

Index

Agenda	1
Services Required	1
Pipeline Architecture	2
Azure Blob Storage Creation	3
Azure SQL Database Creation	9
Azure Data Factory Creation	13
Transformation of data using Data Flow	15

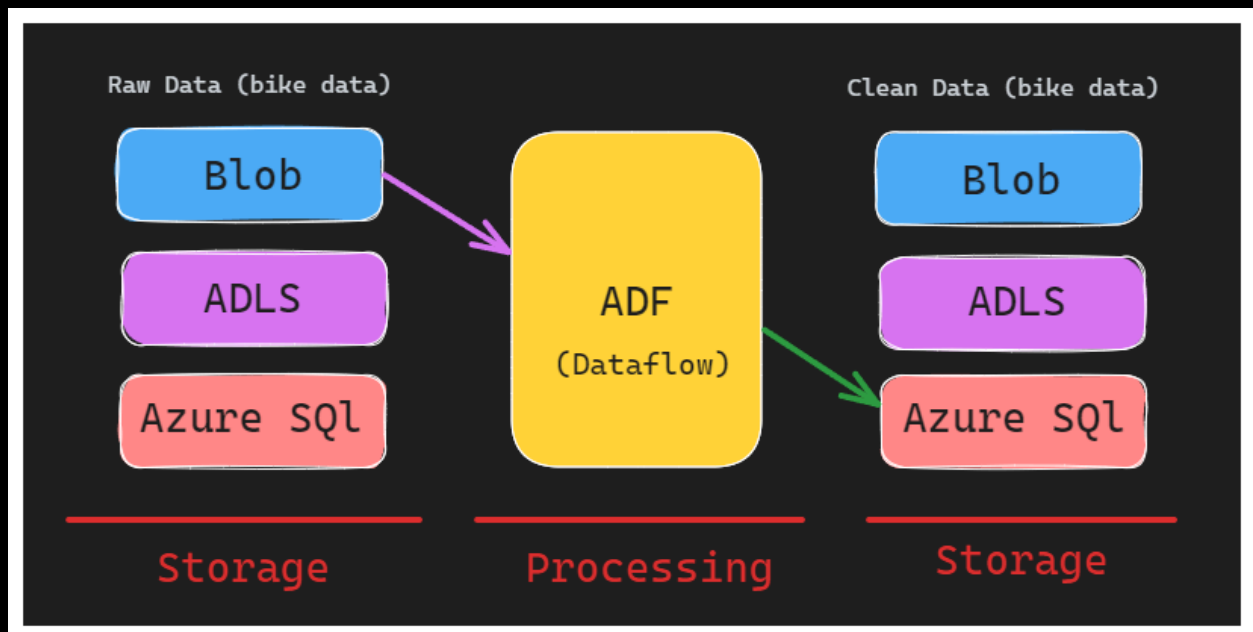
Agenda

Create a data pipeline to transform the bike data from azure blob storage using azure data factory and store the clean data into azure sql database.

Services Required :

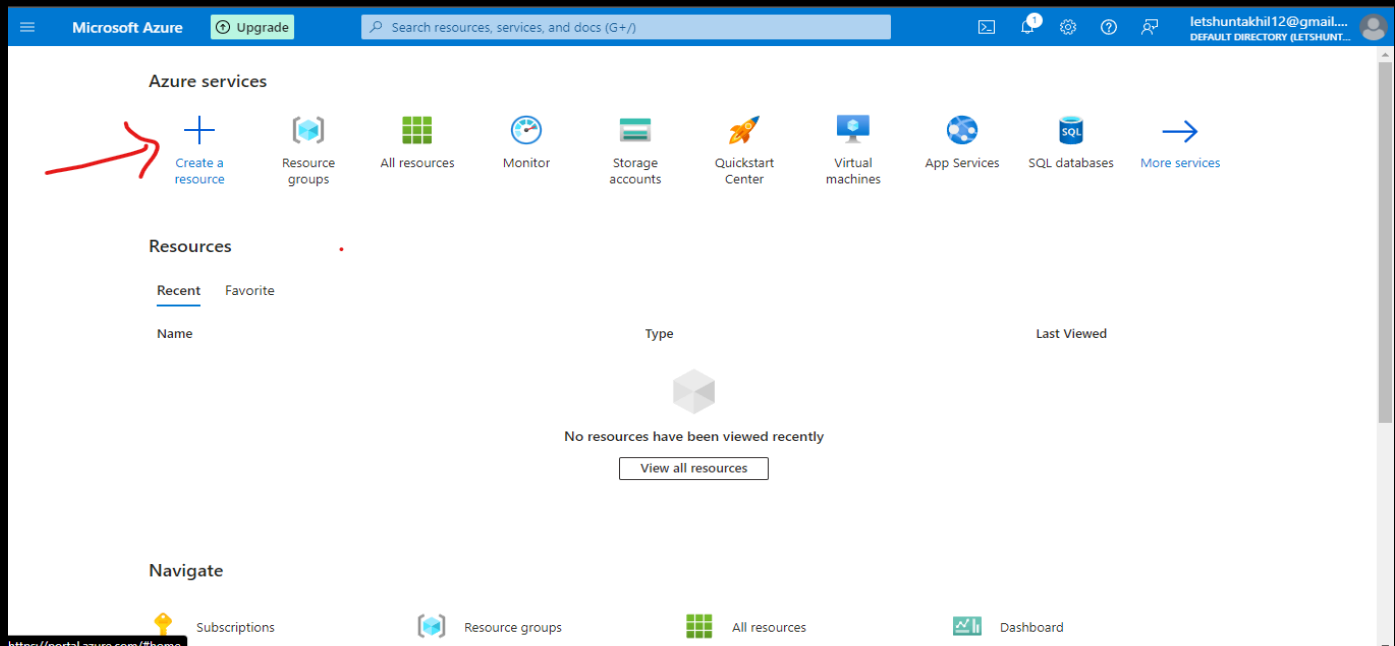
1. Azure Blob Storage
2. Azure Data Factory
3. Azure SQL Database

Pipeline Architecture :

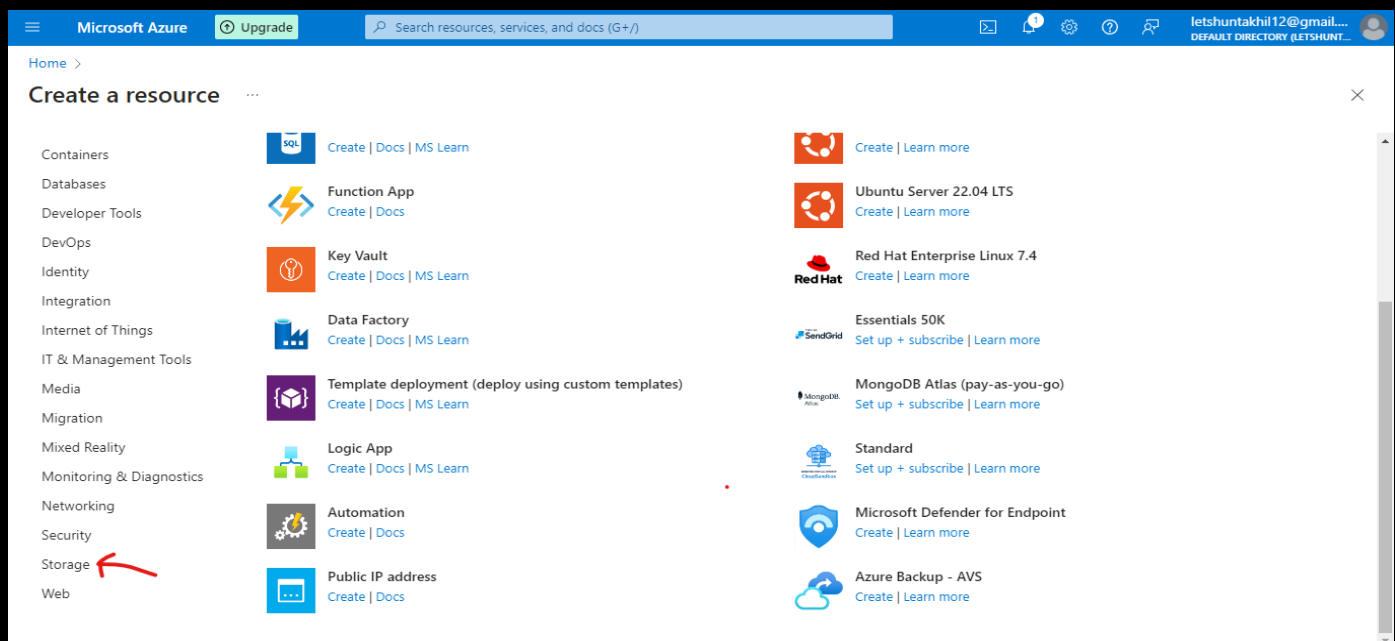


Azure blob storage Creation:

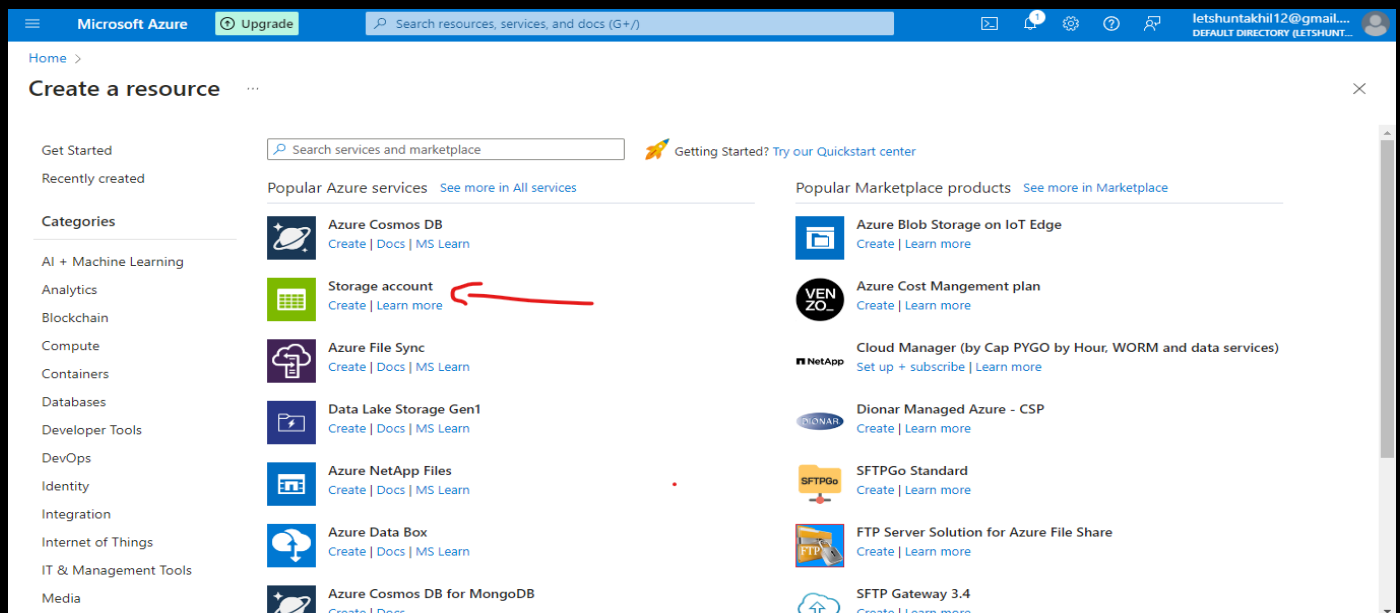
- Login the azure portal portal.azure.com by providing credentials.
- Home page will appear after successful login.
- Click on 'create a resource'.



- Click on Storage & search for the storage account in the azure marketplace.



- Click on the storage account from the list.



- You will get a form to fill up, so that we can make a request to create a storage account.

The screenshot shows the 'Create a storage account' form in the Microsoft Azure portal. The form has six tabs: Basics, Advanced, Networking, Data protection, Encryption, and Tags. The 'Basics' tab is selected. The form contains the following fields:

- Subscription ***: A dropdown menu with 'Free Trial' selected.
- Resource group ***: A dropdown menu with 'Select existing item...' selected. Below it is a link 'Create new'.
- Instance details**:
 - Storage account name ***: A text input field.
 - Region ***: A dropdown menu with '(US) East US' selected. Below it is a link 'Deploy to an Azure Extended Zone'.

At the bottom, there are three buttons: 'Previous', 'Next', and 'Review + create'. A 'Give feedback' link is also present in the bottom right corner.

- In this form, we have six tabs to fill up the fields such as Basics, Advanced, Data Protection, Encryption, Tags and Review+Create.

- In the **Basics Tab** fill the details.
 - **Project details**
 - **Subscription** : Choose ' **Free Trial** ' or ' Pay As You Go ' .
 - Choose Pay as You go if you don't have the free trial
 - **Resource Group** : Create a new resource group if you don't have one.
 - **Instance details**
 - **Storage account name** : provide an account name such that **The field can contain only lowercase letters and numbers. Name must be between 3 and 24 characters and it should be unique.**
 - **Region** : choose a region such that this particular service i.e. azure blob storage can be created. Suppose if we choose East (US) , this service is created in this region.
 - **Performance** :
 - **Standard** : Recommended for most scenarios (general purpose v2 account).
 - **Premium** : Recommended for scenarios that require low latency.
 - **Redundancy (Backup)** : select anyone based on requirement.
 - **Locally redundant Storage (LRS).**
 - **Geo Redundant Storage (GRS).**
 - **Zone Redundant Storage (ZRS).**
 - **Geo-Zone Redundant Storage (GZRS).**

The screenshot shows the 'Create a storage account' wizard in the Microsoft Azure portal. The 'Basics' tab is selected. The 'Project details' section shows 'Subscription' as 'Free Trial' and 'Resource group' as '(New) blob2asq(PQL)rg'. The 'Instance details' section shows 'Storage account name' as 'blob2asqblob', 'Region' as '(US) East US', 'Performance' as 'Standard: Recommended for most scenarios (general-purpose v2 account)', and 'Redundancy' as 'Locally-redundant storage (LRS)'. Navigation buttons at the bottom include 'Review', '< Previous', and 'Next: Advanced >'. A 'Give feedback' link is in the bottom right corner.

