CASE-STUDY (NUMBER - 7) – SOLUTION SUBMISSION

ON

AZURE ANALYTICS

BY

NAME –SUPRIYA BHARATHA ROLL NUMBER DXCAB1211

BATCH:DXC-262-ANALYTICS-B12-AZURE COMPANY – DXC TECHNOLOGY

TRAINING UNDER – MANIPAL PRO LEARN TRAINER NAME – MR. AJAY KUMAR DATE OF SUBMISSION – 7th JUNE 2022 TOTAL NUMBER OF QUESTIONS: 10 EMPLOYEE DOMAIN - AZURE ANALYTICS

* THE QUESTIONS:

1. Explain what are various components of SPARK with block diagram?

explain functionality of every components?

2. Explain Spark core in details & how RDD is related to Spark core - explain with Spark program ?

3. Explain various Mlib algorithms Spark is supporting ?

4. Explain benifits Spark SQL & how relational data will be inserted into SPARK ?

5. Explain Spark streaming in detail ?

6. Explain SPARK architecure? what is Master - Slave architecure ?

7. Explain various cluster managers in SPARK?

8. Explain with sceenshots & steps how to create Cosmos DB ?

9. Explain with sceenshots & step how to insert data into Cosmos DB?

10. Explain with sceenshots & step how to create Azure SQL Db & also explain how to

insert data into Azure SQL D?

1. Explain what are various components of SPARK with block diagram? explain functionality of every components?

SOLUTION:

**GraphX**

**MLlib**

**Spark Streaming**

**Spark SQL**

**Spark Core**

**Apache spark**

* Spark core: Spark core is the base engine for large-scale parallel and distributed data processing.
* Spark SQL: Spark SQL framework components is used for structured and semi-structured data processing.
* Spark streaming: Spark streaming is a lightweight API that allows developers to perform batch processing and real time streaming of data with ease.
* MLlib: MLlib is a low- level machine learning library that is simple to use, is scalable, and compatible with various programming languages.
* GraphX: GraphX is sparks own graph computation engine and data store.

1. Explain Spark core in details & how RDD is related to Spark core - explain with Spark program ?

SOLUTION:

Spark core: Spark core is the base engine for large-scale parallel and distributed data processing.

* It is responsible for:
* Memory management.
* Fault recovery.
* Scheduling, distributing and monitoring jobs on a cluster.
* Interacting with storage systems.
* Resilient Distributed Dataset :

Spark Core is embedded with RDDs (Resilient Distributed Datasets), an immutable fault-tolerant, distributed collection of objects that can be operated on in parallel.

**RDD**

**Action**

**Transformation**

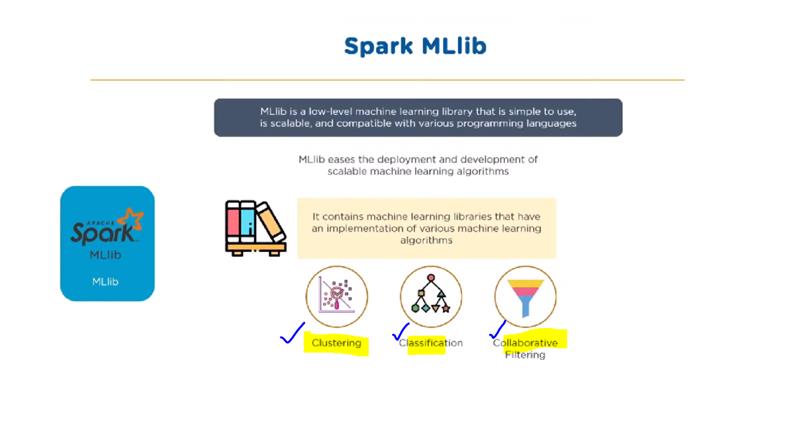
These are operations (such as reduce, first, count) that return a value after running a computation on an RDD

These are operations (such as map, filter, join, union) that are performed on an RDD that yields a new RDD containing the result.

1. Explain various Mlib algorithms Spark is supporting ?

SOLUTION:

* MLlib is a low-level machine learning library that is simple to use, is scalable, and compatible with various programming languages.
* MLlib eases the deployment and development of scalable machine learning algorithms.
* It contains machine learning libraries that have an implementation of various machine learning algorithms.



