



TECHNOLOGY

| Unext

Presentation
for

DXC - Realtime Project

Covered under

UNext Learning

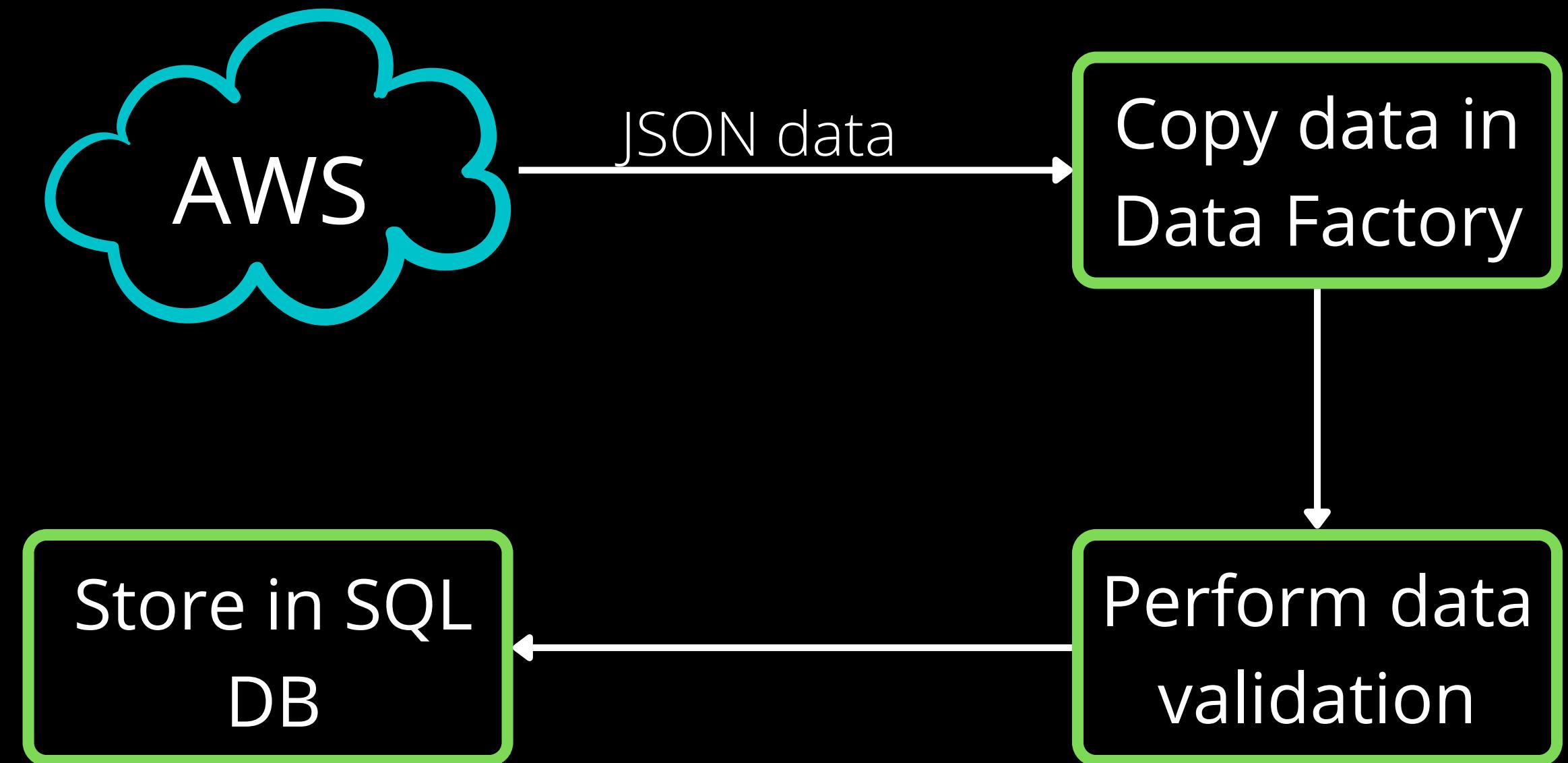
Introduction to the Projects

- • • •

- Projects Assigned
 - Smart Vehicle
 - This project is designed to store IOT data coming from AWS platform in the JSON format which will be **validated** and moved to Azure's **SQL database**
 - AP Morgan Data Platform
 - The project is designed to work as a **validation and integration** system wherein the incoming data from multiple sources will undergo data integrity check and thereby stored in **Delta table** of Azure Databricks.

Smart vehicles

- Project Description and Working
 - Smart vehicles compromise of creating a highly advance driving system with the use of IOT Data.



