

Sulaiman Ali Data Engineer Test Akal Karya

Data and Query Task (Basic)

1. Create script to create table for each object
 - a. Employee
 - b. PositionHistory
2. Create insert script to inserting data into each table (*Employee and PositionHistory*)
3. Create query to display all employee (*EmployeeId, FullName, BirthDate, Address*) data with their **current** position information (*PosId, PosTitle, EmployeeId, StartDate, EndDate*).

1. Create script to create table for each object
 - a. Employee
 - b. PositionHistory

```
USE [TesAkalkarya]
GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[PositionHistory](
    [Id] [int] IDENTITY (1,1) NOT NULL PRIMARY KEY,
    [PosId] [nchar](10) NOT NULL,
    [PosTitle] [nchar](100) NOT NULL,
    [EmployeeId] [nchar](10) NOT NULL,
    [StartDate] [date] NOT NULL,
    [EndDate] [date] NOT NULL
) ON [PRIMARY]
GO
```

```
1 USE [TesAkalkarya]
2 GO
3
4 SET ANSI_NULLS ON
5 GO
6
7 SET QUOTED_IDENTIFIER ON
8 GO
9
10 Create TABLE [dbo].[Employee](
11     [Id] [int] IDENTITY (1,1) NOT NULL,
12     [EmployeeId] [nchar](10) NOT NULL,
13     [FullName] [nchar](100) NOT NULL,
14     [BirthDate] [date] NOT NULL,
15     [Address] [nchar](500) NULL
16 ) ON [PRIMARY]
17 GO
18
```

2. Create insert script to inserting data into each table (*Employee and PositionHistory*)

```
1 USE [TesAkaKarya]
2
3 INSERT INTO [dbo].[Employee] ([EmployeeId], [FullName], [BirthDate], [Address])
4 VALUES
5 ('10105001', 'Ali Anton', '1982-08-19', 'Jakarta Utara'),
6 ('10105002', 'Rara Siva', '1982-01-01', 'Mandalika'),
7 ('10105003', 'Rini Aini', '1982-02-20', 'Sumbawa Besar'),
8 ('10105004', 'Budi', '1982-02-22', 'Mataram Kota');
9
10 select * from Employee
11
12 INSERT INTO [dbo].[PositionHistory] ([PosId], [PosTitle], [EmployeeID], [StartDate], [EndDate])
13 VALUES
14 ('50000', 'IT Manager', '10105001', '2022-01-01', '2022-02-28'),
15 ('50001', 'IT Sr. Manager', '10105001', '2022-03-01', '2022-12-31'),
16 ('50002', 'Programmer Analyst', '10105002', '2022-01-01', '2022-02-28'),
17 ('50003', 'Sr. Programmer Analyst', '10105002', '2022-03-01', '2022-12-31'),
18 ('50004', 'IT Admin', '10105003', '2022-01-01', '2022-02-28'),
19 ('50005', 'IT Secretary', '10105003', '2022-03-01', '2022-12-31');
20
21 select * from PositionHistory
22 select * from Employee
```

3.

Create query to display all employee (*EmployeeId*, *FullName*, *BirthDate*, *Address*) data with their **current** position information (*PosId*, *PosTitle*, *EmployeeId*, *StartDate*, *EndDate*).

```
1 SELECT
2     a.EmployeeId,
3     a.FullName,
4     a.BirthDate,
5     a.Address,
6     a.PosId,
7     a.PosTitle,
8     a.StartDate,
9     a.EndDate
10 FROM (
11     SELECT
12         E.EmployeeId,
13         E.FullName,
14         E.BirthDate,
15         E.Address,
16         PH.PosId,
17         PH.PosTitle,
18         PH.StartDate,
19         PH.EndDate,
20         ROW_NUMBER() OVER (PARTITION BY e.EmployeeId ORDER BY ph.EndDate DESC) AS RowNum
21     FROM
22         Employee e
23     LEFT JOIN
24         PositionHistory ph ON e.EmployeeId = ph.EmployeeId
25 ) a
26 WHERE
27     RowNum = 1;
```

