

# Links

Friday, February 3, 2023 12:45 PM

Microsoft documentation	<a href="https://learn.microsoft.com/en-us/azure/data-factory/">https://learn.microsoft.com/en-us/azure/data-factory/</a>
Azure data Factory	<a href="#">#8. Azure Data Factory - Load data from On Premise SQL table to Azure Blob Storage</a>  All About BI channel on Youtube has 155 plus videos

<https://azurelib.freshlearn.com/member/#/login>

For examples of data factory  
<https://www.youtube.com/@cloudguru3178/videos>

Pending  
CI/CD pipeline build - adding paramter file  
Connect adf to git

**Lookup**  
-- can we lookup for a value in table like customer id from customers exists in Lookup\_C table  
-- Lookup at the target if table exists or create it and then insert the data

# Course summary

Thursday, February 2, 2023 3:39 PM

1. Storage account - create blob storage and ADLS (gen2)
  2. Within storage create a container / Root directory - name staging/ landing / raw .
  3. Set the public access level and upload the file
  4. For ADLS create a directory and sub directory and then upload a file  
\*\*\*NOTE - usually we do not come and create a storage, containers, upload the files . It is already setup for automatic load. Example facebook app is linked directly to store images in BLOB.  
Or we have some database from which we pull the data using pipeline /ADF , store the data in Blob/ ADLS , then we use databricks /sql to prepare a result
  5. Not down the the access key
- 
1. Note down the shared access signature SAS \*\*\*\*
  5. Download storage account explorer and login via subscription so that you will see all underlying storages .
  6. Create a SQL server and database
  7. Create azure data factory account and open azure data factory studio
  8. Create a new Azure linked service to azure sql db
  9. Create another azure linked service to connect to ADLS storage
  10. Create a dataset from sql db employee table and link to DB service
  11. Create a dataset from ADLS and link to ADLS storage - select target as a CSV
  12. Create a pipeline with Copy data source task . Add source, Sink , target sink set row header

## Video 4

### Designing Dynamic flow

1. Select dynamic Source DB - using parameterized schema and db name
2. Select a Target ADLS , and add a parameter to select a path to create or select a dynamic folder ... Paramter value should be DeptFolder Contact with date time , so everytime new file is added it will not overwrite existing one , it will create new folder
3. Create a pipeline and add above source and target. At source append a paramter value for file path as

```
@concat('DeptFolder/',formatDateTime(utcnow(),'yyyy'),  
formatDateTime(utcnow(),'mm'),  
formatDateTime(utcnow(),'dd'),  
formatDateTime(utcnow(),'hh'),  
formatDateTime(utcnow(),'mm'),  
formatDateTime(utcnow(),'ss'))
```

## Video 5

--Parameterized linked service - not so common

Used to make linked service common to multiple database. Here we will need 3 params - db name,uname,password

At dataset In ADF new link service property will be shown to enter dbname,uname,pwd - At dataset level - add a one more parameter to enter this db name dynamically at pipeline level

-Lookup Activity in ADF

Example - Create lookup activity to check if employee table count is greater than 1 , Create a If activity , True create a copy activity from Source DB to ADLS file

--Nested Lookup - If else if else

Check If empCount>100 then Copy emp data and within this task Check if Product count is >100 then copy product data

For this Inside first If True activity call another pipeline which will check => If pdorcut count >100 then copy product data . We do not have Nested if else , but we can call pipeline within tasks

## --Video 6

**For Loop** -- SELECT SEQUENTIAL OPTION IF WE WANT TO RUN THE ETL ONE AFTER OTHER , IF YOU WANT TO RUN IT IN PARALLEL THEN LEAVE THE OPTION

With sequential and batch count set to 2 - it will perform the 2 parallel ETL at a time

Example 1 - **Lookup\_department pipeline**

Apply lookup function on department - pass the department id to for loop - for each department fetch employee details and create department wise list

