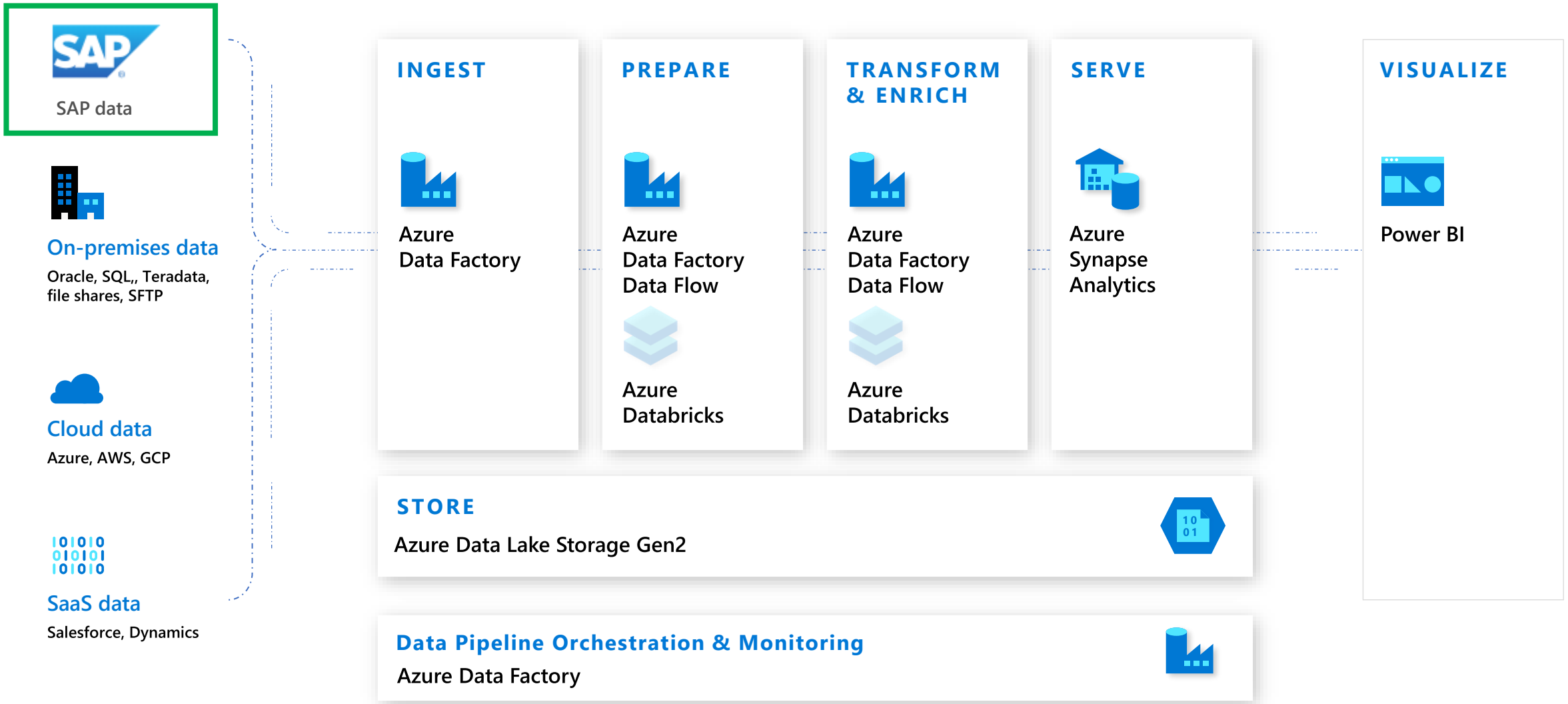
Microsoft

