

**UseCase:** To get the insights from the equity stocks based on daily OHLC(open,high,low,close) data feed from NSE (which hosted in AWS S3) and daily customers transactions (which hosted in Azure ADLS Gen2). So that client can show these insights in the trading terminal while user open any stock for trading which help users to look in to some trading , prices based statistics.

To achieve the use case described, you can follow these general steps:

Data Ingestion: Set up connections to fetch daily OHLC data from NSE hosted in AWS S3 and daily customer transactions from Azure ADLS Gen2 using **Azure Data Factory**

Data Processing: Process the OHLC data and customer transactions separately to handle missing values, outliers, and any other data quality issues. Merge or join the relevant information from both datasets based on common identifiers (e.g., **ISIN**) using **Azure Databricks**.

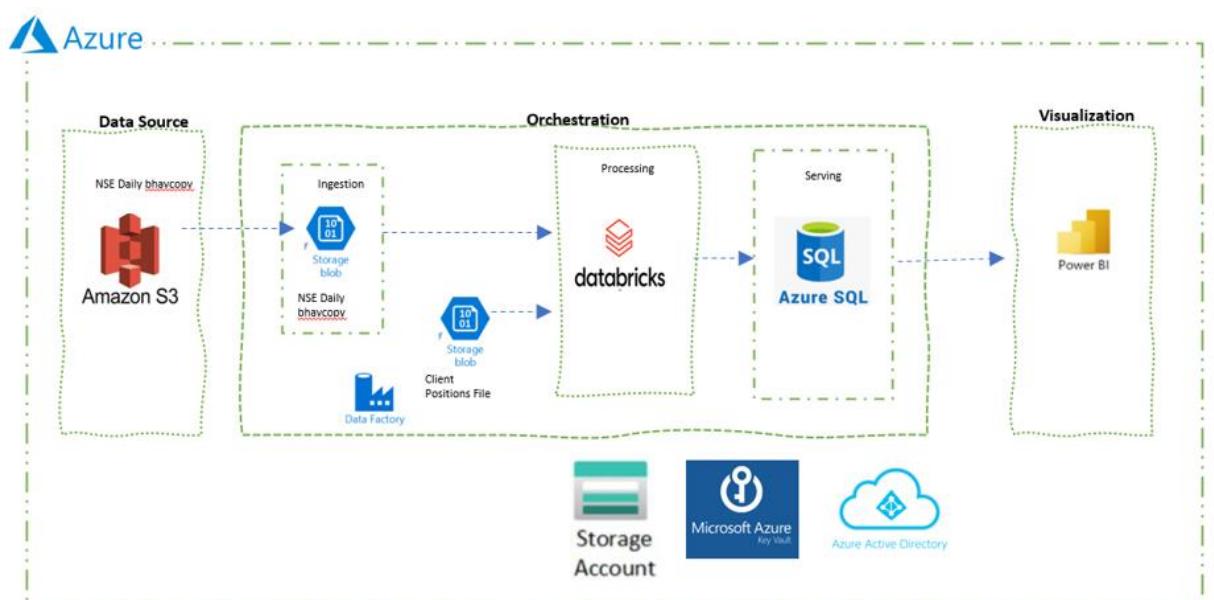
Data Storage: Store the processed data in a suitable data storage solution, such as a database (e.g., **Azure SQL Database**) and in data lake (e.g., **Azure Data Lake Storage Gen2**).

Security and Compliance: Implement security measures to protect sensitive financial data. This may include encryption, access controls (**Azure KeyVault, Azure Active Directory, Azure SAS key**)

Scaling and Performance: Design the system to scale as the amount of data and users grow. Consider using scalable cloud services and optimizing queries for performance (**Scaling in Databricks**)

Documentation: Document the entire process, including data sources, processing steps, and any configurations made. This documentation is crucial for troubleshooting, maintenance, and future enhancements.

## **Architecture:**



## **Implementation:**

Resource group: rg-stock-analysis-dev

Screenshot of the Microsoft Azure portal showing the creation of a new resource group named "rg-stock-analysis-dev".

**Basics** tab selected. Project details: Subscription is "Free Trial" and Resource group is "rg-stock-analysis-dev". Region is set to "(US) East US".

Bottom navigation bar includes "Review + create", "< Previous", "Next : Tags >", and a search bar.

Create Storage account: stockanalysissestorageacct

Create a storage account

Subscription \*: Free Trial

Resource group \*: rg-stock-analysis-dev

Storage account name \*: stockanalysistorageacct1

Region \*: (US) East US

Review < Previous Next : Advanced > Give feedback

Enable Hierarchical Namespace under Advanced TAB to be eligible to create sub-folders inside containers.

Advanced

Permitted scope for copy operations (preview): From any storage account

Hierarchical Namespace

Hierarchical namespace, complemented by Data Lake Storage Gen2 endpoint, enables file and directory semantics, accelerates big data analytics workloads, and enables access control lists (ACLs). [Learn more](#)

Enable hierarchical namespace

Access protocols

Blob and Data Lake Gen2 endpoints are provisioned by default. [Learn more](#)

Review < Previous Next : Networking > Give feedback

Created containers – abhisai-stockanalysis and creates folders with nse\_daily\_bhavcopy  
zerodha\_daily\_positions and load with old data.

The screenshot shows the Microsoft Azure Storage account overview for the container 'abhisai-stockanalysis'. It displays two blobs: 'nse\_daily\_bhavcopy' and 'zerodha\_daily\_positions'. The interface includes tabs for Overview, Diagnose and solve problems, Access Control (IAM), Settings (Shared access tokens, Manage ACL, Access policy, Properties, Metadata), Authentication method (Access key, Microsoft Entra user account), Location (abhisai-stockanalysis), and a search bar for blobs by prefix. The status bar at the bottom shows the date and time as 11/9/2023, 12:48 AM.

Created S3 bucket in AWS to upload data - abhisai-stockanalysis and created bhavcopy folder in it.

The screenshot shows the AWS S3 console for the bucket 'abhisai-stockanalysis'. It displays one object named 'bhavcopy/'. The interface includes a sidebar with options like Buckets, Access Points, Object Lambda Access Points, Multi-Region Access Points, Batch Operations, IAM Access Analyzer for S3, Block Public Access settings, Storage Lens, Dashboards, AWS Organizations settings, and Feature spotlight. The main area shows the Objects tab with a list of objects, including a 'Create folder' button and an 'Upload' button. The status bar at the bottom shows the date and time as 11/9/2023, 12:43 AM.

## Created ADF for ingestion and scheduling: **adf-stock-analysis-dev**

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	rg-stock-analysis-dev
Name	adf-stock-analysis-dev
Region	East US
Version	V2

**Networking**

Connect via	Public endpoint
-------------	-----------------

Previous    Next    Create    Give feedback

## Create key-vault for storing all keys: **kv-stockanalysis-dev**

**Basics**

Subscription	Free Trial
Resource group	rg-stock-analysis-dev
Key vault name	kv-stockanalysis-dev
Region	East US
Pricing tier	Standard
Soft-delete	Enabled
Purge protection during retention period	Disabled
Days to retain deleted vaults	90 days

**Access configuration**

Azure Virtual Machines for deployment	Disabled
---------------------------------------	----------

Previous    Next    Create    Give feedback

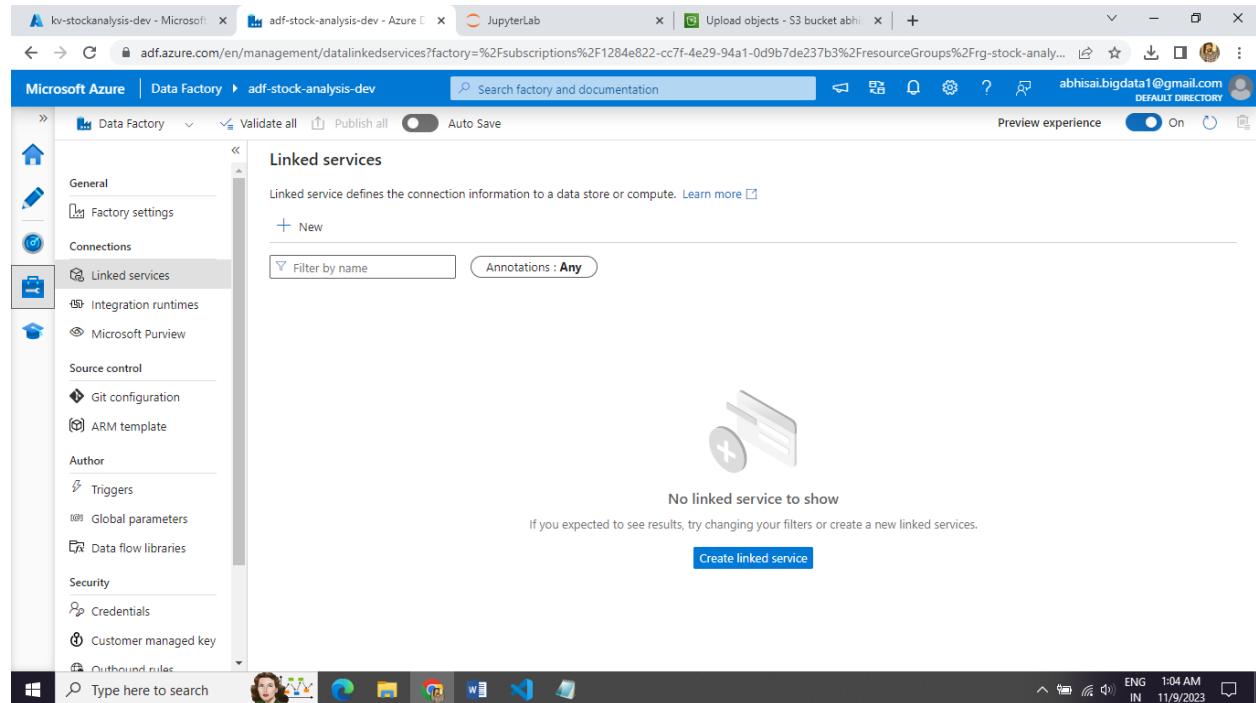
## Building Pipeline in ADF:

- Once daily transactions file loaded to Azure we need to trigger pipeline
- Once step a is completed we need to ingest data from AWS S3 for bhavcopy for that day
- Once step b is completed we need to trigger Azure Databricks for processing
- Processing output should be loaded back to Azure SQL for reporting and PowerBI connectivity

## Steps to ingest data from AWS:

- Create Linked Service for AWS S3(source) and Azure ADLS Gen2 (sink)
- Create Datasets for AWS S3(source) and Azure ADLS Gen2 (sink)
- Add metadata activity to check latest file
- IF condition for validations and then Copy Step for ingesting data.

## Creation of Linked Services – Source and Sink:



## Source:

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the navigation menu is open with 'General' selected under 'Linked services'. In the center, a 'New linked service' dialog is displayed. The 'Data store' tab is selected. A search bar at the top right contains the text 'Search'. Below it, tabs for 'All', 'Azure', 'Database', 'File', 'Generic protocol', 'NoSQL', and 'Services and apps' are shown. Under 'Services and apps', there are six options: 'Amazon Marketplace Web Service', 'Amazon RDS for Oracle', 'Amazon RDS for SQL Server', 'Amazon Redshift', 'Amazon S3', and 'Amazon S3 Compatible'. The 'Amazon S3' option is highlighted with a blue border. At the bottom of the dialog are 'Continue' and 'Cancel' buttons.

This screenshot shows the same 'New linked service' dialog as the previous one, but with more detailed configuration. The 'Name' field is filled with 'Is\_AWS\_S3\_bhavcopy'. The 'Connect via Integration runtime' dropdown is set to 'AutoResolveIntegrationRuntime'. The 'Authentication type' dropdown is set to 'Access key'. The 'Access key ID' section is active, showing 'Azure Key Vault' as the provider. The 'Access key ID' field is empty. The 'Secret access key' section is also active, showing 'Azure Key Vault' as the provider. The 'Secret access key' field is empty. At the bottom are 'Create', 'Back', 'Test connection', and 'Cancel' buttons.

Here if we see while creating linked service for AWS S3 we need to provide access key ID and value.

Go to IAM in AWS and create user and attach policy to read S3 contents and create a access for that user so using those ID/value we can access AWS S3 from outside.

kv-stockanalysis-dev - Microsoft | adf-stock-analysis-dev - Azure | JupyterLab | Create user | IAM | Global

us-east-1.console.aws.amazon.com/iamv2/home?region=us-east-1#/users/create

Services Search [Alt+S]

EC2 S3

IAM > Users > Create user

Step 1 Specify user details

Step 2 Set permissions

Step 3 Review and create

## Specify user details

User details

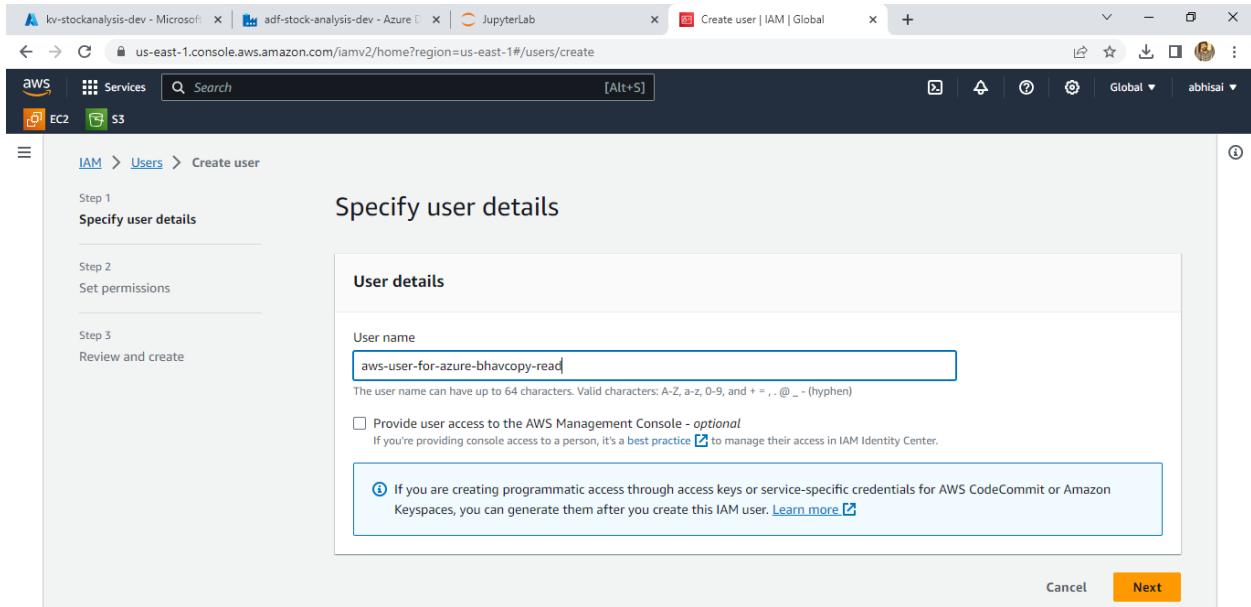
User name: aws-user-for-azure-bhavcopy-read

The user name can have up to 64 characters. Valid characters: A-Z, a-z, 0-9, and + = , . @ \_ ~ (hyphen)

Provide user access to the AWS Management Console - optional  
If you're providing console access to a person, it's a best practice [to manage their access in IAM Identity Center.](#)

If you are creating programmatic access through access keys or service-specific credentials for AWS CodeCommit or Amazon Keyspaces, you can generate them after you create this IAM user. [Learn more](#)

Cancel Next

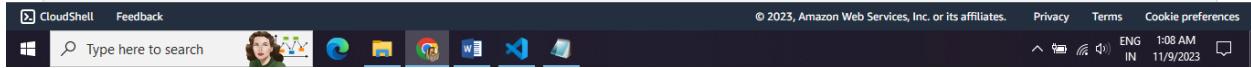


CloudShell Feedback

Type here to search

© 2023, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

1:08 AM ENG IN 11/9/2023



kv-stockanalysis-dev - Microsoft | adf-stock-analysis-dev - Azure | JupyterLab | Create user | IAM | Global

us-east-1.console.aws.amazon.com/iamv2/home?region=us-east-1#/users/create

Services Search [Alt+S]

EC2 S3

IAM > Users > Create user

Step 2 Set permissions

Step 3 Review and create

Add user to an existing group or create a new one. Using groups is a best-practice way to manage user's permissions by job functions. [Learn more](#)

## Permissions options

Add user to group  
Add user to an existing group, or create a new group. We recommend using groups to manage user permissions by job function.

Copy permissions  
Copy all group memberships, attached managed policies, and inline policies from an existing user.

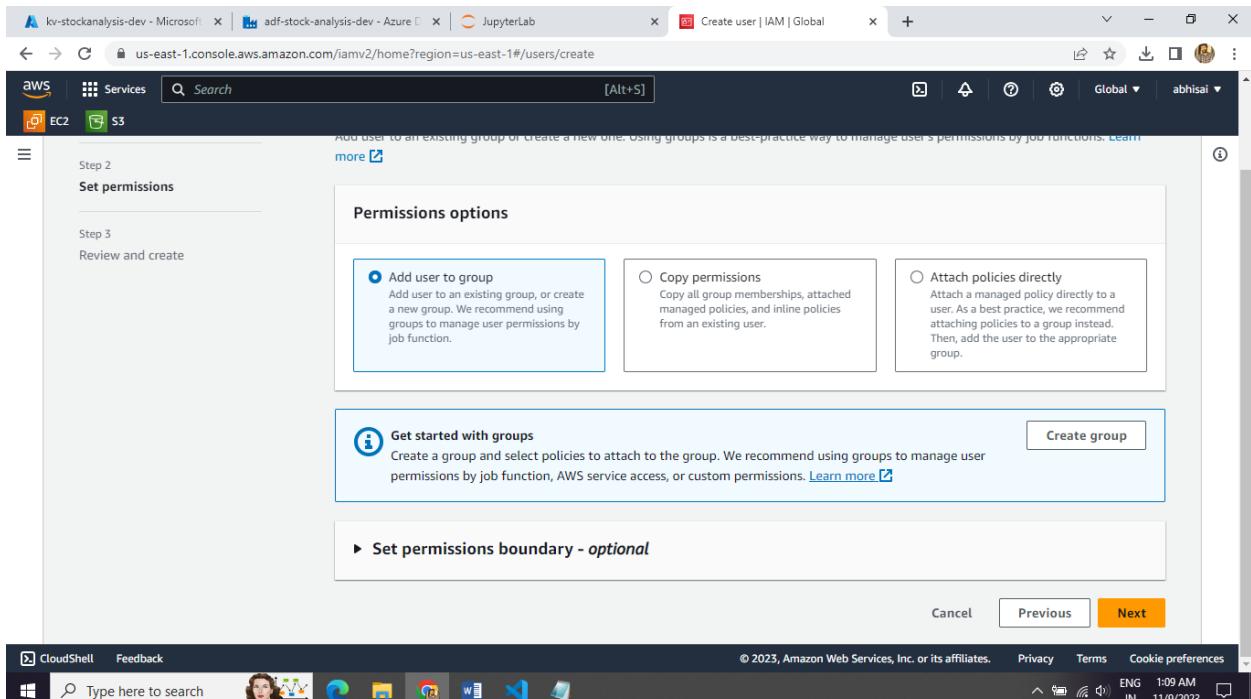
Attach policies directly  
Attach a managed policy directly to a user. As a best practice, we recommend attaching policies to a group instead. Then, add the user to the appropriate group.

Get started with groups  
Create a group and select policies to attach to the group. We recommend using groups to manage user permissions by job function, AWS service access, or custom permissions. [Learn more](#)

Set permissions boundary - optional

Create group

Cancel Previous Next

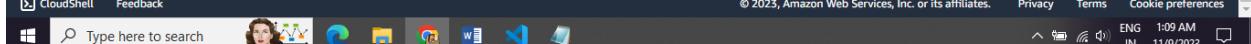


CloudShell Feedback

Type here to search

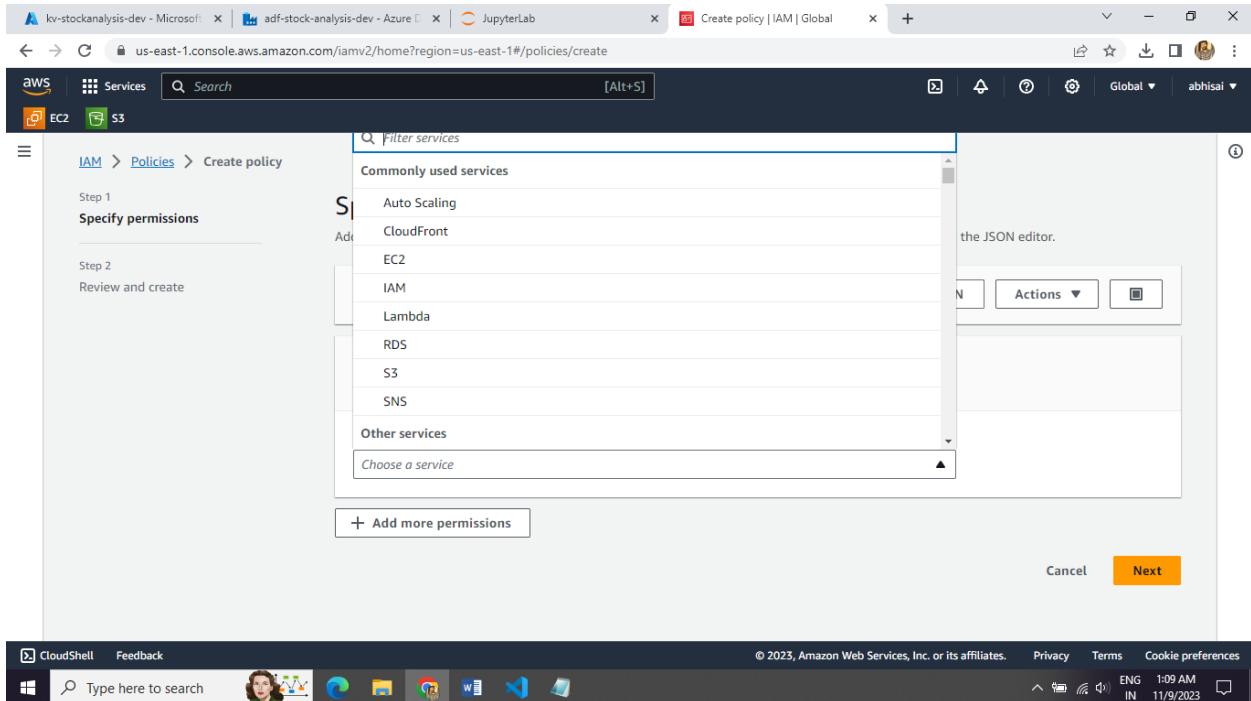
© 2023, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

1:09 AM ENG IN 11/9/2023

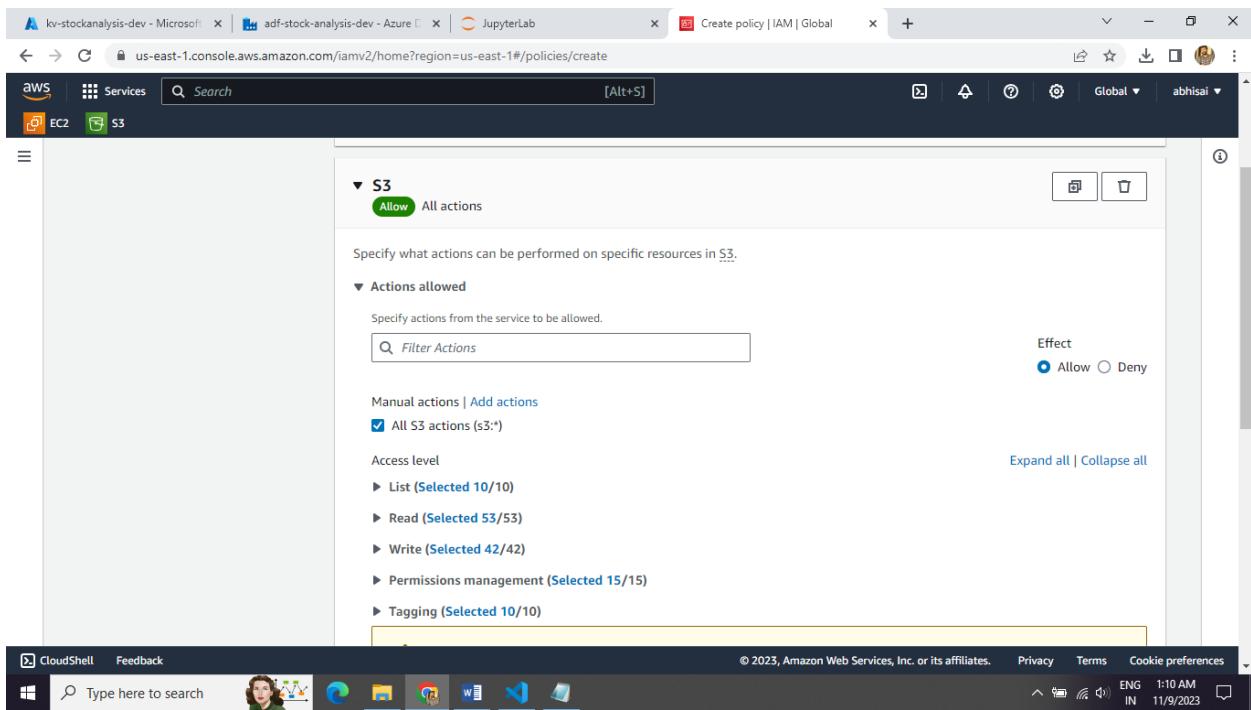


Creating policy for that user:

Select S3 in the below dropdown



The screenshot shows the 'Create policy' wizard in the AWS IAM console. The current step is 'Step 1: Specify permissions'. A dropdown menu is open, showing 'Commonly used services' and 'Other services'. Under 'Commonly used services', 'S3' is selected. Other options include Auto Scaling, CloudFront, EC2, IAM, Lambda, RDS, and SNS. Below the dropdown, there's a button '+ Add more permissions'. At the bottom right, there are 'Cancel' and 'Next' buttons.



