

Azure Data Pipeline Lab

Outline

- Lab1: Setup & query data from Azure SQL Database
- Lab2: ETL pipeline in Azure Data Factory (ADF) with Azure Data Lake (ADL)
 - Lab2-1: Create Dataflow “IMDBDataflow”
 - Lab2-2: Create Pipeline “IMDBPipeline”
 - Add “Web node” to send email
 - Need to setup LogicApps
 - Lab2-3: Set triggers (scheduler)

Lab1: Setup & query data from Azure SQL Database

Steps

- Create “Storage account” (lake)
- Upload data source “AllElectronics.bacpac” to Container “data”
- Create “Azure SQL server” (serverless)
- Create “Azure SQL database”
- Import database (file.bacpac) from the Storage account (lake)
- Now ready to query data using “Query Editor”

Create “Storage account”

- Check to enable “Data Lake Storage Gen2” in Advanced tab

Home > Storage accounts >

Create a storage account

Basics Advanced Networking Data protection Encryption Tags Review

Instance details

If you need to create a legacy storage account type, please click [here](#).

Storage account name ⓘ *

Region ⓘ *

[Deploy to an edge zone](#)

Performance ⓘ * ☒ **Standard:** Recommended for most scenarios (general-purpose v2 account)
☐ **Premium:** Recommended for scenarios that require low latency.

Redundancy ⓘ *

Create a storage account

Basics **Advanced** Networking Data protection Encryption Tags Review

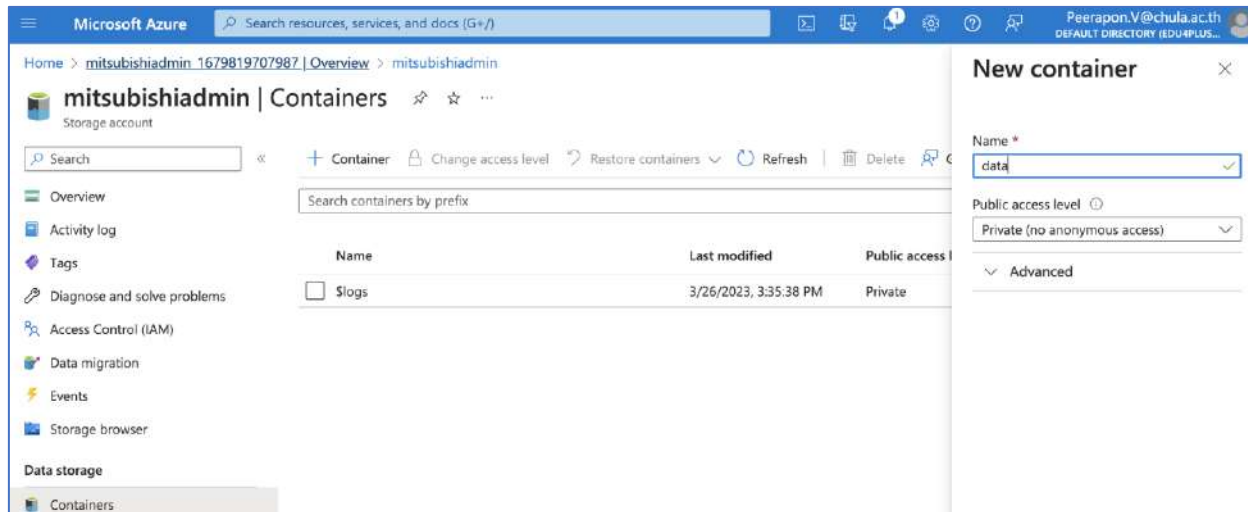
Data Lake Storage Gen2

The Data Lake Storage Gen2 hierarchical namespace accelerates big data analytics workloads and enables file-level access control lists (ACLs). [Learn more](#)

Enable hierarchical namespace ☒

Blob storage

Upload data source “AllElectronics.bacpac” to Container “data”



Create Azure SQL Server

User: student
Password: Mitsubishi2023

Microsoft Azure

Search resources, services, and docs (G+)

[Home](#) > [SQL servers](#) >

Create SQL Database Server

Microsoft

Server name *

az-mitsubishi-sql-admin ✓

.database.windows.net

Location *

(Asia Pacific) Southeast Asia ▼

Authentication

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 Azure AD authentication [Learn more](#) using an existing Azure AD user, group, or application as Azure AD admin [Learn more](#), or select both SQL and Azure AD authentication.