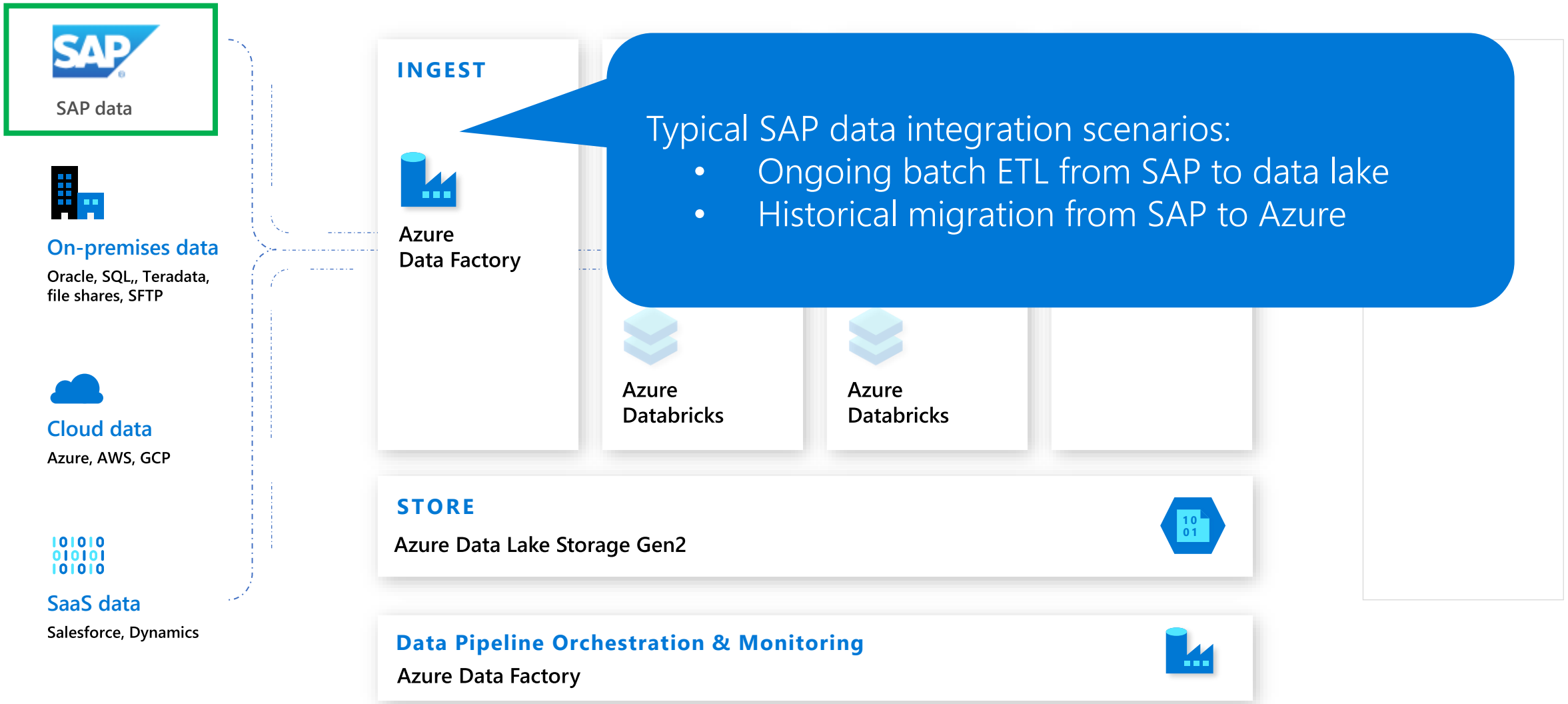
SAP Data Integration Using Azure Data Factory

