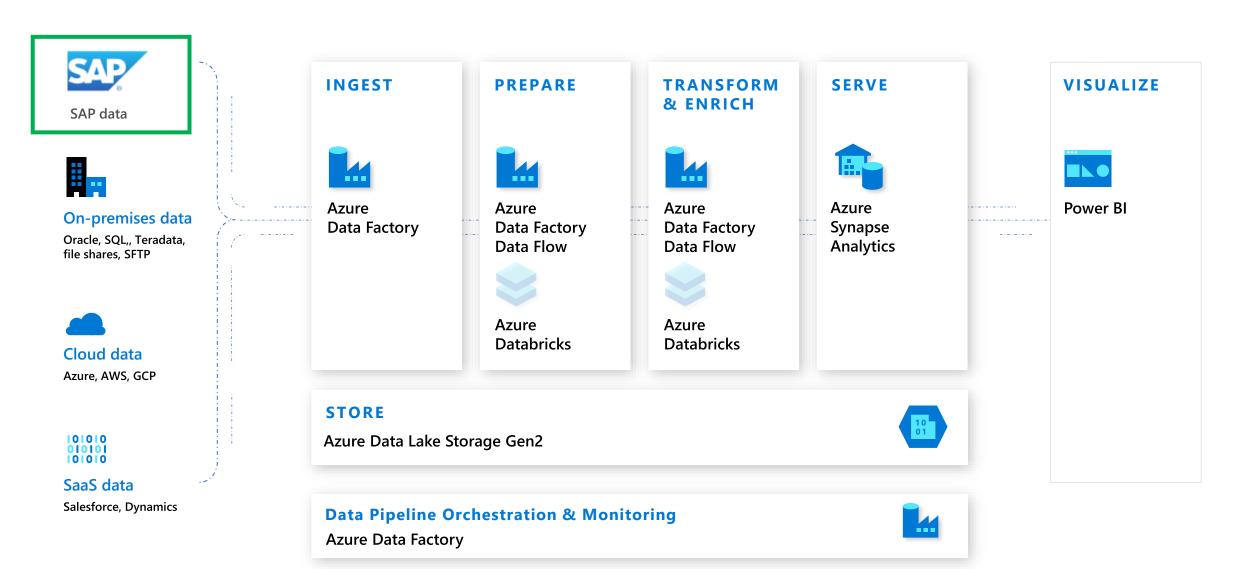


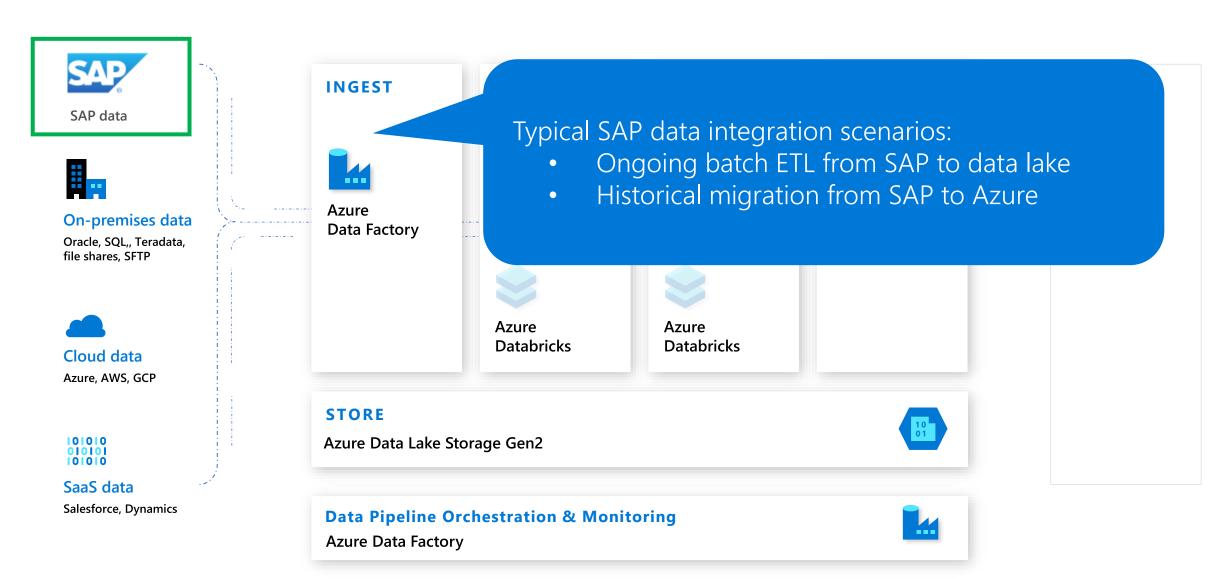
SAP Data Integration Using Azure Data Factory