- Click next (**Advanced**) .
- In the **advanced** tab :
 - Leave the fields as it is but change the Access tier to Cool.
 - **Access Tier** :
 - **Hot** : Optimized for frequently accessed data and everyday usage scenarios.
 - **Cool** : Optimized for infrequently accessed data and backup scenarios.

The screenshot shows the 'Create a storage account' wizard in the Microsoft Azure portal, now on the 'Advanced' tab. The 'Blob and Data Lake Gen2 endpoints are provisioned by default' message is at the top. Under 'Enable SFTP', there is a checkbox and a note: 'To enable SFTP, 'hierarchical namespace' must be enabled.' Under 'Enable network file system v3', there is a checkbox and a note: 'To enable NFS v3 'hierarchical namespace' must be enabled. Learn more about NFS v3'. In the 'Blob storage' section, 'Allow cross-tenant replication' has a checkbox. The 'Access tier' section has two radio buttons: 'Hot: Optimized for frequently accessed data and everyday usage scenarios' (unselected) and 'Cool: Optimized for infrequently accessed data and backup scenarios' (selected). In the 'Azure Files' section, 'Enable large file shares' has a checkbox. Navigation buttons at the bottom include 'Review', '< Previous', and 'Next: Networking >'. A 'Give feedback' link is in the bottom right corner.

- Click next (**Networking**) .

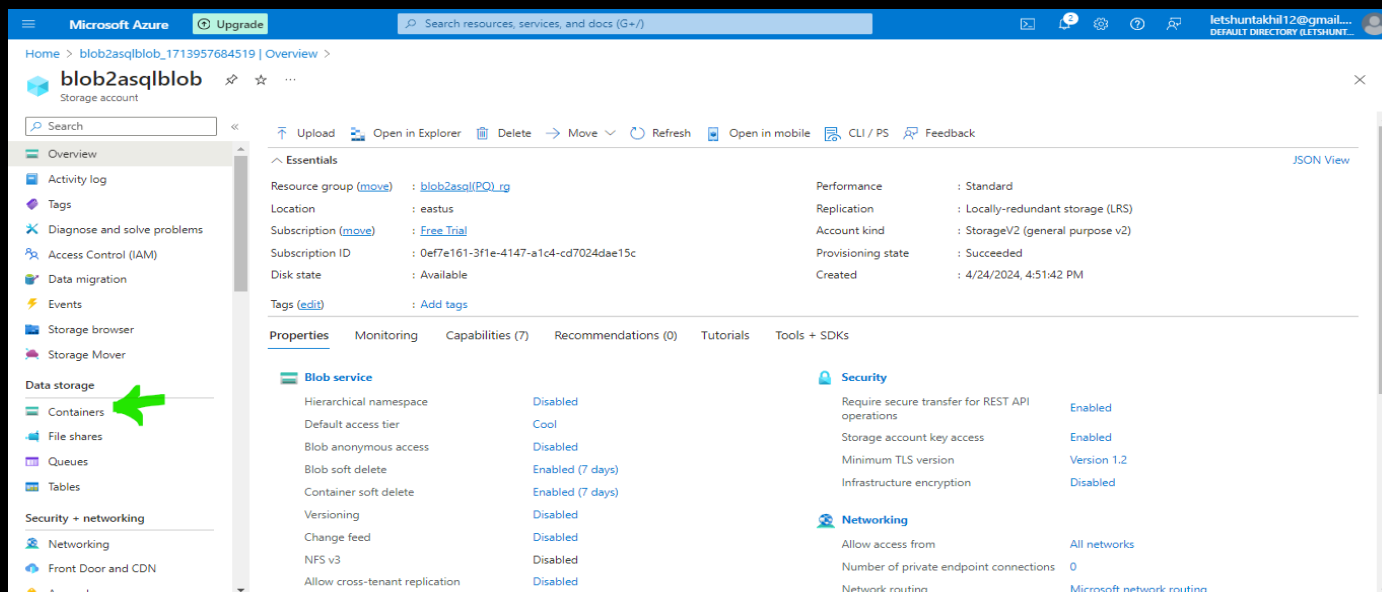
- **Networking** : Keep it default -> click next (**Data protection**) .
- **Data protection** : Keep it default -> click next (**Encryption**) .
- **Encryption** : Keep it default -> click next (**Tags**) .
- **Tags** : Keep it default -> click next (**Review + Create**) .
- In **Review + Create** verify the details and click **create**.

The screenshot shows the 'Create a storage account' page in the Microsoft Azure portal, specifically the 'Review' tab. The page displays configuration details for a new storage account named 'blob2asqblob' in the 'eastus' location. The 'Basics' section includes details like Subscription (Free Trial), Resource Group (blob2asql(PQ)_rg), and Replication (Locally-redundant storage (LRS)). The 'Advanced' section shows various features like 'Enable hierarchical namespace' and 'Enable SFTP' are disabled. At the bottom, there is a 'Create' button highlighted with a green arrow, along with 'Previous' and 'Next' navigation buttons.

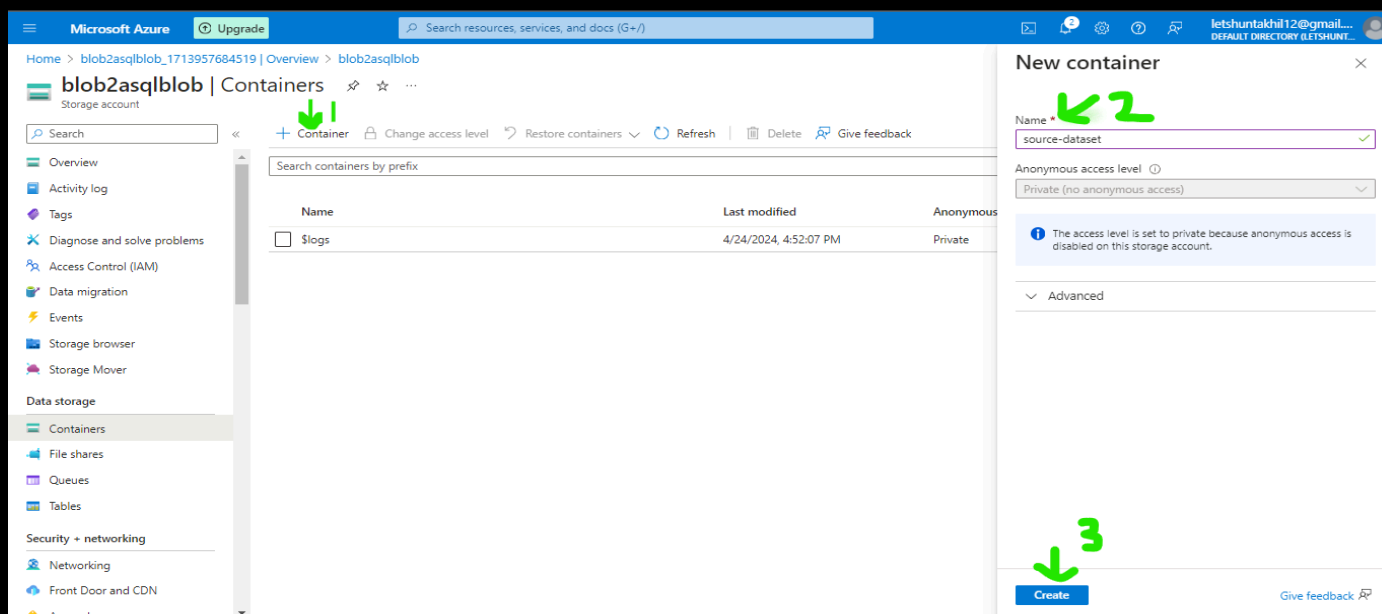
- Azure blob storage has been created and click on ' **Go to Resource** ' .

The screenshot shows the 'Overview' page for the deployment 'blob2asqblob_1713957684519'. A green checkmark indicates 'Your deployment is complete'. The page displays deployment details such as the deployment name, subscription (Free Trial), and resource group (blob2asql(PQ)_rg). The 'Next steps' section includes a 'Go to resource' button, which is highlighted with a green arrow. The right sidebar contains links for 'Cost Management', 'Microsoft Defender for Cloud', and 'Free Microsoft tutorials'.

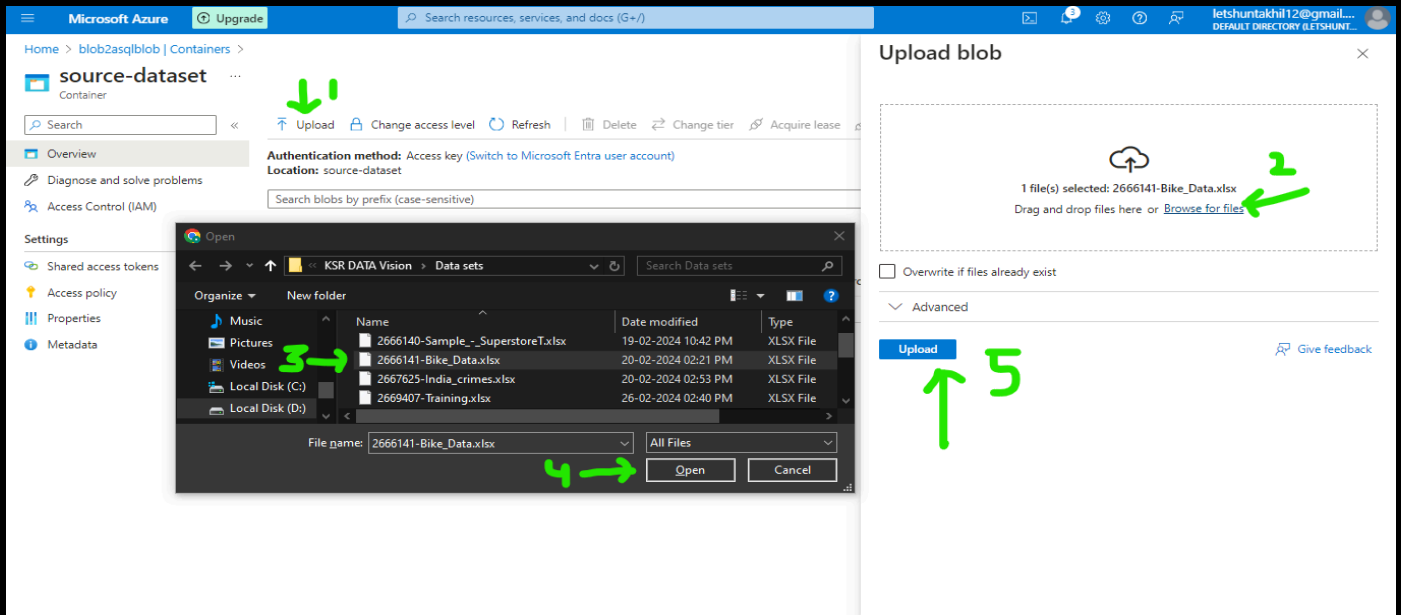
- Click on the 'containers' in the menu list.



- Create a container by clicking '+ Container'.
- Fill the form after clicking '+ Container' and then click 'create'.



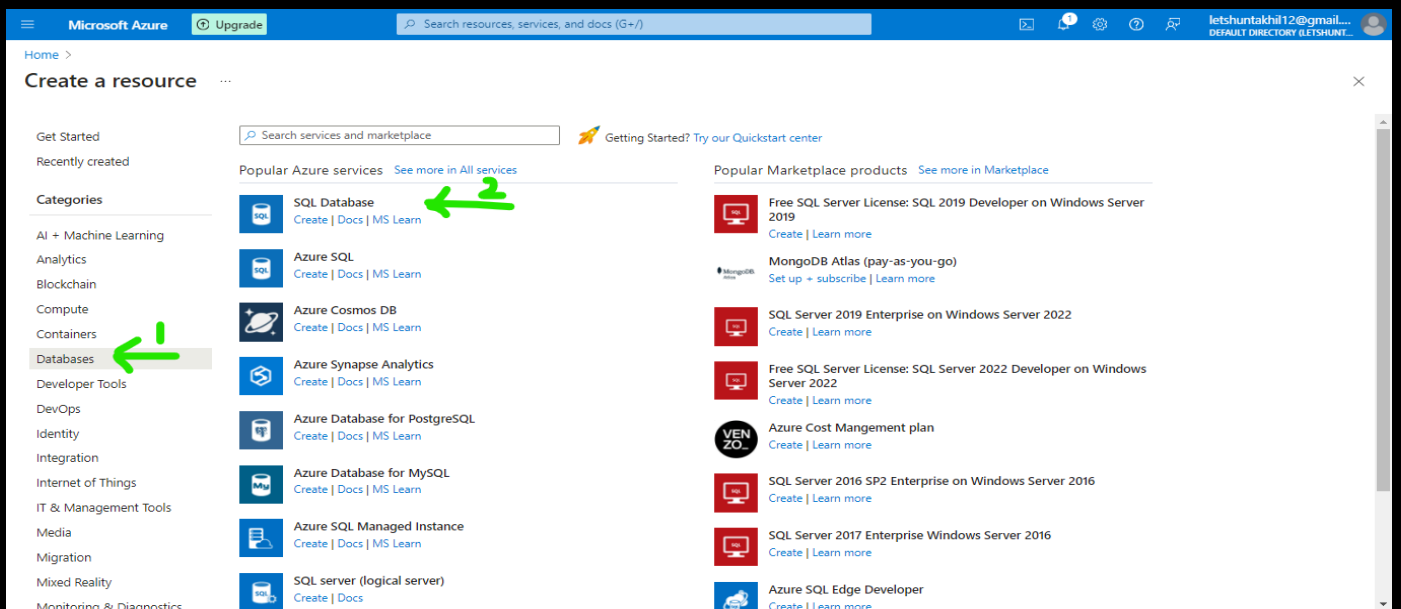
- Click on the container 'source-dataset' after creating.
- Click Upload -> Browse for files -> select the file -> open -> upload.
- The bike data is uploaded.