## FAIL EXCEPTION activity

**IF pipeline is failed then call this activity to give customize exception**

## --Video 7

Metadata activity - file is greater than 10mb, if file is modified one hour back / one day back / if file exists , list of child items -list of folder files and scan

### VARIABLES

We can use it to record start and end time

### SET VARIABLE FUNCTION

## --Video 8

Link service - REST API |REST GET METHOD

USE BASE URL and then relative url to filter the data example [www.data.com](http://www.data.com) is base url and then [www.data.com/country/india](http://www.data.com/country/india) So \Country\parameter will become relative URL

Link service - REST DATA TYPE

## Video 9

Incremental pipeline

Copy data after specific timestamp only

## Video 10

Pipeline parameters - are defined at pipeline level and can be edited at run time .

When we defined inside at data source level or any other element,,you will be able to run it at debug level and define it in advance .

### Global parameter

This is global across the data factory and can be consumed by any pipeline . Not at underlying data set levels (@pipeline.globalname.databasename())

Debug - here we can click on red circle top on each activity and change it to debug mode , so that pipeline will run only till that point

Trigger- when u set a trigger , it means u want to run entire pipeline

## Video 11

Create a key vault - secret key

## Video 12

Scheduling a pipeline using triggers

4 types of triggers

Scheduled - regular pipeline for ETL

Tumbling window

Storage events - Run it when storage is available . Means file is placed in source | specify storage name and container name and path | Ignore empty file | blob create or delete  
For this under subscription , Microsoft eventGrid and Hub must be registered under Resource Providers

Custom events

One trigger can be attached to multiple pipelines

If file upload is failed then we need to manually load it after , by clicking on restart

## Video 13

Tumbling window trigger - frequency based trigger ,we can set dependency , failover retry policy | attached to one pipeline only

**Integration runtime** - provides computation power . Information is in different format so we need some mechanism for converting the format. So this comes in picture

For any linked service default is - AutoResolveIntegrationRuntime. At run time it checks which region is efficient to provide IR , either at source region or target region .

For faster performance IR should be near target . It try to provide nearest region. We ca

3 types - Azure integration runtime ( )

Self hosted Integration runtime ( on premise - private network / office network)

Azure -SSIS IR - lift and shift SSIS package to azure

## Video 14

Virtual machine creation - we can create windows machine . We can download software on this machins , visual studio virtual machine

Self hosted Integration runtime ( on premise - private network / office network) - So this we need to install integration run time on local server . Or virtual machine where we may have our target DB

**IF source is - self hosted IR and target is auto IR then by default self hosted IR will be used for operation**

## Video 15

**Dataflows - Data flows allow data engineers to develop data transformation logic without writing code . No need to know more spark, Scala or language it uses spark cluster at the backend**

Data flow debug creates a spark cluster at the back. Mainly used for datasets which is not a table where we can simply fire a query for transformation

Similar to Tableau Prep where we can work on data preparation and transformation

Dataflow is customize activity ,so to run it we need to create a data pipeline

So behind all, it runs on spark which is on big data , which is high performance engine . Also instead of creating data bricks we can use dataflow as we don't need separate maintenance.

But it can become lengthy as thing which we can do in one sql , we might have to add multiple steps

## Video 16

**Code management using - GIT (Internal ) or Git Hub ( Cloud -public or private repo )**

- Azure DevopsGit

- o Create a new orgs- add projects - new repo

- GitHub

Create dev repo

Create UAT repo

Create a CI/CD deployment pipeline

## Video 17 - PENDING

**Logic apps -similar to data factory where we can create a workflow / orchestration - step by step workflow where we can integrate different services**

## Video 18,19,20

Python Intro

## Video 21

**Add annotation** - to identify similar type of work / pipelines / team name . So in run history annotation may help

- to pipeline , trigger

**User properties** - we can add user properties to each step in data pipeline , data flow . We can use auto generate option which may generate source and target

we can mention query used in step or any other things which can help us to debug immediately at run time

Annotation is at pipeline level, user properties are at individual activity level

