ASSIGNMENT------SOLUTION SUBMISSION
ON
AZURE ANALYTICS
BY

NAME : SAI KIRAN ANCHE
ROLL NO: DXC262AB12021
BATCH:DXC-262-ANALYTICS-B12-AZURE
COMPANY – DXC TECHNOLOGY
TRAINING UNDER : MANIPAL PRO LEARN
TRAINER NAME – MR. AJAY KUMAR
DATE OF SUBMISSION : 07-06-2022
NO OF QUESTIONS :10
EMPLOYEE DOMAIN - AZURE ANALYTICS

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?

INTRODUCTION

This Assignment is given by manipal pro learn team on the basis of the training done in the forenoon session of this morning. The main objective behind this assignment is to master the theory and enhance knowledge over creating cosmos database and also sql database...

There are 10 questions and they are of easy to moderately difficult level. All the questions have been focused on what the trainer taught in the earlier sessions.Some questions have been answered partially due to unavailability of access.

This assignment gave me immense confidence in mastering the domain that has been assigned to me.

Solutions

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

A:

The fundamental components of spark include:

SPARK SQL	*SPARK STREAMING REAL-TIME*	*MILB MACHINE LEARNING*	*GRAPH X GRAPH PROCESSING*

SPARK CORE

SPARK CORE:

It is the heart if the spark frame work and it looks after the core functionality. It holds various components required for performing various actions.

SPARK SQL:

The Spark SQL is build on the spark core and it also provides support to the structured data

SPARK STREAMING:

Spark Streaming is a Spark component that supports scalable and fault-tolerant processing of streaming data.

MILB MACHINE LEARNING:

It is a Machine Learning Library that has various machine learning algorithms.

GRAPH X GRAPH PROCESSING:

It is a library that is used to manipulate the graphs and perform graph- parallel operations.

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

A:

Spark Core:

- The spark core is the heart of the spark frame work. It is the base engine for the large-scale parallel and distributed data processing.
- It is responsible for the memory management.
- It also helps in the fault recovery.
- It interacts with storage systems.
- It also performs scheduling , distributing and also it monitors jobs on a cluster.

3. Explain various Mlib algorithms Spark is supporting ?

A.

MLib is a machine learning library of low level which is simple to use . this library helps us in the development and deployment of various machine learning algorithms.

Some of the algorithms include :

- Clustering
- Classification.
- Collaborative Filtering.

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

A:

Some of the benefits of the spark sql are:

- Integrated
- Unified data access
- Highly compatible.
- Standard connectivity.
- Scalability.

- Performance Optimization.

The relational data can be inserted into spark by creating a dataframes and inserting the data into the dataframe.

5. Explain Spark streaming in detail ?

A:

The spark streaming is an API of light weight that allows the developers to do real-time streaming of data and batch processing very easily.

The spark streaming also provides fast , secure and reliable processing of data streams.

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

A:

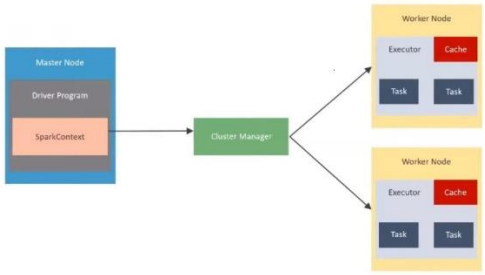
In general the spark architecture has three main components that include:

- Master Node
- Cluster manager
- Worker Node.

The master node has two components that are:

- SPARK context
- Driver program.

Here all the tasks are given by the master node to the cluster manager to run them and the cluster manager will assign the work to the worker nodes and if the worker node fails to do the task the again the data is redirected to the cluster manager and a new worker node is assigned again until the program complete.



7. Explain various cluster managers in SPARK?

A:

There are three Spark cluster manager, Standalone cluster manager, Hadoop YARN and Apache Mesos. Apache Spark supports these three type of cluster manager.

Standalone Cluster Manager

It is a part of spark distribution and available as a simple cluster manager to us. Standalone cluster manager is resilient in nature, it can handle work failures. It has capabilities to manage resources according to the requirement of applications.

2. Hadoop Yarn

This cluster manager works as a distributed computing framework. It also maintains job scheduling as well as resource management. In this cluster, masters and slaves are highly available for us. We are also available with executors and pluggable scheduler.

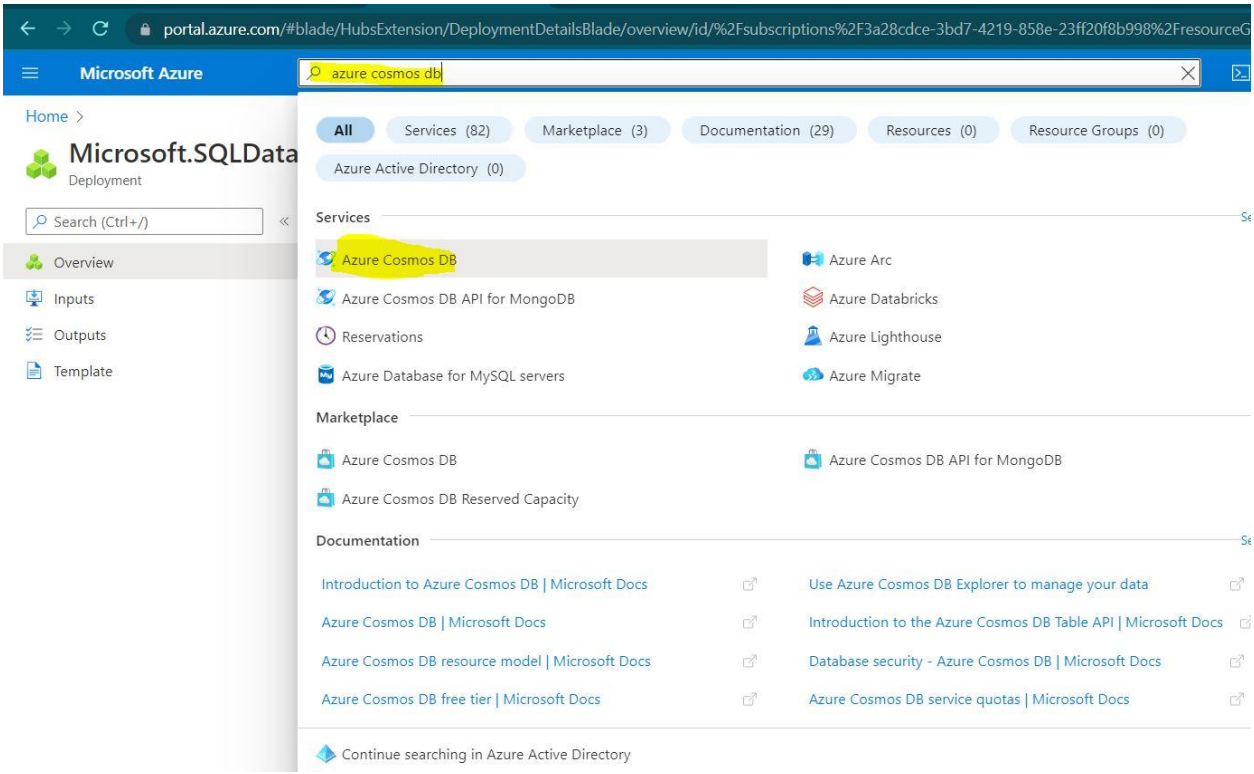
3. Apache Mesos

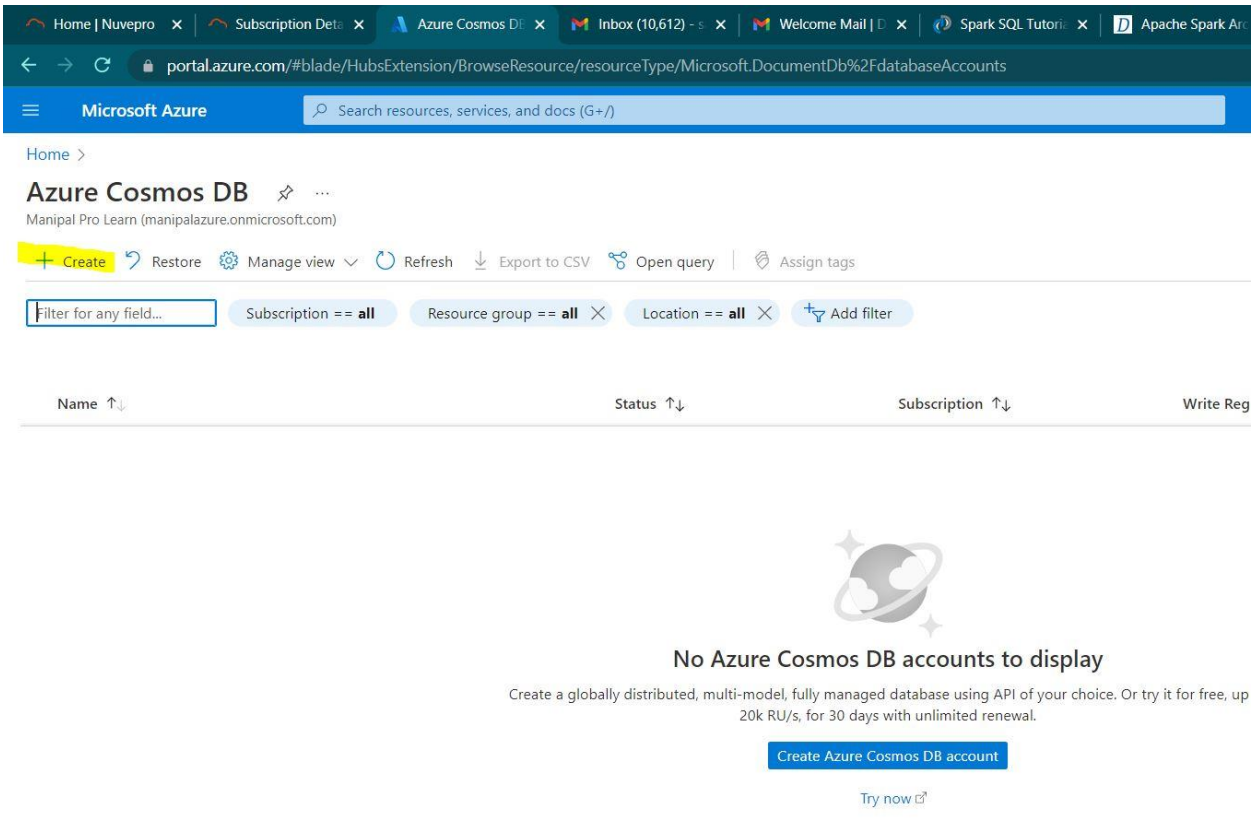
It is a distributed cluster manager. As like yarn, it is also highly available for **maste**r and **slaves**. It can also manage resource per application. We can run spark jobs, Hadoop MapReduce or any other service applications easily.