- Successfully we have created a storage account (blob) , container and uploaded the file.

Azure SQL Database Creation :

- Click on 'Go To Resource' by opening the homepage in the new tab.
- In the menu list , search for 'Databases' and click on it.
- Look for 'SQL Database' in the azure marketplace —> click.



- Fillup the form to request creation of SQL Database .
- In the form we have 6 tabs → **Basics**, **Networking**, **Security**, **Additional settings**, **Tags** and ' **Review + Create** '.
- Basics form :
 - **Project Details**
 - **Subscription** : Free Trial
 - **Resource Group** : Select from the list
 - **Database details**
 - **Database name** : Provide the database name'
 - **Server** : Create a new server if you didn't create it. You need to fill the form when you create a new server.

The screenshot shows the 'Create SQL Database Server' form in the Microsoft Azure portal. The form is titled 'Create SQL Database Server' and includes a sub-header 'Microsoft Azure SQL Database'. Below the title, there is a note: '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.'

The form contains the following fields and sections:

- Server name ***: A text input field containing 'blob2asqlserver'. A green arrow points to this field.
- Location ***: A dropdown menu showing '(US) East US'.
- Authentication**: A section with a blue banner stating 'Azure Active Directory (Azure AD) is now Microsoft Entra ID. Learn more'. Below this, there is a note: '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 Microsoft Entra authentication Learn more or using an existing Microsoft Entra user, group, or application as Microsoft Entra admin Learn more or, or select both SQL and Microsoft Entra authentication.'
- Authentication method**: Three radio buttons are present:
 - ☐ Use Microsoft Entra-only authentication
 - ☐ Use both SQL and Microsoft Entra authentication
 - ☒ Use SQL authentication
 A green arrow points to the 'Use SQL authentication' option.
- Server admin login ***: A text input field containing 'serveradmin'. A green arrow points to this field.
- Password ***: A password input field with a checkmark. A green arrow points to this field.
- Confirm password ***: A password input field with a checkmark. A green arrow points to this field. A green arrow also points to the 'Password and confirm password must match.' message.

At the bottom of the form, there is a blue 'OK' button. A green arrow points to this button. The URL 'https://aka.ms/AADRebrandFAQ' is visible at the bottom left.

- Select the Workload environment as development and configure the compute and storage as per the requirement.
- **Workload environment** : Development
- **Compute + Storage** : DTU → Basic (for less demanding workloads).

Microsoft Azure Upgrade Search resources, services, and docs (G+7) letshuntakhi12@gmail... DEFAULT DIRECTORY (LETSHUNT...)

Create SQL Database

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

Database name * blob2asqldb ✓

Server * (new) blob2asqserver (East US) ✓
[Create new](#)

Want to use SQL elastic pool? ☐ Yes ☒ No

Workload environment ☒ Development ☐ Production

Default settings provided for Development workloads. Configurations can be modified as needed.

Compute + storage * **General Purpose - Serverless**
Standard-series (Gen5), 1 vCore, 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 ☒ Locally-redundant backup storage ☐ Zone-redundant backup storage ☐ Geo-redundant backup storage

[Review + create](#) [Next: Networking >](#)

Microsoft Azure Upgrade Search resources, services, and docs (G+7) letshuntakhi12@gmail... DEFAULT DIRECTORY (LETSHUNT...)

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](#) ⓘ

SQL Database Hyperscale: Low price, high scalability, and best feature set. [Learn more](#) ⓘ

Service tier **General Purpose (Most budget friendly)** ✓

Compute tier **vCore-based purchasing model**
General Purpose (Most budget friendly)
Hyperscale (Highly scalable compute and storage) ⓘ
Business Critical (Highest availability and performance) ⓘ

Compute Hardware **DTU-based purchasing model**
Basic (For less demanding workloads) ✓
Standard (Budget friendly)
Premium (Highest availability and performance) ⓘ
up to 80 vCores, up to 240 GB memory
[Change configuration](#)

Hardware Configuration

Max vCores 1

Min vCores 0.5 vCores

[Apply](#)

Cost summary

General Purpose (GP 5 Gen5 1)
Cost per GB (in INR) 9.57
Max storage selected (in GB) x 41.6

ESTIMATED STORAGE COST / MONTH 398.00 INR
COMPUTE COST / VCORE SECOND 0.012058 INR

NOTES
1. Serverless databases are billed in vCore seconds based on a combination of CPU and memory utilization. [Learn more](#) about serverless billing.

Microsoft Azure Upgrade Search resources, services, and docs (G+7) letshuntakhi12@gmail... DEFAULT DIRECTORY (LETSHUNT...)

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](#) ⓘ

SQL Database Hyperscale: Low price, high scalability, and best feature set. [Learn more](#) ⓘ

Service tier **Basic (For less demanding workloads)**

DTUs [Compare DTU options](#) ⓘ

5 (Basic)

Data max size (GB) 2

[Apply](#)

Cost summary

Basic (Basic)
Cost per DTU (in INR) 81.49
DTUs selected x 5

ESTIMATED COST / MONTH 407.46 INR

<https://portal.azure.com/#>

- Backup storage redundancy
 - Backup storage redundancy : Locally redundant backup storage
- Click next (Networking).

The screenshot shows the 'Create SQL Database' wizard in the Microsoft Azure portal. The 'Backup storage redundancy' section is expanded, showing three options: 'Locally-redundant backup storage' (selected), 'Zone-redundant backup storage', and 'Geo-redundant backup storage'. A green arrow points to the 'Locally-redundant backup storage' option. Another green arrow points to the 'Review + create' button at the bottom left of the form.

- Networking form :
 - Network connectivity
 - Connectivity method : public endpoint
 - Firewall Rules
 - Allow azure services and resources to access this server : Yes
 - Add current client ip address : Yes
 - Connection Policy
 - Connection policy : default
 - Encrypted Connections
 - Minimum TLS version : TLS 1.2
- Click on next (Security)

Home > Create a resource > Upgrade

Search resources, services, and docs (G+V)

Microsoft Azure

Create SQL Database

Microsoft

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 *

☐ No ☒ Yes

Add current client IP address *

☐ No ☒ Yes

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 (except Private Endpoint connections) 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

[Review & create](#) [< Previous](#) [Next > Security](#)

- **Security form** : Keep it default → click next (**Additional settings**).

Home > Create a resource > Upgrade

Search resources, services, and docs (G+V)

Microsoft Azure

Create SQL Database

Microsoft

Microsoft Defender for SQL

Protect your data using Microsoft Defender for SQL, a unified security package including vulnerability assessment and advanced threat protection for your server. [Learn more](#)

Get started with a 30 day free trial period, and then 1247.9202 INR/server/month.

Enable Microsoft Defender for SQL *

- ☐ Start free trial
- ☒ Not now

Ledger

Ledger cryptographically verifies the integrity of your data and detects any tampering that might have occurred. [Learn more](#)

Ledger

Not configured
[Configure ledger](#)

Server identity

Use system assigned and user assigned managed identities to enable central access management between this database and other Azure resources. [Learn more](#)

Server identity

Not enabled
[Configure identities](#)

Transparent data encryption key management

Transparent data encryption encrypts your databases, backups, and logs at rest without any changes to your application. To enable encryption, go to each database. Database level settings, if enabled, will override the server level setting. [Learn more](#)

Server level key

☐ Service-managed key selected
[Configure transparent data encryption](#)

[Review & create](#) [< Previous](#) [Next > Additional settings](#)

- **Additional Settings form** : Keep it default → click next (**Tags**).

Home > Create a resource > Upgrade

Search resources, services, and docs (G+V)

Microsoft Azure

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

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 *

[Find a collation](#)

Maintenance window

Select a preferred maintenance window from the drop-down. During maintenance, databases remain available, but some updates may require a failover. The system default maintenance window (5pm to 8am) limits most activities to this time, but urgent updates may occur outside of it. To ensure all updates occur only during the maintenance window, select a non-default option. [Learn more](#)

Maintenance window

[Review & create](#) [< Previous](#) [Next > Tags](#)

- **Tags form** : Keep it default → click next (**Review + Create**).

Microsoft Azure | Upgrade

Home > Create a resource >

Create SQL Database

Microsoft

Basics Networking Security Additional settings Tags Review + create

Tags are name/value pairs that enable you to categorize and view consolidated billing by applying the same tag to multiple resources and resource groups. [Learn more](#)

Note that if you create tags and then change resource settings on other tabs, your tags will be automatically updated.

Name	Value	Resource
		2 selected

Cost summary

Basic (Basic)	
Cost per DTU (in INR)	81.49
DTUs selected	x 5
ESTIMATED COST / MONTH	407.46 INR

Review + create < Previous Next: Review + create >

- Review the form and click '**create**'.

Microsoft Azure | Upgrade

Home > Create a resource >

Create SQL Database

Microsoft

Basics Networking Security Additional settings Tags Review + create

Product details

SQL database by Microsoft

Estimated cost per month: 407.46 INR

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	Free Trial
Resource group	blob2asq(PQULrg)
Region	East US
Database name	blob2asqfdb
Server	(new) blob2asqserver
Authentication method	SQL authentication
Server admin login	serveradmin
Compute + storage	Basic: 2 GB storage
Backup storage redundancy	Locally-redundant backup storage

Networking

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

- After creating, wait for sometime to deploy and then click '**Go To Resource**'.

Microsoft Azure | Upgrade

Home >

Microsoft.SQLDatabase.newDatabaseNewServer_54b3868e18d5461eac3d | Overview

Deployment

Search Delete Cancel Redeploy Download Refresh

Overview

✓ Your deployment is complete

Deployment name : Microsoft.SQLDatabase.newDatabaseNewServer_54b3868e18d5461eac3d Start time : 4/24/2024, 8:25:47 PM
 Subscription : Free Trial Correlation ID : fb16e9e2-34c8-4a18-936d-acdbbd121f3e
 Resource group : blob2asq(PQULrg)

> Deployment details

▼ Next steps

Go to resource

Give feedback
 Tell us about your experience with deployment

Cost management

Get notified to stay within your budget and prevent unexpected charges on your bill. [Set up cost alerts >](#)

Microsoft Defender for Cloud

Secure your apps and infrastructure [Go to Microsoft Defender for Cloud >](#)

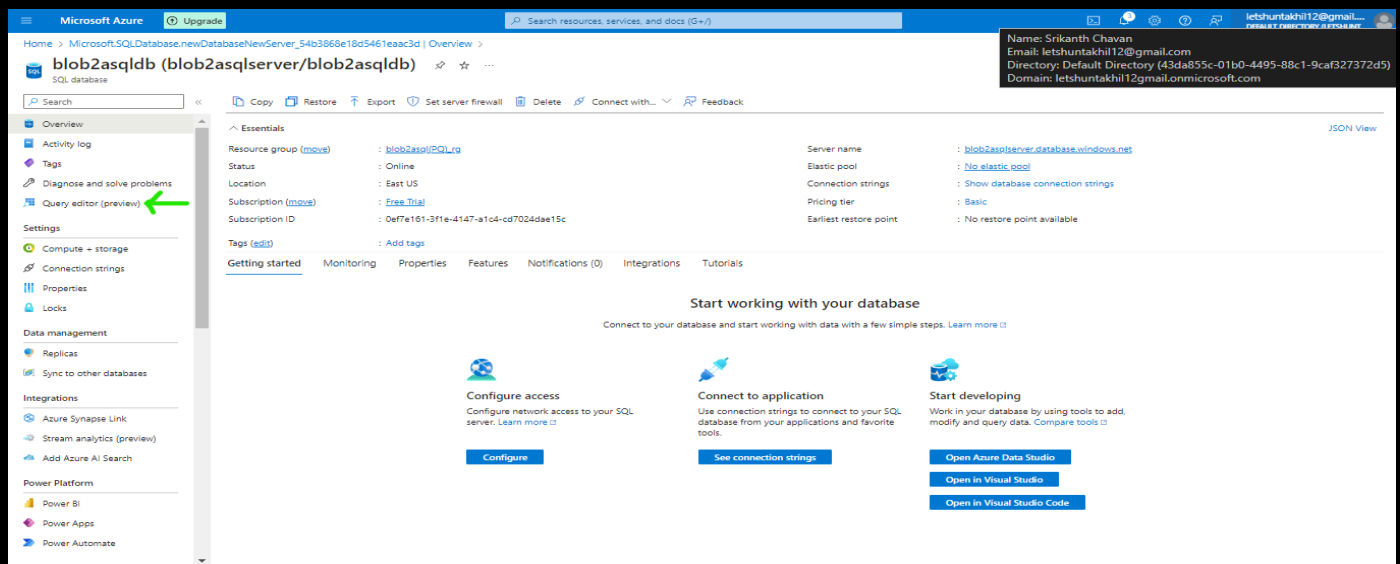
Free Microsoft tutorials

[Start learning today >](#)

