curry’,DepartmentID='2',Semester='2'
  + gcloud spanner rows insert --instance=sonali-midterm-instance --database=midterm-db --table=Student --data=StudentId=3,FirstName=‘Kevin’,LastName=‘Durant’,DepartmentID='3',Semester='3'
  + gcloud spanner rows insert --instance=sonali-midterm-instance --database=midterm-db --table=Student --data=StudentId=4,FirstName=‘Sonali’,LastName=‘Sabnam’,DepartmentID='1',Semester='1'
  + gcloud spanner rows insert --instance=sonali-midterm-instance --database=midterm-db --table=Student --data=StudentId=5,FirstName=‘Kate’,LastName=‘Durant’,DepartmentID='2',Semester='1'

A black background with white lines

Description automatically generated

* Now the student table has 5 rows of data.
* Execute the below commands to insert data into Department table
  + gcloud spanner rows insert --instance=sonali-midterm-instance --database=midterm-db --table=Department --data=DepartmentID=1,SubjectID=1,Faculty='Lakers',Score=100,AssesmenType='midterm'
  + gcloud spanner rows insert --instance=sonali-midterm-instance --database=midterm-db --table=Department --data=DepartmentID=3,SubjectID=2,Faculty='Lebron',Score=90,AssesmenType='midterm'
  + gcloud spanner rows insert --instance=sonali-midterm-instance --database=midterm-db --table=Department --data=DepartmentID=2,SubjectID=3,Faculty='Kati',Score=80,AssesmenType='midterm'
  + gcloud spanner rows insert --instance=sonali-midterm-instance --database=midterm-db --table=Department --data=DepartmentID=4,SubjectID=4,Faculty='Jhon',Score=90,AssesmenType='midterm'
  + gcloud spanner rows insert --instance=sonali-midterm-instance --database=midterm-db --table=Department --data=DepartmentID=5,SubjectID=1,Faculty='Tim',Score=100,AssesmenType='midterm'

A black and white screen

Description automatically generated

* Now the Department table has five rows of data.

**Step 5: Query Data**

* Go to the search and type “Spanner”.
* Click on the instance

A screenshot of a computer

Description automatically generated

* Click on the database
* Click on the Spanner Studio from the left pan.

A screenshot of a computer

Description automatically generated

* We can see the tables in the left side.
* Write the below query in the editor to select the rows from Department table.
  + Select \* from Department
* The below screenshot displays the data which we entered earlier.

A screenshot of a computer

Description automatically generated

* Write the below select query to select data from Student table.
  + Select \* from Student
* The below screenshot displays the data which we entered earlier.

A screenshot of a computer

Description automatically generated

**Step 6: Delete database and instance**

* Go to the cloud shell and set the instance using the below command
  + gcloud config set spanner/instance sonali-midterm-instance
* To delete the DB, run the below command and enter Y
  + gcloud spanner databases delete midterm-db

A black screen with white text

Description automatically generated

* Delete the instance using the below command.
  + gcloud spanner instances delete sonali-midterm-instance

A black and white screen with white text

Description automatically generated

* Now, the instance, database, tables and the data are deleted. We need to create them again if we need them.

**Question 8: BigQuery**

**Objective:**

* Create a dataset, and a table.
* Download data from canvas and load it into the table.
* Verify if the data was loaded successfully.

**Step 1: Download the data file and unzip it**

* Download the data from Canvas🡪5240🡪Modules🡪DataSets🡪babynames.zip
* Unzip the data and read the NationalReadme.

**Step 2: Create a dataset**

* In the search bar type big query and click on “BigQuery”.