### Reduce cost of pipeline

- No of DIU \* duration \* cost

- DIU - data integration unit (memory , machines ,processors )

- Higher the DIU higher the cost . Keep DIU to 2 or try to reduce it

- Monitor Data movement activities at run window s

- 

- Enable staging (valid for only few data source ) - this improves the performance by staging a data in intermediate space from source and then copy to target

### How will you manage the errors

- In data copy operation we may encounter bad records
- In setting we have **fault tolerance** option - skip incompatible rows, skip missing files , etc
- Enable logging - to catch the lines which are failing
- Logging level - warning for skipped file - specify the location for storing log files
- After running the pipeline we can check the run log how many rows read, how may rows skipped

# Blob Storage

Wednesday, February 22, 2023

1:14 PM

**Blob storage** account -

**ADLS** Select data lake storage gen2 to create ADLS - data lake storage account

**FileShare** - is used by java dev/ cloud developer and used at application end. Data engineer uses only above two . You can map your local drive over network to file share, so whatever file you add to local driver will get synced

**Queue service**- Applications which talk to each other | JMS

**Azure tables** - no sql database . **Column based database** - we use **COSMOS DB** instead which is more advance

# Step1 Create Resource Group

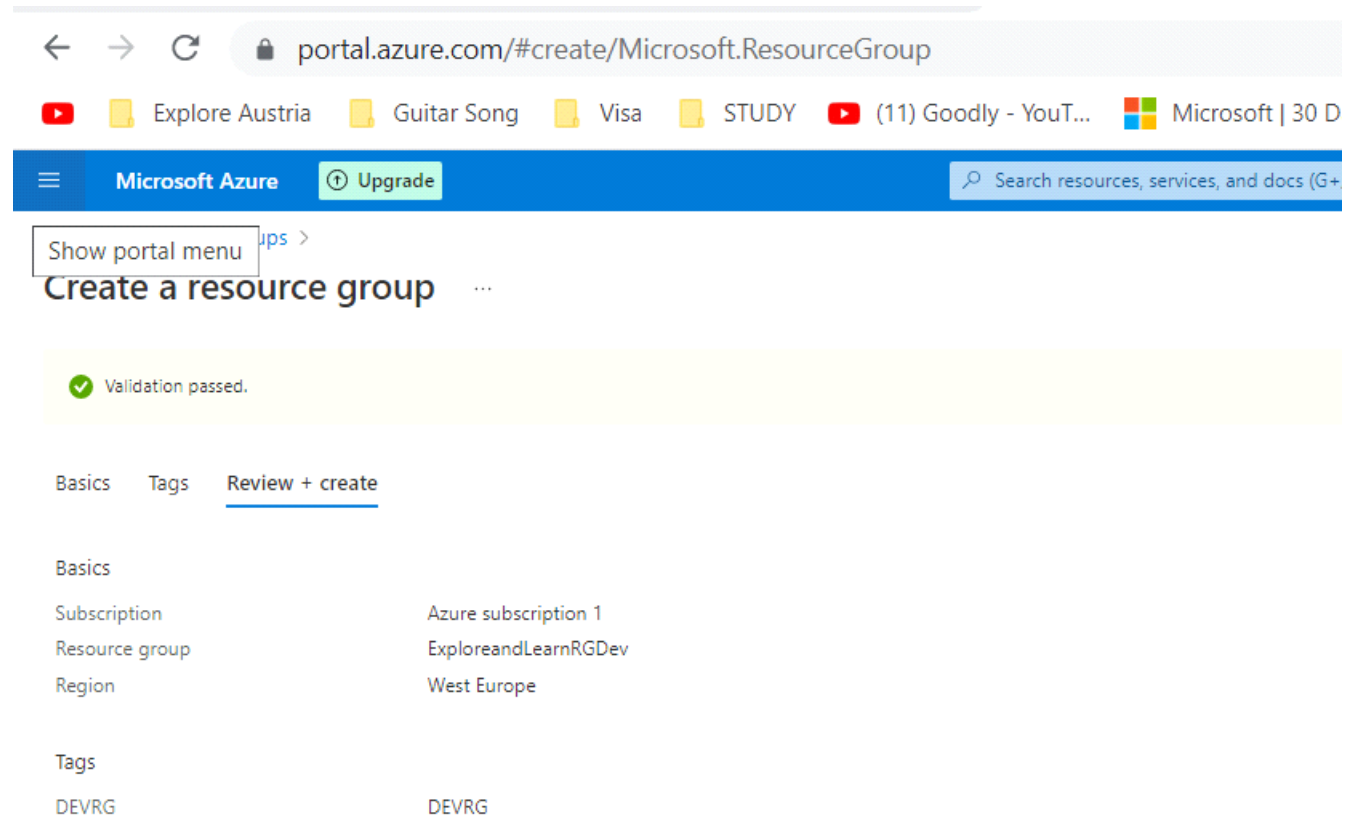
Wednesday, February 22, 2023 2:32 PM

1 Enter Resource group name **ExploreandLearnRGDev**

2 Select the region closer to your physical location to get better performance

Skip Intermediate options

3 Add a tag for identification DEVRG



portal.azure.com/#create/Microsoft.ResourceGroup

Explore Austria Guitar Song Visa STUDY (11) Goodly - YouT... Microsoft | 30 D

