

Varthya Akhil

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 PowerQuery	15

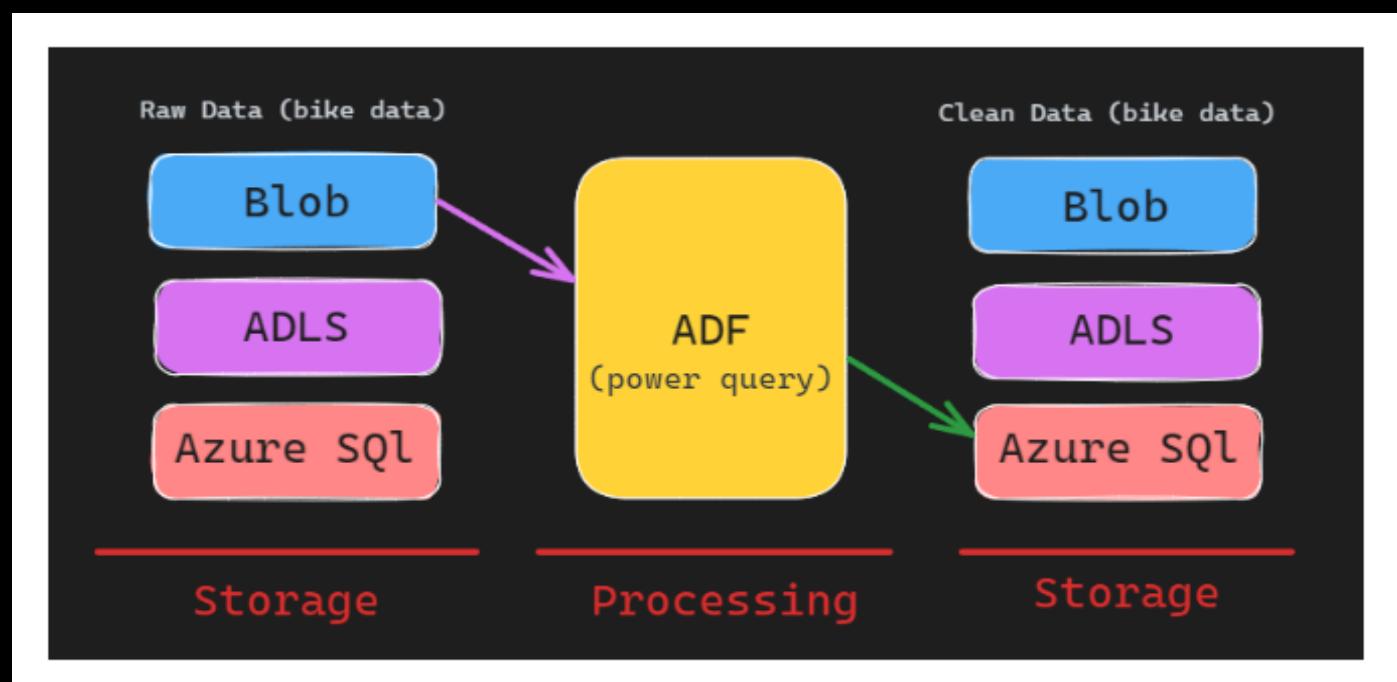
Agenda

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'.

The screenshot shows the Microsoft Azure home page. At the top, there's a navigation bar with 'Microsoft Azure', 'Upgrade', a search bar, and user information. Below the bar, under 'Azure services', there's a 'Create a resource' button with a red arrow pointing to it. Other service icons include Resource groups, All resources, Monitor, Storage accounts, Quickstart Center, Virtual machines, App Services, and SQL databases. A large central area is labeled 'Resources' with sections for 'Recent' and 'Favorite'. It shows a message 'No resources have been viewed recently' with a 'View all resources' button. At the bottom, there's a 'Navigate' section with links to Subscriptions, Resource groups, All resources, and Dashboard.

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

The screenshot shows the 'Create a resource' marketplace. On the left, there's a sidebar with categories like Containers, Databases, Developer Tools, DevOps, Identity, Integration, Internet of Things, IT & Management Tools, Media, Migration, Mixed Reality, Monitoring & Diagnostics, Networking, Security, and Storage. A red arrow points to the 'Storage' category. The main area lists various Azure services with their icons and creation links. Some services shown include SQL, Function App, Key Vault, Data Factory, Template deployment (deploy using custom templates), Logic App, Automation, Public IP address, Ubuntu Server 22.04 LTS, Red Hat Enterprise Linux 7.4, Essentials 50K, SendGrid, MongoDB Atlas (pay-as-you-go), Standard, Microsoft Defender for Endpoint, and Azure Backup - AVS.

- Click on the storage account from the list.

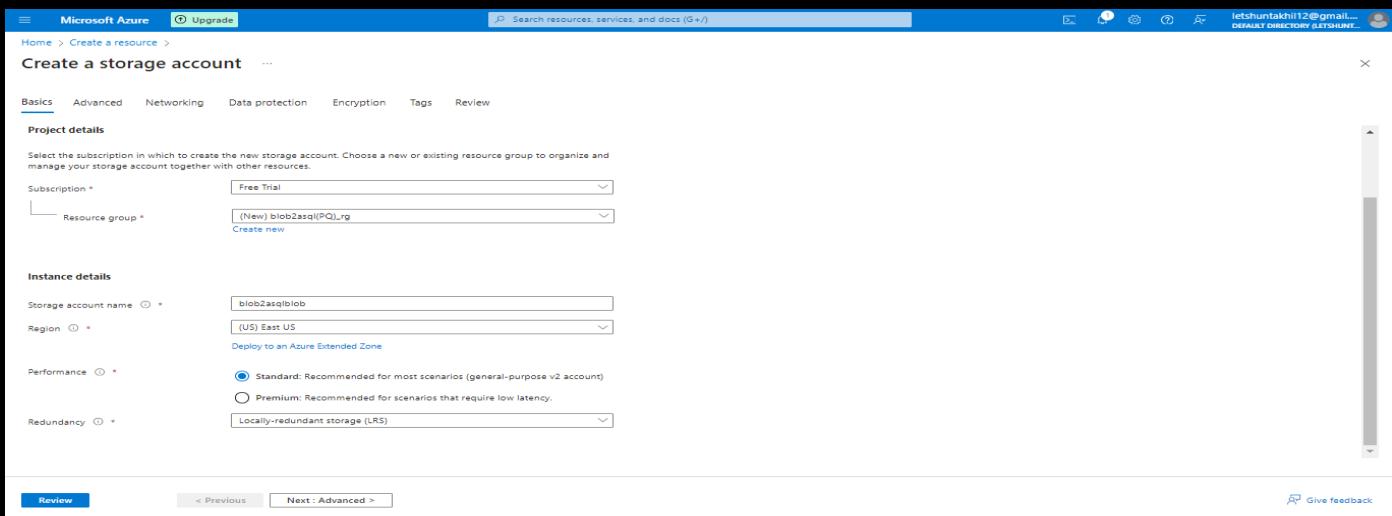
The screenshot shows the Microsoft Azure portal's 'Create a resource' interface. On the left, there's a sidebar with categories like AI + Machine Learning, Analytics, Blockchain, Compute, Containers, Databases, Developer Tools, DevOps, Identity, Integration, Internet of Things, IT & Management Tools, and Media. In the center, there's a search bar and sections for 'Popular Azure services' and 'Popular Marketplace products'. The 'Storage account' option under 'Popular Azure services' is highlighted with a red arrow. It has a 'Create' and 'Learn more' button next to it. Other services listed include Azure Cosmos DB, Azure File Sync, Azure Data Lake Storage Gen1, Azure NetApp Files, Azure Data Box, and Azure Cosmos DB for MongoDB.

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

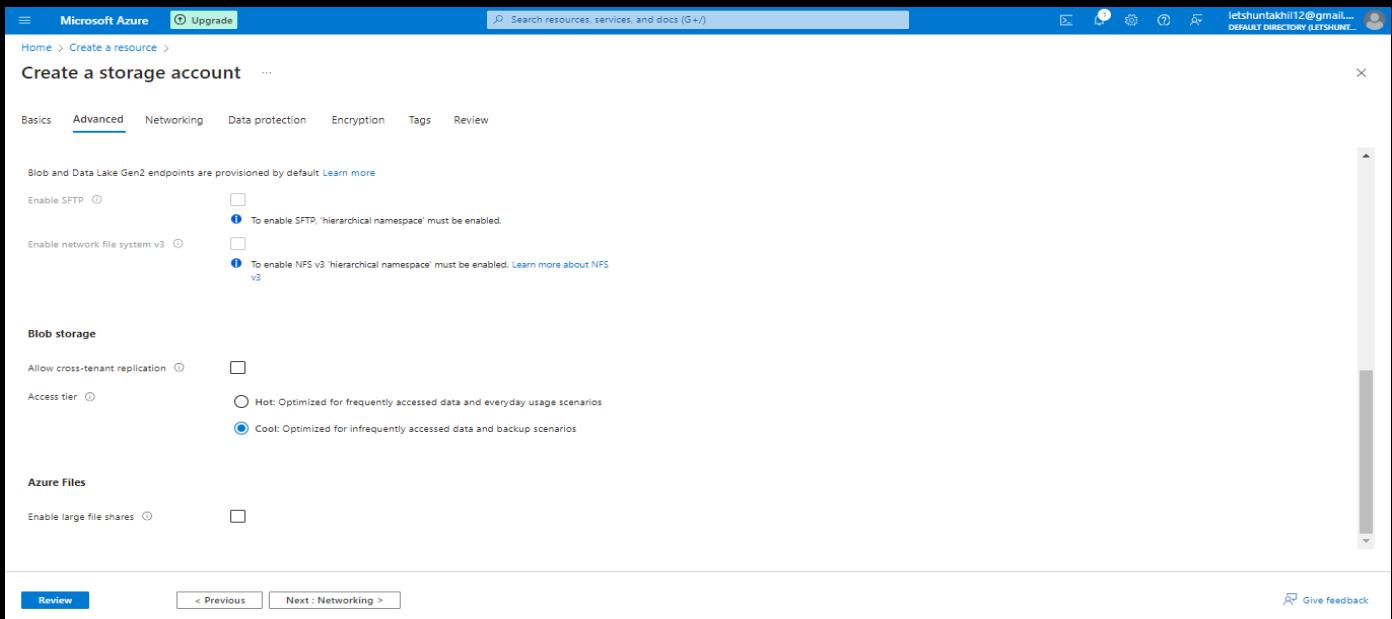
The screenshot shows the 'Create a storage account' form. At the top, there are tabs for Basics, Advanced, Networking, Data protection, Encryption, Tags, and Review + create. The Basics tab is selected. Below the tabs, there's a brief description of Azure Storage and a 'Project details' section where users can choose a subscription and resource group. Under 'Instance details', users can enter the storage account name and select a region. At the bottom, there are 'Previous', 'Next', and 'Review + create' buttons, along with a 'Give feedback' link.

- 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)**.

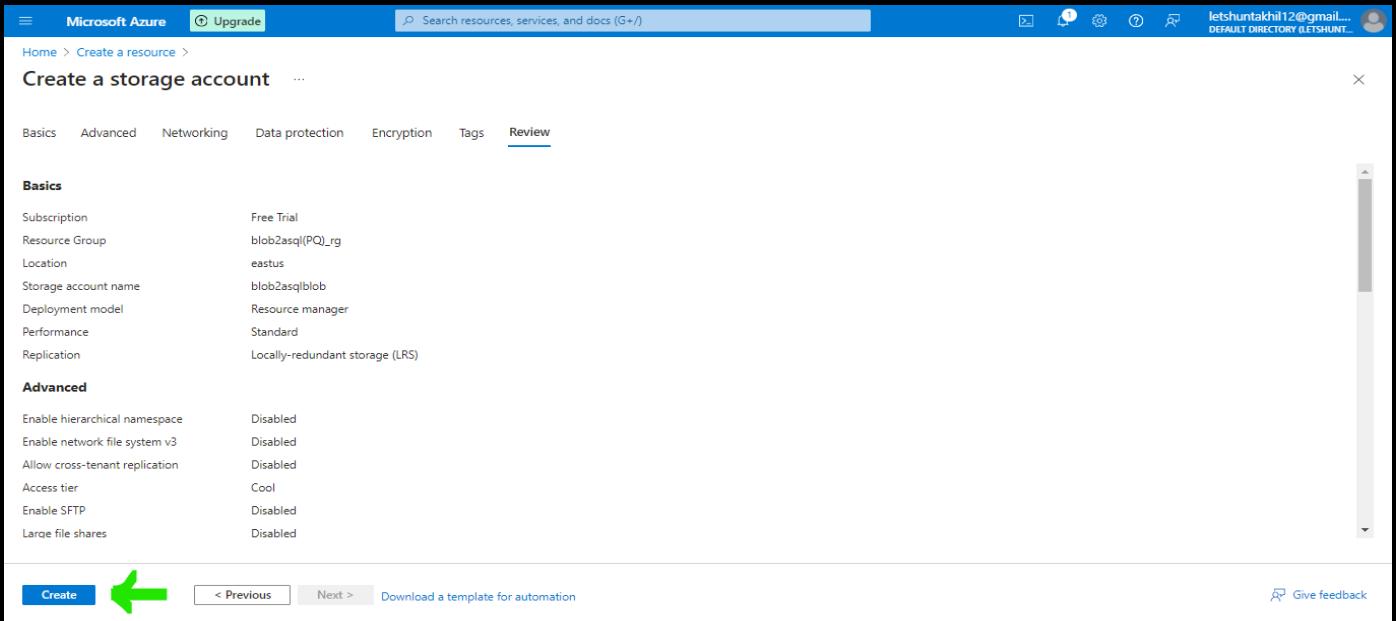


- Click next (Advanced).
- In the advanced tab :
 - Leave the fields as it is but change the Access tier to Cool.
 - Access Tier :
 - Hot : Optimized for frequently accessed data and everyday usage scenarios.
 - Cool : Optimized for infrequently accessed data and backup scenarios.



