

Coding Challenge-4

Date: 10/12/2024

Name: Harish E.R

Q) Copy activity from datalake to datalake in azure datafactory

Copying from datalakegen1hexaware to datalakegen2hexaware:

- datalakegen1hexaware:

The screenshot displays the Azure portal interface for the 'datalakegen1hexaware' storage account. The left sidebar shows the navigation menu with 'Data storage' expanded. The main content area is divided into 'Essentials' and 'Properties' tabs. The 'Essentials' tab shows key information: Resource group (rg-azuser2365_mmllocal-Ta7JA), Location (westus), Subscription (MML Learners), Subscription ID (2a3c6418-97b9-4d96-a24b-2c2d7633d375), Disk state (Available), and Tags (Add tags). The 'Properties' tab is active, showing 'Data Lake Storage' settings (Hierarchical namespace: Enabled, Default access tier: Hot, Blob anonymous access: Disabled, Blob soft delete: Enabled (7 days), Container soft delete: Enabled (7 days), Versioning: Disabled, Change feed: Disabled) and 'Security' settings (Require secure transfer for REST API operations: Enabled, Storage account key access: Enabled, Minimum TLS version: Version 1.2, Infrastructure encryption: Disabled). The 'Networking' section shows 'Allow access from: All networks'.

- datalakegen2hexaware:

The screenshot displays the Azure portal interface for the 'datalakegen2hexaware' storage account. The left sidebar shows the navigation menu with 'Data storage' expanded. The main content area is divided into 'Essentials' and 'Properties' tabs. The 'Essentials' tab shows key information: Resource group (rg-azuser2365_mmllocal-Ta7JA), Location (westus3), Subscription (MML Learners), Subscription ID (2a3c6418-97b9-4d96-a24b-2c2d7633d375), Disk state (Available), and Tags (Add tags). The 'Properties' tab is active, showing 'Data Lake Storage' settings (Hierarchical namespace: Enabled, Default access tier: Hot, Blob anonymous access: Disabled, Blob soft delete: Enabled (7 days), Container soft delete: Enabled (7 days), Versioning: Disabled, Change feed: Disabled) and 'Security' settings (Require secure transfer for REST API operations: Enabled, Storage account key access: Enabled, Minimum TLS version: Version 1.2, Infrastructure encryption: Disabled). The 'Networking' section shows 'Allow access from: All networks'.

Step 1) Make Datafactory account:

The screenshot shows the Microsoft Azure portal interface. At the top, the header includes the Microsoft Azure logo, a search bar, and the user profile 'azuser2365_mml.local@techademy.com'. The main content area displays the 'Overview' page for a new Data Factory (V2) account named 'azuredatafactoryhexaware'. The left sidebar contains navigation links: Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Settings, Getting started, Monitoring, Automation, and Help. The 'Essentials' section on the right provides key details: Resource group (rg-azuser2365_mml.local-Ta7IA), Status (Succeeded), Location (East US), Subscription (MML Learners), and Subscription ID (2a3c6418-97b9-4d96-a24b-2c2d7633d375). Below this, the 'Azure Data Factory Studio' logo is prominently displayed with a 'Launch studio' button. At the bottom, there are four tiles for 'Quick Starts', 'Tutorials', 'Template Gallery', and 'Training Modules'.

The screenshot shows the Azure Data Factory Studio interface. The top header includes the Microsoft Azure logo, the breadcrumb 'Data Factory > azuredatafactoryhexaware', a search bar, and the user profile 'azuser2365_mml.local@techademy.com'. A notification banner at the top asks if the user wants to see Data Factory inside of Microsoft Fabric. The main content area features a large 3D illustration of a factory. Below this, there are four main action tiles: '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). The left sidebar contains navigation icons for Home, Data flows, Pipelines, Monitoring, and Settings. The 'Recent resources' section at the bottom shows a table with columns for Name, Type, and Last opened by you.