Authentication method

☐ Use only Azure Active Directory (Azure AD) authentication

☐ Use both SQL and Azure AD authentication

☒ Use SQL authentication

Server admin login *

student ✓

Password *

..... ✓

Confirm password *

..... ✓

Review + create

Next : Networking >

[Home](#) > [SQL servers](#) >

Create SQL Database Server

Microsoft

Basics

Networking

Additional settings

Tags

Review + create

Configure networking access for your server.

Firewall rules


Allow Azure services and resources to access this server ⓘ

Yes

No

Set SQL Server to be “serverless”

Configure

 Feedback

Service and compute tier

Select from the available tiers based on the needs of your workload. The vCore model provides a wide range of configuration controls and offers Hyperscale and Serverless to automatically scale your database based on your workload needs. Alternately, the DTU model provides set price/performance packages to choose from for easy configuration. [Learn more](#)

Service tier General Purpose (Most budget friendly, Serverless compute) ▾
[Compare service tiers](#)

Compute tier

☐ **Provisioned** - Compute resources are pre-allocated. Billed per hour based on vCores configured.


☒ **Serverless** - Compute resources are auto-scaled. Billed per second based on vCores used.

Compute Hardware

Select the hardware configuration based on your workload requirements. Availability of compute optimized, memory optimized, and confidential computing hardware depends on the region, service tier, and compute tier.

Hardware Configuration **Standard-series (Gen5)**
up to 40 vCores, up to 120 GB memory
[Change configuration](#)

Max vCores



Cost summary

General Purpose (GP_S_Gen5_1)	
Cost per GB (in USD)	0.14
Max storage selected (in GB)	x 1.3
ESTIMATED STORAGE COST / MONTH	0.18 USD
COMPUTE COST / VCORE SECOND ¹	0.000172 USD

NOTES

¹ Serverless databases are billed in vCore seconds based on a combination of CPU and memory utilization. [Learn more about serverless billing](#)

Import database

Microsoft Azure

Search resources, services, and docs (G+)

[Home](#) > [SQL servers](#) > [az-mitsubishi-sql-admin](#) >

↓

Import database

az-mitsubishi-sql-admin

Storage (Premium not supported) *

AllElectronics.bacpac
mitsubishiadmin/data
[Select backup](#)

Pricing tier * ⓘ

General Purpose - Serverless
Standard-series (Gen5), 1 vCore, 1 GB storage, zone redundant disabled
[Configure database](#)

Database name

AllElectronics

Collation * ⓘ

SQL_Latin1_General_CP1_CI_AS

Authentication type

SQL Server

Server admin login *

student

Password *

.....

OK

Lab2: ETL pipeline in Azure Data Factory (ADF) with Azure Data Lake (ADL)

Steps

- Lab2-0: Upload “imdb-top-1000.csv” into ADL (Storage account)
- Lab2-1: Create Dataflow “IMDBDataflow”
- Lab2-2: Create Pipeline “IMDBPipeline”
 - Add “Web node” to send email
 - Need to setup LogicApps
- Lab2-3: Set triggers (scheduler)

Lab2-1: Create Dataflow “IMDBDataflow”

- Create Link
 - AzureDataLakeStorage
- Create Data Set
 - IMDBsource “imdb-top-1000.csv” into ADL