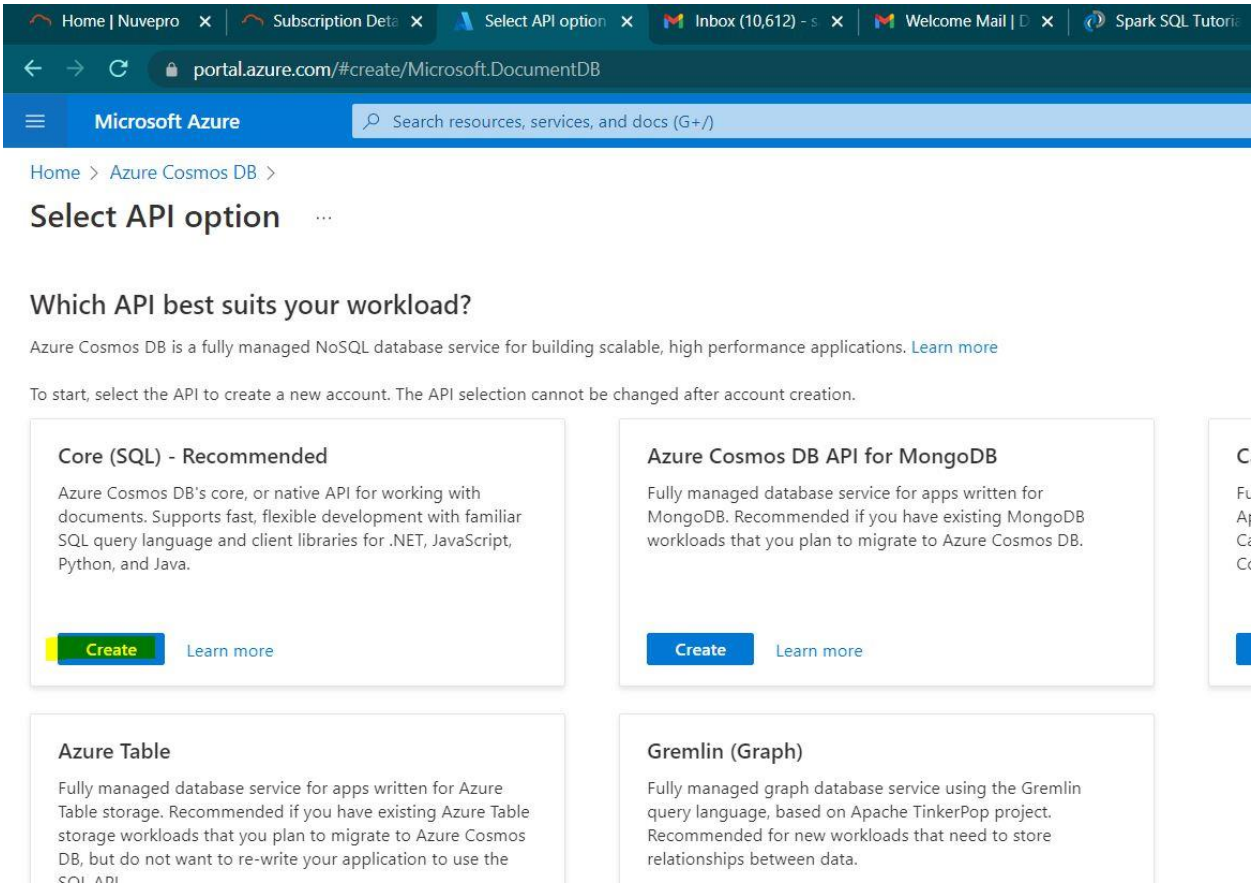
8. Explain with sceenshots & steps how to create Cosmos DB ?
9. Explain with sceenshots & step how to insert data into Cosmos DB?

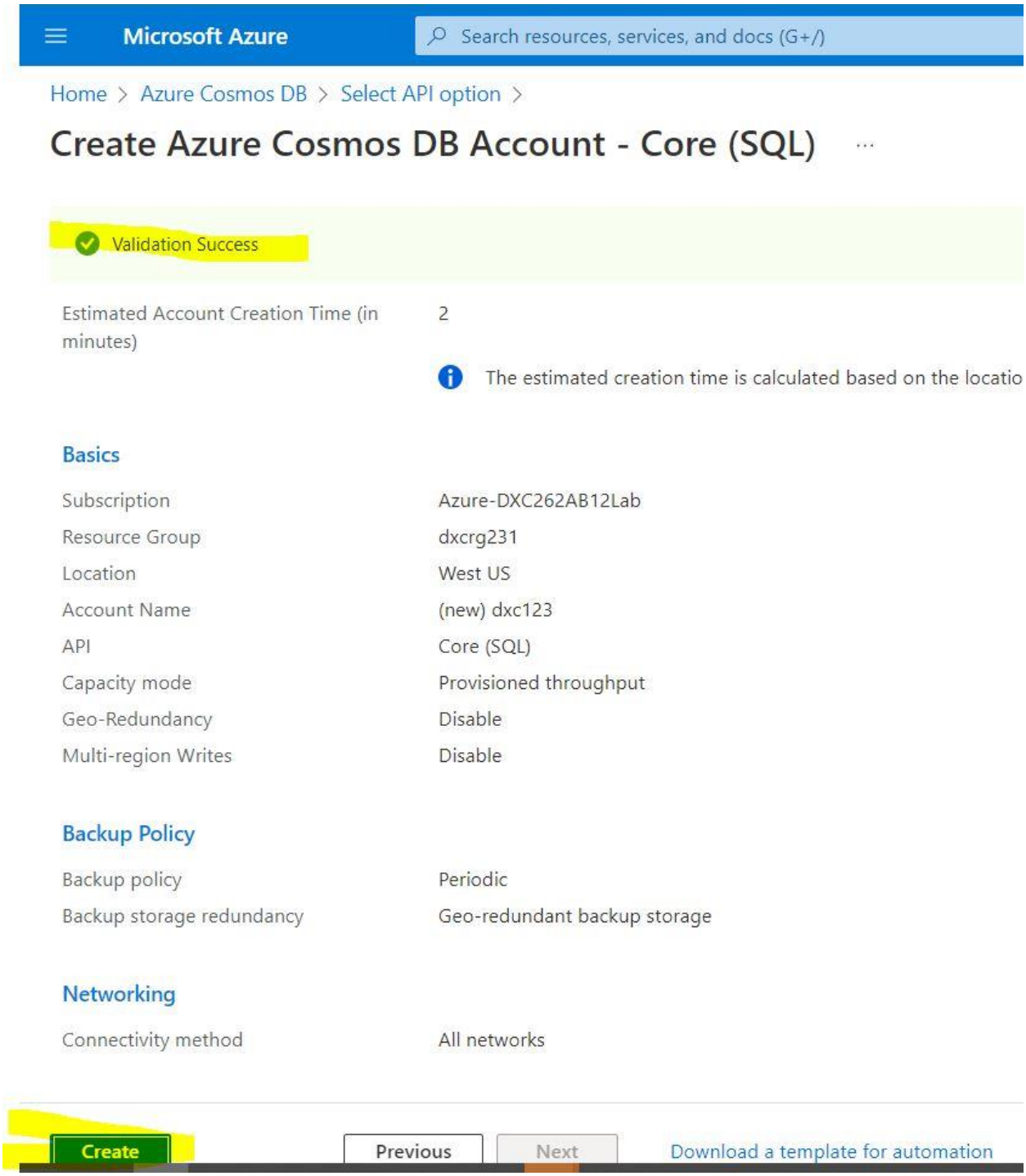
A:

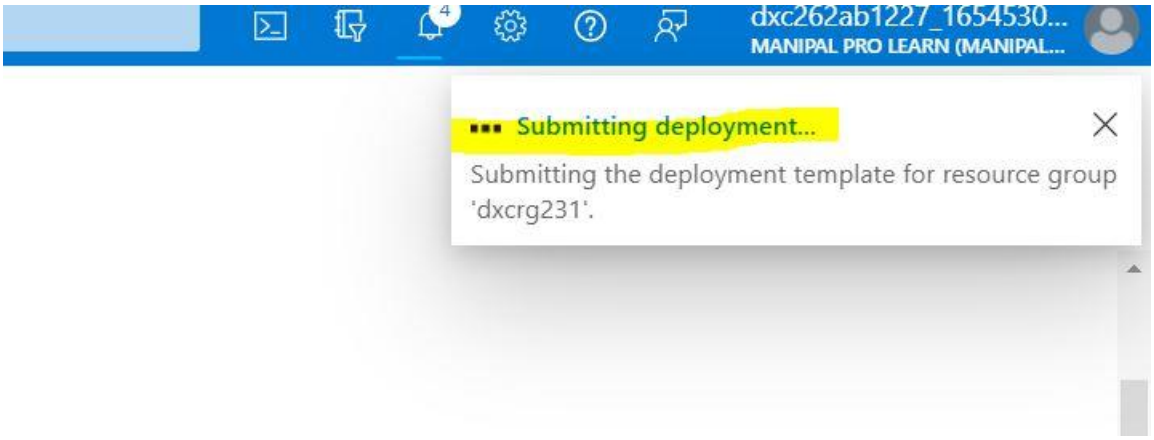
Answer for 8 and 9

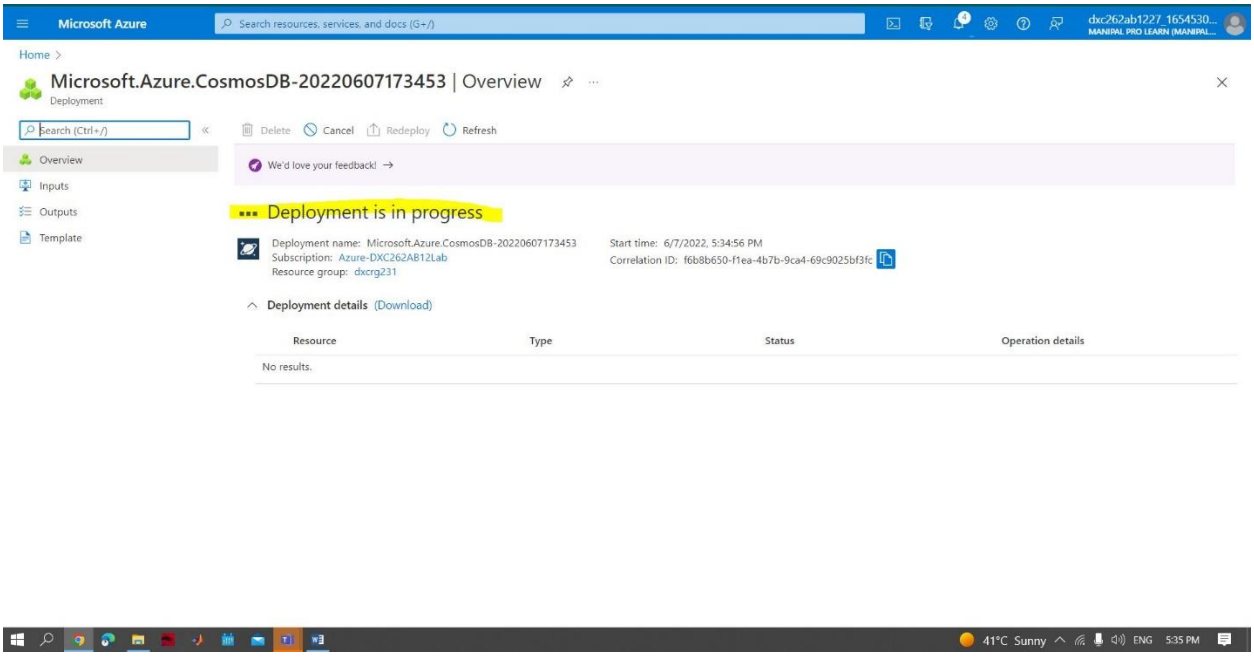

portal.azure.com/#blade/HubsExtension/DeploymentDetailsBlade/overview/id/%2Fsubscriptions%2F3a28cdce-3bd7-4219-858e-23ff20f8b998%2FresourceG
Microsoft Azure
azure cosmos db
Home
Microsoft.SQLData
Deployment
Search (Ctrl+/)
Overview
Inputs
Outputs
Template
All
Services (82)
Marketplace (3)
Documentation (29)
Resources (0)
Resource Groups (0)
Azure Active Directory (0)
Services
Azure Cosmos DB
Azure Cosmos DB API for MongoDB
Reservations
Azure Database for MySQL servers
Azure Arc
Azure Databricks
Azure Lighthouse
Azure Migrate
Marketplace
Azure Cosmos DB
Azure Cosmos DB API for MongoDB
Azure Cosmos DB Reserved Capacity
Documentation
Introduction to Azure Cosmos DB | Microsoft Docs
Azure Cosmos DB | Microsoft Docs
Azure Cosmos DB resource model | Microsoft Docs
Azure Cosmos DB free tier | Microsoft Docs
Use Azure Cosmos DB Explorer to manage your data
Introduction to the Azure Cosmos DB Table API | Microsoft Docs
Database security - Azure Cosmos DB | Microsoft Docs
Azure Cosmos DB service quotas | Microsoft Docs
Continue searching in Azure Active Directory