Smart Vehicles

- The project is covered based on different tasks allotted
 - **TASK 1 - Create Azure Data Factory Account For Data pipelines**
 - Create a new Azure Data Factory

The screenshot shows the Microsoft Azure portal interface. On the left, there is a sidebar with 'Azure services' icons: Create a resource, SQL databases, Storage accounts, Data factories (which is selected and highlighted in blue), Azure Databricks, Azure Synapse Analytics, Azure Active Directory, Virtual networks, Virtual machines, and More services. Below this is a 'Resources' section with 'Recent' and 'Favorite' tabs, showing a list of resources: newsqldatabase (newserv101/newsqldatabase), newresource, storage1011, NetworkWatcherRG, dbstoragev2h4mfceqkezg, and newbricks. At the bottom of this section is a 'See all' link. On the right, the main content area has a header 'Data factories' with a 'Create' button and filter options for Subscription, Type, Resource group, Location, and a 'No grouping' dropdown. A message at the bottom states 'No data factories to display' with a note 'Try changing or clearing your filters.'

Smart Vehicles



- **TASK 1 - Create Azure Data Factory Account For Data pipelines**
 - Enter the details about the Data Factory account and deploy
 - Open Data factory account and We have successfully created azure data factory account.

The image displays two screenshots of the Microsoft Azure portal. The left screenshot shows the 'Create Data Factory' wizard on the 'Basics' step. It requires entering project details such as Subscription (selected: Azure-DXC262AB12Lab), Resource group (selected: newresource), Name (newdata-factory), Region (East US), and Version (V2 (Recommended)). The right screenshot shows the main 'automobile' Data Factory dashboard. It features a central 3D icon of a factory building and four cards with icons and descriptions: 'Ingest' (Copy data at scale once or on a schedule.), 'Orchestrate' (Code-free data pipelines.), 'Transform data' (Transform your data using data flows.), and 'Configure SSIS' (Manage & run your SSIS packages in the cloud.).

Smart vehicles



■ TASK 2 - Create Azure SQL Server and Database

- Search and Open SQL database
- Click on + Create
- Enter the basic details for the account and proceed to networking

The image consists of two side-by-side screenshots of the Microsoft Azure portal. The left screenshot shows the main Azure services dashboard with a 'Create a resource' button, 'Storage accounts', and the 'SQL databases' icon highlighted. Below it, there's a section for 'Free training from Microsoft' and a list of recent resources. The right screenshot shows a 'Create SQL Database' wizard. It has a warning message about changing basic options. Under 'Database details', the 'Database name' is set to 'newsqldatabase' and the 'Server' is set to 'newserver101 (East US)'. There are options for using a SQL elastic pool and selecting a compute + storage tier (General Purpose). At the bottom, there are 'Review + create' and 'Next : Networking >' buttons.

Smart Vehicles

EUR/USD - 1,35379 - 00:00:00 14 giu (EEST)
EUR/USD (Bid), Ticks, # 300 / 300

■ TASK 2 - Create Azure SQL Server and Database

- Give **access** for current IP address for connectivity
- Proceed to deployment after **successful validation**