The screenshot shows the 'Create policy' wizard in the AWS IAM console, Step 2: Set actions. The 'S3' service is expanded, showing the 'Allow All actions' section. It says 'Specify what actions can be performed on specific resources in S3.' and 'Actions allowed'. Under 'Actions allowed', there's a search bar 'Filter Actions' and a list of actions: 'All S3 actions (s3:\*)'. The 'Effect' is set to 'Allow'. At the bottom right, there are 'Expand all' and 'Collapse all' buttons.

We can actually provide only the required access as per documentation.

The screenshot shows the 'Create policy' wizard in the AWS IAM console. The current step is 'Specify permissions'. The 'Policy details' section contains a 'Policy name' field with the value 'aws-policy-for-azure-bhavcopy-read'. Below it is a large text area for 'Description - optional', which is currently empty. At the bottom of the page, there is a 'Permissions defined in this policy' section with an 'Edit' button.

Now attach this policy to that user:

The screenshot shows the 'User Details' page for the user 'aws-user-for-azure-bhavcopy-read'. The 'Permissions' tab is selected. The 'Permissions policies (0)' section indicates that no policies are attached. On the right side, there is a 'Add permissions' button with options for 'Add permissions' and 'Create inline policy'. The left sidebar shows the 'Access management' section with 'Users' selected.

kv-stockanalysis-dev - Microsoft | adf-stock-analysis-dev - Azure | JupyterLab | Add permissions | IAM | Global | +

us-east-1.console.aws.amazon.com/iamv2/home?region=us-east-1#/users/details/aws-user-for-azure-bhavcopy-read/add-permissions

Services Search [Alt+S] Global abhisai

EC2 S3

Add user to group  
Add user to an existing group, or create a new group. We recommend using groups to manage user permissions by job function.

Copy permissions  
Copy all group memberships, attached managed policies, inline policies, and any existing permissions boundaries from an existing user.

Attach policies directly  
Attach a managed policy directly to a user. As a best practice, we recommend attaching policies to a group instead. Then, add the user to the appropriate group.

Permissions policies (1/1140)

Filter by Type: aws-policy-for-azure-bhavcopy-read, All types, 1 match

Policy name	Type	Attached entities
aws-policy-for-azure-bhavcopy-read	Customer managed	0

There are no permissions selected.

Cancel Next

CloudShell Feedback © 2023, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences ENG IN 11/9/2023

kv-stockanalysis-dev - Microsoft | adf-stock-analysis-dev - Azure | JupyterLab | Add permissions | IAM | Global | +

us-east-1.console.aws.amazon.com/iamv2/home?region=us-east-1#/users/details/aws-user-for-azure-bhavcopy-read/add-permissions

Services Search [Alt+S] Global abhisai

EC2 S3

IAM > Users > aws-user-for-azure-bhavcopy-read > Add permissions

Step 1 Add permissions

Step 2 Review

The following policies will be attached to this user. [Learn more](#)

User details

User name  
aws-user-for-azure-bhavcopy-read

Permissions summary (1)

Name	Type	Used as
aws-policy-for-azure-bhavcopy-read	Customer managed	Permissions policy

Cancel Previous Add permissions

CloudShell Feedback © 2023, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences ENG IN 11/9/2023

The screenshot shows the AWS Identity and Access Management (IAM) console. On the left, there's a navigation sidebar with options like Dashboard, Access management (Users, Roles, Policies, Identity providers, Account settings), and Access reports (Archive rules, Analyzers). The main content area displays a user named "aws-user-for-azure-bhavcopy-read". A green banner at the top says "1 policy added". Below it, the "Summary" section shows the ARN (arn:aws:iam::787017564811:user/aws-user-for-azure-bhavcopy-read), Console access (Disabled), and Last console sign-in (November 09, 2023, 01:09 (UTC+05:30)). The "Permissions" tab is selected, showing "Permissions policies (1)". At the bottom right of this section, there's a "Create access key" link. The status bar at the bottom right shows the date (11/9/2023) and time (1:13 AM).

Then we can create a access key using option “Create access key”

The screenshot shows the "Create access key" wizard, Step 1: Access key best practices & alternatives. It includes a note: "Avoid using long-term credentials like access keys to improve your security. Consider the following use cases and alternatives." Below this, there are four options under "Use case":

- Command Line Interface (CLI)  
You plan to use this access key to enable the AWS CLI to access your AWS account.
- Local code  
You plan to use this access key to enable application code in a local development environment to access your AWS account.
- Application running on an AWS compute service  
You plan to use this access key to enable application code running on an AWS compute service like Amazon EC2, Amazon ECS, or AWS Lambda to access your AWS account.
- Third-party service  
You plan to use this access key to enable access for a third-party application or service that monitors or manages your AWS resources.

The status bar at the bottom right shows the date (11/9/2023) and time (1:13 AM).

The screenshot shows the AWS IAM 'Create access key' page. A green header bar at the top says 'Access key created'. Below it, a message states: 'This is the only time that the secret access key can be viewed or downloaded. You cannot recover it later. However, you can create a new access key any time.' The main content area is titled 'Retrieve access keys' and contains three steps: Step 1 (Access key best practices & alternatives), Step 2 - optional (Set description tag), and Step 3 (Retrieve access keys). Step 3 is currently active, showing a table with one row. The table has two columns: 'Access key' and 'Secret access key'. The 'Access key' column contains the value 'AKIA3OPPHAKF3KCTGCT2'. The 'Secret access key' column contains '\*\*\*\*\*' followed by a 'Show' link. To the right of the table is a section titled 'Access key best practices' with three bullet points: 'Never store your access key in plain text, in a code repository, or in code.', 'Disable or delete access key when no longer needed.', and 'Enable least-privilege permissions.'.

Once done access key ID will be shown for the user as below:

The screenshot shows the AWS IAM 'Security credentials' tab for the user 'aws-user-for-azure-bhavcopy-read'. The left sidebar shows the 'Identity and Access Management (IAM)' navigation pane with options like Dashboard, Access management, Users, Policies, Identity providers, and Account settings. The 'Users' section is expanded, showing the user 'aws-user-for-azure-bhavcopy-read'. The main content area displays the user's summary information, including ARN, Console access status (Disabled), and two access keys. Access Key 1 is listed as 'AKIA3OPPHAKF3KCTGCT2 - Active' with a note 'Never used. Created today.' Access Key 2 is listed with a 'Create access key' link. Below this, there are tabs for Permissions, Groups, Tags, Security credentials (which is selected), and Access Advisor. Under the 'Console sign-in' section, there are links for 'Console sign-in link' and 'Console password', along with a 'Enable console access' button.

Now we can add this access key ID and value in Azure Key-Vault so that we can give them in ADF.

Key vaults

kv-stockanalysis-dev | Secrets

The operation is not allowed by RBAC. If role assignments were recently changed, please wait several minutes for role assignments to become effective.

You are unauthorized to view these contents.

First in order to add secrets in azure Key Vault we need to provide access to the user.

kv-stockanalysis-dev - Microsoft | adf-stock-analysis-dev - Azure | JupyterLab | aws-user-for-azure-bhavcopy-re... | +

portal.azure.com/#@abhisaiibigdata1gmail.onmicrosoft.com/resource/subscriptions/1284e822-cc7f-4e29-94a1-0d9b7de237b3/resourceGroups/rg-stock-anal...

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

Home > Key vaults > kv-stockanalysis-dev

## Key vaults

Default Directory (abhisaiibigdata1gmail.onmicrosoft.com)

+ Create ...

Filter for any field...

Name ↑

kv-stockanalysis-dev

...

Search

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Access policies

Events

Objects

Keys

Secrets

Certificates

Settings

Access configuration

Networking

Page 1 of 1

kv-stockanalysis-dev | Access control (IAM)

Key vault

Add role assignment

Role Deny assignments Classic administrators

Add co-administrator Add role assignment

**My access**

View my level of access to this resource.

**View my access**

**Check access**

Review the level of access a user, group, service principal, or managed identity has to this resource. [Learn more](#)

**Check access**

**Grant access to this resource**

Grant access to resources by assigning a role. [Learn more](#)

**Add role assignment**

**View access to this resource**

View the role assignments that grant access to this and other resources. [Learn more](#)

**View**

Type here to search

Microsoft Defender for Cloud

Add role assignment - Microsoft | adf-stock-analysis-dev - Azure | JupyterLab | aws-user-for-azure-bhavcopy-re... | +

portal.azure.com/#view/Microsoft\_Azure\_AD/AddRoleAssignmentsLandingBlade/scope/%2Fsubscriptions%2F1284e822-cc7f-4e29-94a1-0d9b7de237b3%2Fre...

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

Home > Key vaults > kv-stockanalysis-dev | Access control (IAM) >

## Add role assignment

Role Members Review + assign

A role definition is a collection of permissions. You can use the built-in roles or you can create your own custom roles. [Learn more](#)

Assignment type

Job function roles Privileged administrator roles

Grant access to Azure resources based on job function, such as the ability to create virtual machines.

key vault Type : All Category : All

Name ↑	Description ↑	Type ↑	Category ↑	Details
Key Vault Administrator	Perform all data plane operations on a key vault and all objects in it, including certificates, keys, and se...	BuiltinRole	Security	View
Key Vault Certificates Officer	Perform any action on the certificates of a key vault, except manage permissions. Only works for key v...	BuiltinRole	Security	View
Key Vault Contributor	Lets you manage key vaults, but not access to them.	BuiltinRole	Security	View
Key Vault Crypto Officer	Perform any action on the keys of a key vault, except manage permissions. Only works for key vaults t...	BuiltinRole	Security	View
Key Vault Crypto Service Encryption User	Read metadata of keys and perform wrap/unwrap operations. Only works for key vaults that use the '...	BuiltinRole	Security	View
Key Vault Crypto User	Perform cryptographic operations using kevs. Only works for key vaults that use the 'Azure role-based...	BuiltinRole	Security	View

Review + assign Previous Next Feedback

Type here to search

ENG US 1:16 AM 11/9/2023

**Add role assignment**

**Selected role**: Key Vault Administrator

**Assign access to**:  User, group, or service principal  Managed identity

**Members**: + Select members

Name	Object ID	Type
Abhi Sai	4816978f-0ba7-4658-9d9f-65392fa44562	User

**Description**: Optional

**Review + assign** | **Previous** | **Next**

Here actually this key vault should be accessed from ADF as well and Databricks as well so giving same permission for all now itself so that we no need to revisit this page. In Production ready we need to different access to 3<sup>rd</sup> party applications as this is just test providing same access.

**Select members**

Select  adf-stock-analysis-dev

adf-stock-analysis-dev

Selected members:  
No members selected. Search for and add one or more members you want to assign to the role for this resource.

Learn more about RBAC

**Review + assign** | **Previous** | **Next**

**Add role assignment**

**Role** **Members** **Review + assign**

**Selected role** Key Vault Administrator

**Assign access to**  User, group, or service principal  Managed identity

**Members** [+ Select members](#)

Name	Object ID	Type
Abhi Sai	4816978f-0ba7-4658-9d9f-65392fa44562	User
adf-stock-analysis-dev	1396c8d2-192d-42f6-a071-b79dbebf1b...	App

**Description** Optional

**Review + assign** **Previous** **Next** **Select** **Close**

**Add role assignment**

**Role** **Members** **Review + assign**

**Selected role** Key Vault Administrator

**Assign access to**  User, group, or service principal  Managed identity

**Members** [+ Select members](#)

Name	Object ID	Type
Abhi Sai	4816978f-0ba7-4658-9d9f-65392fa44562	User
adf-stock-analysis-dev	1396c8d2-192d-42f6-a071-b79dbebf1b...	App
AzureDatabricks	6c47e086-1fb4-44f5-b425-287966bffb25	App

**Description** Optional

**Review + assign** **Previous** **Next** **Feedback**

Now I can go inside secrets to add key and values for AWS S3.

Screenshot of the Microsoft Azure portal showing the creation of a secret in a Key Vault.

**Create a secret**

Upload options: Manual

Name: aws-s3-bhavcopy-id

Secret value:  (redacted)

Content type (optional):

Set activation date:

Set expiration date:

Enabled: Yes

Tags: 0 tags

**Create** **Cancel**

**kv-stockanalysis-dev | Secrets**

Key vault: kv-stockanalysis-dev

Overview

The secret 'aws-s3-bhavcopy-value' has been successfully created.

Name	Type	Status	Expiration date
aws-s3-bhavcopy-value		✓ Enabled	
aws-s3-bhavcopy-id		✓ Enabled	

Give feedback

Now we can refer to this access key from ADF by creating a linked service to Azure Key-vault

The screenshot shows the Microsoft Azure portal interface for a Data Factory named 'adf-stock-analysis-dev'. The left sidebar navigation bar is visible, showing various options like General, Factory settings, Connections, and Linked services. The main content area is titled 'New linked service' and is specifically for creating a linked service to an Azure Key Vault. The 'Name' field is populated with 'ls\_Azure\_Key\_Vault'. The 'Azure key vault selection method' section has the radio button selected for 'From Azure subscription'. The 'Azure subscription' dropdown shows 'Free Trial (1284e822-cc7f-4e29-94a1-0d9b7de237b3)'. The 'Azure key vault name' dropdown shows 'kv-stockanalysis-dev'. The 'Authentication method' dropdown is set to 'System Assigned Managed Identity'. A note at the bottom indicates that managed identity access is granted to the Data Factory service. At the bottom right of the dialog are 'Create' and 'Cancel' buttons, along with a 'Test connection' link.

This screenshot shows the same Azure Data Factory interface, but the configuration for the linked service is different. The 'Authentication type' dropdown is now set to 'Access key'. The 'AKV linked service' tab is selected, showing 'Secret name' as 'aws-s3-bhavcopy-id'. The 'Secret access key' tab is also visible, showing 'AKV linked service' as 'ls\_Azure\_Key\_Vault' and 'Secret name' as 'aws-s3-bhavcopy-value'. The rest of the interface is identical to the first screenshot, including the 'Create', 'Cancel', and 'Test connection' buttons.

Check the connection using “Test connection” option next to Cancel button.

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the navigation pane is open with 'Data Factory' selected. Under 'Connections', 'Linked services' is also selected. In the main content area, a 'New linked service' dialog is open. The 'Secret access key' tab is selected, showing an 'AKV linked service' named 'ls\_Azure\_Key\_Vault'. The 'Secret name' is set to 'aws-s3-bhavcopy-value'. The 'Service URL' is set to 'https://s3.amazonaws.com'. A 'Test connection' button is visible, and a message indicates 'Connection successful'. The status bar at the bottom right shows the date and time as 11/9/2023.

## Sink:

The screenshot shows the Microsoft Azure Data Factory interface. The left navigation pane is open with 'Data Factory' selected. Under 'Connections', 'Linked services' is selected. In the main content area, a 'New linked service' dialog is open. The 'Data store' tab is selected, showing a grid of options. The 'Azure Data Lake Storage Gen2' icon is highlighted with a blue border. Below the grid, a 'Continue' button is visible. The status bar at the bottom right shows the date and time as 11/9/2023.

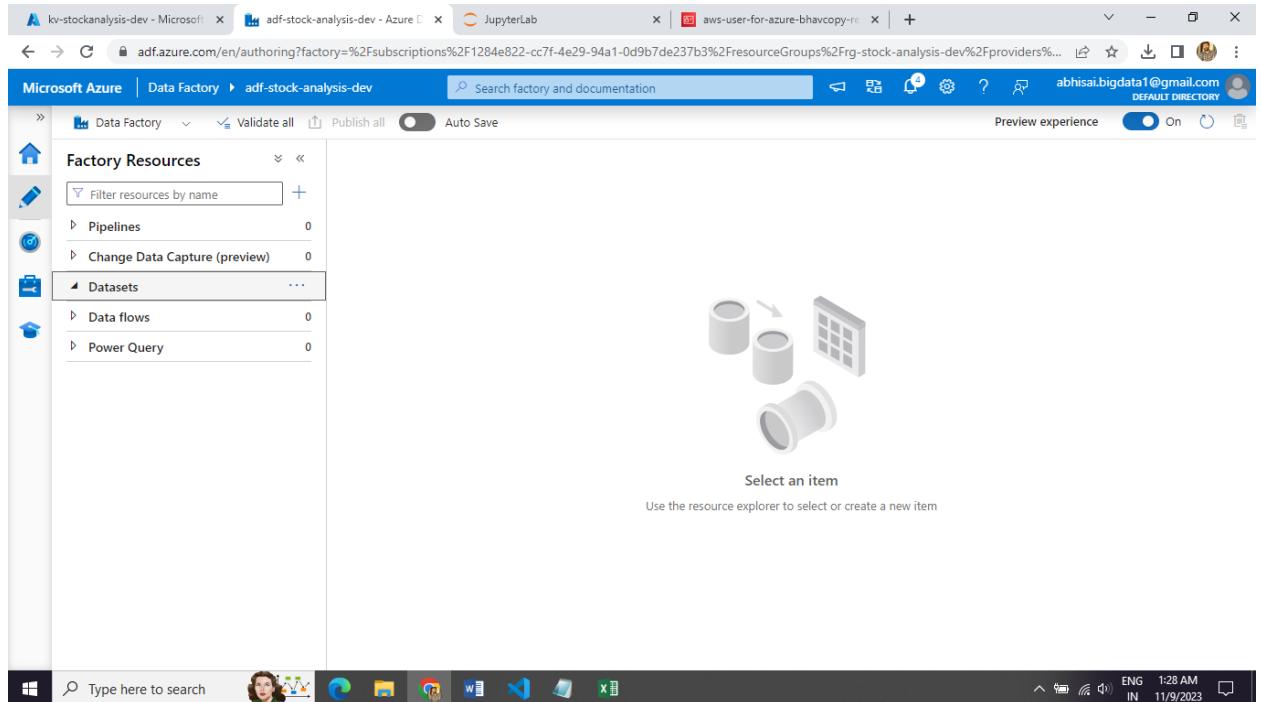
Fill in the storage account details in ADLS and Test Connection

The screenshot shows the Microsoft Azure Data Factory interface. The left sidebar has 'Data Factory' selected under 'Connections'. The main area is titled 'Linked services' with the sub-instruction 'Linked service defines the connection information to a data store or compute.' Below this is a table showing two existing linked services: 'Is\_AWS\_S3\_bhavcopy' (Amazon S3) and 'Is\_Azure\_Key\_Vault' (Azure Key Vault). A 'New' button is visible. On the right, a 'New linked service' form is open, pre-filled with 'Is\_Azure\_dl\_bhavcopy' as the name, 'AutoResolveIntegrationRuntime' as the connect via integration runtime, and 'Account key' as the authentication type. The 'From Azure subscription' radio button is selected, and a dropdown shows 'Free Trial (1284e822-cc7f-4e29-94a1-0d9b7de237b3)'. The 'Storage account name' field contains 'stockanalysistorageacct'. A success message 'Connection successful' is displayed at the bottom.

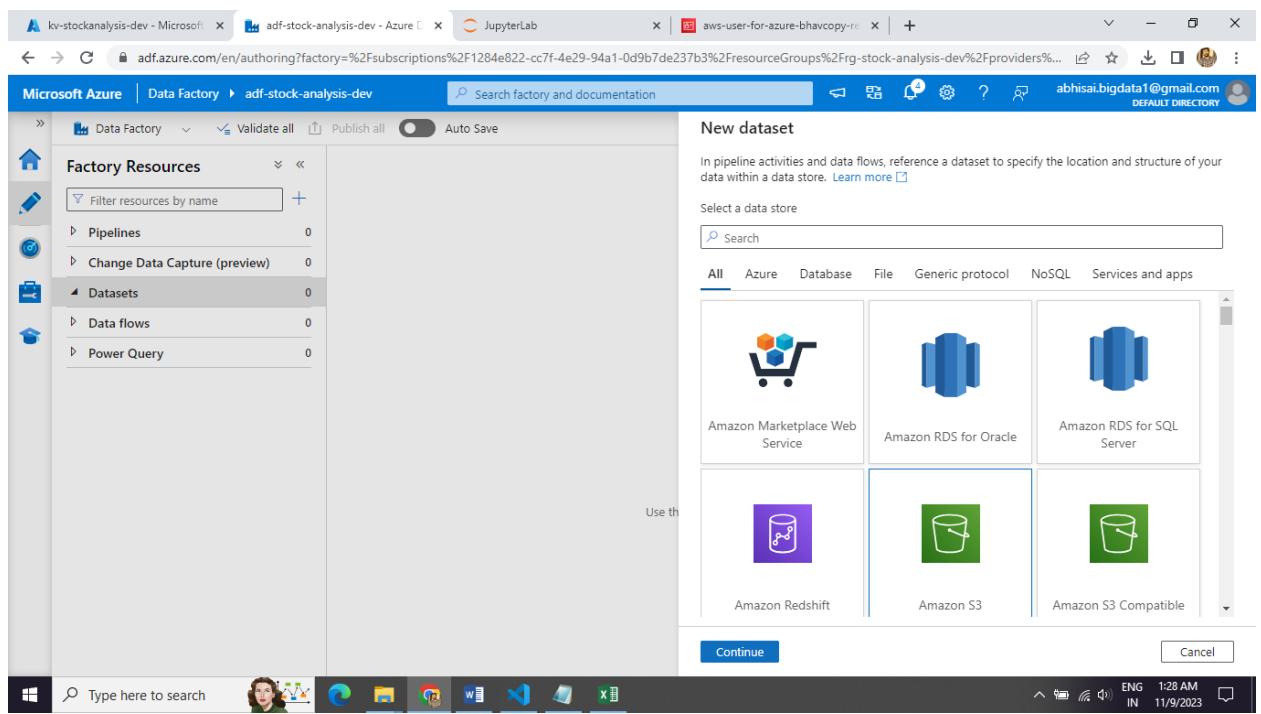
Note: Remember in Linked Services we don't talk about which file we need to copy and what is the schema etc. We just kind of establish the connectivity from ADF to source and sinks.

The screenshot shows the same Microsoft Azure Data Factory interface as before, but now with three linked services listed: 'Is\_AWS\_S3\_bhavcopy', 'Is\_Azure\_dl\_bhavcopy', and 'Is\_Azure\_Key\_Vault'. A success message 'Successfully created Is\_Azure\_dl\_bhavcopy (Linked service)' is displayed in a toast notification. The table columns are 'Name', 'Type', 'Related', and 'Annotations'.