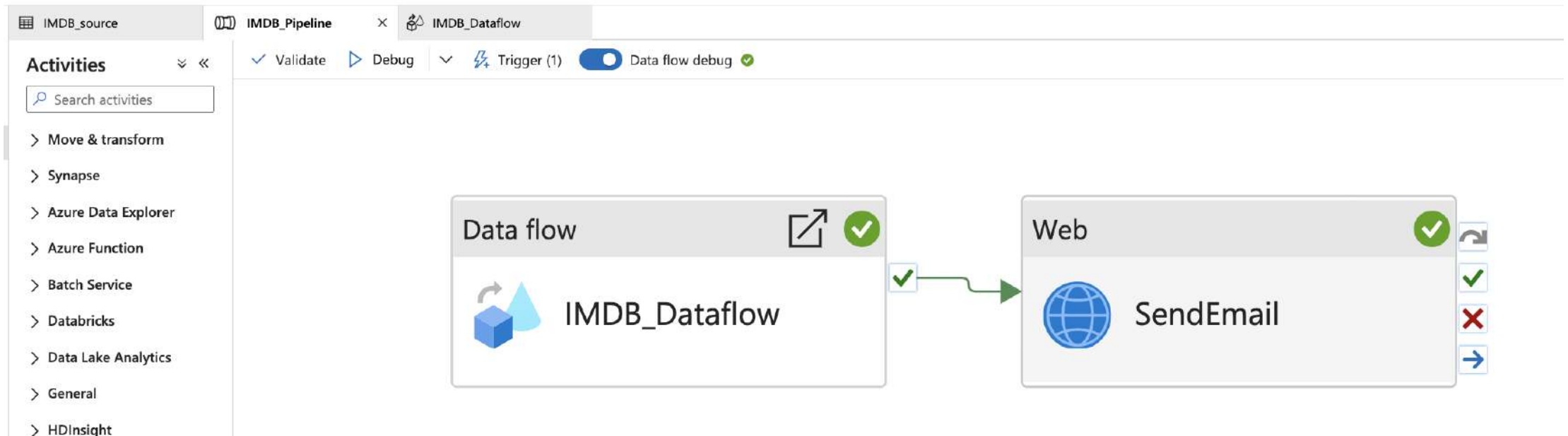
Lab2-1: Create Dataflow “IMDBDataflow” (cont.)

The screenshot displays the Microsoft Azure Data Factory interface for the 'az-mitsubishi-adf-admin' workspace. The left sidebar shows the 'Factory Resources' tree with 'IMDB_Dataflow' selected under 'Data flows'. The main canvas shows the 'IMDB_Dataflow' pipeline with the following steps, each highlighted with a yellow number:

- 1** **IMDBSource**: Import data from IMDB_source
- 2** **RenameFields**: Renaming IMDBSource to RenameFields with columns 'Poster_Link, Series_Title, ReleasedYear, Certificate'
- 3** **NewMovieFilter**: Filtering rows using expressions on columns 'ReleasedYear'
- 4** **ParseGenre**: Creating/updating the columns 'Poster_Link, Series_Title, ReleasedYear, Certificate, Runtime, Genre, Rating'
- 5** **WindowByGenreYear**: Aggregates data based on a window and joins with original data
- 6** **Aggregate**: Aggregating data by 'PrimaryGenre, ReleasedYear' producing columns 'AvgRating, HighestRatedMovie'
- 7** **sink1**: Add sink dataset

The interface also includes a top navigation bar with 'Validate all' and 'Publish all' buttons, a search bar, and a 'Preview experience' toggle. The bottom right corner shows a vertical toolbar with icons for zooming and other actions.

Lab2-2: Create Pipeline “IMDBPipeline”



Create LogicApps

The screenshot displays the Microsoft Azure Logic App Designer interface. The top navigation bar includes the Microsoft Azure logo, a search bar, and various utility icons. The breadcrumb trail shows 'Home > az-mitsubishi-logicapps-admin'. The main header identifies the workspace as 'az-mitsubishi-logicapps-admin | Logic app designer'. A left-hand sidebar contains navigation links for Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, and a 'Development Tools' section with Logic app designer (selected), Logic app code view, Versions, API connections, and Quick start guides. Below this is a 'Settings' section with Workflow settings, Authorization, Access keys, Identity, and Properties. The main workspace features a toolbar with Save, Discard, Run Trigger, Designer (active), Code view, Parameters, Templates, Connectors, Help, and Info. The workflow canvas shows a trigger 'When a HTTP request is received' with an 'HTTP POST URL' field containing a long URL and a 'Request Body JSON Schema' field with a JSON definition. Below the schema is a link 'Use sample payload to generate schema' and an 'Add new parameter' input. An arrow points from the trigger to an action 'Send an email (V2)'.