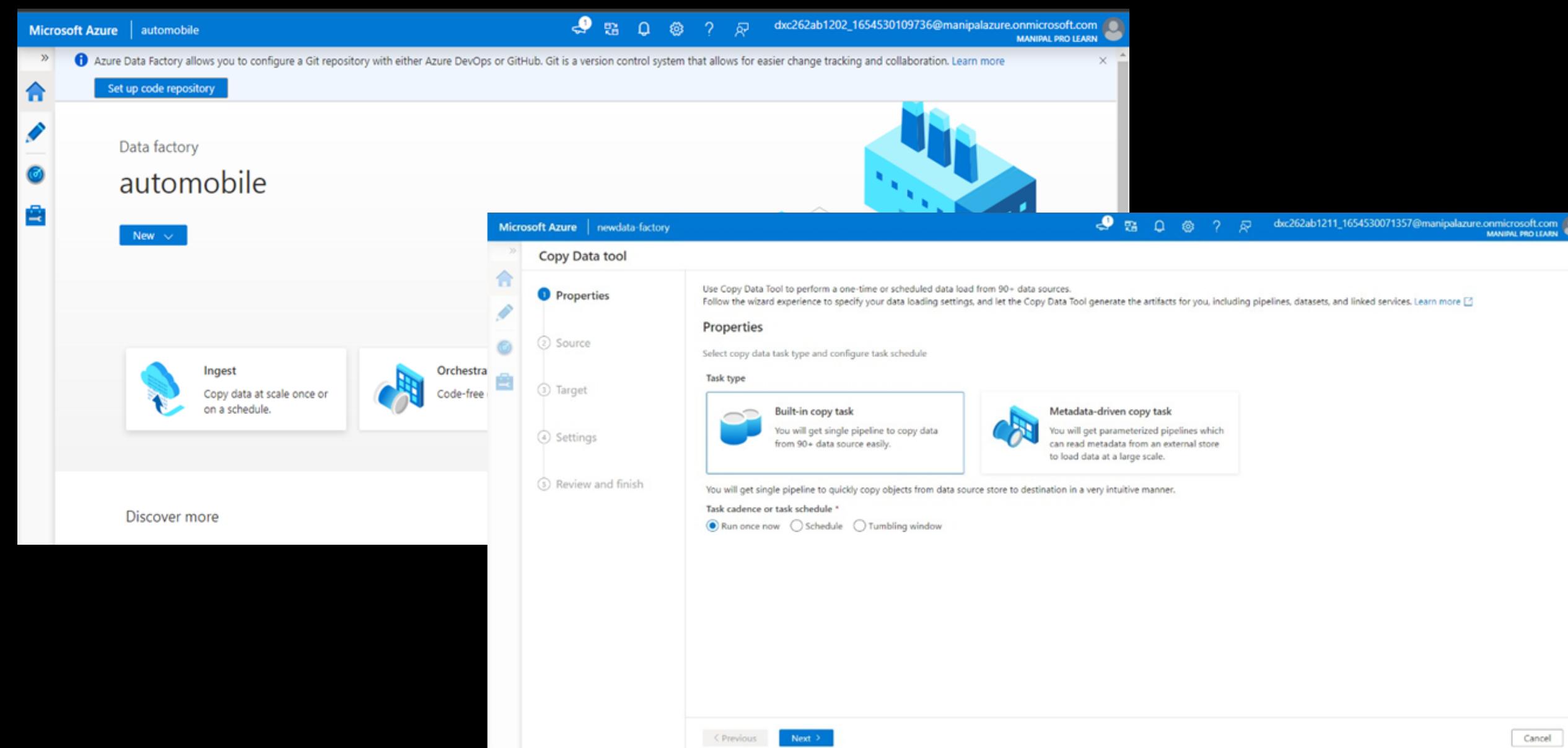
The image shows two screenshots of the Microsoft Azure portal. The left screenshot displays the 'Create SQL Database' wizard, specifically the 'Networking' tab. It includes sections for 'Firewall rules' (allowing Azure services and resources to access the server), 'Add current client IP address' (with a 'Yes' button selected), and 'Private endpoints'. The right screenshot shows the 'Overview' page for a deployed database named 'Microsoft.SQLDatabase.newDatabaseExistingServer_b925ac6280474c34'. The status bar indicates 'Your deployment is complete'. The deployment details show it was started on 6/10/2022 at 5:30:01 PM. The portal also features a sidebar with links for Cost Management, Microsoft Defender for Cloud, Free Microsoft tutorials, and Work with an expert.

Smart Vehicles



■ TASK 3- Create ADF Pipeline End to end pipeline with triggers enable

- Open Data Factory and Click on **Copy data tool**
- Select build-in copy tool and proceed



Smart vehicles



- TASK 3- Create ADF Pipeline End to end pipeline with triggers enable
 - Create a source and destination path for the pipeline

The image contains two side-by-side screenshots of the Microsoft Azure Copy Data tool interface.

Left Screenshot (Source Configuration):

- Properties:** Shows the current step is "Properties".
- Source:** Set to "Azure Blob Storage" with connection "AzureBlobStorage1".
- Dataset:** File or folder specified as "source/transactions.csv".
- Options:** Includes "Binary copy", "Recurvively" (checked), "Enable partition discovery", and "Max concurrent connections".
- Filter by last modified:** Set "Start time (UTC)" to "2014-06-13T00:00:00Z" and "End time (UTC)" to "2014-06-14T00:00:00Z".
- Buttons:** "Next >" and "Previous <".

Right Screenshot (Destination Configuration):

- Properties:** Shows the current step is "Properties".
- Source:** Set to "Azure Blob Storage file".
- Target:** Set to "Azure SQL Database" with connection "AzureSqlDatabase1".
- Dataset:** "transaction" is mapped to "Azure Blob Storage file (auto-create)".
- Options:** Includes "Skip column mapping for all tables".
- Buttons:** "Next >" and "Previous <".

Smart vehicles



- TASK 3- Create ADF Pipeline End to end pipeline with triggers enable
 - After pipeline is ready , review and trigger it.

The top screenshot shows the 'Copy Data tool' wizard in Microsoft Azure Data Factory. The pipeline is configured to copy data from Azure Blob Storage to Azure SQL Database. The task name is 'CopyPipeline_zbw'. The source is defined with connection name 'AzureBlobStorage1', dataset name 'SourceDataset_zbw', and file name 'transactions.csv'. The target is an Azure SQL Database. The bottom screenshot shows the 'Pipeline runs' page, which displays a single run for the 'CopyPipeline_zbw' pipeline, triggered manually on Jun 10, 2022, at 5:41:49 pm, with a duration of 00:00:17 and a status of 'Succeeded'.

Smart Vehicles



- TASK 3- Create ADF Pipeline End to end pipeline with triggers enable
 - End results

newsqldatabase (newserver101/newsqldatabase) | Query editor (preview)

SQL database

Search (Ctrl+Shift+F)

Login New Query Open query Feedback

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

transaction.Azure Blob Storage file

Showing limited object explorer here. For full capability please open SSDT.