100 %

Results Messages

	EmployeeId	FullName	BirthDate	Address	PosId	PosTitle	StartDate	EndDate
1	10105001	Ali Anton	1982-08-19	Jakarta Utara	50001	IT Sr. Manager	2022-03-01	2022-12-31
2	10105002	Rara Siva	1982-01-01	Mandalika	50003	Sr. Programmer Analyst	2022-03-01	2022-12-31
3	10105003	Rini Aini	1982-02-20	Sumbawa Besar	50005	IT Secretary	2022-03-01	2022-12-31
4	10105004	Budi	1982-02-22	Mataram Kota	NULL	NULL	NULL	NULL

ETL, Data Warehouse and Analytics Task

Build and design simple Data Warehouse and analytics data, data source came from 2 different source

1. Employee data, source from Azure
2. Training history, source from GCP (Google Cloud Platform)

Requirement and Details:

1. Azure Employee data is using SQL Server as Database Server. This is database for saving employment data like their employee number, name, birthdate, position/job etc. Each employee has one record for their employment data.
2. Training History data is using Google Worksheet as data platform. The worksheet contains historical data about training that completed by employee. Each employee may have more than one records for their training history data.
3. Design simple ETL Flow that compile these two data sources into Data Warehouse
4. Design simple report that displays historical training data
5. Design simple dashboard that displays
 - a. Total employee completed training each month
 - b. Total training each month

Data Schema

Source Data

Source Data on
Google Platform



Google Sheet



Google Cloud Platform

Source Data
on Premise



Employee



Position
History



SQL Server
(Oreo281)



Microsoft®
SQL Server™
Integration Services

SSIS



Azure SQL Server
(Oreo2811)

Microsoft®
SQL Azure™

Azure



Azure
Data Factory



Sql Azure Cloud



Power BI

Dashboard

Present

SQL Server Azure

- Database Cloud Azure
- NAMA SERVER :oreo2811.database.windows.net
- USER:oreo2811
- PASS:Pa55word@