Home | Nuvepro
Subscription Det
Azure Cosmos DI
Inbox (10,612) -
Welcome Mail | D
Spark SQL Tutori
Apache Spark Ar
portal.azure.com/#blade/HubsExtension/BrowseResource/resourceType/Microsoft.DocumentDb%2FdatabaseAccounts
Microsoft Azure
Search resources, services, and docs (G+/)
Home
Azure Cosmos DB
Manipal Pro Learn (manipalazure.onmicrosoft.com)
Create
Restore
Manage view
Refresh
Export to CSV
Open query
Assign tags
Filter for any field...
Subscription == all
Resource group == all
Location == all
Add filter
Name
Status
Subscription
Write Reg
No Azure Cosmos DB accounts to display
Create a globally distributed, multi-model, fully managed database using API of your choice. Or try it for free, up 20k RU/s, for 30 days with unlimited renewal.
Create Azure Cosmos DB account
Try now

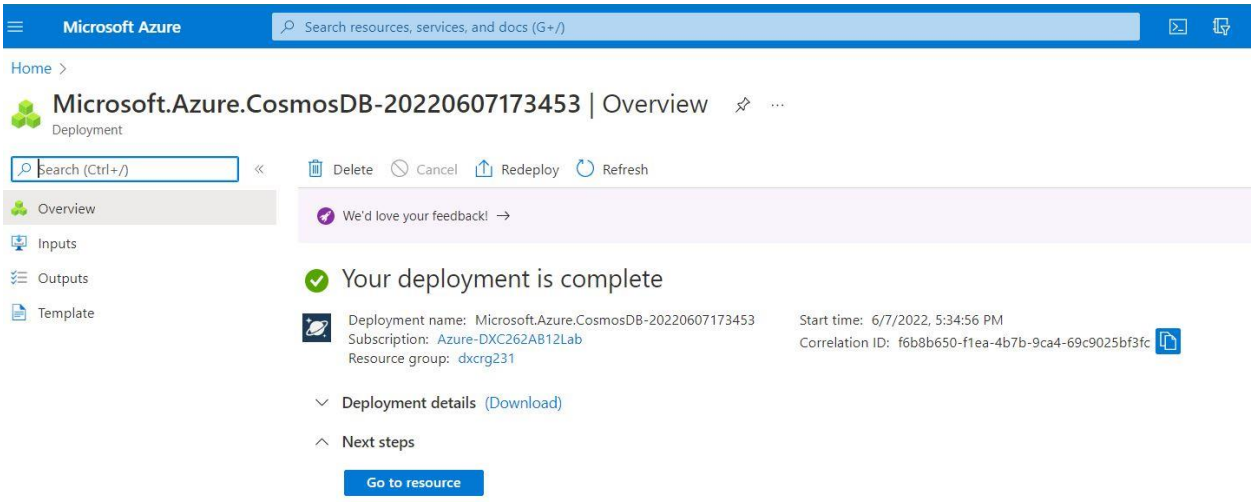

Home | Nuvepro
Subscription Deta
Select API option
Inbox (10,612) - s
Welcome Mail | D
Spark SQL Tutoria
portal.azure.com/#create/Microsoft.DocumentDB
Microsoft Azure
Search resources, services, and docs (G+/)
Home > Azure Cosmos DB >
Select API option
Which API best suits your workload?
Azure Cosmos DB is a fully managed NoSQL database service for building scalable, high performance applications. Learn more
To start, select the API to create a new account. The API selection cannot be changed after account creation.
Core (SQL) - Recommended
Azure Cosmos DB's core, or native API for working with documents. Supports fast, flexible development with familiar SQL query language and client libraries for .NET, JavaScript, Python, and Java.
Create
Learn more
Azure Cosmos DB API for MongoDB
Fully managed database service for apps written for MongoDB. Recommended if you have existing MongoDB workloads that you plan to migrate to Azure Cosmos DB.
Create
Learn more
Azure Table
Fully managed database service for apps written for Azure Table storage. Recommended if you have existing Azure Table storage workloads that you plan to migrate to Azure Cosmos DB, but do not want to re-write your application to use the
Gremlin (Graph)
Fully managed graph database service using the Gremlin query language, based on Apache TinkerPop project. Recommended for new workloads that need to store relationships between data.

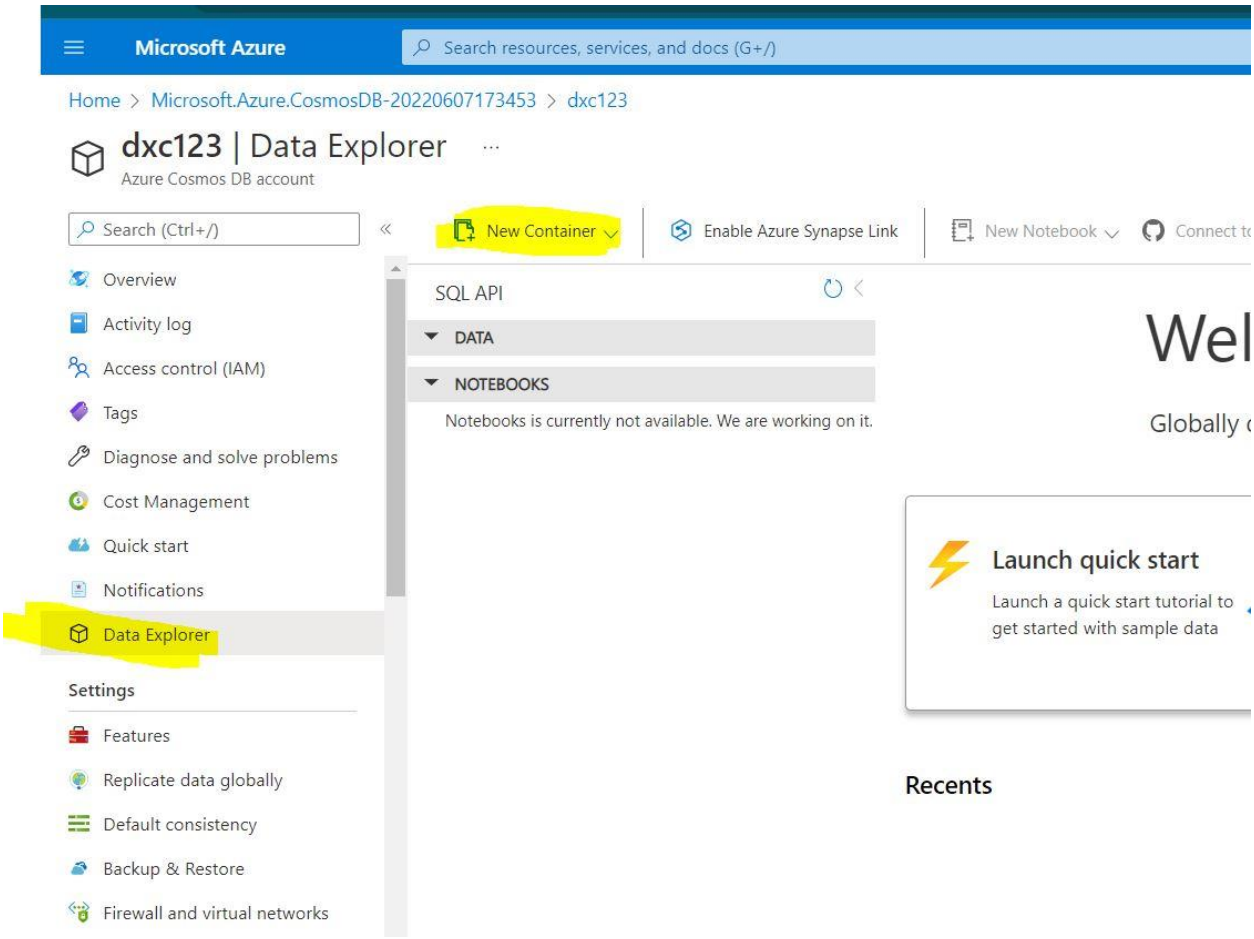

Microsoft Azure
Search resources, services, and docs (G+/)
Home > Azure Cosmos DB > Select API option >
Create Azure Cosmos DB Account - Core (SQL)
Validation Success
Estimated Account Creation Time (in minutes)
2
The estimated creation time is calculated based on the locatio
Basics
Subscription
Azure-DXC262AB12Lab
Resource Group
dxcrg231
Location
West US
Account Name
(new) dxc123
API
Core (SQL)
Capacity mode
Provisioned throughput
Geo-Redundancy
Disable
Multi-region Writes
Disable
Backup Policy
Backup policy
Periodic
Backup storage redundancy
Geo-redundant backup storage
Networking
Connectivity method
All networks
Create
Previous
Next
Download a template for automation