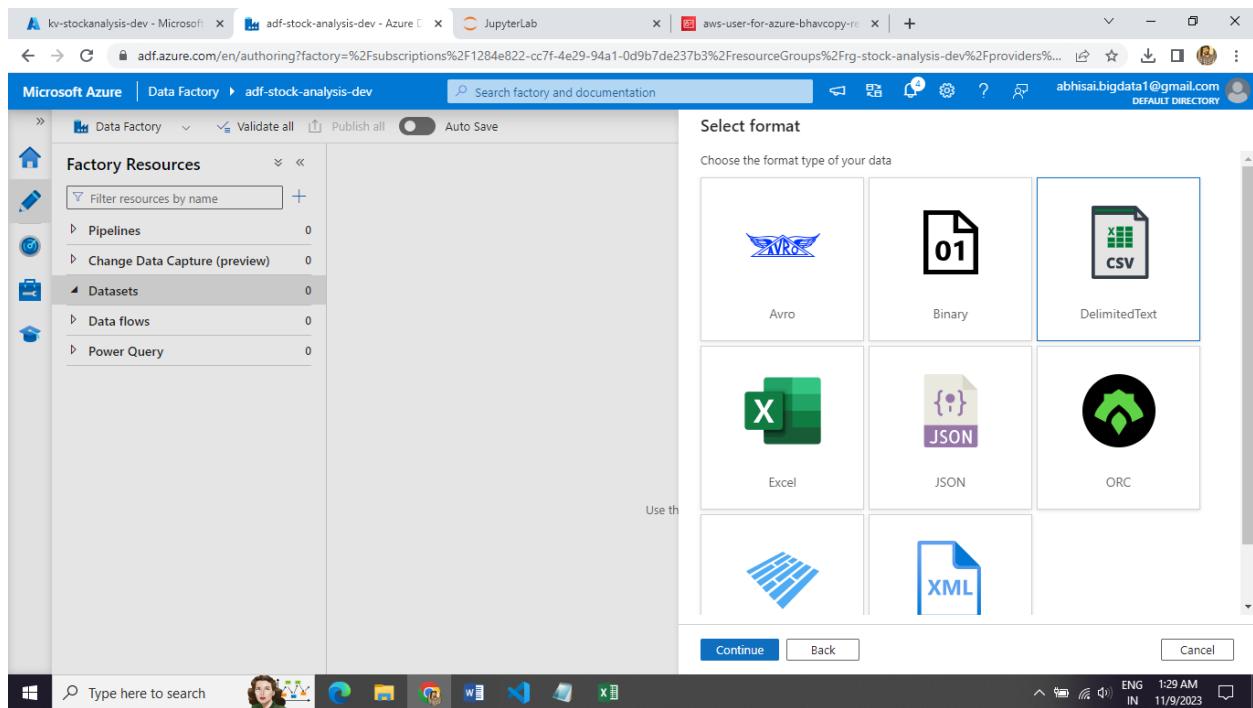
## Creation of Datasets – Source and Sink:



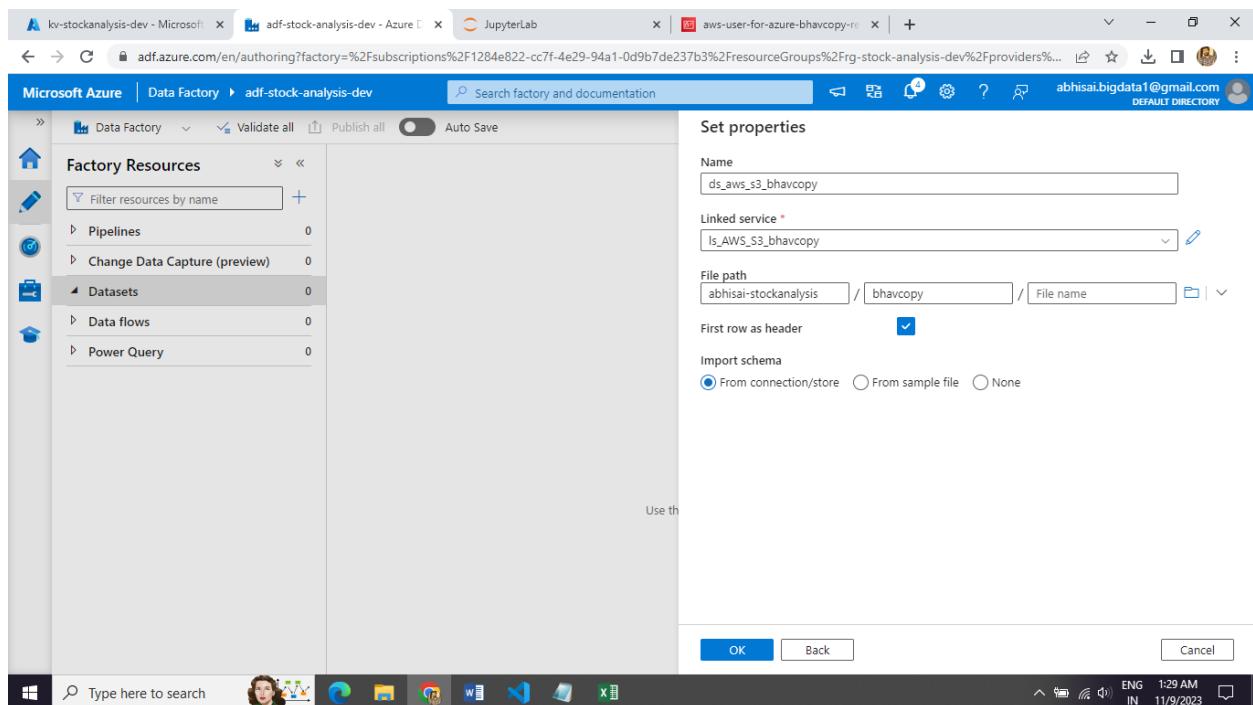
## Source:



Select format of the file:



Then provide the file location details. As I don't want to hardcode the file name I didn't kept the file name. Link the AWS S3 linked service to this dataset.



Sink:

Microsoft Azure | Data Factory | adf-stock-analysis-dev

Search factory and documentation

New dataset

In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. [Learn more](#)

Select a data store

Search

All Azure Database File Generic protocol NoSQL Services and apps

ds\_aws\_s3\_bhavcopy

DelimitedText ds\_aws\_s3\_bhavcopy

Connection Schema Parameters

Linked service \* ls\_AWS\_S3\_bhavcopy

File path \* abhisai-stockanalysis

Compression type Select...

Column delimiter Comma (,)

Row delimiter Default (\r\n, or \n)

Encoding Default(UTF-8)

Quote character Double quote ("")

Azure Cosmos DB for NoSQL

Azure Data Explorer (Kusto)

Azure Data Lake Storage Gen1

Azure Data Lake Storage Gen2

Azure Database for MariaDB

Azure Database for MySQL

Continue Cancel

Type here to search

11:31 AM 11/9/2023

Microsoft Azure | Data Factory | adf-stock-analysis-dev

Search factory and documentation

Select format

Choose the format type of your data

DelimitedText ds\_aws\_s3\_bhavcopy

Connection Schema Parameters

Linked service \* ls\_AWS\_S3\_bhavcopy

File path \* abhisai-stockanalysis

Compression type Select...

Column delimiter Comma (,)

Row delimiter Default (\r\n, or \n)

Encoding Default(UTF-8)

Quote character Double quote ("")

Avro

Binary

CSV

Excel

JSON

ORC

XML

Continue Back Cancel

Type here to search

11:31 AM 11/9/2023

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists Pipelines, Change Data Capture (preview), Datasets, Data flows, and Power Query. Under 'Datasets', 'ds\_aw\_s3\_bhavcopy' is selected. The main pane displays the 'Set properties' dialog for this dataset. The dataset name is 'ds\_aw\_s3\_bhavcopy'. The 'Linked service' is set to 'ls\_Azure\_dl\_bhavcopy'. The 'File path' is 'abhisai-stockanalysis/nse\_daily\_bhavcopy'. The 'First row as header' option is checked. The 'Import schema' section shows 'From connection/store' selected. Below these settings are fields for 'Compression type', 'Column delimiter', 'Row delimiter', 'Encoding', and 'Quote character'. At the bottom right of the dialog are 'OK', 'Back', and 'Cancel' buttons.

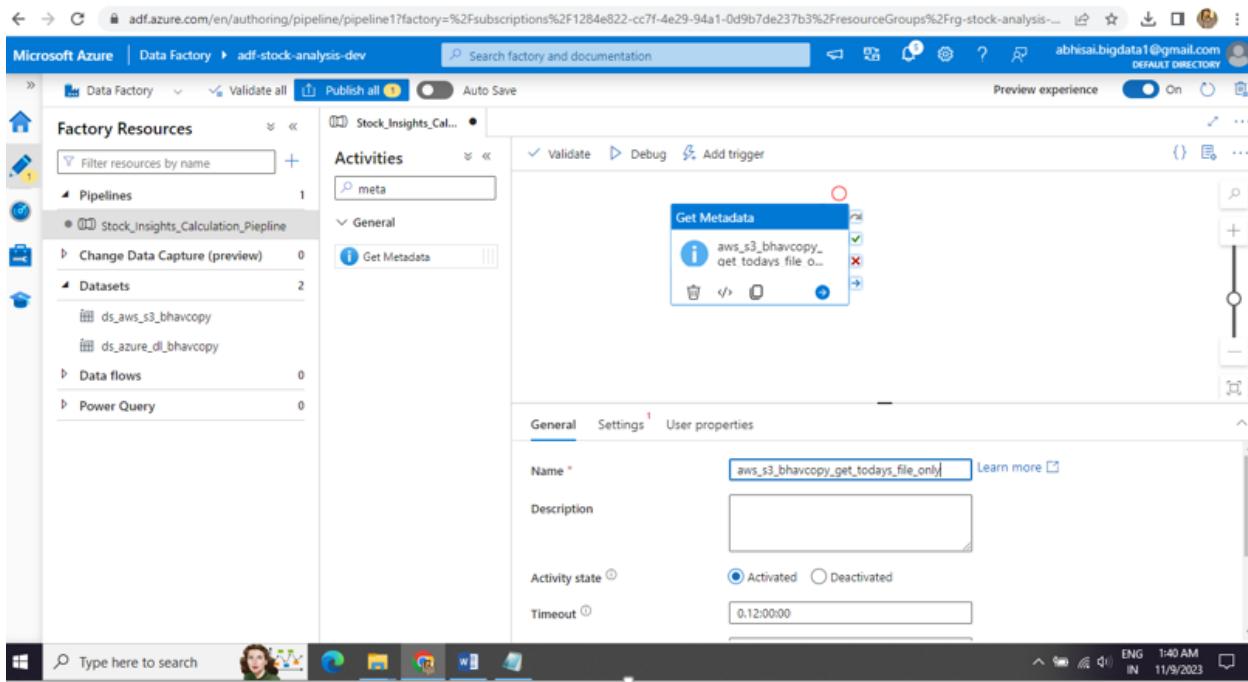
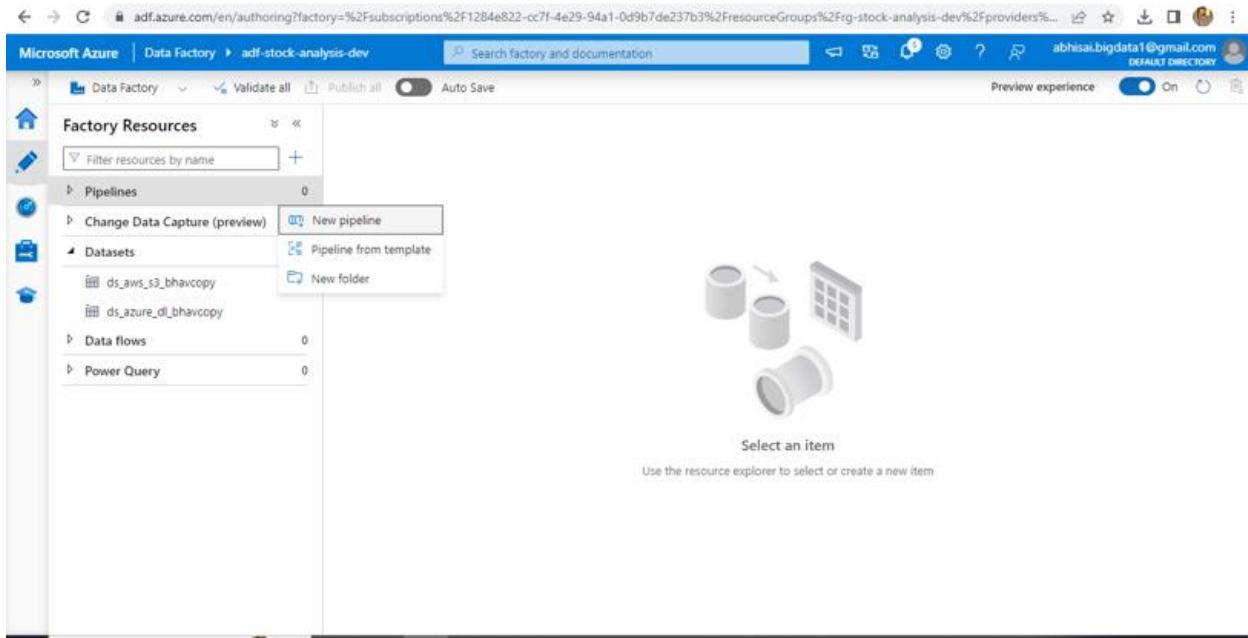
Publish all the changes to save all details in ADF.

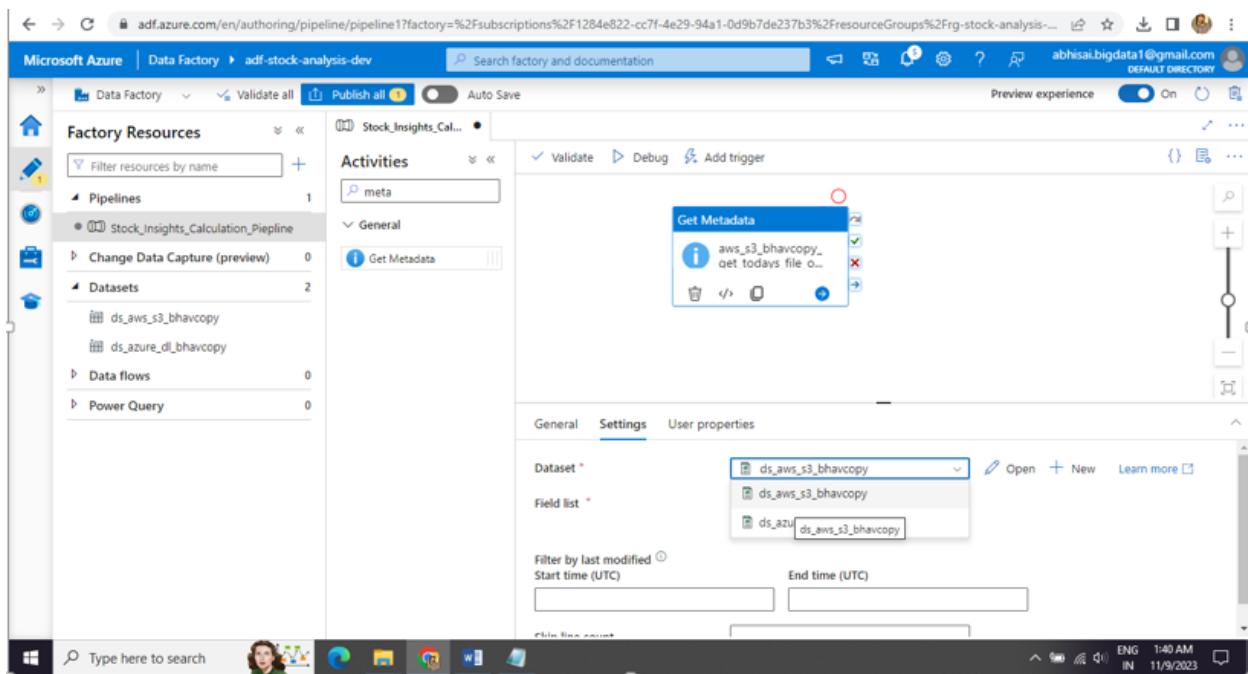
The screenshot shows the Microsoft Azure Data Factory interface with the 'Publishing' tab selected. The 'Pending changes' section indicates there are 2 changes to be published. The changes listed are:

NAME	CHANGE	EXISTING
ds_aw_s3_bhavcopy	(New)	-
ds_azure_dl_bhavcopy	(New)	-

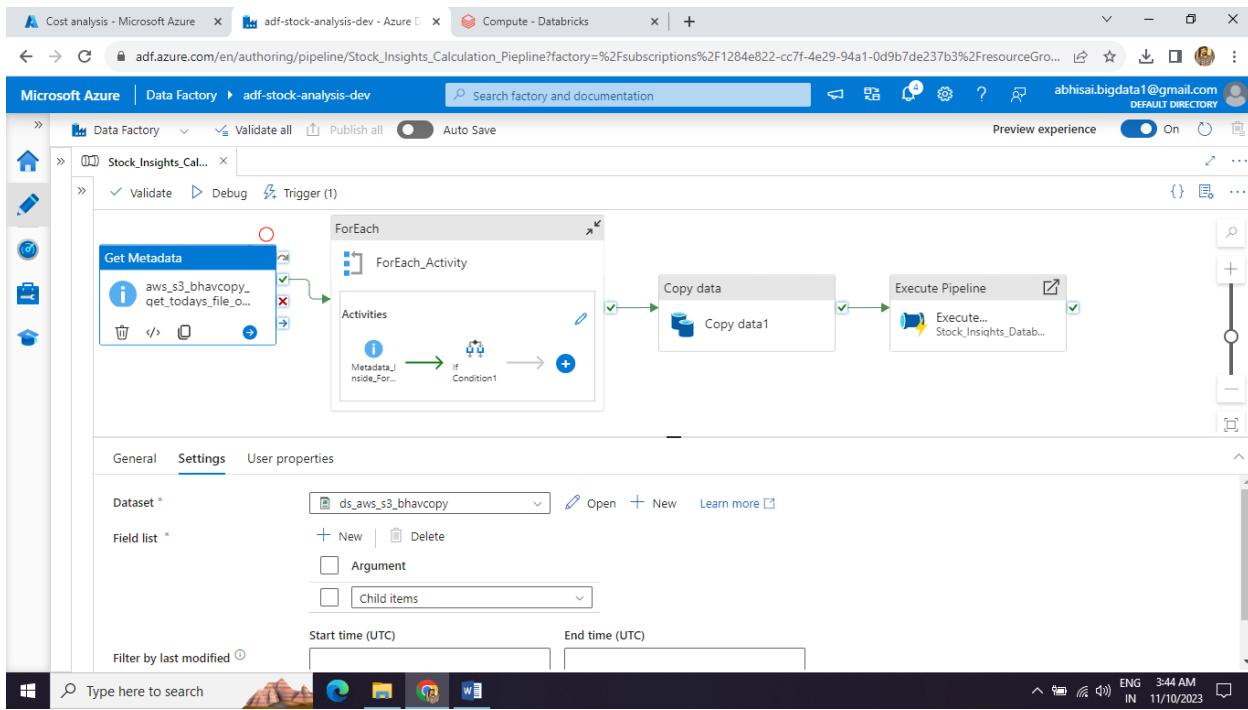
The 'Publish' button is highlighted at the bottom right of the dialog.

Pipeline building: As we need to ingest the latest file from AWS S3 we need to kind of read each file and verify the last\_modified time to see if its latest or not.

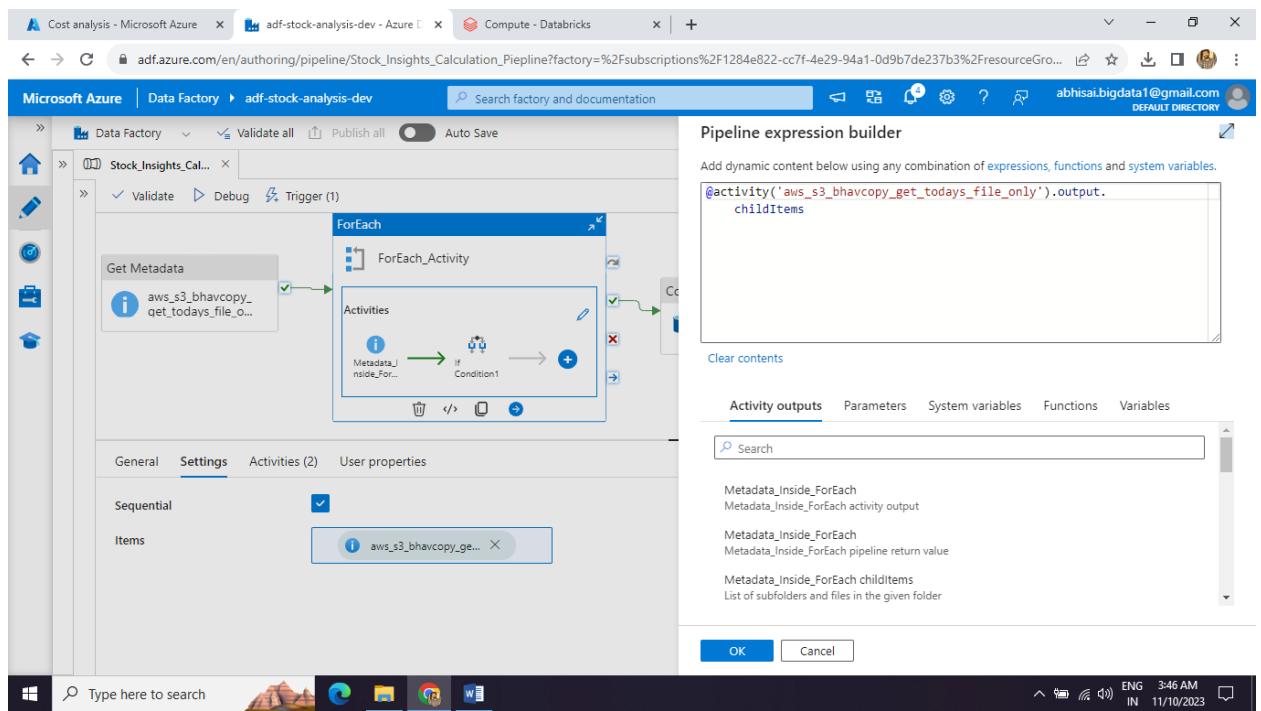
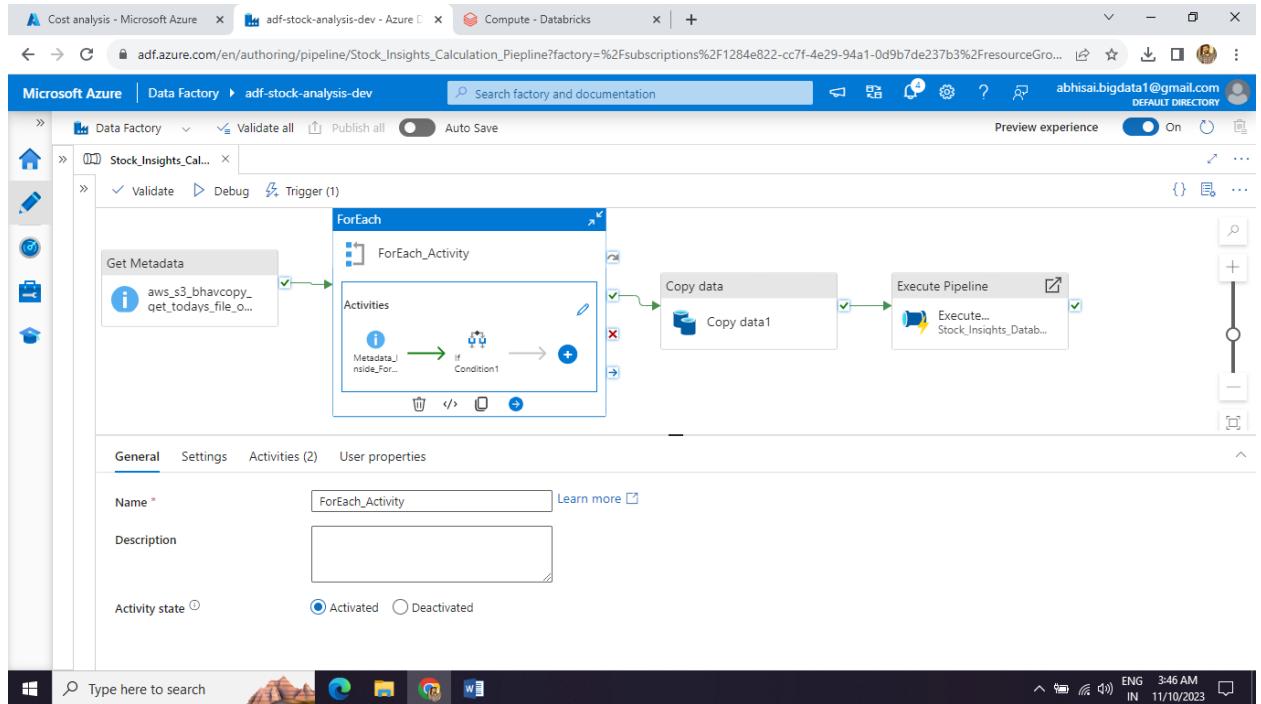