Microsoft Azure Upgrade Search resources, services, and docs (G+

Show portal menu

## Create a resource group

Validation passed.

Basics Tags Review + create

Basics

Subscription	Azure subscription 1
Resource group	ExploreandLearnRGDev
Region	West Europe


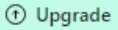
Tags

DEVRG	DEVRG
-------	-------

# Step2 Create Storage Account

Wednesday, February 22, 2023 2:32 PM

1. Click on create storage account and provide name exploreandlearnsdev
2. Select Resource group name **ExploreandLearnRGDev**
3. Select the region closer to your physical location to get better performance
4. Redundancy - For demo purpose select local option , cost efficient
5. Select the check box for **Data Lake Storage Gen2**. It will provide access to ADLS
6. Access tier - select cool for demo purpose
7. Data protection - keep retention period of 20 days for all options
8. Skip intermediate options
9. Add a tag DEVSG

 Microsoft Azure 

Search resources, services, and...

Home > Storage accounts >

## Create a storage account

Basics Advanced Networking Data protection Encryption Tags Review

### Basics

Subscription	Azure subscription 1
Resource Group	ExploreandLearnRGDev
Location	switzerlandnorth
Storage account name	exploreandlearnsdev
Deployment model	Resource manager
Performance	Standard
Replication	Locally-redundant storage (LRS)

### Advanced

Secure transfer	Enabled
Allow storage account key access	Enabled
Allow cross-tenant replication	Disabled
Default to Azure Active Directory authorization in the Azure portal	Disabled
Blob public access	Enabled
Minimum TLS version	Version 1.2
Permitted scope for copy operations (preview)	From any storage account
Enable hierarchical namespace	Enabled
Enable network file system v3	Disabled
Access tier	Cool
Enable SFTP	Disabled
Large file shares	Disabled