A screenshot of a computer

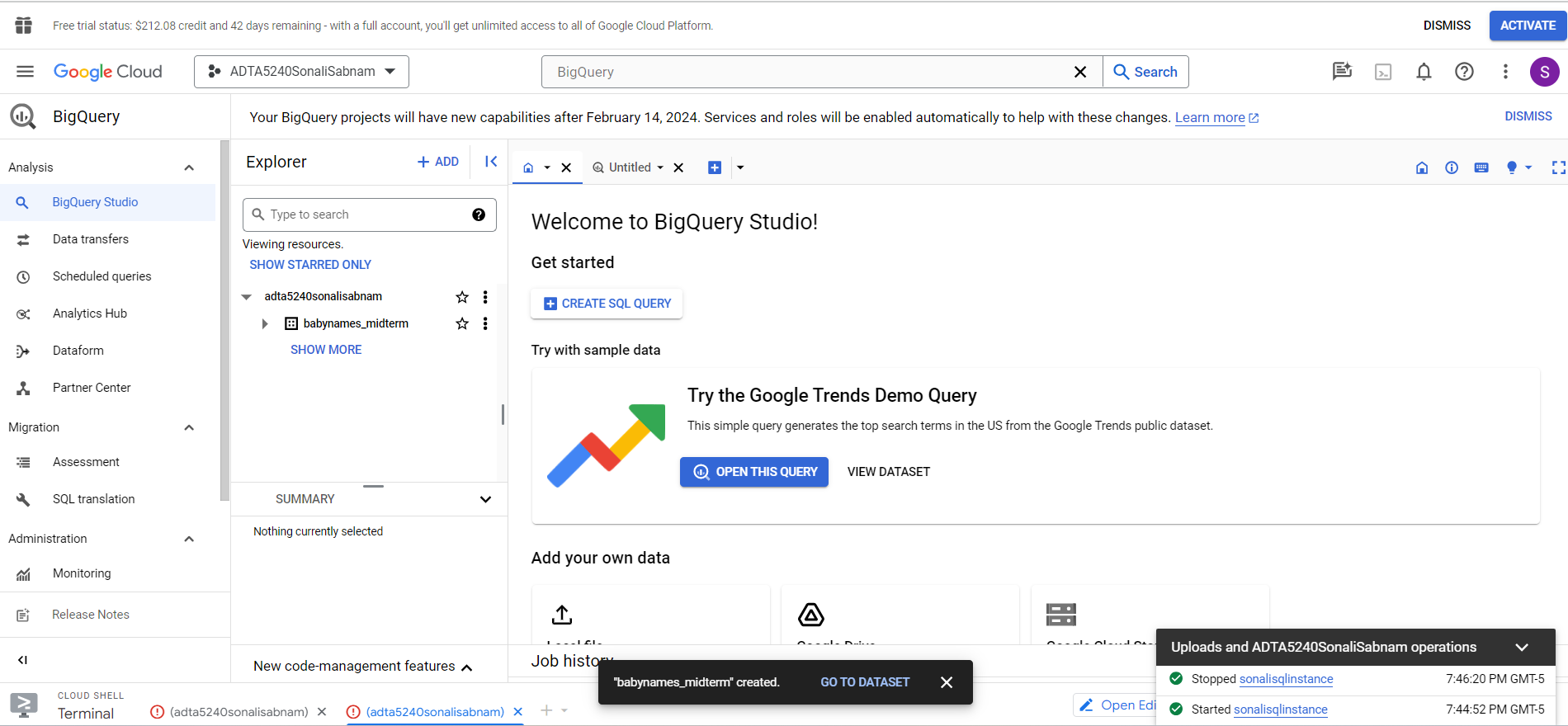
Description automatically generated

* Hover on the project and click on “Create dataset.
* The right pan opens a new screen.
* Fill in the details as per the below screenshot.

A screenshot of a computer

Description automatically generated

* Click on “Create Dataset”.
* Now, “babynames\_midterm” dataset is created.



**Step 3: Create table and upload data**

* Hover on the dataset and select “Create table”.
* Fillin the fields as per the below screenshot.
* In the accessibility option field enter “name: string, gender: string, count: integer”.

A screenshot of a computer

Description automatically generated

* Click on “Create Table”.
* The table is now created.
* Click on the table, the below screen gets loaded.

A screenshot of a computer

Description automatically generated

**Step 4: Verify data**

* Click on “Preview” tab to see the data.
* It displays the data successfully.
* The below screenshot shows the data uploaded using the file.

A screenshot of a computer

Description automatically generated

* With this we successfully validated the data uploaded.

**Step 5: Cleanup**

* To avoid unnecessary charges, we need to cleanup.
* Delete the table using the delete button.

A screenshot of a computer

Description automatically generated

* The table is now deleted.