Update: Jun 28, 2020

Modern Data Warehouse

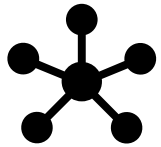


Modern Data Warehouse



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

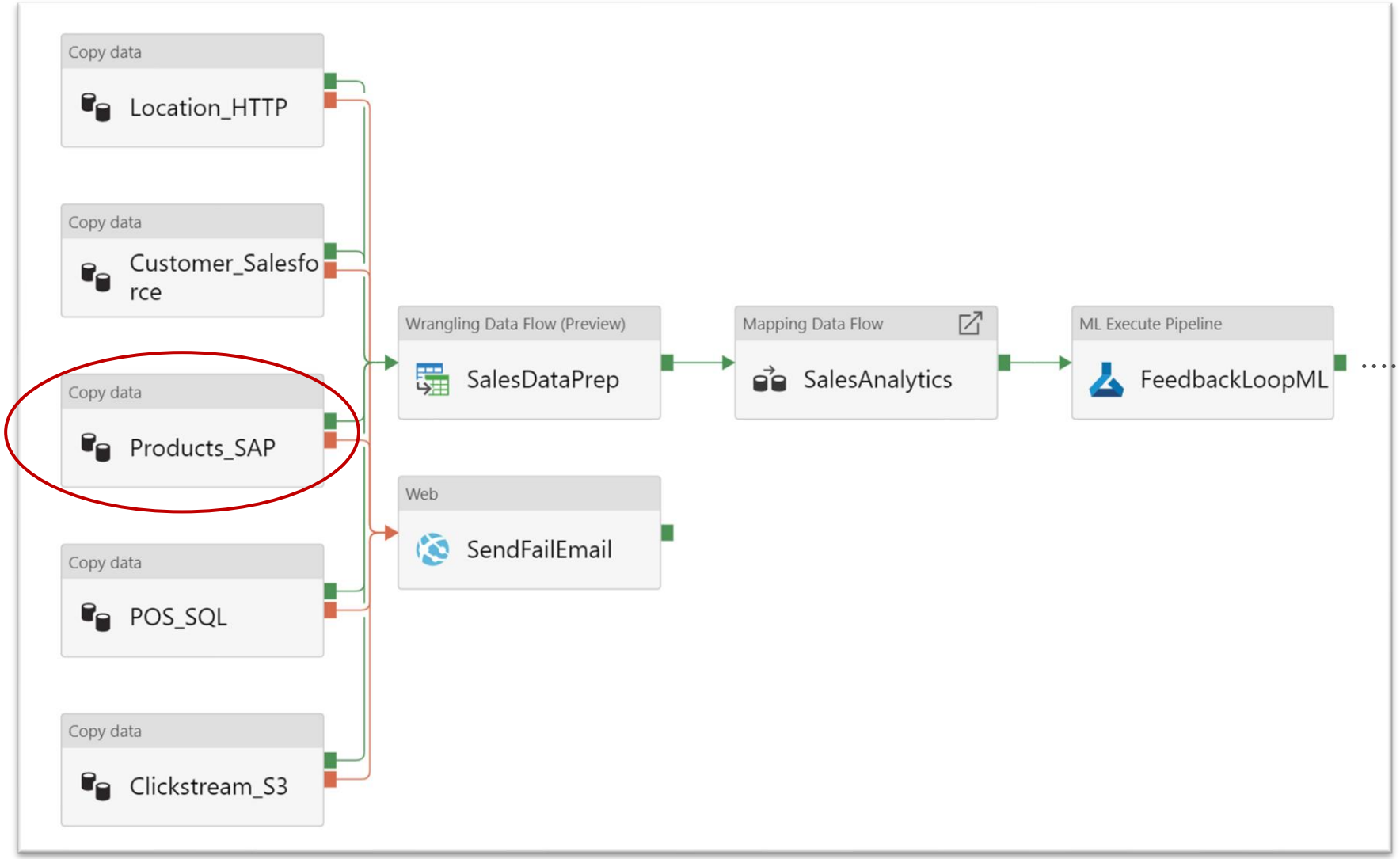
- Certified compliance
- Enterprise grade security
- MSI and AKV support