Create

< Previous

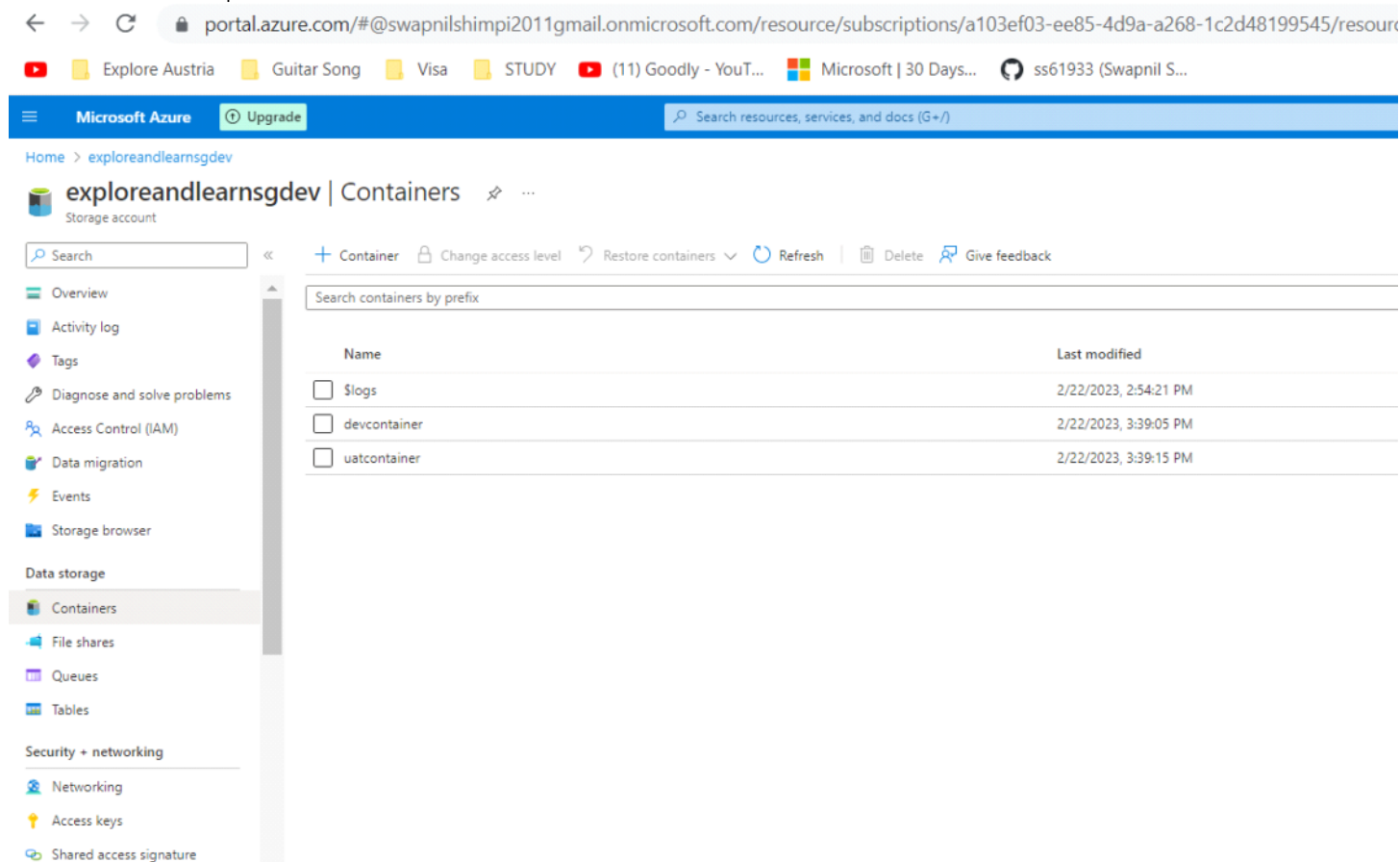
Next >

Download a template for automation

## Step3 Create storage containers

Wednesday, February 22, 2023 3:36 PM

1. Navigate to Storage **exploreandlearnsgdev**
2. Click on container option on left side and add below 2 containers



The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes the Microsoft Azure logo, an 'Upgrade' button, and a search bar. The breadcrumb trail indicates the user is in the 'exploreandlearnsgdev' storage account. The left sidebar contains a navigation menu with categories like 'Overview', 'Data storage', and 'Security + networking'. The 'Containers' option under 'Data storage' is selected. The main content area displays a table of containers with columns for 'Name' and 'Last modified'. The table lists three containers: '\$logs' (modified 2/22/2023, 2:54:21 PM), 'devcontainer' (modified 2/22/2023, 3:39:05 PM), and 'uatcontainer' (modified 2/22/2023, 3:39:15 PM). Each container has a checkbox to its left.