1. Explain benifits Spark SQL & how relational data will be inserted into SPARK ?

SOLUTION:

* Spark SQL: Spark SQL framework components is used for structured and semi-structured data processing.

Spark SQL and HQL

Data Frame DSL

Data Frame API

Data Source Api

JDBC

JSON

CSV

1. Explain Spark streaming in detail ?

SOLUTION:

* Spark Streaming is a lightweight API that allows developers to perform batch processing and real-time streaming of data with ease.
* Provides secure, reliable, and fast processing of live data streams.

Batches of processed data

Batches of input data

Input data stream

Spark Engine

Spark streaming

1. Explain SPARK architecure? what is Master - Slave architecure ?

SOLUTION:

Worker node

Cache

MASTER NODE

Task

Task

Driver Program

CLUSTER MANAGER

Worker node

Spark Context

A job is split into multiple tasks that are distributed over the worker node .

When an RDD is created in Spark context, it can be distributed across various nodes.

Worker nodes are slaves that run different tasks

Cache

Task

Task

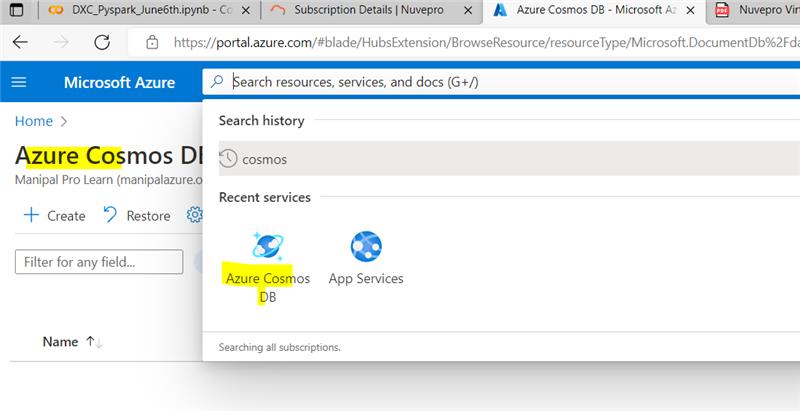
1. Explain various cluster managers in SPARK?

SOLUTION:

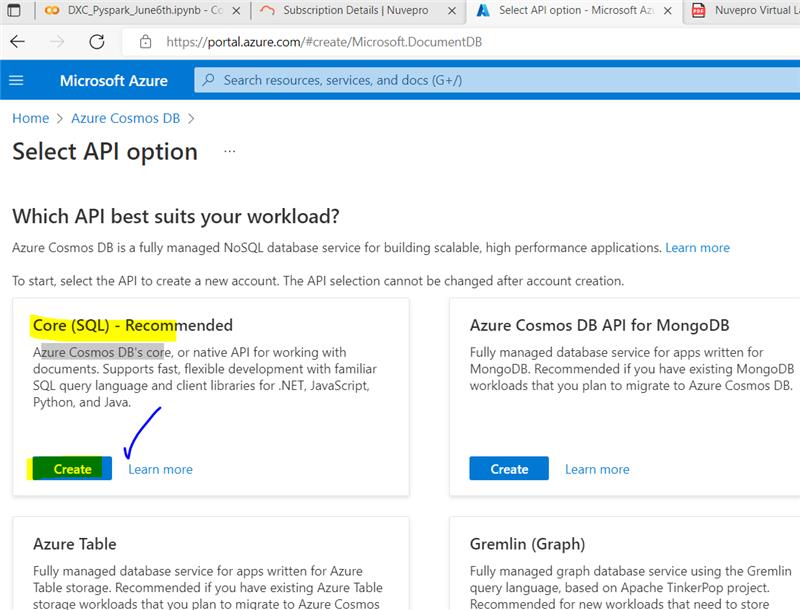


1. Apache spark: By default, applications submitted to the standalone mode cluster will run in FIFO order, and each application will try to use all available nodes.
2. Mesos: Apache Mesos is an open-source project to manage computer clusters, and can also run Hadoop applications.
3. Hadoop Yarn: Apache YARN is the cluster resource manager of Hadoop 2. Spark can be run on YARN.
4. Kubernetes: Kubernetes is an opensource system for automating deployment, scaling, and management of containerized applications.
5. Explain with sceenshots & steps how to create Cosmos DB ?