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

Home > az-mitsubishi-logicapps-admin

az-mitsubishi-logicapps-admin | Logic app designer

Logic app

Search

Save Discard Run Trigger Designer Code view Parameters Templates Connectors Help Info

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Development Tools

Logic app designer

Logic app code view

Versions

API connections

Quick start guides

Settings

Workflow settings

Authorization

Access keys

Identity

Properties

When a HTTP request is received

HTTP POST URL

https://prod-19.southeastasia.logic.azure.com:443/workflows/e9a8fc78e82b4515bdebfd7ab535c...

Request Body JSON Schema

```
{
  "properties": {
    "dataFactoryName": {
      "type": "string"
    },
    "message": {
      "type": "string"
    },
    "pipelineName": {

```

Use sample payload to generate schema

Add new parameter

Send an email (V2)

Create Web node in Pipeline

The screenshot shows the Azure Data Factory pipeline editor interface. At the top, there are tabs for 'IMDB_source', 'IMDB_Pipeline', and 'IMDB_Dataflow'. Below the tabs, there's a toolbar with 'Validate', 'Debug', 'Trigger (1)', and 'Data flow debug' (checked). The main canvas displays a pipeline with two nodes: 'Data flow' (labeled 'IMDB_Dataflow') and 'Web' (labeled 'SendEmail'). A green arrow indicates the flow from 'Data flow' to 'Web'. The 'Web' node is highlighted with a blue border. Below the canvas, there are three tabs: 'General', 'Settings', and 'User properties'. The 'Settings' tab is active, showing the following configuration:

- URL ***: `https://prod-19.southeastasia.logic.azure....`
Information will be sent to the URL specified. Please ensure you trust the URL entered.
- Method ***: `POST`
- Body**: `{ "message": "This is a custom dynamic message from your pipeline with run ID @pipeline().RunId", "dataFactoryName": "..." }`
- Authentication**: `None`
- Headers**: `+ New`
- Advanced**: `>`

The screenshot shows the 'Parameters' tab of the 'Web' node configuration. It has tabs for 'Parameters', 'Variables', 'Settings', and 'Output'. The 'Parameters' tab is active, showing a table with columns for 'Name', 'Type', and 'Default value'. There is a '+ New' button and a 'Delete' button (trash icon).

Name	Type	Default value
receiver	String	peerapon.vateekul@gmail.cc

Lab2-3: Set triggers (scheduler)

Microsoft Azure | Data Factory | az-mitsubishi-adf-admin

Factory Resources

- Pipelines: 1
 - IMDB_Pipeline
- Change Data Capture (preview): 0
- Datasets: 1
 - IMDB_source
- Data flows: 1
 - IMDB_Dataflow
- Power Query: 0

Activities

- Move & transform
- Synapse
- Azure Data Explorer
- Azure Function
- Batch Service
- Databricks
- Data Lake Analytics
- General
- HDInsight
- Iteration & conditionals
- Machine Learning
- Power Query

IMDB_Pipeline

IMDB_Dataflow

Data flow

IMDB_Dataflow

Parameters

Name	Type	Default value
receiver	String	peerapon.vateekul@gmail.cc

Edit trigger

Name *

Every 3 minutes

Description

Type *

ScheduleTrigger

Start date *

3/26/2023, 5:22:00 PM

Time zone *

Bangkok, Hanoi, Jakarta (UTC+7)

Recurrence *

Every 3 Minute(s)