Name	Last modified
<input type="checkbox"/> \$logs	2/22/2023, 2:54:21 PM
<input type="checkbox"/> devcontainer	2/22/2023, 3:39:05 PM
<input type="checkbox"/> uatcontainer	2/22/2023, 3:39:15 PM

**Theory** - storage can be of Blob type or ADLS type  
Blob will not have sub directory



## Step4 Manage access to Container

Wednesday, February 22, 2023 3:45 PM

1. Select storage -> container -> Access control (IAM) -> Role Owner ->
2. Select the member from left menu. ( we can grant a role to groups also if present)

The screenshot shows the 'Add role assignment' page in the Microsoft Azure portal. The 'Members' tab is selected, and a 'Select members' sidebar is open on the right. The sidebar shows a search bar and a list of selected members, including 'swapnil shimpi'.

**Role assignment details:**

- Role:** Owner
- Assign access to:** ☒ User, group, or service principal
- Members:** + Select members
- Description:** Optional

**Select members sidebar:**

- Search by name or email address
- No users, groups, or service principals found.
- Selected members: swapnil shimpi, swapnil.shimpi2011@gmail.com#EXT#...

The screenshot shows the 'Add role assignment' page in the Microsoft Azure portal. The 'Review + assign' tab is selected, and the 'Role assignment' details are displayed. The 'Members' tab is also visible, showing the 'swapnil shimpi' member.

**Role assignment details:**

- Role:** Owner
- Scope:** /subscriptions/a103ef03-ee85-4d9a-a268-1c2d48199545/resourcegroups/ExploreandLearnRGDev/providers/Microsoft.Storage/storageAccounts/...
- Members:** swapnil shimpi
- Description:** No description

**Members tab details:**

Name	Object ID
swapnil shimpi	29cc5492-b334-42ec-9ade-1ac86e8da822

# Step5 Create Sql server and database

Wednesday, February 22, 2023 2:56 PM

1. Click on create Sql server , name the server **exploredbserver**
2. Select Resource group name **ExploreandLearnRGDev**
3. Select the region closer to your physical location to get better performance
4. Select " Use both SQL and Azure AD authentication " authentication method
5. Add server admin credentials- exploreadmin / Ss0061933!
6. Select Azure Active Directory as a "Azure SQL Database"
7. Set firewall rule "Allow Azure services and resources to access this serve" - YES
8. Add a tag DEVDBSERVER
9. Skip intermediate options
10. Add a tag DEVSG

The screenshot shows the 'Create SQL Database Server' wizard in the Azure portal, specifically the 'Review + create' step. The browser address bar shows 'portal.azure.com/#create/Microsoft.SQLServer'. The page has a blue header with the Microsoft Azure logo and a search bar. Below the header, the breadcrumb 'Home > SQL servers >' is visible. The main heading is 'Create SQL Database Server' with a sub-heading 'Microsoft'. The wizard has five tabs: 'Basics', 'Networking', 'Additional settings', 'Tags', and 'Review + create' (which is selected). Under 'Product details', it shows 'SQL Database Server by Microsoft' and 'Estimated cost per month: No additional charges'. A 'Terms' section contains a legal disclaimer. The 'Basics' section lists configuration details: Subscription (Azure subscription 1), Resource group (ExploreandLearnRGDev), Server name (exploreandlearnserver), Authentication method (SQL and Azure Active Directory authentication), Server admin login (exploreadmin), Azure AD Admin (Azure SQL Database), and Location (Sweden Central). The 'Networking' section shows 'Allow Azure services to access server' set to 'Yes'. The 'Additional settings' section shows 'Enable Microsoft Defender for SQL' set to 'Not now'. The 'Tags' section shows a tag 'DEVDBSERVER' with the value 'DEVDBSERVER (Server)'. At the bottom, there is a 'Create' button, a '< Previous' button, and a link 'Download a template for automation'.

portal.azure.com/#create/Microsoft.SQLServer

Microsoft Azure Upgrade Search resources, services, and docs (G+/I)