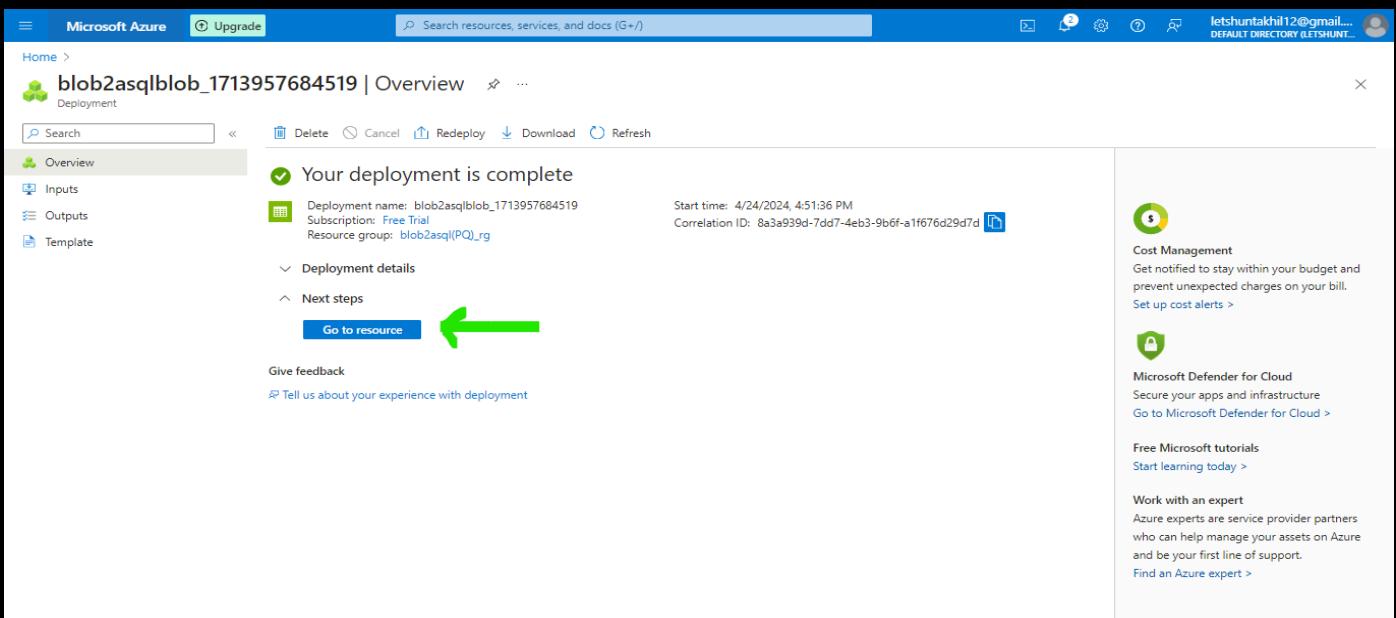
- Click next (Networking).

- **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**.



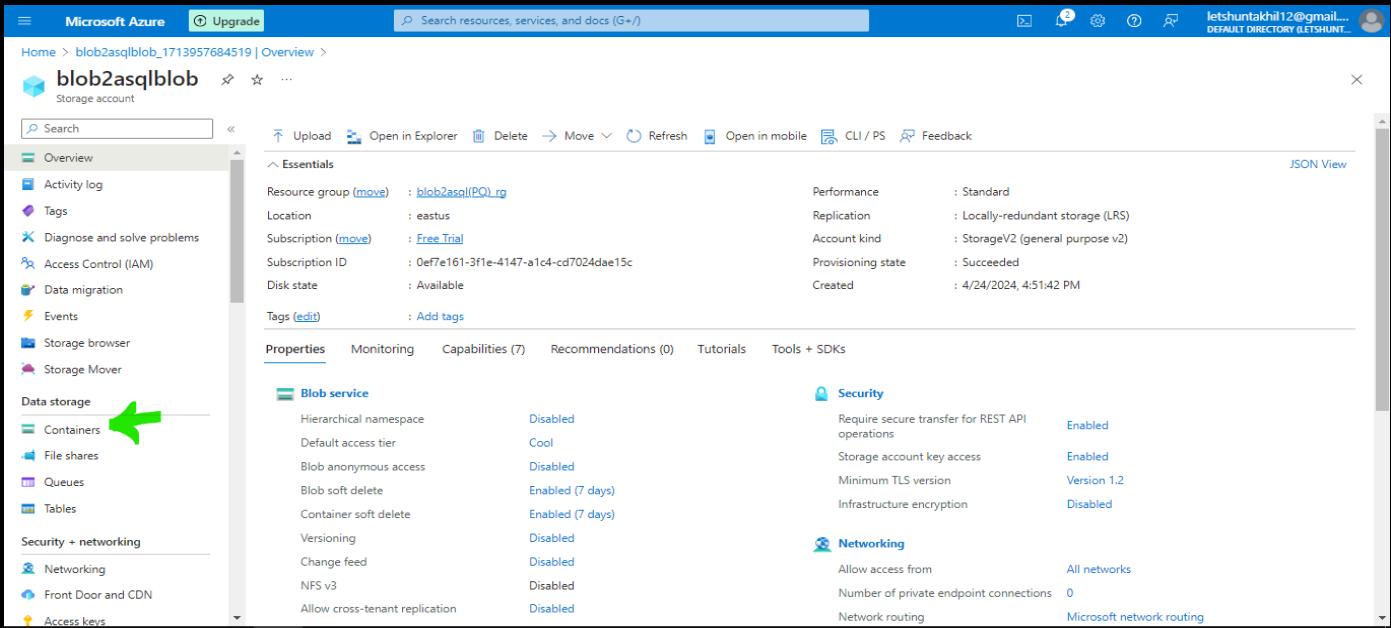
The screenshot shows the 'Create a storage account' wizard in the Microsoft Azure portal. The 'Review' tab is selected. The 'Basics' and 'Advanced' sections are displayed, showing configuration details like Subscription (Free Trial), Resource Group (blob2asql(PQ)_rg), Location (eastus), Storage account name (blob2asqlblob), Deployment model (Resource manager), Performance (Standard), and Replication (Locally-redundant storage (LRS)). At the bottom, there are buttons for 'Create' (highlighted with a green arrow), '< Previous', 'Next >', and 'Download a template for automation'. A 'Give feedback' link is also present.

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



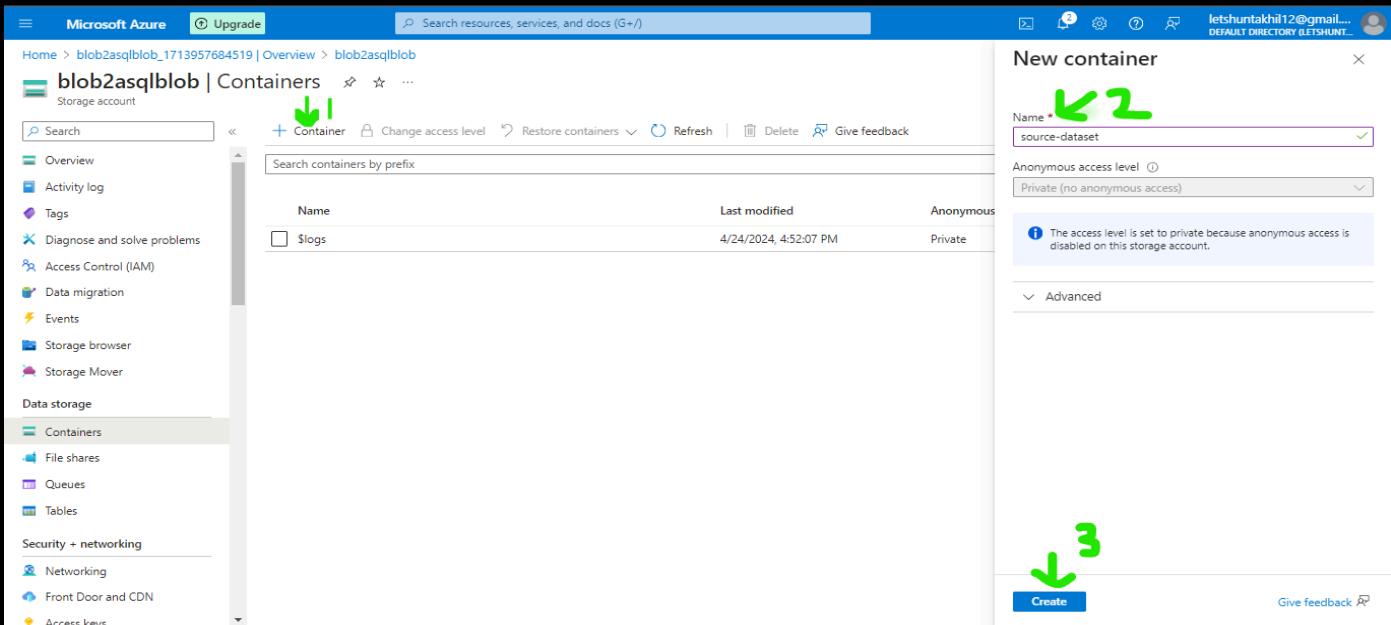
The screenshot shows the 'blob2asqlblob_1713957684519 | Overview' page in the Microsoft Azure portal. The 'Overview' tab is selected. It displays deployment status: 'Your deployment is complete'. Deployment details include name (blob2asqlblob_1713957684519), subscription (Free Trial), and resource group (blob2asql(PQ)_rg). A green checkmark icon is shown next to the completion message. At the bottom, there are buttons for 'Go to resource' (highlighted with a green arrow), 'Give feedback', and 'Tell us about your experience with deployment'. On the right side, there are promotional cards for Cost Management, Microsoft Defender for Cloud, Free Microsoft tutorials, and Work with an expert.

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



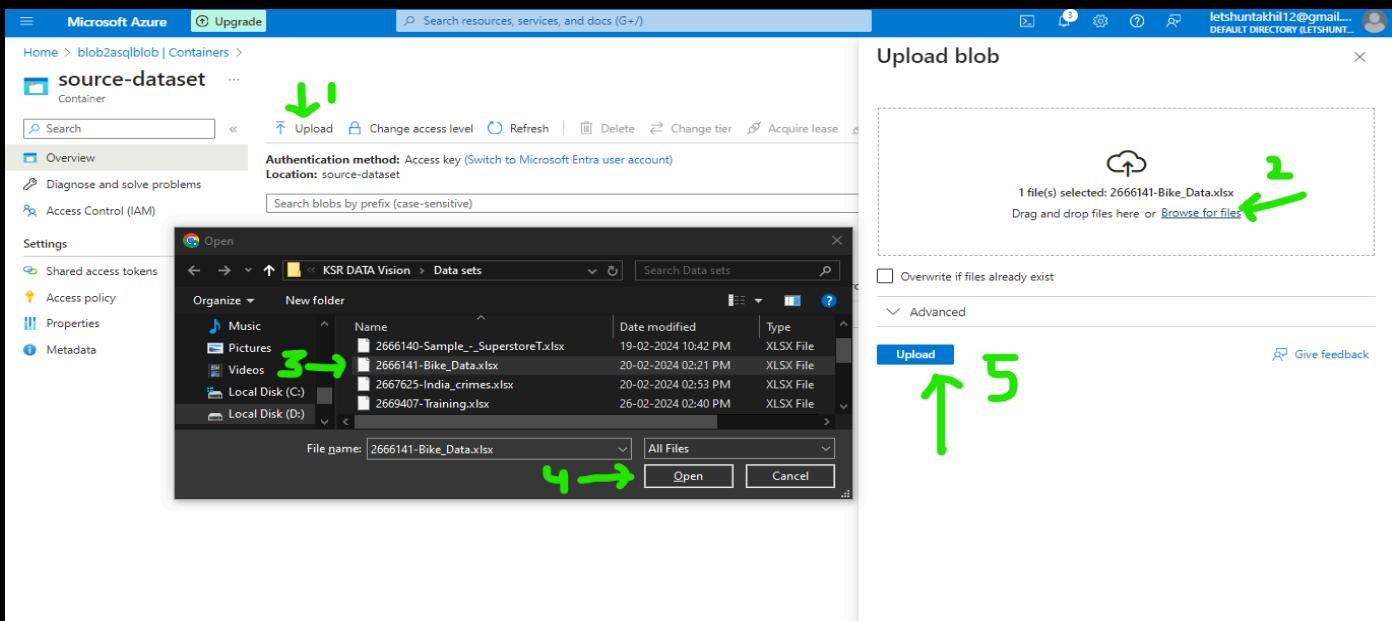
This screenshot shows the Microsoft Azure Storage account overview page for 'blob2asqlblob'. The left sidebar has 'Containers' highlighted with a green arrow. The main content area displays account details under 'Essentials' and various service properties like Blob service, Security, and Networking.

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



This screenshot shows the 'Containers' page for the 'blob2asqlblob' storage account. A 'New container' dialog box is open, prompting for a container name ('source-dataset'). The 'Create' button at the bottom is highlighted with a green arrow. The main table lists one existing container named 'Slogs'.