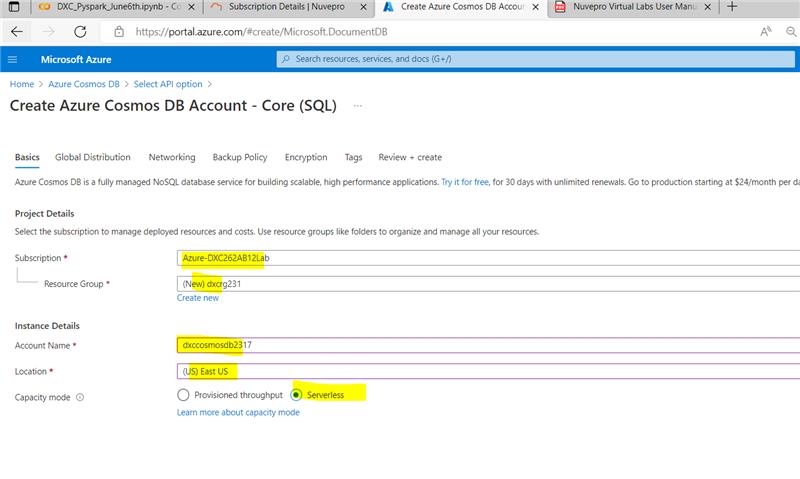
SOLUTION:

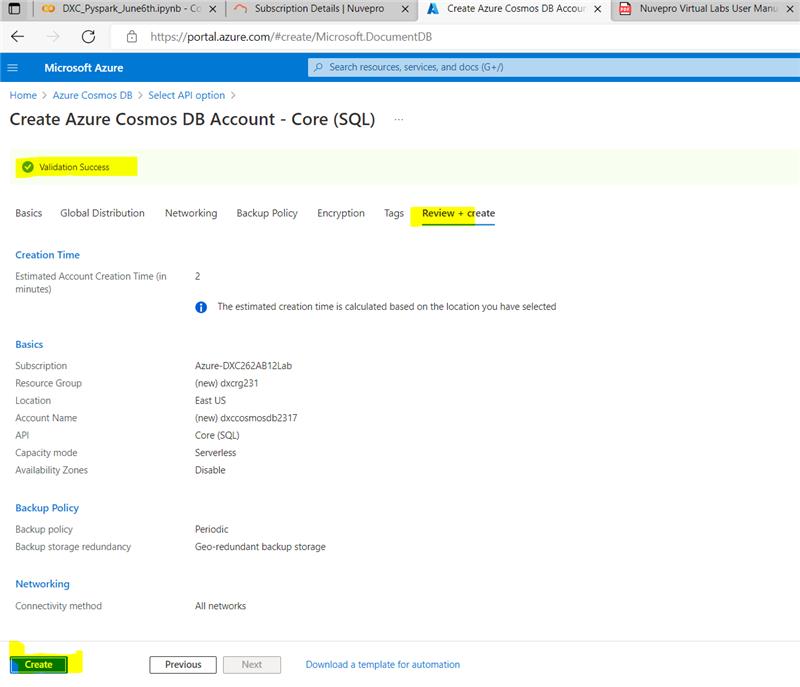
 **STEP 1:** From the Azure portal menu or the Home page, select Create a resource. On the New page, search for and select Azure Cosmos DB.