Tables transaction.Azure Blob Storage file

- Date (nvarchar, null)
- Account (nvarchar, null)
- Transaction (nvarchar, null)
- Amount (nvarchar, null)
- Merchant (nvarchar, null)
- PaymentChannel (nvarchar, null)
- Type (nvarchar, null)
- Category (nvarchar, null)

Views Stored Procedures

Query 1 × transaction.Azure Blob Storage file

Create New Row Save Refresh Discard Delete Row

Search to filter items...

Date	Account	Transaction	Amount	Merchant	PaymentChannel	Type
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	DEPOSITED OR CAS...	-\$150.00			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	AMERICAN EXPRESS...	\$10,935.00			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	BANKCARD MITOT D...	\$109.64			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	CHECK	-\$616.63			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	CHECK	-\$702.15			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	CHECK	-\$590.92			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	CHECK	-\$905.78			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	T-Mobile	-\$442.87	T-Mobile		place
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	BUSINESS TO BUSIN...	-\$3,847.67		ACH	place
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	AMERICAN EXPRESS...	\$636.90			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	CHECK	-\$1,208.50			special
3/18/2022 12:00:00 ...	BUSINESS CHECKIN...	AMERICAN EXPRESS...	\$4,422.12			special

Ready

Conclusion

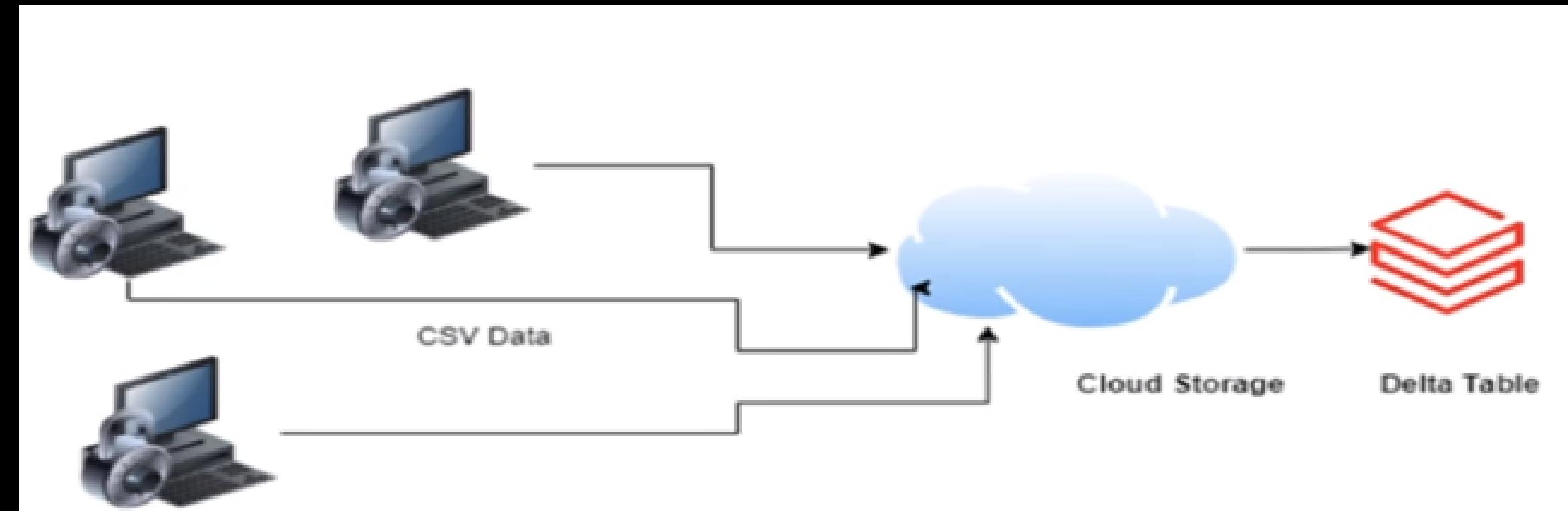
- In this project we successfully created a pipeline that validated and copied the blob data into the SQL database using Azure Data Factory