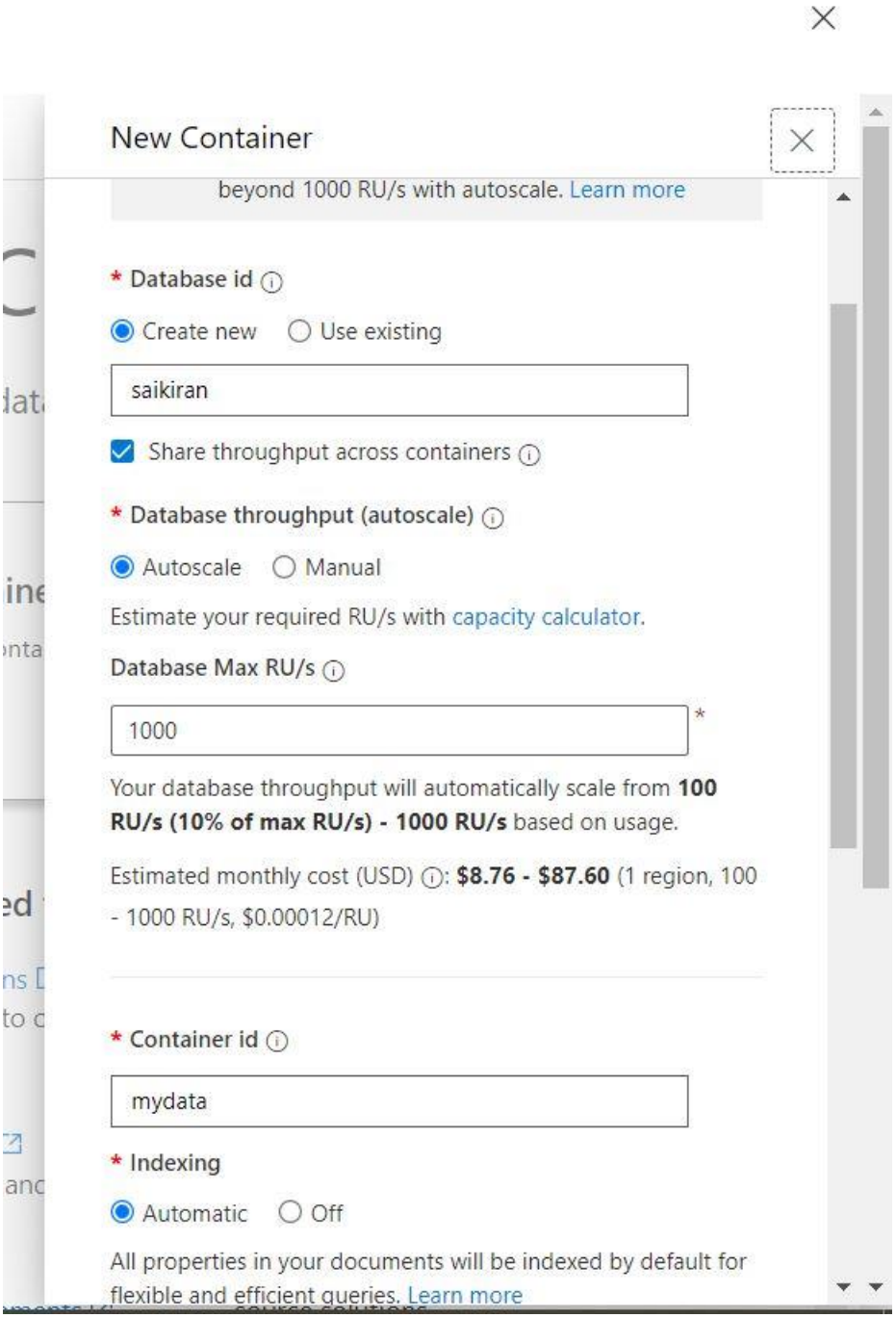

dxc262ab1227_1654530...
MANIPAL PRO LEARN (MANIPAL...
Submitting deployment...
Submitting the deployment template for resource group 'dxcrg231'.

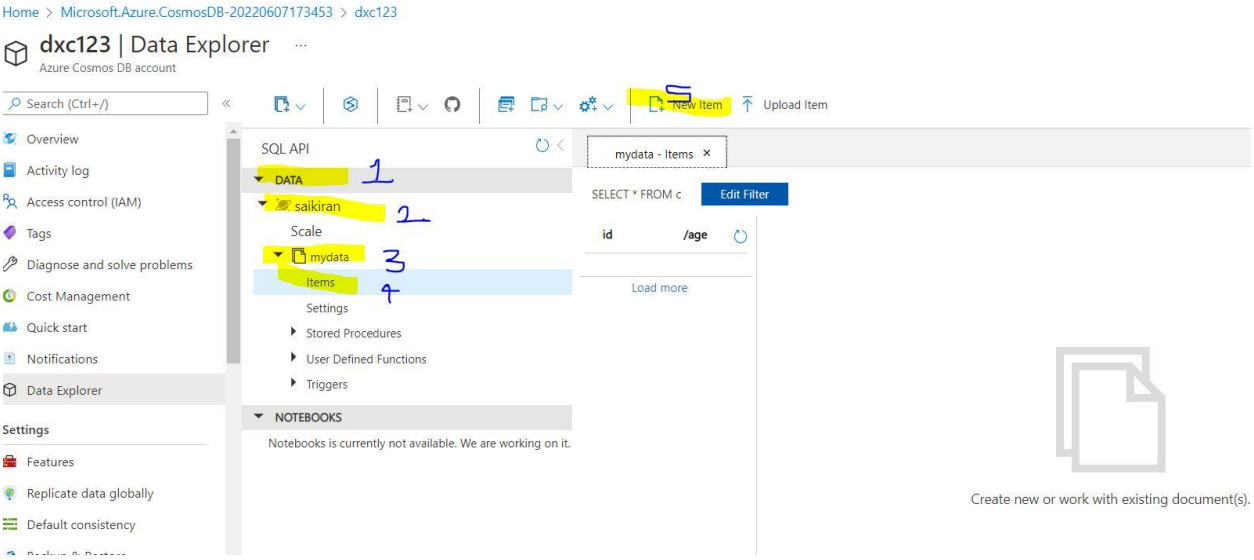

Microsoft Azure
Search resources, services, and docs (G+/)
dxc262ab1227_1654530...
MANIPAL PRO LEARN (MANIPAL...
Home
Microsoft.Azure.CosmosDB-20220607173453 | Overview
Deployment
Search (Ctrl+/)
Overview
Inputs
Outputs
Template
Delete
Cancel
Redeploy
Refresh
We'd love your feedback!
Deployment is in progress
Deployment name: Microsoft.Azure.CosmosDB-20220607173453
Subscription: Azure-DXC262AB12Lab
Resource group: dxcrg231
Start time: 6/7/2022, 5:34:56 PM
Correlation ID: f6b8b650-f1ea-4b7b-9ca4-69c9025bf3fc
Deployment details (Download)
Resource
Type
Status
Operation details
No results.

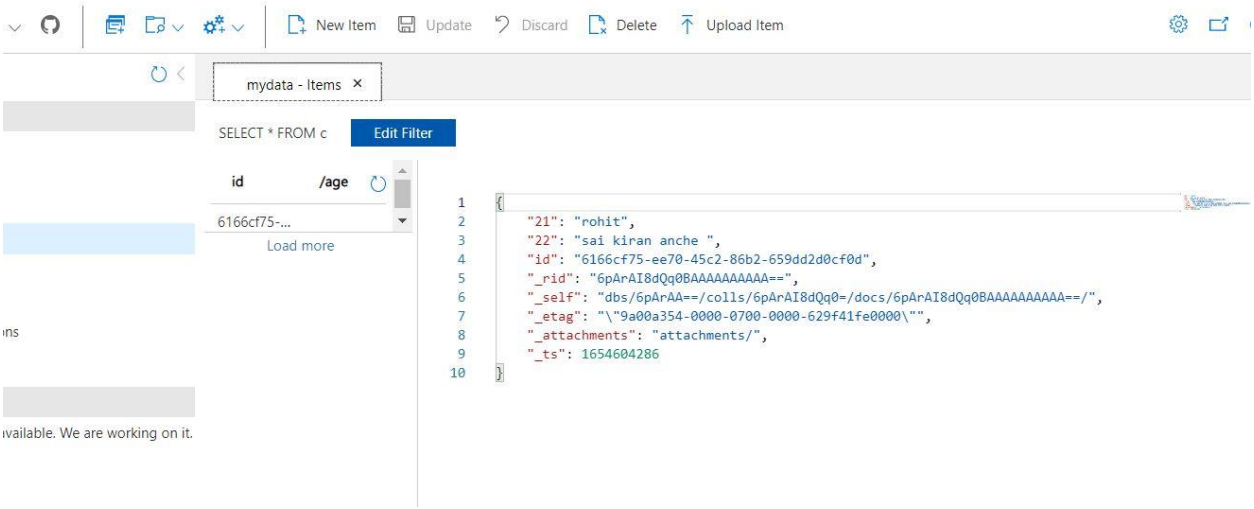
41°C Sunny
ENG
5:35 PM


Microsoft Azure
Search resources, services, and docs (G+/)
Home >
Microsoft.Azure.CosmosDB-20220607173453 | Overview
Deployment
Search (Ctrl+/)
Overview
Inputs
Outputs
Template
Delete
Cancel
Redeploy
Refresh
We'd love your feedback! →
Your deployment is complete
Deployment name: Microsoft.Azure.CosmosDB-20220607173453
Subscription: Azure-DXC262AB12Lab
Resource group: dxcrg231
Start time: 6/7/2022, 5:34:56 PM
Correlation ID: f6b8b650-f1ea-4b7b-9ca4-69c9025bf3fc
Deployment details (Download)
Next steps
Go to resource


Microsoft Azure
Search resources, services, and docs (G+/)
Home > Microsoft.Azure.CosmosDB-20220607173453 > dxc123
dxc123 | Data Explorer
Azure Cosmos DB account
Search (Ctrl+/)
New Container
Enable Azure Synapse Link
New Notebook
Overview
Activity log
Access control (IAM)
Tags
Diagnose and solve problems
Cost Management
Quick start
Notifications
Data Explorer
Settings
Features
Replicate data globally
Default consistency
Backup & Restore
Firewall and virtual networks
SQL API
DATA
NOTEBOOKS
Notebooks is currently not available. We are working on it.
Launch quick start
Launch a quick start tutorial to get started with sample data
Recents


New Container
beyond 1000 RU/s with autoscale. Learn more
* Database id
Create new
Use existing
saikiran
Share throughput across containers
* Database throughput (autoscale)
Autoscale
Manual
Estimate your required RU/s with capacity calculator.
Database Max RU/s
1000
*
Your database throughput will automatically scale from 100 RU/s (10% of max RU/s) - 1000 RU/s based on usage.
Estimated monthly cost (USD): $8.76 - $87.60 (1 region, 100 - 1000 RU/s, $0.00012/RU)
* Container id
mydata
* Indexing
Automatic
Off
All properties in your documents will be indexed by default for flexible and efficient queries. Learn more





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

A:

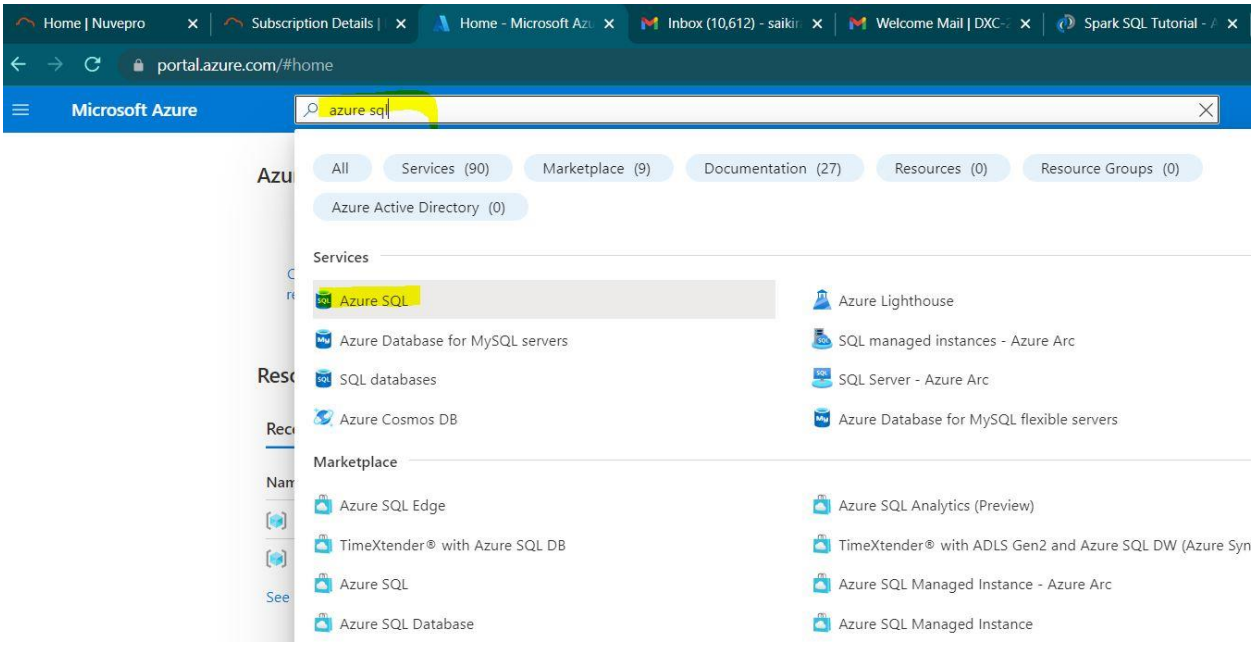
To create azure SQL Db follow the given steps :