Penambahan Source
dan sync_Date untuk
proses etl di azure

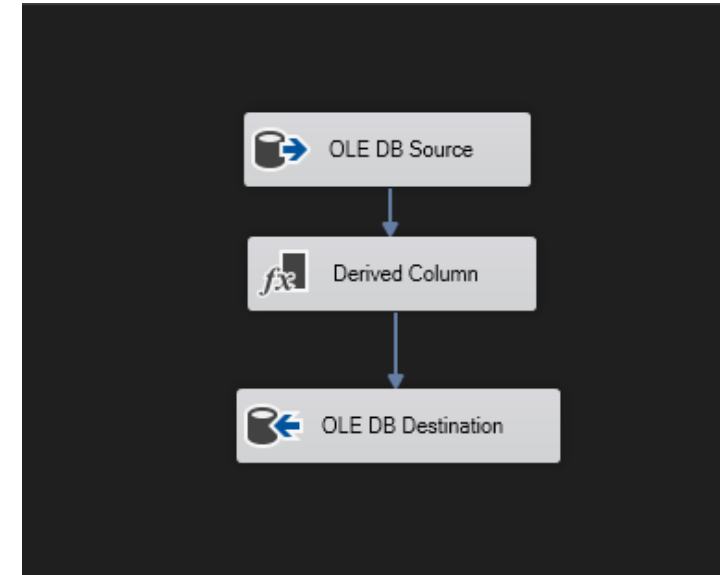
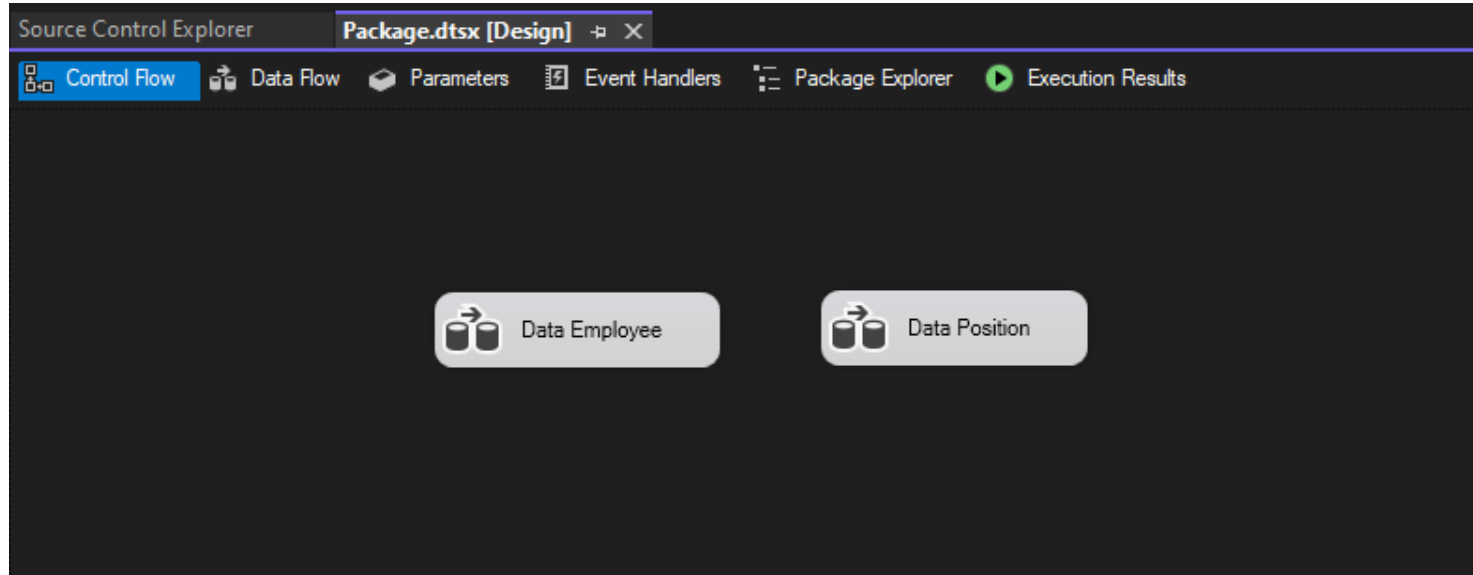
```
20
21 select * from PositionHistory
22 select * from Employee
23
24
25
26
27
```

100 %

	Id	PosId	PosTitle	EmployeeId	StartDate	EndDate	Source	Sync_date
1	1	50000	IT Manager	10105001	2022-01-01	2022-02-28	NULL	NULL
2	2	50001	IT Sr. Manager	10105001	2022-03-01	2022-12-31	NULL	NULL
3	3	50002	Programmer Analyst	10105002	2022-01-01	2022-02-28	NULL	NULL
4	4	50003	Sr. Programmer Analyst	10105002	2022-03-01	2022-12-31	NULL	NULL
5	5	50004	IT Admin	10105003	2022-01-01	2022-02-28	NULL	NULL
6	6	50005	IT Secretary	10105003	2022-03-01	2022-12-31	NULL	NULL

	Id	EmployeeId	FullName	BirthDate	Address	Source	Sync_date
1	1	10105001	Ali Anton	1982-08-19	Jakarta Utara	NULL	NULL
2	2	10105002	Rara Siva	1982-01-01	Mandalika	NULL	NULL
3	3	10105003	Rini Aini	1982-02-20	Sumbawa Besar	NULL	NULL
4	4	10105004	Budi	1982-02-22	Mataram Kota	NULL	NULL