Productive

- Drag & drop UI
- Single-pane-of-glass monitoring
- CICD model

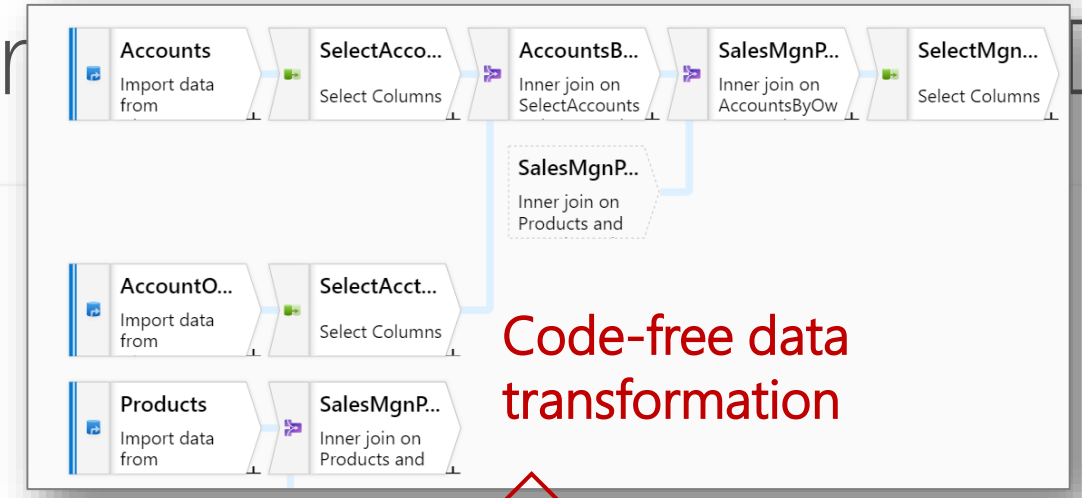
Build Modern Data Warehouse Solution w/ SAP Data