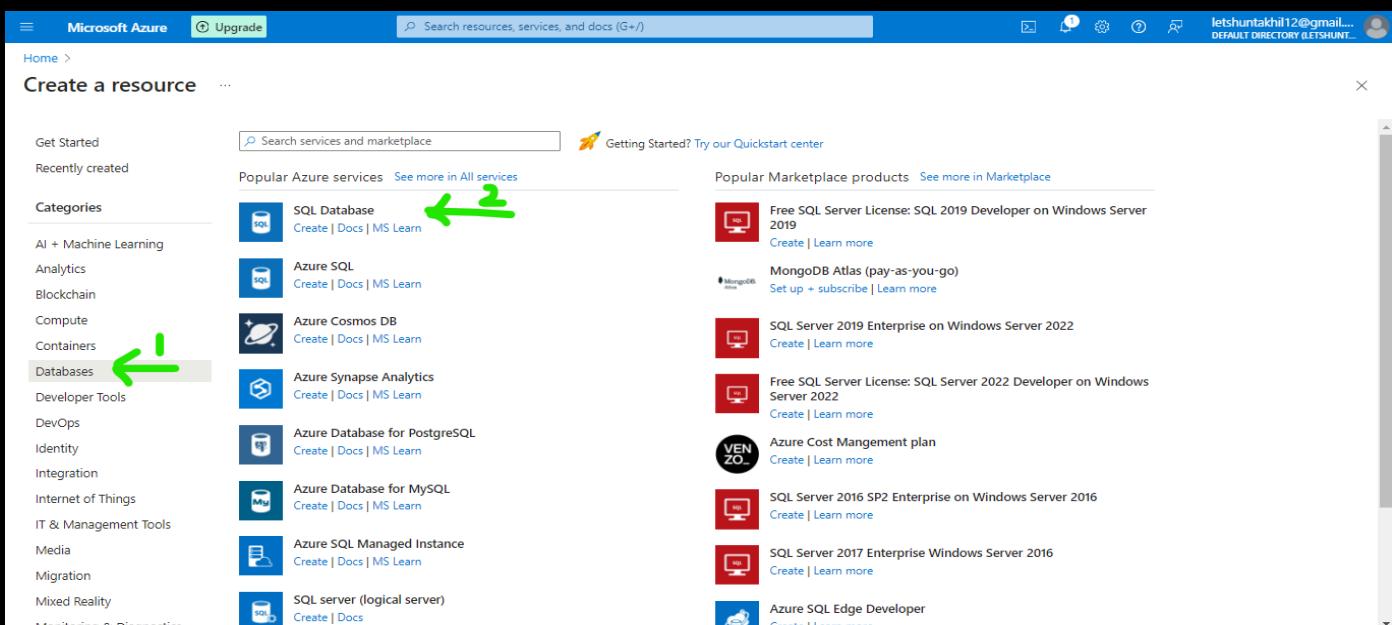
- Click on the container 'source-dataset' after creating.
- Click Upload -> Browse for files -> select the file -> open -> upload .
- The bike data is uploaded.



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

Azure blob storage 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.

Microsoft Azure Upgrade

Home > Create a resource > Create SQL Database > Search resources, services, and docs (G+)

Create SQL Database Server ... Microsoft Server security

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 * blob2asqlserver .database.windows.net

Location * (US) East US

Authentication

Azure Active Directory (Azure AD) is now Microsoft Entra ID. [Learn more](#)

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 select both SQL and Microsoft Entra authentication.

Authentication method Use Microsoft Entra-only authentication Use both SQL and Microsoft Entra authentication **Use SQL authentication**

Server admin login * serveradmin

Password * *****

Confirm password * ***** >Password and confirm password must match.

OK

<https://aka.ms/AADRebrandFAQ>

- **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).

Create SQL Database

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

Database name *: blob2sqldb

Server: (new) blob2sqlserver (East US)

Want to use SQL elastic pool?: No (Development selected)

Workload environment: Development

Compute + storage *: General Purpose - Serverless (Standard-series (Gen5), 1 vCore, 32 GB storage, zone redundant disabled)

Backup storage redundancy: Locally-redundant backup storage

Review + create | **Next : Networking >**

Configure

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

Service tier: General Purpose (Most budget friendly) (selected)

Compute tier: General Purpose (Most budget friendly)

Compute Hardware: DTU-based purchasing model

Hardware Configuration: Premium (Highest availability and performance) (selected)

Max vCores: 1

Min vCores: 0.5 vCores

Cost summary

General Purpose (GP S Gen5 1)	9.57
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: Serverless databases are billed in vCore seconds based on a combination of CPU and memory utilization. Learn more about serverless billing.

Apply

Configure

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

Service tier: Basic (For less demanding workloads) (selected)

DTUs: Compare DTU options

5 (Basic)

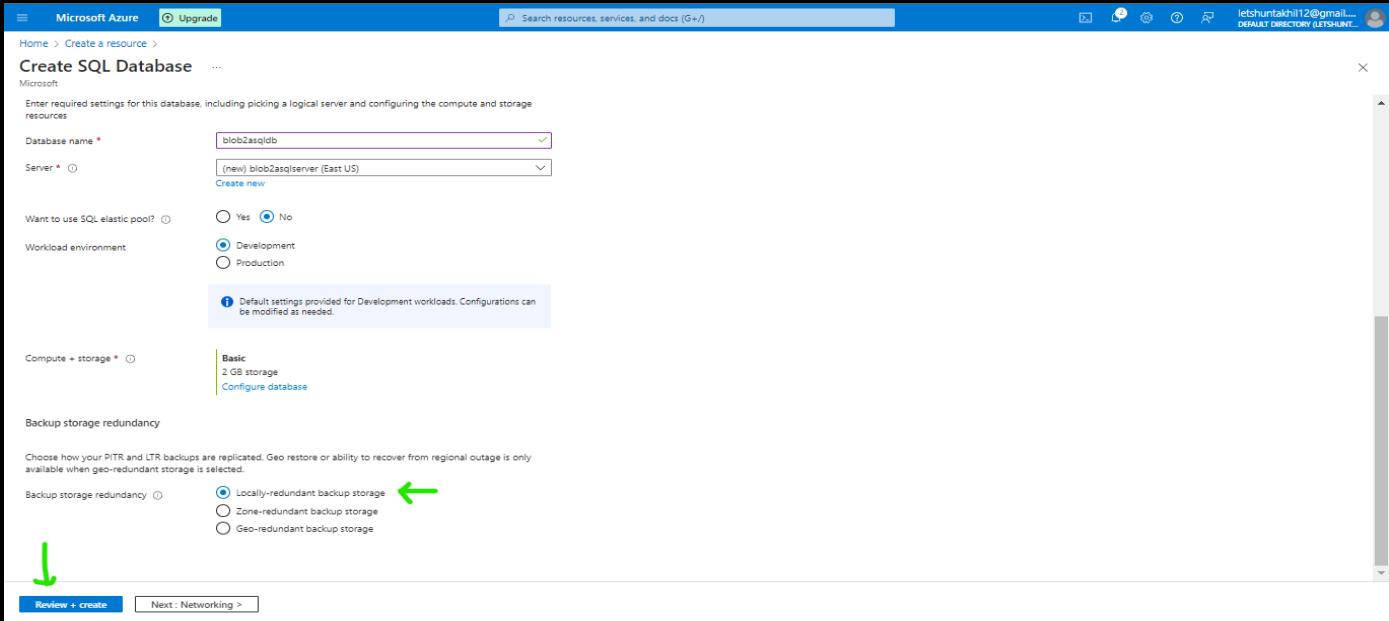
Data max size (GB): 2

Cost summary

Basic (Basic)	81.49
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**).



- **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**)

Cost summary

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

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 2

Add current client IP address: No Yes 3

Connection policy 4 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 provided via the Azure SQL Database gateway.
 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 [TLS/SSL](#). [Learn more](#).

Minimum TLS version: TLS 1.2 5

Review + create < Previous Next: Security >

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

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 1

Ledger

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

Ledger: Not configured 2 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 3 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 4 Configure transparent data encryption

Review + create < Previous Next: Additional settings > 5

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

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 1

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_CI_AS. [Learn more](#).

Collation: SQL_Latin1_General_CI_AS 2 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: System default (5pm to 8am) 3

Review + create < Previous Next: Tags > 4

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

Create SQL Database

Tags

Name: Value

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**' .

Create SQL Database

Estimated cost per month: 407.46 INR

Product details

SQL database by Microsoft. Terms of use | Privacy policy

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	blob2sql(PQ).rg
Region	East US
Database name	blob2sqldb
Server	(new) blob2sqlserver
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.SQLDatabase.newDatabaseNewServer_54b3868e18d5461eaac3d | Overview

Your deployment is complete

Deployment name: Microsoft.SQLDatabase.newDatabaseNewServer_54b3868e18d5461eaac3d
Subscription: Free Trial
Resource group: blob2sql(PQ).rg

Start time: 4/24/2024, 6:25:47 PM
Correlation ID: fb16e9e2-34c8-4a18-936d-ac4bbd121f8e

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.

The screenshot shows the Microsoft Azure portal interface. On the left, there's a sidebar with various navigation options like Overview, Activity log, Tags, and Query editor (preview). The main content area displays the 'blob2asqldb' database details under the 'Essentials' section. It includes information such as Resource group, Status, Location, Subscription, Tags, and server settings. Below this, there are sections for 'Start working with your database', 'Configure access', 'Connect to application', and 'Start developing', each with corresponding buttons to 'Configure' or 'See connection strings'. The top right corner shows user profile information: Name: Srikanth Chavan, Email: letshuntakhi12@gmail.com, Directory: Default Directory (43da855c-01b0-4495-88c1-9caf32372d5), and Domain: letshuntakhi12@gmail.onmicrosoft.com.

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

This screenshot shows the 'Query editor (preview)' page for the 'blob2asqldb' database. The left sidebar is identical to the previous screenshot. The main area features a 'Welcome to SQL Database Query Editor' message. It provides two authentication options: 'SQL server authentication' (with fields for 'Login' (serveradmin) and 'Password') and 'Microsoft Entra authentication' (with a 'Continue as' button for 'letshuntakhi12@gmail.com'). A green arrow points to the 'OK' button at the bottom of the form.

- Successfully we have created the azure sql database.

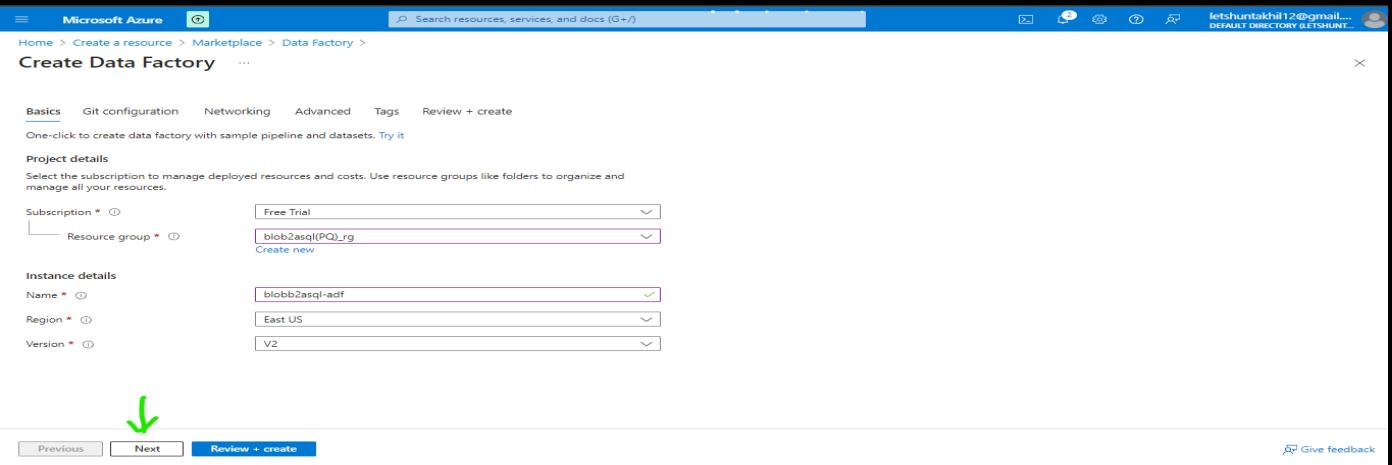
Azure Data Factory Creation :

- 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.

The screenshot shows the Microsoft Azure Marketplace interface. On the left, there's a sidebar with various categories like Get Started, Service Providers, Management, Private Marketplace, and My Marketplace. The main area has a search bar at the top with the query 'azure data factory'. Below the search bar, there are several filters: Pricing : All, Operating System : All, Publisher Type : All, Product Type : All, and Publisher name : All. A message box says 'New! Get AI-generated suggestions for your search. Ask AI to suggest products, articles, and solutions for what you need.' Below the filters, it says 'Showing 1 to 20 of 65 results for "azure data factory"'. There are two rows of tiles, each containing five items. The first item in the first row is 'Data Factory' by Microsoft, which is highlighted with a green checkmark icon. Other items include 'Delphix Compliance Services for Microsoft Azure', 'Cluedin MDM and Data Quality - SaaS', 'Modern Data Mart', 'Data#3 Azure Optimiser', and 'Profisee SaaS Enterprise Master Data Management'. The second row contains items like 'Cluedin MDM and Data Quality - PaaS', 'Profisee Master Data Management - SaaS Solution', 'KoçSistem Azure Data Analytics Service', and 'Smart Factory Platform - Predictive Quality Assurance'.

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

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.