Update: Jun 28, 2020

Modern Data Warehouse



Modern Data Warehouse





Azure Data Factory

A fully-managed data integration service for cloud-scale analytics in Azure



Connected & Integrated

Rich connectivity

Built-in transformation

Flexible orchestration

Full integration with Azure Data services



Scalable & Cost-Effective

Serverless scalability without infra mgmt

Pay for use



Secure & Compliant

Certified compliance

Enterprise grade security

MSI and AKV support



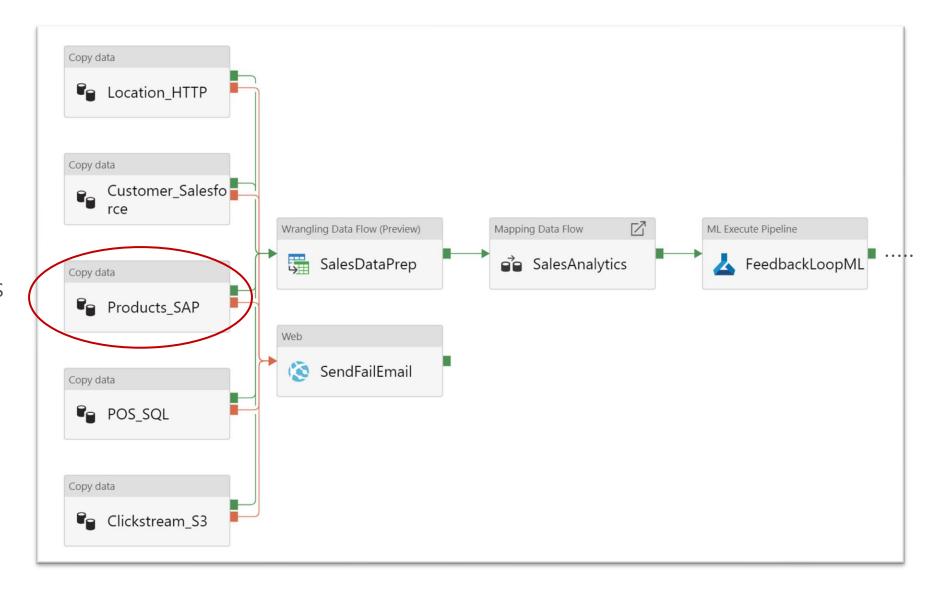
Productive

Drag & drop UI

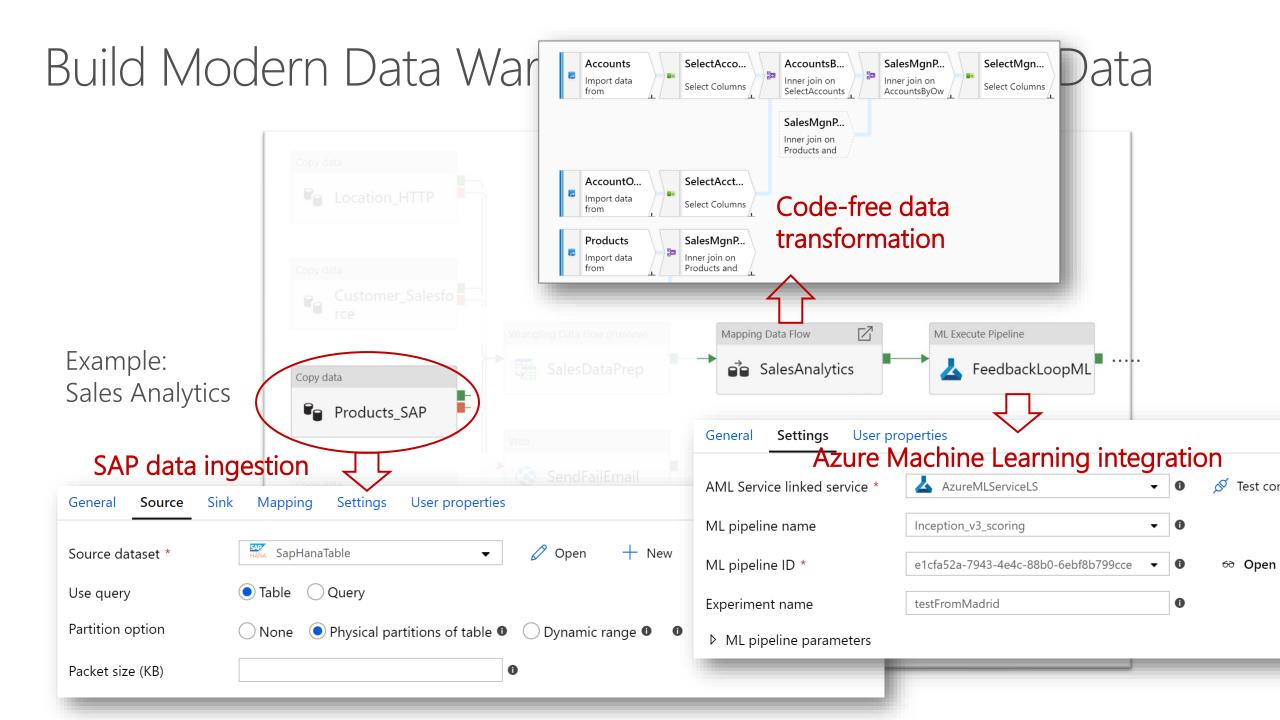
Single-pane-of-glass monitoring

CICD model

Build Modern Data Warehouse Solution w/ SAP Data

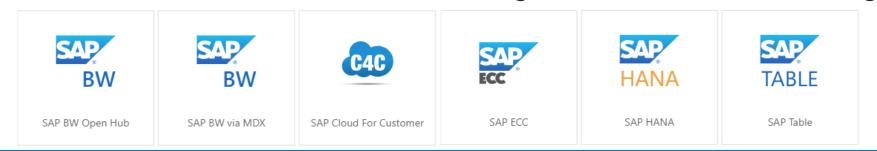


Example: Sales Analytics



Access All Your Data

Single tool to enable data ingestion from SAP as well as other various sources, and data transformation via built-in Data Flow, integration with Databricks/HDInsight/etc.



Azure	Databas	e & DW	File Storage	File Formats	NoSQL	Services	s & Apps	Generic
Blob Storage	Amazon Redshift	Phoenix	Amazon S3	Avro	Cassandra	Amazon MWS	PayPal	HTTP
Cosmos DB – SQL API	DB2	PostgreSQL	File System	Binary	Couchbase	CDS for Apps	QuickBooks	OData
Cosmos DB – MongoDB API	Drill	Presto	FTP	Common Data Model	MongoDB	Concur	Salesforce	ODBC
ADLS Gen1	Google BigQuery	SAP BW Open Hub	Google Cloud Storage	Delimited Text		Dynamics 365	SF Service Cloud	REST
ADLS Gen2	Greenplum	SAP BW MDX	HDFS	Excel		Dynamics AX	SF Marketing Cloud	
Data Explorer	HBase	SAP HANA	SFTP	JSON		Dynamics CRM	SAP C4C	
Database for MariaDB	Hive	SAP Table		ORC		Google AdWords	SAP ECC	
Database for MySQL	Impala	Snowflake		Parquet		HubSpot	ServiceNow	
Database for PostgreSQL	Informix	Spark				Jira	SharePoint List	
File Storage	MariaDB	SQL Server				Magento	Shopify	
SQL Database	Microsoft Access	Sybase				Marketo	Square	
SQL Managed Instance	MySQL	Teradata				Office 365	Web Table	
Synapse Analytics	Netezza	Vertica				Oracle Eloqua	Xero	
Search Index	Oracle					Oracle Responsys	Zoho	
Table Storage						Oracle Service Cloud		

Table of Content

- SAP Data Integration Overview
- SAP HANA Connector
- SAP Table Connector
- SAP BW Open Hub Connector
- SAP ECC Connector
- SAP BW MDX Connector
- More about Azure Data Factory Copy Activity
- Resources

SAP HANA Integration

"I want to extract data from SAP HANA database" >

ADF connector:



(Connector deep-dive)

SAP BW Integration

"I want to extract data from SAP BW" →

		Sugge	sted decision di	rection		_
ADF connector options	SAP Table	TABLE SAP Table	SAP BW Open Hub	BW SAP BW Open Hub	SAP BW via MDX	SAP BW SAP BW via MDX
Objects to extract	` 1	arent, Pooled, e) and View	DSO, InfoCube DataSou		InfoCubes,	QueryCubes
SAP side configuration	N/A		SAP Open Hub Destination		N/A	
Performance	Fast w/ built-in parallel loading based on configurable partitioning		Fast w/ built-in parallel loading based on OHD specific schema		Slower	
Suitable workload	Large volume		Well-thought-through workload Large volume		Exploratory workload Small volume	
	(Connecto	or deep-dive)	(Connecto	r deep-dive)	(<u>Connector</u>	deep-dive)

NOTE: SAP BW4/HANA is not supported now.