**STEP 2:** Select API option as Core(SQL) to create.

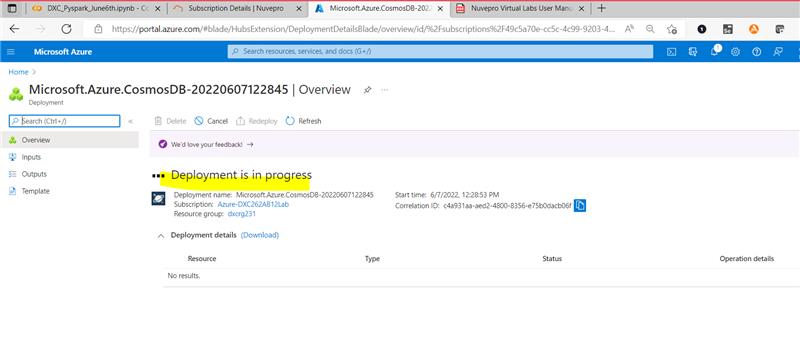
****

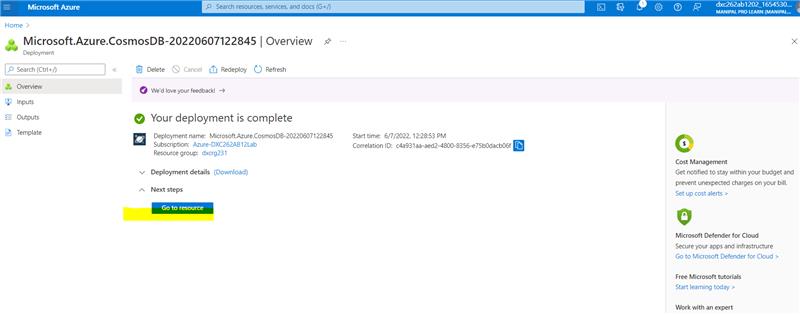
**STEP 3:** Give the resource group name and Account name.



