

ASSIGNMENT – 5 (Azure Databricks)

TRAINEE NAME: Swathi Baskaran

Creating a container in Azure Storage Account

The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes the Microsoft Azure logo, a search bar, and a user profile. The main content area displays the 'Containers' section for the storage account 'storageaccountsynapsep2'. A sidebar on the left lists various Azure services, with 'Containers' selected. The main area shows a table of containers with columns for Name, Last modified, Anonymous access level, and Lease state. The table lists three containers: 'slogs', 'datalake', and 'employeeedata'.

Name	Last modified	Anonymous access level	Lease state
slogs	27/08/2025, 15:28:37	Private	Available
datalake	27/08/2025, 15:29:27	Private	Available
employeeedata	29/08/2025, 12:18:32	Private	Available

Uploading the csv file in the container

The screenshot shows the Microsoft Azure portal interface for the 'employeeedata' container. The top navigation bar includes the Microsoft Azure logo, a search bar, and a user profile. The main content area displays the 'employeeedata' container details. A sidebar on the left lists various Azure services, with 'Containers' selected. The main area shows a table of blobs with columns for Name, Last modified, Access tier, Blob type, Size, and Lease state. The table lists one blob: 'employees.csv'.

Name	Last modified	Access tier	Blob type	Size	Lease state
employees.csv	29/08/2025, 12:21:26	Hot (Inferred)	Block blob	3.69 KiB	Available

Creating a Data Factory in Azure Portal

The screenshot shows the Azure Portal interface for a Data Factory (V2) named 'EmployeesData'. The top navigation bar includes the Microsoft Azure logo, a search bar, and user information. The left sidebar contains a navigation menu with options like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Resource visualizer, Settings, Getting started, Monitoring, Automation, and Help. The main content area displays the 'Overview' tab for the 'EmployeesData' Data factory (V2). It includes a search bar, a 'Delete' button, and a 'JSON View' link. Below this, there's a 'Summary' section with details: Resource group (rg-azuser4030_mml.local-dp88), Status (Succeeded), Location (East US), Subscription (MML Learners), and Subscription ID (2a3c6418-97b9-4d96-a24b-2c2d7633d375). The 'Type' is listed as 'Data factory (V2)' with a 'Getting started' link. A large 'Launch studio' button is prominently displayed. Below the button are four tiles: 'Quick Starts', 'Tutorials', 'Template Gallery', and 'Training Modules'. At the bottom, a small note says 'Add or remove favorites by pressing Ctrl+Shift+F'.

Click on Ingest in the Home Page

The screenshot shows the 'EmployeesData' Data factory home page in the Azure Portal. The top navigation bar includes the Microsoft Azure logo, the breadcrumb 'Data Factory > EmployeesData', a search bar, and user information. The left sidebar contains a navigation menu with options like Home, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Resource visualizer, Settings, Getting started, Monitoring, Automation, and Help. The main content area displays the 'Overview' tab for the 'EmployeesData' Data factory (V2). It includes a search bar, a 'Delete' button, and a 'JSON View' link. Below this, there's a 'Summary' section with details: Resource group (rg-azuser4030_mml.local-dp88), Status (Succeeded), Location (East US), Subscription (MML Learners), and Subscription ID (2a3c6418-97b9-4d96-a24b-2c2d7633d375). The 'Type' is listed as 'Data factory (V2)' with a 'Getting started' link. A large 'Launch studio' button is prominently displayed. Below the button are four tiles: 'Quick Starts', 'Tutorials', 'Template Gallery', and 'Training Modules'. At the bottom, a small note says 'Add or remove favorites by pressing Ctrl+Shift+F'.