EUR/USD - 1,35379 - 00:00:00 14 giu (EEST)
EUR/USD (Bid), Ticks, # 300 / 300

AP Morgan Data Platform

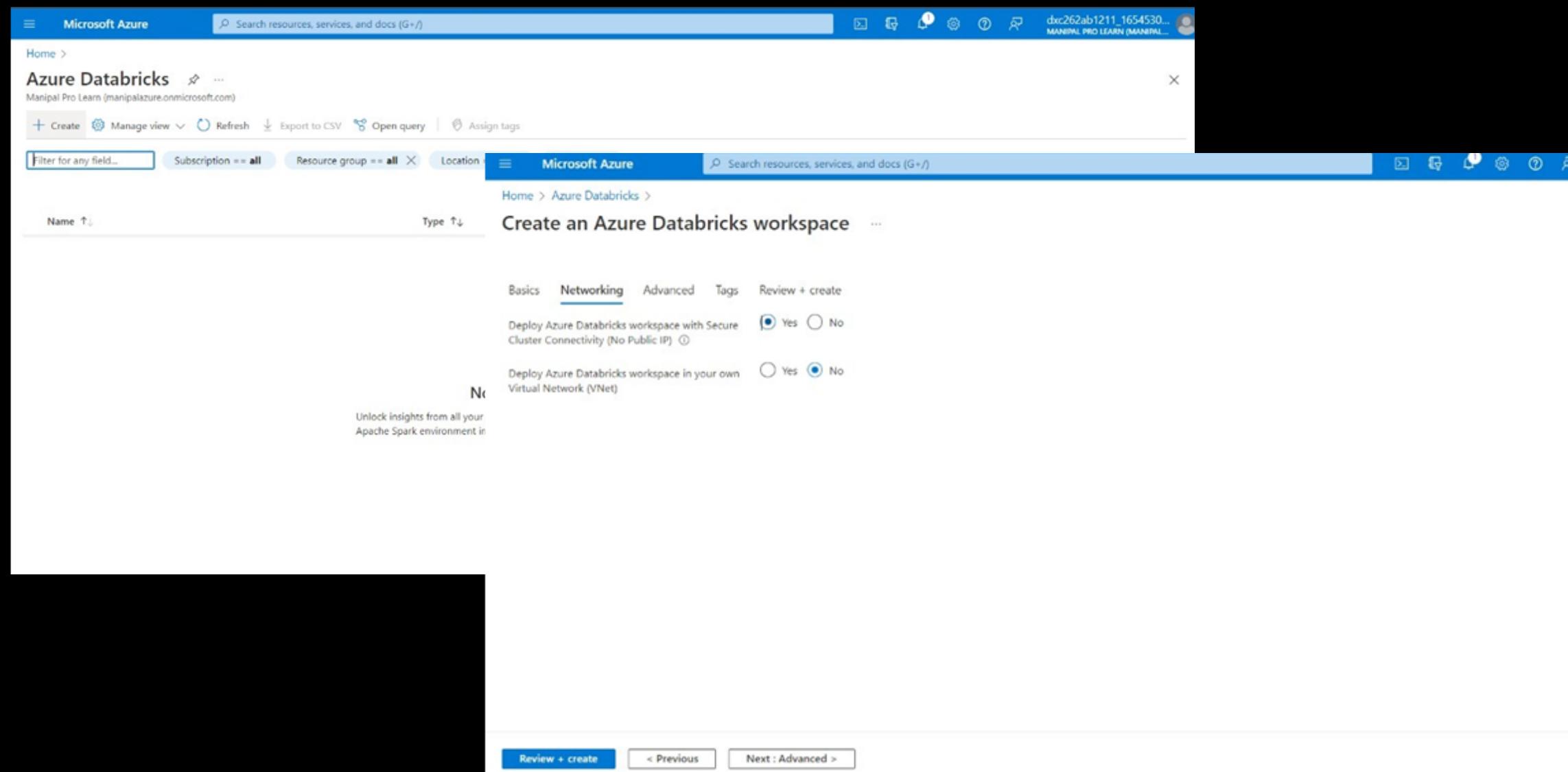
- Project Description and Phases

- The data coming from different sources need to be integrated after validating to get a final database and make data ready for further analysis.



AP Morgan Data Platform

- **TASK 1 - Create a DataBricks workspace and create a cluster into it**
 - Navigate and open databricks and click on +create
 - Give the details about the account required.
 - In networking give access for secure cluster connectivity and proceed for deployment.