**STEP 4:** Now we can continue to create cosmosdb account.

**STEP 5:** Deployment is carried out.

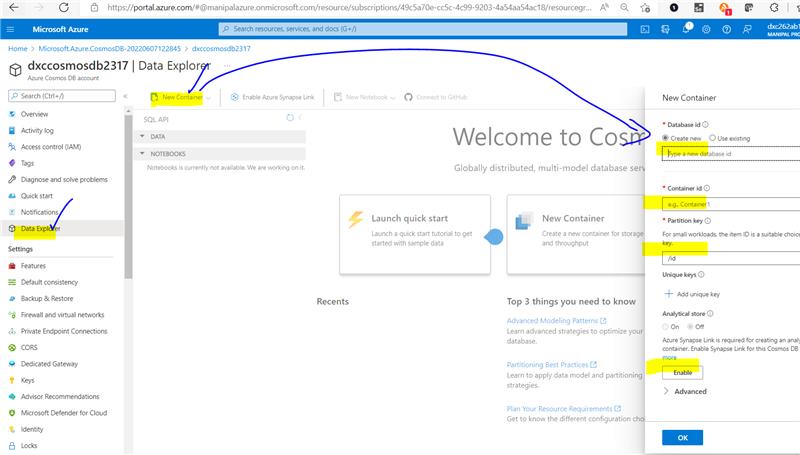
****

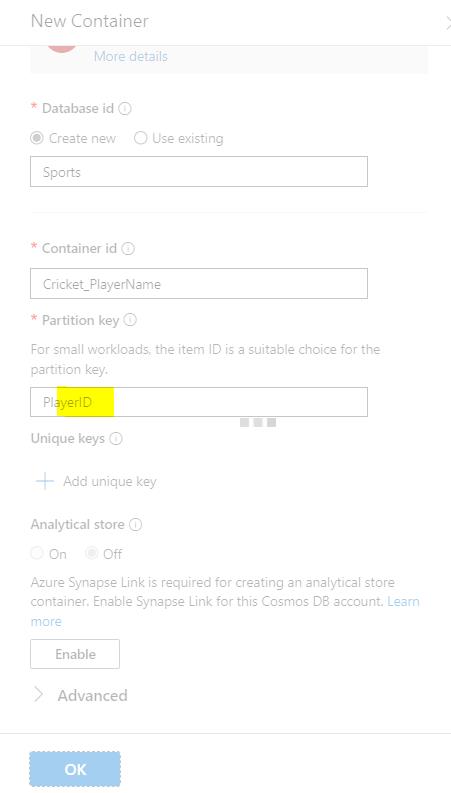


1. Explain with sceenshots & step how to insert data into Cosmos DB?

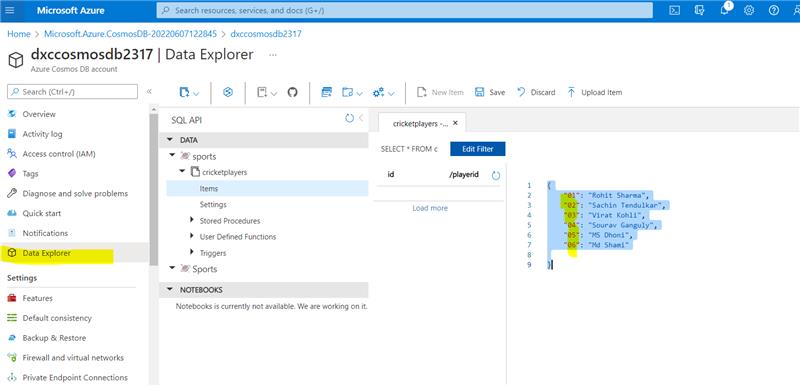


SOLUTION:

**STEP 1:** After creating Cosmos DB.Using data explorer new containers of different content or information can be added.

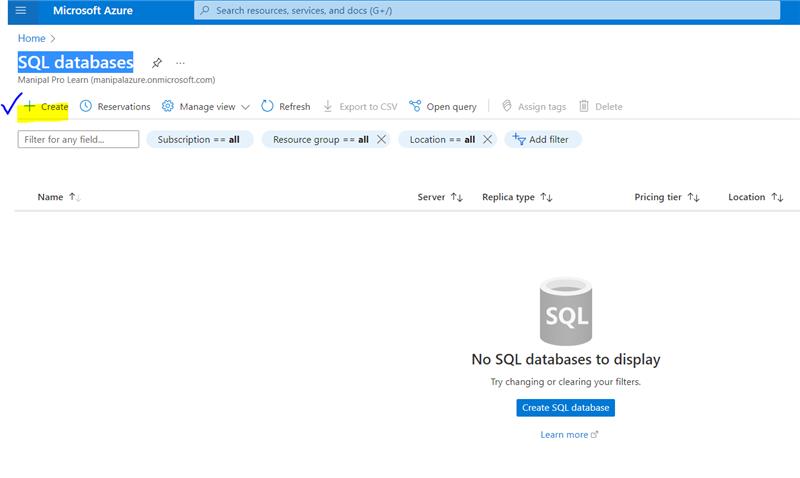
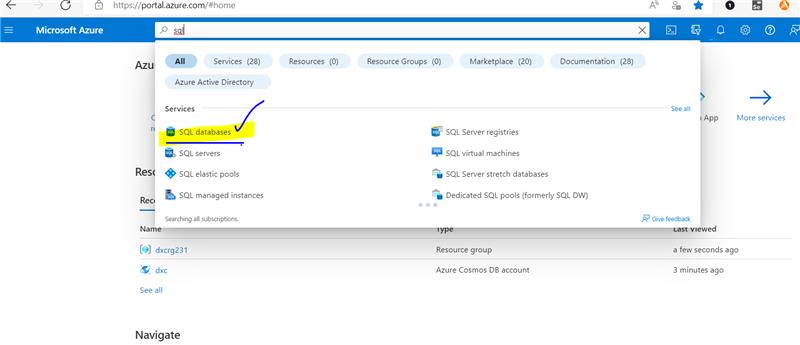


**STEP 2:** Here for inserting data database id, container id, partition key is given .

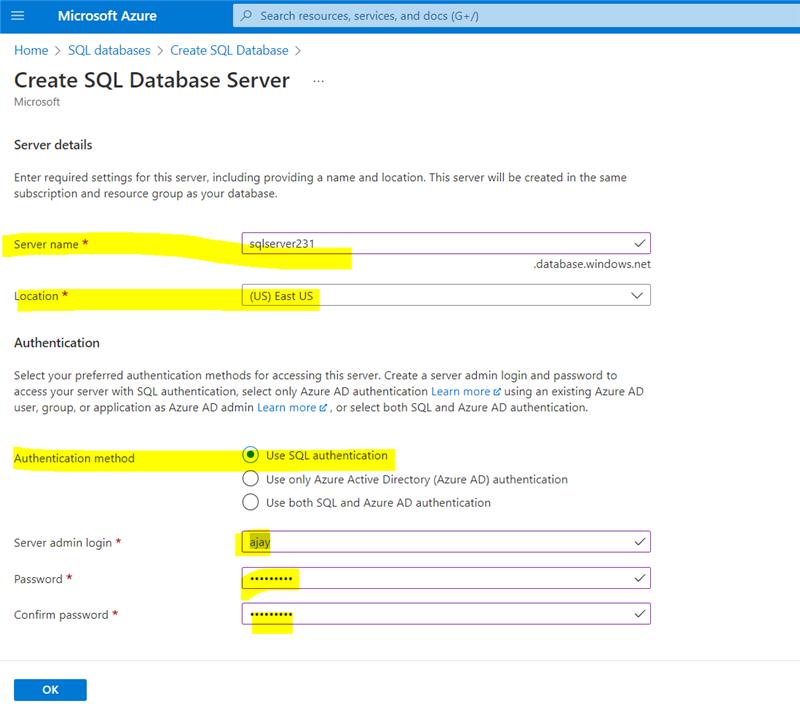
**STEP 3:** After creating data , we have to give the information as shown and save it.