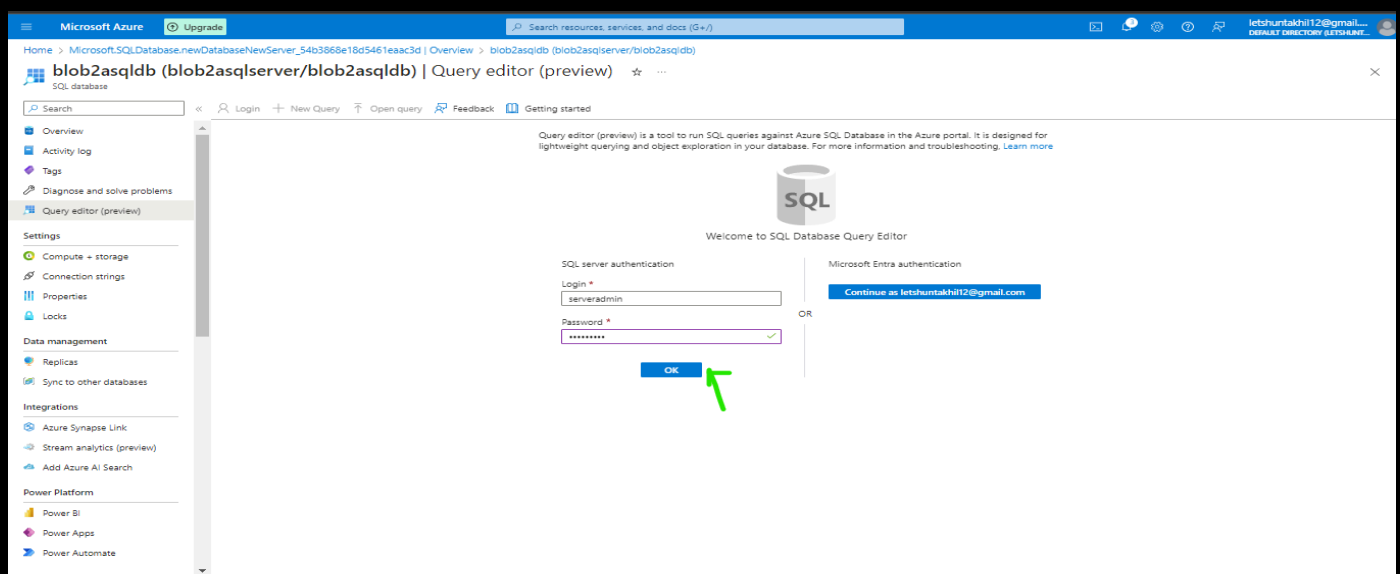
Work with an expert

Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support. [Find an Azure expert >](#)

- Click the **Query editor** in the menu to look for the tables and query the table .



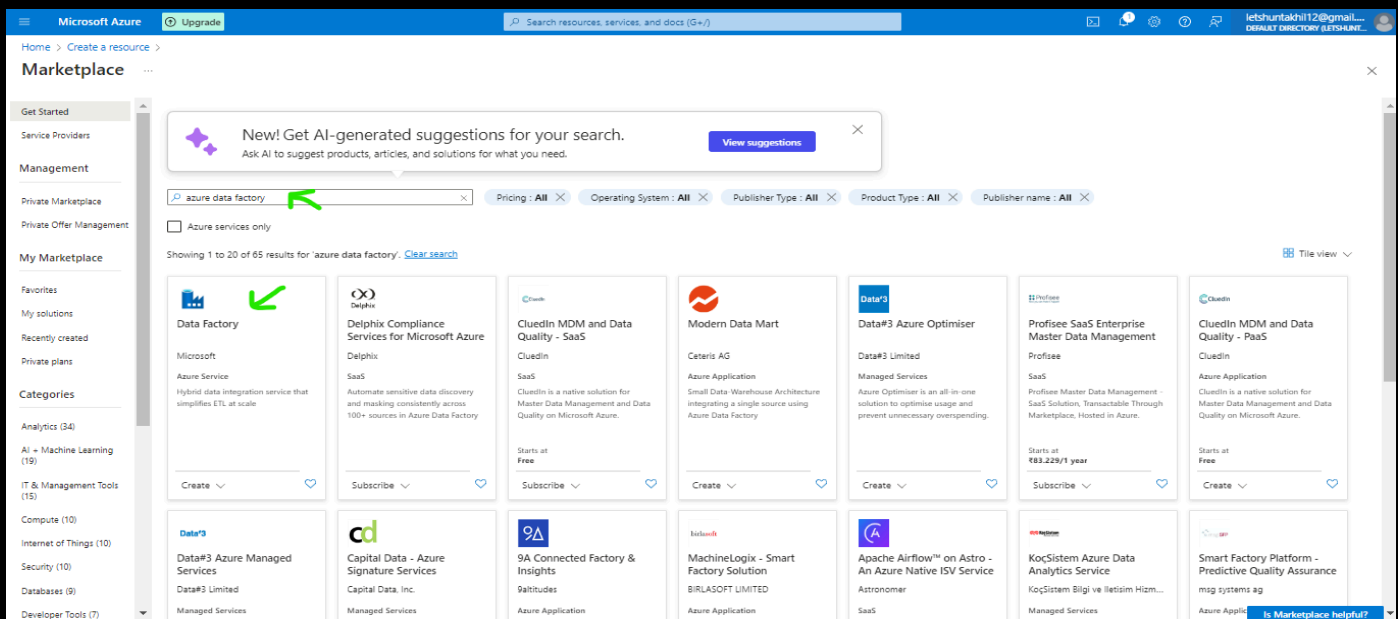
- Login to the database by providing credentials (details provided in the form) .



- Successfully we have created the azure sql database.

Azure Data Factory Creation :

- Click on 'Go To Resource' by opening the homepage in the new tab.
- In the menu list, search for **Azure data factory** and click enter.
- Look for **Data factory** in the azure marketplace and click on it.



- Fill the form to create an azure data factory.
- In the form we have 6 tabs **Basics**, **Git configuration**, **Networking**, **Advanced**, **Tags** and **Review + Create**.
- **Basics** form :
 - **Project details** :
 - **Subscription** : Select from the list .
 - **Resource group** : Select from the list .
 - **Instance details** :
 - **Name**: Select from the list.
 - **Region** : East US.
 - **Version**: V2.

- Click next (**Git Configuration**) .

Microsoft Azure

Home > Create a resource > Marketplace > Data Factory >

Create Data Factory

Basics Git configuration Networking Advanced Tags Review + create

One-click to create data factory with sample pipeline and datasets. [Try it](#)

Project details

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

Subscription *

Resource group *
[Create new](#)

Instance details

Name *

Region *

Version *

[Previous](#) [Next](#) [Review + create](#)

[Give feedback](#)

- Click next (**Networking**) .

Microsoft Azure

Home > Create a resource > Marketplace > Data Factory >

Create Data Factory

Basics Git configuration Networking Advanced Tags Review + create

Managed virtual network

Choose whether you want the default AutoResolveIntegrationRuntime to be provisioned on demand inside an ADF-managed virtual network. If this setting is disabled, after the data factory is created, you can still choose whether to provision explicitly created Azure integration runtime inside an ADF-managed virtual network. [Learn more](#)

Enable Managed Virtual Network on the default AutoResolveIntegrationRuntime ☐

Self-hosted integration runtime inbound connectivity to Azure Data Factory service

Choose whether to connect your self-hosted integration runtime to Azure Data Factory via public endpoint or private endpoint. This applies to self-hosted integration runtime running either on premises or inside customer managed Azure virtual network. [Learn more](#)

Connect via * ☒ Public endpoint ☐ Private endpoint

? You can change this or configure another connectivity method after this resource is created. [Learn more](#)

[Previous](#) [Next](#) [Review + create](#)

[Give feedback](#)

- Click next (**Advanced**) .

Microsoft Azure

Home > Create a resource > Marketplace > Data Factory >

Create Data Factory

Basics Git configuration Networking Advanced Tags Review + create

Managed virtual network

Choose whether you want the default AutoResolveIntegrationRuntime to be provisioned on demand inside an ADF-managed virtual network. If this setting is disabled, after the data factory is created, you can still choose whether to provision explicitly created Azure integration runtime inside an ADF-managed virtual network. [Learn more](#)

Enable Managed Virtual Network on the default AutoResolveIntegrationRuntime ☐

Self-hosted integration runtime inbound connectivity to Azure Data Factory service

Choose whether to connect your self-hosted integration runtime to Azure Data Factory via public endpoint or private endpoint. This applies to self-hosted integration runtime running either on premises or inside customer managed Azure virtual network. [Learn more](#)

Connect via * ☒ Public endpoint ☐ Private endpoint

? You can change this or configure another connectivity method after this resource is created. [Learn more](#)

[Previous](#) [Next](#) [Review + create](#)

[Give feedback](#)

- Click next (Tags).

Microsoft Azure

Home > Create a resource > Marketplace > Data Factory >

Create Data Factory

Basics Git configuration Networking **Advanced** Tags Review + create

Datafactory Encryption

By default, data is encrypted with Microsoft-managed keys. For additional control over encryption keys, you can supply customer-managed keys to use for encryption of blob and file data. Customer-managed keys must be stored in an Azure Key Vault. You can either create your own keys and store them in a key vault, or you can use the Azure Key Vault APIs to generate keys. The storage account and the key vault must be in the same region, but they can be in different subscriptions.

Enable encryption using a Customer Managed Key ☐

Previous **Next** Review + create

Give feedback

- Click next (Review + Create).

Microsoft Azure

Home > Create a resource > Marketplace > Data Factory >

Create Data Factory

Basics Git configuration Networking Advanced **Tags** Review + create

Tags are name/value pairs that enable you to categorize resources and view consolidated billing by applying the same tag to multiple resources and resource groups. [Learn more about tags](#)

Note that if you create tags and then change resource settings on other tabs, your tags will be automatically updated.

Name	Value	Resource
	:	Data factory (V2)

Previous **Next** Review + create

Give feedback

- Review the form and click create.

Microsoft Azure

Home > Create a resource > Marketplace > Data Factory >

Create Data Factory

Basics Git configuration Networking Advanced Tags **Review + create**

[View automation template](#)

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. See the [Azure Marketplace Terms](#) for additional details.

Basics

Subscription	Free Trial
Resource group	blob2asq(POL)_rg
Name	blobb2asqf-adf
Region	East US
Version	V2