SAP ECC, S/4 HANA, SAP Application Integration

"I want to extract data from SAP ECC, S/4 HANA, or other SAP applications" \rightarrow

	Suggested decision direction				
ADF connector options	SAP Table	TABLE SAP Table	SAP ECC	SAP ECC SAP ECC	
Objects to extract	Table (Transparent, Pooled, Cluster Table) and View		· ·	OData entities exposed via SAP Gateway (BAPI, ODP)	
SAP side configuration	N/A		SAP G	SAP Gateway	
Performance	Fast w/ built-in parallel loading		Slo	Slower	
Suitable workload	Large volume		Small volume		
			-	-	

If you push ECC data into SAP HANA/BW, you can also go through SAP HANA/BW connector options.

(Connector deep-dive)

(Connector deep-dive)

SAP HANA Connector

SAP HANA Connector



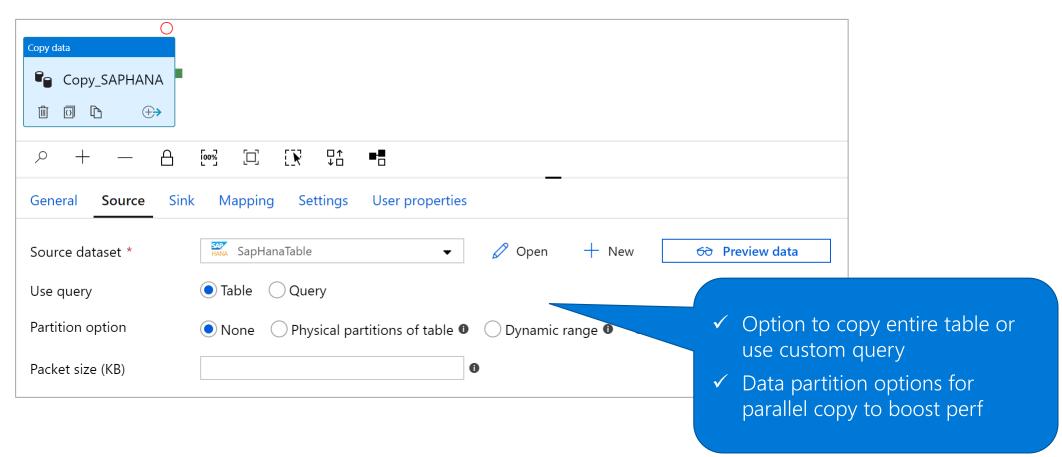
SAP HANA

Suitable scenario: ingest data from SAP HANA.

Supported versions	All SAP HANA versions, on-prem or in the cloud
Supported SAP objects	 HANA Information Models (Analytic/Calculation views) Row & Column Tables
Supported authentications	 Basic – username & password Windows – Single Sign-On via Kerberos-constrained delegation
Mechanism and prerequisites	 Built on top of SAP's HANA ODBC driver Pull data via custom query Run on Self-hosted Integration Runtime
Performance & Scalability	 Built-in parallel loading option based on configurable data partitioning NEW Performant to handle TB level data with hundred millions to billion of rows per run, observed several to several dozens MB/s (varies per customers' data/env.)

SAP HANA Connector





SAP HANA Connector – How It Works



←--- Command and Control → Data **Azure Data Factory Service** Scheduling | Orchestration | Monitoring Pipeline **Azure Data Stores** Azure Outbound On-prem or Port 443 **Azure VNET ADF Self-**SAP **SAP HANA** ODBC hosted **ODBC** HANA Integration Query Driver Runtime **HANA Server** Machine

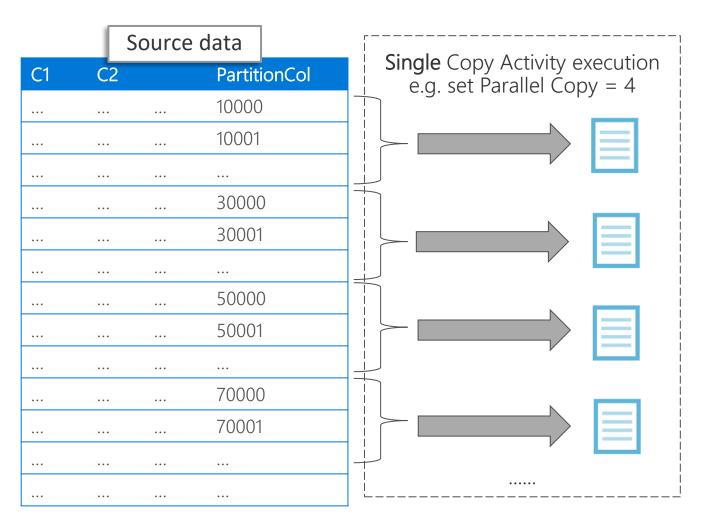
SAP HANA Connector – Built-in Parallel Loading



For each copy activity run, ADF issue the specified query to source to retrieve the data.

Out-of-box optimization for SAP HANA:

- Built-in parallel copy by partitions to boost performance for large table ingestion.
- Options of HANA physical table partition and dynamic range partition.



SAP HANA Connector – Incremental Copy



Pattern I: "my data has timestamp column e.g. last modified time"

<u>Solution:</u> tumbling window trigger + dynamic query with system variables. Get started via Copy Data Tool.