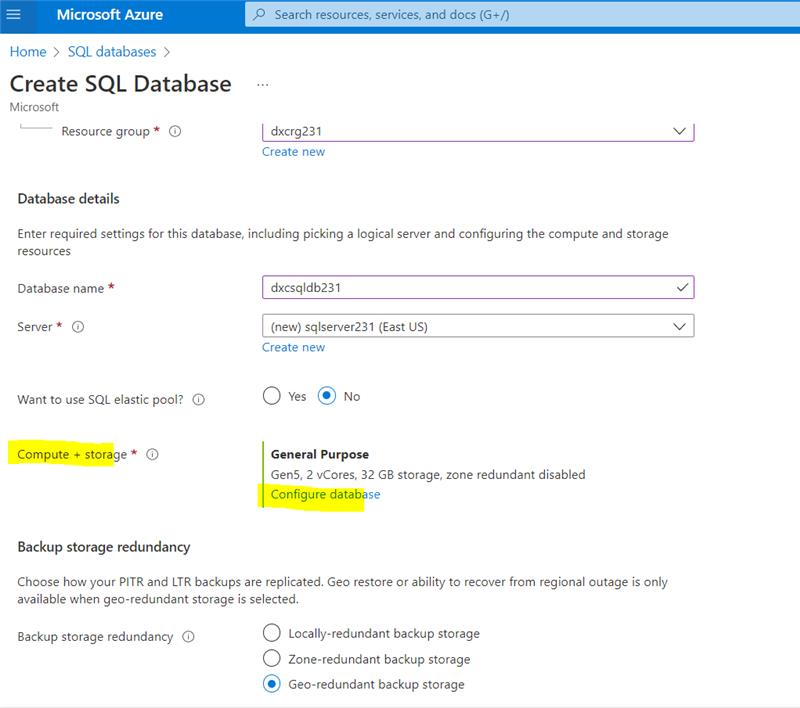
1. Explain with sceenshots & step how to create Azure SQL Db & also explain how to insert data into Azure SQL D?

SOLUTION:

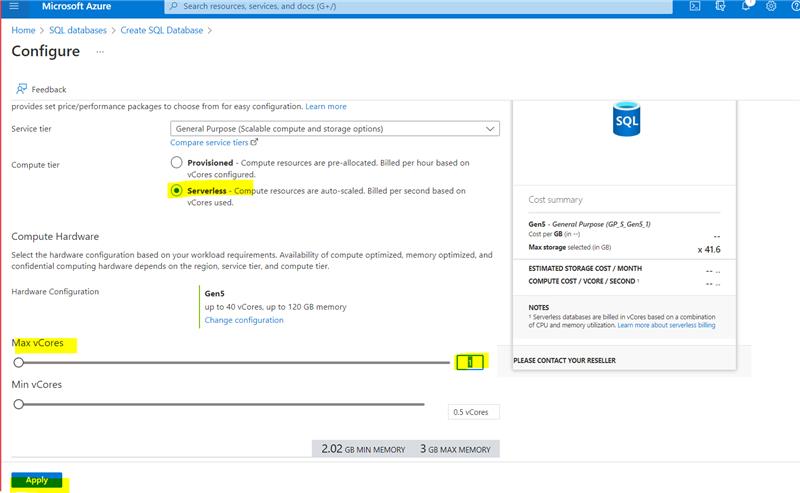
**STEP 1:** From the Azure portal menu or the Home page, select Create a resource. On the New page, search for and select SQL Databases.

