

# Azure Trailblazer Academy

## Azure Data Factory (ADF) Lab

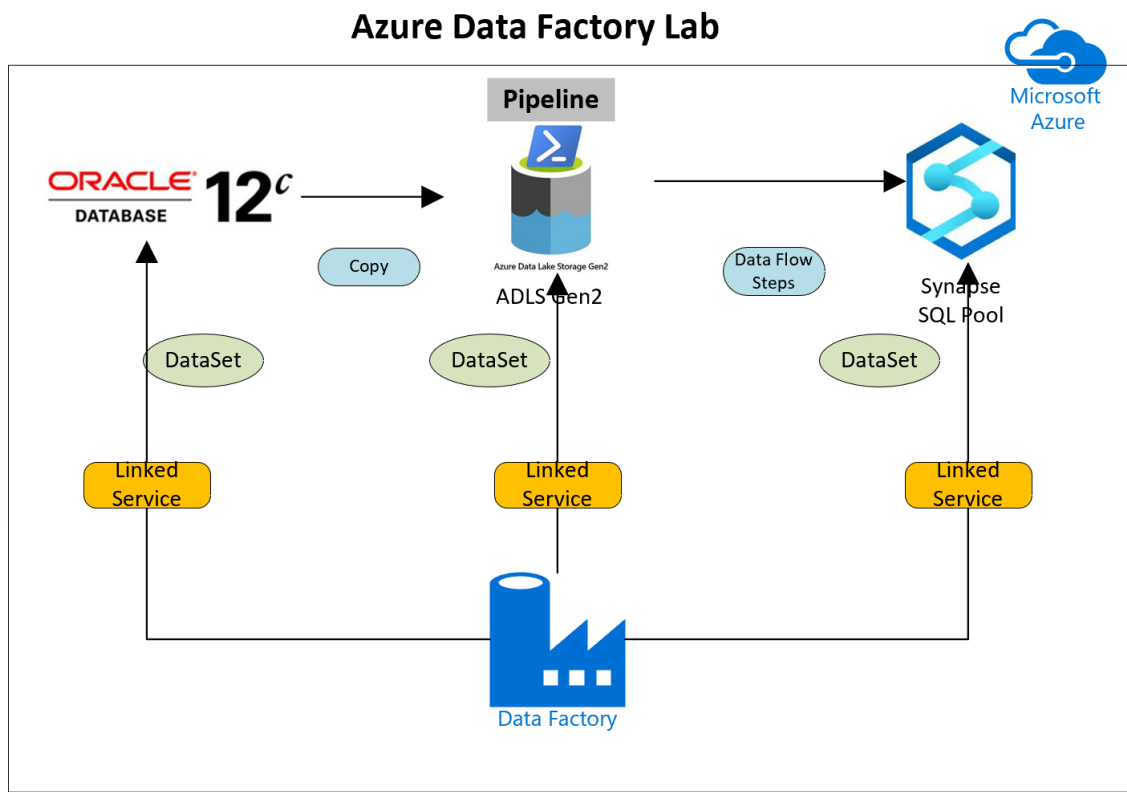
### Overview

Azure Data Factory is a PaaS cloud-based ETL & data integration tool that allows you to create data-driven workflows in the cloud for orchestrating and automating data movement and data transformation.

It is a great tool to orchestrate big data and operationalize processes to refine these enormous stores of raw data into actionable business insights.

### Lab Overview

This lab will help you gain the experience to ingest data from on-premises databases such as Oracle, SAP, Teradata, Hortonworks, DB2, SQL Server and Cloudera to Azure Data storage, databases and data warehouses services. It will showcase the steps to build a pipeline using ADF to ingest the data into ADLS GEN2 storage and secure the PII data using data transformation functions using the Data Flow activity and finally store the data in Synapse SQL Pool (Data warehouse) for building BI dashboards.



## Pre-requisites

- Write Access to Azure Data Lake Storage Account (ADLS Gen2)
- Read Access to Sample HR schema in Oracle Database
- Write Access to Synapse SQL Pool Data warehouse

## Automated Deployment

This automated deployment script will create the required Azure services for this lab. Press the "Deploy to Azure" button below.



[https://portal.azure.com/#create/Microsoft.Template/uri/https%3A%2F%2Fraw.githubusercontent.com%2Fmicrosoft%2FAzureTrailblazerAcademy%2Fmaster%2Fmonth2%2Fflabs%2Fflab\\_adf%2Fscripts%2Fflab2\\_data\\_deployment.json](https://portal.azure.com/#create/Microsoft.Template/uri/https%3A%2F%2Fraw.githubusercontent.com%2Fmicrosoft%2FAzureTrailblazerAcademy%2Fmaster%2Fmonth2%2Fflabs%2Fflab_adf%2Fscripts%2Fflab2_data_deployment.json)

Enter the following information

- Subscription: Enter your subscription.
- Resource group: Select 'Create new' under Resource group.  
Enter: 'ata-adf-lab- <YourName> -rg'
- Region: Select 'East US'.
- SQL\_Server Name: Enter 'ata-adf-lab-sql- <YourName>'
- Server\_location: Enter 'eastus'

# Custom deployment

Deploy from a custom template

Subscription \* ⓘ

Resource group \* ⓘ

[Create new](#)

## Parameters

Region \* ⓘ

SQL\_Server Name ⓘ

Server\_Location ⓘ

Administrator Login \* ⓘ

Administrator Login Password \*

Collation ⓘ

Database Name \* ⓘ

Sku Name ⓘ

Storage Accounts\_Data Lake Gen2 ⓘ

Datalakestorage\_location ⓘ




[Review + create](#)

[< Previous](#)

[Next : Review + create >](#)

- AdministratorLogin: 'azureadmin'
- Administrator Login Password: 'Ataadf123!'
- Database Name: 'ataadflabsqldb'
- Storage\_Accounts\_Data Lake Gen2: 'ataadflabstorage<YourName>'
- Click on 'Review and Create'
- Click on 'Create'

Please check if it created the following services after successful deployment.

<input type="checkbox"/>	 ataadflabsqldb (ata-adf-lab-sql-srini/ataadflabsqldb)	Synapse SQL pool (data warehouse)
<input type="checkbox"/>	 ata-adf-lab-sql-srini	SQL server
<input type="checkbox"/>	 ataadflabstoragesrini	Storage account

Temporary access to an Oracle database for completing this lab will be provided in the class.

# Task List

- Task-1: Create Azure Data Factory Service
- Task-2: Create linked services
- Task-3: Create Copy Activity to ingest Oracle HR Employee data to Azure Storage
- Task-4: Create Data Flows to Transform Employee data and store it in Synapse SQL Analytics
- Task-5: Build a pipeline to connect Copy and Dataflow activities and Test the Execution

## Task-1: Create Azure Data Factory Service

1. Provision an ADF service
  - Type 'Data factories' in the search bar.
  - Select 'Data factories' and select 'add' to create a new service

Provide the following info:

- Subscription: Make sure to select your subscription
- Select 'ata-adf-lab-<YourName>-rg', the resource group you have created with custom deployment.
- Name: 'ata-adf-lab-<YourName>'
- Location: select 'East US'
- Click the 'Next: Git configuration' button
- GIT configuration: check the 'Configure Git later' checkbox

# Create Data Factory

 Changes on this step may reset later selections you have made. Review all options prior to deployment.

- Basics
- Git configuration
- Networking
- Tags
- Review + create

## Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription \* ⓘ

Microsoft Azure Internal Consumption BP

Resource group \* ⓘ

ata-adf-lab-brpham2-rg

Create new

## Instance details

Region \* ⓘ

East US

Name \*

ata-adf-lab-brpham2

Version \* ⓘ

V2

# Create Data Factory

- Basics
- Git configuration
- Networking
- Tags
- Review + create

Azure Data Factory allows you to configure a Git repository with either Azure DevOps or GitHub. Git is a version control system that allows for easier change tracking and collaboration.  
[Learn more about Git integration in Azure Data Factory](#)

Configure Git later ⓘ

☒

## - Click on 'Create' button

[Home](#) > [Data factories](#) >

### Create Data Factory

✓ Validation Passed

[Basics](#) [Git configuration](#) [Networking](#) [Tags](#) [Review + create](#)

#### TERMS

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. See the [Azure Marketplace Terms](#) for additional details.

#### Basics

Subscription	Microsoft Azure Internal Consumption BP
Resource group	ata-adf-lab-brpham2-rg
Region	East US
Name	ata-adf-lab-brpham2
Version	V2

#### Networking

Connect via	Public endpoint
-------------	-----------------

Create

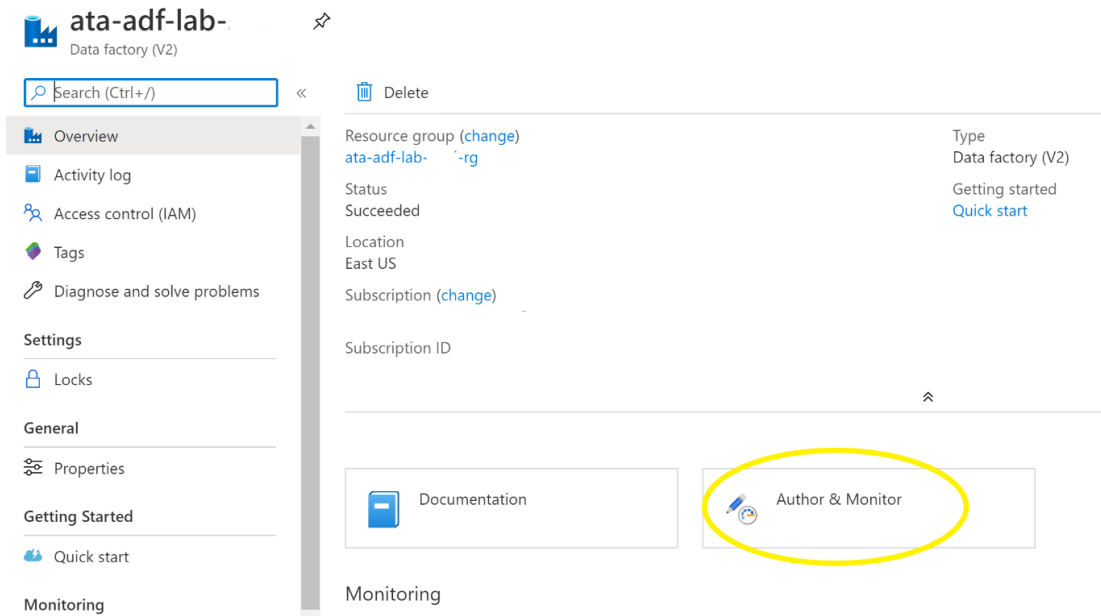
< Previous

Next

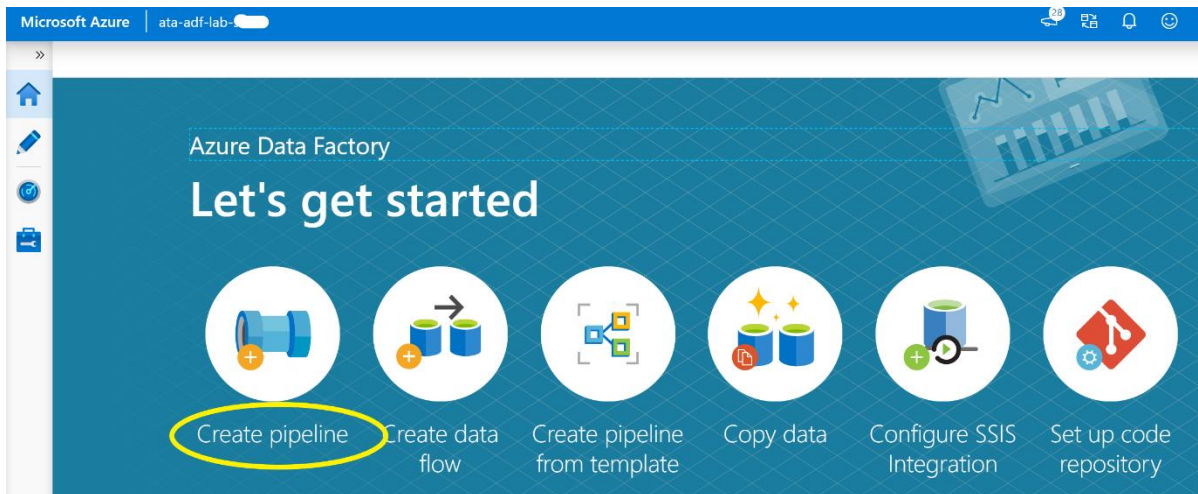
[Download a template for automation](#)

## 2. Open ADF Author Tool

- Select 'Go to resource' when it completes the deployment.
- Select 'Author & Monitor' in the middle of the screen.

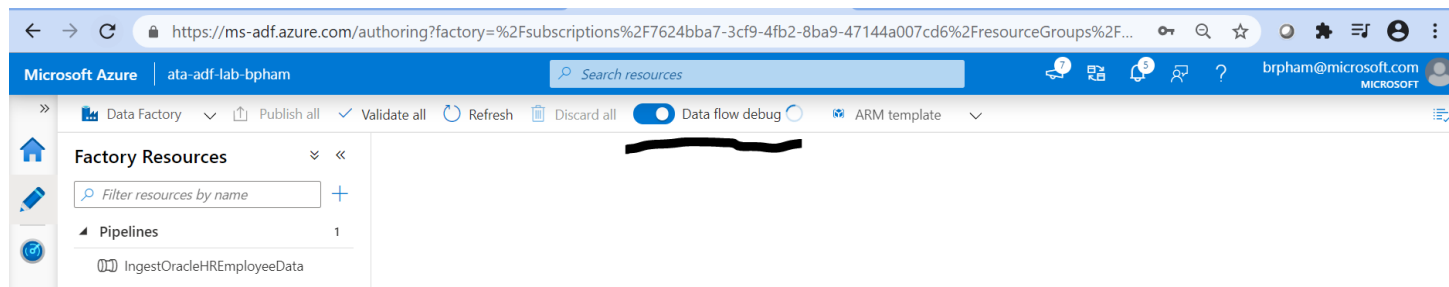


Opens up a new tab introducing the drag and drop interface to build pipelines.



Turn on Dataflow Debug functionality to test the modifications.

- To preview data, we need to turn on the 'data flow debug' option

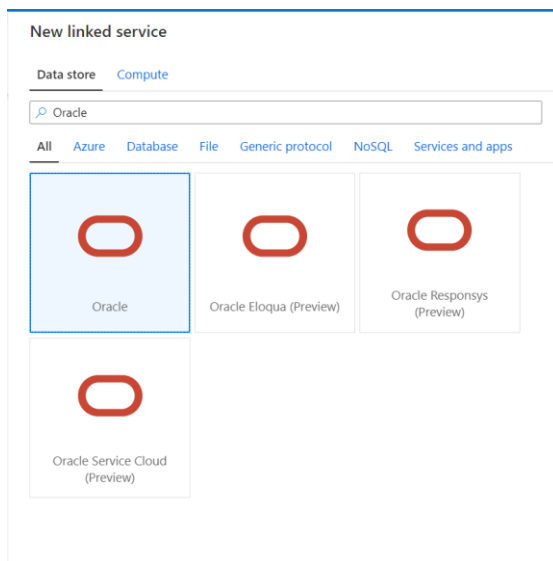


## Task-2: Create linked services

You will be creating connection linked services to sync and source systems such as Oracle, Azure Data Lake Storage (ADLS) Gen2 and Synapse Analytics (formerly SQL DW) in this task.

### 1. Create Oracle Linked Service

- Select Management hub (Toolbox) icon on the left and select 'Linked services' under 'Connections' section.
- Select '+ New' under Linked Services
- Search for 'Oracle' under 'Data Store' and select 'Oracle Database' and click on 'Continue'



Enter the following Oracle Connect Info:

- Name: Enter 'OracleDB12cHR'
- Leave the default 'Connection string' option
- Host: Enter Instructor provided server IP address – 40.84.58.15
- Port: 1521 (Default Oracle Port)
- Connection type: Select 'Oracle Service Name'
- Service name: 'nonpdb'
- Username: hr
- Password: hr2



## New linked service (Oracle)

Name \*

OracleDB12cHR

Description

Connect via integration runtime \*

AutoResolveIntegrationRuntime



Connection string

Azure Key Vault

Host \*

Port

1521

Connection type

Oracle Service Name



Service name \*

pdb1

User name \*

hr

Password

Azure Key Vault

Password \*

..

Create

Back



Connection successful



Test connection

Cancel

- Select 'test connection' to verify the successful connection
- Click on 'Create' button to create the Oracle linked service.

## 2. Create Azure Data Lake Storage (ADLS) Gen2 Linked Service

- Select 'New' under Linked Services
- Search for 'Gen2' under data store and select Azure Data Lake Storage Gen2 Click on 'Continue' and enter the following information:
- Name: Enter 'Gen2HRStorage'

- Authentication Method: Leave the default 'Account Key' selection
- Account Selection method: Leave the default 'From Azure subscription'
- Storage account name: select 'ataadflabstorage<YourName>'
- Click on 'Test connection' Click on 'Create' after successful connection to create the ADLS Gen2 storage linked service
- Click Create button

## New linked service (Azure Data Lake Storage Gen2)

Hosted integration runtime is higher than version 4.0 if connecting via self-hosted integration runtime.

**Name \***

Gen2HRStorage

Description

**Connect via integration runtime \***

AutoResolveIntegrationRuntime

**Authentication method**

Account key

**Account selection method**

☒ From Azure subscription ☐ Enter manually

Azure subscription

**Storage account name \***

ataadflabstoragesrini

**Test connection**

☒ To linked service ☐ To file path

**Annotations**

+ New

Create

Back

✓ Connection successful

Test connection

Cancel

### 3. Create Synapse Analytics Linked Service

- Select 'New' under Linked Services
- Search for 'Synapse' under data store and select 'Azure Synapse Analytics (formerly SQL DW)'
- Click on 'Continue'

Enter the following info:

- Name: Enter 'SynapseDBHR'
- Server name: select 'ata-adf-lab-sql-<yourname>'
- Database name: select 'ataadflabsqldb'
- Username: enter 'azureadmin'
- Password: enter the default 'ataadf123!'
- Click on 'test connect' to test the connection.

Click on 'Create' after the successful connection to create the Synapse linked service

## New linked service (Azure Synapse Analytics (formerly SQL DW))

Name \*

SynapseDBHR

Description

Connect via integration runtime \*

AutoResolveIntegrationRuntime

Connection string

Azure Key Vault

Account selection method

☒ From Azure subscription ☐ Enter manually

Azure subscription

Server name

ata-adf-lab-sql-srini

Database name

ataadflabsqldb

Authentication type \*

SQL authentication

User name \*

azureadmin

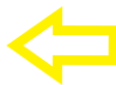
Password

Azure Key Vault

Password \*

Create

Back



Test connection

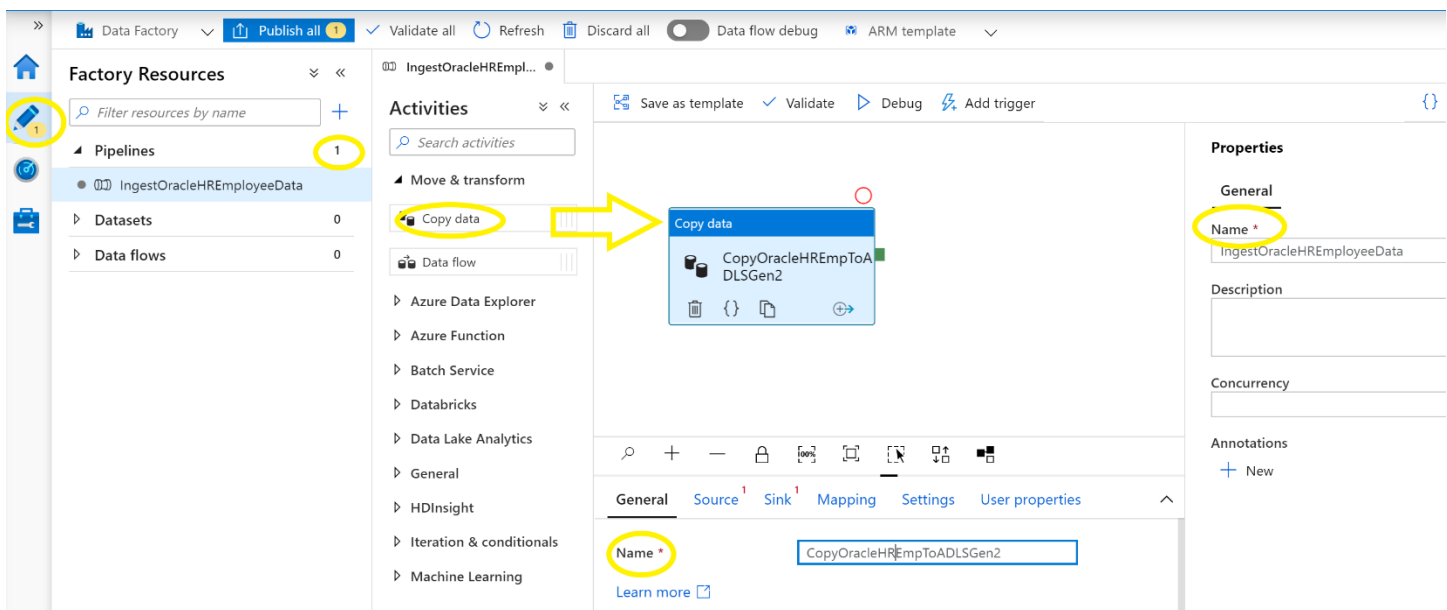
Cancel

You have successfully created connection linked services to Oracle, ADLS Gen2 and Synapse SQL Pool.

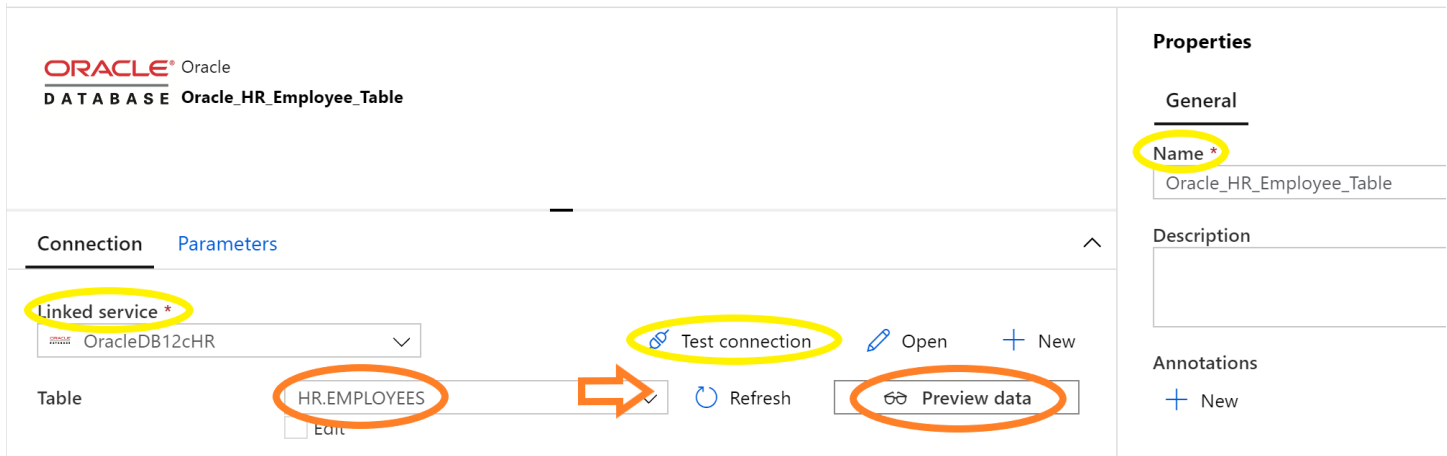
## Task-3: Create Copy Activity to ingest Oracle HR Employee data to Azure Storage

We have established the connection services to the source Oracle DB and the sink Azure Data Lake Storage, you will create a copy activity to ingest the data from Oracle to Azure Data Lake Storage in this task.

1. Select 'Pencil' icon on the left and select three dots next to pipelines to select 'new pipeline' action.
  - Name the pipeline as 'IngestOracleHREmployeeData' under the properties section on the right side.
  - Drag the 'Copy data' from 'Move & transform' section under 'Activities' list to the canvas in the middle of the screen.
  - Name the copy activity as 'CopyOracleHREmpToGen2' below the canvas under the 'General' tab



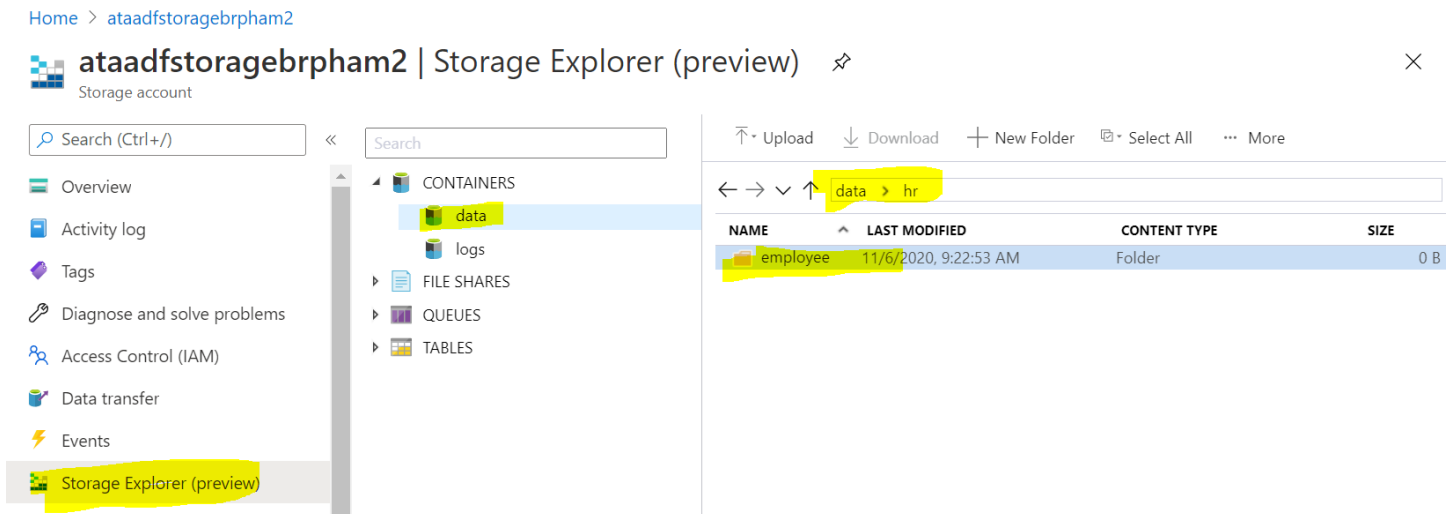
2. Select 'Source' tab next to 'General' to define the source system.
  - Select 'New' to create a new source dataset
  - Select 'Oracle database' as the data store after filtering with 'Oracle' and click on 'Continue'
  - Select 'Open' to define the source dataset
  - Name the dataset as 'Oracle\_HR\_Employee\_table' under the properties section on the right side.
  - Select the 'OracleDB12cHR' linked service and click on 'Test connection' to test the connectivity.
  - Filter table list by typing 'hr.emp' and select 'HR.EMPLOYEES' table and select 'Preview data'.



Make sure you are able to see the employee folder

You will have to enter the file system and directory path

- First, we need to create the directory path in the Azure storage account. Leave this browser tab as is and switch to the Azure Services tab in the browser.
- Access storage account and open up the 'storage explorer' to create 'hr' under 'data' file system and create 'employee' as a subfolder under 'hr' folder.**



3. Select 'Sink' tab next to 'Source' to define the sink system

- Select 'New' to create a new sink data set
- Select 'Azure Data Lake Storage Gen2' as the data store after filtering with 'Gen2' and click on 'Continue'

- Select 'Open' to define the sink dataset
- Select 'Delimited Text' as the format and click on 'Continue'
- Name the dataset as 'Gen2HREmpData'
- Select the 'Gen2HRStorage' linked service.
- Enter File path below

## Set properties

### Name

Gen2HREmpData

### Linked service \*

Gen2HRStorage

### File path

data

/ hr/employee

/ File

First row as header ☐

### Import schema

☒ From connection/store ☐ From sample file ☐ None

▶ Advanced

OK

Back

Cancel

- Select 'Gen2HREmpData' under Datasets section on the left
- Check the 'First row as header' checkbox

**Factory Resources**

Filter resources by name

**Pipelines** 1

- IngestOracleHREmployeeData

**Datasets** 2

- Gen2HREmpData**
- Oracle\_HR\_Employee\_Table

**Data flows** 0

**Global parameters**

**Connections**

**Gen2HREmpData**

**DelimitedText**

**Gen2HREmpData**

**Connection** **Schema** **Parameters**

Linked service \*  
Gen2HRStorage

File path \*  
data / hr/Employee / File

Compression type  
none

Column delimiter  
Comma (,)

Row delimiter  
Auto detect (\r\n, or \n)

Encoding  
Default(UTF-8)

Escape character  
Backslash (\)

Quote character  
Double quote (")

First row as header ☒

Null value

**Properties**

**General**

Name \*  
Gen2HREmpData

Description

Annotations  
+ New

## 5. Test the copy activity

- Select the 'IngestOracleHREmployeeData' under the Pipelines section
- Click on 'Debug' option just above the canvas.
- It will start the process and put it in the queue. Wait till it finishes.
- Check for the status and make sure it is successful

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' pane lists 'Pipelines' and 'Datasets'. The 'IngestOracleHREmployeeData' pipeline is selected. The main canvas shows the pipeline design with a 'Copy data' activity named 'CopyOracleHREmpToGen2'. Below the canvas, the 'Output' tab displays the pipeline run details.

Name	Type	Run start	Duration	Status	Integration runtime	Run ID
CopyOracleHREmpToGen2	Copy	2020-11-06T15:41:44.785	00:00:06	Succeeded	DefaultIntegrationRuntime (East US)	f3a3d

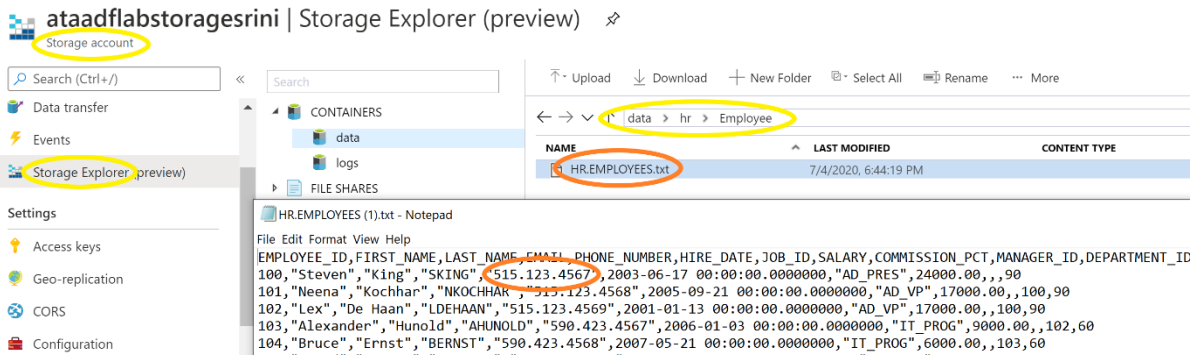
## 6. Verify the data ingestion in ADLS Gen2 storage.

Switch to Azure services tab and access the storage account.

- Open up the 'Storage Explorer' and access the data file system and drill down to 'hr' and 'employee' folder.
- Confirm the 'HR.EMPLOYEES.txt' file. Double click on the file to download and view the data.

You can see how the phone numbers are ingested as text. This is PII data and we should protect this data.



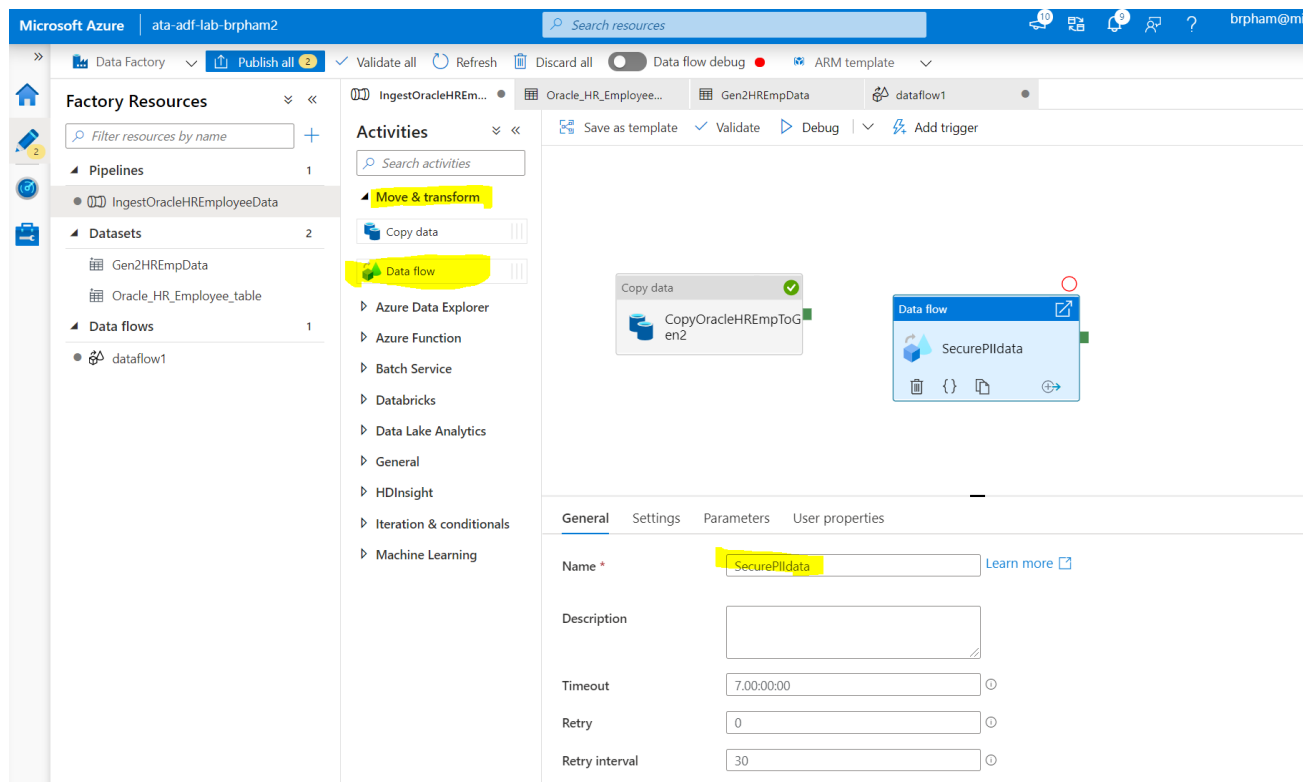


## Task-4: Create Data flow to transform Employee data and store it in Synapse SQL Analytics

We have noticed the PII data we just ingested into ADLS Gen2 storage. You will secure the PII data using the Data flow functionality in this task.

### 1. Create Source data

- Drag and drop the data flow activity into the canvas from the 'Move & Transform' section.
- Select 'create new data flow' and select 'Data Flow' option and click on 'OK'.



## Adding data flow

☐ Use existing data flow ☒ Create new data flow



### Data flow

Code free data transformation at scale



### Wrangling Data Flow (Preview)

Code free data preparation at scale

OK

Cancel

- Select "Source" data
- Name the data flow as 'SecurePIIData'.

Provide the following information under 'Source settings' tab:

- Output stream name: Gen2HrEmpDataOut
- Source type - Select 'Dataset'
- Dataset - Select 'Gen2HREmpData'

Validate

Gen2HREmpDataOut  
Columns: 0 total

**Properties**  
**General**  
**Name \***  
SecurePIIData  
Description

**Source settings** Source options Projection Optimize Inspect Data preview Description

**Output stream name \*** Gen2HREmpDataOut [Learn more](#)

**Source type \*** Dataset

**Dataset \*** Gen2HREmpData

Test connection Open New

**Options**

- ☒ Allow schema drift ⓘ
- ☐ Infer drifted column types ⓘ
- ☐ Validate schema ⓘ

**Skip line count**

**Sampling \*** ☐ Enable ☒ Disable ⓘ

- To preview data, we need to turn on the dataflow debug option.
- Click on 'Import projections' in the 'Projection' tab to view all the columns. The 'Data flow debug' must be enabled for the 'Import projection' button to be active

Validate all Refresh Discard all **Data flow debug** ARM template

Validate Debug Settings

Gen2HREmpDataOut  
Columns: 11 total

**Source settings** Source options **Projection** Optimize Inspect Data preview Description

Define default format Detect data type **Import projection** Reset schema

Column name	Type	Format
EMPLOYEE_ID	12s short	Specify format
FIRST_NAME	abc string	Specify format
LAST_NAME	abc string	Specify format
EMAIL	abc string	Specify format
PHONE_NUMBER	abc string	Specify format
HIRE_DATE	abc string	Specify format
JOB_ID	abc string	Specify format
SALARY	1.2 double	Specify format
COMMISSION_PCT	1.2 double	Specify format
MANAGER_ID	12s short	Specify format
DEPARTMENT_ID	12s short	Specify format

2. Create a data flow step to hash the first 6 digits of the phone number.

- Add another data flow step to secure the phone number by selecting '+' sign and select 'Derived Column' under 'Schema modifier' section.

The screenshot shows the 'Gen2HREmpData' stream with 11 columns. A context menu is open, showing the 'Derived Column' option under the 'Schema modifier' section. The data preview table is visible, showing columns: LAST\_NAME, EMAIL, and PHONE\_NUMBER. The 'PHONE\_NUMBER' column is highlighted.

LAST_NAME	EMAIL	PHONE_NUMBER
King	SKING	515.123.4567
Kochhar	NKOCHHAR	515.123.4568
De Haan	LDEHAAN	515.123.4569
Hunold	AHUNOLD	590.423.4567
Ernst	BERNST	590.423.4568
Austin	DAUSTIN	590.423.4569
Valli	VPATABAL	590.423.4560

- Name the data flow as 'HREmpSecurePhoneNumber'
- Select 'PHONE\_NUMBER' column from the list
- Click on 'Enter Expression' textbox
- Click on 'Open expression builder'

The screenshot shows the 'Derived column's settings' pane for the 'HREmpSecurePhoneNumber' step. The 'Output stream name' is 'HREmpSecurePhoneNumber', the 'Incoming stream' is 'Gen2HREmpData', and the 'Columns' list contains 'PHONE\_NUMBER'. The 'Enter expression...' field is highlighted with a yellow arrow.

Write RegExpression to replace the digits before the '.' with '#' as shown in the picture:

```
regexReplace({PHONE_NUMBER}, '\\d+\\. ', '###.')
```

- Click on 'Refresh' to verify the output of the function.

- Click 'Save and finish' button after you satisfy the result

Visual expression builder

Expression reference documentation [🔗](#) [🔗](#) [✕](#)

OUTPUT SCHEMA << FUNCTIONS << EXPRESSION FOR FIELD "PHONE\_NUMBER" [Save](#)

ANY PHONE\_NUMBER

Filter...

regexReplace({PHONE\_NUMBER}, '\\d+\\. ', '###.')

< All Functions Input schema >

123 EMPLOYEE\_ID  
abc FIRST\_NAME  
abc LAST\_NAME  
abc EMAIL  
abc PHONE\_NUMBER  
abc HIRE\_DATE

+ - \* / || && ! ^ == === <=> != > < >= <= []

Data preview [Refresh](#)

Output: PHONE\_NUMBER abc

PHONE_NUMBER abc
###.###.4567
###.###.4568
###.###.4569

[Save and finish](#) [Cancel](#) [Clear contents](#)

3. Create a data flow step to store the transformed data into Synapse SQL Pool.

We are now ready to output the transformed data into Synapse SQL Pool

- Select '+' and select 'Sink' in the 'Destination' section. Last one in the list.
- Enter 'sinkToSynapseDb' for Output stream name
- Click the '+ New' next to Dataset to create new Azure Synapse Analytics (formerly SQL DW) dataset

IngestOracleHREmp...SecurePIIData

✓ Validate

Debug Settings

Gen2HrEmpDataOut

Import data from Gen2HREmpData

+

Reference: 1

Columns: 11 total

+

sinkToSynapseDb

Columns: 11 total

Add Source

Sink

Settings

Mapping

Optimize

Inspect

Data preview

Output stream name \*

sinkToSynapseDb

Learn more

Incoming stream \*

HREmpSecurePhoneNumber

Sink type \*

Dataset

Dataset \*

Select...

+ New

Options

☒ Allow schema drift

☐ Validate schema

New dataset

In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. [Learn more](#)

Select a data store

Search

All

Azure

Database

File

Generic protocol

NoSQL

Services and apps

Azure Blob Storage

Azure Cosmos DB (SQL API)

Azure Data Lake Storage Gen1

Azure Data Lake Storage Gen2

Azure SQL Database

Azure SQL Database Managed Instance

Azure Synapse Analytics (formerly SQL DW)

Snowflake

Amazon Marketplace Web Service

- For dataset name, enter 'Synape\_HR\_Employee\_table'
- Select 'SynapseDBHR' Linked service
- Check the 'Create new table' radio button
- Check the 'Edit' and enter schema: dbo and table name: employee
- Click 'OK'

## Set properties

Name

Synape\_HR\_Employee\_table

Linked service \*

SynapseDBHR

☐ Select from existing table ☒ Create new table

Schema and table name

dbo

employee

▸ Advanced

OK

Back

Cancel

- Click the 'Settings' tab
- Check the 'Recreate table' radio button
- Uncheck the 'Enable staging' checkbox

IngestOracleHREmp...SecurePIIDataSynapse\_HR\_Employe...

✓ Validate ⚙ Debug Settings

Gen2HrEmpDataOut

Import data from Gen2HREmpData

+

HREmpSecurePhoneNu...

Creating/updating the columns 'EMPLOYEE\_ID, FIRST\_NAME, LAST\_NAME, EMAIL, PHONE\_NUMBER, HIRE\_DATE,

+

⚙ sinkToSynapseDb

Columns: 11 total

Add Source

SinkSettingsMappingOptimizeInspectData preview ●

📘 We recommend enabling staging to improve performance with Azure Synapse Analytics datasets.

Update method

☒ Allow insert  
☐ Allow delete  
☐ Allow upsert  
☐ Allow update

Table action

☐ None ☒ Recreate table ☐ Truncate table

Enable staging

☐

Batch size

ⓘ

Pre SQL scripts

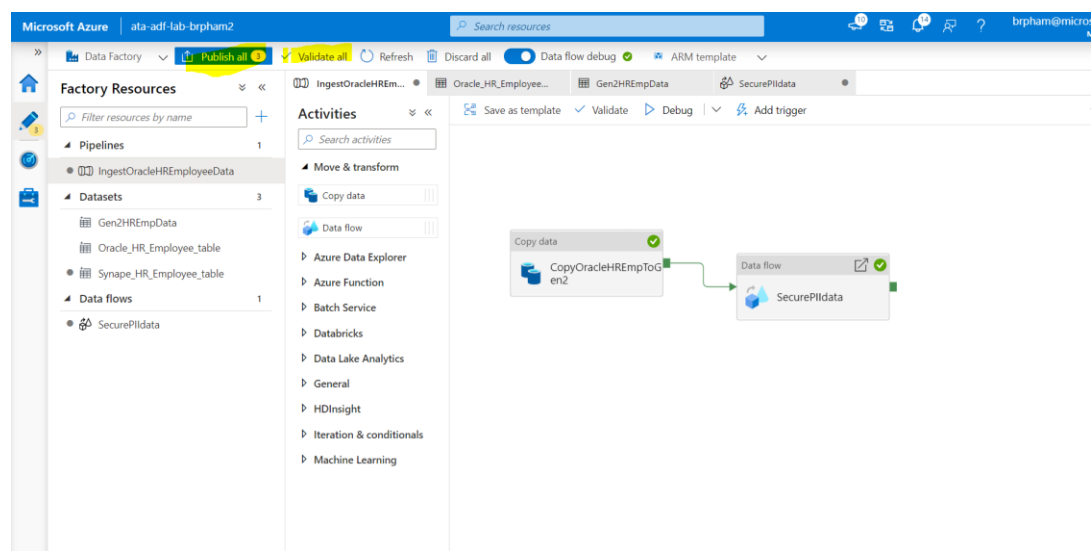
Post SQL scripts



# Task-5: Build a pipeline to connect Copy and Data flow activities and Test the Execution

We have a create copy activity to ingest the data from Oracle to Azure Data Lake Storage and Dataflow activity to secure PII data. You will connect both activities to build a pipeline to execute the end to end functionality.