Example: scheduled daily incremental copy starting at midnight

C1	C2		LastModifiedDate
•••	•••	•••	
•••	•••	•••	2019/03/18
• • •	• • •	•••	2019/03/18
• • •	• • •	•••	
• • •	• • •	•••	2019/03/18
•••	•••	• • •	2019/03/19
•••	•••	• • •	2019/03/19
•••	• • •	•••	
•••	•••	•••	2019/03/19
•••	•••	•••	

SELECT * FROM MyTable

WHERE LastModifiedDate >= @{formatDateTime(pipeline().parameters.windowStartTime, 'yyyy/MM/dd')
AND LastModifiedDate < @{formatDateTime(pipeline().parameters.windowEndTime, 'yyyy/MM/dd')

Execution start time: 2019/03/19 00:00:00 (window end time)

Delta extraction: last modified time between 2019/03/18 – 2019/03/19

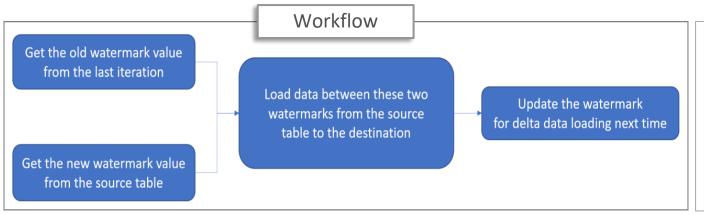
Execution start time: 2019/03/20 00:00:00 (window end time)

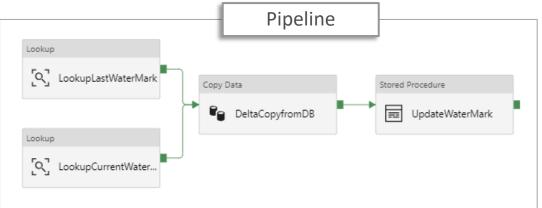
Delta extraction: last modified time between 2019/03/19 – 2019/03/20

SAP HANA Connector – Incremental Copy



<u>Pattern II:</u> "my data has an <u>incremental column</u> e.g. ID" <u>Solution:</u> external control table/file + high watermark. Get started via solution template "Delta copy from Database".





Pattern III: "my data is small in size as dimension data"

Solution: full copy and overwrite

SAP Table Connector

SAP Table Connector

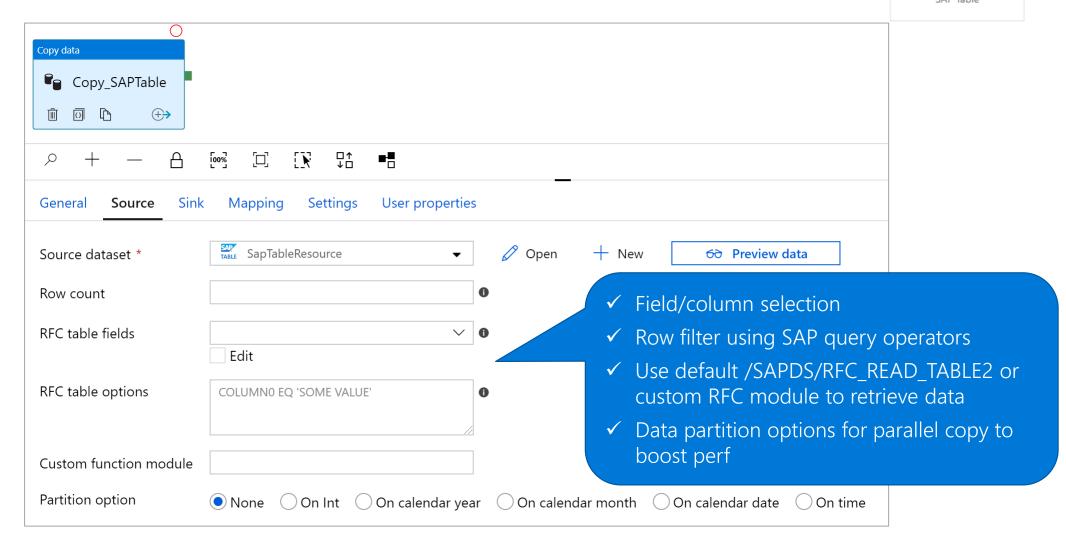


Suitable scenario: ingest data from SAP Table for SAP ECC, S/4 HANA, BW, or other application in Business Suite.

Supported versions	 SAP ECC or other applications in Business Suite version 7.01 and above, on-prem or in the cloud S/4 HANA
Supported SAP objects	 SAP Transparent Table, Pooled Table, Cluster Table and View
Supported server type	Connect to Application Server or Message Server
Supported authentications	 Basic – username & password SNC (Secure Network Communications)
Mechanism and prerequisites	 Built on top of SAP .NET Connector 3.0, pull data via NetWeaver RFC w/ field selection & row filter Run on Self-hosted Integration Runtime
Performance & Scalability	 Built-in parallel loading option based on configurable data partitioning Performant to handle TB level data, with per run dozen millions to billion of rows & observed several to 20s MB/s (varies per customers' data/env.)

SAP Table Connector





SAP Table Connector – How It Works

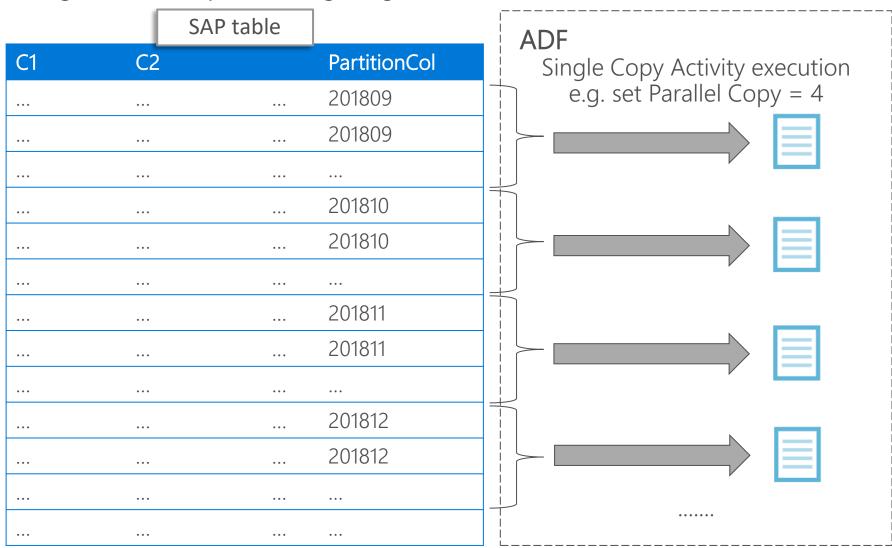


←---→ Command and Control → Data **Azure Data Factory Service** Scheduling | Orchestration | Monitoring Pipeline **Azure Data Stores** Azure Outbound On-prem or Port 443 **Azure VNET** Capabilities: **ADF Self-**✓ Field selection RFC SAP .NET hosted ✓ Row filter (SAP query operators) Connector Integration Runtime ✓ Default or custom RFC func Machine ✓ Built-in partition + parallel load

SAP Table Connector – Built-in Parallel Loading



Configurable data partitioning on given column (INT, Calendar Year/Month/Date) + parallel copies -





Enable partitioning when ingesting large dataset, e.g. dozen millions of rows.

To speed up, choose the proper partition column and partition numbers, and adjust parallel copies.