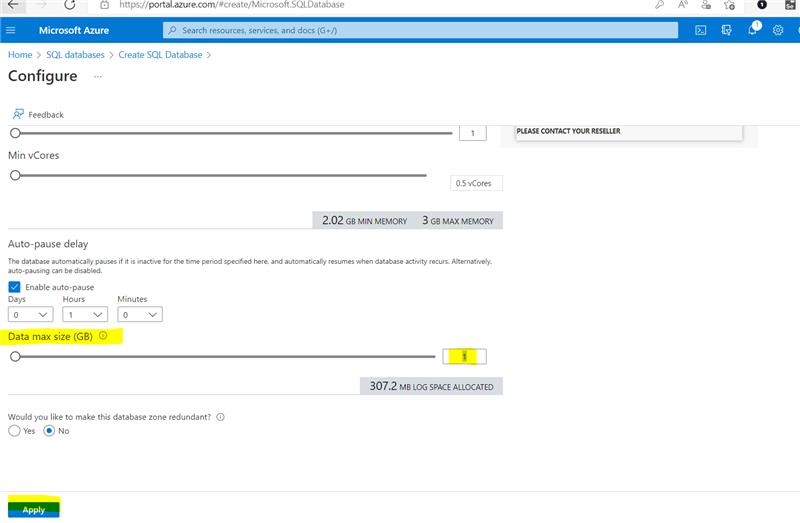
**STEP 2:** Creating SQL database database id, container id, partition key is given and , for server we have to create SQL Database server by giving the information shown below.