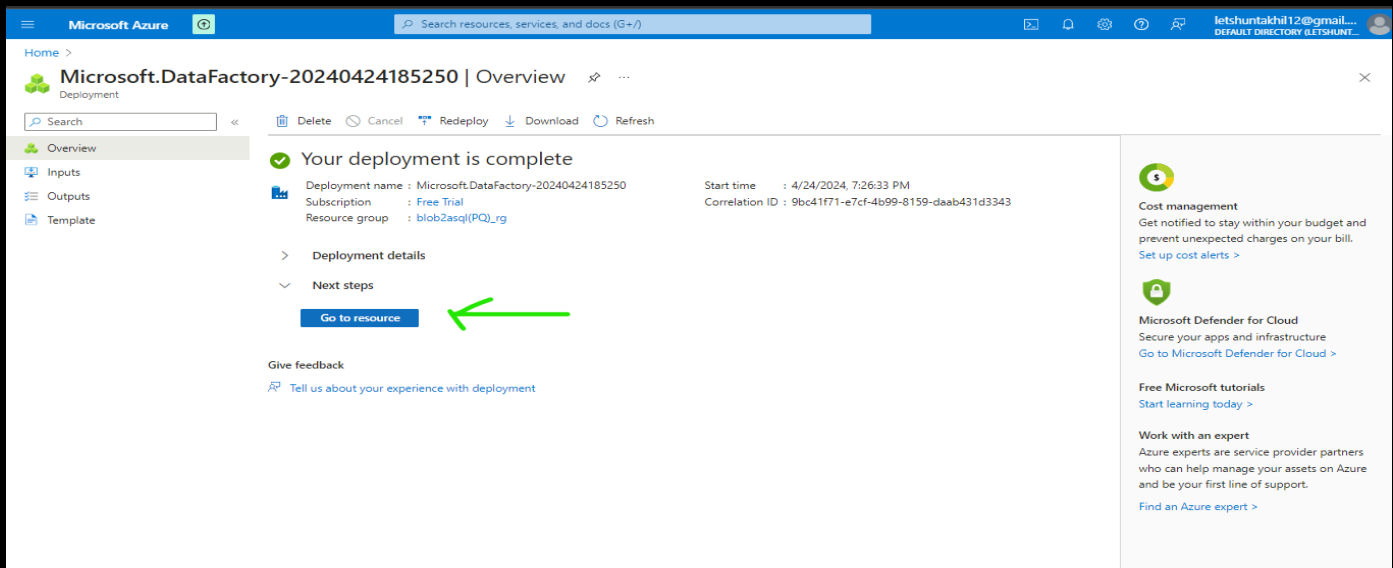
Networking

Connect via Public endpoint

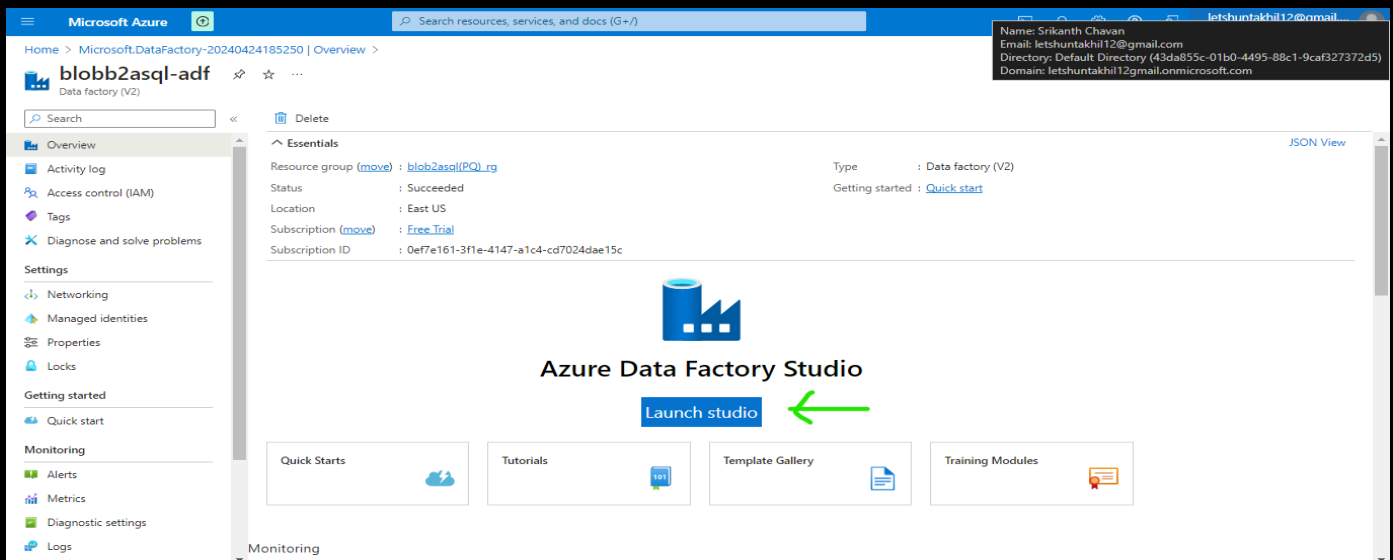
Previous **Next** **Create**

Give feedback


- After successful creation and deployment of azure data factory, click **Go to Resource**.

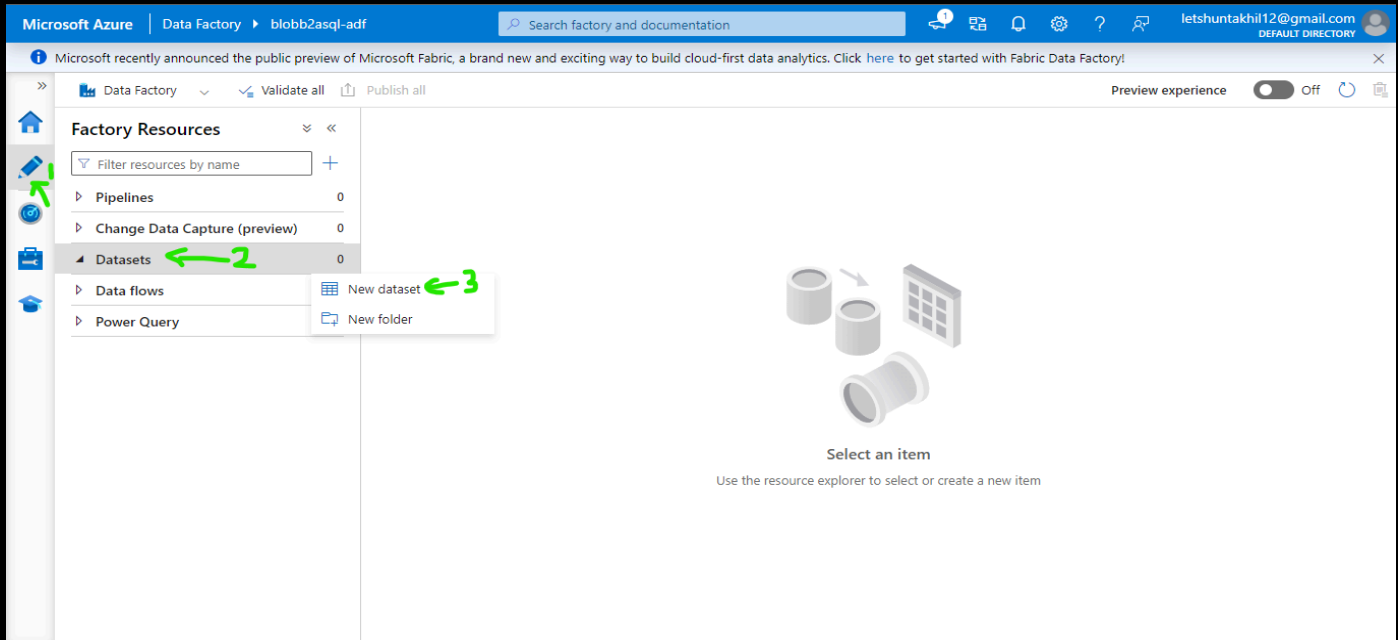


- Let's transform the bike data by clicking the **launch studio**.

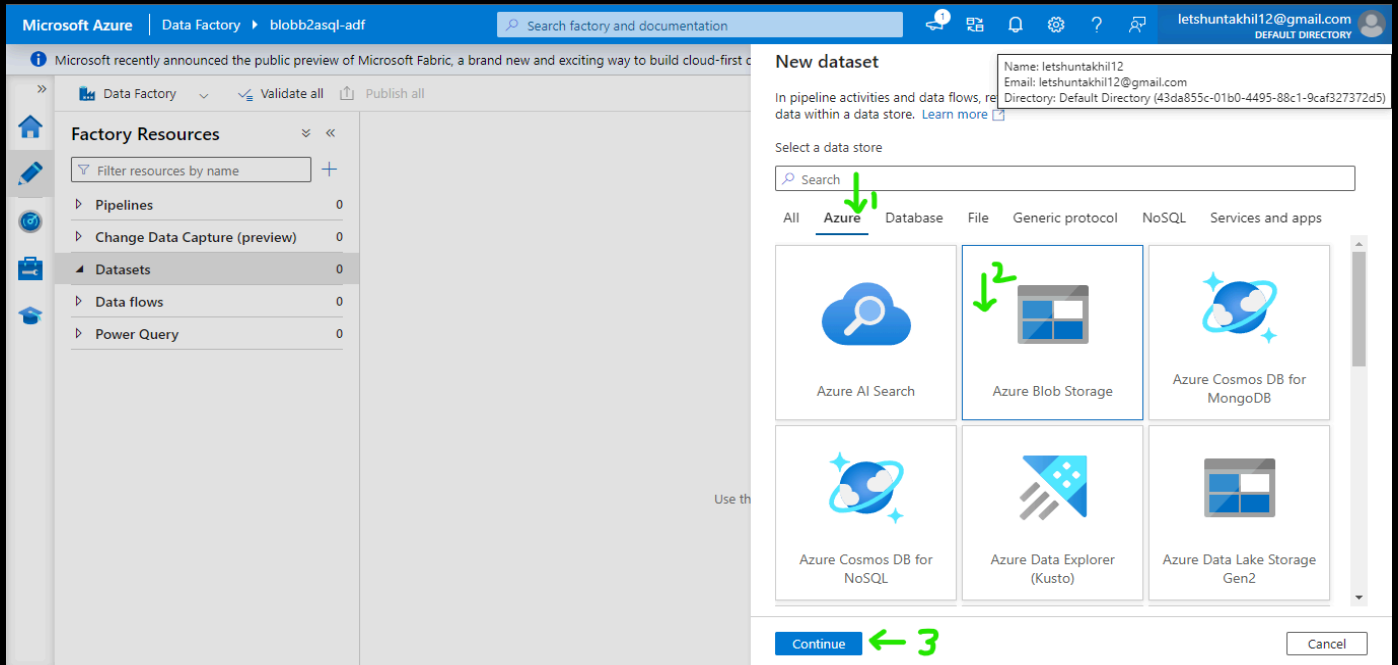


Transformation of data using Data Flow:

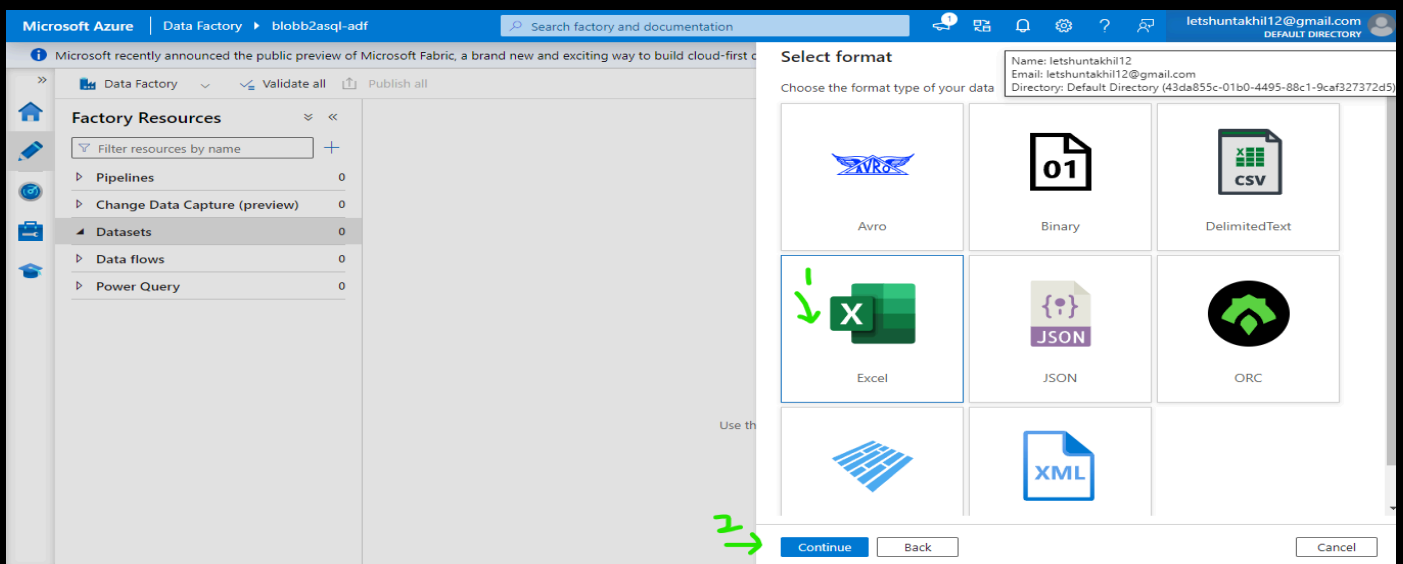
- Azure data factory studio will open in a new tab.
- Create a dataset (path) for the azure blob storage (source data) by clicking the **pencil icon** () → datasets → new datasets.



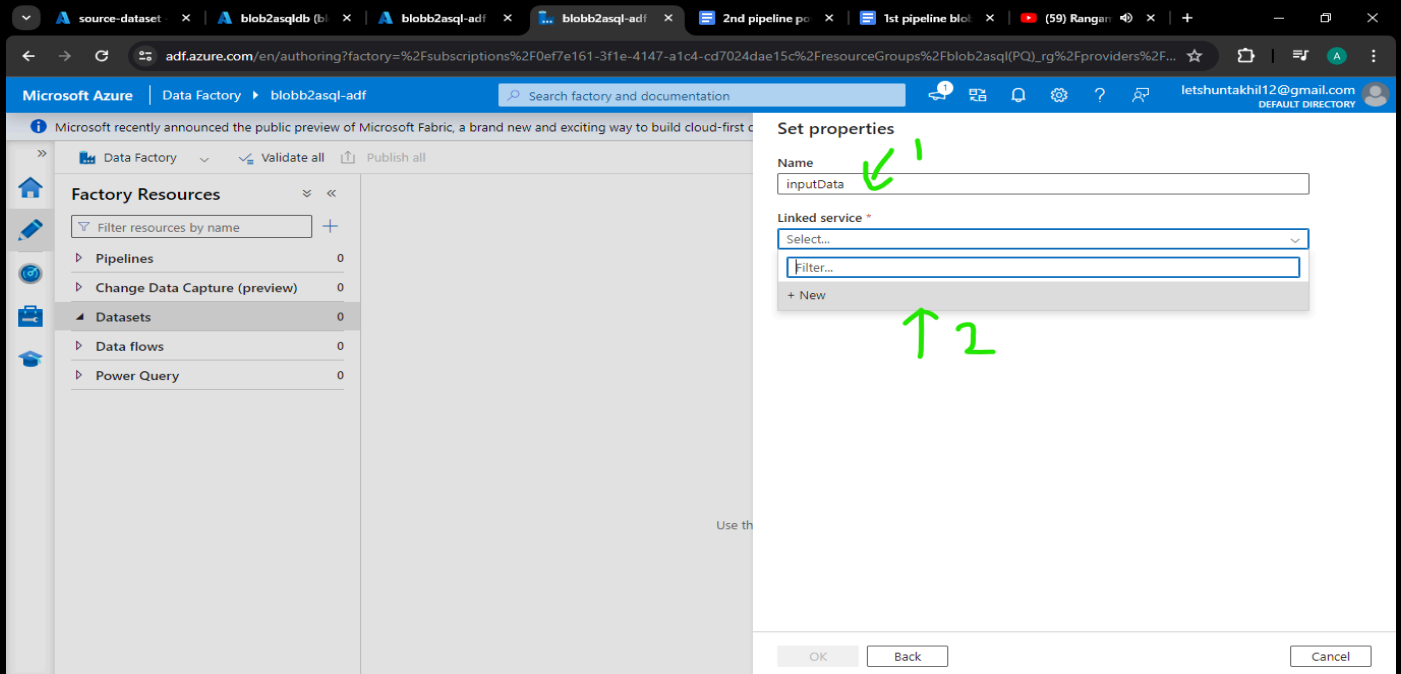
- Select the **azure blob storage** in the azure tab and then click **continue**.