Example:
Sales Analytics



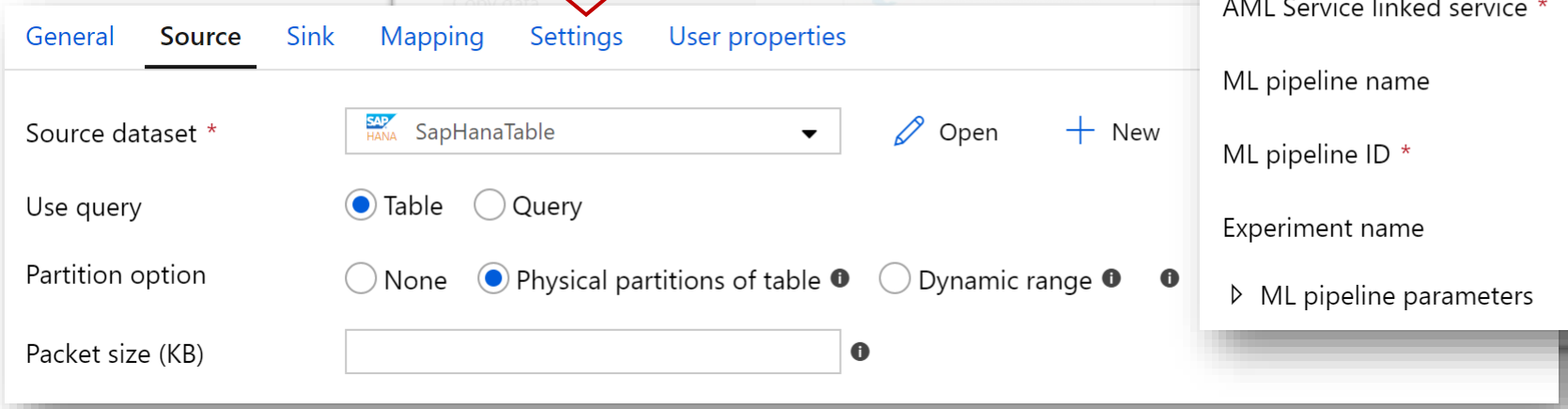
Build Modern Data War … Data

Example:
Sales Analytics

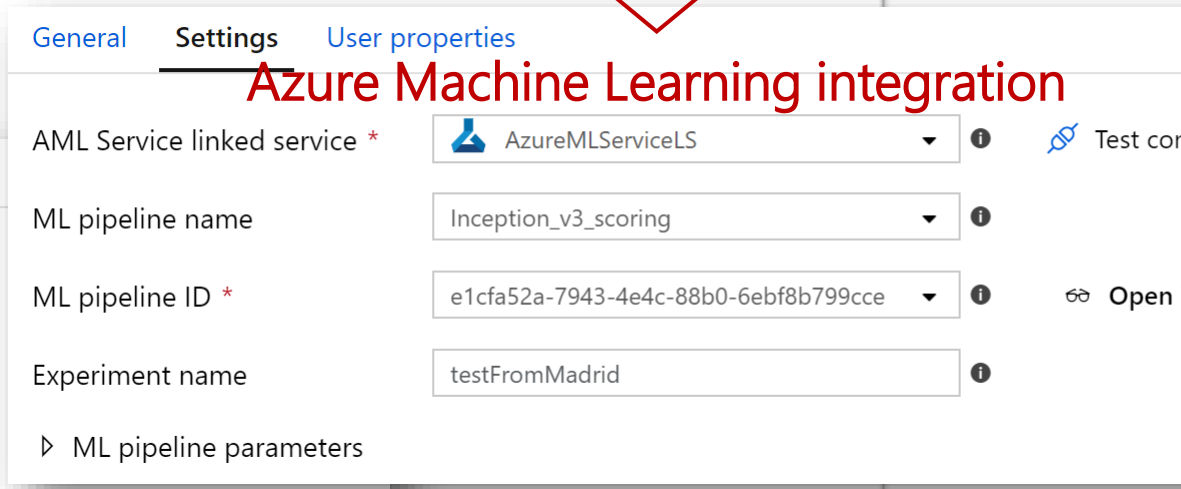


Code-free data transformation

SAP data ingestion

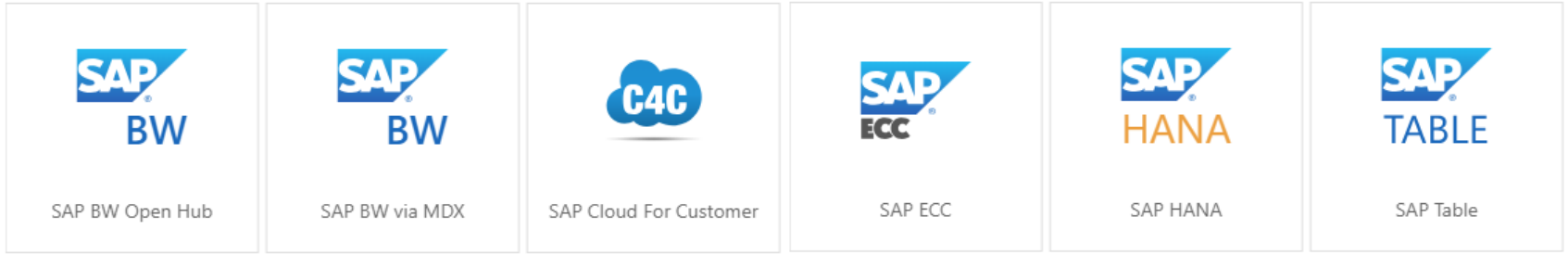


Azure Machine Learning integration



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	Database & DW		File Storage	File Formats	NoSQL	Services & Apps		Generic
Blob Storage	Amazon Redshift	Oracle	Amazon S3	AVRO	Cassandra	Amazon MWS	PayPal	HTTP
Cosmos DB – SQL API	DB2	Phoenix	File System	Delimited Text	Couchbase	CDS for Apps	QuickBooks	OData
Cosmos DB – MongoDB API	Drill	PostgreSQL	FTP	Excel	MongoDB	Concur	Salesforce	ODBC
ADLS Gen1	Google BigQuery	Presto	Google Cloud Storage	JSON		Dynamics 365	SF Service Cloud	REST
ADLS Gen2	Greenplum	SAP BW Open Hub	HDFS	ORC		Dynamics AX	SF Marketing Cloud	
Data Explorer	HBase	SAP BW MDX	SFTP	Parquet		Dynamics CRM	SAP C4C	
Database for MariaDB	Hive	SAP HANA				Google AdWords	SAP ECC	
Database for MySQL	Impala	SAP Table				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 Responsys	Zoho	
Table Storage						Oracle Service Cloud		

Table of Content

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" →

Suggested decision direction

ADF connector options	SAP Table 	SAP BW Open Hub 	SAP BW via MDX 
Objects to extract	Table (Transparent, Pooled, Cluster Table) and View	DSO, InfoCube, MultiProvider, DataSource, etc	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
	(Connector deep-dive)	(Connector deep-dive)	(Connector 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" →