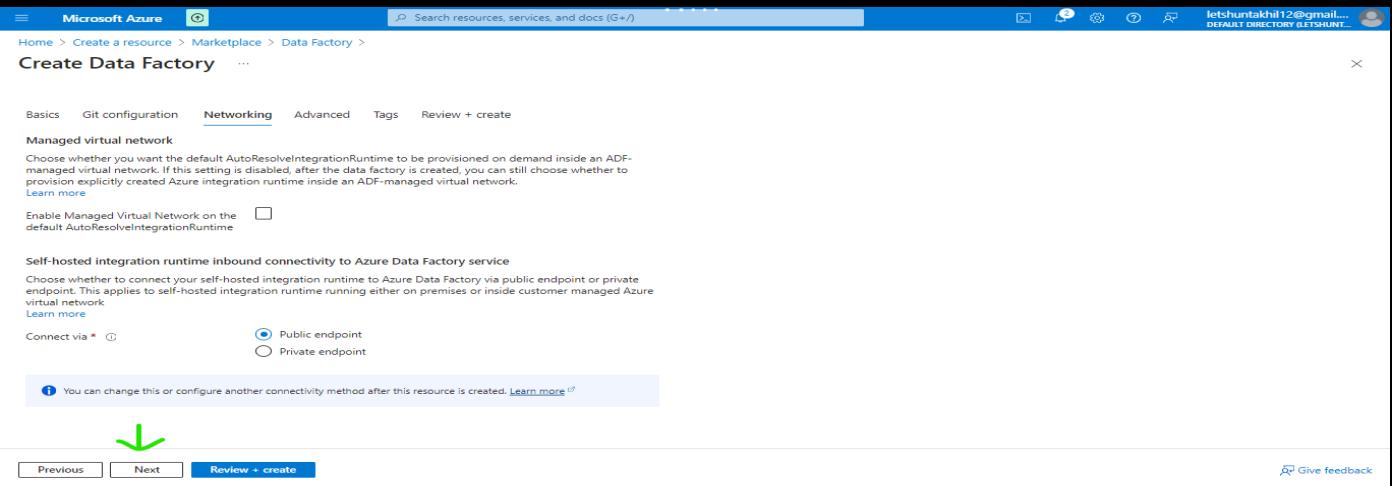
Subscription * Resource group * [Create new](#)

Instance details

Name * Region * Version *

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

- Click next (Networking).



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

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.

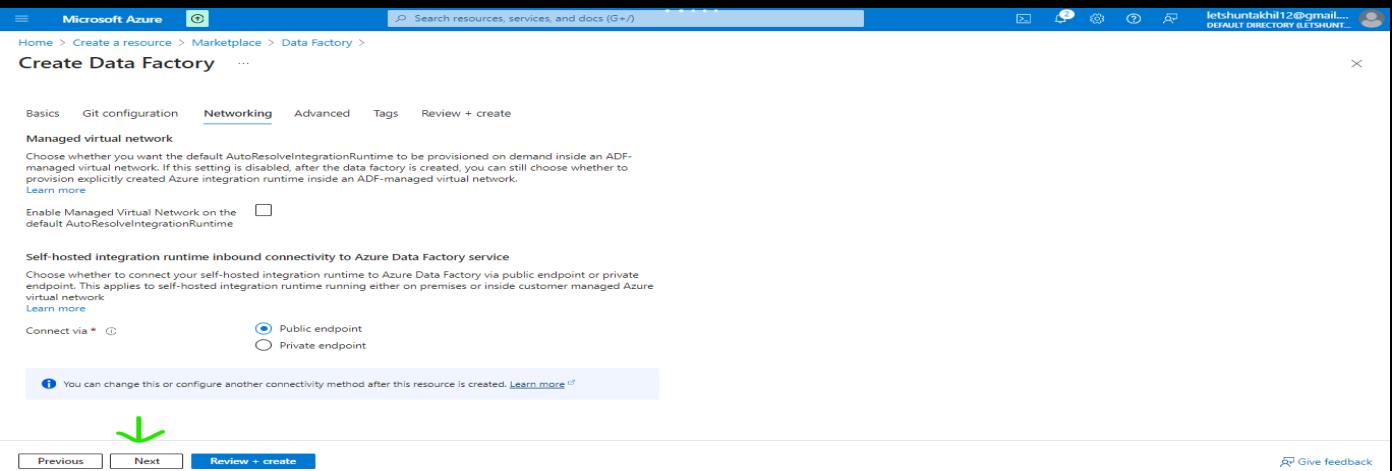
[Learn more](#)

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 Search resources, services, and docs (G+)

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 Search resources, services, and docs (G+) 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 Search resources, services, and docs (G+) Home > Create a resource > Marketplace > Data Factory > Create Data Factory ...

Name: Srikanth Chavan
Email: letshuntakhill12@gmail.com
Directory: Default Directory (43da855c-01b0-4495-88c1-9caf327372d5)
Domain: letshuntakhill12@gmail.onmicrosoft.com

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 Search resources, services, and docs (G+) 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	blob2asql(PQ).rg
Name	blobb2asql-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.

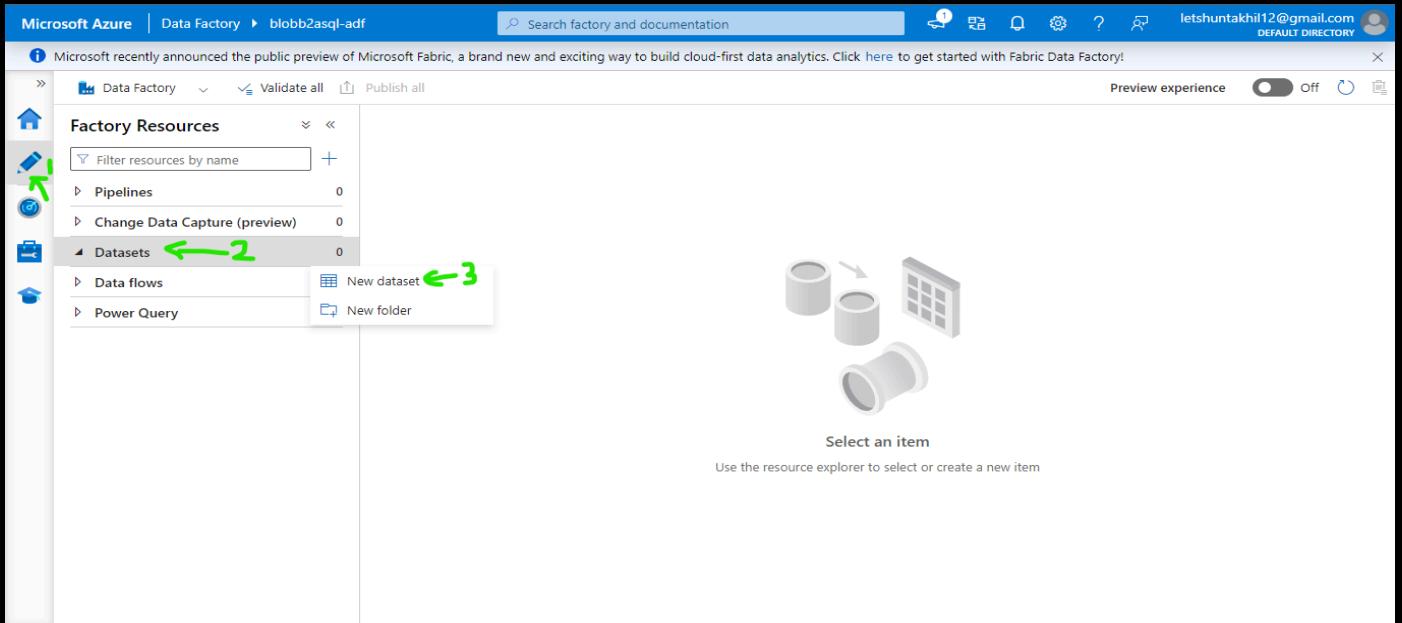
The screenshot shows the Microsoft Azure Data Factory Overview page for a deployment named "Microsoft.DataFactory-20240424185250". The status is "Your deployment is complete". Below the status, there are sections for "Deployment details" and "Next steps". A green arrow points to the "Go to resource" button in the "Next steps" section. To the right of the main content, there are promotional cards for Cost management, Microsoft Defender for Cloud, and Free Microsoft tutorials.

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

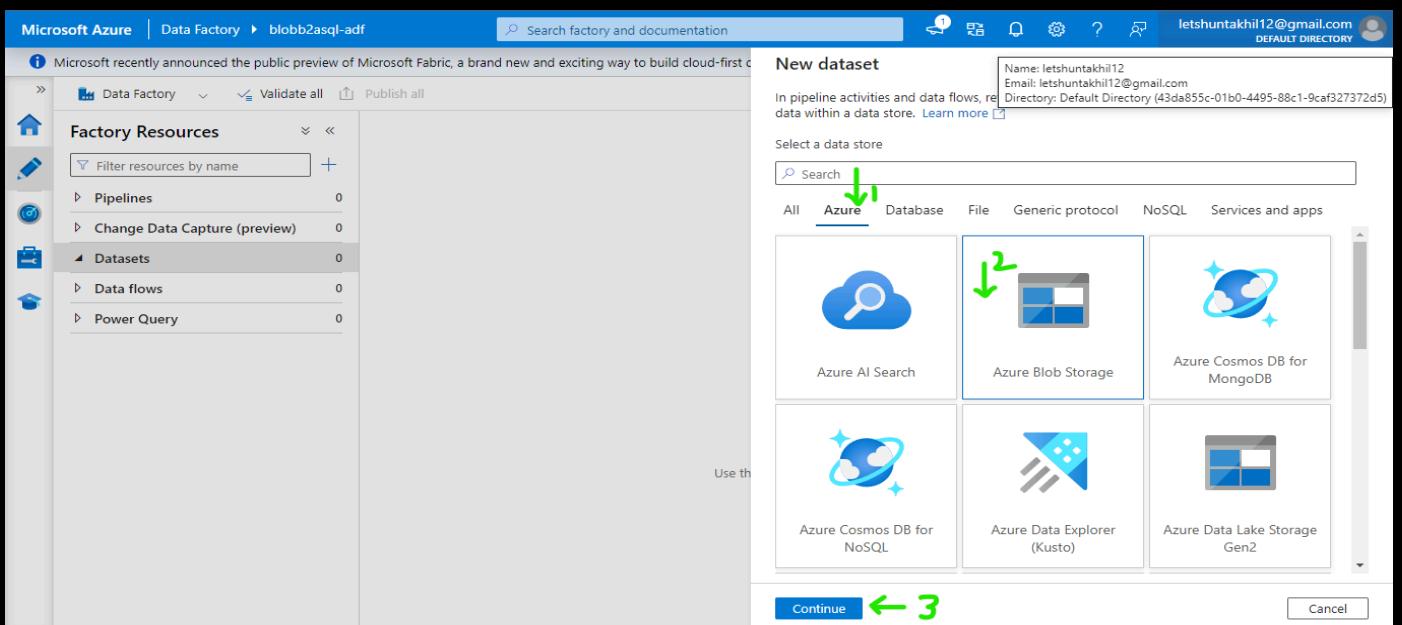
The screenshot shows the Microsoft Azure Data Factory Studio landing page for a data factory named "blobb2asql-adf". The page includes a summary of the data factory's essentials (Resource group, Status, Location, Subscription) and a "Launch studio" button. Below the button are four cards: "Quick Starts", "Tutorials", "Template Gallery", and "Training Modules". A green arrow points to the "Launch studio" button.

Transformation of data using power query:

- 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.

Microsoft Azure | Data Factory > blobb2asql-adf

Select format

Choose the format type of your data

Use the arrow keys to select a different format.

Continue Back Cancel

- 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 .