Learn more

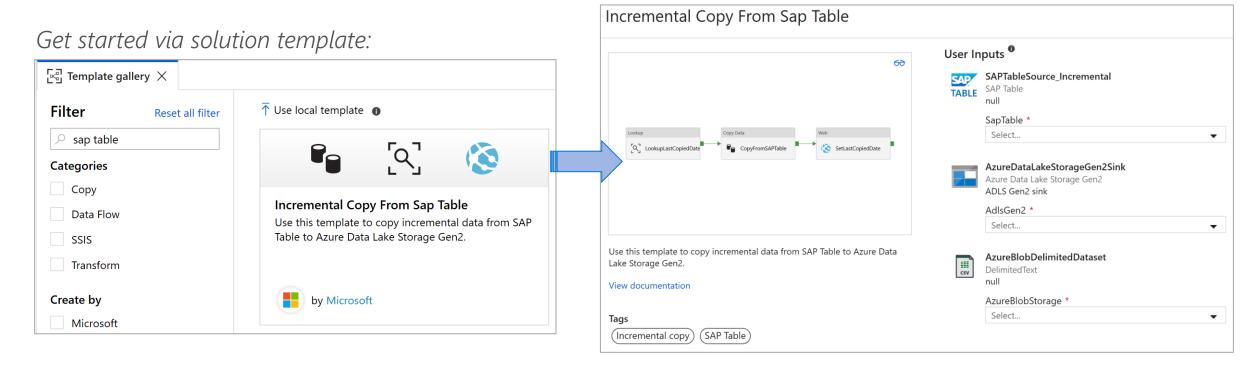
SAP Table Connector – Incremental Copy



Pattern I: "my data has timestamp column e.g. calendar date"

<u>Solution:</u> tumbling window trigger + dynamic query with system variables via SAP table option (filter)

<u>Pattern II:</u> "my data has an incremental column e.g. id/last copied date" <u>Solution:</u> external control table/file + high watermark.



SAP BW Open Hub Connector

SAP BW Open Hub Connector



Suitable scenario: ingest data from SAP BW with targeted/well-thought-through workload.

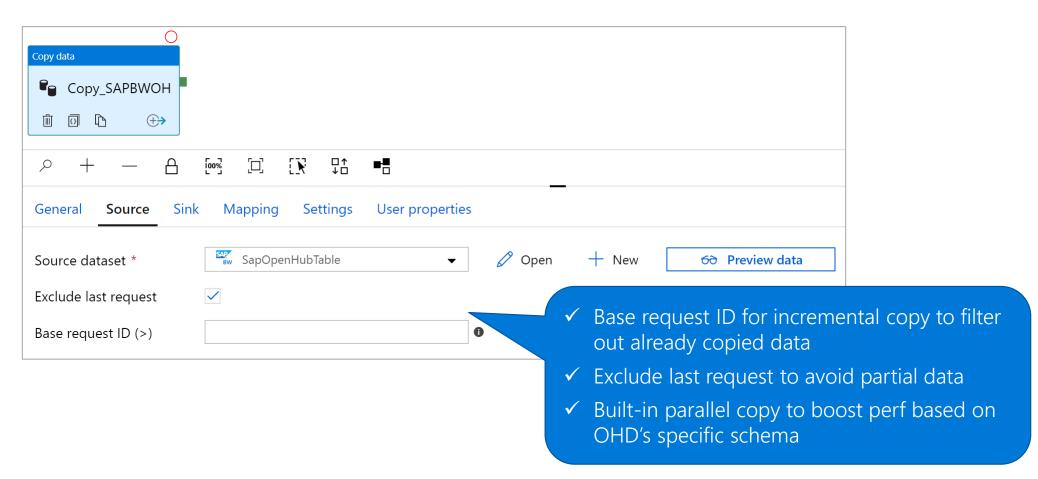
Supported versions	SAP BW version 7.01 and above, on-prem or in the cloud*		
Supported SAP objects	Open Hub Destination (OHD) local table Underneath objects can be DSO, InfoCube, MultiProvider, DataSource etc.		
Supported server type	Connect to Application Server or Message Server NEW		
Supported authentications	Basic – username & password		
Mechanism and prerequisites	 Built on top of SAP .NET Connector 3.0, pull data via NetWeaver RFC Run on ADF Self-hosted Integration Runtime SAP side config: create SAP OHD in SAP BW to expose data 		
Performance & Scalability	 Built-in parallel loading option based on OHD specific schema Performant to handle TB level data, with per run dozens millions to billion of rows & 		

observed several to 20s MB/s (varies per customers' data/env.)

^{*}NOTE: currently SAP BW4/HANA is not supported now. Workaround – flow data to other ADF supported data stores e.g. via Open Hub Destination as ADF supported database.

SAP BW Open Hub Connector





SAP BW Open Hub – How It Works

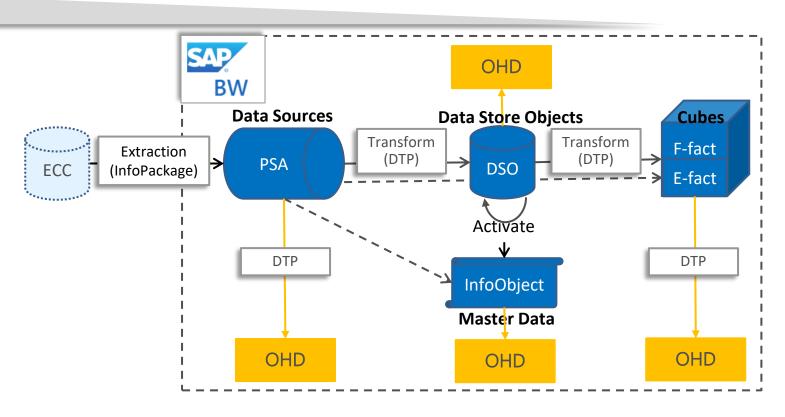




SAP Open Hub Destination (OHD):