Click on Built-in copy task

Microsoft Azure | Data Factory | EmployeesData | Search

Copy Data tool

1 Properties

2 Source

3 Destination

4 Settings

5 Review and finish

Use Copy Data Tool to perform a one-time or scheduled data load from over 100 data sources. Follow the wizard experience to specify your data loading settings, and let the Copy Data Tool generate the artifacts for you, including pipelines, datasets, and linked services. [Learn more](#)

Properties

Select copy data task type and configure task schedule

Task type

Built-in copy task

You will get a single pipeline which is capable of smoothly copying data from over 100 different data sources.

Metadata-driven copy task

You will get parameterized pipelines which can read metadata from an external store to load data at a large scale.

You will get single pipeline to quickly copy objects from data source store to destination in a very intuitive manner.

Task cadence or task schedule *

☒ Run once now ☐ Schedule ☐ Tumbling window

< Previous Next > Cancel

Moving to the source data source page

Microsoft Azure | Data Factory | EmployeesData | Search

Copy Data tool

1 Properties

2 Source

3 Dataset

4 Configuration

5 Destination

6 Settings

7 Review and finish

Source data store

Specify the source data store for the copy task. You can use an existing data store connection or specify a new data store.

Source type All

Connection * Select... + New connection

< Previous Next > Cancel

Creating a new connection

New connection

Azure Blob Storage

Learn more

Name *

EmployeeBS

Description

Connect via integration runtime * ⓘ

✓

AutoResolveIntegrationRuntime

Authentication type

Account key

Connection string

Azure Key Vault

Account selection method ⓘ

☒ From Azure subscription

☐ Enter manually

Azure subscription ⓘ

Select all

Storage account name *

Additional connection properties

+ New

Create

Back

Test connection

Cancel

Testing the connection after filling all the details

New connection

Azure Blob Storage

Learn more

Name *

EmployeeBS

Description

Connect via integration runtime * ⓘ

✓

AutoResolveIntegrationRuntime

Authentication type

Account key

Connection string

Azure Key Vault

Account selection method ⓘ

☒ From Azure subscription

☐ Enter manually

Azure subscription ⓘ

MML Learners (2a3c6418-97b9-4d96-a24b-2c2d7633d375)

Storage account name *

hexadatastoragegen3

Additional connection properties

Connection successful

Create

Back

Test connection

Cancel

Filling the other details on Source Data Source Page

The screenshot shows the 'Source data store' configuration page in the Microsoft Azure Data Factory 'Copy Data tool'. The left sidebar contains a navigation menu with steps: Properties, Source, Dataset, Configuration, Destination, Settings, and Review and finish. The 'Source' step is currently selected. The main content area is titled 'Source data store' and includes the following fields and options:

- Source type:** A dropdown menu set to 'All'.
- Connection:** A dropdown menu set to 'EmployeeBS', with links for 'Edit' and 'New connection'.
- File or folder:** A text input field containing 'employeeData/input/employees.csv', with a 'Browse' button.
- Options:**
 - ☒ **Binary copy**
 - Compression type:** A dropdown menu set to 'None'.
 - ☒ **Recursively**
 - ☐ **Delete files after completion**
- Max concurrent connections:** A text input field.
- Filter by last modified:** Two text input fields for 'Start time (UTC)' and 'End time (UTC)'.

At the bottom of the page, there are '< Previous' and 'Next >' buttons, and a 'Cancel' button on the right side.

Filling the details on the Destination Data Source Page

The screenshot shows the 'Destination data store' configuration page in the Microsoft Azure Data Factory 'Copy Data tool'. The left sidebar contains a navigation menu with steps: Properties, Source, Destination, Dataset, Configuration, Settings, and Review and finish. The 'Destination' step is currently selected. The main content area is titled 'Destination data store' and includes the following fields and options:

- Destination type:** A dropdown menu set to 'All'.
- Connection:** A dropdown menu set to 'EmployeeBS', with links for 'Edit' and 'New connection'.
- Folder path:** A text input field containing 'employeeData/output', with a 'Browse' button.
- File name:** A text input field.
- Compression type:** A dropdown menu set to 'None'.
- Copy behavior:** A dropdown menu set to 'Select...'.
- Max concurrent connections:** A text input field.
- Block size (MB):** A text input field.