Step 1:

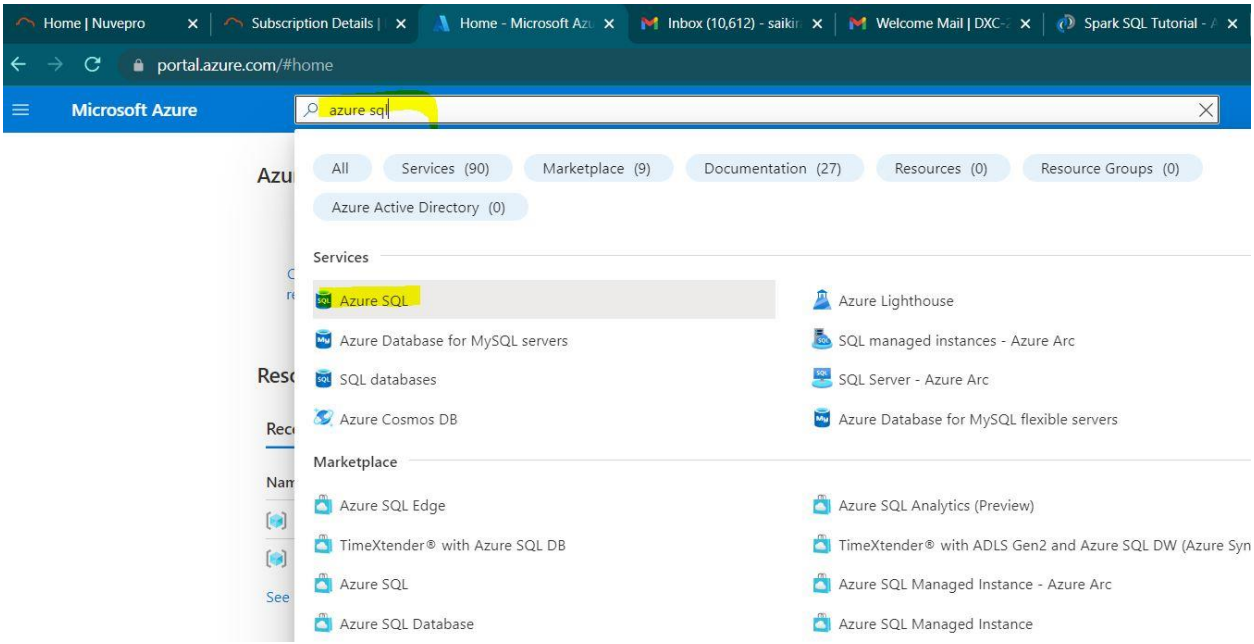
Open URL: https://portal.azure.com/#home

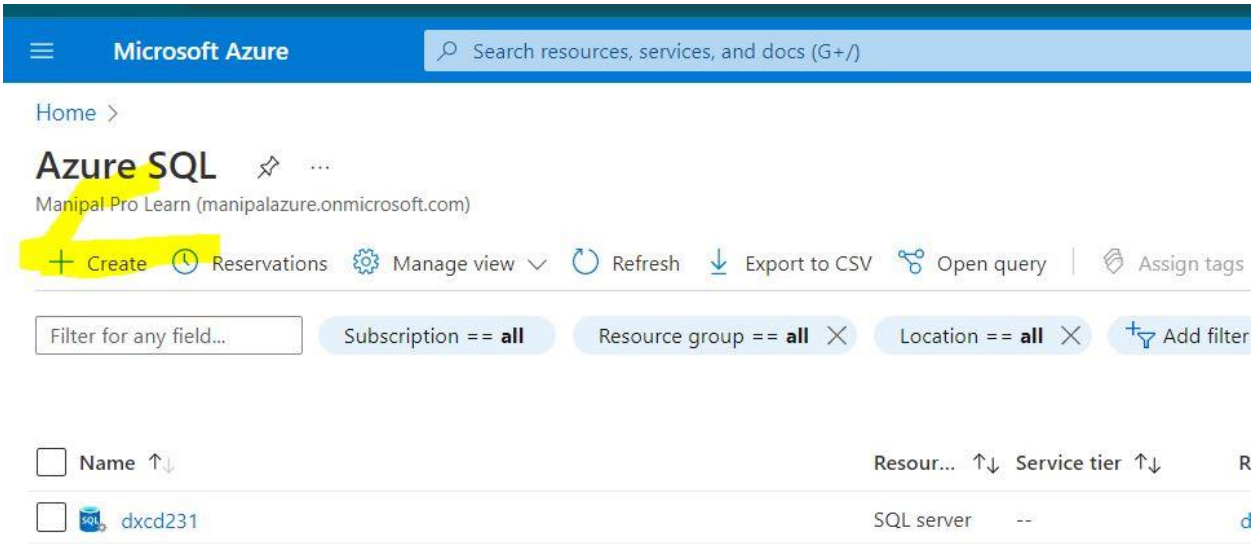
Step2:

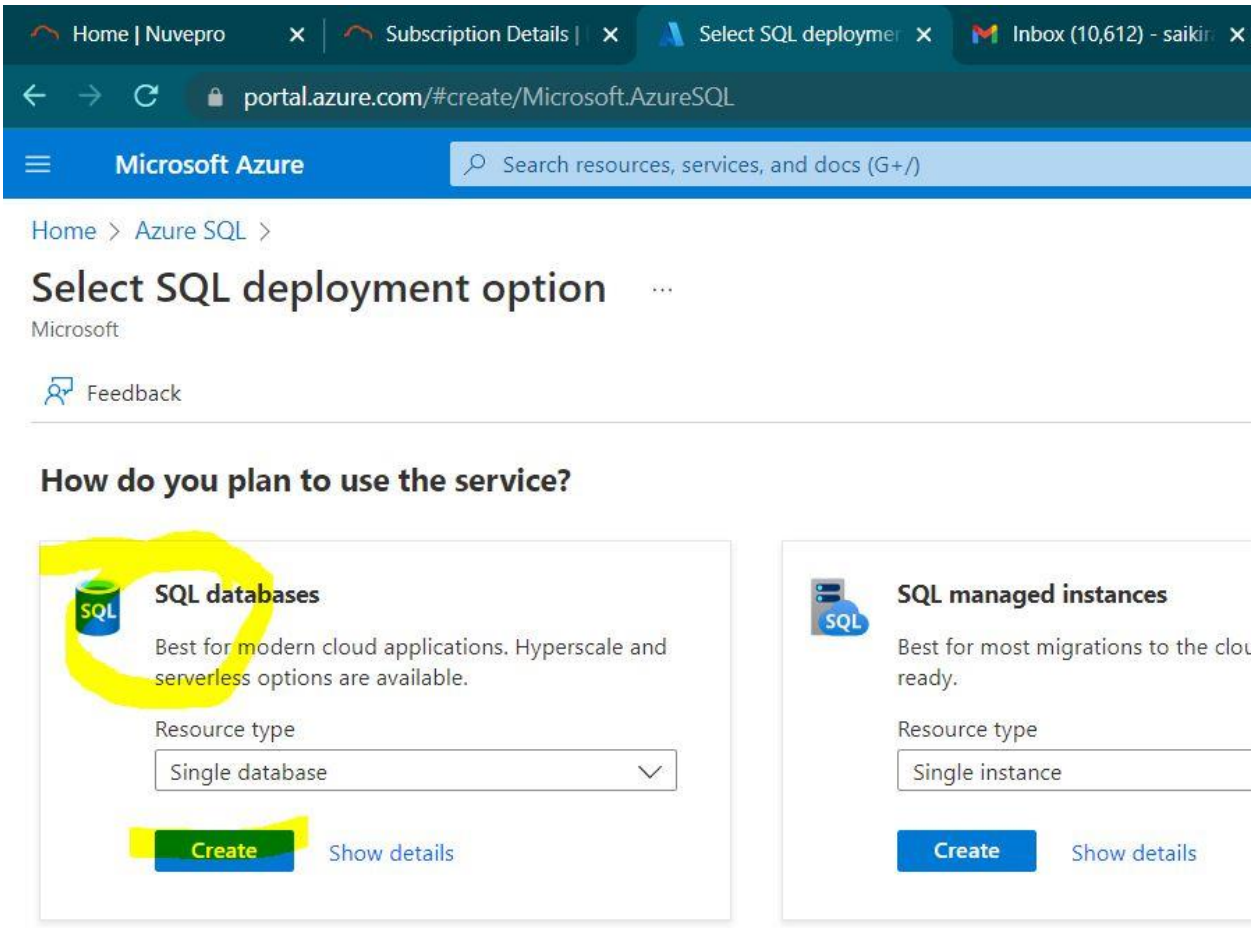
Log in and search Azure Sql in the search bar.



Step 3:




Microsoft Azure
Search resources, services, and docs (G+/)
Home >
Azure SQL
Manipal Pro Learn (manipalazure.onmicrosoft.com)
Create
Reservations
Manage view
Refresh
Export to CSV
Open query
Assign tags
Filter for any field...
Subscription == all
Resource group == all
Location == all
Add filter
Name
Resour...
Service tier
R
dxcd231
SQL server
--
d


Home | Nuvepro
Subscription Details |
Select SQL deploymer
Inbox (10,612) - saikir
portal.azure.com/#create/Microsoft.AzureSQL
Microsoft Azure
Search resources, services, and docs (G+/)
Home > Azure SQL >
Select SQL deployment option
Microsoft
Feedback
How do you plan to use the service?
SQL databases
Best for modern cloud applications. Hyperscale and serverless options are available.
Resource type
Single database
Create
Show details
SQL managed instances
Best for most migrations to the clou
ready.
Resource type
Single instance
Create
Show details

Create SQL Database ...

Microsoft

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * ⓘ: Azure-DXC262AB12Lab

Resource group * ⓘ: dxcrg231

Create new

Database details

Enter required settings for this database, including picking a logical server and configuring the compute and storage resources