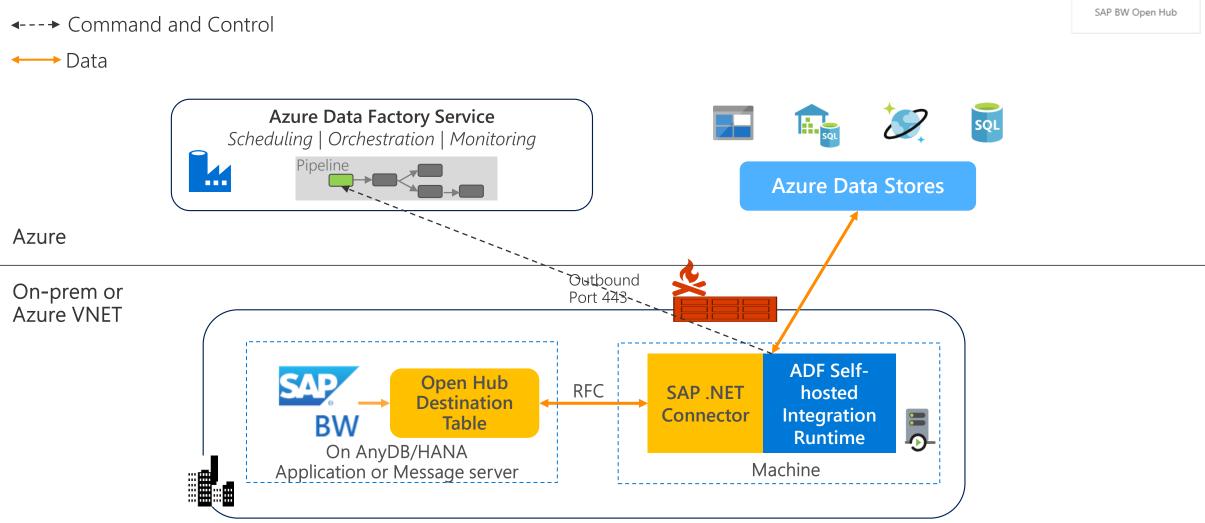
- What is OHD: defines the target to which the data is relayed.
- Supported data: any objects supported by SAP Data Transfer Process (DTP) can be used as open hub data sources.
- OHD types: database tables (local or remote) and flat files.

The connector support OHD local table in BW.



SAP BW Open Hub Connector – How It Works





SAP BW Open Hub Connector – Built-in Parallel Loading







SAP BW Open Hub

<u>Solution:</u> external control table/file + high watermark (max copied request ID).



SAP OHD supports "delta" extraction mode to load incremental data into OHD table.



SAP BW Open Hub

Solution: external control table/file + high watermark (max copied request ID).

Get the last max copied request ID

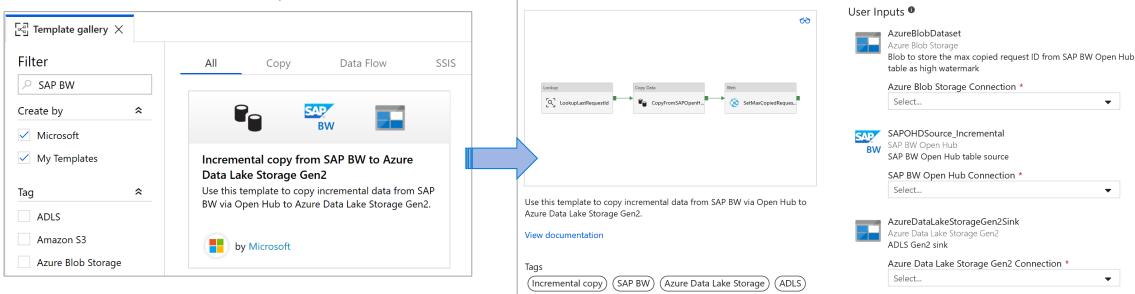
Load data since last copied

request ID from SAP BW

Open Hub to destination

Store the new max copied request
ID returned by Copy activity for incremental data loading next time

Get started via solution template:



Incremental copy from SAP BW to Azure Data Lake Storage Gen2



SAP BW Open Hub

Solution: external control table/file + high watermark (max copied request ID).



• baseRequestId: The ID of request for delta loading. Once it is set, only data with requestId larger than the value of this property will be retrieved.

Copy activity output:





SAP BW Open Hub

Solution: external control table/file + high watermark (max copied request ID).

Load data since last copied Store the new max copied request Get the last max copied ID returned by Copy activity for request ID from SAP BW request ID Open Hub to destination incremental data loading next time

excludeLastRequestId: Whether to exclude the records of the last request. Default value is true.

Request ID	Package ID	Record ID	•••
			•••
100		•••	• • •
	•••	•••	•••
200		•••	•••
300		•••	•••
300		•••	•••

Exclude Last request ID:Applicable if DTP and Copy may run at the same time

Include Last request ID:Applicable if Copy is always invoked after DTP is done

SAP BW Open Hub Connector – Best Practice



SAP BW Open Hub

- SAP BW OHD configurations and how it chains with ADF copy (guidance).
 - Extraction mode full vs delta
 - DTP and ADF scheduling
 - Housekeeping on SAP server

SAP ECC Connector

SAP ECC Connector



Suitable scenario: ingest data from SAP Applications other than SAP Table.

Supported	versions

Supported SAP objects

Supported authentications

Mechanism and prerequisites

- SAP ECC version 7.0 and above
- Any entities exposed by SAP ECC OData services
- Entities exposed by SAP OData services
- BAPI, ODP (DataExtractors/DataSource), etc.
- Basic user name & password
- Though OData + SAP Gateway
- Run on Self-hosted Integration Runtime if SAP in private network
- SAP side config: set up SAP Gateway, activate OData service, and expose entities

SAP ECC Connector – How Connector Works



←---→ Command and Control → Data **Azure Data Factory Service** Scheduling | Orchestration | Monitoring Pipeline **Azure Data Stores** Azure Outbound On-prem or Port 443 **Azure VNET** If your ECC is publicly accessible, you can use managed Azure Integration OData **OData ADF Self-hosted** Runtime instead of Self-hosted SAP **Integration Runtime** Query Gateway Integration Runtime. Tip: per run limit to under 1 Machine million rows

SAP ECC Connector – Incremental Copy



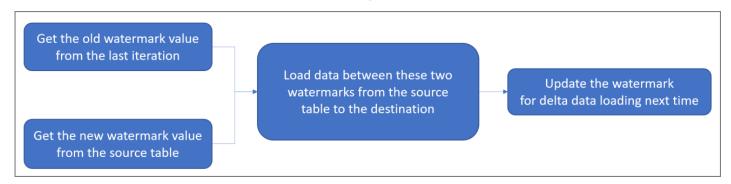
(in general, same as HANA in earlier slides)

Pattern I: "my data has timestamp column e.g. last modified time"

<u>Solution:</u> tumbling window trigger + dynamic query with system variables via OData query

Pattern II: "my data has an incremental column e.g. ID"

Solution: external control table/file + high watermark.