Create ETL SSIS



Data yang telah dilakukan proses etl ke sql azure menggunakan ssis

43	select * from Employee
44	

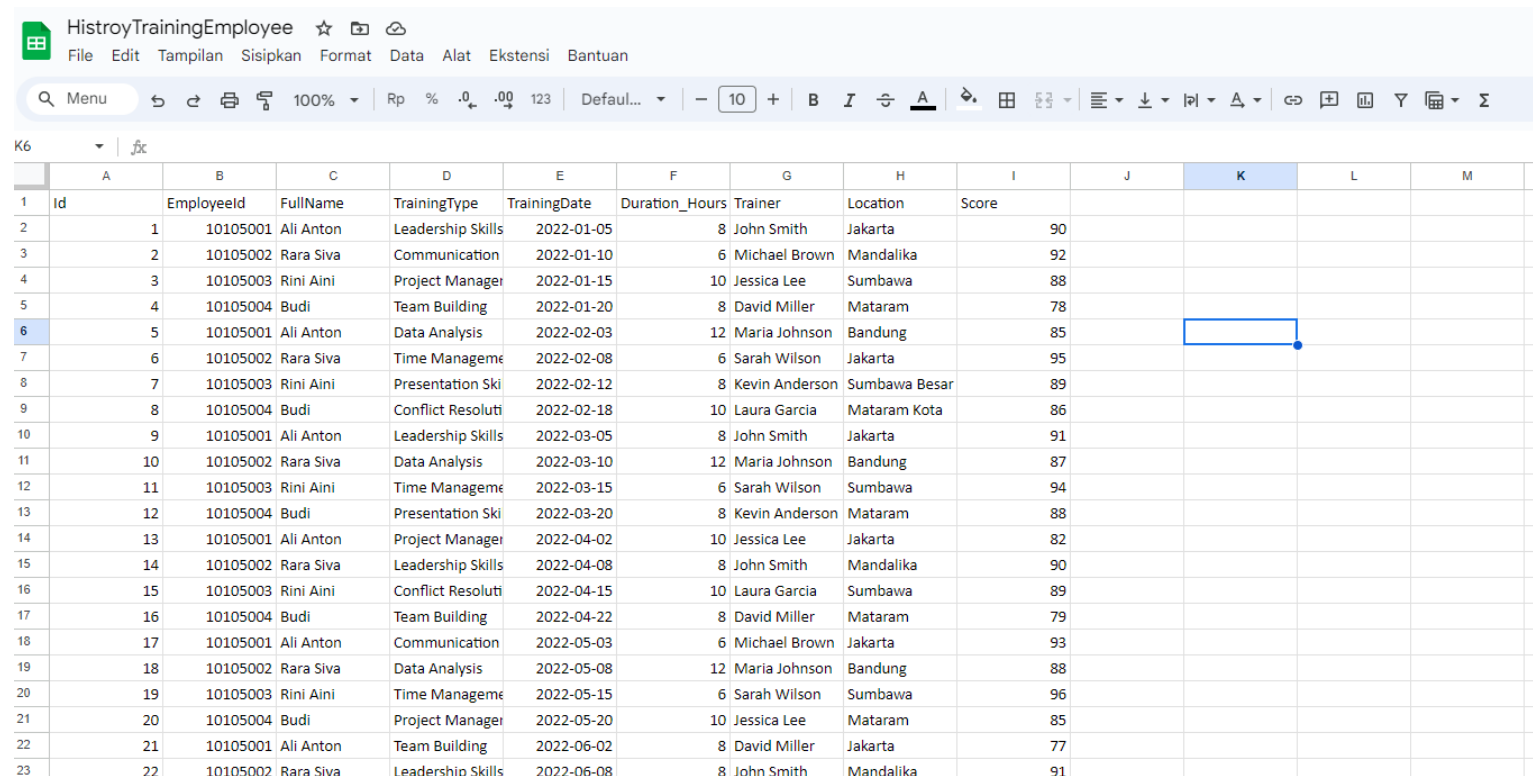
	Id	EmployeeId	FullName	BirthDate	Address	Source	Synch_date
1	1	10105001	Ali Anton	1982-08-19	Jakarta Utara	MSSQL_OREO281	2023-12-19
2	2	10105002	Rara Siva	1982-01-01	Mandalika	MSSQL_OREO281	2023-12-19
3	3	10105003	Rini Aini	1982-02-20	Sumbawa Besar	MSSQL_OREO281	2023-12-19
4	4	10105004	Budi	1982-02-22	Mataram Kota	MSSQL_OREO281	2023-12-19