Database name *: dxcdata123

Server * ⓘ: Select a server

Create new

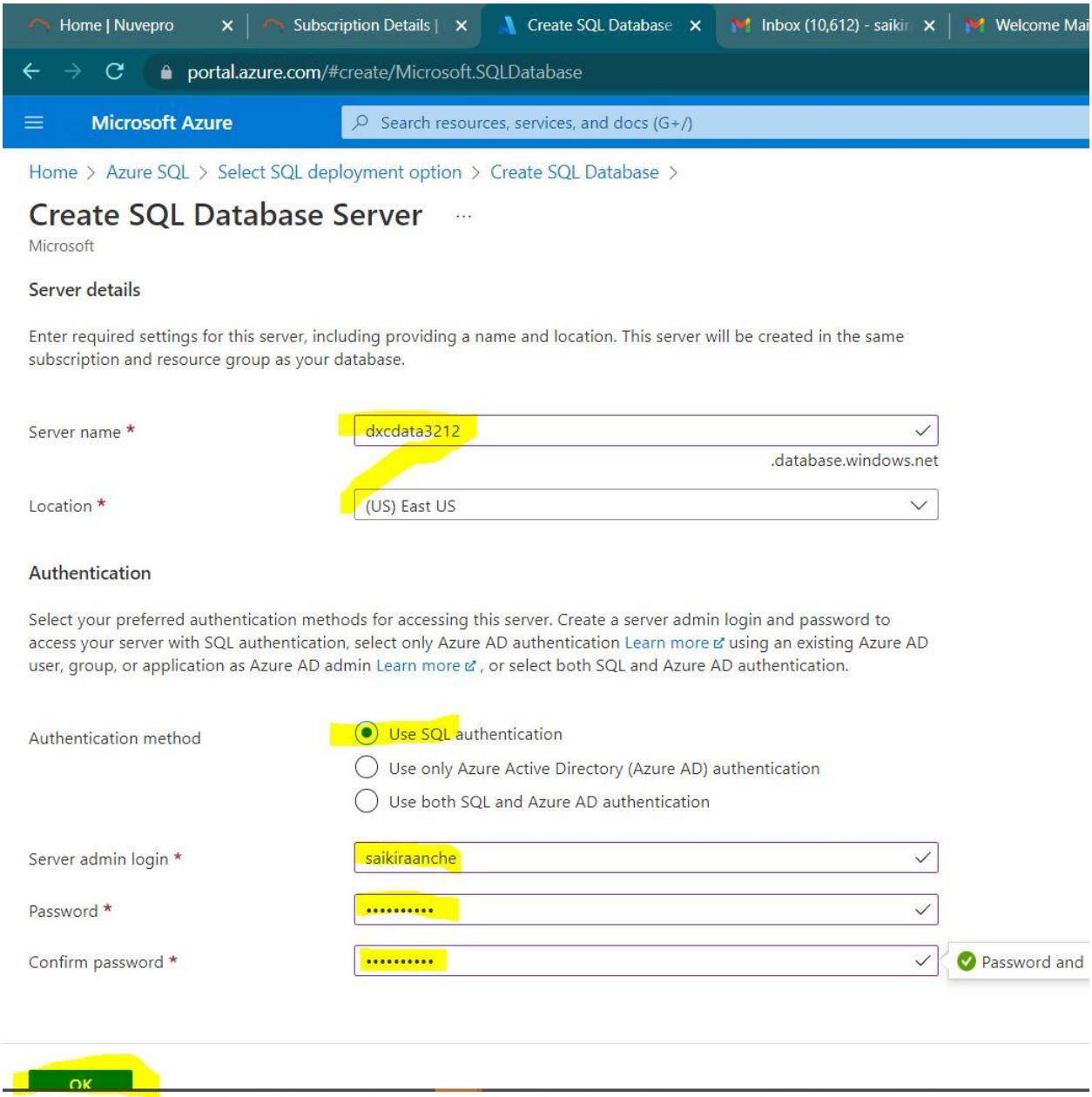
The value must not be empty.

Want to use SQL elastic pool? ⓘ: ○ Yes ◉ No

Compute + storage * ⓘ: **Please select a server first.**

Configure database

Backup storage redundancy


Home | Nuvepro
Subscription Details |
Create SQL Database
Inbox (10,612) - saiki
Welcome Mai
portal.azure.com/#create/Microsoft.SQLDatabase
Microsoft Azure
Search resources, services, and docs (G+/)
Home > Azure SQL > Select SQL deployment option > Create SQL Database >
Create SQL Database Server
Microsoft
Server details
Enter required settings for this server, including providing a name and location. This server will be created in the same subscription and resource group as your database.
Server name *
dxcdata3212
.database.windows.net
Location *
(US) East US
Authentication
Select your preferred authentication methods for accessing this server. Create a server admin login and password to access your server with SQL authentication, select only Azure AD authentication Learn more using an existing Azure AD user, group, or application as Azure AD admin Learn more , or select both SQL and Azure AD authentication.
Authentication method
Use SQL authentication
Use only Azure Active Directory (Azure AD) authentication
Use both SQL and Azure AD authentication
Server admin login *
saikiraanche
Password *
Confirm password *
Password and
OK

Create SQL Database ...

Microsoft

Enter required settings for this database, including picking a logical server and configuring the compute and storage resources

Database name *: dxcdata123

Server * (i): (new) dxcdata3212 (East US)

Create new

Want to use SQL elastic pool? (i): ○ Yes ◉ No

Compute + storage * (i): **General Purpose**
Gen5, 2 vCores, 32 GB storage, zone redundant disabled
Configure database

Backup storage redundancy

Choose how your PITR and LTR backups are replicated. Geo restore or ability to recover from regional outage is only available when geo-redundant storage is selected.

Backup storage redundancy (i):
- ○ Locally-redundant backup storage
- ○ Zone-redundant backup storage
- ◉ Geo-redundant backup storage

⚠ Selected value for backup storage redundancy is Geo-redundant backup storage. Database backups will be geo-replicated which might impact your data residency requirements. Learn more

Review + create | Next : Networking >

Home > Azure SQL > Select SQL deployment option > Create SQL Database >

Configure ...

Feedback

Service and compute tier

Select from the available tiers based on the needs of your workload. The vCore model provides a wide range of configuration controls and offers Hyperscale and Serverless to automatically scale your database based on your workload needs. Alternately, the DTU model provides set price/performance packages to choose from for easy configuration. Learn more

Service tier: General Purpose (Scalable compute and storage options)

Compare service tiers

Compute tier

- **Provisioned** - Compute resources are pre-allocated. Billed per hour based on vCores configured.
- **Serverless** - Compute resources are auto-scaled. Billed per second based on vCores used.

Compute Hardware

Select the hardware configuration based on your workload requirements. Availability of compute optimized, memory optimized, and

Feedback

1

PLEASE

Min vCores

0.5 vCores

2.02 GB MIN MEMORY 3 GB MAX MEMORY

Auto-pause delay

The database automatically pauses if it is inactive for the time period specified here, and automatically resumes when database activity recurs. Alternatively, auto-pausing can be disabled.

Enable auto-pause

Days: 0 Hours: 1 Minutes: 0

Data max size (GB)

1

307.2 MB LOG SPACE ALLOCATED

Would you like to make this database zone redundant?

Yes No

Apply

Create SQL Database ...

Microsoft

Database name * dxcdata123

Server * (new) dxcdata3212 (East US)

Create new

Want to use SQL elastic pool? ○ Yes ◉ No

Compute + storage * **General Purpose - Serverless**
Gen5, 1 vCore, 1 GB storage, zone redundant disabled
Configure database

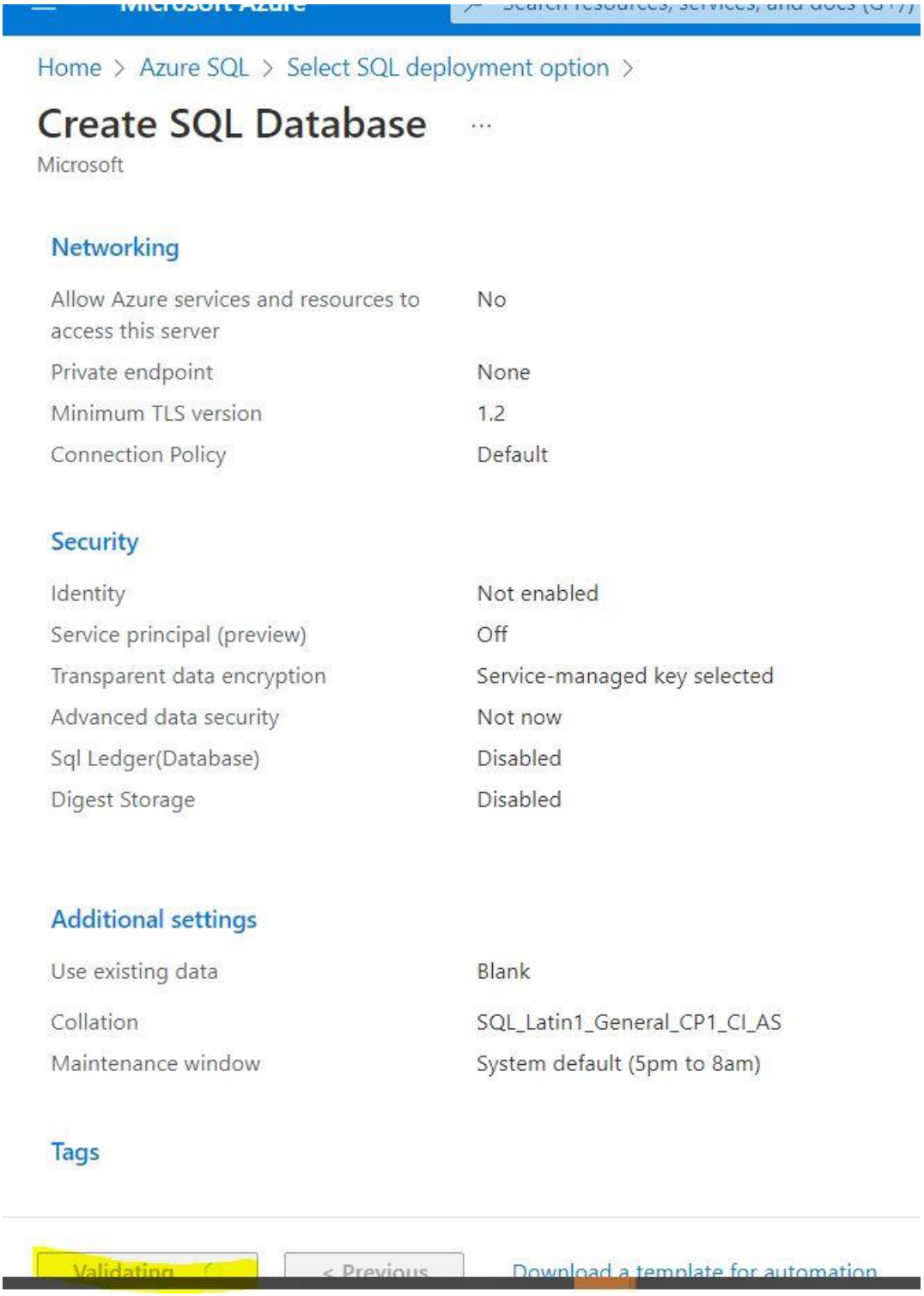
Backup storage redundancy

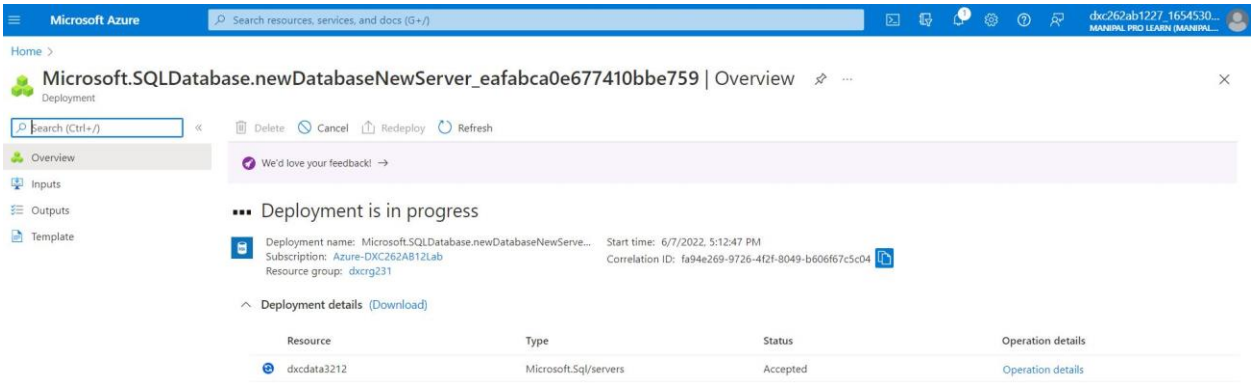
Choose how your PITR and LTR backups are replicated. Geo restore or ability to recover from regional outage is only available when geo-redundant storage is selected.