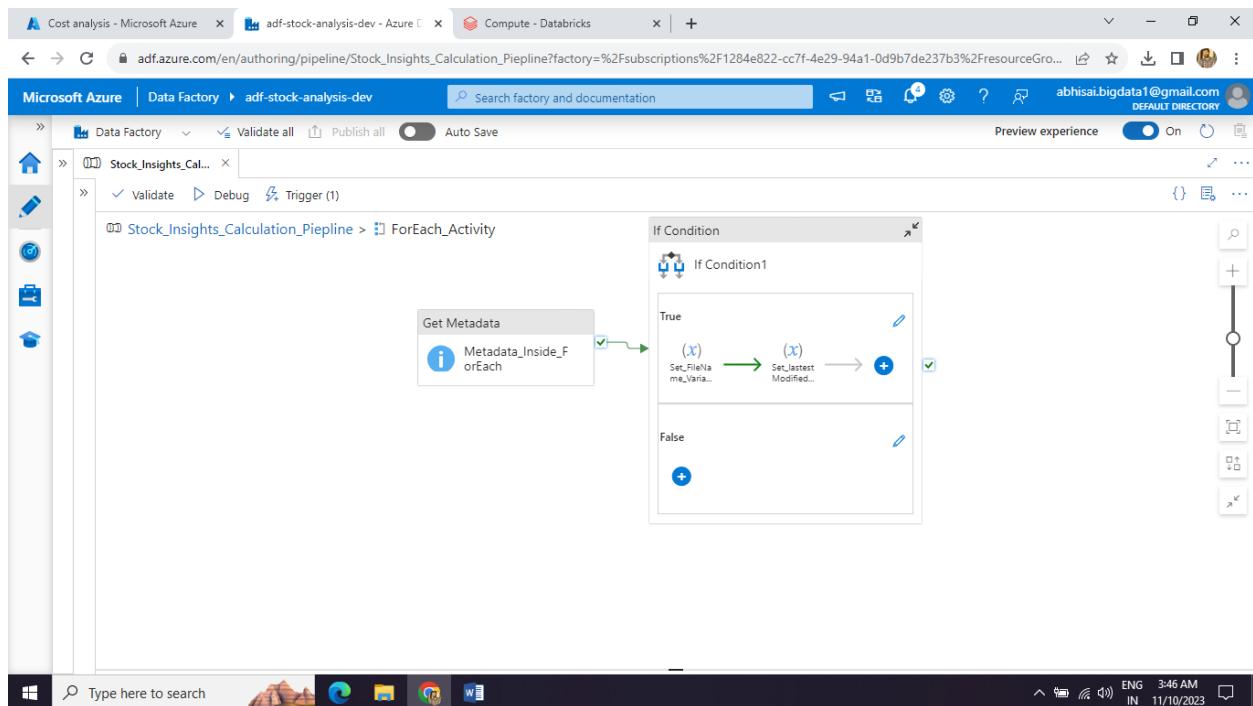
As we need child items from the S3 , we selected file as child so that output of this metadata we can see that child files in that s3 bucket.



Next For each activity for every file in that folder to know which is latest file



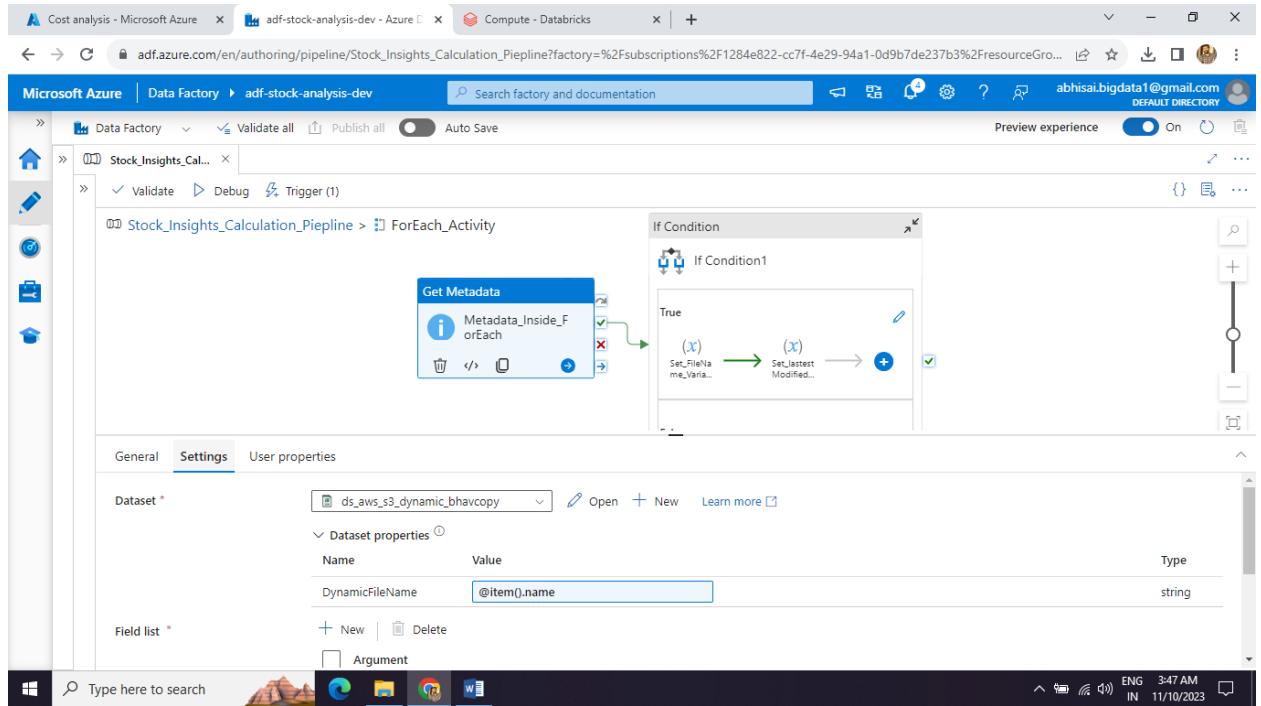
Inside For-each



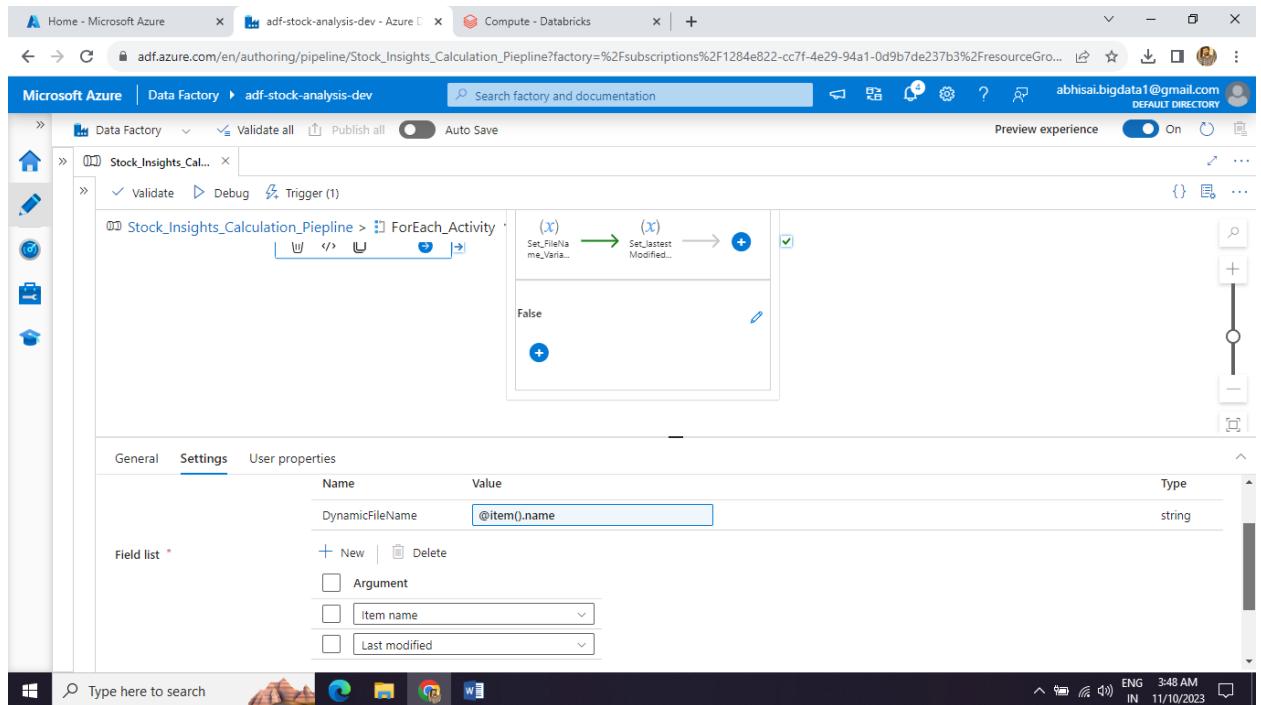
Again Metadata to know each file and its modified time

The screenshot shows the Azure Data Factory pipeline editor with the "General" tab selected for the "Metadata\_Inside\_Forceach" activity. The activity's name is "Metadata\_Inside\_Forceach". It is described as "Get Metadata activity to get file metadata inside a folder." The activity state is set to "Activated" and the timeout is 0.12:00:00. The pipeline editor interface includes a toolbar, a preview experience toggle, and a status bar at the bottom.

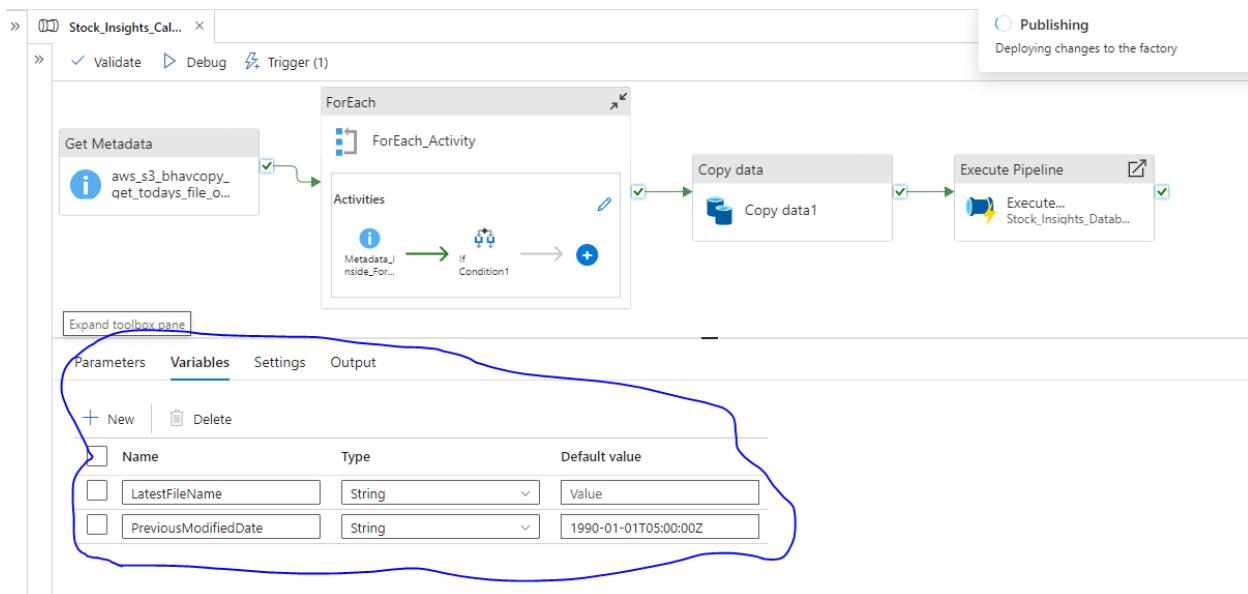
General	Settings	User properties
Name *	Metadata_Inside_Forceach <a href="#">Learn more</a>	
Description	<input type="text"/>	
Activity state	<input checked="" type="radio"/> Activated	<input type="radio"/> Deactivated
Timeout	0.12:00:00	



We need file name and time



Create pipeline level Variables:



We created one more dataset – with “DynamicFileName” as parameter

Cost analysis - Microsoft Azure | adf-stock-analysis-dev - Azure | Compute - Databricks

adf.azure.com/en/authoring/dataset/ds\_aws\_s3\_dynamic\_bhavcopy?factory=%2Fsubscriptions%2F1284e822-cc7f-4e29-94a1-0d9b7de237b3%2FresourceGroups%...

Microsoft Azure | Data Factory > adf-stock-analysis-dev

Search factory and documentation

Preview experience On

Data Factory Validate all Publish all Auto Save

Factory Resources Filter resources by name

Pipelines Stock\_Insights\_Databricks Stock\_Insights\_Calculation\_Pipeline Change Data Capture (preview) 0

Datasets ds\_aws\_s3\_bhavcopy ds\_aws\_s3\_dynamic\_bhavcopy ds\_azure\_dl\_bhavcopy 3

Data flows 0

Power Query 0

DelimitedText ds\_aws\_s3\_dynamic\_bhavcopy

CSV

Connection Schema Parameters

+ New Delete

Name Type Default value

DynamicFileName String Value

Type here to search ENG 3:47 AM IN 11/10/2023

Home - Microsoft Azure | adf-stock-analysis-dev - Azure | Compute - Databricks

adf.azure.com/en/authoring/pipeline/Stock\_Insights\_Calculation\_Pipeline?factory=%2Fsubscriptions%2F1284e822-cc7f-4e29-94a1-0d9b7de237b3%2FresourceGro...

Microsoft Azure | Data Factory > adf-stock-analysis-dev

Search factory and documentation

Preview experience On

Data Factory Validate all Publish all Auto Save

Stock\_Insights\_Cal...

Validate Debug Trigger (1)

Stock\_Insights\_Calculation\_Pipeline > ForEach\_Activity

Get Metadata Metadata\_Inside\_ForEach

If Condition1

True

(x) Set\_FileName... → (x) Set\_Lastest... → +

False

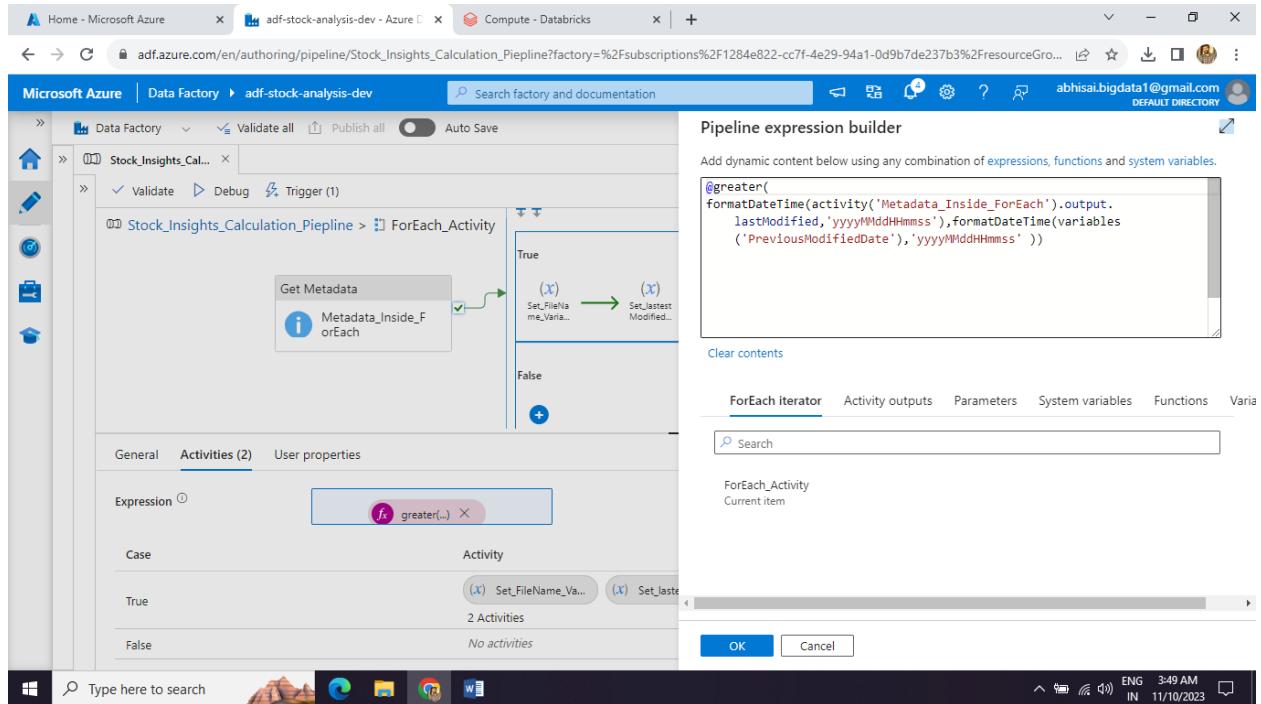
General Activities (2) User properties

Name: If Condition1

Description:

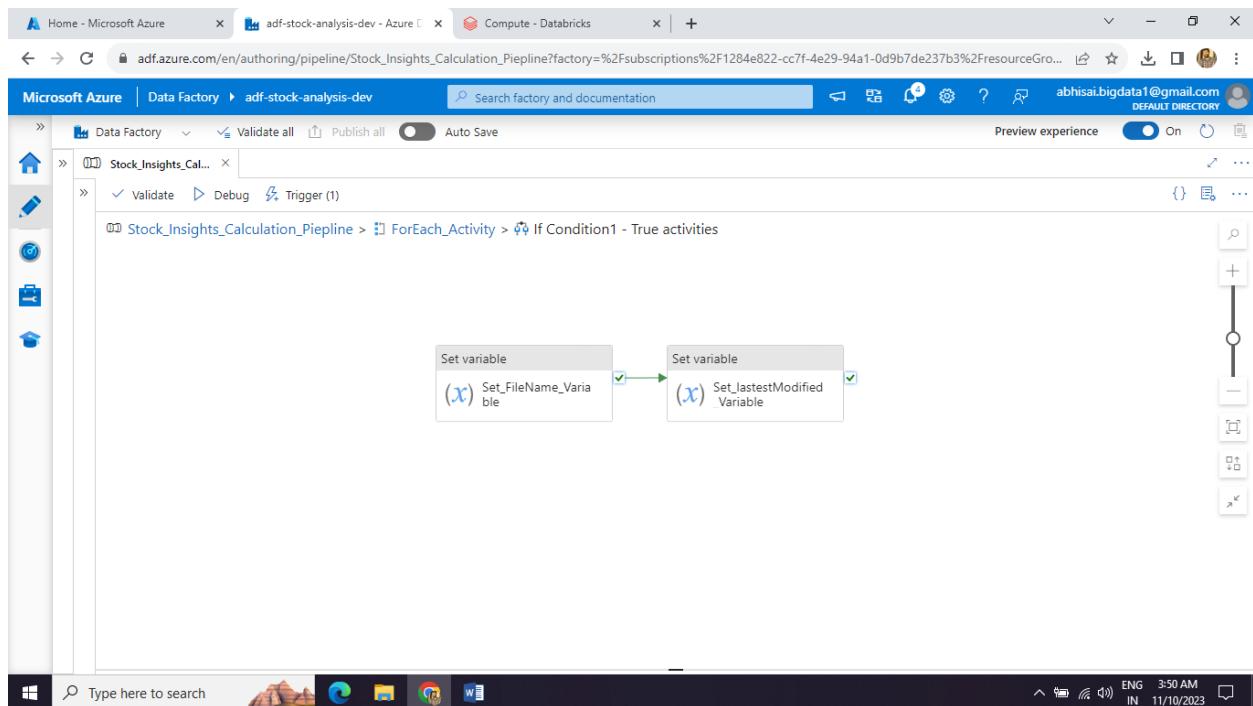
Activity state: Activated

Type here to search ENG 3:49 AM IN 11/10/2023



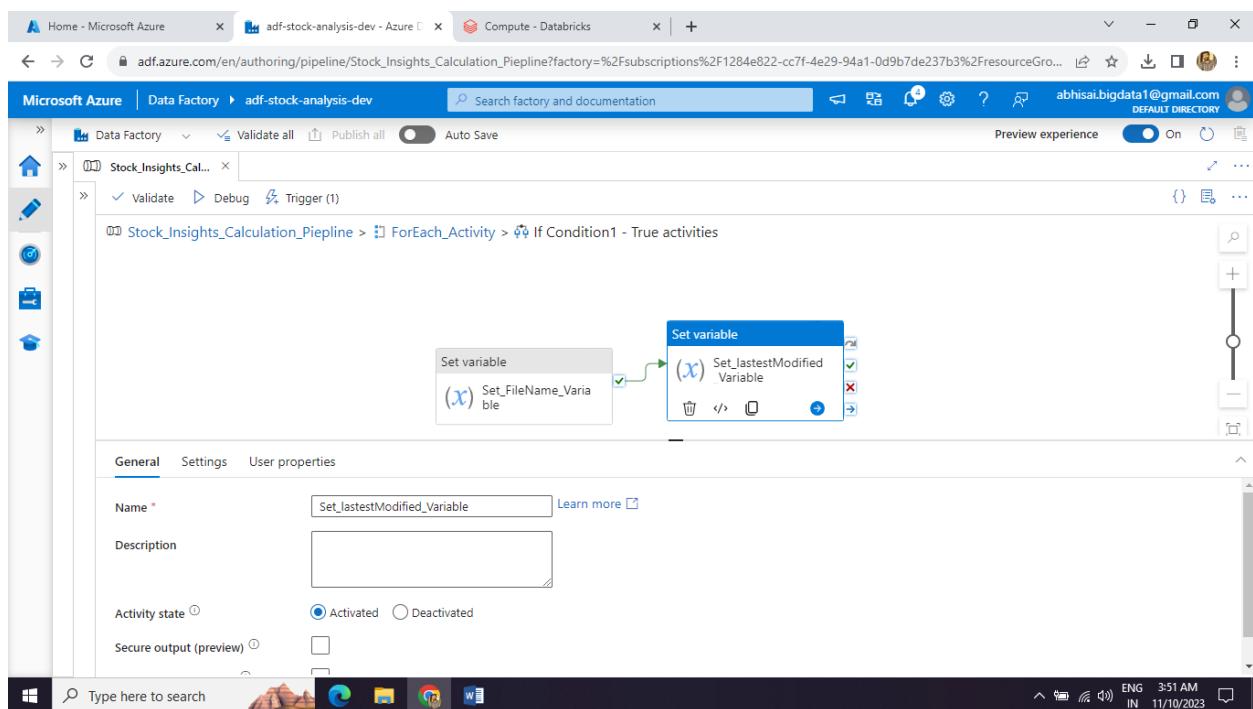
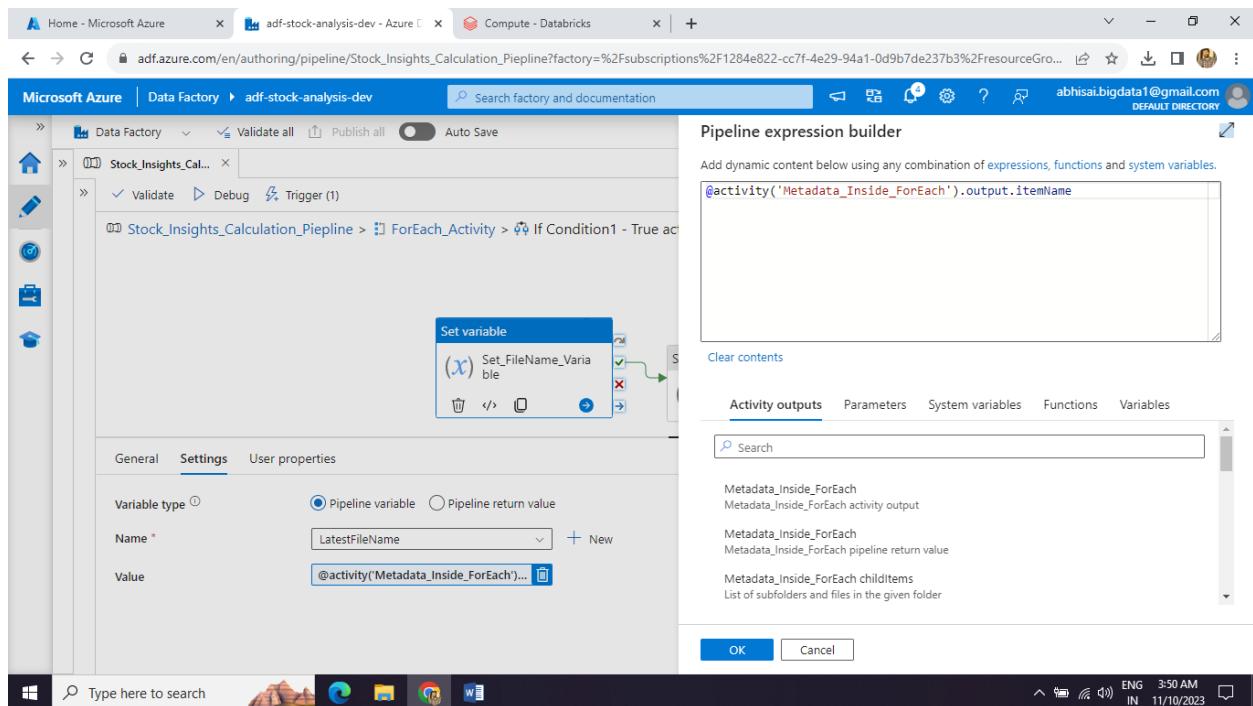
```
@greater(  
formatDateTime(activity('Metadata_Inside_ForEach').output.lastModified, 'yyyyMMddHHmmss'),formatDateTime(variables('PreviousModifiedDate'), 'yyyyMMddHHmmss'))
```