Pattern III: "my data is small in size as dimension data"

Solution: full copy and overwrite

SAP BW via MDX Connector

SAP BW via MDX Connector



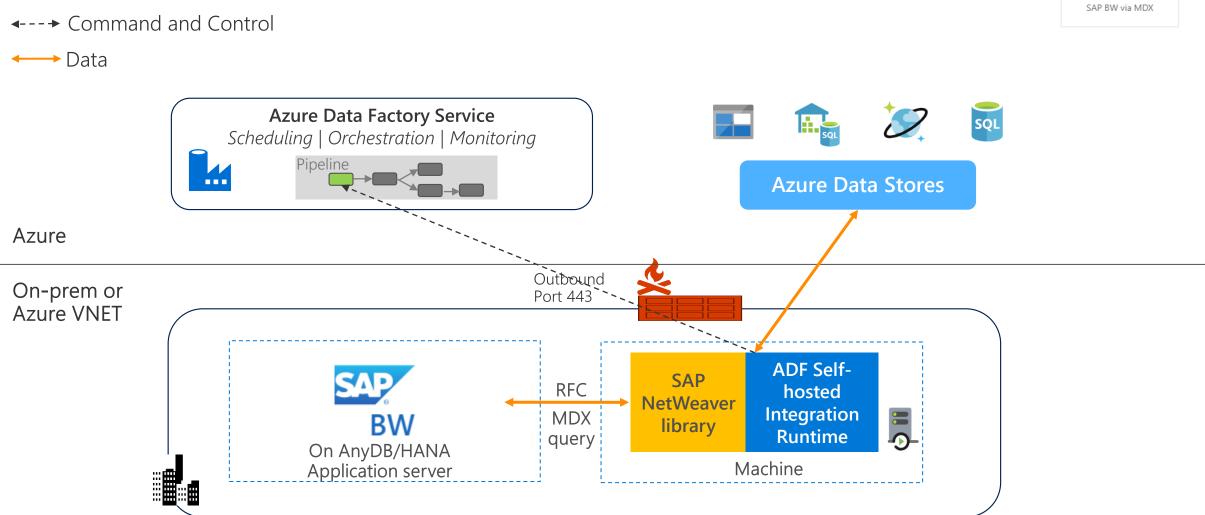
Suitable scenario: ingest data from SAP BW, with exploratory use case.

Supported versions	SAP BW version 7.x, on-prem or in the cloud e.g. on Azure
Supported server type	Connect to Application Server
Supported SAP objects	InfoCubes and QueryCubes (including BEx queries)
Supported authentications	Basic – username & password
Mechanism and prerequisites	 Built on top of SAP NetWeaver library, pull data via RFC Run on Self-hosted Integration Runtime

NOTE: SAP BW4/HANA is not supported now.

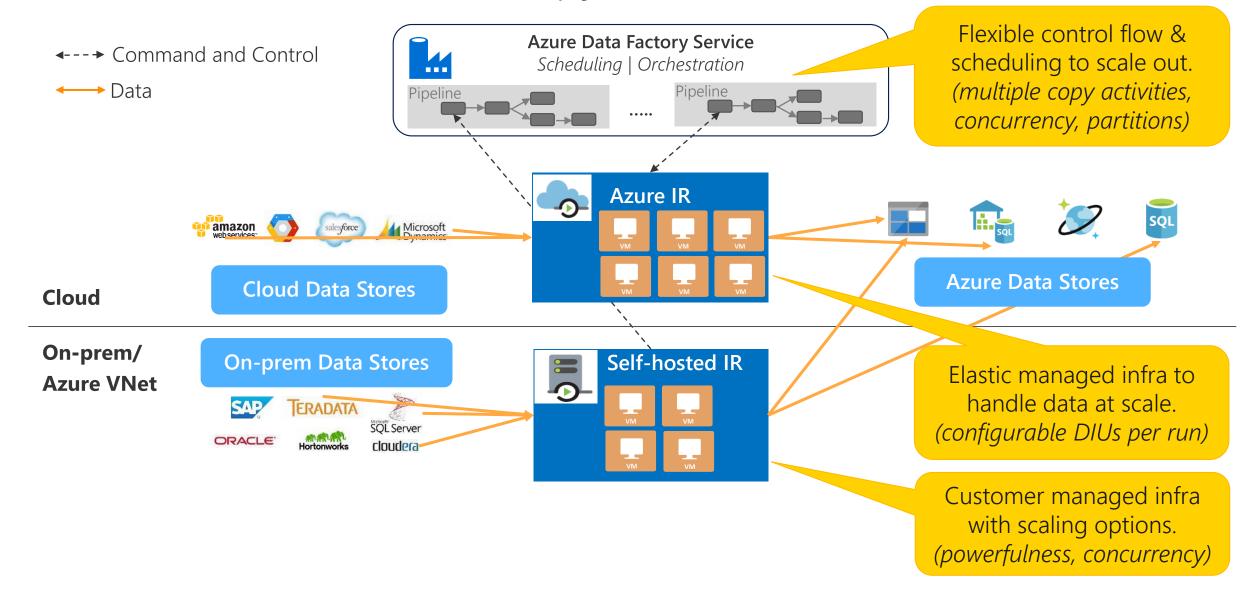
SAP BW via MDX Connector – How It Works





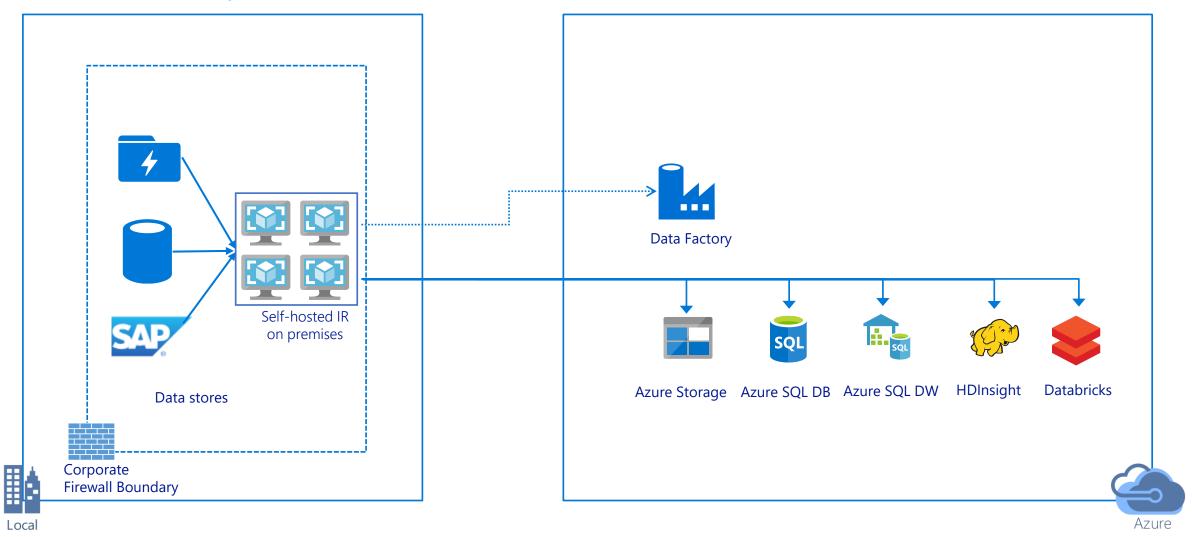
More about Azure Data Factory Copy Activity