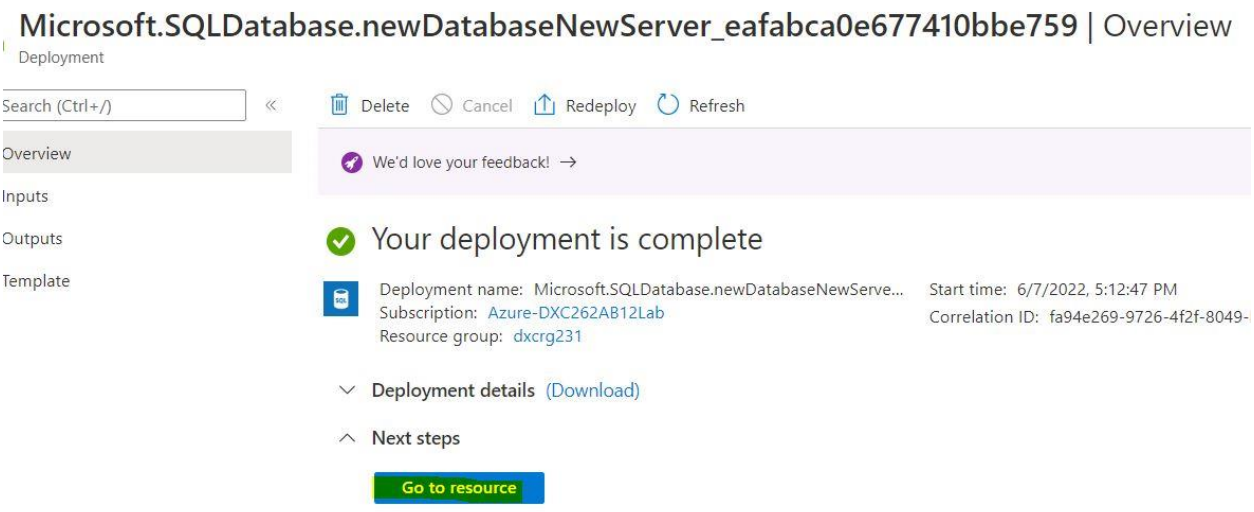
Backup storage redundancy
- ○ Locally-redundant backup storage
- ○ Zone-redundant backup storage
- ◉ Geo-redundant backup storage

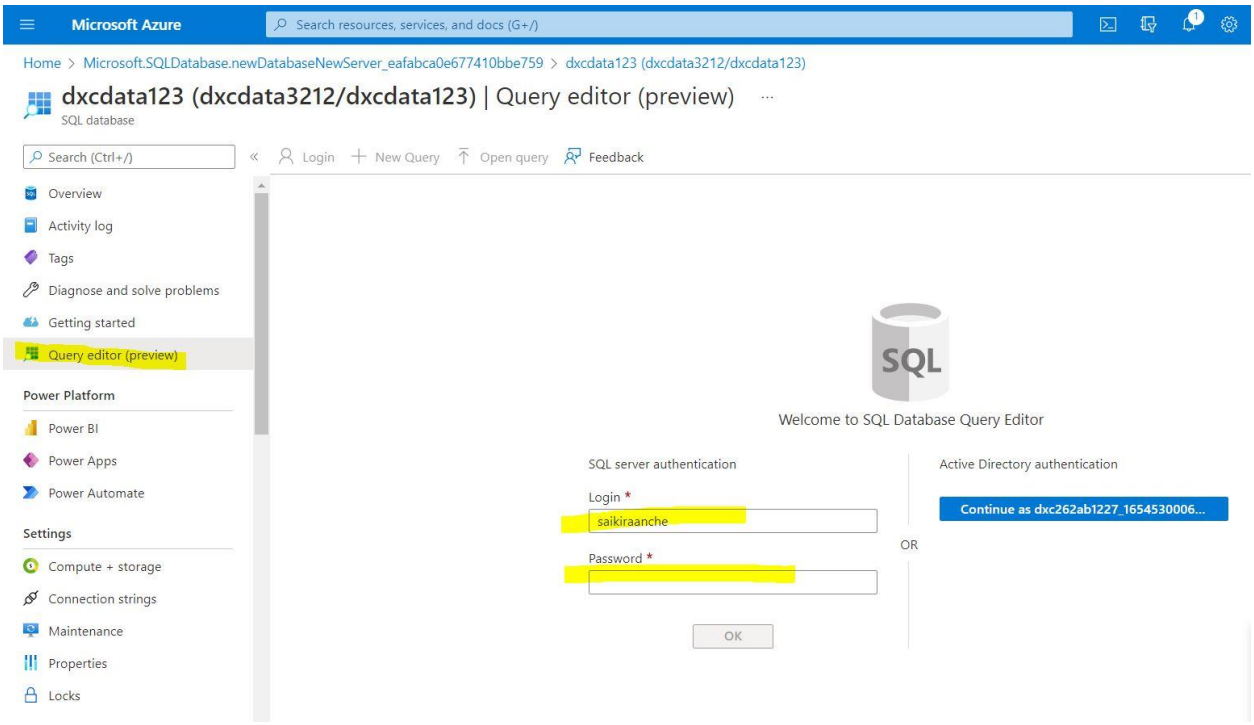
⚠ Selected value for backup storage redundancy is Geo-redundant backup storage. Database backups will be geo-replicated which might impact your data residency requirements. Learn more

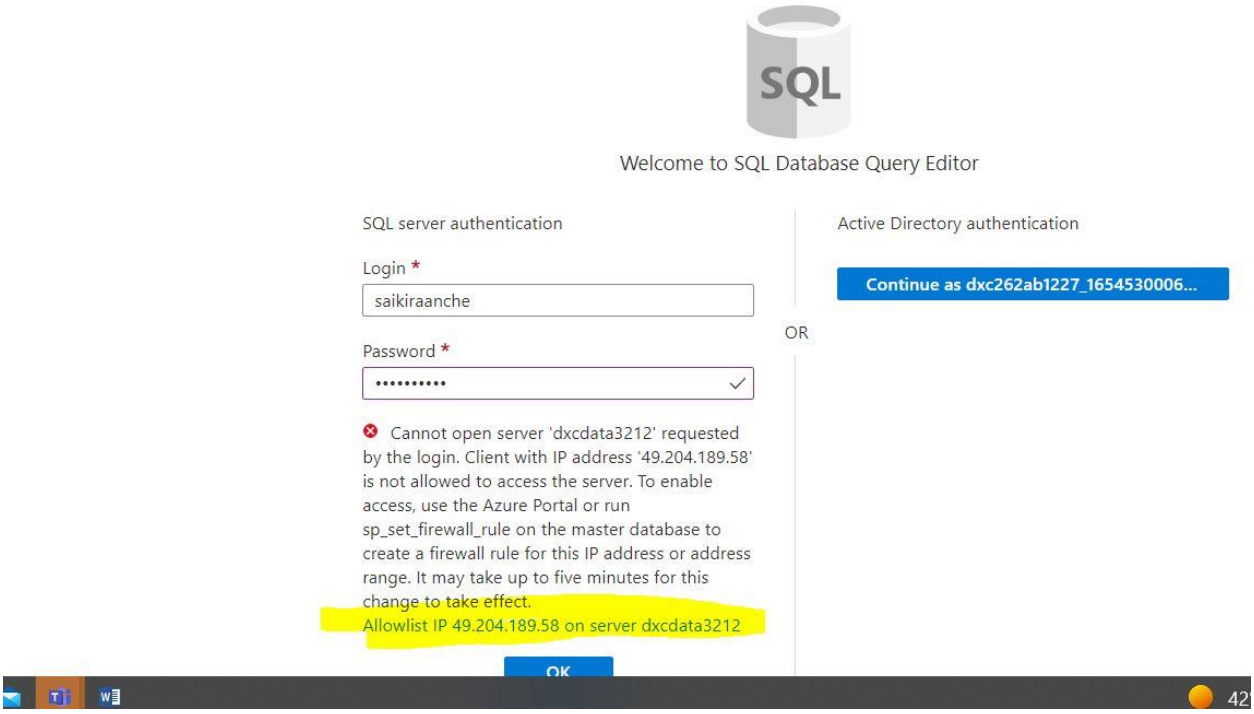
Review + create Next : Networking >

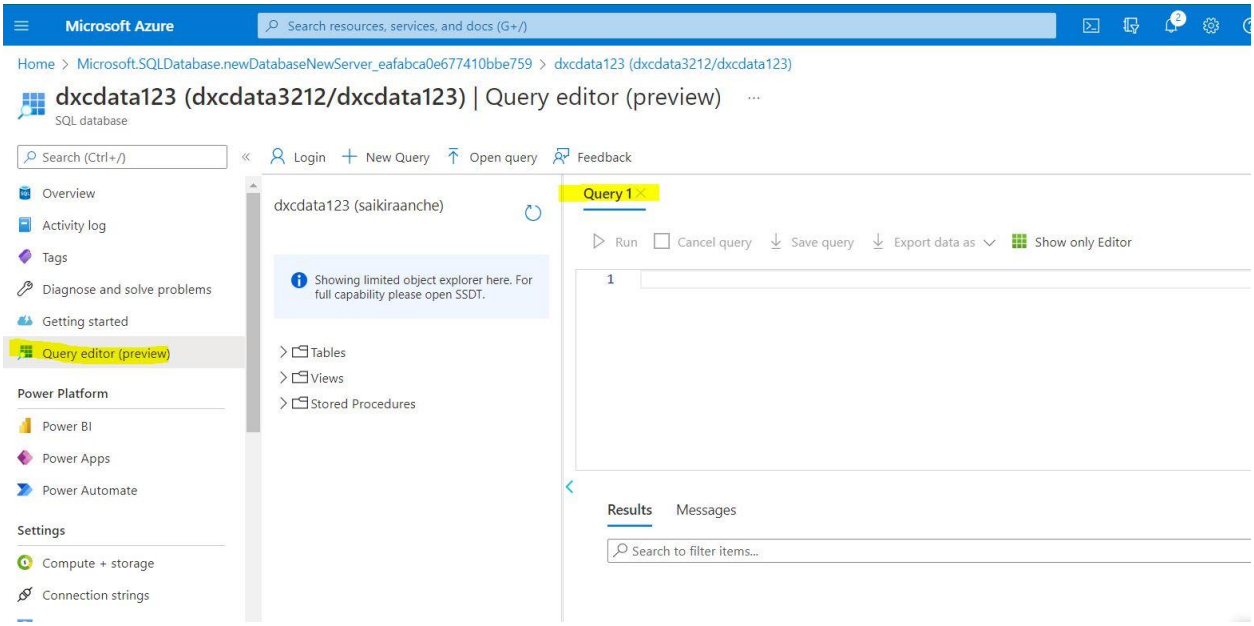

Home > Azure SQL > Select SQL deployment option >
Create SQL Database
Microsoft
Networking
Allow Azure services and resources to access this server
No
Private endpoint
None
Minimum TLS version
1.2
Connection Policy
Default
Security
Identity
Not enabled
Service principal (preview)
Off
Transparent data encryption
Service-managed key selected
Advanced data security
Not now
Sql Ledger(Database)
Disabled
Digest Storage
Disabled
Additional settings
Use existing data
Blank
Collation
SQL_Latin1_General_CP1_CI_AS
Maintenance window
System default (5pm to 8am)
Tags
Validating
< Previous
Download a template for automation