EUR/USD - 1,35379 - 00:00:00 14 giu (EEST)
EURUSD (Bid), Ticks, # 300 / 300

AP Morgan Data Platform

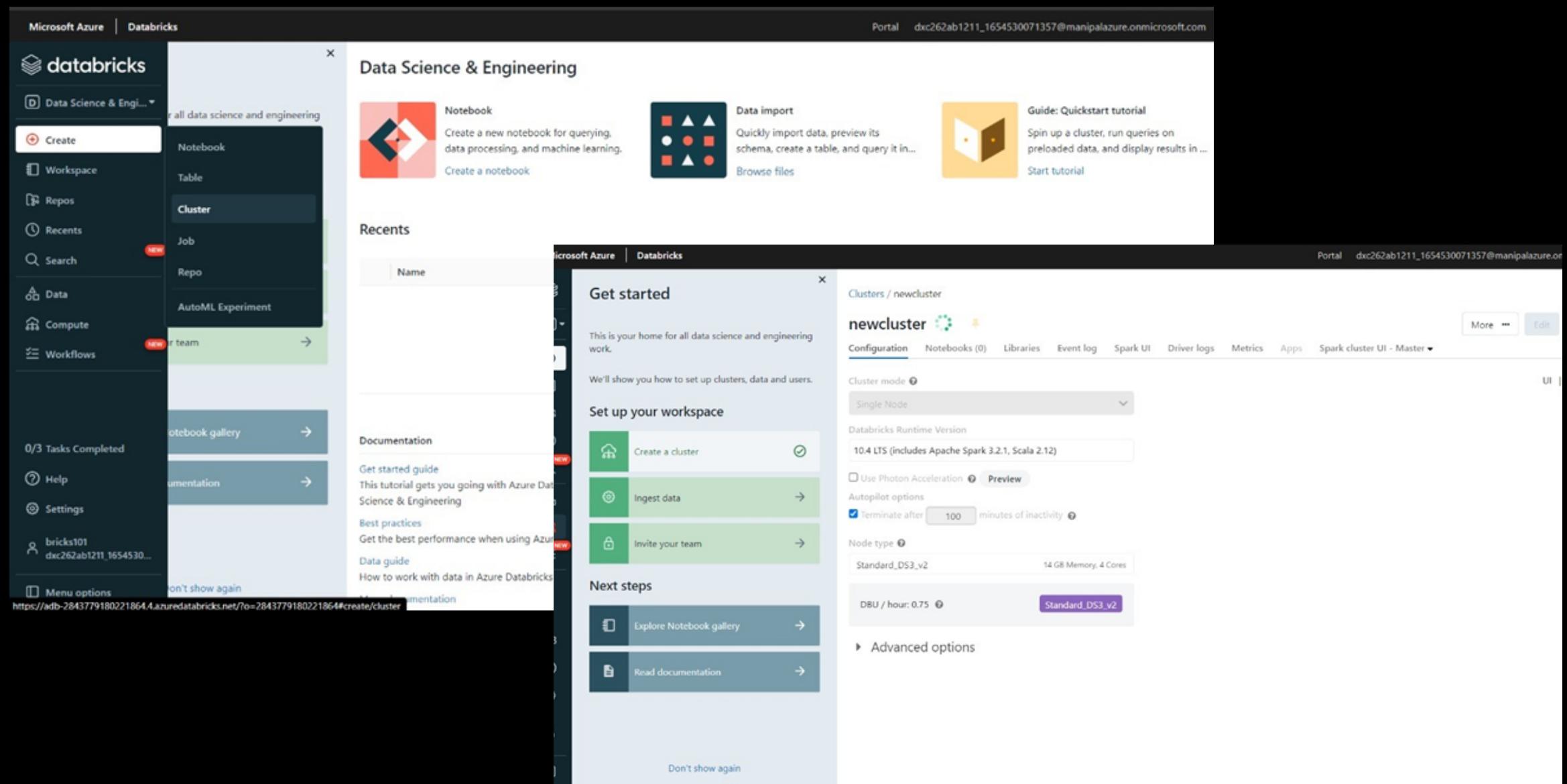
- TASK 1 - Create a DataBricks workspace and create a cluster into it
 - After the deployment is done proceed to launching the Data bricks workspace.

The screenshot displays two main Azure portal pages. The top page shows the 'resourcegroup_bricks101 | Overview' of a deployment, indicating a successful deployment with a green checkmark. The bottom page shows the 'bricks101' Azure Databricks Service overview, detailing its status as Active, resource group as 'resourcegroup', location as East US, and subscription as 'Azure-DXC262AB12Lab'. It also lists virtual network peering, encryption, properties, locks, automation tasks, and support options. A large red 'Launch Workspace' button is prominently displayed at the bottom right. The URL for the workspace is shown as <https://adb-28437791802218644.azuredatabricks.net>.

EUR/USD - 1,35379 - 00:00:00 14 giu (EEST)
EUR/USD (Bid), Ticks, # 300 / 300

AP Morgan Data Platform

- **TASK 1 - Create a DataBricks workspace and create a cluster into it**
 - Navigate to Cluster present in create dropdown
 - Give the details about cluster and proceed to create