Microsoft Azure | Data Factory > blobb2asql-adf

Set properties

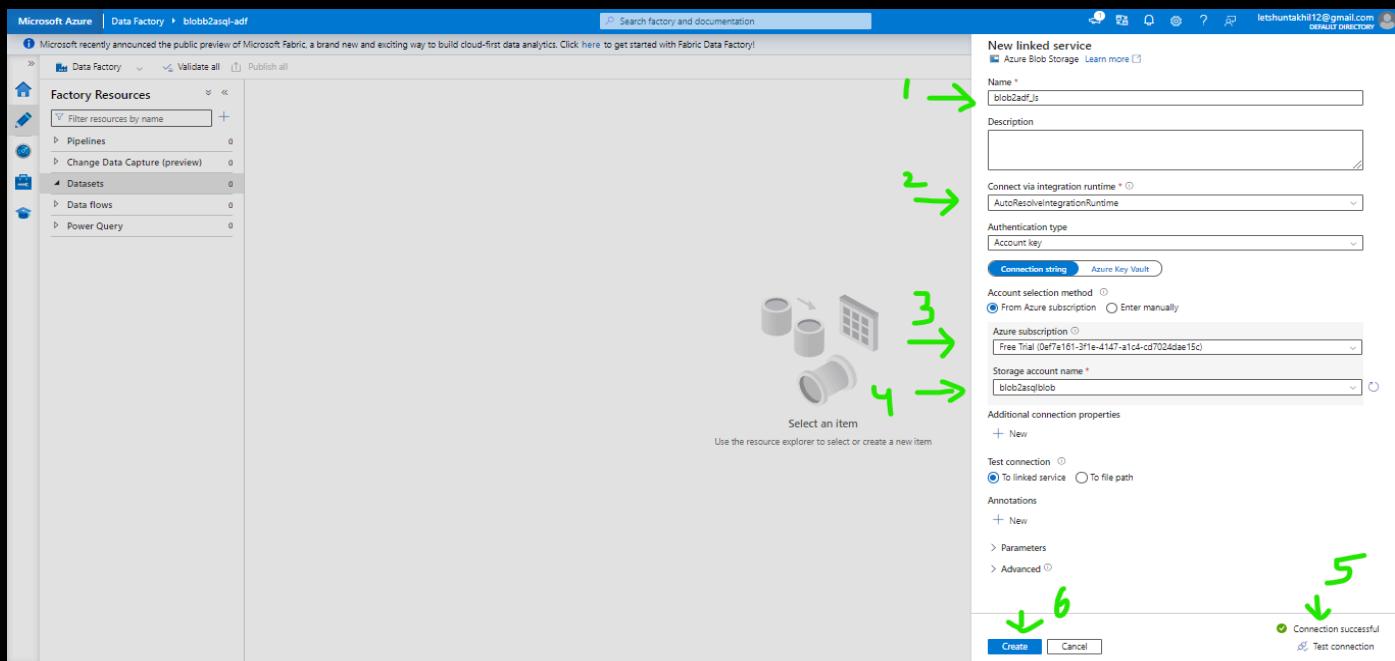
Name

Linked service *

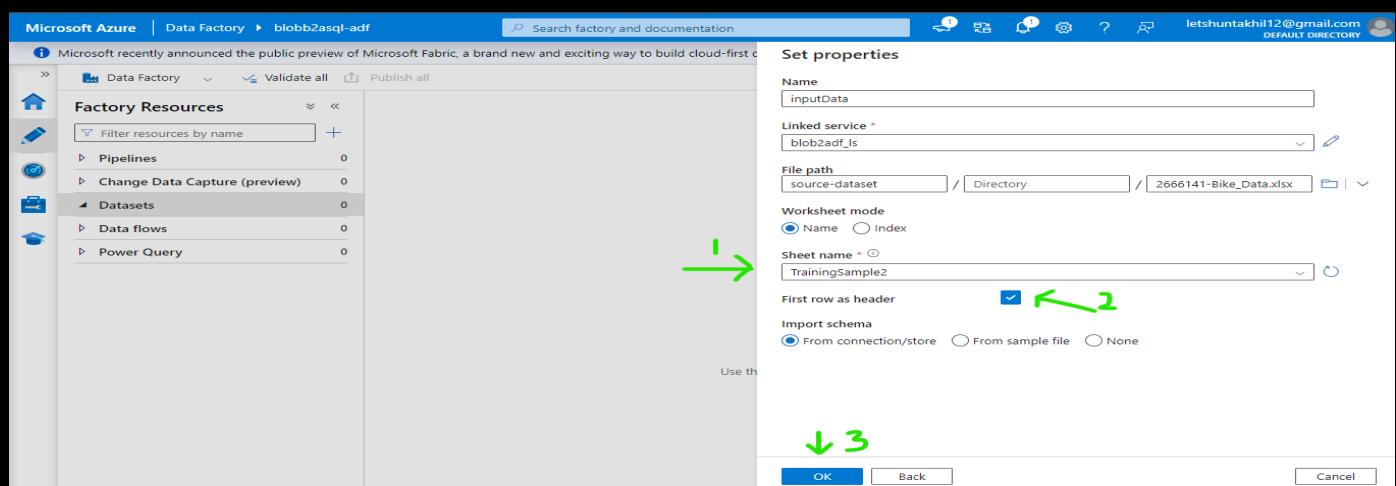
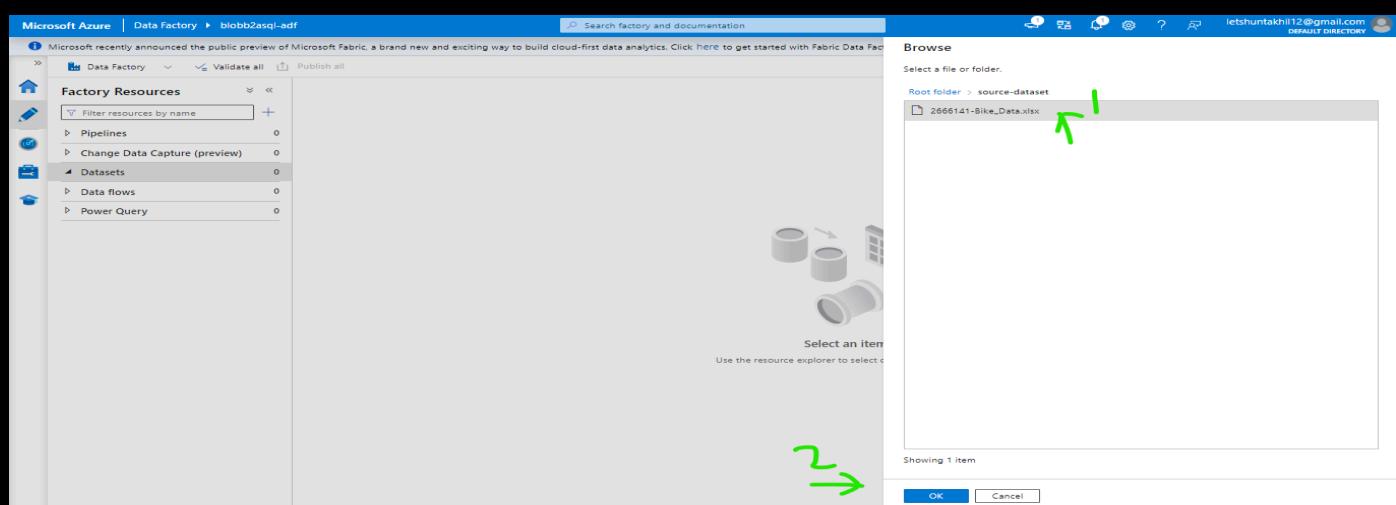
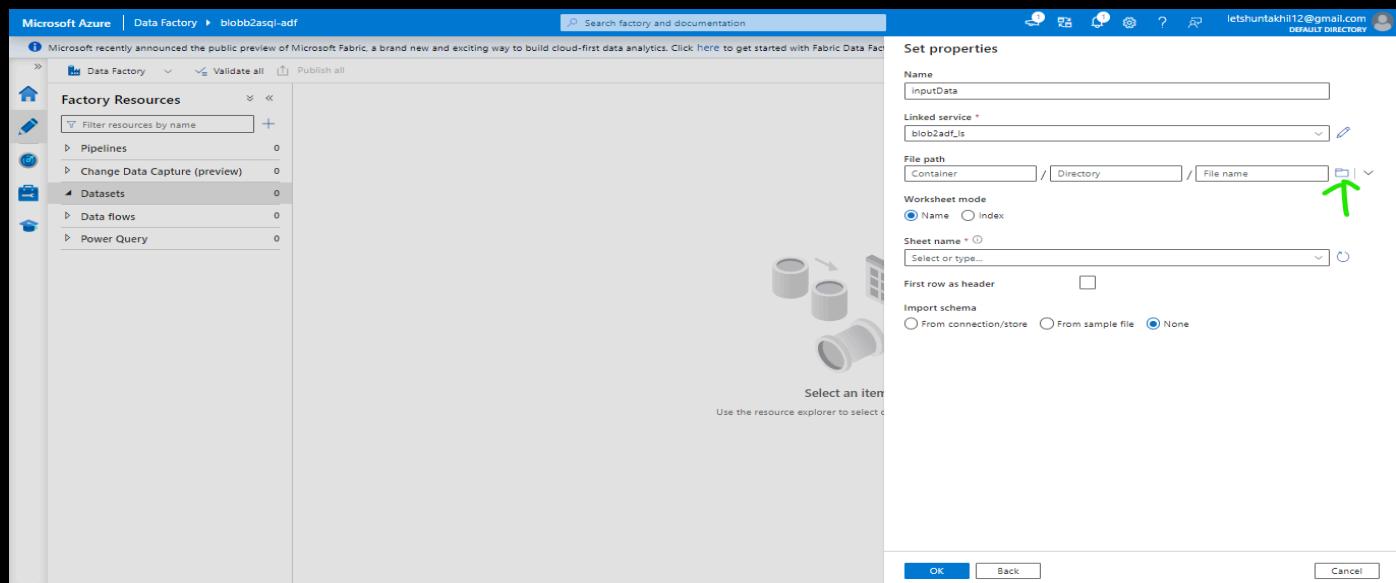
Select... Filter... + New

OK Back Cancel

- 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**.

Microsoft Azure | Data Factory > blob2asql-adf

Search factory and documentation

Publishing 1

Factory Resources

Datasets

inputData

Excel inputData

Connection Schema Parameters

Linked service: blob2adf_ls

File path: source-dataset

Compression type: Select...

Worksheet mode: Name (radio button selected)

Sheet name: TrainingSample2

Pending changes (1)

NAME	CHANGE	EXISTING
inputData	(New)	-

Publish all

You are about to publish all pending changes to the live environment. [Learn more](#)

Published

Cancel

- Successfully brought bike data to the data factory to transform.
- Create a power query → **new power query** which is in the **pencil icon** (✏️) menu.

Microsoft Azure | Data Factory > blob2asql-adf

Search factory and documentation

Preview experience: Off

Factory Resources

Datasets

inputData

Excel inputData

Connection Schema Parameters

New power query

File path: source-dataset / Directory / 2666141-Bike_Data.xlsx

Compression type: Select...

Worksheet mode: Name (radio button selected)

Sheet name: TrainingSample2

Range: e.g. A3:H5

Null value:

- You will get a form to give the name for the power query .
- Settings → Dataset → select the file (inputData) .

Microsoft Azure | Data Factory > blobb2asql-adf

Search factory and documentation

Preview experience Off

Factory Resources

- Pipelines 0
- Change Data Capture (preview) 0
- Datasets 1
 - inputData
- Data flows 0
- Power Query 1
 - bikeData_PQ

No source dataset

Add a source dataset below to start editing your Power Query
Please ensure that all source datasets have a valid schema

Properties

General Related

Name: bikeData_PQ

Description:

Settings

Source dataset*

Dataset

- Power query Editor will appear after selecting the dataset.

Microsoft Azure | Data Factory > blobb2asql-adf

Search factory and documentation

Preview experience Off

inputData bikeData_PQ

Home Transform Add column View Help

Enter data Manage connections Options Manage parameters Options Refresh Advanced editor Choose columns Remove columns Manage columns Keep rows Remove rows Filter rows Sort Split column Group by Replace values Data type: Text Use first row as headers Append queries Merge queries Append Combine files Export template Share

Queries [2]

#	Region	Country	Customer	Business Segment	Category	Model	Color	SalesDate	ListPrice
1	North America	United States	Advanced Bike Components	Components	Road Frames	LL Road Frame	Red	4/1/2020	337.22
2	North America	United States	Central Discount Store	Bikes	Mountain Bikes	Mountain-100	Silver	4/1/2020	3399.95
3	North America	United States	Leading Sales & Repair	Clothing	Jerseys	Long-Sleeve Logo Jersey	Multi	4/1/2020	49.95
4	North America	United States	Paint Supply	Components	Mountain Frames	HL Mountain Frame	Black	4/1/2020	1349.€
5	North America	United States	Scooters and Bikes Store	Bikes	Road Bikes	Road-450	Red	4/1/2020	1457.95
6	North America	United States	Scooters and Bikes Store	Bikes	Road Bikes	Road-450	Black	4/1/2020	782.95
7	North America	United States	Scooters and Bikes Store	Bikes	Road Bikes	Road-450	Red	4/1/2020	782.95
8	North America	United States	Modern Bike Store	Bikes	Road Bikes	Road-450	Red	4/1/2020	1457.95
9	North America	United States	Modern Bike Store	Bikes	Road Bikes	Road-650	Black	4/1/2020	782.95
10	North America	Canada	Corner Bicycle Supply	Accessories	Helmets	Sport-100	Black	4/1/2020	34.95
11	North America	Canada	Metal Processing Company	Bikes	Road Bikes	Road-650	Red	4/1/2020	782.95
12	North America	United States	Excellent Riding Supplies	Bikes	Road Bikes	Road-650	Red	4/1/2020	782.95
13	North America	United States	Excellent Riding Supplies	Bikes	Road Bikes	Road-450	Red	4/1/2020	1457.95
14									

Completed (4.76 s) Columns: 11 Rows: 99+

Query settings

- Properties Name: UserQuery
- Applied steps

- Data cleaning steps :
 - Change the data type of selected columns.
 - OrderQty → ABC 123 → Whole number
 - UnitPrice → ABC 123 → decimal
 - ListPrice → ABC 123 → decimal

- Create a column (**UnitPrice*OrderQty**).

- Select columns to multiply → add column → standard → multiply .

- After selecting multiply , a new column is created with the column name as multiplication.
- Change the column name to cost by right click on the selected column.

Microsoft Azure | Data Factory > blobb2asql-adf

Search factory and documentation

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

inputData bikeData_PQ

Refresh

Home Transform Add column View Help

Column from examples+ Custom function Conditional column Rank column Duplicate column General Format Extract from text Statistics Standard Scientific Rounding Information From number Date and time column

Queries [2]

Table.AddColumn("Changed column type", "Multiplication", each [UnitPrice] * [OrderQty], type number)

	Business Segment	Category	Model	Color	123 S...	123 ListPrice	1.2 UnitPrice	123 OrderQty	1.2 Multiplication
1	Components	Road Frames	LL Road Frame	Red	4/1/2020	337.22	183.94	1	
2	Bikes	Mountain Bikes	Mountain-100	Silver	4/1/2020	3399.99	2039.99	1	
3	Clothing	Jerseys	Long-Sleeve Logo Jersey	Multi	4/1/2020	49.99	28.84	6	
4	Components	Mountain Frames	HL Mountain Frame	Black	4/1/2020	1349.6	714.7	2	
5	Bikes	Road Bikes	Road-450	Red	4/1/2020	1457.99	874.79	2	
6	Bikes	Road Bikes	Road-650	Black	4/1/2020	782.99	419.46	2	
7	Bikes	Road Bikes	Road-650	Red	4/1/2020	782.99	419.46	4	
8	Bikes	Road Bikes	Road-450	Red	4/1/2020	1457.99	874.79	1	
9	Bikes	Road Bikes	Road-650	Black	4/1/2020	782.99	419.46	2	
10	Accessories	Helmets	Sport-100	Black	4/1/2020	34.99	20.19	5	
11	ny	Bikes	Road-650	Red	4/1/2020	782.99	419.46	1	
12	Bikes	Road Bikes	Road-650	Red	4/1/2020	782.99	419.46	1	
13	Bikes	Road Bikes	Road-450	Red	4/1/2020	1457.99	874.79	1	

Completed (2.04 s) Columns: 12 Rows: 99+

Settings Parameters

Copy preview data
Remove columns
Remove other columns
Duplicate column
Add column from examples...
Remove duplicates
Remove errors
Split column
Replace values...
Replace errors...
Change type
Transform column
Group by...
Fill
Unpivot columns
Unpivot other columns
Unpivot only selected columns
Rename...
Move
Drill down
Add as new query

	Business Segment	Category	Model	Color	123 S...	123 ListPrice	1.2 UnitPrice	123 OrderQty	1.2 Cost
1	Components	Components	Road Frames	LL Road Frame	Red	4/1/2020	337.22	183.94	1 183.94
2	nt Store	Bikes	Mountain Bikes	Mountain-100	Silver	4/1/2020	3399.99	2039.99	1 2039.99
3	x Repair	Clothing	Jerseys	Long-Sleeve Logo Jersey	Multi	4/1/2020	49.99	28.84	6 173.04
4		Components	Mountain Frames	HL Mountain Frame	Black	4/1/2020	1349.6	714.7	2 1429.4
5	Bikes Store	Bikes	Road Bikes	Road-450	Red	4/1/2020	1457.99	874.79	2 1749.58
6	Bikes Store	Bikes	Road Bikes	Road-650	Black	4/1/2020	782.99	419.46	2 838.92
7	Bikes Store	Bikes	Road Bikes	Road-650	Red	4/1/2020	782.99	419.46	4 1677.84
8	ore	Bikes	Road Bikes	Road-450	Red	4/1/2020	1457.99	874.79	1 874.79
9	ore	Bikes	Road Bikes	Road-650	Black	4/1/2020	782.99	419.46	2 838.92
10	Supply	Accessories	Helmets	Sport-100	Black	4/1/2020	34.99	20.19	5 100.95
11	ng Company	Bikes	Road Bikes	Road-650	Red	4/1/2020	782.99	419.46	1 419.46
12	g Supplies	Bikes	Road Bikes	Road-650	Red	4/1/2020	782.99	419.46	1 419.46
13	g Supplies	Bikes	Road Bikes	Road-450	Red	4/1/2020	1457.99	874.79	1 874.79

- Create a column (**ListPrice * OrderQty**).