**Question 9: MYSQL**

**Objective:**

* To create MYSQL instance, database, tables
* Load and retrieve data
* Cleanup

**Step 1: Verify if Cloud Admin API is enabled**

* In the search bar type “Cloud SQL Admin API” and verify if it is enabled
* It was already enabled for me.
* Please enable it if it is not the same for you.

A screenshot of a computer

Description automatically generated

**Step 2: Create SQL instance**

**Note**: I already have a Cloud MySQL instance and I am unable to create another due to free credit restrictions. Here are the steps to create cloud MySQL instance if you are creating for the first time.

* in the search bar type “SQL” and click on SQL
* Click on “Create Instance”.

**A screenshot of a computer

Description automatically generated**

* Select MySQL and fill in the required details like
  + instance ID
  + password.
  + “Enterprise” for Cloud SQL edition
  + “Sandbox” as the environment
  + “Single Zone”
* Now click on “CREATE INSTANCE”. MySQL instance should get created and running.

**Step 3: Connect to MYSQL Shell and connect to MYSQL instance.**

* Open the cloud shell and type the below command to connect to your instance
  + gcloud sql connect sonalisqlinstance
* Authorize and type the password to connect to MYSQL.

A computer screen with text

Description automatically generated

* Now, connection is successfully established with the MYSQL instance.

**Step 4: Create new database and add table and data**

* Run the below command to create the database “midterm”.
  + CREATE DATABASE midterm;
* To create the table “Student” in the database, execute the below command.
  + USE midterm;
  + CREATE TABLE Student ( StudentId INT NOT NULL, FirstName VARCHAR(255), LastName VARCHAR(255), DepartmentID int, Semester int, Isinternational char, Degree VARCHAR(255), Major VARCHAR(255), Concentration VARCHAR(255));

A black and white text

Description automatically generated

**Step 5: Add data using insert query**

* To insert data run the below queries
  + INSERT INTO Student (StudentId, FirstName, LastName,DepartmentID, Semester, Isinternational, Degree, Major,Concentration) values (1, "Tony", "Barone", "1", "1", "N", "Master of Science", "Data Science", "Deep Learning");
  + INSERT INTO Student (StudentId, FirstName, LastName,DepartmentID, Semester, Isinternational, Degree, Major,Concentration) values (2, "Tamy", "Salt", "1", "1", "N", "Master of Science", "Advanced Data Analytics", "Sports Data Analytics");
  + INSERT INTO Student (StudentId, FirstName, LastName,DepartmentID, Semester, Isinternational, Degree, Major,Concentration) values (3, "Raj", "Kumar", "1", "1", "N", "Master of Science", "Business Analytics", "Project Management");
  + INSERT INTO Student (StudentId, FirstName, LastName,DepartmentID, Semester, Isinternational, Degree, Major,Concentration) values (4, "Irfan", "Khan", "1", "1", "N", "Master of Science", "Advanced Data Analytics", "Healthcare Analytics");
  + INSERT INTO Student (StudentId, FirstName, LastName,DepartmentID, Semester, Isinternational, Degree, Major,Concentration) values (5, "Irfan", "Khan", "1", "1", "N", "Master of Science", "Advanced Data Analytics", "Healthcare Analytics");
  + INSERT INTO Student (StudentId, FirstName, LastName,DepartmentID, Semester, Isinternational, Degree, Major,Concentration) values (6, "Pooja", "Kumari", "1", "1", "N", "Master of Science", "Advanced Data Analytics", "Applied Artificial Intelligence");

A black and white text

Description automatically generated

* The query gets executed successfully and data is inserted successfully.

**Step 6: Retrieve the data**

* Execute select query to retrieve the data.
  + select \* from Student;

A black and white screen

Description automatically generated

**Step 7: Cleanup**

* Close the terminal.
* From all instances, select the instance and click on STOP.

**A screenshot of a computer

Description automatically generated**

* Now, the instance is stopped.

**Question 10: Upload a file to a folder in Cloud storage using cloud console**

**Objective:**

* To create a folder in Cloud storage and upload a file using cloud console

**Step 1: Create a new storage bucket and load file**

* To create a new storage bucket, type bucket in the search bar
* Click on bucket and it displays the previously created buckets as shown below.

**A screenshot of a computer

Description automatically generated**

* Click on “create” to create a new bucket if you do not have one.