AP Morgan Data Platform

■ TASK 2 - Add notebook in Databricks and Implement the Business Logic

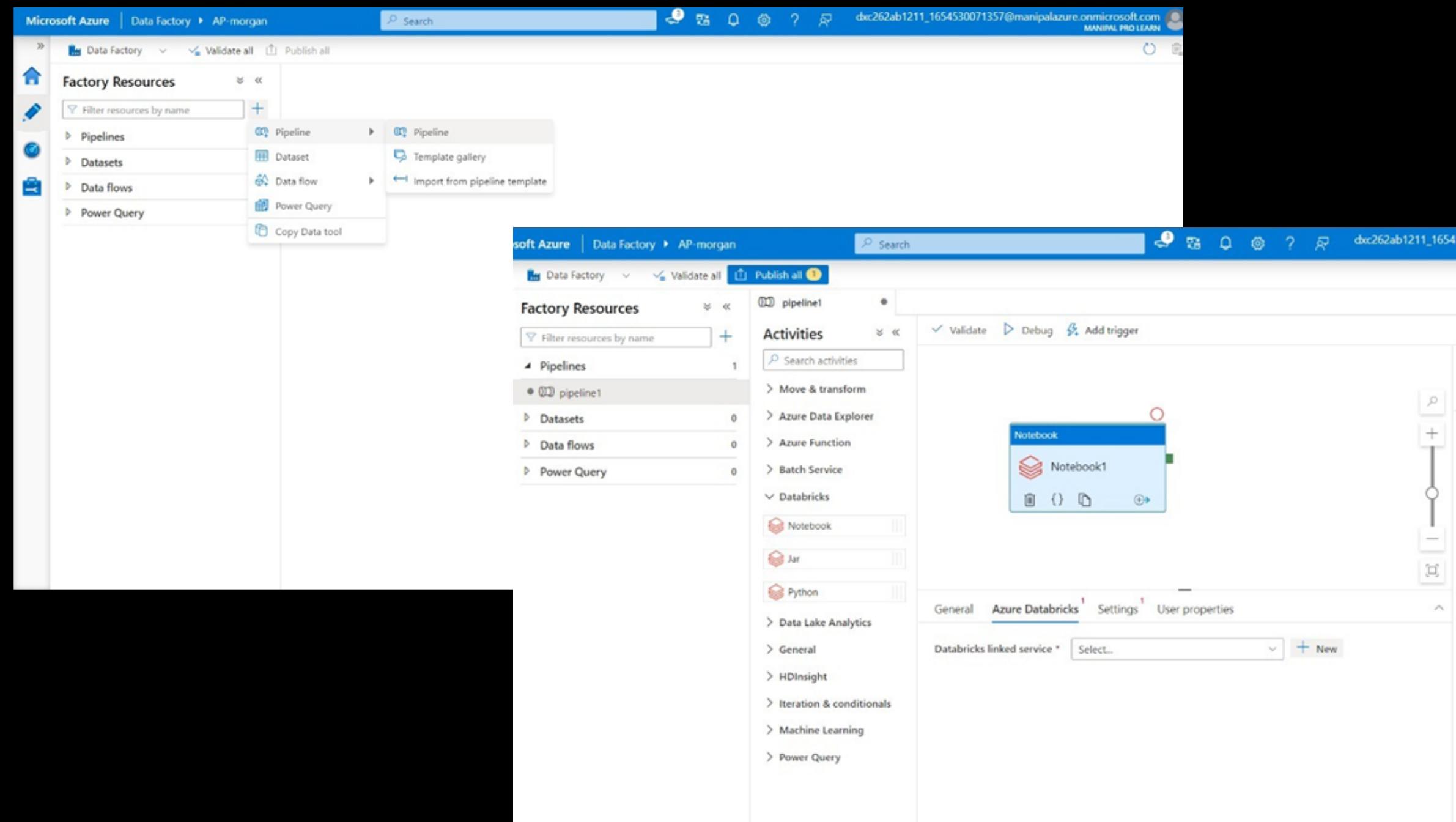
- Navigate towards creating a Notebook and open it
- Here we are suppose to perform some data manipulation and validation with respect to data.

The image consists of three screenshots of the Microsoft Azure Databricks interface:

- Screenshot 1:** The main Databricks dashboard. The sidebar on the left has sections for Data Science & Engineering, Compute, and Notebooks. The 'Compute' section is currently selected. The main area shows a 'newcluster' cluster with 0 notebooks attached.
- Screenshot 2:** The 'Clusters / newcluster' page. It shows a table with columns for Name, Status, Last Command Run, and Location. There are no rows in the table.
- Screenshot 3:** The 'Notebook1' page. It shows a 'Get started' section with steps like 'Create a cluster', 'Ingest data', and 'Invite your team'. Below that is a 'Next steps' section with 'Explore Notebook gallery' and 'Read documentation'. The main area contains a code cell with the Python command: `print("Do some computation on data")`. The output of the cell shows the message "Do some computation on data".