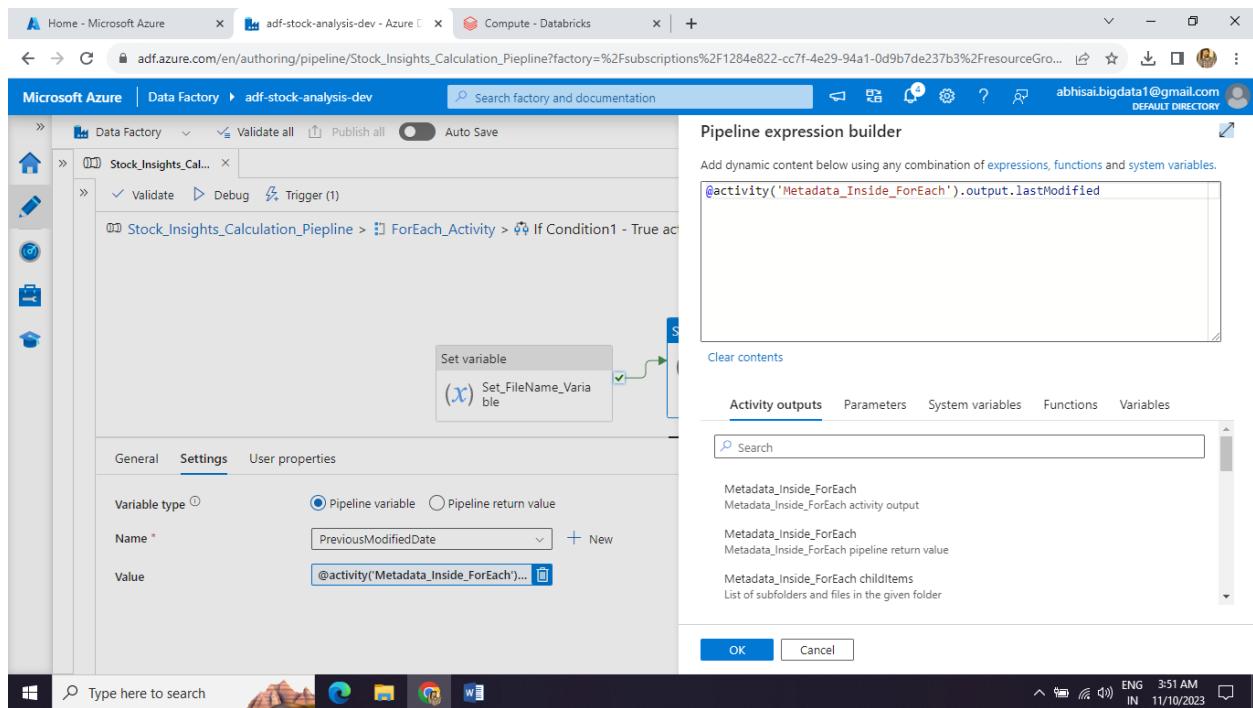
If TRUE



I am setting pipeline variables for file name

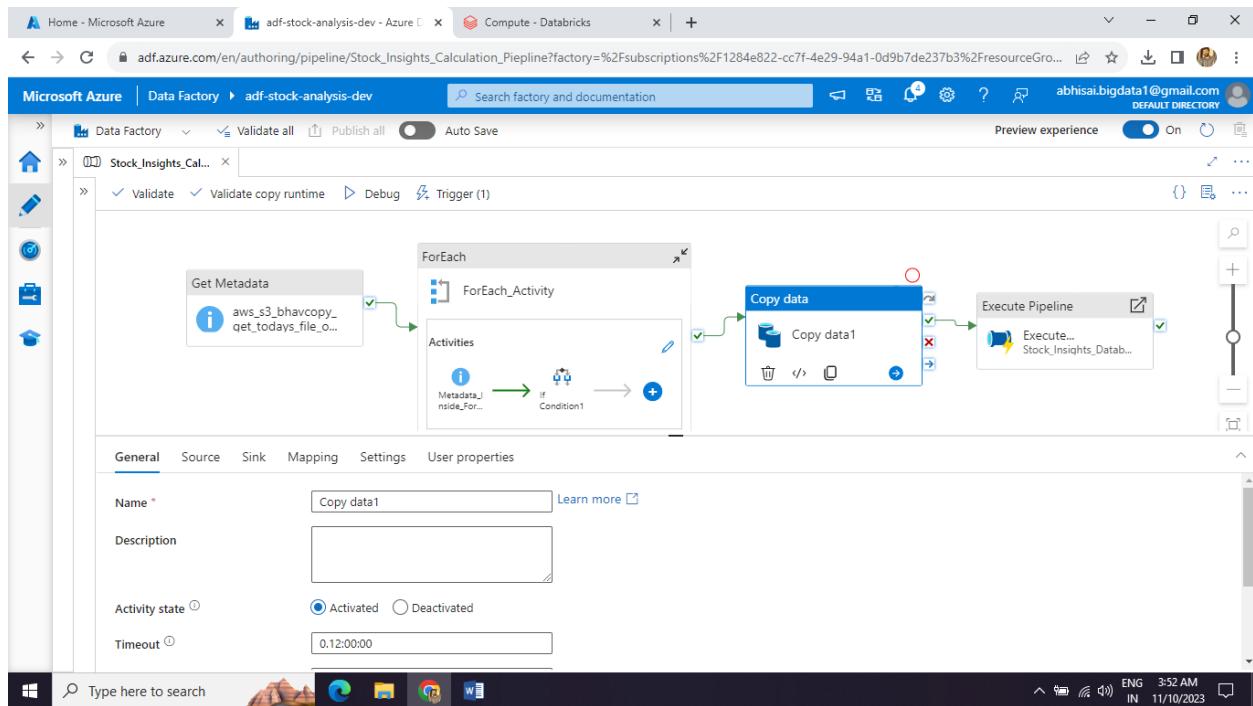
This screenshot shows the same pipeline setup as the previous one, but with the "General" tab of the "Set variable" activity properties panel visible. The activity is named "Set\_FileName\_Variable". It is currently activated. The "Description" field is empty. The "Secure output (preview)" checkbox is unchecked. The pipeline editor interface is identical to the first screenshot, showing the overall pipeline structure and the sequence of activities.



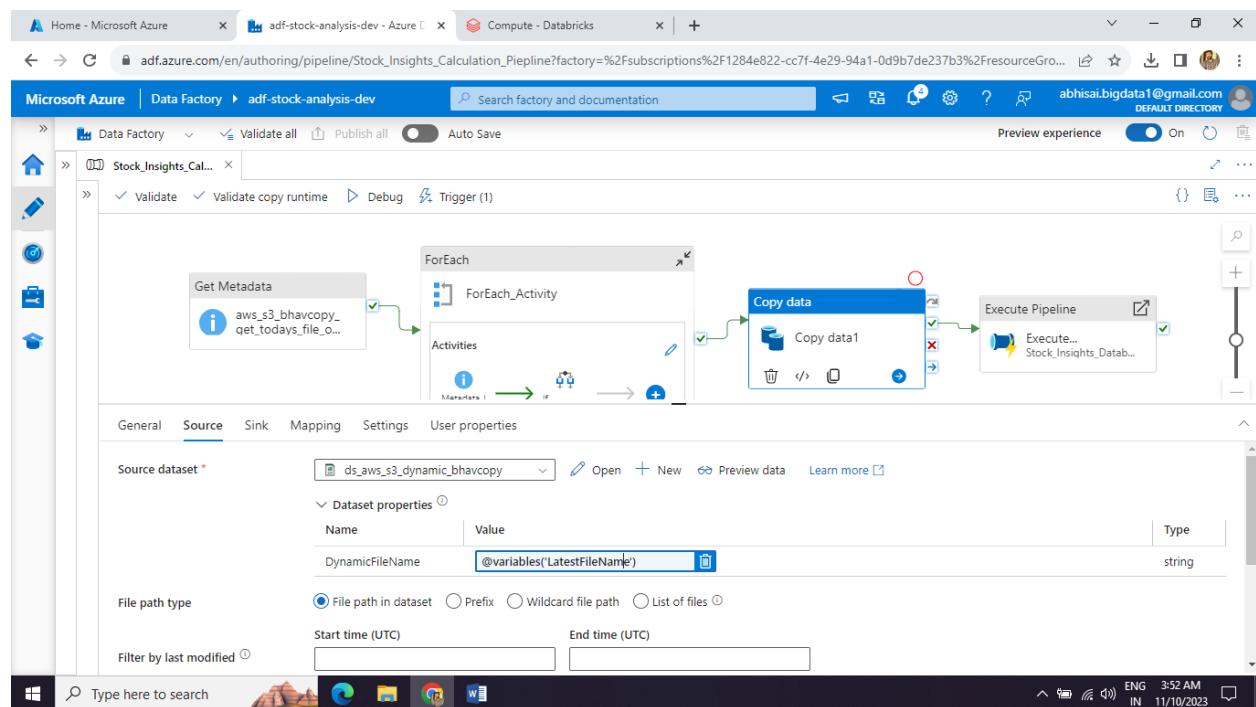
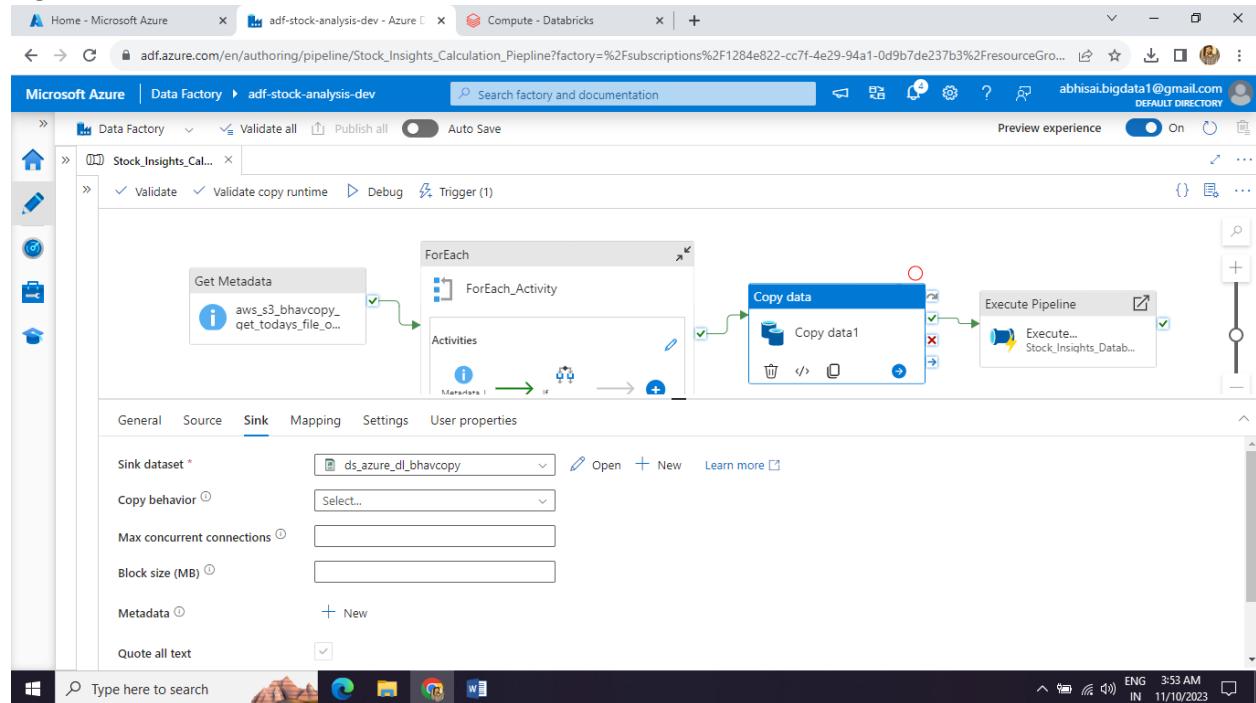


Reason for saving “LastModified” is we need to compare this last modified everytime with previous file to know if this is latest or not.

Next : Copy activity for that Dynamic file



As we need to ingest only that latest file , we are using pipeline variable to know which file we need to ingest.



Till now, it's all about Ingesting the data. Now we need to process the data using Databricks.

Now Create Databricks:

databricks-stock-analysis-dev

The screenshot shows the Azure portal interface for creating a new Databricks workspace. The current step is 'Project Details'. The workspace name is set to 'databricks-stock-analysis-dev'. The region is 'Canada Central' and the pricing tier is 'Standard (Apache Spark, Secure with Microsoft Entra ID)'. A managed resource group named 'databricks-stock-analysis-managed-dev' is selected. At the bottom, the 'Review + create' button is highlighted, indicating the next step in the process.

**Create an Azure Databricks workspace**

**Validation Succeeded**

**Basics**   **Networking**   **Encryption**   **Tags**   **Review + create**

**Summary**

**Basics**

Workspace name	databricks-stock-analysis-dev
Subscription	Free Trial
Resource group	rg-stock-analysis-dev
Region	Canada Central
Pricing Tier	standard
Managed Resource Group name	databricks-stock-analysis-managed-dev

**Networking**

Deploy Azure Databricks workspace with	No
Secure Cluster Connectivity (No Public IP)	
Deploy Azure Databricks workspace in	No

**Create**   < Previous   Download a template for automation

In order to trigger Databricks notebook from ADF we need to provide access key associated with Azure Databricks in ADF (using Azure KeyVault)

**databricks-stock-analysis-dev**   Azure Databricks Service

**Overview**   **Activity log**   **Access control (IAM)**   **Tags**   **Diagnose and solve problems**

**Settings**

- Virtual Network Peering
- Encryption
- Networking
- Properties
- Locks

**Automation**

- CLI / PS
- Tasks (preview)

**Launch Workspace**   **Upgrade to Premium**

To generate access key in Databricks

Go to user settings

The screenshot shows the Microsoft Azure Databricks interface. The left sidebar has sections like Workspace, Recents, Catalog, Workflows, Compute, Data Engineering, Job Runs, Delta Live Tables, Machine Learning, Experiments, Features, Models, Serving, and Partner Connect. The main area is titled 'Profile' under 'Settings'. It shows a 'Display name' field with 'Abhi Sai (abhisai.bigdata1@gmail.com)' and a 'Groups' section. A context menu is open on the right, showing options like User Settings, Admin Settings, Azure Portal, Manage Account, Privacy Policy, and Log out. The URL in the address bar is <https://adb-320085348881679.19.azuredatabricks.net/settings/user/profile/?o=320085348881679>.

Go to developer

The screenshot shows the Microsoft Azure Databricks interface. The left sidebar is identical to the previous screenshot. The main area is titled 'Developer' under 'Settings'. It shows an 'Access tokens' section with a 'Manage' button. Below it is an 'Editor settings' section with 'General' and 'Notebook Notifications' (which is turned On). Further down are 'Spark tips' (On) and 'Databricks Advisor' (On). The URL in the address bar is <https://adb-320085348881679.19.azuredatabricks.net/settings/user/.../access-tokens/?o=...>.

Generate Access keys:

The screenshot shows the Databricks Settings - Access tokens page. The left sidebar includes sections like Workspace, Recents, Catalog, Workflows, Compute, Data Engineering, Job Runs, Delta Live Tables, Machine Learning, Experiments, Features, Models, Serving, and Partner Connect. The main area is titled "Access tokens" under "User settings > Developer". It displays a message about using personal access tokens for secure authentication to the Databricks API instead of passwords. A "Generate new token" button is present. Below it, a table shows columns for Comment, Creation, and Expiration, with a note stating "No tokens exist". The status bar at the bottom shows "ENG IN 11:38 AM 11/9/2023".

The screenshot shows the "Generate new token" dialog box overlaid on the previous screen. The dialog has fields for "Comment" (containing "databricks-key-for-adf-stock-analysis") and "Lifetime (days)" (set to 90). There are "Cancel" and "Generate" buttons at the bottom. The background shows the same Databricks Settings - Access tokens page as the first screenshot.

The screenshot shows the Databricks Settings - Access tokens page. The left sidebar has a 'New' button and sections for Workspace, Recents, Catalog, Workflows, Compute, Data Engineering, Job Runs, Delta Live Tables, Machine Learning, Experiments, Features, Models, Serving, and Partner Connect. A 'Collapse menu' button is at the bottom. The main area shows 'User settings > Developer > Access tokens'. It says 'Personal access tokens can be used for secure authentication to the [Databricks API](#) instead of passwords.' A 'Generate new token' button is present. A table lists one token:

Comment	Creation	Expiration
databricks-key-for-adf-stock-analysis	2023-11-09 11:38:47 IST	2024-02-07 11:38:47 IST

The status bar at the bottom shows 'Type here to search' and system icons.

Now add this access key to Azure key vault

kv-stockanalysis-dev - Microsoft | Access tokens - Settings - Data | portal.azure.com/#@abhisaibigdata1gmail.onmicrosoft.com/resource/subscriptions/1284e822-cc7f-4e29-94a1-0d9b7de237b3/resourceGroups/r...

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

Home > Key vaults > kv-stockanalysis-dev

## Key vaults

Default Directory (abhisaibigdata1@gmail.onmicrosoft.com)

+ Create ...

Filter for any field...

Name ↑

kv-stockanalysis-dev ...

Search

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Access policies

Events

Objects

Keys

Secrets

Certificates

Settings

Access configuration

Networking

Generate/Import Refresh Restore Backup View sample code Manage deleted secrets

Name	Type	Status	Expiration date
aws-s3-bhavcopy-id		✓ Enabled	
aws-s3-bhavcopy-value		✓ Enabled	

Page 1 of 1

Give feedback

Type here to search

Windows Start button

11:39 AM ENG IN 11/9/2023

Create a secret - Microsoft Azure | Access tokens - Settings - Data | portal.azure.com/#view/Microsoft\_Azure\_KeyVault/CreateSecretBlade/secret~/null/vaultId/%2Fsubscriptions%2F1284e822-cc7f-4e29-94a1-0d9b7de237b3%2...

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

Home > Key vaults > kv-stockanalysis-dev | Secrets >

## Create a secret

Upload options

Name \* ⓘ databricks-access-key

Secret value \* ⓘ  ⚡

Content type (optional)

Set activation date ⓘ

Set expiration date ⓘ

Enabled Yes

Tags 0 tags

Create Cancel

Type here to search

Windows Start button

11:40 AM ENG IN 11/9/2023

The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes 'kv-stockanalysis-dev - Microsoft' and 'Access tokens - Settings - Data...'. The main search bar says 'Search resources, services, and docs (G+)'. The user is signed in as 'abhisai.bigdata1@gmail...'.

The left sidebar shows 'Key vaults' under 'kv-stockanalysis-dev'. A message at the top right says 'The secret 'databricks-access-key' has been successfully created.' Below is a table of secrets:

Name	Type	Status	Expiration date
databricks-access-key		✓ Enabled	
aws-s3-bhavcopy-id		✓ Enabled	
aws-s3-bhavcopy-value		✓ Enabled	

The status bar at the bottom shows 'Page 1 of 1' and a 'Give feedback' link. The taskbar at the bottom of the screen includes icons for File Explorer, Task View, Edge, Google Chrome, File, Word, and Power BI.

How to access Azure Storage accounts from Databricks:

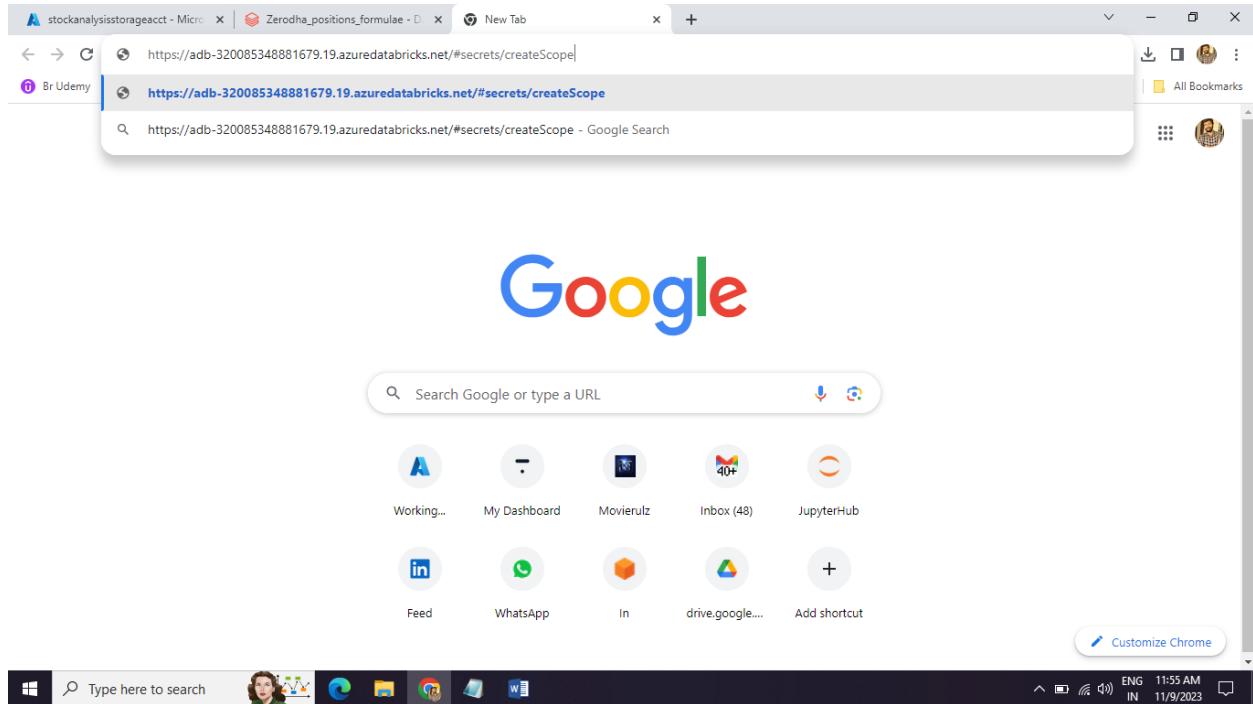
```
spark.conf.set("fs.azure.account.auth.type.<storage-account>.dfs.core.windows.net", "SAS")
```

```
spark.conf.set("fs.azure.sas.token.provider.type.<storage-account>.dfs.core.windows.net",
"org.apache.hadoop.fs.azurebfs.sas.FixedSASTokenProvider")
```

```
spark.conf.set("fs.azure.sas.fixed.token.<storage-account>.dfs.core.windows.net",
dbutils.secrets.get(scope="<scope>", key="<sas-token-key>"))
```

Instead of harding the SAS key in the above we need to basically create a scope and use the below method to get the key using scope - dbutils.secrets.get(scope="<scope>", key="<sas-token-key>")

**Databricks.com/#secrets/createScope**



HomePage / Create Secret Scope

Create Secret Scope | Cancel Create

A store for secrets that is identified by a name and backed by a specific store type. [Learn more](#)

Scope Name

Manage Principal

Azure Key Vault

DNS Name

Resource ID

New

Workspace

Recents

Catalog

Workflows

Compute

Data Engineering

Job Runs

Delta Live Tables

Machine Learning

Experiments

Features

Models

Serving

Partner Connect

Collaborate menu

Waiting for adb-320085348881679.19.azuredatabricks.net...