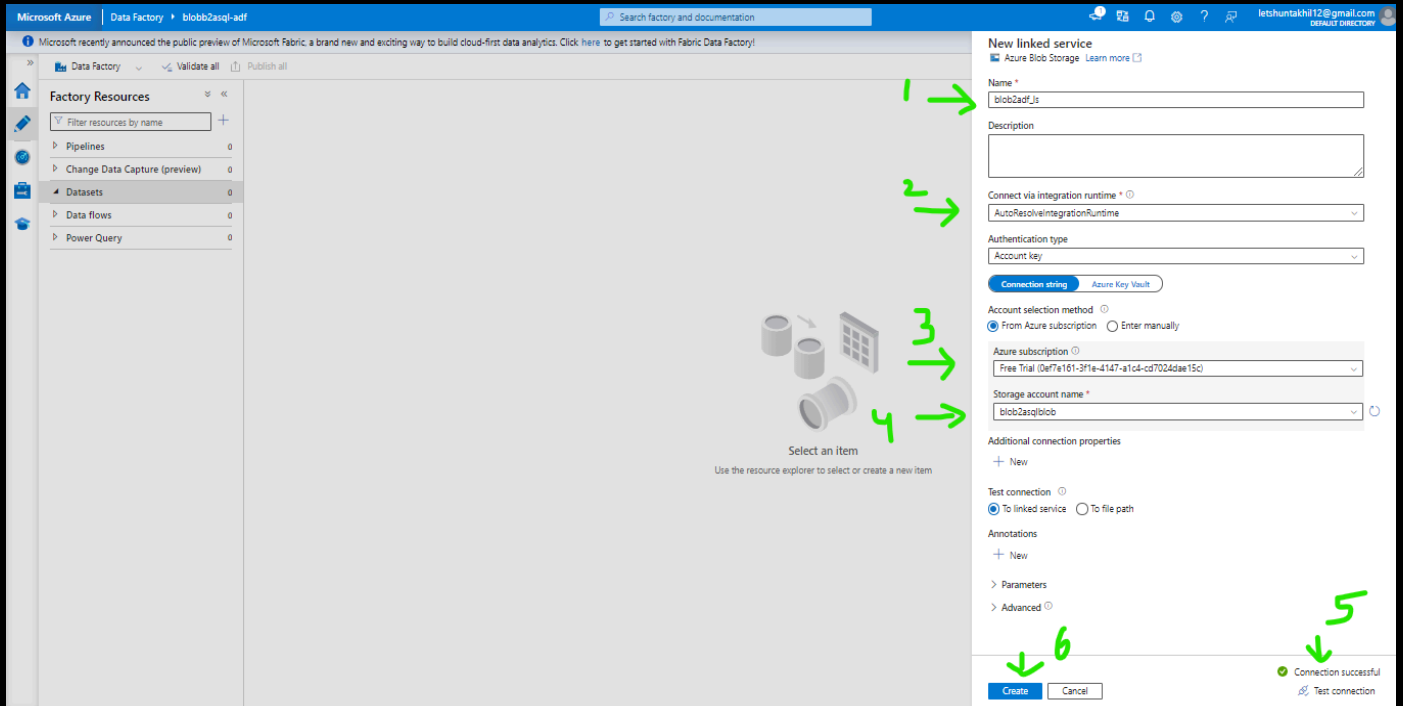
- Select the format (**Excel**) of the data to transform which is in the source dataset (azure blob storage) and then click **continue**.




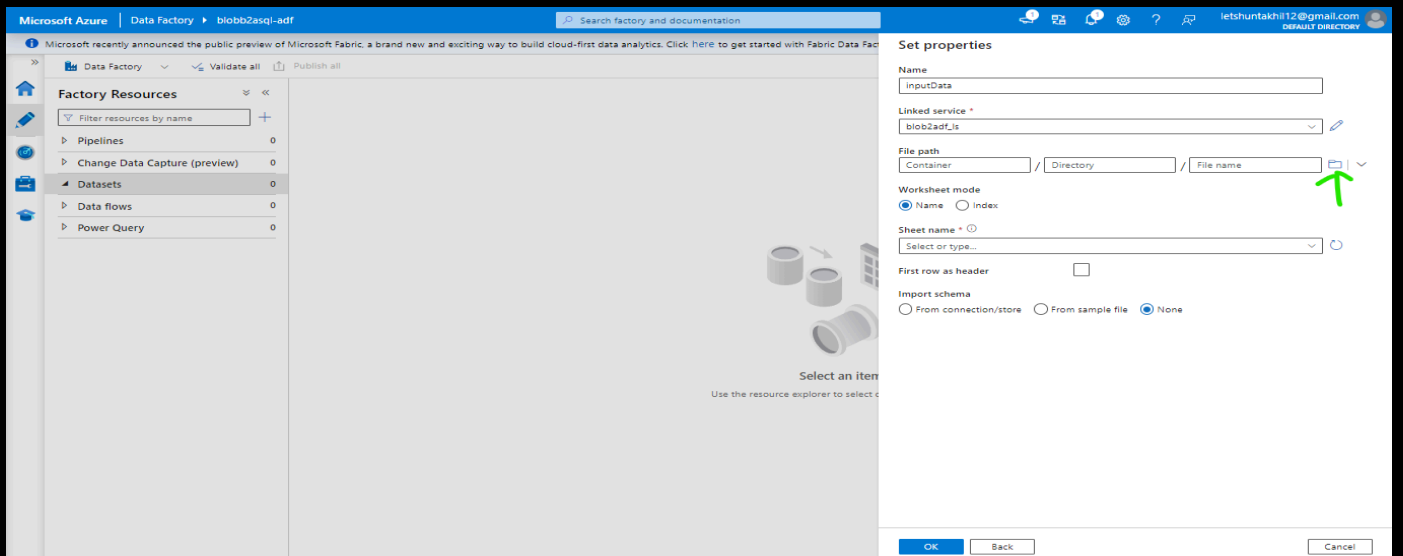
- To make a connection between blob storage and azure data factory we need a linked service to connect or make a bridge between them. So that communication can happen.
- The next step is to choose or create a linked service .

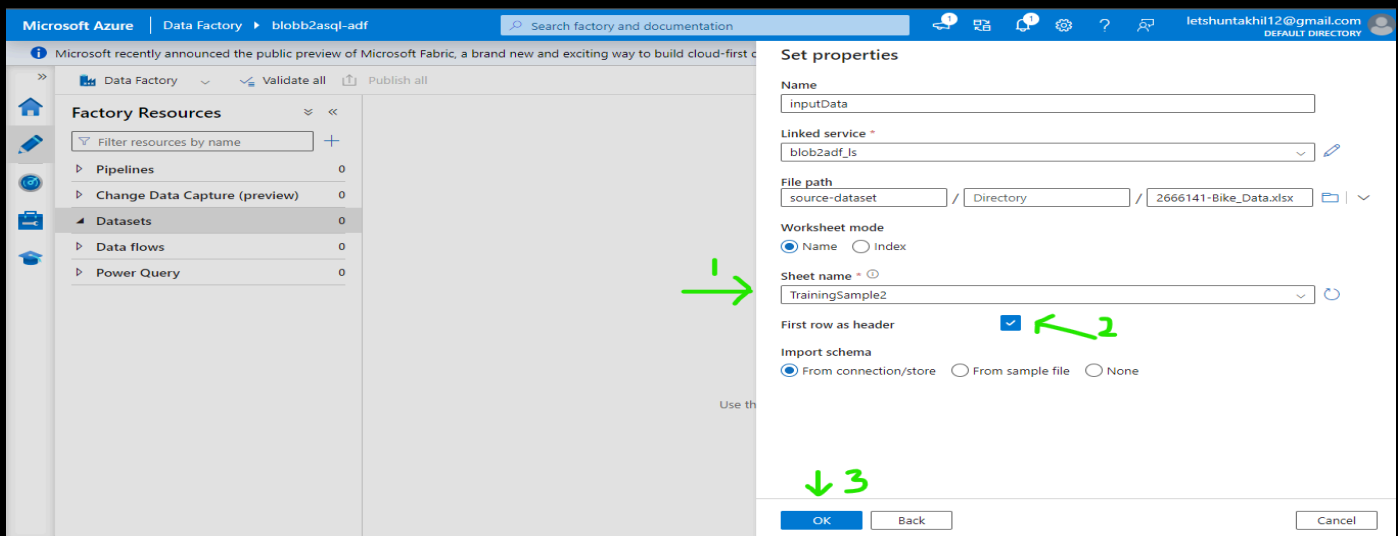
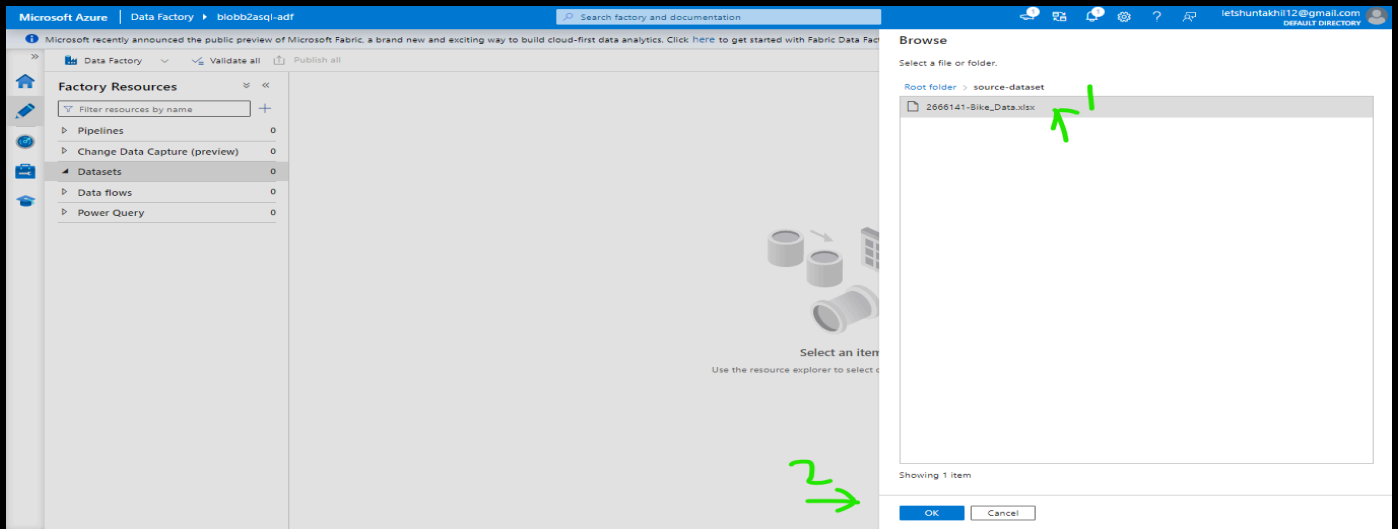


- Fill the form to create a linked service.
- In the form we have properties to fill in **name**, **description**, **connect via integration runtime**, **authentication type**, **subscription**, **storage account name** and some other properties.
- **Connect via integration runtime** → integration runtime is responsible for creating a connection between different environments.
 - **Self Integration runtime** : It is for the on premises environment.
 - **AutoResolveIntegrationRuntime** : It is for within the environment (azure to azure).
- **Azure subscription** : select the subscriptions where the file is located.
- **Storage account name** : select the storage account where the file is located.
- **Test connection** → **successful** → **Create**.

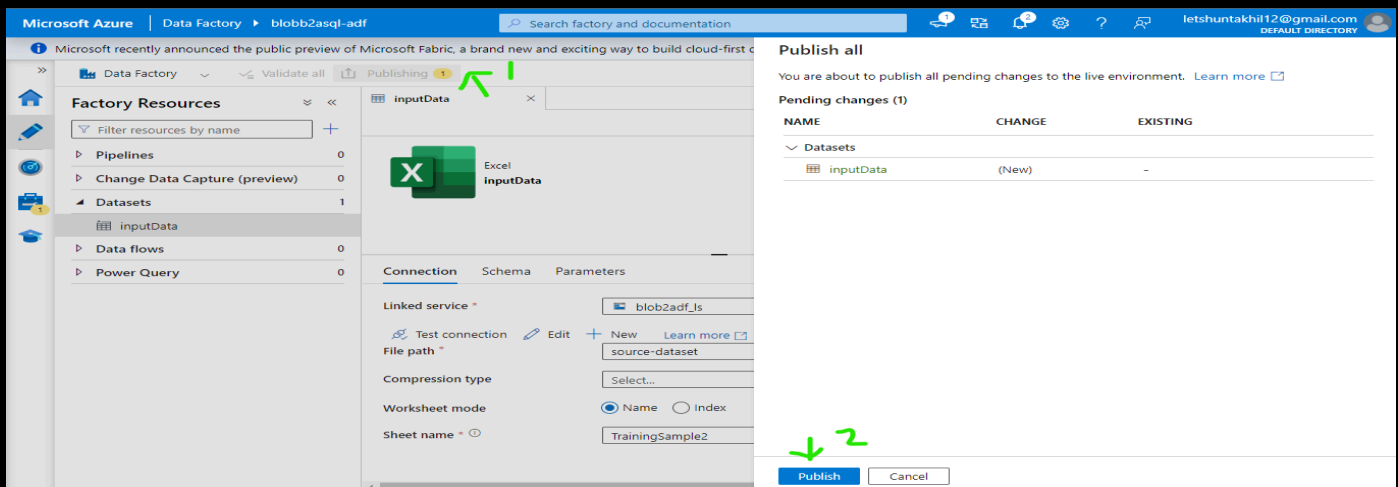



- Choose the file path by clicking the **file icon** () → **select the first row as header** → **select the sheet name** → **ok**.