Name	Type	Last opened by you
Configuration tool	Pipeline	4 hours ago

Step 2: Check the properties in the Copy Data Tool:

The screenshot shows the 'Copy Data tool' wizard in the 'Properties' step. The left sidebar contains a navigation pane with steps: 1 Properties, 2 Source, 3 Destination, 4 Settings, and 5 Review and finish. The main area is titled 'Properties' and includes a description: 'Use Copy Data Tool to perform a one-time or scheduled data load from 90+ 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](#)'. Below this is the 'Task type' section with two options: 'Built-in copy task' (selected) and 'Metadata-driven copy task'. The 'Built-in copy task' description states: 'You will get a single pipeline which is capable of smoothly copying data from over 100 different data sources.' The 'Metadata-driven copy task' description states: 'You will get parameterized pipelines which can read metadata from an external store to load data at a large scale.' Below the task types is the 'Task cadence or task schedule' section with three radio buttons: 'Run once now' (selected), 'Schedule', and 'Tumbling window'. At the bottom are buttons for '< Previous', 'Next >', and 'Cancel'.

Step 3: Check and click the Source for copying:

The screenshot shows the 'Copy Data tool' wizard in the 'Source' step. The left sidebar contains a navigation pane with steps: 1 Properties, 2 Source, 3 Destination, 4 Settings, and 5 Review and finish. The main area is titled 'Source data store' and includes a description: 'Specify the source data store for the copy task. You can use an existing data store connection or specify a new data store.' Below this is the 'Source type' dropdown menu set to 'Azure Data Lake Storage Gen2'. The 'Connection' dropdown menu is set to 'Select...'. To the right of the 'Connection' dropdown is a '+ New connection' button. Below the 'Source type' and 'Connection' dropdowns are buttons for '< Previous', 'Next >', and 'Cancel'. On the right side of the screen is a 'New connection' panel. It includes a 'Name' field set to 'AzureDataLakeStorage2', a 'Description' field, a 'Connect via integration runtime' dropdown set to 'AutoResolveIntegrationRuntime', an 'Authentication type' dropdown set to 'Account key', an 'Account selection method' section with 'From Azure subscription' selected, an 'Azure subscription' dropdown set to 'MML Learners (2a3c6418-97b9-4d96-a24b-2c2d7633d375)', a 'Storage account name' dropdown set to 'datalakegen1hexaware', and a 'Test connection' section with 'To linked service' selected. At the bottom of the 'New connection' panel are buttons for 'Create', 'Cancel', and 'Test connection'.

The screenshot shows the 'Copy Data tool' wizard in the 'Dataset' step. The left sidebar contains a navigation pane with steps: 1 Properties, 2 Source, 3 Dataset, 4 Settings, and 5 Review and finish. The main area is titled 'Source data store' and includes a description: 'Specify the source data store for the copy task. You can use an existing data store connection or specify a new data store.' Below this is the 'Source type' dropdown menu set to 'Azure Data Lake Storage Gen2'. The 'Connection' dropdown menu is set to 'AzureDataLakeStorage2'. To the right of the 'Connection' dropdown is an 'Edit' button and a '+ New connection' button. Below the 'Source type' and 'Connection' dropdowns is the 'File or folder' section. It includes a description: 'If the identity you use to access the data store only has permission to subdirectory instead of the entire account, specify the path to browse. Append a slash (/) at the end if the path refers to a folder.' Below this is a text field set to 'codingchallenge4/mudah-apartment-kl-selangor.csv' and a 'Browse' button. Below the text field is the 'Options' section with a 'Binary copy' checkbox checked, a 'Compression type' dropdown set to 'None', a 'Recursively' checkbox checked, a 'Delete files after completion' checkbox unchecked, and a 'Max concurrent connections' field. Below the 'Max concurrent connections' field is the 'Filter by last modified' section. At the bottom are buttons for '< Previous', 'Next >', and 'Cancel'.