46	select * from PositionHistory
----	-------------------------------

	Id	PosId	PosTitle	EmployeeId	StartDate	EndDate	Source	Synch_date
1	1	50000	IT Manager	10105001	2022-01-01	2022-02-28	MSSQL_OREO281	2023-12-19
2	2	50001	IT Sr. Manager	10105001	2022-03-01	2022-12-31	MSSQL_OREO281	2023-12-19
3	3	50002	Programmer Analyst	10105002	2022-01-01	2022-02-28	MSSQL_OREO281	2023-12-19
4	4	50003	Sr. Programmer Analyst	10105002	2022-03-01	2022-12-31	MSSQL_OREO281	2023-12-19
5	5	50004	IT Admin	10105003	2022-01-01	2022-02-28	MSSQL_OREO281	2023-12-19
6	6	50005	IT Secretary	10105003	2022-03-01	2022-12-31	MSSQL_OREO281	2023-12-19

Source from Google Sheet

- https://docs.google.com/spreadsheets/d/1Akupq7_J_Cp61tRmJFGYC_N54-Id6QptDwPbXXz7CP-I/edit#gid=0



HistroyTrainingEmployee

File Edit Tampilan Sisipkan Format Data Alat Ekstensi Bantuan

Menu 100% Rp % 123 Default... 10 + B I A

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Id	Employeeid	FullName	TrainingType	TrainingDate	Duration_Hours	Trainer	Location	Score				
2	1	10105001	Ali Anton	Leadership Skills	2022-01-05	8	John Smith	Jakarta	90				
3	2	10105002	Rara Siva	Communication	2022-01-10	6	Michael Brown	Mandalika	92				
4	3	10105003	Rini Aini	Project Manager	2022-01-15	10	Jessica Lee	Sumbawa	88				
5	4	10105004	Budi	Team Building	2022-01-20	8	David Miller	Mataram	78				
6	5	10105001	Ali Anton	Data Analysis	2022-02-03	12	Maria Johnson	Bandung	85				
7	6	10105002	Rara Siva	Time Manageme	2022-02-08	6	Sarah Wilson	Jakarta	95				
8	7	10105003	Rini Aini	Presentation Ski	2022-02-12	8	Kevin Anderson	Sumbawa Besar	89				
9	8	10105004	Budi	Conflict Resoluti	2022-02-18	10	Laura Garcia	Mataram Kota	86				
10	9	10105001	Ali Anton	Leadership Skills	2022-03-05	8	John Smith	Jakarta	91				
11	10	10105002	Rara Siva	Data Analysis	2022-03-10	12	Maria Johnson	Bandung	87				
12	11	10105003	Rini Aini	Time Manageme	2022-03-15	6	Sarah Wilson	Sumbawa	94				
13	12	10105004	Budi	Presentation Ski	2022-03-20	8	Kevin Anderson	Mataram	88				
14	13	10105001	Ali Anton	Project Manager	2022-04-02	10	Jessica Lee	Jakarta	82				
15	14	10105002	Rara Siva	Leadership Skills	2022-04-08	8	John Smith	Mandalika	90				
16	15	10105003	Rini Aini	Conflict Resoluti	2022-04-15	10	Laura Garcia	Sumbawa	89				
17	16	10105004	Budi	Team Building	2022-04-22	8	David Miller	Mataram	79				
18	17	10105001	Ali Anton	Communication	2022-05-03	6	Michael Brown	Jakarta	93				
19	18	10105002	Rara Siva	Data Analysis	2022-05-08	12	Maria Johnson	Bandung	88				
20	19	10105003	Rini Aini	Time Manageme	2022-05-15	6	Sarah Wilson	Sumbawa	96				
21	20	10105004	Budi	Project Manager	2022-05-20	10	Jessica Lee	Mataram	85				
22	21	10105001	Ali Anton	Team Building	2022-06-02	8	David Miller	Jakarta	77				
23	22	10105002	Rara Siva	Leadership Skills	2022-06-08	8	John Smith	Mandalika	91				