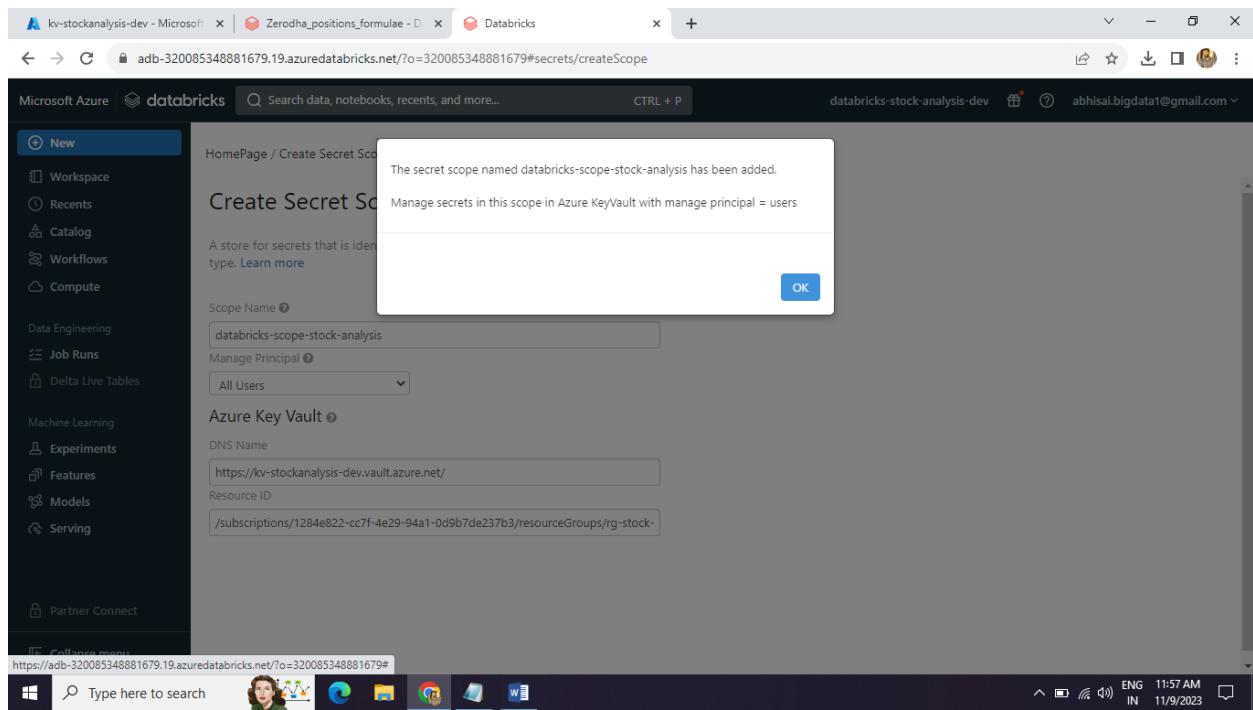
Type here to search

ENG IN 11:56 AM 11/9/2023

Three screenshots of a Microsoft Azure Databricks interface showing the creation of a Secret Scope.

The screenshots show the following steps:

- Screenshot 1:** The "Create Secret Scope" page. The "Scope Name" field contains "databricks-scope-stock-analysis". The "Manage Principal" dropdown is set to "All Users". Under "Azure Key Vault", the "DNS Name" is "https://kv-stockanalysis-dev.vault.azure.net/" and the "Resource ID" is "https://rg-stock-analysis-dev/providers/Microsoft.KeyVault/vaults/kv-stockanalysis-dev".
- Screenshot 2:** The same page, but the "Resource ID" field now shows the full resource ID: "/subscriptions/1284e822-cc7f-4e29-94a1-0d9b7de237b3/resourceGroups/rg-stock-".
- Screenshot 3:** The page is in the process of verification, indicated by the "Verifying..." status next to the "Create" button.



Create SAS key for storage account so that we can add that SAS key in Azure Keyvault and Databricks can access that SAS key

**Storage accounts** | stockanalysissstorageacct

Containers

Name	Last modified	Anonymous access l...	Lease state
Slogs	11/8/2023, 1:44:06 AM	Private	Available
abhisai-stockanalysis	11/9/2023, 12:46:07 ...	Private	Available

Shared access signature

A shared access signature (SAS) is a URI that grants restricted access rights to Azure Storage resources. You can provide a shared access signature to clients who should not be trusted with your storage account key but whom you wish to delegate access to certain storage account resources. By distributing a shared access signature URI to these clients, you grant them access to a resource for a specified period of time.

An account-level SAS can delegate access to multiple storage services (i.e. blob, file, queue, table). Note that stored access policies are currently not supported for an account-level SAS.

Learn more about creating an account SAS

Allowed services: Blob, File, Queue, Table

Allowed resource types: Service, Container, Object

Allowed permissions: Read, Write, Delete, List, Add, Create, Update, Process, Immutable storage, Permanent delete

Blob versioning permissions: Enables deletion of versions

The screenshot shows the Microsoft Azure Storage accounts page for the account 'stockanalysissstorageacct'. The 'Shared access signature' section is selected. The configuration includes:

- Start and expiry date/time:
  - Start: 11/09/2023
  - End: 11/23/2023
- Allowed IP addresses: (UTC+05:30) Chennai, Kolkata, Mumbai, New Delhi
- Allowed protocols: HTTPS only
- Preferred routing tier: Basic (default)
- Signing key: key1

A 'Generate SAS and connection string' button is visible at the bottom.

The screenshot shows the Microsoft Azure Storage accounts page for the account 'stockanalysissstorageacct'. The 'Shared access signature' section is selected. The generated SAS token is displayed in the 'SAS token' field:

```
?sv=2022-11-02&ss=bfqt&srt=sco&sp=rw&lt;long_hex_string>&se=2023-11-23T14:29:36Z&st=2023-11-09T06:29:36Z&spr=https&sig=8HxKi16Edg4%2BAoCRrkUalzwSITTy3ztRK0rv1kVqtaM%3D
```

The 'Copy' button next to the SAS token has been clicked, as indicated by the 'Copied' message.

?sv=2022-11-02&ss=bfqt&srt=sco&sp=rw&lt;long\_hex\_string>&se=2023-11-23T14:29:36Z&st=2023-11-09T06:29:36Z&spr=https&sig=8HxKi16Edg4%2BAoCRrkUalzwSITTy3ztRK0rv1kVqtaM%3D

Add above SAS key in Azure KeyVault

Screenshot of Microsoft Azure portal showing the creation of a secret in a Key Vault.

**Create a secret**

Upload options: Manual

Name: storageaccount-sas-key

Secret value: (redacted)

Content type (optional):

Set activation date: (checkbox)

Set expiration date: (checkbox)

Enabled: Yes

Tags: 0 tags

**Create** **Cancel**

**kv-stockanalysis-dev | Secrets**

The secret 'storageaccount-sas-key' has been successfully created.

Name	Type	Status	Expiration date
storageaccount-sas-key		✓ Enabled	
aws-s3-bhavcopy-id		✓ Enabled	
aws-s3-bhavcopy-value		✓ Enabled	
databricks-access-key		✓ Enabled	

```
spark.conf.set("fs.azure.account.auth.type.stockanalysissecretstorageacctdfs.core.windows.net", "SAS")
spark.conf.set("fs.azure.sas.token.provider.type.stockanalysissecretstorageacctdfs.core.windows.net", "org.apache.hadoop.fs.azurebfs.sas.FixedSASTokenProvider")
```

```
spark.conf.set("fs.azure.sas.fixed.token.stockanalysissestorageacct.dfs.core.windows.net", dbutils.secrets.get(scope="databricks-scope-stock-analysis", key="storageaccount-sas-key"))
```

### Create Azure SQL:

Azure SQL - Microsoft Azure

portal.azure.com/#view/HubsExtension/BrowseResource/resourceType/Microsoft.Sql%2Fazuresql

Microsoft Azure

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

abhisai.bigdata1@gmail...  
DEFAULT DIRECTORY (ABHISAI...)

Home >

Azure SQL

Default Directory (abhisai.bigdata1@gmail.onmicrosoft.com)

+ Create Reservations Manage view Refresh Export to CSV Open query Assign tags Delete

Filter for any field... Subscription equals all Resource group equals all Location equals all Add filter

No grouping List view

Showing 0 to 0 of 0 records.

Name ↑↓ Resour... ↑↓ Service tier ↑↓ Resource group ↑↓ Location ↑↓ Subscription ↑↓

SQL

No Azure SQL resources to display

Try changing or clearing your filters.

Create Azure SQL resource

Learn more Create

Give feedback

Type here to search

Windows Start button

11:32 AM  
ENG IN 11/9/2023

Select “SQL Databases”

A Select SQL deployment option - +

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

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

Home > Azure SQL > Select SQL deployment option

How do you plan to use the service?

**SQL databases**  
Best for modern cloud applications. Hyperscale and serverless options are available.  
Resource type: Single database  
**Create** **Hide details** **Create**

**SQL managed instances**  
Best for most migrations to the cloud. Lift-and-shift ready.  
Resource type: Single instance  
**Create** **Show details**

**SQL virtual machines**  
Best for migrations and applications requiring OS-level access. Lift-and-shift ready.  
Image: **Create** **Show details**  High availability

**Single database**  
Single databases are a great fit for modern, cloud-born applications that need a fully managed database with predictable performance.  
**Featured capabilities:**

- Hyperscale storage (up to 100TB)
- Serverless compute

**Elastic pool**  
Elastic pools provide a cost-effective solution for managing the performance of multiple databases with variable usage patterns.  
**Featured capabilities:**

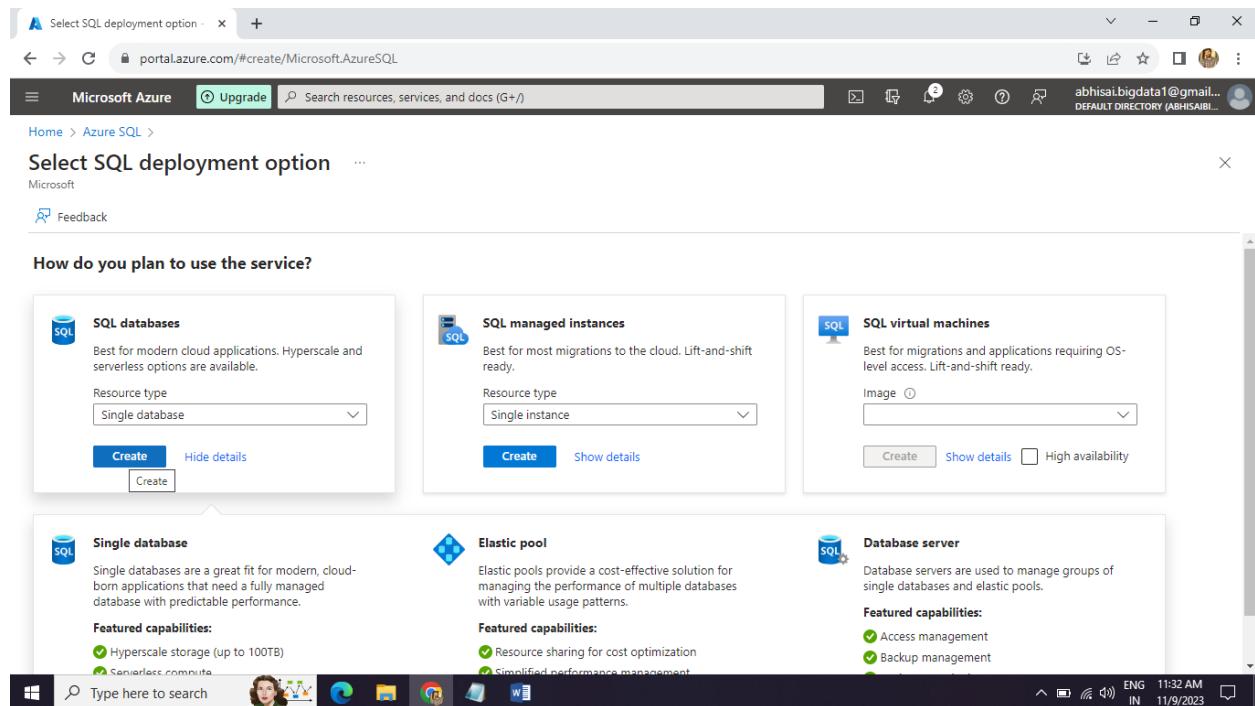
- Resource sharing for cost optimization
- Simplified performance management

**Database server**  
Database servers are used to manage groups of single databases and elastic pools.  
**Featured capabilities:**

- Access management
- Backup management

Type here to search

11:32 AM ENG IN 11/9/2023



A Create SQL Database Server - M +

portal.azure.com/#view/SqAzureExtension/SharedCreateServer/isNewServer~/true/subscriptionId/1284e822-cc7f-4e29-94a1-0d9b7de237b3/resourceGroup...

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

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

Microsoft

Server details

Enter required settings for this server, including providing a name and location. This server will be created in the same subscription and resource group as your database.

Server name \*  .database.windows.net

Location \*

Authentication

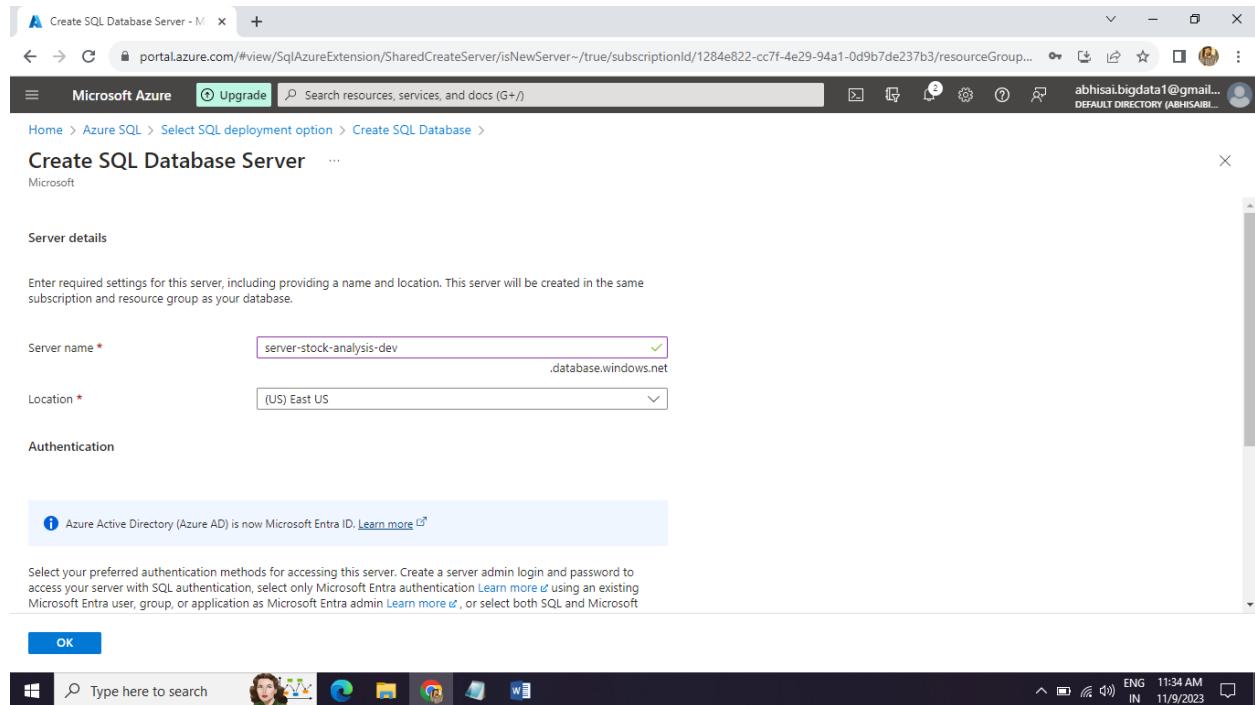
**Info** Azure Active Directory (Azure AD) is now Microsoft Entra ID. [Learn more](#)

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](#) using an existing Microsoft Entra user, group, or application as Microsoft Entra admin [Learn more](#), or select both SQL and Microsoft

**OK**

Type here to search

11:34 AM ENG IN 11/9/2023



A Create SQL Database Server - Microsoft Azure

portal.azure.com/#view/SqAzureExtension/SharedCreateServer/isNewServer~/true/subscriptionId/1284e822-cc7f-4e29-94a1-0d9b7de237b3/resourceGroup... abhisai.bigdata1@gmail... DEFAULT DIRECTORY (ABHISAI...)

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

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

## Create SQL Database Server

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](#) 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 \*: sqladmin

Password \*:

Confirm password \*:

OK

A Create SQL Database - Microsoft Azure

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

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

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

## Create SQL Database

Project details

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

Subscription \*: Free Trial

Resource group \*: rg-stock-analysis-dev

General Purpose (GP\_S\_Gen5\_1)

Cost per GB (in INR)	9.03
Max storage selected (in GB)	x 41.6

ESTIMATED STORAGE COST / MONTH: 375.71 INR

COMPUTE COST / VCORE SECOND: 0.011382 INR

NOTES

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

Database details

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

Database name \*: db-stock-analysis-dev

Server \*: (new) server-stock-analysis-dev (East US)

Want to use SQL elastic pool?  Yes  No

Review + create Next : Networking >

Windows Type here to search ENG IN 11:34 AM 11/9/2023

Create SQL Database

Want to use SQL elastic pool?  Yes  No

Workload environment  Development  Production

Compute + storage \*  Basic 2 GB storage [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.

Locally-redundant backup storage  Zone-redundant backup storage

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

Estimated cost per month  
384.63 INR

Cost summary

Basic (Basic)	76.93
Cost per DTU (in INR)	76.93
DTUs selected	x 5
ESTIMATED COST / MONTH	384.63 INR

Basics Networking Security Additional settings Tags [Review + create](#)

Product details

SQL database by Microsoft

Terms of use | Privacy policy

Estimated cost per month  
384.63 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 rg-stock-analysis-dev

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

Once activated we need to enable azure services access

The screenshot shows the Azure portal interface for managing a SQL server named 'server-stock-analysis-dev'. The left sidebar lists 'Azure SQL' under 'Default Directory' with options like '+ Create', 'Reservations', and 'Locks'. The main content area is titled 'server-stock-analysis-dev | Networking' and contains a table for 'Virtual network rules'. A note says 'No virtual network rules found.' Below this is a section for 'Firewall rules' with options to 'Add your client IPv4 address' or 'Add a firewall rule'. Under 'Exceptions', there is a checked checkbox for 'Allow Azure services and resources to access this server'. At the bottom are 'Save' and 'Discard' buttons.

Create table stock\_analytics:

```
create table stock_analytics (isin varchar(12),crt_date TIMESTAMP,todays_buy_quantity int,todays_sell_quantity int,previous_month_buy_quantity int,previous_month_sell_quantity int,mf_buy_change float,mf_sell_change float,change_from_previous_day float,change_from_previous_month float,change_from_52week_high float,change_from_52week_low float)
```

Add user ID and password to Azure Key Vault

The screenshot shows the Microsoft Azure Key Vault interface for the 'kv-stockanalysis-dev' vault. On the left, there's a sidebar with options like 'Create', 'Events', 'Objects' (with 'Keys' selected), 'Secrets', 'Certificates', 'Settings', 'Properties', 'Locks', and 'Monitoring'. The main area displays a table of secrets:

Name	Type	Status	Expiration date
azure-sql-user-password		✓ Enabled	
azure-sql-user-id		✓ Enabled	
aws-s3-bhavcopy-id		✓ Enabled	
aws-s3-bhavcopy-value		✓ Enabled	
databricks-access-key		✓ Enabled	
storageaccount-sas-key		✓ Enabled	

A message at the top right says: 'Creating the secret 'azure-sql-user-password'. The secret 'azure-sql-user-password' has been successfully created.'

The screenshot shows the Microsoft Azure SQL Database Query Editor for the 'db-stock-analysis-dev' database. The left sidebar shows the object explorer with 'Tables' expanded, showing 'dbo.stock\_analytics'. The main area contains a query editor window with the following code:

```

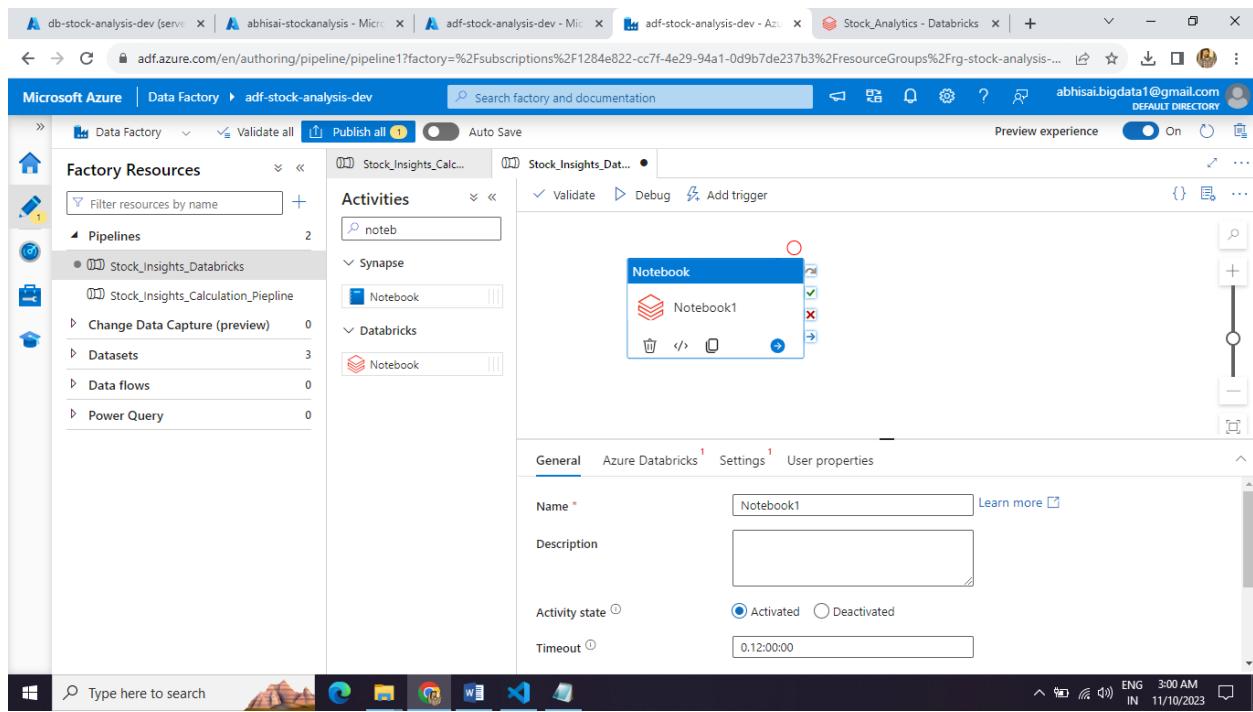
1 create table stock_analytics (isin varchar(12),crt_date TIMESTAMP,todays_buy_quantity int,todays_sell_quantity int,previous_month_buy_quantity int,previous_month_sell_quantity int,mf_buy_change float,mf_sell_change float,change from previous day)
2
3 select * from [dbo].[stock_analytics];
4
5
6
7
8

```

The 'Messages' tab shows the result: 'Query succeeded: Affected rows: 0'. The status bar at the bottom indicates the date as 11/10/2023.