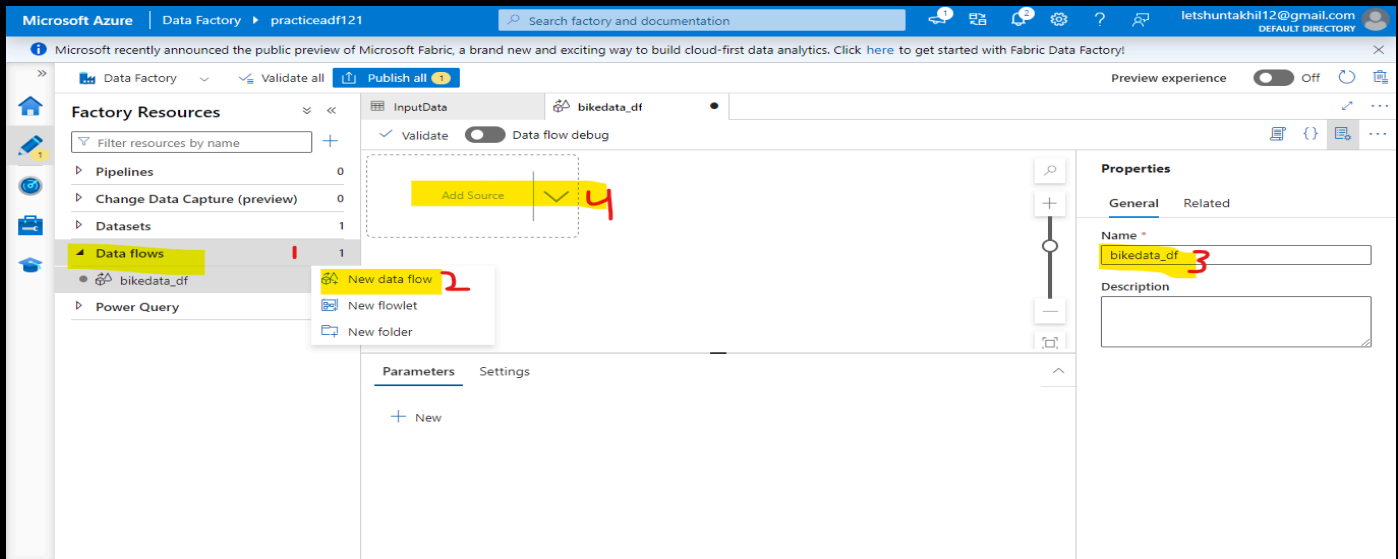




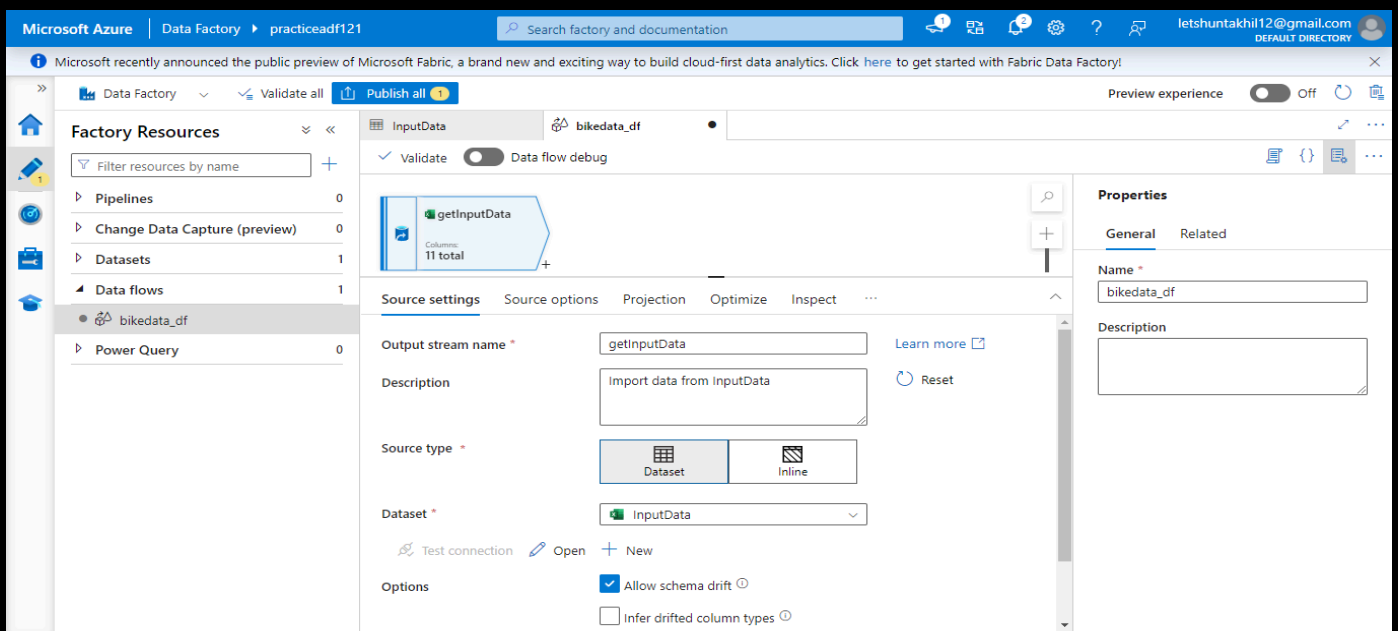
- Save the dataset by clicking **publish all** → **publish**.



- Successfully brought bike data to the data factory to transform.
- Create a Dataflow → **new Data Flow (bikeData_df)** which is in the **pencil icon ()** menu.
- Click **add source** in the workspace.



- You will get a form to provide the details of source data.
- **bikeData_df → Settings → Dataset → select the file (inputData) .**



- Click on the **data preview** .

- It will ask you to turn **on** the Data flow debug.
- Click **ok**.
- **Projection** → **import projection**

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' pane lists 'Data flows' with 'bikedata_df' selected. The main canvas displays a data flow named 'bikedata_df' with a 'getInputData' activity. The 'Projection' tab is active, showing a table of columns to be projected. The 'Data flow debug' toggle is turned on. The 'Properties' pane on the right shows the 'General' tab with the name 'bikedata_df' and a description field.

Column name	Type	Format
Region	abc string	Specify format
Country	abc string	Specify format
Customer	abc string	Specify format
Business Segment	abc string	Specify format
Category	abc string	Specify format
Model	abc string	Specify format
Color	abc string	Specify format
SalesDate	abc string	Specify format

- **Data preview** by clicking **refresh**

The screenshot shows the same Microsoft Azure Data Factory interface as the previous one, but with a context menu open over the 'Derived Column' option in the 'Projection' tab. The 'Data flow debug' toggle is still on. The 'Properties' pane on the right remains the same. The context menu includes options like 'Multiple inputs/outputs', 'Join', 'Conditional Split', 'Exists', 'Union', 'Lookup', 'Schema modifier', 'Derived Column' (highlighted with a red '2'), 'Select', 'Aggregate', 'Surrogate Key', and 'Pivot'.

- Click on the **+** button to create a **Derived column** (new column)
- Create a new column cost and sales by multiplying $\text{UnitPrice} \times \text{OrderQty}$ and $\text{ListPrice} \times \text{OrderQty}$ respectively.

Microsoft Azure | Data Factory | practiceadf121

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Data Factory | Validate all | Publish all

Preview experience: Off

InputData | bikedata_df

Validate | Data flow debug | Debug Settings

Derived column's settings | Optimize | Inspect | Data preview

Output stream name: **colCostSales**

Description: Creating/updating the columns 'Region, Country, Customer, Business Segment, Category, Model, Color, SalesDate, ...

Incoming stream: **getInputData**

Columns:

Column	Expression
Cost	$\text{UnitPrice} \times \text{OrderQty}$
Sales	$\text{ListPrice} \times \text{OrderQty}$

Properties: General | Related

Name: bikedata_df

Description:

- Create a **profit column** → click on **+** after the colCostSales → **derived column**.

Microsoft Azure | Data Factory | practiceadf121

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Data Factory | Validate all | Publish all

Preview experience: Off

InputData | bikedata_df

Validate | Data flow debug | Debug Settings

Derived column's settings | Optimize | Inspect | Data preview

UPsert: 0 | LOOKUP: 0 | ERROR: 0 | TOTAL: 1000

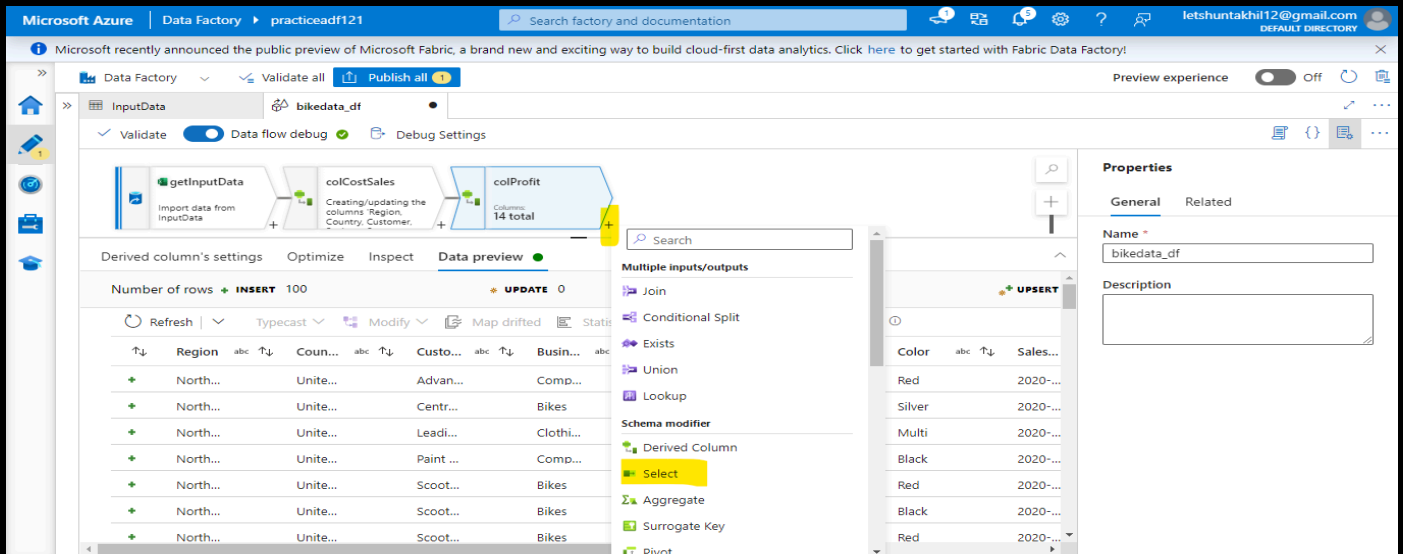
Color	Sales	ListPrice	UnitPrice	OrderQty	Cost	Sales	Profit
Red	2020-...	337.22	183.94	1	183.94	337.22	153.2...
Silver	2020-...	3399.99	2039.99	1	2039.99	3399.99	1359....
Multi	2020-...	49.99	28.84	6	173.04	299.94	126.9
Black	2020-...	1349.6	714.7	2	1429.4	2699.2	1269....
Red	2020-...	1457.99	874.79	2	1749.58	2915.98	1166.4
Black	2020-...	782.99	419.46	2	838.92	1565.98	727.0...
Red	2020-...	782.99	419.46	4	1677.84	3131.96	1454....