☒ Specify an end date

End On *

3/26/2023, 5:30:00 PM

Annotations

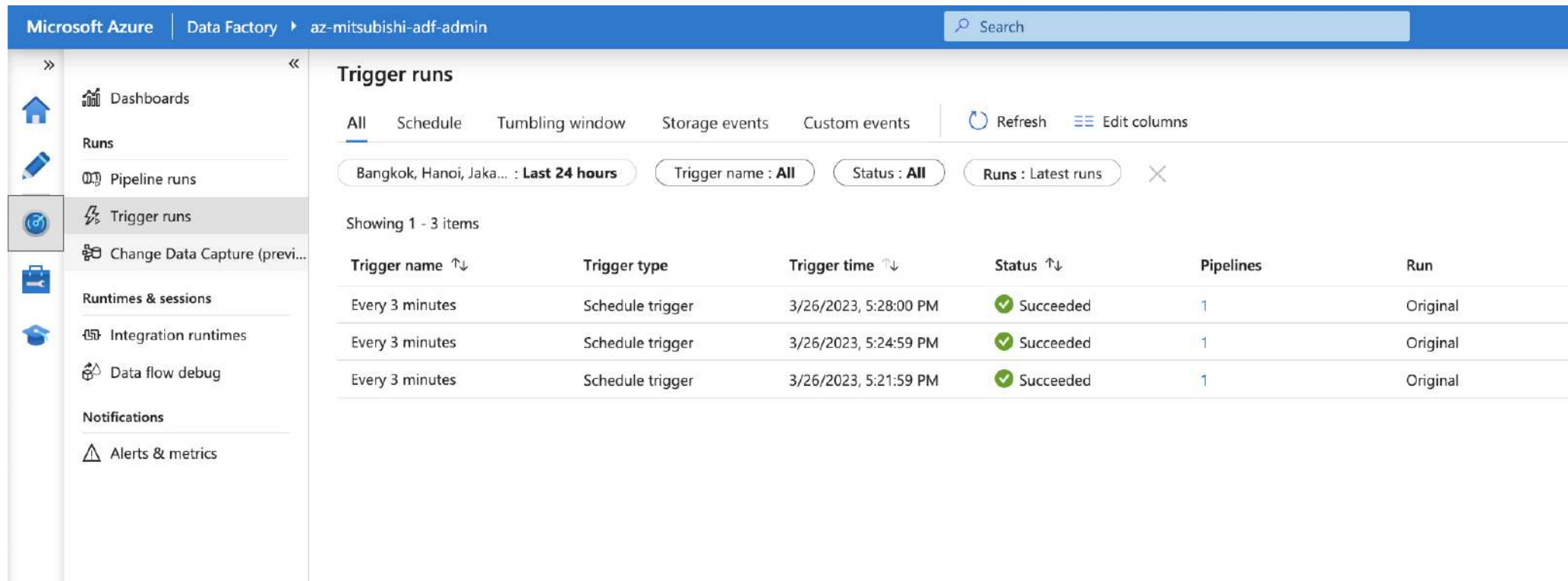
+ New

Status

☒ Started ☐ Stopped

Finally, you can check your trigger here 😊

- Monitor → Trigger runs



The screenshot displays the 'Trigger runs' page in the Microsoft Azure Data Factory portal. The left sidebar shows the navigation menu with 'Trigger runs' selected. The main content area shows a table of trigger runs. The table has columns for Trigger name, Trigger type, Trigger time, Status, Pipelines, and Run. Three runs are listed, all with a status of 'Succeeded'.

Trigger name	Trigger type	Trigger time	Status	Pipelines	Run
Every 3 minutes	Schedule trigger	3/26/2023, 5:28:00 PM	✓ Succeeded	1	Original
Every 3 minutes	Schedule trigger	3/26/2023, 5:24:59 PM	✓ Succeeded	1	Original
Every 3 minutes	Schedule trigger	3/26/2023, 5:21:59 PM	✓ Succeeded	1	Original