Now again Pipeline Creation for this Databricks:

Create a new pipeline with NOTEBOOK as shown:



Add Azure Databricks dataset and linked service

Microsoft Azure | Data Factory > adf-stock-analysis-dev

Search factory and documentation

New linked service

Azure Databricks [Learn more](#)

Name \* Is\_Azure\_Databricks

Description

Connect via integration runtime \* [AutoResolveIntegrationRuntime](#)

Account selection method \*  From Azure subscription  Enter manually

Azure subscription \* [Free Trial \(1284e822-cc7f-4e29-94a1-0d9b7de237b3\)](#)

Databricks workspace \* [databricks-stock-analysis-dev](#)

Select cluster  New job cluster  Existing interactive cluster  Existing instance pool

**Create** **Cancel**

Connection successful Test connection

Type here to search

11/10/2023 3:02 AM ENG IN

Microsoft Azure | Data Factory > adf-stock-analysis-dev

Search factory and documentation

New linked service

Azure Databricks [Learn more](#)

Select cluster  New job cluster  Existing interactive cluster  Existing instance pool

Databrick Workspace URL \* <https://adb-320085348881679.19.azuredatabricks.net>

Authentication type \* Access Token

Access token **Azure Key Vault**

AKV linked service \* [Is\\_Azure\\_Key\\_Vault](#)

Secret name \* databricks-access-key

Secret version \* Latest version

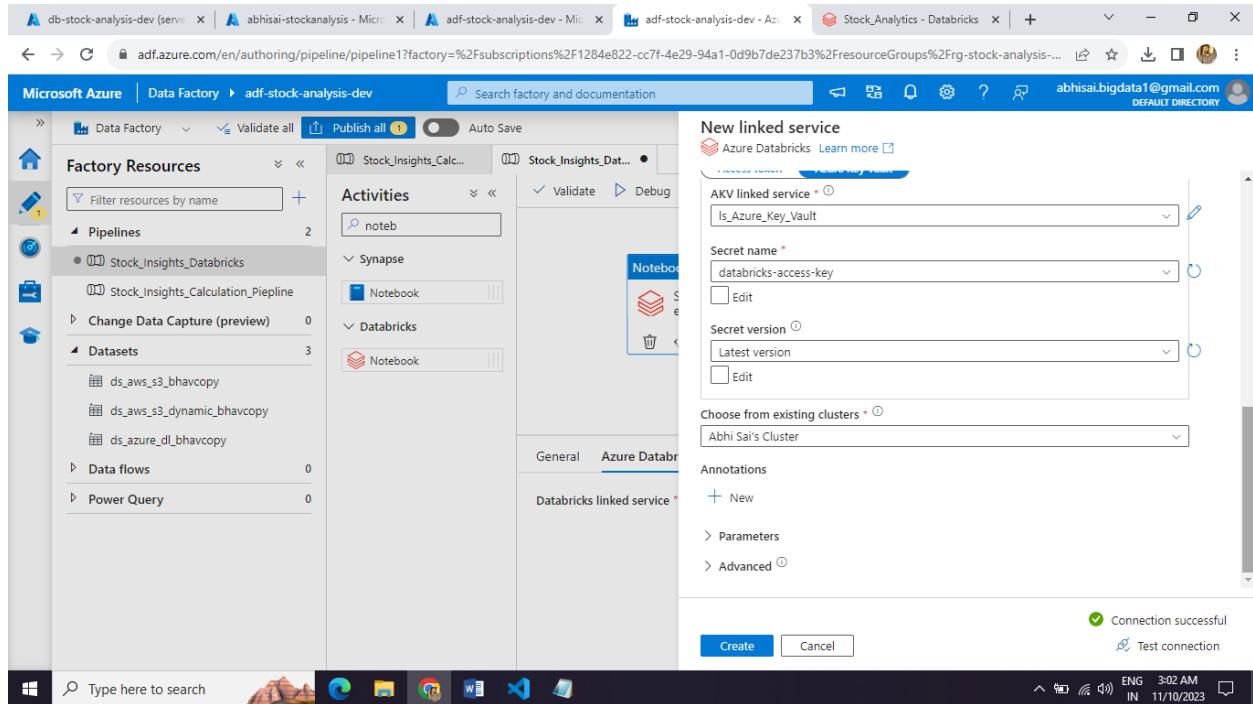
Change from existing cluster \* [Create new cluster](#)

**Create** **Cancel**

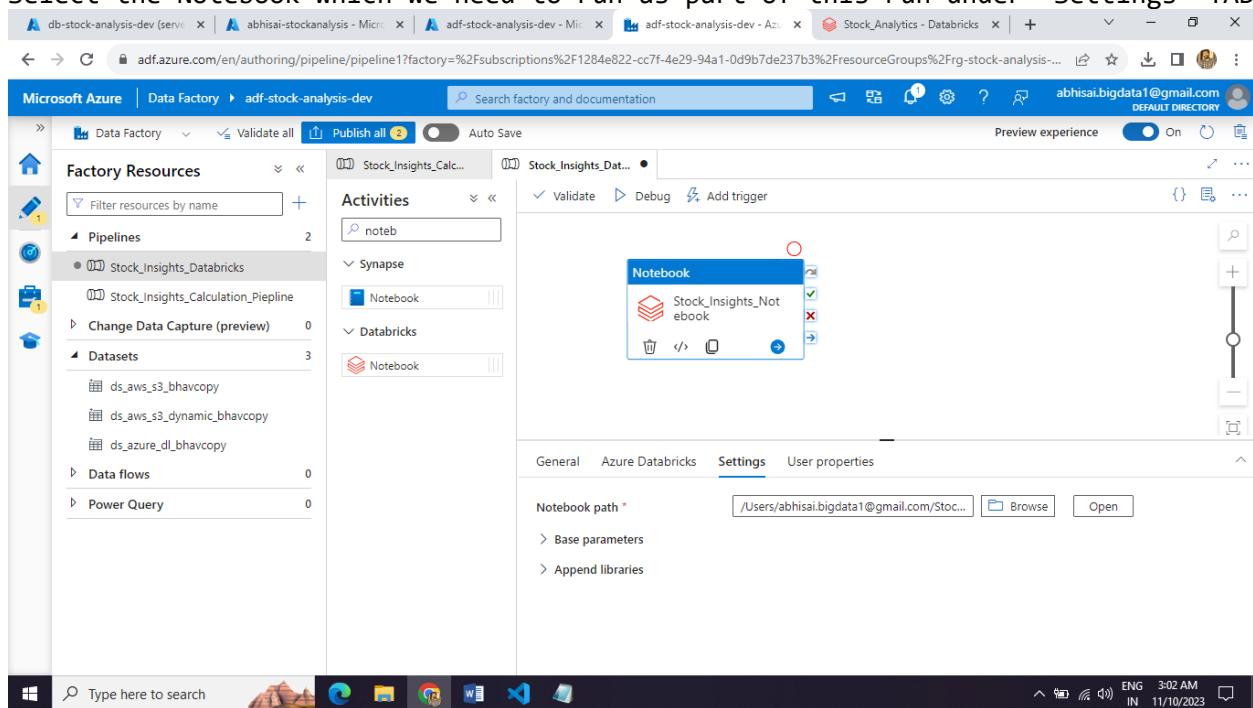
Connection successful Test connection

Type here to search

11/10/2023 3:02 AM ENG IN



Select the Notebook which we need to run as part of this run under “Settings” TAB



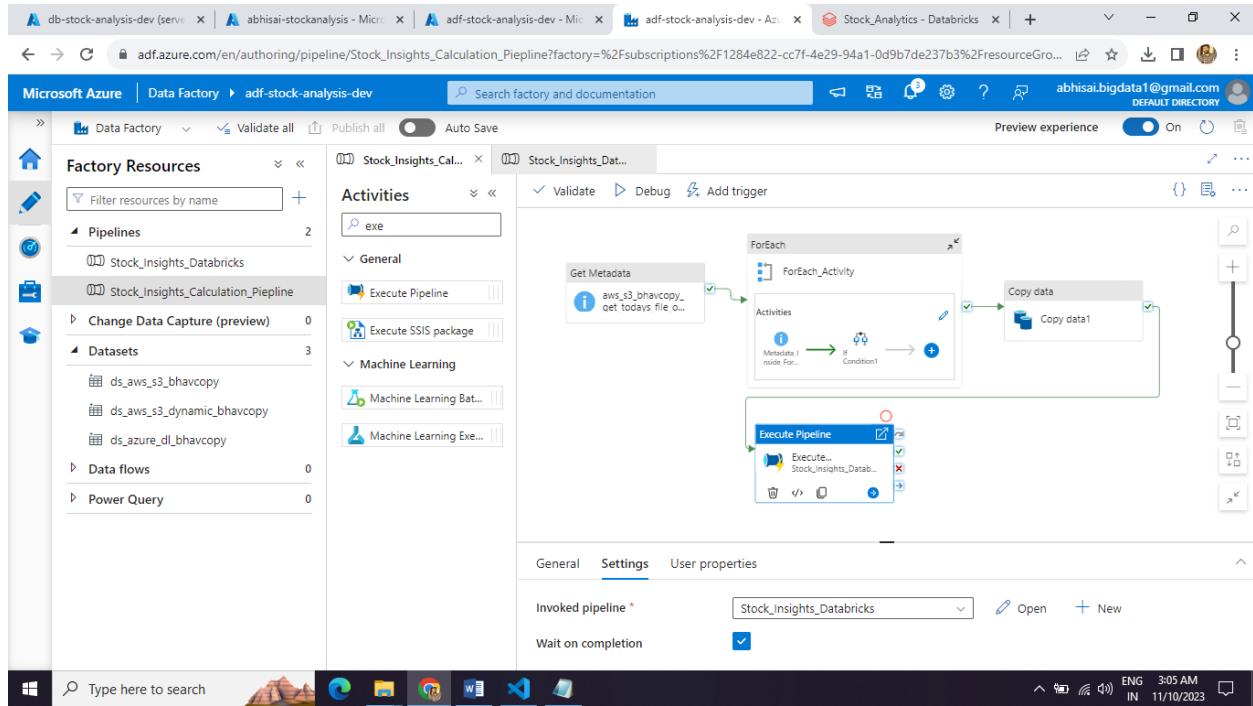
Now add this new pipeline to our original pipeline

The screenshot shows the Microsoft Azure Data Factory pipeline editor. The left sidebar displays 'Factory Resources' with sections for Pipelines, Datasets, and Data flows. The 'Activities' pane on the right shows the current pipeline structure:

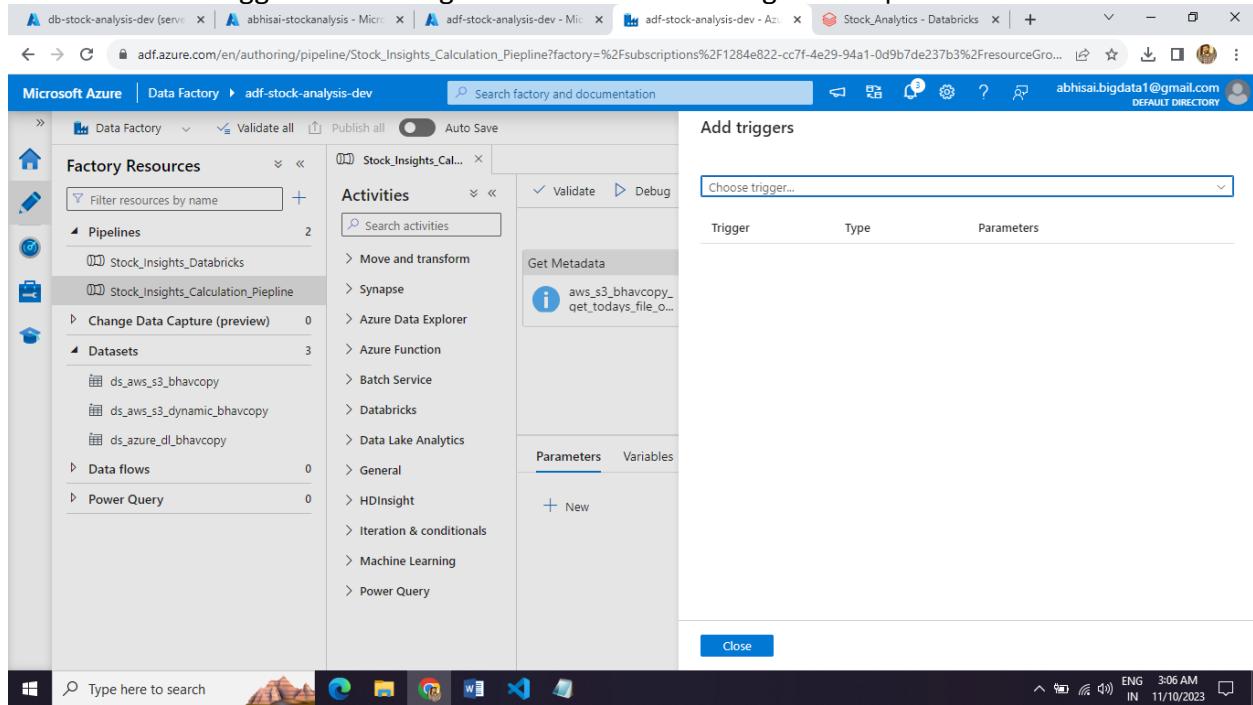
```
graph TD; GetMetadata[Get Metadata] --> ForEach[ForEach]; ForEach --> IfCondition[If Condition]; IfCondition --> CopyData1[Copy data]; IfCondition --> ExecutePipeline1[Execute Pipeline]; ExecutePipeline1 --> End(( ));
```

The pipeline is named 'Stock\_Insights\_Calculation\_Pipeline'. In the 'Settings' tab of the pipeline details, the 'Name' field is set to 'Execute Stock\_Insights\_Databricks\_Pipeline'.

Select the pipeline under “Settings”



Now add the Trigger as Storage event for this original Pipeline:



The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists Pipelines, Datasets, and Data flows. The main area displays a pipeline named 'Stock\_Insights\_Calculation\_Pipeline'. A 'Get Metadata' activity is selected. To the right, a 'New trigger' dialog is open, configured to trigger on 'Storage events' from an 'Azure subscription' (Free Trial) using the 'From Azure subscription' method. The trigger is set to monitor the 'stockanalyststorageacct' container for blobs starting with 'zerodha\_daily\_positions/' and ending with '.csv'. The 'Event' section includes a checked checkbox for 'blob created'.

The screenshot shows the same Microsoft Azure Data Factory interface. The 'Data preview' dialog is open, displaying a list of 22 blobs matched in the 'abhisai-stockanalysis' container. The blobs are listed as follows:

Blob name
1 zerodha_daily_positions/transactions_2023-10-09.csv
2 zerodha_daily_positions/transactions_2023-10-10.csv
3 zerodha_daily_positions/transactions_2023-10-11.csv
4 zerodha_daily_positions/transactions_2023-10-12.csv
5 zerodha_daily_positions/transactions_2023-10-13.csv
6 zerodha_daily_positions/transactions_2023-10-13.csv

Upload bhavcopy to AWS S3 and transactions file to Azure container.  
Before Run:

The screenshot shows the Microsoft Azure Storage Container Overview page for the 'abhisai-stockanalysis' container. The browser tab bar includes multiple tabs for different Azure services like DB, AD, and Storage. The main navigation bar has 'Microsoft Azure' and 'Upgrade' buttons, and a search bar. The left sidebar has 'Overview', 'Diagnose and solve problems', 'Access Control (IAM)', and 'Settings' sections. The 'Settings' section is expanded, showing 'Shared access tokens', 'Manage ACL', 'Access policy', 'Properties', and 'Metadata'. The main content area displays blob details with columns: Name, Modified, Access tier, Archive status, Blob type, Size, and Lease state. Two blobs are listed: 'nse\_daily\_bhavcopy' and 'zerodha\_daily\_positi...'. A search bar at the top allows filtering by prefix.

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
nse_daily_bhavcopy					-	***
zerodha_daily_positi...					-	***



The image shows two side-by-side screenshots of Microsoft Azure services.

**Left Screenshot: Microsoft Azure SQL Query editor (preview)**

This screenshot shows the Azure SQL Query editor interface. The top navigation bar includes tabs for Home, Azure SQL, server-stock-analysis-dev | SQL databases, db-stock-analysis-dev (server-stock-analysis-dev/db-stock-analysis-dev), and Query editor (preview). The main area displays a query editor titled "Query 1" containing the following T-SQL code:

```
1 create table stock_analytics (isin varchar(12),crt_date TIMESTAMP,todays_buy_quantity int,todays_sell_quantity int,previous_month_buy_quantity int,previous_month_sell_quantity int,mtf_buy_change float,mtf_sell_change float,change from previous day)
2
3 select * from [dbo].[stock_analytics];
```

The results pane below shows the message: "Query succeeded: Affected rows: 0".

**Right Screenshot: Microsoft Azure Data Factory Pipeline runs**

This screenshot shows the Azure Data Factory Pipeline runs interface. The top navigation bar includes tabs for Microsoft Azure, Data Factory, and adf-stock-analysis-dev. The main area displays a "Pipeline runs" list with the following details:

- Triggered tab is selected.
- Filter by run ID or name: Chennai, Kolkata, Mu... : Last 24 hours.
- Pipeline name: All.
- Status: All.
- Showing 0 - 0 items.
- Table headers: Pipeline name ↑, Run start ↓, Duration, Triggered by, Status ↑, Parameters, Run ID.
- A magnifying glass icon with the text "No results to show".
- A note: "If you expected to see results, try changing your filters."

Uploaded to S3

The screenshot shows the AWS S3 console interface. At the top, there are several tabs open, including 'db-stock-analysis-de', 'abhisai-stockanalysis', 'adf-stock-analysis-de', 'adf-stock-analysis-de', 'Stock\_Analytics - Data', and 'Upload objects - S3 b...'. The main window title is 'Upload objects - S3 b...' and the URL is 's3.console.aws.amazon.com/s3/upload/abhisai-stockanalysis?region=us-east-1&prefix=bhavcopy/'. The navigation bar includes 'Services' (selected), 'EC2', and 'S3'. The status bar shows '2023-11' and '1/1'. A green banner at the top says 'Upload succeeded' with a link to 'View details below.' Below this, a summary table shows 'Destination s3://abhisai-stockanalysis/bhavcopy/' with 'Succeeded' (1 file, 1.9 MB (100.00%)) and 'Failed' (0 files, 0 B (0%)). A table titled 'Files and folders (1 Total, 1.9 MB)' lists the uploaded file 'transactions\_2023-11-09.csv' with details: Name (transactions\_2023-11-09.csv), Folder (-), Type (text/csv), Size (1.9 MB), Status (Succeeded), and Error (-). The bottom of the screen shows a Windows taskbar with various icons.

We can see bhavcopy 09112023 is not available in Azure

The screenshot shows the Microsoft Azure Storage Container blade. The URL in the address bar is 'https://portal.azure.com/#view/Microsoft\_Azure\_Storage/ContainerMenuBlade/~/overview/storageAccountId/%2Fsubscriptions%2F1284e822-cc7f-4e29-9...'. The left sidebar shows 'Overview', 'Diagnose and solve problems', 'Access Control (IAM)', 'Shared access tokens', 'Manage ACL', 'Access policy', 'Properties', and 'Metadata'. The main area displays a table of files in the 'abhisai-stockanalysis' container. The columns are Name, Modified, Access tier, Archive status, Blob type, and Size. There are 11 entries, all named 'NSE\_CM\_bhavcopy\_xx112023.csv' where xx is a two-digit number from 01 to 11. All files were modified on 11/9/2023 at different times between 3:29:15 PM and 3:32:15 PM, and they are all 'Hot (Inferred)' in the Access tier, 'Block blob' in Blob type, and have a size of 276, 277, or 281 MB. The bottom of the screen shows a Windows taskbar with various icons.

Uploaded transactions:

**Container:** abhisai-stockanalysis

**Authentication method:** Access key (Switch to Microsoft Entra user account)

**Location:** abhisai-stockanalysis / zerotha\_daily\_positions

Name	Modified	Access tier	Archive status	Blob type	Size
[..]					
transactions_2023-10-09.csv	11/10/2023, 1:59:17 AM	Hot (Inferred)		Block blob	1.9 MiB
transactions_2023-10-10.csv	11/10/2023, 1:59:16 AM	Hot (Inferred)		Block blob	1.9 MiB
transactions_2023-10-11.csv	11/10/2023, 1:59:11 AM	Hot (Inferred)		Block blob	1.9 MiB
transactions_2023-10-12.csv	11/10/2023, 1:59:11 AM	Hot (Inferred)		Block blob	1.9 MiB
transactions_2023-10-13.csv	11/10/2023, 1:59:11 AM	Hot (Inferred)		Block blob	1.9 MiB
transactions_2023-10-16.csv	11/10/2023, 1:59:13 AM	Hot (Inferred)		Block blob	1.9 MiB
transactions_2023-10-17.csv	11/10/2023, 1:59:12 AM	Hot (Inferred)		Block blob	1.9 MiB
transactions_2023-10-18.csv	11/10/2023, 1:59:12 AM	Hot (Inferred)		Block blob	1.9 MiB
transactions_2023-10-19.csv	11/10/2023, 1:59:13 AM	Hot (Inferred)		Block blob	1.9 MiB

## Pipeline Triggered:

**Pipeline runs**

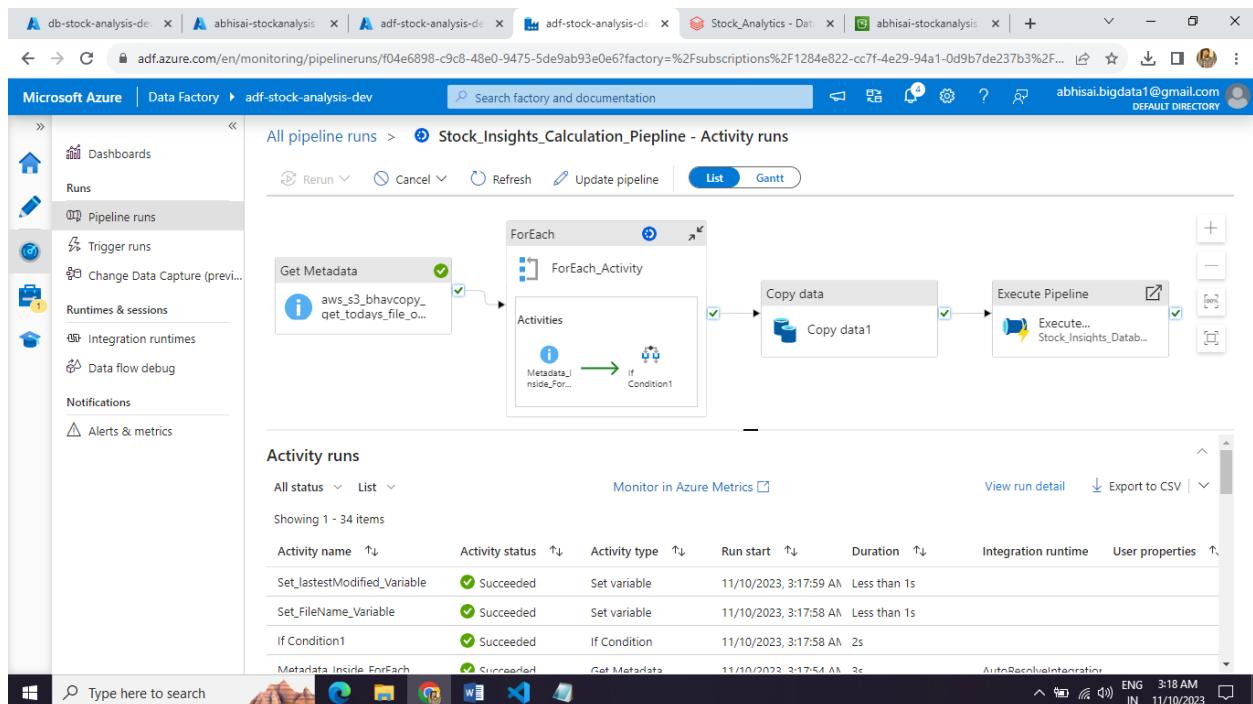
**Triggered**   **Debug**   **Rerun**   **Cancel options**   **Refresh**   **Edit columns**   **List**   **Gantt**

**Filter by run ID or name**: Chennai, Kolkata, Mu... : Last 24 hours   **Pipeline name**: All   **Status**: All

**Runs**: Latest runs   **Triggered by**: All   **Add filter**

Last refreshed 0 minutes ago

Pipeline name	Run start	Duration	Triggered by	Status	Parameters	Run ID
Stock_Insights_Calculation_Pi...	11/10/2023, 3:16:58 AM	15s	Stock_Insights_Trigger	In progress		f04e6898-c9c8-48



The screenshot shows the Microsoft Azure Data Factory interface. On the left, a sidebar lists navigation options: Dashboards, Runs, Pipeline runs (selected), Trigger runs, Change Data Capture (previous), Runtimes & sessions, Integration runtimes, Data flow debug, Notifications, and Alerts & metrics. The main area displays the 'Details' page for a specific activity run. The title is 'Copy Activity completed from AWS Latest File.' The activity run ID is dd810e1d-0e7d-4d0f-af4e-2132780ce734. The activity details show a successful transfer from 'Amazon S3' to 'Azure Data Lake Storage Gen2' in the 'Region: East US'. Key metrics include:

- Data read: 1.989 MB
- Files read: 1
- Peak connections: 1
- Data written: 1.989 MB
- Files written: 1
- Peak connections: 1

Performance metrics:

- Copy duration: 00:00:56
- Throughput: 663.028 KB/s

Log details for the transfer:

- Start time: 11/10/2023, 3:19:58 AM
- Used DIUs: 4
- Used parallel copies: 1
- Duration: 00:00:56

At the bottom, there is a satisfaction survey: 'How satisfied or dissatisfied are you with the performance of this copy activity?' with a scale from 1 to 5.