Understand How ADF Copy Scales



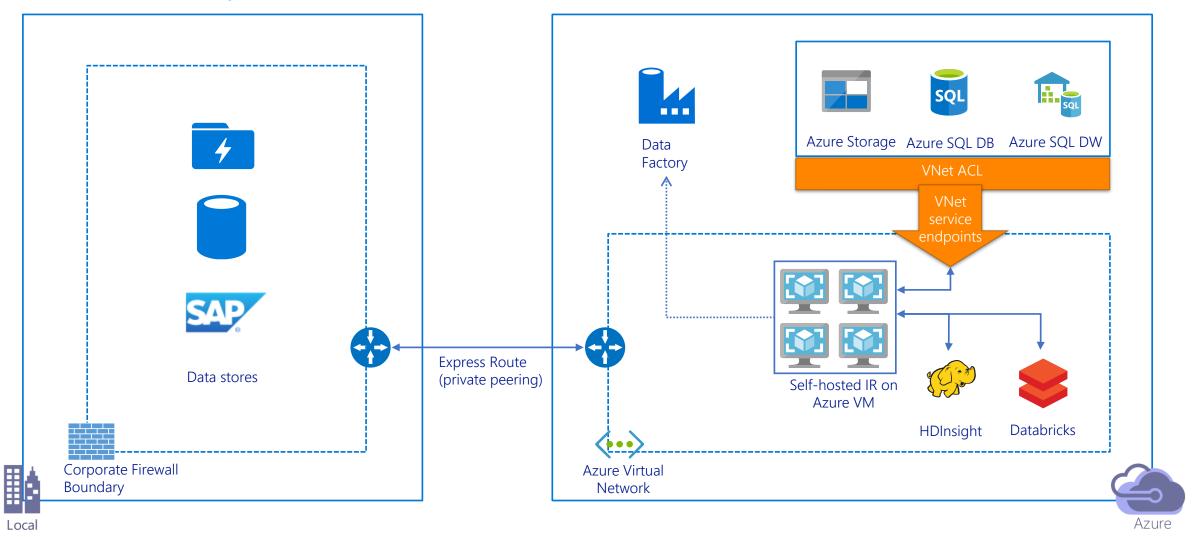
On-prem connected to Azure through public internet

Self-hosted IR deployed on premises



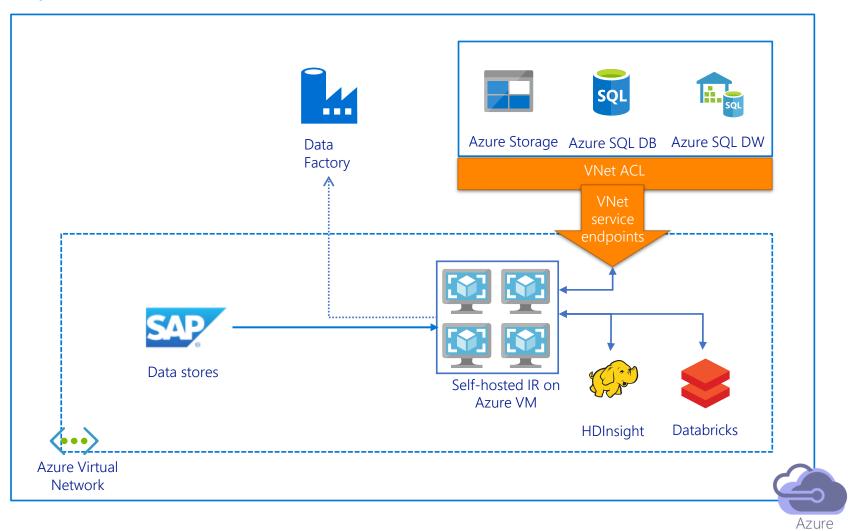
On-prem connected to Azure VNet through ExpressRoute

Self-hosted IR deployed on Azure VM

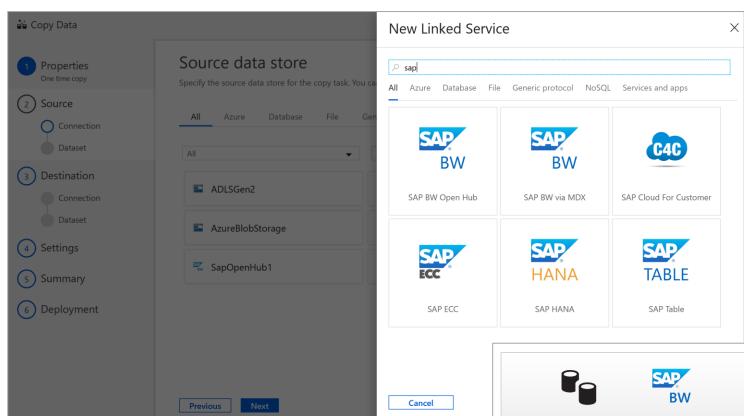


SAP on Azure

Self-hosted IR deployed on Azure VM



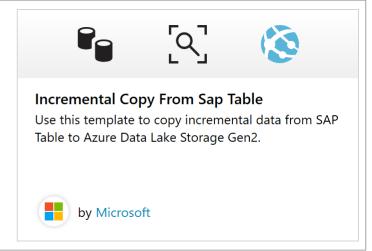
Get Started



Copy Data Tool

Solution Template





Resources

ADF Copy Activity Overview	https://docs.microsoft.com/azure/data-factory/copy-activity-overview
SAP HANA Connector	https://docs.microsoft.com/azure/data-factory/connector-sap-hana
SAP Table Connector	https://docs.microsoft.com/azure/data-factory/connector-sap-table
SAP BW Open Hub Connector	https://docs.microsoft.com/azure/data-factory/connector-sap-business-warehouse-open-hub
SAP BW MDX Connector	https://docs.microsoft.com/azure/data-factory/connector-sap-business-warehouse
SAP ECC Connector	https://docs.microsoft.com/azure/data-factory/connector-sap-ecc
SAP C4C Connector	https://docs.microsoft.com/azure/data-factory/connector-sap-cloud-for-customer
Customer case study	 Analytics and Integration for SAP Global Instance running on-premises with ADF Reckitt Benckiser (RB): https://customers.microsoft.com/story/newell-brands-consumer-goods-azure Newell: https://customers.microsoft.com/story/newell-brands-consumer-goods-azure