A screenshot of a computer

Description automatically generated

* Fill in the details as per the screenshot below and click on create.
* Now the new bucket is created.

A screenshot of a computer

Description automatically generated

* Click on “create folder” to add a new folder in the bucket.
* In the new pop-up, provide a name and click OK.
* Now, the folder is created in the bucket.

A screenshot of a computer

Description automatically generated

* Now click on the newly created folder and select “Upload Files”.
* Select the “Iris.csv” file from your drive (it was downloaded from canvas previously and saved in my drive) and click upload.
* Now the file is successfully uploaded into the folder inside the bucket.

A screenshot of a computer

Description automatically generated

* Now we have the bucket, folder and the file available for us to be used in HDFS.

**Note**: the above steps can also be performed by using cloud console but for the sake of simplicity I have done it through UI. I will be performing the HDFS steps in cloud console.

**Step 2: Start the cluster**

* Type “cluster” in the search bar and click on “Cluster Dataproc”.
* It displays the previously created cluster. You can create one if you do not have it.

**A screenshot of a computer

Description automatically generated**

* Click on Start to start the cluster.

**A screenshot of a computer

Description automatically generated**

**Step 3: Connect to the terminal**

* Click on the cluster, scroll down and click on SSH.
* A new window opens up, click on “Authorize”.
* Now the terminal is open.

A screenshot of a computer

Description automatically generated

* Type “clear” to clear the terminal.
* Type the below commands to know the current user, present working directory and list the contents of the directory with permissions.
  + Whoami
  + Pwd
  + hdfs dfs -ls /

A computer screen shot

Description automatically generated

* Run the below command to get inside the “user” folder.
  + hdfs dfs -ls /user

**Step 4: Create a folder and add file in HDFS and in the VM**

* Write the below command to create a new folder inside “user” folder.
  + hdfs dfs -mkdir /user/sonalisabnam2024midterm
  + hdfs dfs -ls /user

A screen shot of a computer screen

Description automatically generated

* Now the folder is created.
* Execute the below command to create a sub-folder.
  + hdfs dfs -mkdir /user/sonalisabnam2024midterm/data
  + hdfs dfs -ls /user/sonalisabnam2024midterm

A screenshot of a computer

Description automatically generated

* Go to the GCP console and type “Compute Engine”.
* Click on SSH, to connect to the terminal and Authorize the popup.

A screenshot of a computer

Description automatically generated

* Now we are connected to the Linux Virtual Machine.
* Type the below command to create “MIDTERMDATA” folder
  + mkdir MIDTERMDATA
  + ls -l

A screenshot of a computer program

Description automatically generated

* Go inside the folder and list the contents using the below commands
  + mkdir MIDTERMDATA
  + cd MIDTERMDATA

A screenshot of a computer program

Description automatically generated

* We can see that MIDTERMDATA folder is empty.
* To copy the data from bucket to this folder, execute the below command
  + gsutil cp gs://adta5240sonalibucketmidterm/midtermdata/Iris.csv Iris.csv

A screen shot of a computer program

Description automatically generated

* Now the Iris.csv file is copied successfully to the MIDTERMDATA folder.

**Step 5: Copy file from master node to HDFS**

* Type the below command in the Linux terminal to copy the Iris.csv file and list the contents.
  + hdfs dfs -put Iris.csv /user/sonalisabnam2024midterm/midtermdata
  + hdfs dfs -ls /user/sonalisabnam2024midterm/midtermdata

A screenshot of a computer program

Description automatically generated

* We can see that the file copied successfully.
* With this we have successfully copied the “Iris.csv” file from the master node to HDFS.

**Step 6: Cleanup**

* To avoid unnecessary charges, we need to close stop the clusters
* In the GCP console, go to Cluster and click on “STOP”.

A screenshot of a computer

Description automatically generated

* The cluster is now stopped.

**Conclusion:** In question 10, we created a bucket and a folder inside it. We added “Iris.csv” file into the bucket. Then we connected to the HDFS using terminal and created a folder and sub-folder. Then we connected to the master node using Linux terminal, created a folder, copied “Iris.csv” file from the bucket to the master node folder. Finally, we moved the file “Iris.csv” from the master node to the HDFS. After completing, we stopped the cluster.