- Click **add columns** → select columns (**ListPrice , OrderQty**) → **standard** → **multiply** → rename column from “ multiplication ” to “ Sales ”.

	Business Segment	Category	Model	Color	123 S...	1.2 ListPrice	1.2 UnitPrice	123 OrderQty	1.2 Cost	1.2 Sales
1	nts	Components	Road Frames	LL Road Frame	Red	4/1/2020	337.22	183.94	1	183.94
2		Bikes	Mountain Bikes	Mountain-100	Silver	4/1/2020	3399.99	2039.99	1	2039.99
3		Clothing	Jerseys	Long-Sleeve Logo Jersey	Multi	4/1/2020	49.99	28.84	6	173.04
4		Components	Mountain Frames	HL Mountain Frame	Black	4/1/2020	1349.6	714.7	2	1429.4
5		Bikes	Road Bikes	Road-450	Red	4/1/2020	1457.99	874.79	2	1749.58
6		Bikes	Road Bikes	Road-650	Black	4/1/2020	782.99	419.46	2	838.92
7		Bikes	Road Bikes	Road-650	Red	4/1/2020	782.99	419.46	4	1677.84
8		Bikes	Road Bikes	Road-450	Red	4/1/2020	1457.99	874.79	1	874.79
9		Bikes	Road Bikes	Road-650	Black	4/1/2020	782.99	419.46	2	838.92
10		Accessories	Helmets	Sport-100	Black	4/1/2020	34.99	20.19	5	100.95
11	ny	Bikes	Road Bikes	Road-650	Red	4/1/2020	782.99	419.46	1	419.46
12		Bikes	Road Bikes	Road-650	Red	4/1/2020	782.99	419.46	1	419.46
13		Bikes	Road Bikes	Road-450	Red	4/1/2020	1457.99	874.79	1	874.79

- Create a column name Profit (Sales - Cost) .
- Add column → select columns (Sales , Cost) → Standard → Subtract → Rename the column name from “ Subtraction ” to “ Profit ” .

	Business Segment	Category	Model	Color	123 Sales	1.2 ListPrice	1.2 UnitPrice	123 OrderQty	1.2 Cost	1.2 Sales	1.2 Profit
1	Components	Road Frames	LL Road Frame	Red	4/1/2020	337.22	183.94	1	183.94	62028.2468	11409475.72
2	Bikes	Mountain Bikes	Mountain-100	Silver	4/1/2020	3399.99	2039.99	1	2039.99	6935945.6	14149259664
3	Accessories	Jerseys	Long-Sleeve Logo Jersey	Multi	4/1/2020	49.99	28.84	6	173.04	1441.7116	249473.7753
4	Components	Mountain Frames	HL Mountain Frame	Black	4/1/2020	1349.6	714.7	2	1429.4	964559.12	1378740806
5	Bikes	Road Bikes	Road-450	Red	4/1/2020	1457.99	874.79	2	1749.58	1275435.072	2231475693
6	Bikes	Road Bikes	Road-650	Black	4/1/2020	782.99	419.46	2	838.92	328432.9854	275529000.1
7	Bikes	Road Bikes	Road-650	Red	4/1/2020	782.99	419.46	4	1677.84	328432.9854	551058000.2
8	Bikes	Road Bikes	Road-450	Red	4/1/2020	1457.99	874.79	1	874.79	1275435.072	1115737846
9	Bikes	Road Bikes	Road-650	Black	4/1/2020	782.99	419.46	2	838.92	328432.9854	275529000.1
10	Accessories	Helmets	Sport-100	Black	4/1/2020	34.99	20.19	5	100.95	706.4481	71315.9357
11	Bikes	Road Bikes	Road-650	Red	4/1/2020	782.99	419.46	1	419.46	328432.9854	137764500.1
12	Bikes	Road Bikes	Road-650	Red	4/1/2020	782.99	419.46	1	419.46	328432.9854	137764500.1
13	Bikes	Road Bikes	Road-450	Red	4/1/2020	1457.99	874.79	1	874.79	1275435.072	1115737846
14											

- There are null values in the color column, replace null values with (N/A) .
- Color column → right click → Replace values
 - Value to find → (null)
 - Replace with → N/A

The screenshot shows the Power BI Data Editor interface. A context menu is open over the 'Color' column header. The 'Replace values...' option is highlighted with a green arrow. Other options like 'Remove duplicates' and 'Replace errors...' are also visible.

The screenshot shows the 'Replace values' dialog box. It has two tabs: 'Basic' (selected) and 'Advanced'. The 'Value to find' field contains 'null' and the 'Replace with' field contains 'N/A'. The 'OK' button is highlighted with a green arrow. The background shows the Power BI Data Editor interface with a table of data and a query editor.

- The SalesDate column data type is in string, we cannot convert it to date datatype(not supported). So we need to extract the last 4 characters and create a column for the year.

Screenshot of Power BI Editor showing the 'Add column' ribbon tab selected. The formula bar shows an error message: 'UserQuery : Expression.Error: Unsupported constant null.' The formula being edited is 'Table.TransformColumnTypes(#"Replaced value 1", {{"SalesDate", type date}})'. The 'Format' ribbon tab is also visible, showing various options like 'Length', 'First characters', and 'Last characters'.

UnitPrice	OrderQty	Cost	Sales	Profit
183.94	1	183.94	62028.2468	11409475.72
2039.99	1	2039.99	6935945.6	1414925964
28.84	6	173.04	1441.7116	249473.7753
714.7	2	1429.4	964559.12	1378740806
874.79	2	1749.58	1275435.072	2231475693
419.46	2	838.92	328432.9854	275529000.1
419.46	4	1677.84	328432.9854	551058000.2
874.79	1	874.79	1275435.072	1115737846
419.46	2	838.92	328432.9854	275529000.1
20.19	5	100.95	706.4481	71315.9357
419.46	1	419.46	328432.9854	137764500.1

- Extract the last 4 characters (year).

■ Add column → Extract → Last Characters → 4 → ok

Screenshot of Power BI Editor showing the 'Add column' ribbon tab selected. The formula bar shows an error message: 'UserQuery : Expression.Error: Unsupported constant null.' The formula being edited is 'Table.TransformColumnTypes(#"Replaced value 1", {{"SalesDate", type date}})'. The 'Format' ribbon tab is also visible, showing various options like 'Length', 'First characters', and 'Last characters'. A green arrow points to the 'Last characters' button, and a green number '4' is placed next to it.

UnitPrice	OrderQty	Cost	Sales	Profit
183.94	1	183.94	62028.2468	11409475.72
2039.99	1	2039.99	6935945.6	1414925964
28.84	6	173.04	1441.7116	249473.7753
714.7	2	1429.4	964559.12	1378740806
874.79	2	1749.58	1275435.072	2231475693
419.46	2	838.92	328432.9854	275529000.1
419.46	4	1677.84	328432.9854	551058000.2
874.79	1	874.79	1275435.072	1115737846
419.46	2	838.92	328432.9854	275529000.1
20.19	5	100.95	706.4481	71315.9357
419.46	1	419.46	328432.9854	137764500.1

Screenshot of Power BI Editor showing the 'Add column' ribbon tab selected. The formula bar shows an error message: 'UserQuery : Expression.Error: Unsupported constant null.' The formula being edited is 'Table.TransformColumnTypes(#"Replaced value 1", {{"SalesDate", type date}})'. A green arrow points to the 'Last characters' button, and a green number '4' is placed next to it. A modal dialog box titled 'Insert last characters' is open, asking 'Enter how many ending characters to keep.' with a 'Count *' input field containing '4'. A green arrow points to the 'OK' button.

UnitPrice	OrderQty	Cost	Sales	Profit
183.94	1	183.94	62028.2468	11409475.72
2039.99	1	2039.99	6935945.6	1414925964
28.84	6	173.04	1441.7116	249473.7753
714.7	2	1429.4	964559.12	1378740806
874.79	2	1749.58	1275435.072	2231475693
419.46	2	838.92	328432.9854	275529000.1
419.46	4	1677.84	328432.9854	551058000.2
874.79	1	874.79	1275435.072	1115737846
20.19	5	100.95	706.4481	71315.9357
419.46	1	419.46	328432.9854	137764500.1

- Rename column from “Last characters” to “Year”.

	Category	Model	Color	SalesDate	1.2 ListPrice	1.2 UnitPrice	1.2 OrderQty	1.2 Cost	1.2 Sales	1.2 Profit	1.2 Year
1	Road Frames	LL Road Frame	Red	4/1/2020	337.22	183.94	1	183.94	62028.2468	11409475.72	2020
2	Mountain Bikes	Mountain-100	Silver	4/1/2020	3399.99	2039.99	1	2039.99	6935945.6	14149259664	2020
3	Jerseys	Long-Sleeve Logo Jersey	Multi	4/1/2020	49.99	28.84	6	173.04	1441.7116	249473.7753	2020
4	Mountain Frames	HL Mountain Frame	Black	4/1/2020	1349.6	714.7	2	1429.4	964559.12	1378740806	2020
5	Road Bikes	Road-450	Red	4/1/2020	1457.99	874.79	2	1749.58	1275435.072	2231475693	2020
6	Road Bikes	Road-650	Black	4/1/2020	782.99	419.46	2	838.92	328432.9854	275529000.1	2020
7	Road Bikes	Road-650	Red	4/1/2020	782.99	419.46	4	1677.84	328432.9854	551058000.2	2020
8	Road Bikes	Road-450	Red	4/1/2020	1457.99	874.79	1	874.79	1275435.072	1115737846	2020
9	Road Bikes	Road-650	Black	4/1/2020	782.99	419.46	2	838.92	328432.9854	275529000.1	2020
10	Helmets	Sport-100	Black	4/1/2020	34.99	20.19	5	100.95	706.4481	71315.9357	2020
11	Road Bikes	Road-650	Red	4/1/2020	782.99	419.46	1	419.46	328432.9854	137764500.1	2020
12											

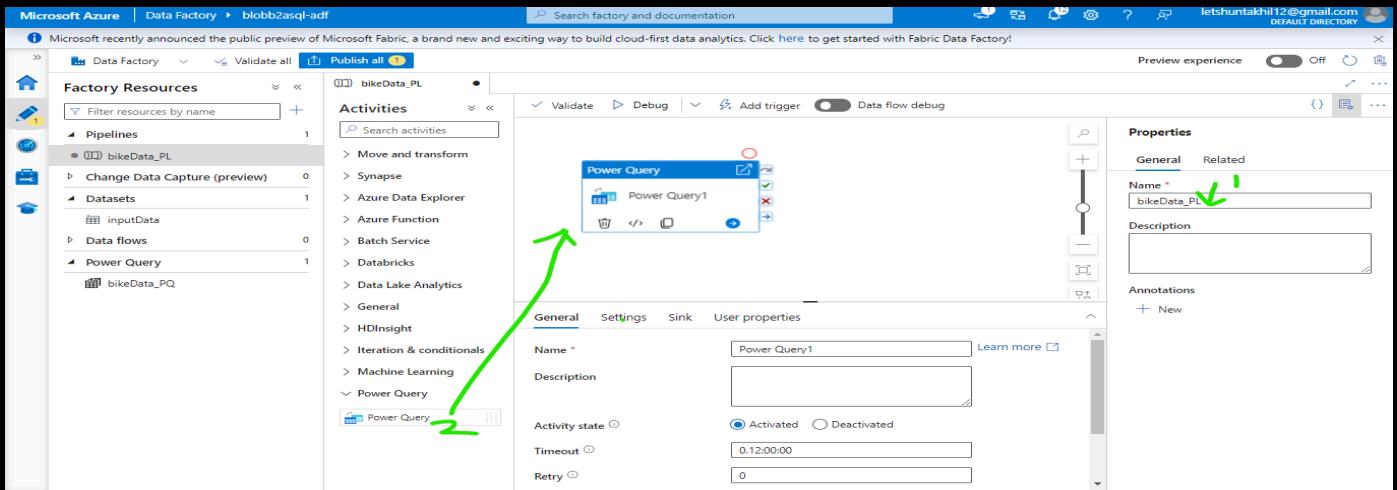
- Save the transformation by publish all → publish.

The screenshot shows the Microsoft Azure Data Factory interface. In the top navigation bar, there is a 'Publishing' button with a green arrow pointing to it. Below the navigation bar, there is a 'Pending changes' section with a table showing a single change for 'bikeData_PQ'. At the bottom right of the Power Query Editor window, there is a 'Publish' button with a green arrow pointing to it.