Make Credential to Access Sheet

Google Cloud

HistoryEmployeeTraining

Search (/) for resources, docs, products, and more

Search

3

APIs & Services

Enabled APIs & services

Library

Credentials

OAuth consent screen

Page usage agreements

Credentials

+ CREATE CREDENTIALS

DELETE

RESTORE DELETED CREDENTIALS

Create credentials to access your enabled APIs. [Learn more](#)

Remember to configure the OAuth consent screen with information about your application.

CONFIGURE CONSENT SCREEN

API Keys

<input type="checkbox"/>	Name	Creation date ↓	Restrictions	Actions
<input type="checkbox"/>	<div><div></div>API key 1</div>	Dec 19, 2023	None	SHOW KEY ⋮

OAuth 2.0 Client IDs

<input type="checkbox"/>	Name	Creation date ↓	Type	Client ID	Actions
No OAuth clients to display					

Service Accounts

[Manage service accounts](#)

<input type="checkbox"/>	Email	Name ↑	Actions
No service accounts to display			

Create Linked Services dari GCP and From AzureSQL Database

The screenshot displays the Microsoft Azure Data Factory portal interface. The top navigation bar includes the Microsoft Azure logo, the breadcrumb 'Data Factory > Oreo2811', a search bar, and user information 'sulaimanali281@gmail.com'. A notification banner at the top left mentions the public preview of Microsoft Fabric. The left-hand navigation pane lists various components: General (Factory settings), Connections (Linked services, Integration runtimes, Microsoft Purview), Source control (Git configuration, ARM template), Author (Triggers, Global parameters, Data flow libraries), and Security (Credentials, Customer managed key). The 'Linked services' section is currently selected. The main content area, titled 'Linked services', provides a description and a '+ New' button. Below this is a table listing existing linked services. The table has columns for Name, Type, and Related. Two services are listed: 'AzureSqlDatabase1' (Azure SQL Database) with 3 related items, and 'HistoryTrainingEmployee' (Google Sheets (Preview)) with 1 related item. On the right side, the 'Edit linked service' panel is open for 'HistoryTrainingEmployee'. It shows the service name, a description field, a dropdown for 'Connect via integration runtime' set to 'AutoResolveIntegrationRuntime', and tabs for 'Token' and 'Azure Key Vault'. The 'Token' tab is active, showing a 'Token' field with masked characters. Below these are sections for 'Annotations', 'Parameters', and 'Advanced'.

Microsoft Azure | Data Factory > Oreo2811

Search factory and documentation

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

>> Data Factory > Validate all > Publish all

General

Factory settings

Connections

Linked services

Integration runtimes

Microsoft Purview

Source control

Git configuration

ARM template

Author

Triggers

Global parameters

Data flow libraries

Security

Credentials

Customer managed key

Linked services

Linked service defines the connection information to a data store or compute. [Learn more](#)

+ New

Filter by name

Annotations : Any

Showing 1 - 2 of 2 items

Name	Type	Related
AzureSqlDatabase1	Azure SQL Database	3
HistoryTrainingEmployee	Google Sheets (Preview)	1

Edit linked service

Google Sheets (Preview)