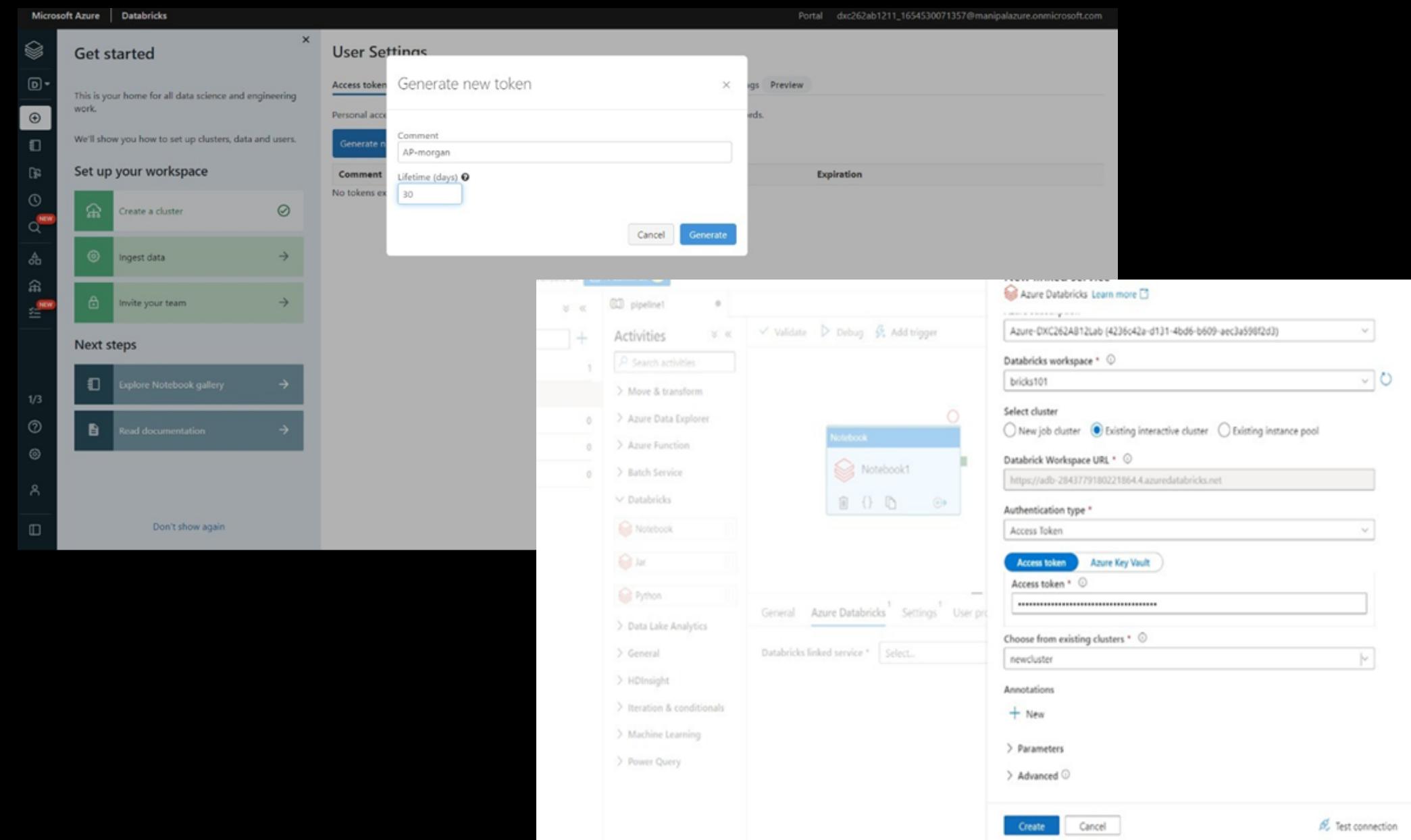
AP Morgan Data Platform

- TASK 3 - Add Azure data bricks linked service in Data Factory
 - Creating a pipeline by dropping and linking notebook that is created in the data bricks



AP Morgan Data Platform

- TASK 3 - Add Azure data bricks linked service in Data Factory
 - Give the details required to connect Notebook including Access Token.



EUR/USD - 1,35379 - 00:00:00 14 giu (EEST)
EURUSD (Bid), Ticks, # 300 / 300

AP Morgan Data Platform

■ TASK 3 - Final Result

- Navigate towards monitor and here we have our pipeline created and triggered linking data bricks with azure data factory

The screenshot shows the 'Pipeline runs' page in the Azure Data Factory interface. The left sidebar has a 'Runs' section selected, which is highlighted in blue. The main area displays a table of pipeline runs. The table has columns for Pipeline name, Run start, Run end, Duration, Triggered by, Status, Error, and Run. There is one item listed:

Pipeline name	Run start	Run end	Duration	Triggered by	Status	Error	Run
pipeline1	Jun 11, 2022, 10:52:42 pm	Jun 11, 2022, 10:53:04 pm	00:00:21	Manual trigger	Succeeded		Original

At the top of the page, there are several filter and search options: 'Triggered' (selected), 'Debug', 'Rerun', 'Cancel', 'Refresh', 'Edit columns', 'List' (selected), 'Gantt', 'Filter by run ID or name' (with 'Chennai, Kolkata, Mu...' and 'Last 24 hours' dropdowns), 'Pipeline name : All', 'Status : All', 'Runs : Latest runs', 'Copy filters', and 'Export to CSV'.

Conclusion

- In this project we were able to successfully link and trigger azure Data Bricks notebook using Data Factory.

Learning Outcome

- Hands on experience on Azure Data bricks , Data Factory , Azure SQL database , Azure Blob Storage.
- By linking connectivity within azure services enabled to learn and get the most out of Azure databricks and Azure Data Factory.
- Diving into Data factory trigger functionality and Databricks code utility gave an opportunity to learn more about automating workflow.

**THANK
YOU**