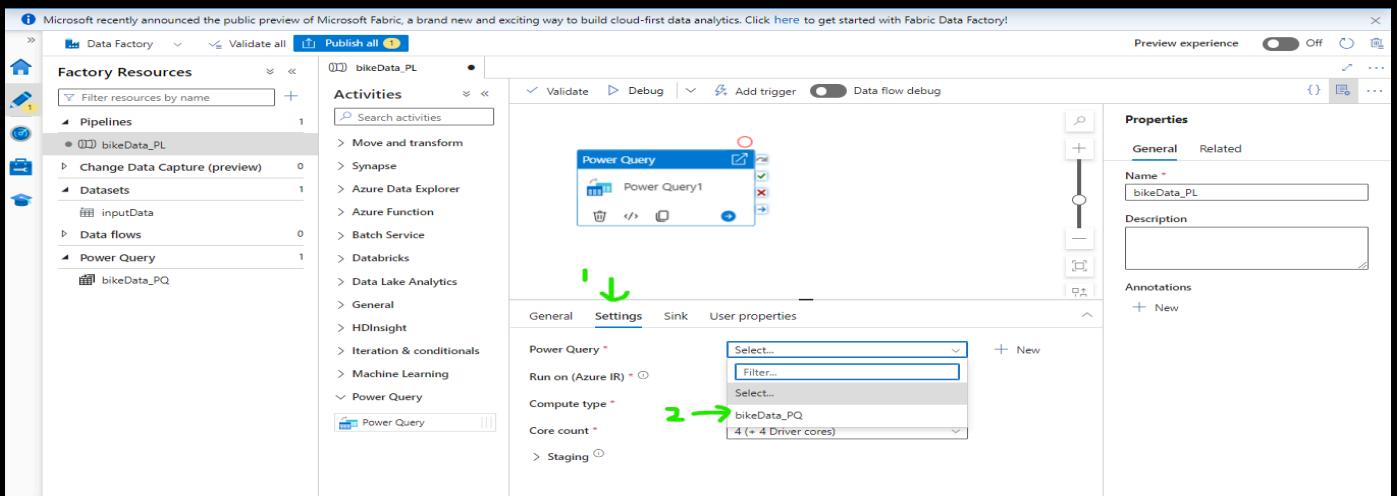
- Create pipeline name as bikeData_PL by closing the power query editor .

The screenshot shows the Microsoft Azure Data Factory 'Factory Resources' page. On the left, there is a sidebar with icons for Home, Transform, Add column, View, Help, Column from examples, Custom column, Invoke custom function, Conditional column, Duplicate column, and General. The main area shows 'Pipelines' with a count of 0. There is a 'New pipeline' button with a green arrow pointing to it. Below the pipelines, there are sections for 'Change Data Capture (preview)', 'Datasets' (with 'inputData'), 'Data flows', and 'Power Query' (with 'bikeData_PQ').

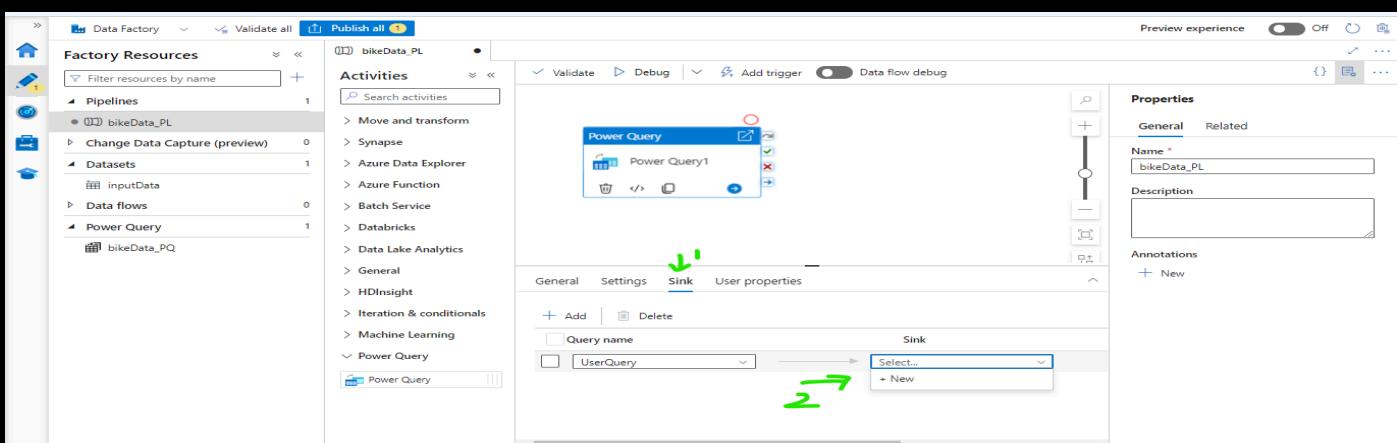
- Drag and drop Power Query Activity to the workspace.



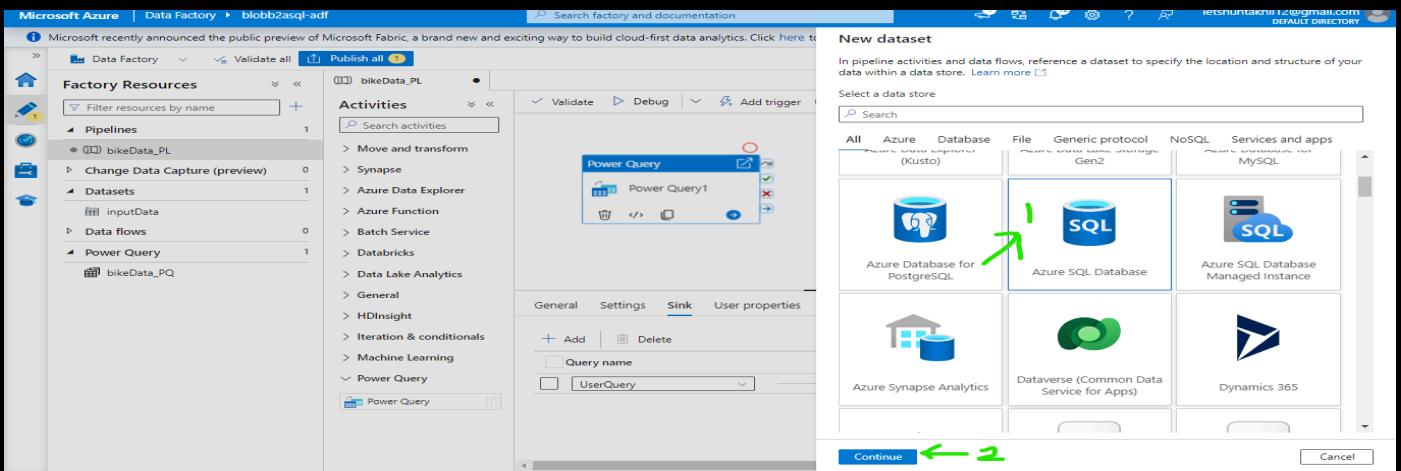
- Power Query → Settings → select the power query (bikeData_PQ).



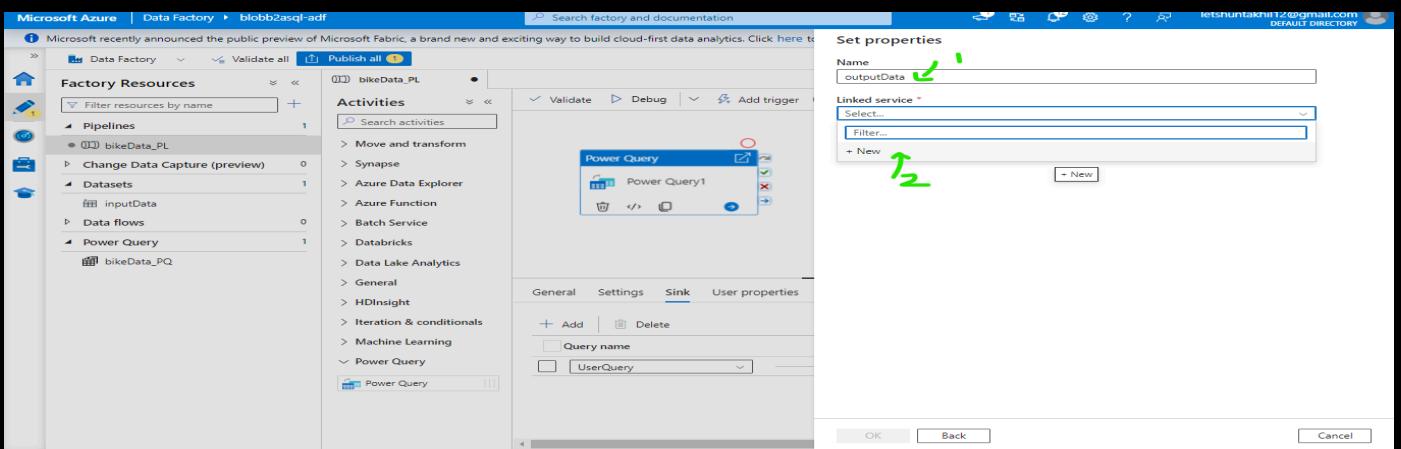
- Power Query → Sink → Queryname → sink → + New .



- New Dataset → Azure SQL Database → continue.
- Here we are selecting the database to store our output data.

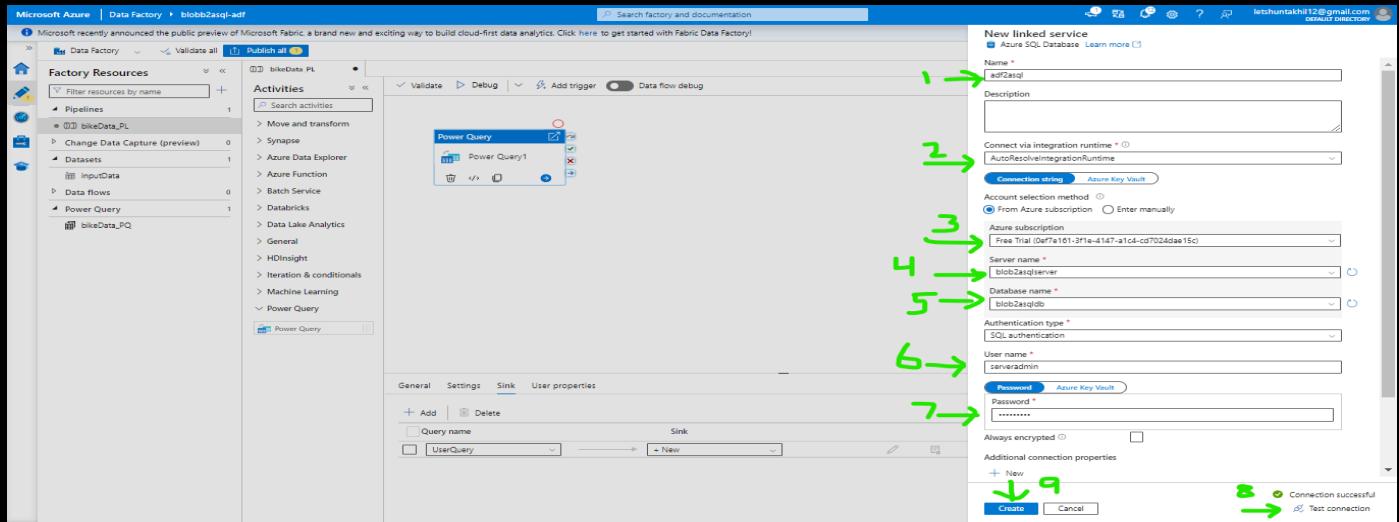


- Give the dataset name and create a new linked service to connect adf with azure sql database.

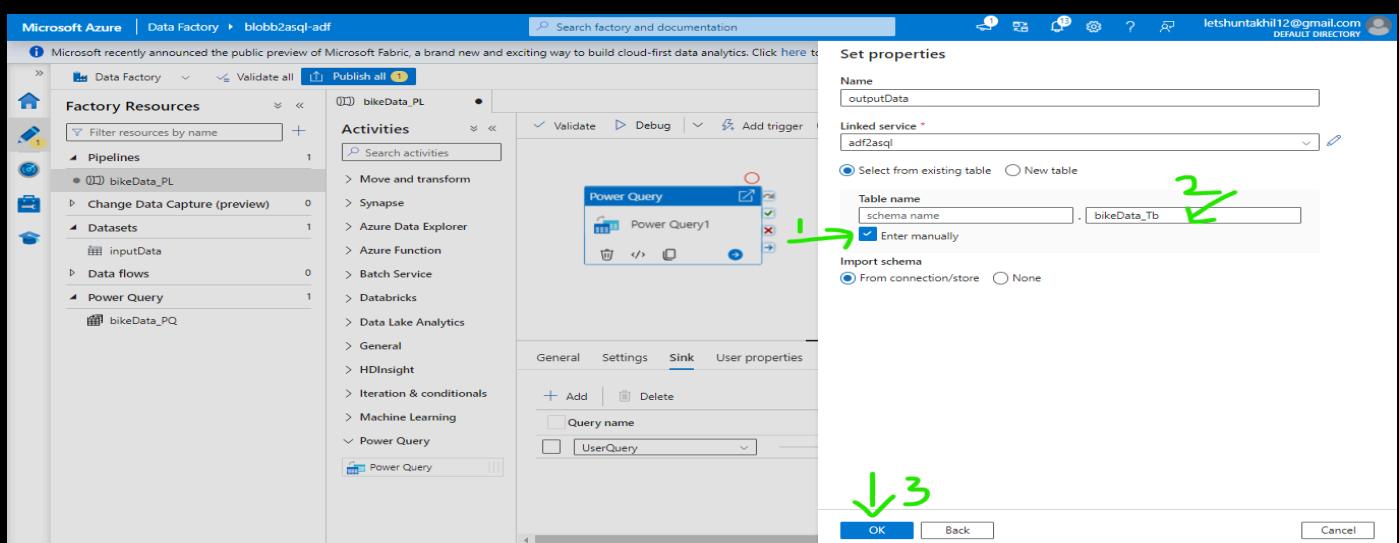


- Fill the form to create a linked service. Provide the details of azure sql database , so that output data can be stored.
- New linked service :
 - **Name** : Give a name for linked service
 - **Connect via integration runtime** : Autoresolveintegrationruntime (Adf - Sql).
 - **Subscription** : select from list .
 - **Server name** : select from list .