Now we can see the file in Azure

The screenshot displays two windows side-by-side. The left window shows the Microsoft Azure Storage Container 'abhisai-stockanalysis' with a list of CSV files. The right window shows the Microsoft Azure Data Factory 'Stock\_Insights\_Calculation\_Pipeline' activity runs.

**Microsoft Azure Storage Container (Left Window):**

- Overview:** Shows 18 files in the container.
- Metadata:** Selected tab, showing detailed information for each file, such as name, last modified, inferred type, and size.

File Name	Last Modified	Type	Size
NSE_CM_bhavcopy_04102023.csv	11/9/2023, 3:29:15 PM	Hot (Inferred)	Block blob 277.22 KiB
NSE_CM_bhavcopy_05102023.csv	11/9/2023, 3:29:15 PM	Hot (Inferred)	Block blob 275.1 KiB
NSE_CM_bhavcopy_06102023.csv	11/9/2023, 3:29:15 PM	Hot (Inferred)	Block blob 277.9 KiB
NSE_CM_bhavcopy_06112023.csv	11/9/2023, 3:29:16 PM	Hot (Inferred)	Block blob 236.74 KiB
NSE_CM_bhavcopy_07112023.csv	11/9/2023, 3:30:21 PM	Hot (Inferred)	Block blob 278.65 KiB
NSE_CM_bhavcopy_08112023.csv	11/9/2023, 3:32:15 PM	Hot (Inferred)	Block blob 277.92 KiB
NSE_CM_bhavcopy_09102023.csv	11/9/2023, 3:29:16 PM	Hot (Inferred)	Block blob 278.74 KiB
<b>NSE_CM_bhavcopy_09112023.csv</b>	<b>11/10/2023, 3:20:54 AM</b>	<b>Hot (Inferred)</b>	<b>Block blob 1.9 MiB</b>
NSE_CM_bhavcopy_10102023.csv	11/9/2023, 3:29:15 PM	Hot (Inferred)	Block blob 276.72 KiB
NSE_CM_bhavcopy_11102023.csv	11/9/2023, 3:29:14 PM	Hot (Inferred)	Block blob 277.58 KiB
NSE_CM_bhavcopy_12102023.csv	11/9/2023, 3:29:14 PM	Hot (Inferred)	Block blob 277.61 KiB
NSE_CM_bhavcopy_13102023.csv	11/9/2023, 3:29:14 PM	Hot (Inferred)	Block blob 276.26 KiB
NSE_CM_bhavcopy_16102023.csv	11/9/2023, 3:29:14 PM	Hot (Inferred)	Block blob 281.11 KiB
NSE_CM_bhavcopy_17102023.csv	11/9/2023, 3:29:14 PM	Hot (Inferred)	Block blob 279.38 KiB
18102023.csv	11/9/2023, 3:29:14 PM	Hot (Inferred)	Block blob 277.71 KiB

**Microsoft Azure Data Factory (Right Window):**

- Pipeline Runs:** Shows the 'Stock\_Insights\_Calculation\_Pipeline' run details.
- Activity Runs:** Shows a list of activity runs for the pipeline, including 'Copy data1', 'If Condition1', 'Metadata\_InsideForEach', and 'If Condition1'.

We can see Databricks run stats in Databricks as well:

Microsoft Azure | databricks | Search data, notebooks, recents, and more... CTRL + P databricks-stock-analysis-dev abhisai.bigdata1@gmail.com

**Workflows**

**Jobs Job runs**

Only my job runs Run status 2023-11-08 03:30:00 → 2023-11-10 03:30:00

**Create job**

1  
1  
0 8 Nov, 12 PM 9 Nov, 12 AM 9 Nov, 12 PM 10 Nov, 12 AM  
Failed Skipped Succeeded

Start time	Job	Run as	Launched	Duration	Status	Run parameters	⋮
Nov 10, 2023, 03:21 AM	ADF_adf-stock-analys...	Abhi Sai	By runs submit API	56s	Succeeded		⋮

Previous Next ↻ ↺ ENG 3:23 AM IN 11/10/2023 ↻ ↺

Type here to search

Microsoft Azure | databricks | Search data, notebooks, recents, and more... CTRL + P databricks-stock-analysis-dev abhisai.bigdata1@gmail.com

**Workflows > Runs >**

**ADF\_adf-stock-analysis-dev\_Stock\_Insights\_Databricks\_Stock\_Insights\_Notebook\_141768b1-7d3a-4a03-99d9-9d6bd0285a23 run**  Delete job run

**Output**

Hide code  Export as HTML

```
from pyspark.sql import SparkSession
from pyspark.sql.functions import *
from pyspark.sql.functions import input_file_name
from pyspark.sql.types import IntegerType, BooleanType, DateType
from datetime import *
from pyspark.sql.functions import expr, col
```

Command took 0.04 seconds

```
spark = SparkSession.builder.getOrCreate()
```

Command took 0.02 seconds

```
spark.conf.set("fs.azure.account.auth.type", "stockanalyststorageacctdfs.core.windows.net", "SAS")
spark.conf.set("fs.azure.sas.token.provider.type", "stockanalyststorageacctdfs.core.windows.net", "org.apache.hadoop.fs.azure.SimpleSasProvider")
```

**Task run details**

Job ID	366894868494281
Task run ID	54637095337156
Run as	Abhi Sai
Launched	By runs submit API
Started	11/10/2023, 03:21:00 AM
Ended	11/10/2023, 03:21:57 AM
Duration	56s
Queue duration	-
Status	Succeeded

**Notebook**

[/Users/abhisai.bigdata1@gmail.com/Stock\\_Analytics](#)

Type here to search

The screenshot shows the Microsoft Azure Storage Container Overview page for the 'abhisai-stockanalysis' container. The container name is displayed at the top. On the left, there's a sidebar with options like Overview, Diagnose and solve problems, Access Control (IAM), Settings (Shared access tokens, Manage ACL, Access policy, Properties, Metadata), and a search bar. The main area displays a table of blobs. The table has columns: Name, Modified, Access tier, Archive status, Blob type, and Size. The blobs listed are: final\_visual\_output, nse\_daily\_bhavcopy, positions\_analytics, price\_analytics, and zerodha\_daily\_positions. Each blob has a small checkbox icon next to it. At the bottom of the table, there are navigation arrows.

**Problem:** When we checked SQL its showing 0 records so verified and found that prices\_df is empty.

Lets see what is wrong.

I wrong uploaded transactions file to S3 instead of Bhav file

The screenshot shows the AWS S3 console interface. On the left, there's a sidebar with options: Buckets, Object Lambda Access Points, Multi-Region Access Points, Batch Operations, IAM Access Analyzer for S3, Block Public Access settings for this account, Storage Lens, Dashboards, AWS Organizations settings, Feature spotlight, CloudShell, and Feedback. The main area is titled 'Objects (27)' and shows a list of CSV files. The table has columns: Name, Type, Last modified, Size, and Storage class. The files listed are: NSE\_CM\_bhavcopy\_09 102023.csv, NSE\_CM\_bhavcopy\_09 112023.csv, NSE\_CM\_bhavcopy\_10 102023.csv, and NSE\_CM\_bhavcopy\_11 102023.csv. The 'NSE\_CM\_bhavcopy\_09 112023.csv' file is selected, indicated by a blue border around its row. At the top of the object list, there are buttons for Upload, Copy S3 URI, Copy URL, Download, Open, Delete, Actions, and Create folder. There's also a search bar labeled 'Find objects by prefix'. The bottom of the screen shows a Windows taskbar with various icons and a status bar indicating the date and time.

Lets delete and rerun the pipeline.

Deleted all outputs

**abhisai-stockanalysis** Container

Search

Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

Authentication method: Access key (Switch to Microsoft Entra user account)  
Location: abhisai-stockanalysis

Search blobs by prefix (case-sensitive) Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
nse_daily_bhavcopy					-	---
zerodha_daily_positi...					-	---

Type here to search

Uploaded again:

2023-11-09

abhisai-stockanalysis Container

Search

Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

Name	Modified	Access tier	Archive status	Blob type	Size
transactions_2023-10-20.csv	11/10/2023, 1:59:13 AM	Hot (Inferred)		Block blob	1.9 MiB
transactions_2023-10-23.csv	11/10/2023, 1:59:15 AM	Hot (Inferred)		Block blob	1.9 MiB
transactions_2023-10-25.csv	11/10/2023, 1:59:14 AM	Hot (Inferred)		Block blob	1.9 MiB
transactions_2023-10-26.csv	11/10/2023, 1:59:14 AM	Hot (Inferred)		Block blob	1.9 MiB
transactions_2023-10-27.csv	11/10/2023, 1:59:15 AM	Hot (Inferred)		Block blob	1.9 MiB
transactions_2023-10-30.csv	11/10/2023, 1:59:15 AM	Hot (Inferred)		Block blob	1.9 MiB
transactions_2023-10-31.csv	11/10/2023, 1:59:17 AM	Hot (Inferred)		Block blob	1.9 MiB
transactions_2023-11-01.csv	11/10/2023, 1:59:16 AM	Hot (Inferred)		Block blob	1.9 MiB
transactions_2023-11-02.csv	11/10/2023, 1:59:17 AM	Hot (Inferred)		Block blob	1.9 MiB
transactions_2023-11-03.csv	11/10/2023, 1:59:17 AM	Hot (Inferred)		Block blob	1.9 MiB
transactions_2023-11-06.csv	11/10/2023, 1:59:19 AM	Hot (Inferred)		Block blob	1.9 MiB
transactions_2023-11-07.csv	11/10/2023, 1:59:19 AM	Hot (Inferred)		Block blob	1.9 MiB
transactions_2023-11-08.csv	11/10/2023, 1:59:18 AM	Hot (Inferred)		Block blob	1.9 MiB
transactions_2023-11-09.csv	11/10/2023, 3:31:21 AM	Hot (Inferred)		Block blob	1.9 MiB

Type here to search

So that automatically our pipeline runs.

3:31 AM IN 11/10/2023

The screenshot shows the Microsoft Azure Data Factory interface. The left sidebar has a 'Runs' section with 'Pipeline runs' selected. The main area displays 'Pipeline runs' for the 'Stock\_Insights\_Calculation\_Pipeline'. It lists three runs:

Pipeline name	Run start	Duration	Triggered by	Status	Run ID
Stock_Insights_Calculation_Pi...	11/10/2023, 3:31:24 AM	41s	Stock_Insights_Trigger	In progress	c85b6231-c373-4f...
Stock_Insights_Databricks	11/10/2023, 3:20:57 AM	1m 7s	5d9a5b9-3132-4d9f...	Succeeded	8f901b46-a412-4c...
Stock_Insights_Calculation_Pi...	11/10/2023, 3:16:58 AM	5m 7s	Stock_Insights_Trigger	Succeeded	f04e6898-c9c8-48...

The screenshot shows the Microsoft Azure Data Factory interface, specifically the 'Activity runs' page for the 'Stock\_Insights\_Calculation\_Pipeline'. The pipeline structure is visible at the top:

```

graph LR
    GetMetadata[Get Metadata] --> ForEach[ForEach]
    ForEach --> ForEachActivity[ForEach_Activity]
    ForEachActivity --> IfCondition[If Condition1]
    IfCondition --> CopyData[Copy data]
    CopyData --> ExecutePipeline[Execute Pipeline]
    ExecutePipeline --> End[...]
  
```

The activity runs table below shows 148 items:

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User properties
If Condition1	Succeeded	If Condition	11/10/2023, 3:33:11 AM	Less than 1s		
Metadata_Inside_Foreach	Succeeded	Get Metadata	11/10/2023, 3:33:06 AM	4s	AutoResolveIntegratio...	
Set_LastModified_Variable	Succeeded	Set variable	11/10/2023, 3:33:05 AM	Less than 1s		
Set_FileName_Variable	Succeeded	Set variable	11/10/2023, 3:33:05 AM	Less than 1s		
If Condition1	Succeeded	If Condition	11/10/2023, 3:33:04 AM	2s		

Now correctly copied bhavfile

Screenshot of the Microsoft Azure Data Factory Details page for a copy activity run.

**Activity run details:**

- Source:** Amazon S3
- Destination:** Azure Data Lake Storage Gen2
- Status:** Succeeded
- Azure IR region:** East US

**Data read:** 286.313 KB (1 file, 1 connection)

**Data written:** 286.313 KB (1 file, 1 connection)

**Performance metrics:**

- Copy duration: 00:00:11
- Throughput: 143.156 KB/s

**Log message:** How satisfied or dissatisfied are you with the performance of this copy activity?

Screenshot of the Microsoft Azure Data Factory Pipeline Runs page for the "Stock\_Insights\_Calculation\_Pipeline".

**Pipeline Run Details:**

- Pipeline:** Stock\_Insights\_Calculation\_Pipeline
- Run Type:** Activity runs

**Activities in the Pipeline:**

```

graph LR
    GetMetadata[Get Metadata] --> ForEach[ForEach]
    ForEach --> CopyData[Copy data]
    CopyData --> ExecutePipeline[Execute Pipeline]
    
```

**Activity Runs:**

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User properties
Execute Stock_Insights_Datab...	Succeeded	Execute Pipeline	11/10/2023, 3:35:07 AM	1m 23s		
Copy data1	Succeeded	Copy data	11/10/2023, 3:34:51 AM	15s	AutoResolveIntegration	
If Condition1	Succeeded	If Condition	11/10/2023, 3:34:45 AM	Less than 1s		
Metadata_Inside_ForEach	Succeeded	Get Metadata	11/10/2023, 3:34:40 AM	4s	AutoResolveIntegration	
If Condition1	Succeeded	If Condition	11/10/2023, 3:34:38 AM	1s		

Final Output :

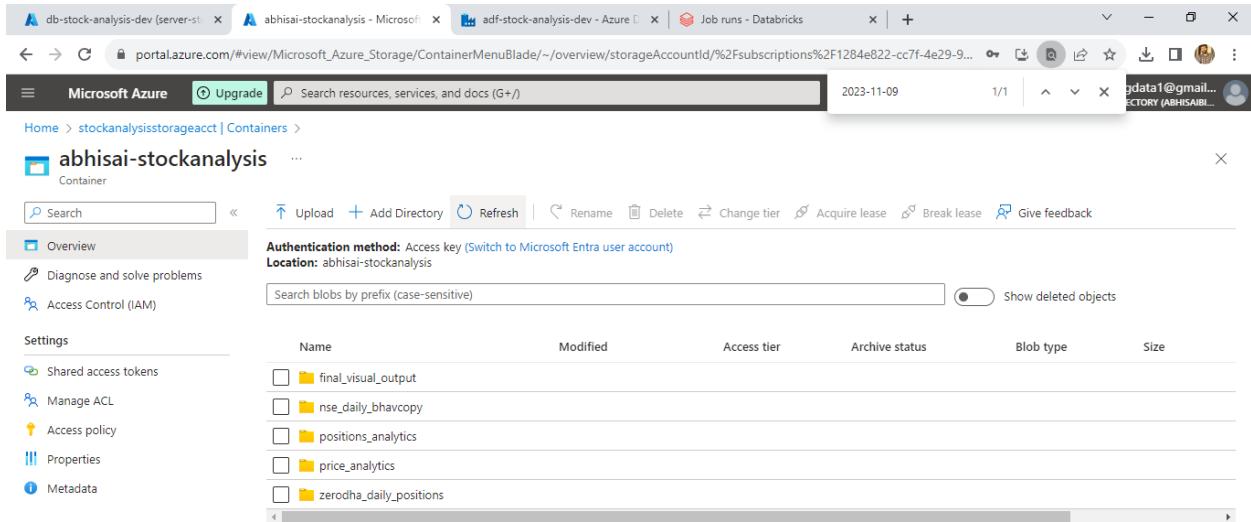
The screenshot shows the Microsoft Azure SQL Query editor interface. The title bar reads "db-stock-analysis-dev (server-stock-analysis-dev/db-stock-analysis-dev) | Query editor (preview)". The main area displays a table with the following columns: isin, crt\_date, todays\_buy\_total\_positions, todays\_sell\_total\_positions, previous\_month\_buy\_total\_pos..., previous\_month\_sell\_total\_pos..., and mf\_buy\_c. The data consists of eight rows of stock information. Below the table, a message bar indicates "Query succeeded | 2s". The system tray at the bottom right shows the date as 11/10/2023 and the time as 3:37 AM.

isin	crt_date	todays_buy_total_positions	todays_sell_total_positions	previous_month_buy_total_pos...	previous_month_sell_total_pos...	mf_buy_c
IN9155A01020	2023-11-09T22:05:53.323000	10708	4204	154818	144046	325.69
INE001B01026	2023-11-09T22:05:53.323000	10173	3632	145571	155835	14.01
INE002A01018	2023-11-09T22:05:53.323000	6111	9145	174689	186783	460.49
INE002L01015	2023-11-09T22:05:53.323000	4239	4392	150670	159178	
INE002S01010	2023-11-09T22:05:53.323000	4224	10409	138168	163375	
INE003A01024	2023-11-09T22:05:53.323000	9850	5995	185122	152851	-19.83
INE005I01014	2023-11-09T22:05:53.323000	9432	6125	156358	146215	11.5

The screenshot shows the Microsoft Azure SQL Query editor interface. The title bar reads "db-stock-analysis-dev (server-stock-analysis-dev/db-stock-analysis-dev) | Query editor (preview)". The main area displays a table with the following columns: total\_pos..., mf\_buy\_change, mf\_sell\_change, change\_from\_previous\_day, change\_from\_last\_month, change\_from\_high, and change\_from\_low. The data consists of seven rows of summary statistics. Below the table, a message bar indicates "Query succeeded | 2s". The system tray at the bottom right shows the date as 11/10/2023 and the time as 3:37 AM.

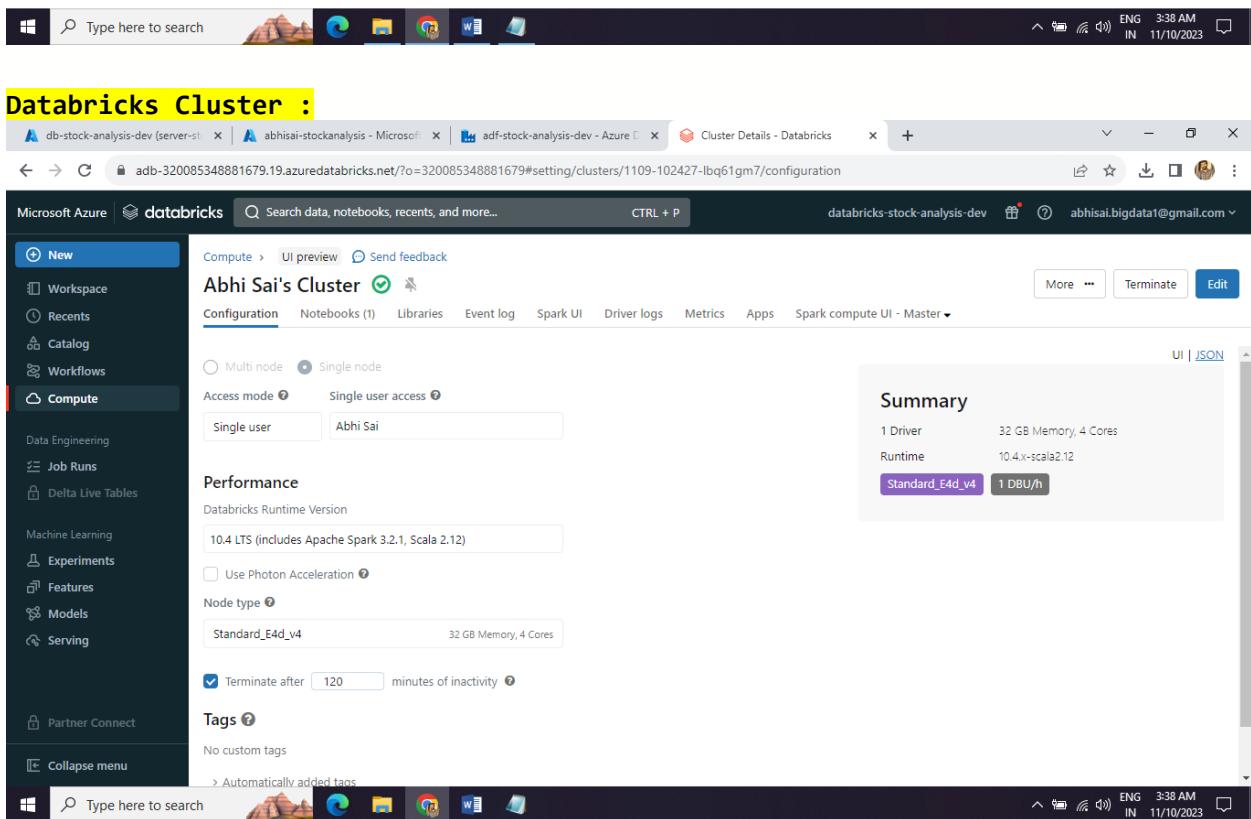
total_pos...	mf_buy_change	mf_sell_change	change_from_previous_day	change_from_last_month	change_from_high	change_from_low
325.69		2.05	6.54	0.89	9.03	
14.01		-8.96	-8.1	15.85	1.9	
460.49	324.02	-1.09	0.09	2.44	4.06	
		-3.12	4.86	5.5	17.51	
		-1.74	-7.21	11.99	5.1	
-19.83	1072.97	-1.01	-5.62	8.34	4.42	
11.5		-4.96	37.26	7.16	47.43	

Outputs are written to data lake as well:



A screenshot of the Microsoft Azure Storage Container Overview page. The container name is "abhisai-stockanalysis". The table lists blobs with the following details:

Name	Modified	Access tier	Archive status	Blob type	Size
final_visual_output					
nse_daily_bhavcopy					
positions_analytics					
price_analytics					
zerodha_daily_positions					



A screenshot of the Databricks Cluster configuration page. The cluster is named "Abhi Sai's Cluster". The configuration settings include:

- Compute mode: Single node
- Access mode: Single user access
- User: Abhi Sai
- Databricks Runtime Version: 10.4 LTS (includes Apache Spark 3.2.1, Scala 2.12)
- Node type: Standard\_E4d\_v4
- Termination settings: Terminate after 120 minutes of inactivity
- Summary: 1 Driver, 32 GB Memory, 4 Cores, Runtime 10.4-x-scal2.12, Standard\_E4d\_v4, 1 DBU/h

## Notebook:



