

# Unlock IBM Data Insights with Skytap on Azure + Azure Synapse Analytics

A step-by-step guide

**Omer Ahmad**

Microsoft Senior Specialist GBB  
Azure Data & AI Americas

**Madiha Khan**

Microsoft Cloud Solution Architect  
Azure Data & AI

**Ricardo Galvan**

Microsoft Senior Program Manager  
Azure Core Engineering

**Matthew Romero**

Skytap Technical Product Marketing  
Manager

April 2022

# Contents

Background ..... 3

Introduction ..... 3

Overview ..... 4

Detailed Steps ..... 4

Learn more ..... 16

## Background

**Skytap on Azure:** The migration of business-critical applications running on IBM Power on-premises to the cloud is often seen as a difficult and challenging move involving re-platforming. With Skytap on Azure, Microsoft brings the unique capabilities of IBM Power9 servers to Azure, directly integrating with Azure networking enabling Skytap to provide its platform with minimal connectivity latency to Azure native services.

Skytap on Azure is a cloud service purpose-built to natively run traditional IBM Power and x86 workloads in Azure. Skytap on Azure is available for purchase and provisioning directly through the [Azure Marketplace](#). Skytap's services include consumption-based pricing, on-demand access to compute and storage resources, self-service provision and REST APIs for extensibility.

**Azure Synapse Analytics:** Azure Synapse Analytics is a limitless analytics service that brings together data integration, enterprise data warehousing, and big data analytics. It gives you the freedom to query data on your terms, using either serverless or dedicated options—at scale.

Azure Synapse brings these worlds together with a unified experience to ingest, explore, prepare, transform, manage, and serve data for immediate BI and machine learning needs.

## Introduction

This article outlines how to extract data from physical and logical files or a DB2 database stored within your IBM AS/400 libraries hosted on Skytap on Azure using Azure Synapse Analytics DB2 connector ([Copy data from DB2 - Azure Data Factory & Azure Synapse | Microsoft Docs](#)).

This connector can be used to land the AS/400 data in ADLS Gen 2 (Azure Data Lake Storage Gen2) which can open endless possibilities to consume the data. A few examples include:

- using code-free transformations and merging it with data from other sources to gain insights ([Data flows - Azure Synapse Analytics | Microsoft Docs](#))
- sharing it using Azure Data Share ([What is Azure Data Share? | Microsoft Docs](#))
- querying it through the Serverless SQL Pool ([Serverless SQL pool - Azure Synapse Analytics | Microsoft Docs](#))
- creating reports through Power BI ([Data Visualization | Microsoft Power BI](#)),
- creating machine learning prediction models ([What is Azure Machine Learning? - Azure Machine Learning | Microsoft Docs](#))
- utilizing Azure Purview ([Introduction to Azure Purview - Azure Purview | Microsoft Docs](#)) to create a holistic, up-to-date map of your data landscape with automated data discovery, sensitive data classification, and end-to-end data lineage

## Overview

Below are high-level steps for extracting data from Skytap on Azure to ADLS Gen 2:

- Create an Azure Synapse Analytics workspace.
- Create linked service to Skytap on Azure using the DB2 connector.
- Create linked service to ADLS Gen 2 connector.
- Create Synapse Pipeline with copy activity having source as Skytap through DB2 linked service and sink as ADLS Gen 2.

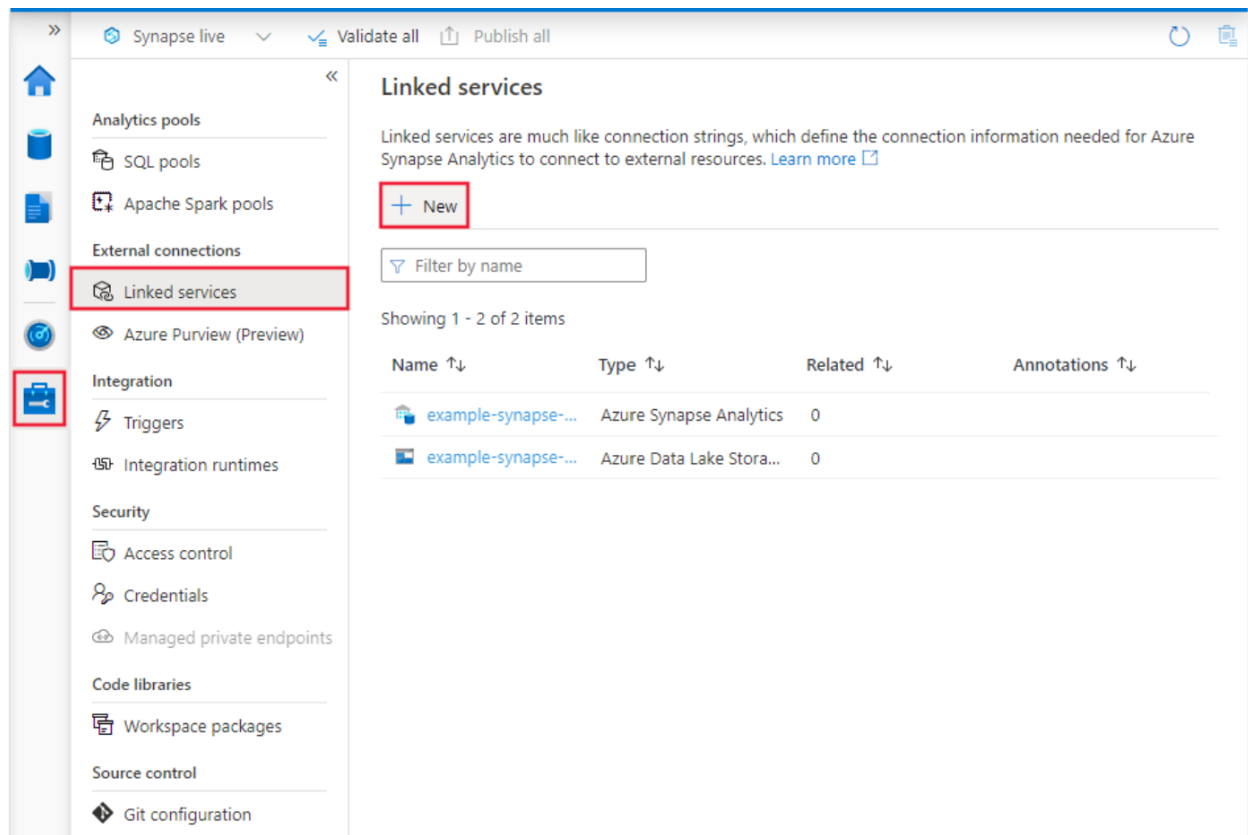
## Detailed Steps

### **Step 1:** Create an Azure Synapse Analytics Workspace

1. Follow steps in this tutorial to create an Azure Synapse Analytics workspace: [Quickstart: create a Synapse workspace - Azure Synapse Analytics | Microsoft Docs](#)

### **Step 2:** Create link service to Skytap on Azure using the DB2 connector.

1. Browse to the Manage tab in your Azure Synapse workspace and select Linked Services, then click New:



2. Search for DB2 and select the DB2 connector.

### New linked service

Data store Compute

db2

All Azure Database File Generic protocol NoSQL Services and apps

**DB2**


DB2

Continue

Cancel

3. Configure the service details, test the connection, and create the new linked service.

## Edit linked service

**DB2** DB2 [Learn more](#) 

Name \*

SkytapDB2

Description

Connect via integration runtime \* 

AutoResolveIntegrationRuntime



Connection string

Azure Key Vault

Server name \* 

Database name \*

Package collection

Certificate common name

Authentication type

Basic



User name \*

Password

Azure Key Vault

Password \*

.....

Additional connection properties


 New


Annotations

 New

Apply

Cancel

 Connection successful

 Test connection

**Step 3:** Create linked service to ADLS Gen 2 connector.

1. Browse to the Manage tab in your Azure Synapse workspace and select Linked Services, then click New:

Synapse Analytics | example-synapse-workspace

Synapse live | Validate all | Publish all

Analytics pools

- SQL pools
- Apache Spark pools

External connections

- Linked services**
- Azure Purview (Preview)

Integration

- Triggers
- Integration runtimes

Security

- Access control
- Credentials
- Managed private endpoints

Code libraries

- Workspace packages

Source control

- Git configuration

### Linked services

Linked services are much like connection strings, which define the connection information needed for Azure Synapse Analytics to connect to external resources. [Learn more](#)

**+ New**

Filter by name

Showing 1 - 2 of 2 items

Name ↑↓	Type ↑↓	Related ↑↓	Annotations ↑↓
example-synapse-...	Azure Synapse Analytics	0	
example-synapse-...	Azure Data Lake Storage	0	


2. Search for Data Lake and select the Azure Data Lake Storage Gen2 connector.

**New linked service**


**Data store**    Compute

data lake


All    Azure    Database    File    Generic protocol    NoSQL    Services and apps




Azure Data Lake Storage  
Gen1



Azure Data Lake Storage  
Gen1 for Cosmos  
Structured Stream



Azure Data Lake Storage  
Gen2



Azure Data Lake Storage  
Gen2 for Cosmos  
Structured Stream



3. Configure the service details, test the connection, and create the new linked service.

## Edit linked service

 Azure Data Lake Storage Gen2 [Learn more](#) 

Name \*

oteststorage

Description

Connect via integration runtime \* ⓘ

 AutoResolveIntegrationRuntime

Authentication method

Account key

Account selection method ⓘ

☐ From Azure subscription ☒ Enter manually

URL \*

https://oteststorage.dfs.core.windows.net

Storage account key

Azure Key Vault

Storage account key \*

\*\*\*\*\*

Test connection ⓘ

☒ To linked service ☐ To file path

Annotations

+ New


> Parameters

> Advanced ⓘ

Apply

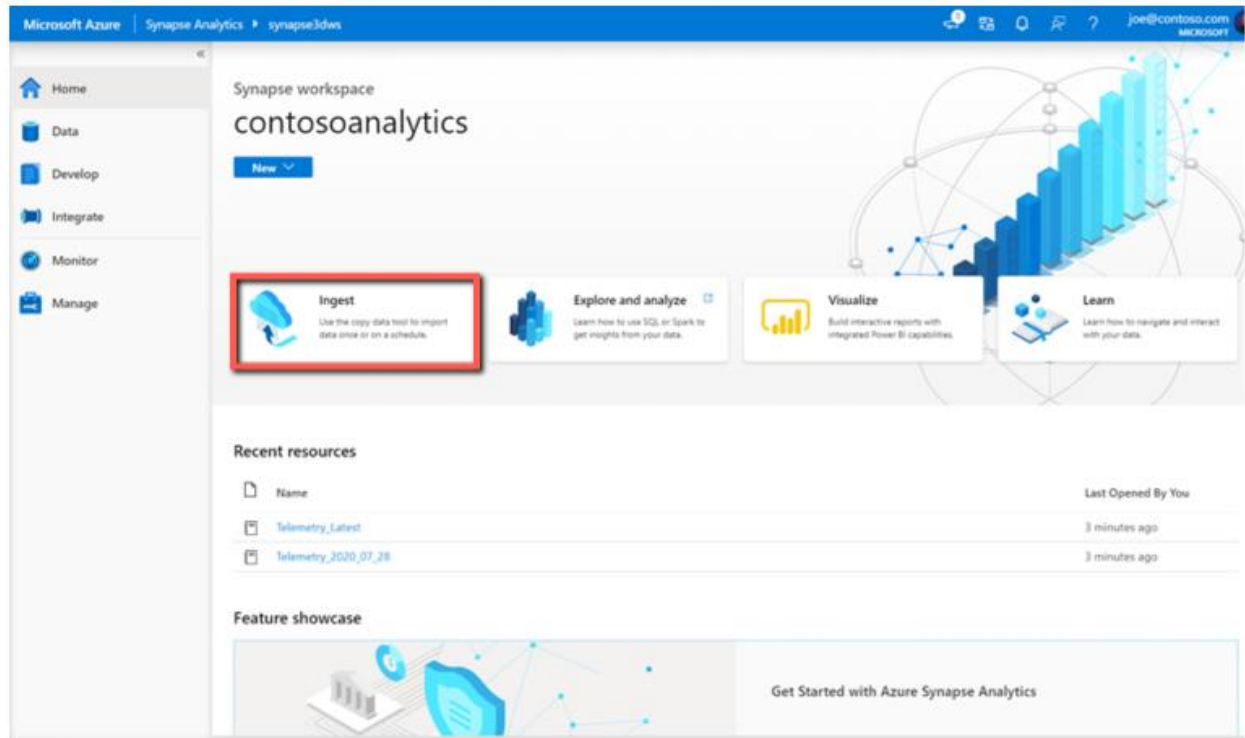
Cancel

 Connection successful

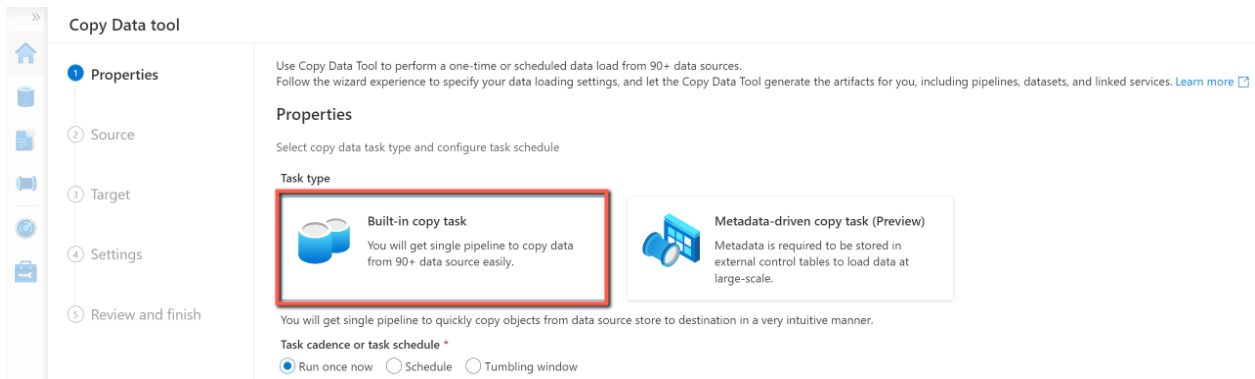
 Test connection

**Step 4:** Create Synapse Pipeline with copy activity having source as Skytap and sink as ADLS Gen 2.

1. Click on the ingest tab.



2. Select “Built-in-copy task” and hit next.



3. Select “Source Type” as DB2 and “Connection” as the name that you provided for your linked service in Step 1. Choose the tables to be copied from the list of existing tables.

Copy Data tool

Properties

Source

Dataset

Configuration

Target

Settings

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: DB2

Connection \*: SkytapDB2

Integration runtime \*: AutoResolveIntegrationRuntime

Source tables: Existing tables

Filter by name...

Show views Refresh

Showing 1886 out of 1886 tables, 0 out of 1068 views (2 selected)

RIGALVAN.CREDIT

RIGALVAN.CUSTOMER

RIGALVAN.EMPAREA

RIGALVAN.EMPLOYEE

RIGALVAN.ORDERS

RIGALVAN.ORDDETAIL

RIGALVAN.PARTNER

RIGALVAN.PRODUCT

RIGALVAN.QCLSRC

RIGALVAN.REGION

RIGALVAN.SHIPPER

RIGALVAN.SUPPLIER

SYSIBM.MQPOLICY\_TABLE

SYSIBM.MQSERVICE\_TABLE

Preview data

4. Select “Preview data” if you would like to see a sample dataset from the selected table.

Copy Data tool

Properties

Source

Dataset

Configuration

Target

Settings

Review and finish

Apply filter

Filter by name...

Advanced

Preview data

RIGALVAN.PARTNER

RIGALVAN.PRODUCT

5. Choose the “Target Type” as Azure Data Lake Storage Gen2, for “Connection” choose the linked service create for ADLS Gen2 in Step 3, provide a folder path to where the tables will be copied within ADLS Gen 2 and also a file name suffix e.g. .txt, .parquet etc.

**Copy Data tool**

Properties

Source

**Target**

Dataset

Configuration

Settings

Review and finish

**Destination data store**

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

Target type: Azure Data Lake Storage Gen2

Connection \*: oteststorage

Integration runtime \*: AutoResolveIntegrationRuntime

**Folder path**

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.

test

**File name**

File name is defined by source table name

> Advanced settings

**File name suffix**

.txt

6. Review the “File format settings” and after choosing the appropriate options, check the “Add header to file” checkbox.

The screenshot shows the 'Copy Data tool' configuration interface. On the left, a vertical sidebar contains icons for home, source, target, dataset, configuration, settings, and review/finish. The main panel is titled 'Copy Data tool' and shows a progress bar with steps: Properties, Source, Target (selected), Dataset, Configuration, Settings, and Review and finish. The 'File format settings' section is active, showing options for File format (DelimitedText), Column delimiter (Comma (,)), Row delimiter (Default (\r\n, or \r\n)), and a checkbox for 'Add header to file' which is checked and highlighted with a red box. Below this is an 'Advanced' link. Other settings include Compression type (None), Max rows per file, and File name prefix.

>>

Copy Data tool

✓ Properties

✓ Source

3 Target

• Dataset

• Configuration

4 Settings

5 Review and finish

File format settings

File format ⓘ  
DelimitedText

Column delimiter  
Comma (,)  
☐ Edit

Row delimiter  
Default (\r\n, or \r\n)  
☐ Edit

☒ Add header to file ⓘ

> Advanced

Compression type  
None

Max rows per file

File name prefix

7. Provide an appropriate "Task Name".

The screenshot shows the 'Copy Data tool' configuration interface, specifically the 'Settings' section. The sidebar on the left is the same as the previous screenshot, but the 'Settings' step is now selected in the progress bar. The 'Settings' section is titled 'Settings' and includes a subtitle 'Enter name and description for the copy data task, more options for data movement'. The 'Task name \*' field is highlighted with a red box and contains the text 'CopyPipeline\_Skytapdemo'. Below this is a 'Task description' text area. Other settings include 'Data consistency verification' (checkbox), 'Fault tolerance' (dropdown), 'Enable logging' (checkbox), and 'Enable staging' (checkbox). An 'Advanced' link is at the bottom.

>>

Copy Data tool

✓ Properties

✓ Source

✓ Target

4 Settings

5 Review and finish

Settings

Enter name and description for the copy data task, more options for data movement

Task name \* CopyPipeline\_Skytapdemo

Task description

Data consistency verification ⓘ ☐

Fault tolerance ⓘ

Enable logging ⓘ ☐

Enable staging ⓘ ☐

> Advanced

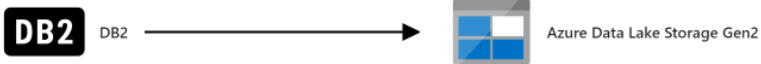
## 8. Review the “Summary” and click next.

>> Copy Data tool

✓ Properties  
✓ Source  
✓ Target  
✓ Settings  
5 Review and finish  
● Review  
○ Deployment

### Summary

You are running pipeline to copy data from DB2 to Azure Data Lake Storage Gen2.



**DB2** → **Azure Data Lake Storage Gen2**

#### Properties

[Edit](#)

Task name	CopyPipeline_Skytapdemo
Task description	

#### Source

[Edit](#)

Connection name	SkytapDB2
Dataset name	SourceDataset_rv8
Number of tables	2

#### Target

[Edit](#)

Connection name	oteststorage
Dataset name	DestinationDataset_rv8
Column delimiter	,
Escape character	\
Quote char	"
First row as header	true

#### Copy settings

[Edit](#)

Timeout	7:00:00:00
Retry	0
Retry interval (sec)	30
Secure output	false

< Previous   Next >

9. Once the deployment is complete, click “Finish” to close the copy data tool. If you want to review the logs for the pipeline run click “Monitor”.

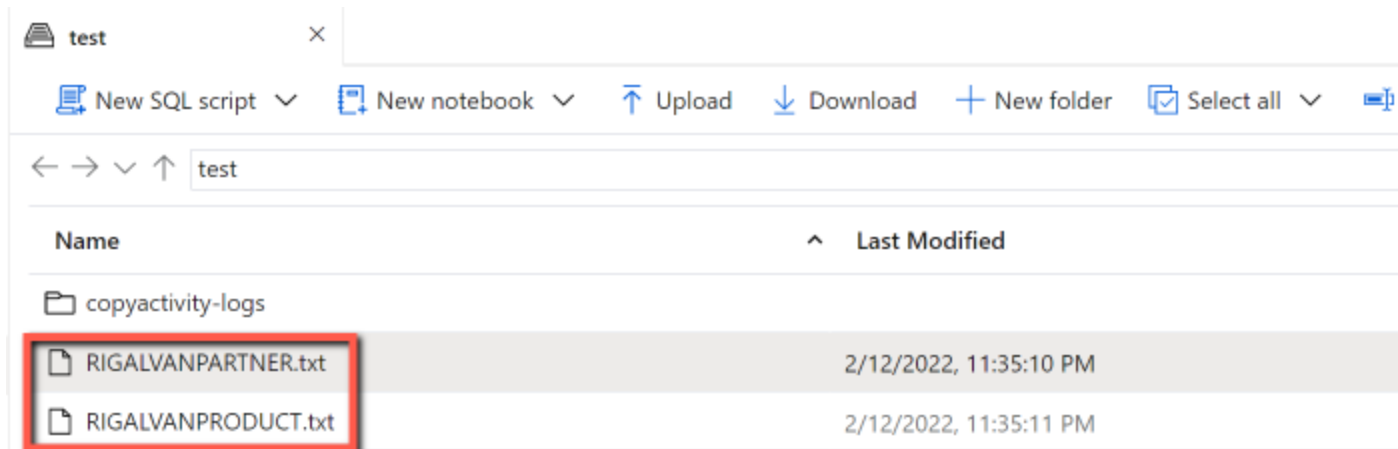
The screenshot displays the 'Copy Data tool' interface. On the left, a vertical sidebar contains a list of steps: Properties, Source, Target, Settings, Review and finish (highlighted with a blue circle and a checkmark), Review, and Deployment. The main area shows a diagram of data flow from 'DB2' to 'Azure Data Lake Storage Gen2'. Below this, the text 'Deployment complete' is displayed. 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 of the interface, three buttons are visible: 'Finish' (highlighted with a red box), 'Edit pipeline', and 'Monitor'.

10. Verify tables have been copied to the desired folder in ADLS Gen 2.



## Learn more

Skytap on Azure is a fast, flexible solution that natively runs IBM Power and x86 workloads in Azure. You can migrate your workloads to Azure without refactoring, rewriting or rearchitecting. Getting started is easy. It takes just a few minutes and pay-as-you-go pricing lets you work with Skytap risk-free. Find Skytap on Azure in the [Azure Marketplace](#). Learn more about Skytap on Azure [here](#).

Azure Synapse Analytics is a limitless analytics service that brings together data integration, enterprise data warehousing, and big data analytics. It gives you the freedom to query data on your terms, using either serverless or dedicated options- at scale. Azure Synapse brings these worlds together with a unified experience to ingest, explore, prepare, transform, manage, and serve data for immediate BI and machine learning needs. Learn more about Azure Synapse Analytics here: [What is Azure Synapse Analytics? - Azure Synapse Analytics | Microsoft Docs](#)