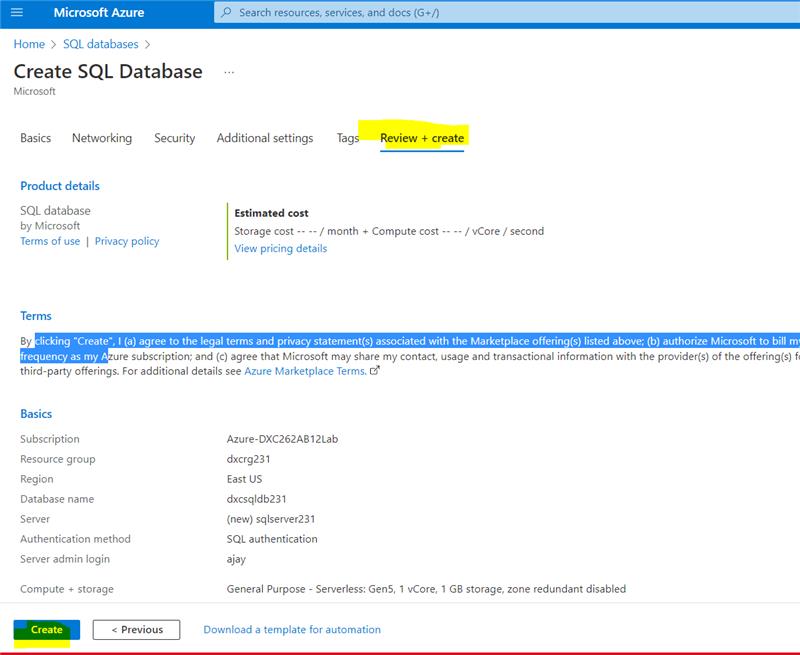


**STEP 3:** For comput + storage the MaxVCores and Data max size should be 1.

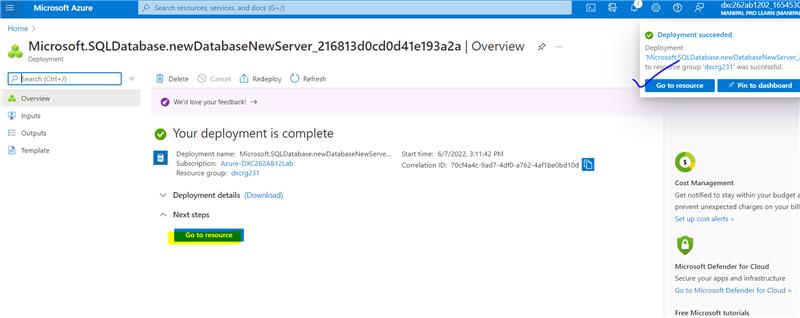


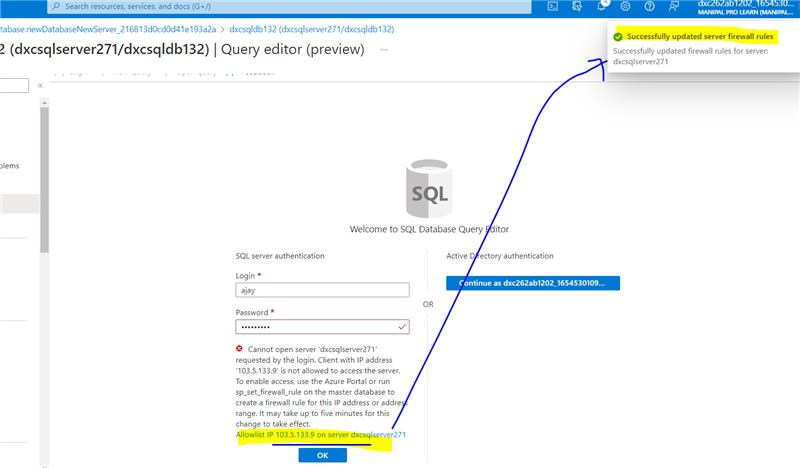


**STEP 4:** Now the SQL database will be created.

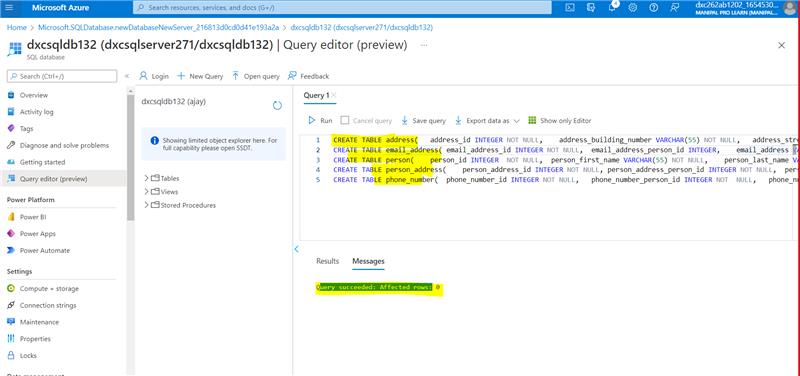


**STEP 5:** For creating the database deployment takes place and for sql the login password should be given as in previous but we need to allow it due to firewall.

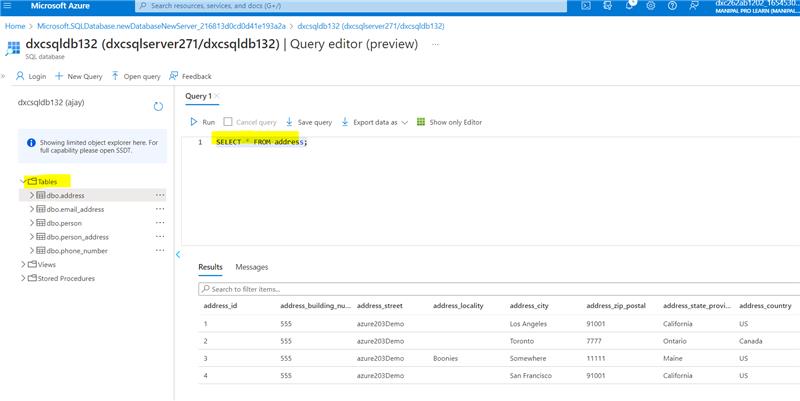




**STEP 6:** Now we need to create table and insert and can view the table by select.







RESULTS:

I HAVE SUCCESSFULLY ANSWERED ALL THE QUESTIONS AS PER ASSIGNMENT REQUIREMENT.

CONCLUSIONS:

All the questions have been solved successfully with all the concepts that have been covered in the training session. It’s really a great experience of learning while solving the cases. This assignment gave me immense confidence regarding my ability to upskill in new technologies.