Name *

HistoryTrainingEmployee

Description

Connect via integration runtime *

AutoResolveIntegrationRuntime

Token Azure Key Vault

Token *

Annotations

+ New

> Parameters

> Advanced

Create Table HistoryTrainingEmployee

```
49 Create TABLE [dbo].[HistroyTrainingEmployee](
50     [Id] [int] IDENTITY (1,1) NOT NULL,
51     [EmployeeId] [nchar](10) NOT NULL,
52     [FullName] [nchar](100) NOT NULL,
53     [TrainingType] [nchar](200) NOT NULL,
54     [TrainingDate] [date] NOT NULL,
55     [DurationHours] [nchar](10) NOT NULL,
56     [Trainer] [nchar](100) NOT NULL,
57     [Location] [nchar](100) NOT NULL,
58     [Score] [nchar](100) NOT NULL
59 ) ON [PRIMARY]
60 GO
61 select * from HistroyTrainingEmployee
```

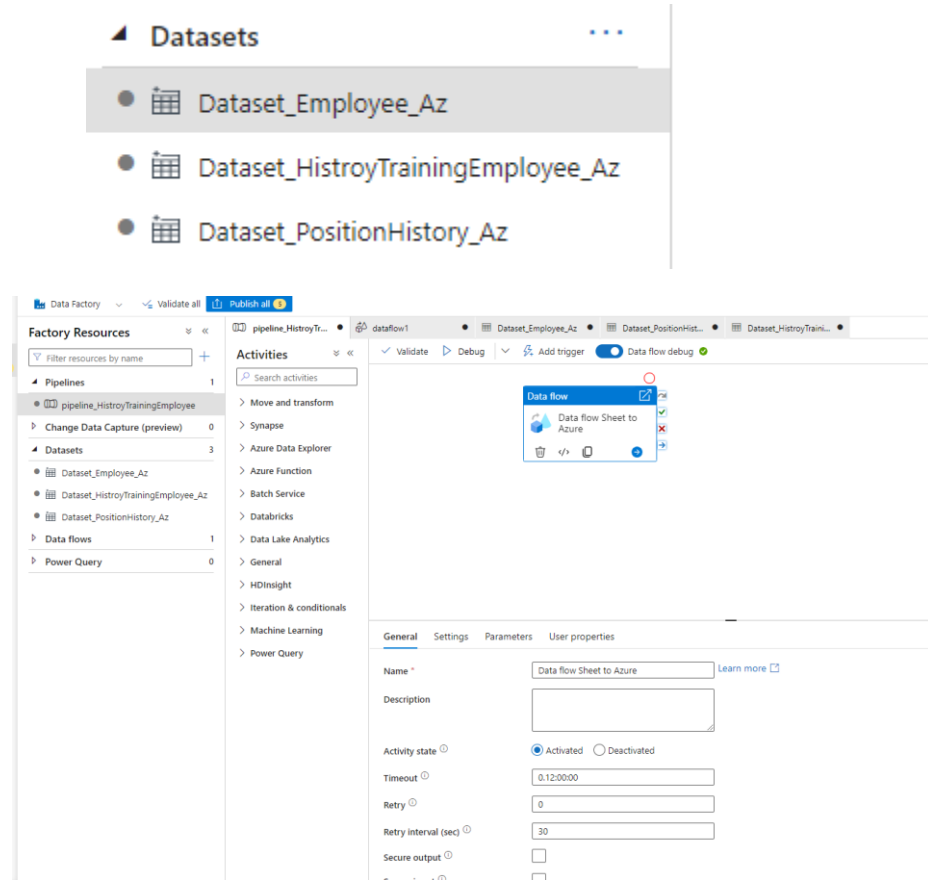
100 %

Results Messages

Id	EmployeeId	FullName	TrainingType	TrainingDate	DurationHours	Trainer	Location	Score
----	------------	----------	--------------	--------------	---------------	---------	----------	-------