Home > SQL servers >

## Create SQL Database Server

Microsoft

Basics Networking Additional settings Tags **Review + create**

**Product details**

SQL Database Server  
by Microsoft  
[Terms of use](#) | [Privacy policy](#)

**Estimated cost per month**  
No additional charges

**Terms**

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payr usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party c

**Basics**

Subscription	Azure subscription 1
Resource group	ExploreandLearnRGDev
Server name	exploreandlearnserver
Authentication method	SQL and Azure Active Directory authentication
Server admin login	exploreadmin
Azure AD Admin	Azure SQL Database
Location	Sweden Central

**Networking**

Allow Azure services to access server	Yes
---------------------------------------	-----

**Additional settings**

Enable Microsoft Defender for SQL	Not now
-----------------------------------	---------

**Tags**

DEVDBSERVER	DEVDBSERVER (Server)
-------------	----------------------

Create < Previous Download a template for automation

## CREATING DATABASE

11. Click on create Sql database, name the server exploredbserver **under above server**
12. Service tier - Basic and Locally-redundant backup storage
13. Networking as below

Create SQL Database - Microsoft | portal.azure.com/#create/Microsoft.SQLDatabase

Microsoft Azure | Upgrade | Search resources, services, and docs (G+)

## Create SQL Database

Choose an option for configuring connectivity to your server via public endpoint or private endpoint. Choosing no access creates with defaults and you can configure connection method after server creation. [Learn more](#)

Connectivity method \*

- ☐ No access
- ☒ Public endpoint
- ☐ Private endpoint

Firewall rules

Setting 'Allow Azure services and resources to access this server' to 'Yes' allows communications from all resources inside the Azure boundary, that may or may not be part of your subscription. [Learn more](#)

Setting 'Add current client IP address' to 'Yes' will add an entry for your client IP address to the server firewall.

Allow Azure services and resources to access this server \*

Add current client IP address \*

Connection policy

Configure how clients communicate with your SQL database server. [Learn more](#)

Connection policy \*

- ☒ Default - Uses Redirect policy for all client connections originating inside of Azure and Proxy for all client connections originating outside Azure
- ☐ Proxy - All connections are proxied via the Azure SQL Database gateways
- ☐ Redirect - Clients establish connections directly to the node hosting the database

Encrypted connections

This server supports encrypted connections using Transport Layer Security (TLS). For information on TLS version and certificates, refer to connecting with TLS/SSL. [Learn more](#)

Minimum TLS version \*

Cost summary

Basic (Basic)	
Cost per DTU (in USD)	0.98
DTUs selected	x 5
<b>ESTIMATED COST / MONTH</b>	<b>4.90 USD</b>

Review + create | < Previous | Next: Security >

#### 14. To create a dummy schema and tables

Microsoft Azure | Upgrade | Search resources, services, and docs (G+)

## Create SQL Database

Microsoft

Basics Networking Security **Additional settings** Tags Review + create

Customize additional configuration parameters including collation & sample data.

Data source

Start with a blank database, restore from a backup or select sample data to populate your new database.

Use existing data \*

None Backup **Sample**

AdventureWorksLT will be created as the sample database.

Database collation

Database collation defines the rules that sort and compare data, and cannot be changed after database creation. The default database collation is SQL\_Latin1\_General\_CP1\_CI\_AS. [Learn more](#)

Collation \*

SQL\_Latin1\_General\_CP1\_CI\_AS

Cost summary

Basic (Basic)	
Cost per DTU (in USD)	0.98
DTUs selected	x 5
<b>ESTIMATED COST / MONTH</b>	<b>4.90 USD</b>

15. Add a tag DEVDB
16. Skip intermediate options
17. Add a tag DEVSG

portal.azure.com/#create/Microsoft.SQLDatabase

Explore Austria Guitar Song Visa STUDY (11) Goodly - YouT... Microsoft | 30 Days... ss61933 (Swapnil S...