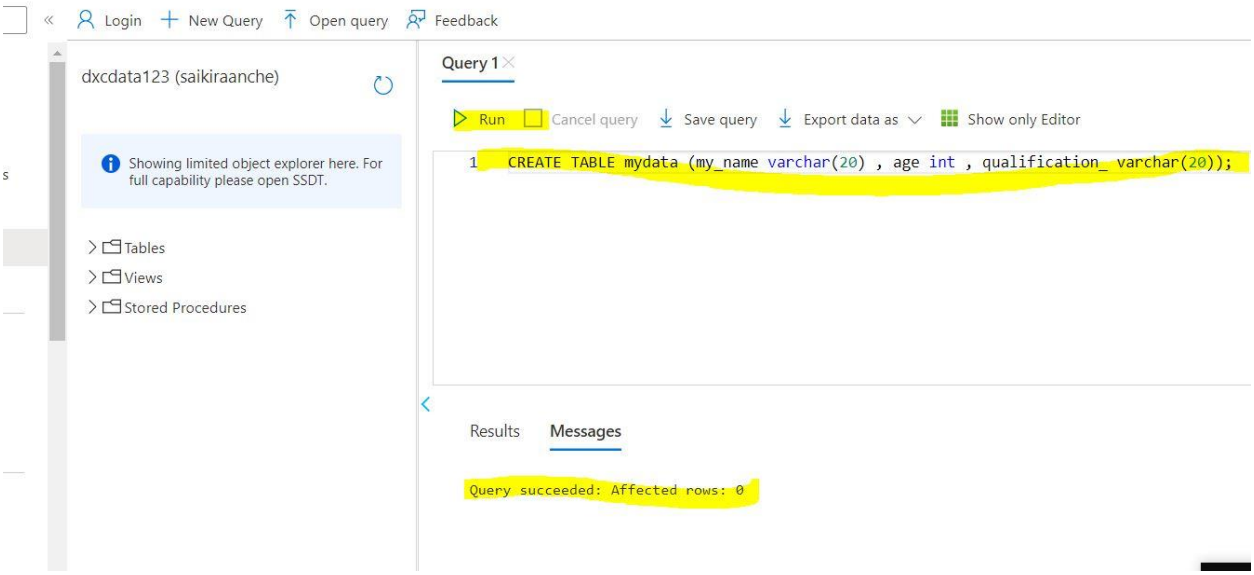

Microsoft Azure
Search resources, services, and docs (G+/)
Home
Microsoft.SQLDatabase.newDatabaseNewServer_eafabca0e677410bbe759 | Overview
Deployment
Search (Ctrl+/)
Overview
Inputs
Outputs
Template
Delete
Cancel
Redeploy
Refresh
We'd love your feedback!
Deployment is in progress
Deployment name: Microsoft.SQLDatabase.newDatabaseNewServe...
Subscription: Azure-DXC262AB12Lab
Resource group: dxcrg231
Start time: 6/7/2022, 5:12:47 PM
Correlation ID: fa94e269-9726-4f2f-8049-b606f67c5c04
Deployment details (Download)
Resource
Type
Status
Operation details
dxcdata3212
Microsoft.Sql/servers
Accepted
Operation details

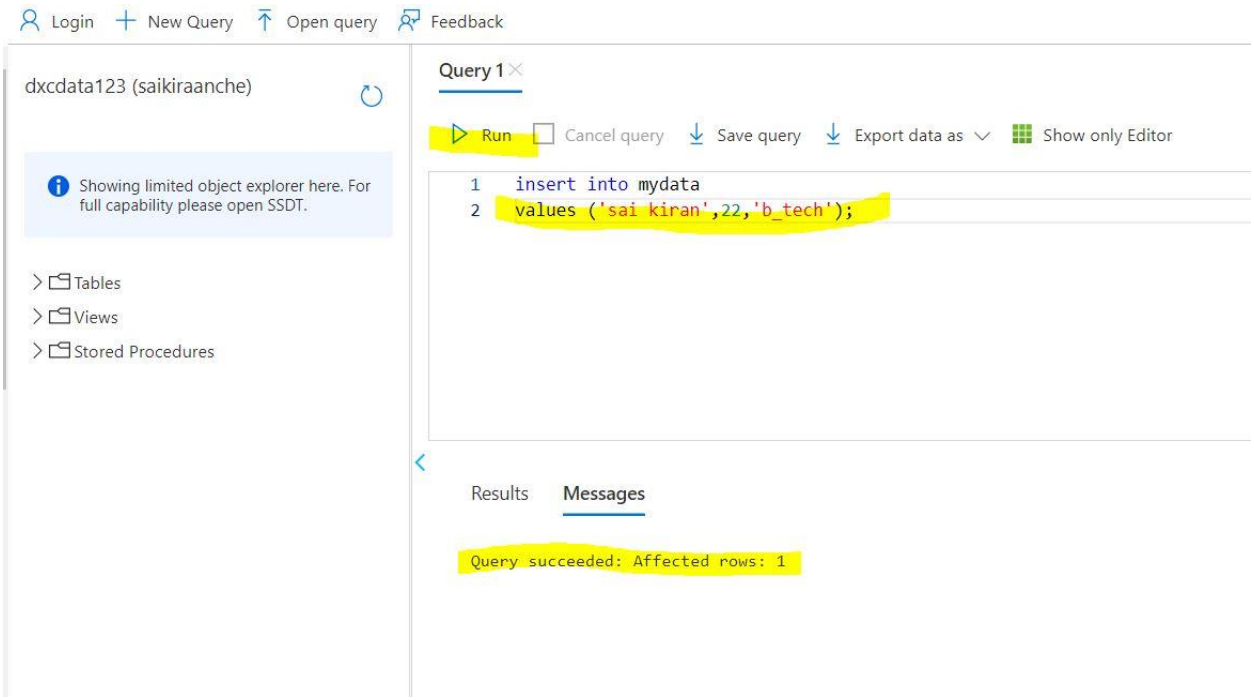

Microsoft.SQLDatabase.newDatabaseNewServer_eafabca0e677410bbe759 | Overview
Deployment
Search (Ctrl+/)
Overview
Inputs
Outputs
Template
Delete
Cancel
Redeploy
Refresh
We'd love your feedback!
Your deployment is complete
Deployment name: Microsoft.SQLDatabase.newDatabaseNewServe...
Subscription: Azure-DXC262AB12Lab
Resource group: dxcrg231
Start time: 6/7/2022, 5:12:47 PM
Correlation ID: fa94e269-9726-4f2f-8049-
Deployment details (Download)
Next steps
Go to resource

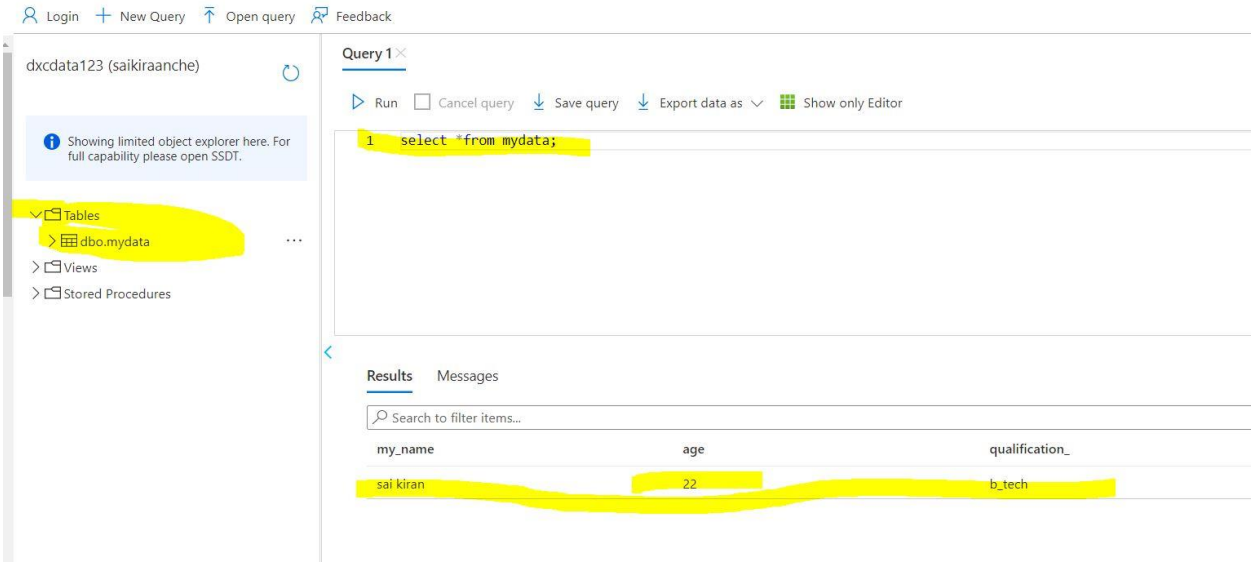

Microsoft Azure
Search resources, services, and docs (G+/)
Home > Microsoft.SQLDatabase.newDatabaseNewServer_eafabca0e677410bbe759 > dxcdata123 (dxcdata3212/dxcdata123)
dxcdata123 (dxcdata3212/dxcdata123) | Query editor (preview)
SQL database
Search (Ctrl+/)
Login
New Query
Open query
Feedback
Overview
Activity log
Tags
Diagnose and solve problems
Getting started
Query editor (preview)
Power Platform
Power BI
Power Apps
Power Automate
Settings
Compute + storage
Connection strings
Maintenance
Properties
Locks
SQL
Welcome to SQL Database Query Editor
SQL server authentication
Login *
saikiraanche
Password *
OK
OR
Active Directory authentication
Continue as dxc262ab1227_1654530006...


SQL
Welcome to SQL Database Query Editor
SQL server authentication
Login *
saikiraanche
Password *
Cannot open server 'dxcdata3212' requested by the login. Client with IP address '49.204.189.58' is not allowed to access the server. To enable access, use the Azure Portal or run sp_set_firewall_rule on the master database to create a firewall rule for this IP address or address range. It may take up to five minutes for this change to take effect.
Allowlist IP 49.204.189.58 on server dxcdata3212
OK
OR
Active Directory authentication
Continue as dxc262ab1227_1654530006...


Microsoft Azure
Search resources, services, and docs (G+/)
Home > Microsoft.SQLDatabase.newDatabaseNewServer_eafabca0e677410bbe759 > dxcdata123 (dxcdata3212/dxcdata123)
dxcdata123 (dxcdata3212/dxcdata123) | Query editor (preview)
SQL database
Search (Ctrl+/)
Overview
Activity log
Tags
Diagnose and solve problems
Getting started
Query editor (preview)
Power Platform
Power BI
Power Apps
Power Automate
Settings
Compute + storage
Connection strings
Login
New Query
Open query
Feedback
dxcdata123 (saikiraanche)
Showing limited object explorer here. For full capability please open SSDT.
Tables
Views
Stored Procedures
Query 1
Run
Cancel query
Save query
Export data as
Show only Editor
1
Results
Messages
Search to filter items...


Login
New Query
Open query
Feedback
dxcdata123 (saikiraanche)
Showing limited object explorer here. For full capability please open SSDT.
Tables
Views
Stored Procedures
Query 1
Run
Cancel query
Save query
Export data as
Show only Editor
1 CREATE TABLE mydata (my_name varchar(20) , age int , qualification_ varchar(20));
Results
Messages
Query succeeded: Affected rows: 0


Login
New Query
Open query
Feedback
dxcdata123 (saikiraanche)
Showing limited object explorer here. For full capability please open SSDT.
Tables
Views
Stored Procedures
Query 1
Run
Cancel query
Save query
Export data as
Show only Editor
1 insert into mydata
2 values ('sai kiran',22,'b_tech');
Results
Messages
Query succeeded: Affected rows: 1


Login
New Query
Open query
Feedback
dxcdata123 (saikiraanche)
Showing limited object explorer here. For full capability please open SSDT.
Tables
dbo.mydata
Views
Stored Procedures
Query 1
Run
Cancel query
Save query
Export data as
Show only Editor
1 select *from mydata;
Results
Messages
Search to filter items...
my_name
age
qualification_
sai kiran
22
b_tech

RESULT

Almost all the test questions have been solved and presented successfully in the present document except few due to lack of data .

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.