- Create dataset

- Create pipeline to flow data from google sheet to azure dataset



- Run Pipeline
- Check on Sql Azure

63 `select * from HistroyTrainingEmployee`

100 %

Results Messages

	Id	EmployeeId	FullName	TrainingType	TrainingDate	DurationHours	Trainer	Location	Score
1	1	10105001	Ali Anton	Leadership Skills	2022-01-05	8	John Smith	Jakarta	90
2	2	10105002	Rara Siva	Communication Skills	2022-01-10	6	Michael Brown	Mandalika	92
3	3	10105003	Rini Aini	Project Management	2022-01-15	10	Jessica Lee	Sumbawa	88
4	4	10105004	Budi	Team Building	2022-01-20	8	David Miller	Mataram	78
5	5	10105001	Ali Anton	Data Analysis	2022-02-03	12	Maria Johnson	Bandung	85
6	6	10105002	Rara Siva	Time Management	2022-02-08	6	Sarah Wilson	Jakarta	95
7	7	10105003	Rini Aini	Presentation Skills	2022-02-12	8	Kevin Anderson	Sumbawa Besar	89
8	8	10105004	Budi	Conflict Resolution	2022-02-18	10	Laura Garcia	Mataram Kota	86
9	9	10105001	Ali Anton	Leadership Skills	2022-03-05	8	John Smith	Jakarta	91
10	10	10105002	Rara Siva	Data Analysis	2022-03-10	12	Maria Johnson	Bandung	87
11	11	10105003	Rini Aini	Time Management	2022-03-15	6	Sarah Wilson	Sumbawa	94
12	12	10105004	Budi	Presentation Skills	2022-03-20	8	Kevin Anderson	Mataram	88
13	13	10105001	Ali Anton	Project Management	2022-04-02	10	Jessica Lee	Jakarta	82
14	14	10105002	Rara Siva	Leadership Skills	2022-04-08	8	John Smith	Mandalika	90
15	15	10105003	Rini Aini	Conflict Resolution	2022-04-15	10	Laura Garcia	Sumbawa	89
16	16	10105004	Budi	Team Building	2022-04-22	8	David Miller	Mataram	79
17	17	10105001	Ali Anton	Communication Skills	2022-05-03	6	Michael Brown	Jakarta	93
18	18	10105002	Rara Siva	Data Analysis	2022-05-08	12	Maria Johnson	Bandung	88
19	19	10105003	Rini Aini	Time Management	2022-05-15	6	Sarah Wilson	Sumbawa	96
20	20	10105004	Budi	Project Management	2022-05-20	10	Jessica Lee	Mataram	85
21	21	10105001	Ali Anton	Team Building	2022-06-02	8	David Miller	Jakarta	77
22	22	10105002	Rara Siva	Leadership Skills	2022-06-08	8	John Smith	Mandalika	91
23	23	10105003	Rini Aini	Communication Skills	2022-06-15	6	Michael Brown	Sumbawa Besar	92
24	24	10105004	Budi	Data Analysis	2022-06-22	12	Maria Johnson	Mataram Kota	84
25	25	10105001	Ali Anton	Presentation Skills	2022-07-05	8	Kevin Anderson	Jakarta	89
26	26	10105002	Rara Siva	Time Management	2022-07-10	6	Sarah Wilson	Bandung	93
27	27	10105003	Rini Aini	Team Building	2022-07-15	8	David Miller	Sumbawa	76
28	28	10105004	Budi	Conflict Resolution	2022-07-20	10	Laura Garcia	Mataram	87
29	29	10105001	Ali Anton	Leadership Skills	2022-08-02	8	John Smith	Jakarta	90

✓ Validate ▶ Debug ▼ ⚡ Add trigger Data flow debug

Data flow

Data flow Sheet to Azure

Parameters Variables Settings **Output**

Pipeline run ID: 9edfd3eb-14fa-4cea-816f-129a71fb1476 **Pipeline status:** ✔ Succeeded [View debug run consumption](#)

ⓘ Data flow activity for this debug run will start as soon as the data flow debug session is ready. ✕

All status ▼ [Monitor in Azure Metrics](#) [Export to CSV](#) ⌵

Showing 1 - 1 of 1 items

	Activity status	Activity type	Run start	Duration	Integration runtime	User properties
re	✔ Succeeded	Data flow	12/20/2023, 1:39:49 AM	3m 19s	AutoResolveIntegration	

Dashboard Simple with Power BI