Microsoft Azure Upgrade Search resources, services, and docs (G+)

Home > SQL databases >

## Create SQL Database

Microsoft

Basics Networking Security Additional settings Tags Review + create

**Product details**

SQL database  
by Microsoft  
[Terms of use](#) | [Privacy policy](#)

**Estimated cost per month**  
4.90 USD

**Terms**

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. For additional details see [Azure Marketplace Terms](#).

**Basics**

Subscription	Azure subscription 1
Resource group	ExploreandLearnRGDev
Region	North Europe
Database name	exploredb
Server	(new) exploredbserver
Authentication method	SQL and Azure Active Directory authentication
Server admin login	exploreadmin
Azure AD Admin	Azure SQL Database
Compute + storage	Basic: 1 GB storage
Backup storage redundancy	Locally-redundant backup storage

**Networking**

Allow Azure services and resources to access this server	Yes
Add current client IP address	Yes

**Cost summary**

<b>Basic (Basic)</b>	
Cost per DTU (in USD)	0.98
DTUs selected	x 5
<b>ESTIMATED COST / MONTH</b>	<b>4.90 USD</b>

**Create** < Previous [Download a template for automation](#)

## Validating database

### 1. Login to database using credentials exploreadmin/Ss0061933!

Home > Microsoft SQL Database new Database New Server\_c139226c04f6495190f14 | Overview > exploredb (exploredbserver/exploredb)

exploredb (exploredbserver/exploredb) | Query editor (preview)

SQL database

Search Login + New Query Open query Feedback

Overview Activity log Tags Diagnose and solve problems Getting started **Query editor (preview)**

**Settings**

- Compute + storage
- Connection strings
- Properties
- Locks

**Data management**

- Replicas
- Sync to other databases

**Integrations**

- Azure Synapse Link
- Stream analytics (preview)
- Add Azure Search

**Power Platform**

- Power BI

Welcome to SQL Database Query Editor

SQL server authentication

Login \*  
exploreadmin

Password \*  
Ss0061933! ✓

OK

Active Directory authentication

Login failed for user 'token-identified principal'.

OR

[Continue as swapnilshimpi2011@gm...](#)

## 2. Query the database

Home > Microsoft SQL Database newDatabaseNewServer\_c139226c047b495190f14 | Overview > exploredb (exploredbserver/exploredb)

exploredb (exploredbserver/exploredb) | Query editor (preview)

SQL database

Search

Overview  
Activity log  
Tags  
Diagnose and solve problems  
Getting started  
Query editor (preview)

Settings  
Compute + storage  
Connection strings  
Properties  
Locks  
Data management  
Replicas  
Sync to other databases  
Integrations  
Azure Synapse Link  
Stream analytics (preview)  
Add Azure Search  
Power Platform  
Power BI  
Power Apps

exploredb (exploreadmin)

Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

Tables

- dbo.BuildVersion
- dbo.ErrorLog
- SalesLT.Address
- SalesLT.Customer
- SalesLT.CustomerAddress
- SalesLT.Product
- SalesLT.ProductCategory
- SalesLT.ProductDescription
- SalesLT.ProductModel
- SalesLT.ProductModelProductDesc
- SalesLT.SalesOrderProductDesc
- SalesLT.SalesOrderHeader
- Views
- Stored Procedures

Query 1

Run Cancel query Save query Export data as Show only Editor

```
1 select * from
2 [SalesLT].[Customer]
```

Results Messages

Search to filter items...

CustomerID	NameStyle	Title	FirstName	MiddleName	LastName
1	False	Mr.	Orlando	N.	Gee
2	False	Mr.	Keith		Harris
3	False	Ms.	Donna	F.	Carrenas
4	False	Ms.	Janet	M.	Gates
5	False	Ms.	Lucy		Harrington
6	False	Ms.	Rosmarie	J.	Carroll
-	-	-	-	-	-

## Step6 Uploading files to ADLS account

Wednesday, February 22, 2023 4:16 PM

1. Select a storage account -> container -> sub container as shown below and upload a file