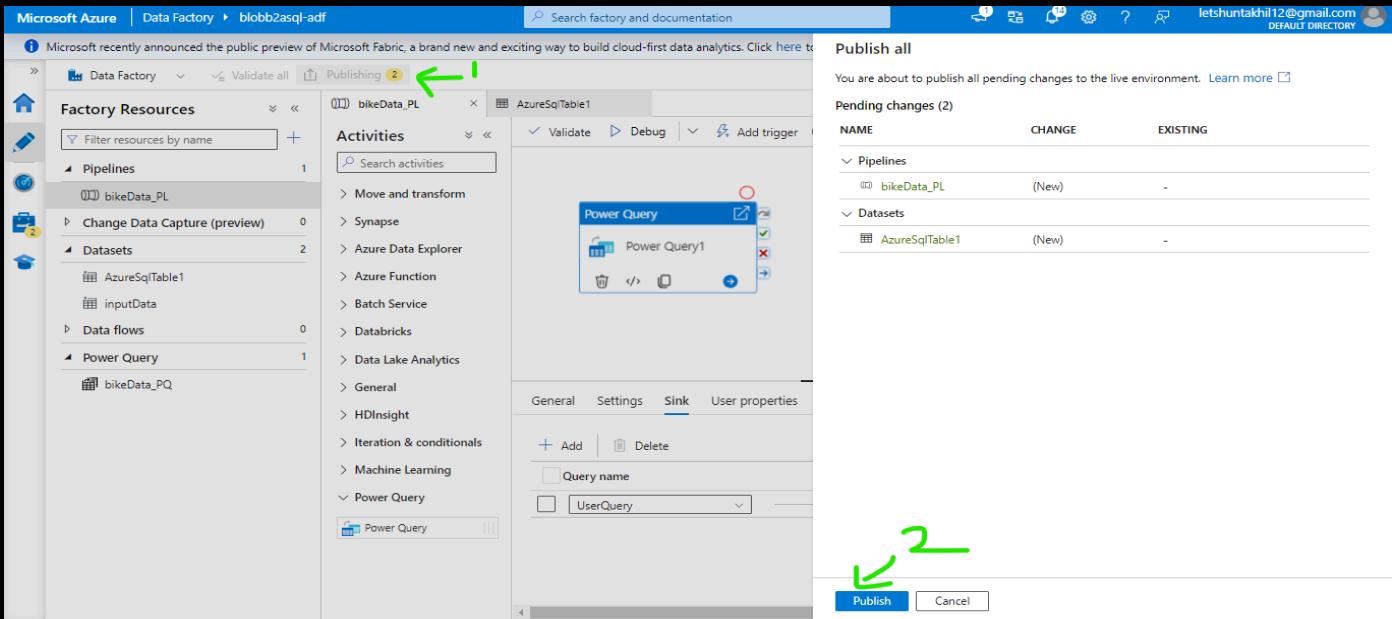
- **Database name** : select
- **Authentication** : Sql Authentication.
- **Username** : Provide the username we have created.
- **Password** : Provide the password.
- Click **test connection** → successful → **create**



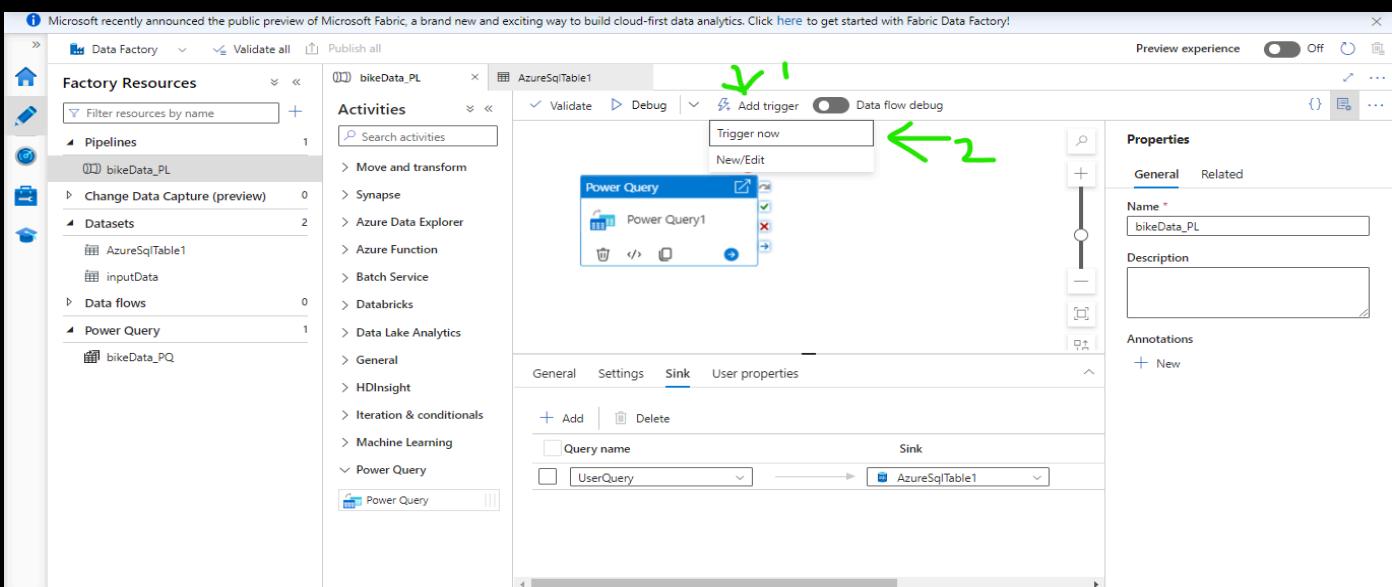
- Select **New Table**.
- Provide the **name of the table** you want to store by clicking **enter manually** and then click **ok**.

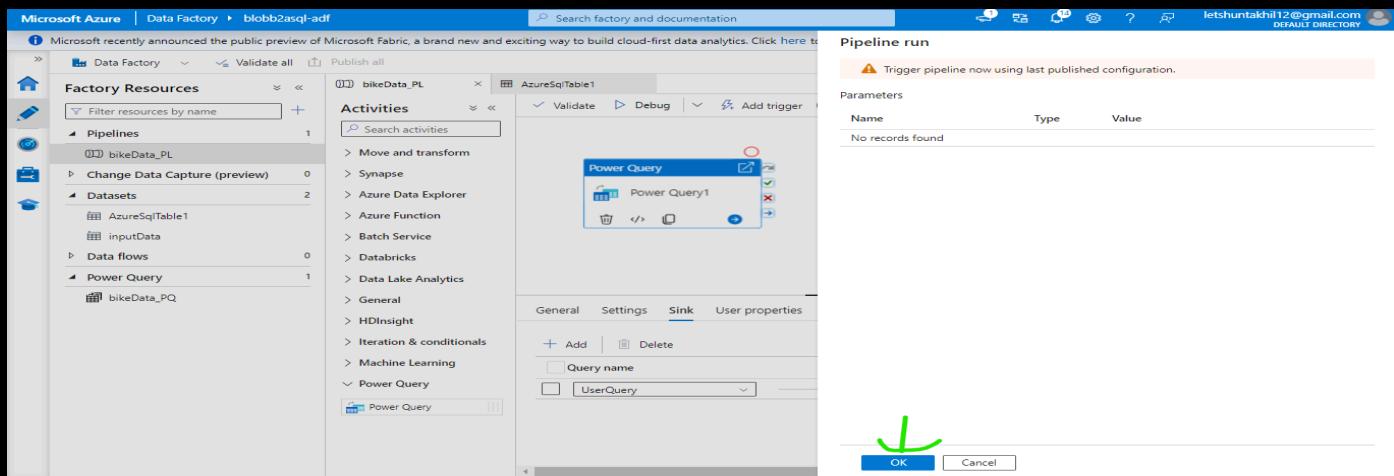


- Publish all → publish.
- Now we have made a successful connection between adf and azure sql database after publishing.

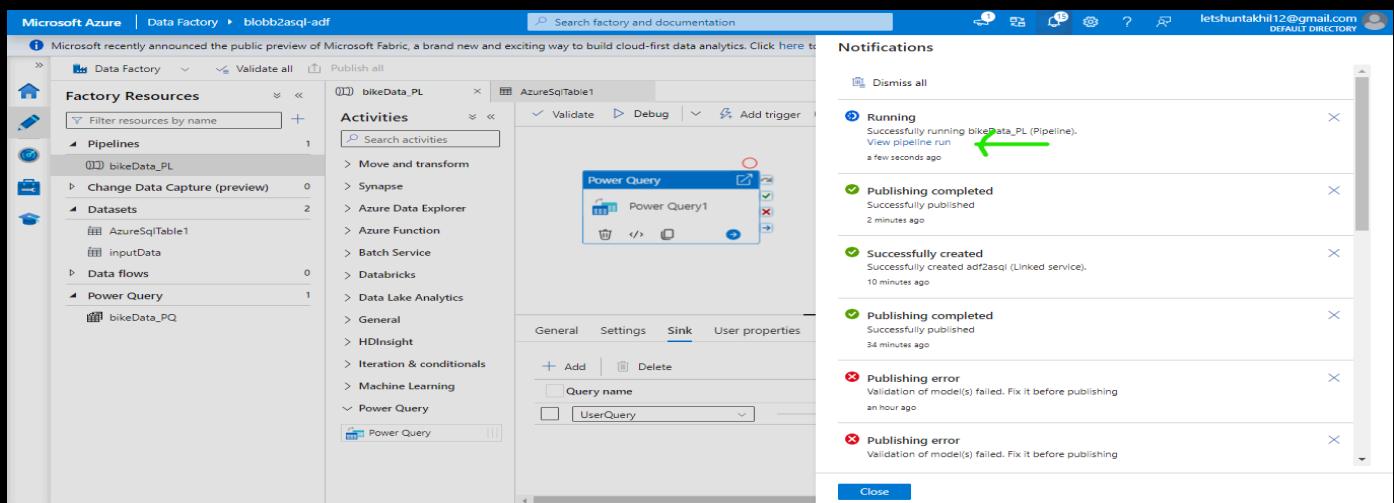


- Let's run the pipeline → Add Trigger → Trigger now → ok

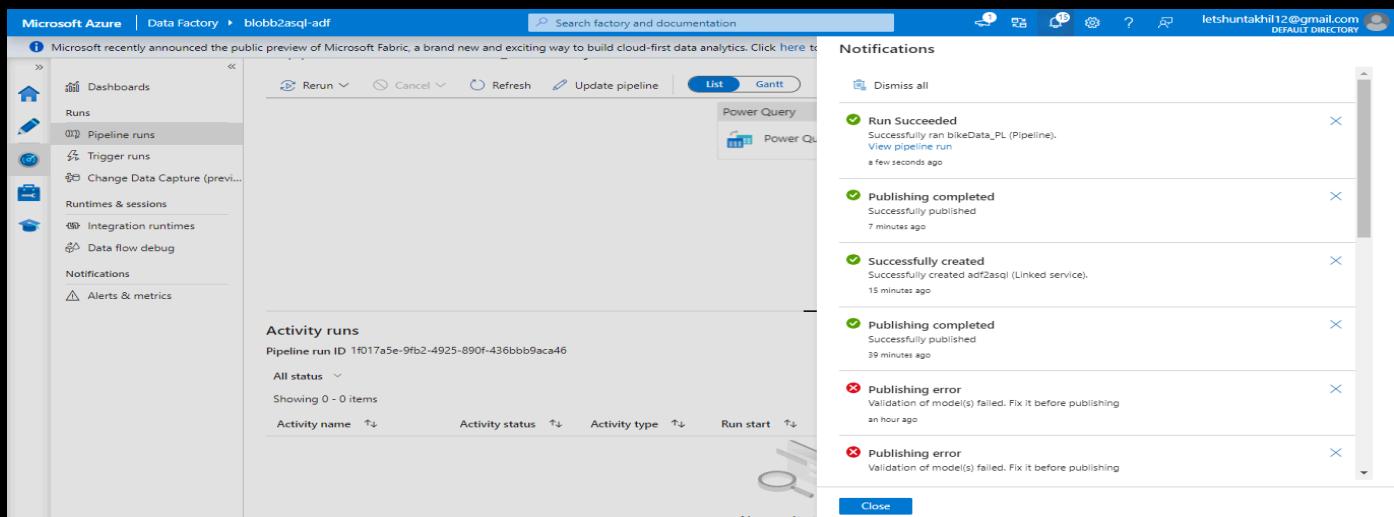




- See the status of the pipeline → view pipeline run .



- bikeData_PL has successfully executed.



- See the Table and Query the data.

The screenshot shows the Microsoft Azure SQL Database Query editor (preview) interface. The left sidebar shows the database structure with 'Tables' expanded, displaying the table 'dbo.bikeData_TB'. A context menu is open over this table, with the option 'Select Top 1000 Rows' highlighted. The main area shows the query editor with the following SQL command:

```
1 SELECT TOP (1000) * FROM [dbo].[bikeData_TB]
```

The results pane displays a table with the following data:

	Cost	Sales	Profit	Year
1	183.94	337.22	153.28000000000003	4-01
2	2039.99	3399.99	1359.9999999999998	4-01
3	173.04	299.94	126.9	4-01
4	1429.4	2699.2	1269.7999999999997	4-01

The status bar at the bottom indicates 'Query succeeded | 1s'.