Properties: General | Related

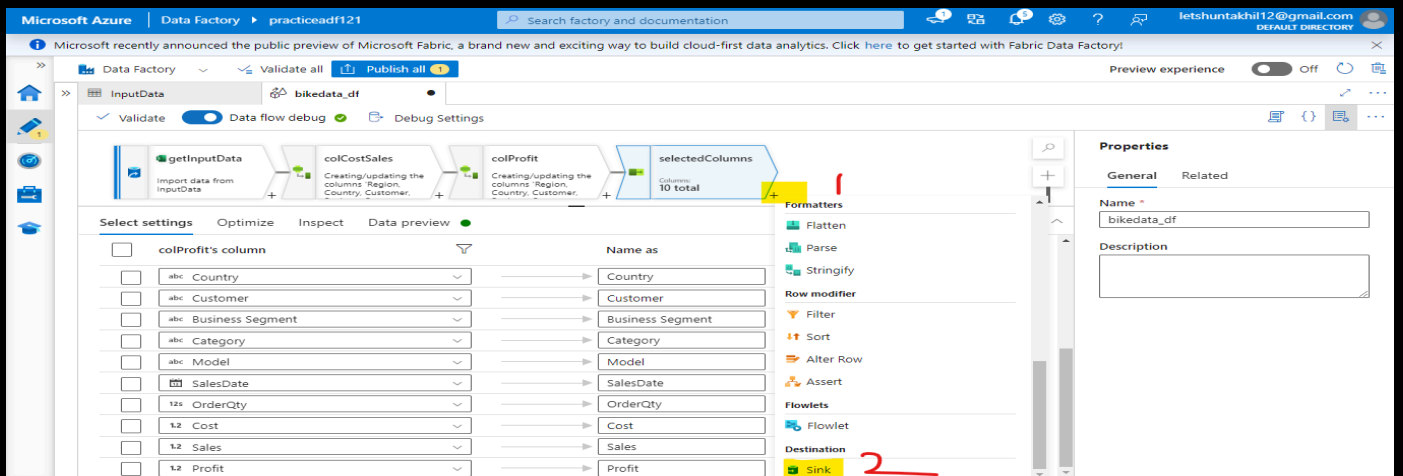
Name: bikedata_df

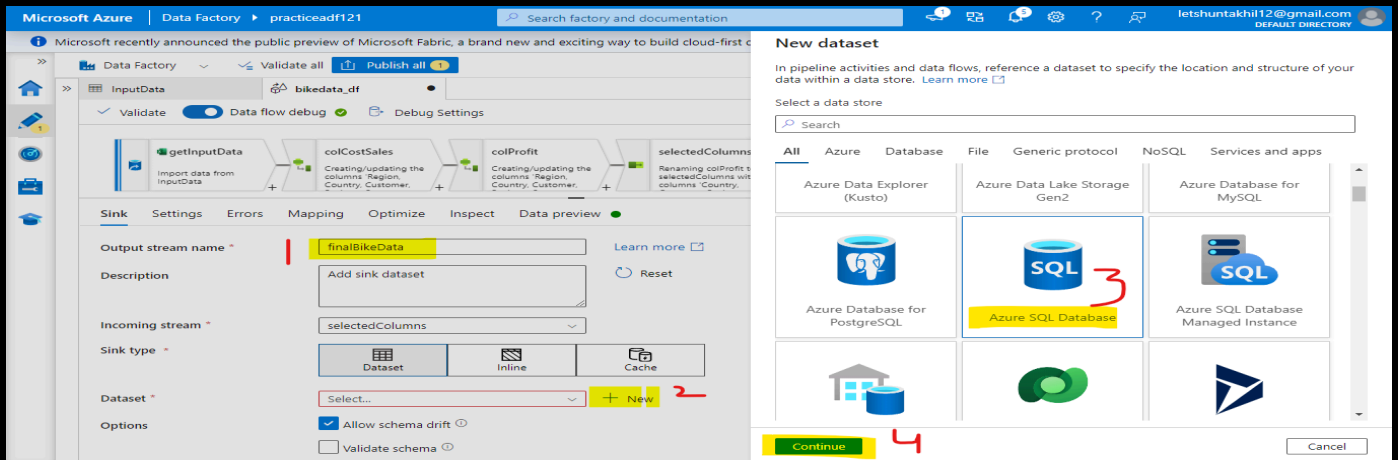
Description:

- Drop the columns which are not frequently used like UnitPrice , ListPrice,Region and Color.
- We can do it by clicking the + button and search for the 'select' option.

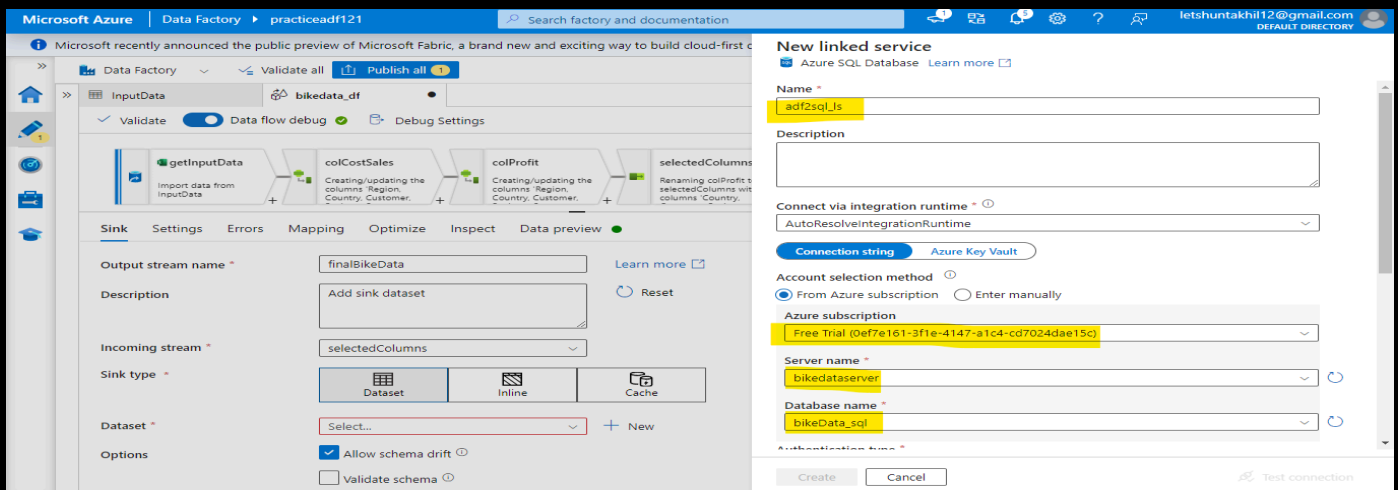


- Drop the selected columns by clicking the delete icon.
- Store the clean data in azure sql database by choosing the sink in dataflow after the selectedColumns.





- Click on the **new linked service**.
- Create a new linked service by entering the details. Linked service is to connect or make a bridge between adf and azure sql database.



Microsoft Azure | Data Factory | practiceadf121

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!

InputData | bikedata_df | Publish all

Validate | Data flow debug | Debug Settings

getInputData | colCostSales | colProfit | selectedColumns

Sink | Settings | Errors | Mapping | Optimize | Inspect | Data preview

Output stream name: finalBikeData

Description: Add sink dataset

Incoming stream: selectedColumns

Sink type: Dataset

Dataset: Select...

Options: Allow schema drift

New linked service: bikeData_sql

Authentication type: SQL authentication

User name: bikedata_admin

Password: [masked]

Always encrypted: [unchecked]

Additional connection properties: + New

Annotations: + New

Parameters: >

Advanced: >

Create | Cancel | Test connection

- Save the Target dataset path by **publish all** → **publish**

Microsoft Azure | Data Factory | practiceadf121

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!

InputData | bikedata_df | OutputData | Publish all

Validate | Data flow debug | Debug Settings

getInputData | colCostSales | colProfit | selectedColumns | finalBikeData

Sink | Settings | Errors | Mapping | Optimize | Inspect | Data preview

Incoming stream: selectedColumns

Sink type: Dataset


Dataset: OutputData

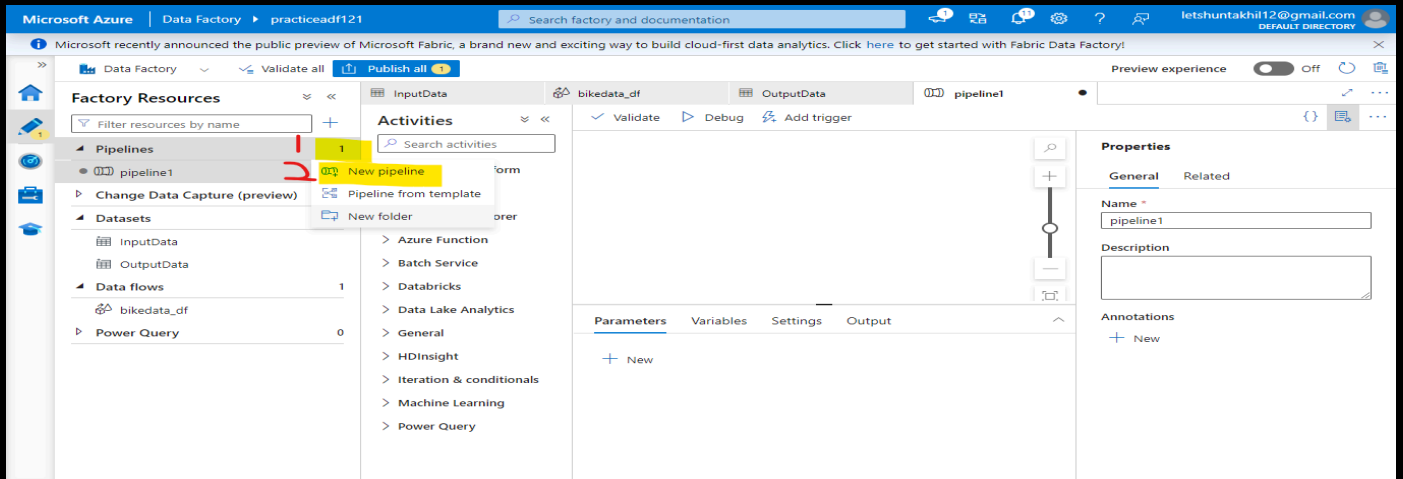
Options: Allow schema drift

Properties: General | Related

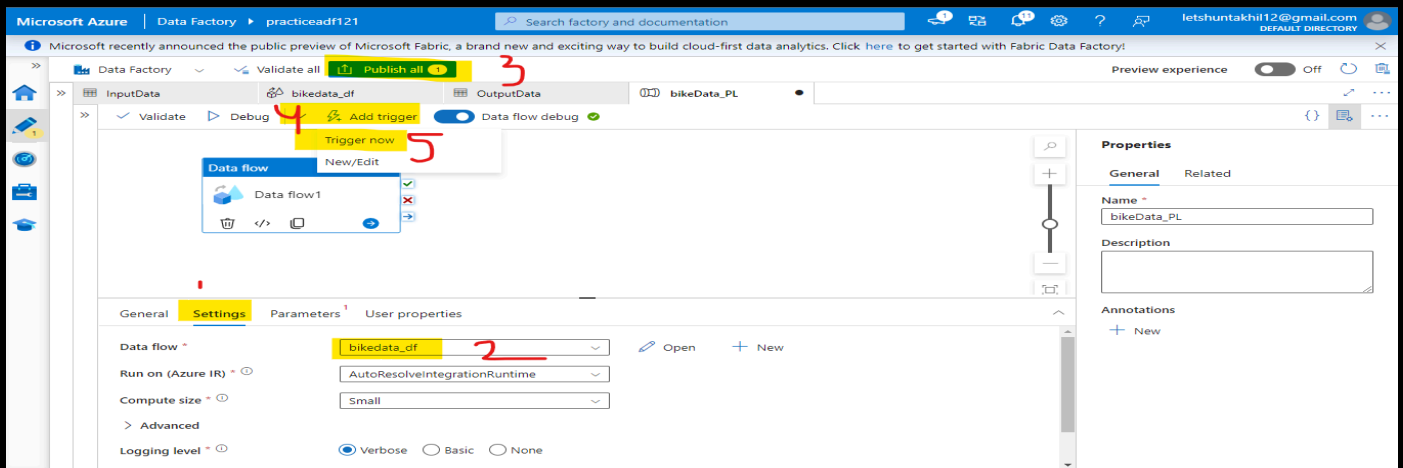
Name: bikedata_df

Description: [empty]

- Save the transformation by **publish all** → **publish**. Let's execute the dataflow
- Pencil icon () → **new pipeline**



- Drag and drop the **data flow activity** to the workspace and click **publish all** → **publish**.



- Let's run the pipeline → **Add Trigger** → **Trigger now** → **ok**.
- See the status of the pipeline → **view pipeline run**.

Microsoft Azure | Data Factory | practiceadf121

Search factory and documentation

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data solutions.

Notifications

- Dismiss all
- Running**
Successfully running bikeData_PL (Pipeline).
[View pipeline run](#)
a few seconds ago
- Publishing completed**
Successfully published
2 minutes ago
- Publishing completed**
Successfully published
15 minutes ago
- Publishing error**
Validation of model(s) failed. Fix it before publishing
16 minutes ago
- Successfully deleted**
Successfully deleted OutputData12 (Dataset).
17 minutes ago

Activity runs

Pipeline run ID: 3c3fb561-215d-405e-bd41-2ab80edba1

All status

Showing 1 - 1 items

Activity name	Activity status	Activity type
Data flow1	In progress	Data flow

- Now after successful pipeline execution, see the table in the azure sql database and Query the Data.

Microsoft Azure

Search resources, services, and docs (G+)

Home > bikeData_sql (bikedataserver/bikeData_sql)

bikeData_sql (bikedataserver/bikeData_sql) | Query editor (preview)

SQL database

Search

Login New Query Open query Feedback Getting started

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

Settings

- Compute + storage
- Connection strings
- Properties
- Locks

Data management

- Replicas
- Sync to other databases

Integrations

- Azure Synapse Link

bikeData_sql (bikedata_admin)

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

- Tables
 - dbo.data_tb
- Views
- Stored Procedures

Query 1

Run Cancel query Save query Export data as Show only Editor

```
1 select * from [dbo].[data_tb];
2
```

Results Messages

Search to filter items...

Country	Customer	Business Segment	Category
United States	Good Bike Shop	Clothing	Jerseys
United States	Remarkable Bike Store	Components	Wheels
United States	Retail Sporting Equipment	Components	Wheels

Query succeeded | 5s