- 1. Connect the 'Copy data' activity with 'Data Flow' activity
- 2. Click on 'Validate all' to confirm that datasets, dataflow, and pipeline are valid
- 3. Click on 'Publish all' to publish the data factory artifacts to Azure Data Factory service



## Publish all

You are about to publish all pending changes to the live environment. [Learn more](#)

### Pending changes (3)

NAME	CHANGE	EXISTING
Pipelines		
IngestOracleHREmployee...	(Edited)	IngestOracleHREmployeeData
Datasets		
Synape_HR_Employee_table	(New)	-
Data flows		
SecurePIIData	(New)	-

#### 4. Trigger the pipeline execution.

- Click on 'Debug' to test the data transformation to secure the PII data and storing it to Synapse Analytics SQL database.

The screenshot shows the Azure Data Factory interface. On the left, the 'Factory Resources' pane lists 'Pipelines' (1), 'Datasets' (3), and 'Data flows' (1). The 'Activities' pane shows 'Move & transform' activities: 'Copy data' and 'Data flow'. The main canvas displays a pipeline diagram with a 'Copy data' activity (labeled 'CopyOracleHREmpToGen2') connected to a 'Data flow' activity (labeled 'SecurePIIData'). The 'Data flow' activity has a red circle icon in the top right corner.

- Execution will be queued and wait till it finishes.

The screenshot shows the Azure Data Factory interface with the pipeline execution output table. The table has columns: NAME, TYPE, RUN START, DURATION, STATUS, and INTEGRATION RUNTIME. The pipeline run ID is 645cc886-28a6-4fc2-952b-d1b0518492ff. The table shows two activities: 'SecurePIIData' and 'OracleHREmpToGen2', both with a status of 'Succeeded'.

NAME	TYPE	RUN START	DURATION	STATUS	INTEGRATION RUNTIME
SecurePIIData	ExecuteDataFlk	2020-07-05T23:16:28.182	00:01:03	Succeeded	DefaultIntegrationRunti
OracleHREmpToGen2	Copy	2020-07-05T23:16:20.674	00:00:06	Succeeded	DefaultIntegrationRunti

#### 4. Verify the transformed data was inserted into Azure Synapse Analytics SQL Pool.

- Select 'Firewalls and virtual networks' from 'Security' section
- Click the '+ Add client IP'
- Click on 'Save' button after you added the rule.

Home > Resource groups > ata-adf-lab-bpham > ata-adf-lab-sql-bpham

### ata-adf-lab-sql-bpham | Firewalls and virtual networks

SQL server

Search (Ctrl+/)

Save Discard + Add client IP

Deny public network access Yes No

To set Deny Public Network Access, click here to create a new private endpoint.

Minimum TLS Version > 1.0 > 1.1 > 1.2

You are setting the Minimal TLS Version property for all SQL Database and SQL Data Warehouse databases associated with the server. Any login attempts from clients using TLS version less than the Minimal TLS Version shall be rejected.

Connection Policy Default Proxy Redirect

Allow Azure services and resources to access this server Yes No

Connections from the IPs specified below provides access to all the databases in ata-adf-lab-sql-bpham.

Client IP address 68.170.94.159

Rule name	Start IP	End IP
ClientIPAddress_2020-10...	68.170.94.159	68.170.94.159

Virtual networks + Add existing virtual network + Create new virtual network

Rule name	Virtual network	Subnet	Address Range	Endpoint
No vnet rules for this server.				

- Switch to Azure Services and access the Synapse SQL Analytics to verify the employee data with secured phone numbers
- Select the Overview and scroll down to select the 'ataadflabsqldb' database
- Select the 'Query editor (preview)' from the 'Common Tasks' section
- Login with 'azureadmin' and password 'Ataadf123!'
- With successful login, you will see the tables and views.

Home > Resource groups > ata-adf-lab-bpham > ata-adf-lab-sql-bpham > ataadflabsqldb (ata-adf-lab-sql-bpham/ataadflabsqldb)

**ataadflabsqldb (ata-adf-lab-sql-bpham/ataadflabsqldb)** | Query editor (preview)

Synapse SQL pool (data warehouse)

Search (Ctrl+/) << Login + New Query ↑ Open query Feedback

Overview

Activity log

Tags

Diagnose and solve problems

Settings

Workload management

Maintenance schedule

Quick start

Geo-backup policy

Connection strings

Properties

Locks

Security

Auditing

Data Discovery & Classification

Dynamic Data Masking

Firewalls and virtual networks

Security Center

Transparent data encryption

Common Tasks

View streaming jobs

Load Data

**Query editor (preview) ✓**

Build dashboards + reports

Model and cache data

Open in Visual Studio

Welcome to SQL Database Query Editor

SQL server authentication

Login \*

Password \*

OK ✓

Active Directory authentication

Continue as brpham@microsoft.com

OR

- Enter 'select \* from dbo.employee' and click the 'Run' button to view the ingested data from the Oracle database.
- You should see all the employee data with the secured phone number.

ataadflabsqldb (ata-adf-lab-sql-srini/ataadflabsqldb) | Query editor (preview)

Synapse SQL pool (data warehouse)

Search (Ctrl+/) Login + New Query ↑ Open query Feedback

Security

- Advanced data security
- Auditing
- Firewalls and virtual networks
- Transparent data encryption

Common Tasks

- View streaming jobs
- Load Data
- Query editor (preview)
- Build dashboards + reports
- Model and cache data
- Open in Visual Studio

Monitoring

- Query activity
- Alerts
- Metrics

ataadflabsqldb (azureadmin)

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

- Tables
- Views
- Stored Procedures

Query 1

Run Cancel query Save query Export data as Show only Editor

1 select \* from [HR].[Employee]

Results Messages

Search to filter items...

EMPLOYEE_ID	FIRST_NAME	LAST_NAME	EMAIL	PHONE_NUMBER	HIRE_DATE
100	Steven	King	SKING	###,###,4567	2003-06-17 00:00:00
101	Neena	Kochhar	NKOCHHAR	###,###,4568	2005-09-21 00:00:00
102	Lex	De Haan	LDEHAAN	###,###,4569	2001-01-13 00:00:00
103	Alexander	Hunold	AHUNOLD	###,###,4567	2006-01-03 00:00:00

Congratulations!! You have successfully ingested the data from Oracle to Azure Data Lake Storage, transformed the data as per your needs such as securing the data and stored the data into Azure Synapse Analytics for your business analytics needs!!!