Step 4: Setup Destination path for copying:

This screenshot shows the 'New connection' dialog in the Microsoft Azure Data Factory interface. The 'Destination data store' section is active, showing 'Azure Data Lake Storage Gen2' as the destination type and 'AzureDataLakeStorage3' as the connection. The 'Connect via integration runtime' is set to 'AutoResolveIntegrationRuntime'. The 'Authentication type' is 'Account key', and the 'Account selection method' is 'From Azure subscription'. The 'Azure subscription' is 'MML Learners (2a3c6418-97b9-4d96-a24b-2c2d7633d375)'. The 'Storage account name' is 'dstalokegen2hexaware'. The 'Test connection' option is 'To linked service'. The 'Create' button is visible at the bottom right.

This screenshot shows the 'Destination data store' configuration page in the Microsoft Azure Data Factory interface. The 'Destination type' is 'Azure Data Lake Storage Gen2' and the 'Connection' is 'AzureDataLakeStorage3'. The 'Folder path' is 'hexawaredata/'. The 'File name' is 'codingchallenge4'. The 'Compression type' is 'None'. The 'Copy behavior' is 'Select...'. The 'Max concurrent connections' is '1'. The 'Block size (MB)' is '1'. The 'Next' button is visible at the bottom right.

Step 5: Setup settings:

This screenshot shows the 'Settings' page in the Microsoft Azure Data Factory interface. The 'Task name' is 'Copyactivity_CodingChallenge4'. The 'Task description' is empty. The 'Data consistency verification' checkbox is unchecked. The 'Enable logging' checkbox is unchecked. The 'Enable staging' checkbox is unchecked. The 'Advanced' section is expanded. The 'Next' button is visible at the bottom right.

Step 6: Review and click Deploy:

Microsoft Azure | Data Factory | azuredatafactoryhexaware

Search factory and documentation

Copy Data tool

Properties

Source

Destination

Settings

Review and finish

Review

Deployment

Summary

You are running pipeline to copy data from Azure Data Lake Storage Gen2 to Azure Data Lake Storage Gen2.

Azure Data Lake Storage Gen2 → Azure Data Lake Storage Gen2

Properties

Task name: Copyactivity_CodingChallenge4

Task description

Source

Connection name: AzureDataLakeStorage2

Dataset name: SourceDataset_sgw

File name: mudah-apartment-kd-selangor.csv

Destination

Connection name: AzureDataLakeStorage3

< Previous Next >

Cancel

Step 7: Run the pipeline:

Microsoft Azure | Data Factory | azuredatafactoryhexaware

Search factory and documentation

Would you like to see Data Factory inside of Microsoft Fabric, Microsoft's newest cloud-first data analytics SaaS platform? Click here to get started with Fabric Data Factory!

All pipeline runs > Copyactivity_CodingChallenge4 - Activity runs

Run Cancel Refresh Update pipeline List Gantt

Copy data

Copy_sgw

Activity runs

Pipeline run ID: 528dc63a-851a-4157-9af2-400570f59778

All status

Showing 1 - 1 items

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User properties	Destination	Source
Copy_sgw	Succeeded	Copy data	12/10/2024, 4:41:42 PM	17s	AutoResolveIntegrator		hexawaredata/coding	coding

- datalakegen2hexaware before copying:

Microsoft Azure

Search resources, services, and docs (G+)

Copilot

azuser2365.mml.local@TECHADEMY LEARNING SOLUTL

Home > datalakegen2hexaware | Containers >

hexawaredata

Container

Search

UploadAdd DirectoryRefreshRenameDeleteChange tierAcquire leaseBreak leaseGive feedback

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

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

Location: hexawaredata

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
No results						

- datalakegen2hexaware after copying:

Microsoft Azure

Search resources, services, and docs (G+)

Copilot

azuser2365.mml.local@TECHADEMY LEARNING SOLUTL

Home > datalakegen2hexaware | Containers >

hexawaredata

Container

Search

UploadAdd DirectoryRefreshRenameDeleteChange tierAcquire leaseBreak leaseGive feedback

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

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

Location: hexawaredata

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
<input type="checkbox"/> mudah-apartment-kl-selangor.csv	12/10/2024, 4:45:08 PM	Hot (Inferred)		Block blob	4.78 MiB	Available