At the bottom of the page, there are '< Previous' and 'Next >' buttons, and a 'Cancel' button on the right side.

Validating the Process

The screenshot shows the 'Validating' page in the Microsoft Azure Data Factory 'Copy Data tool'. The left sidebar contains a navigation menu with steps: Properties, Source, Destination, Settings, Review and finish, Review, and Deployment. The 'Review and finish' step is currently selected. The main content area is titled 'Validating' and includes a diagram showing data flow from 'Azure Blob Storage' to 'Azure Blob Storage'. Below the diagram, there is a table showing the deployment steps and their status.

Deployment step	Status
Validating copy runtime environment	In progress

At the bottom of the table, there is a link to 'Skip validation'.

Completion of Deployment

The screenshot shows the 'Copy Data tool' interface in Microsoft Azure Data Factory. The left sidebar contains a navigation menu with options: Properties, Source, Destination, Settings, Review and finish, Review, and Deployment. The main area displays a diagram of data flow from 'Azure Blob Storage' to 'Azure Blob Storage'. Below the diagram, the text 'Deployment complete' is shown. A table lists the deployment steps and their status:

Deployment step	Status
Validating copy runtime environment	Succeeded
> Creating datasets	Succeeded
> Creating pipelines	Succeeded
> Running pipelines	Succeeded

Below the table, a message states: 'Datasets and pipelines have been created. You can now monitor and edit the copy pipelines or click finish to close Copy Data Tool.' At the bottom, there are buttons for 'Finish', 'Edit pipeline', and 'Monitor'.

Monitoring the execution of Pipeline

The screenshot shows the 'Pipeline runs' interface in Microsoft Azure Data Factory. The left sidebar contains a navigation menu with options: Dashboards, Runs, Pipeline runs, Trigger runs, Change Data Capture (previ..., Runtime & sessions, Integration runtimes, Data flow debug, Notifications, Alerts & metrics. The main area displays the 'Pipeline runs' section. A filter bar shows 'Filter by run ID or name' with a search box containing 'Chennai, Kolkata, Ma...', a 'Last 24 hours' filter, and a 'Pipeline name: PL_EmployeesCopy' filter. The status is set to 'All'. Below the filter bar, there are buttons for 'Runs: Latest runs', 'Triggered by: All', and 'Add filter'. A table shows the pipeline runs:

Pipeline name	Run start	Run end	Duration	Triggered by	Status	Run	Parameters
PL_EmployeesCopy	8/29/2025, 12:34:10 PM	8/29/2025, 12:34:35 PM	26s	Manual trigger	Succeeded	Original	

The table indicates that the pipeline run 'PL_EmployeesCopy' was triggered manually and succeeded. The status is 'Succeeded'.

Viewing the pipeline run

The screenshot shows the 'Activity runs' interface in Microsoft Azure Data Factory for the pipeline run 'PL_EmployeesCopy'. The left sidebar contains a navigation menu with options: Dashboards, Runs, Pipeline runs, Trigger runs, Change Data Capture (previ..., Runtime & sessions, Integration runtimes, Data flow debug, Notifications, Alerts & metrics. The main area displays the 'Activity runs' section. A filter bar shows 'All pipeline runs > PL_EmployeesCopy - Activity runs'. Below the filter bar, there are buttons for 'Rerun', 'Cancel', 'Refresh', 'Update pipeline', and 'List'. A table shows the activity runs:

Activity name	Activity st...	Activit...	Run start	Duration	Integration runtime	User prop...	Destination
Copy_to1	Succeeded	Copy data	8/29/2025, 12:34:19 PM	16s	AutoResolveIntegrationRuntime (Central India)		employeeData/outp

The table indicates that the activity run 'Copy_to1' was triggered manually and succeeded. The status is 'Succeeded'.