Suggested decision direction

ADF connector options	SAP Table 	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 Gateway
Performance	Fast w/ built-in parallel loading	Slower
Suitable workload	Large volume	Small volume
	(Connector deep-dive)	(Connector deep-dive)

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

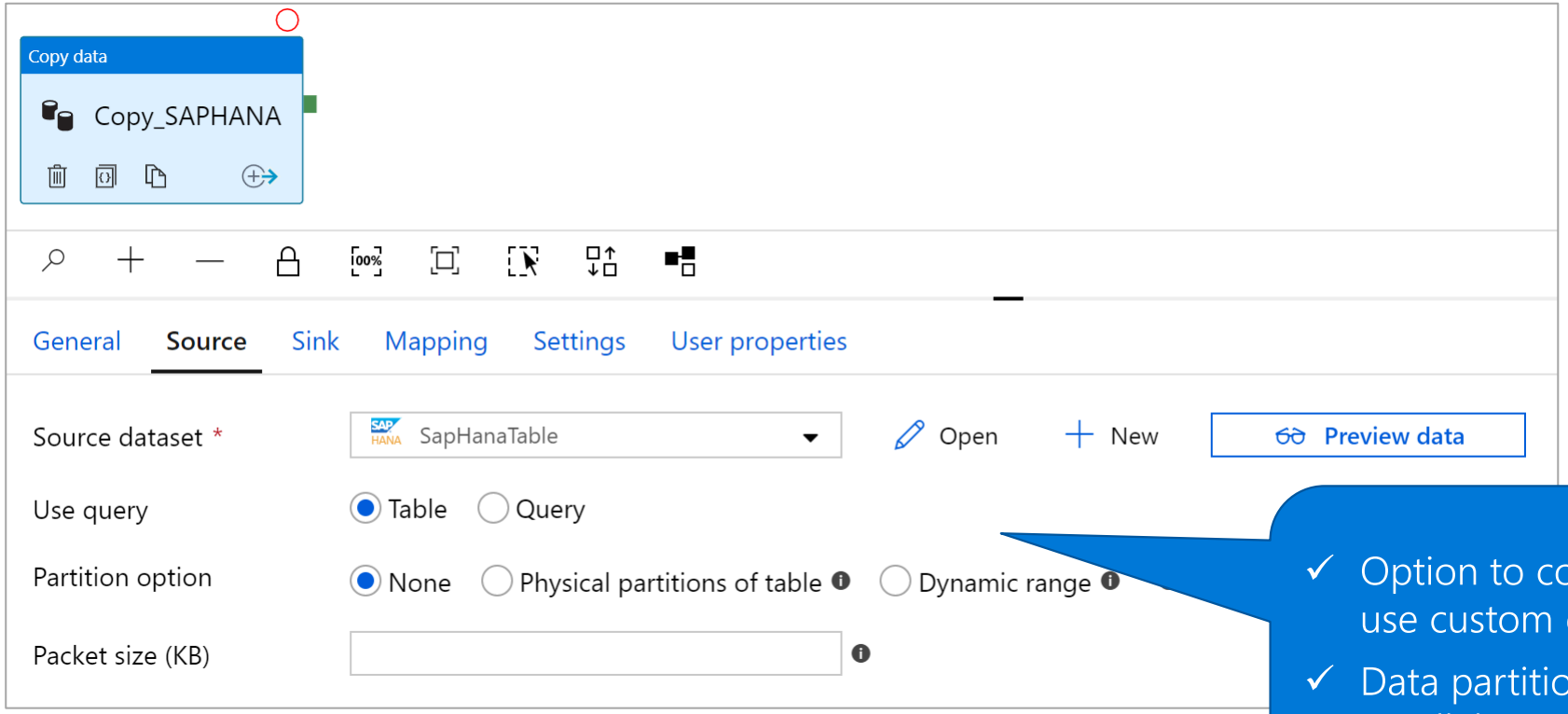
SAP HANA Connector

SAP HANA Connector

Suitable scenario: ingest data from **SAP HANA**.

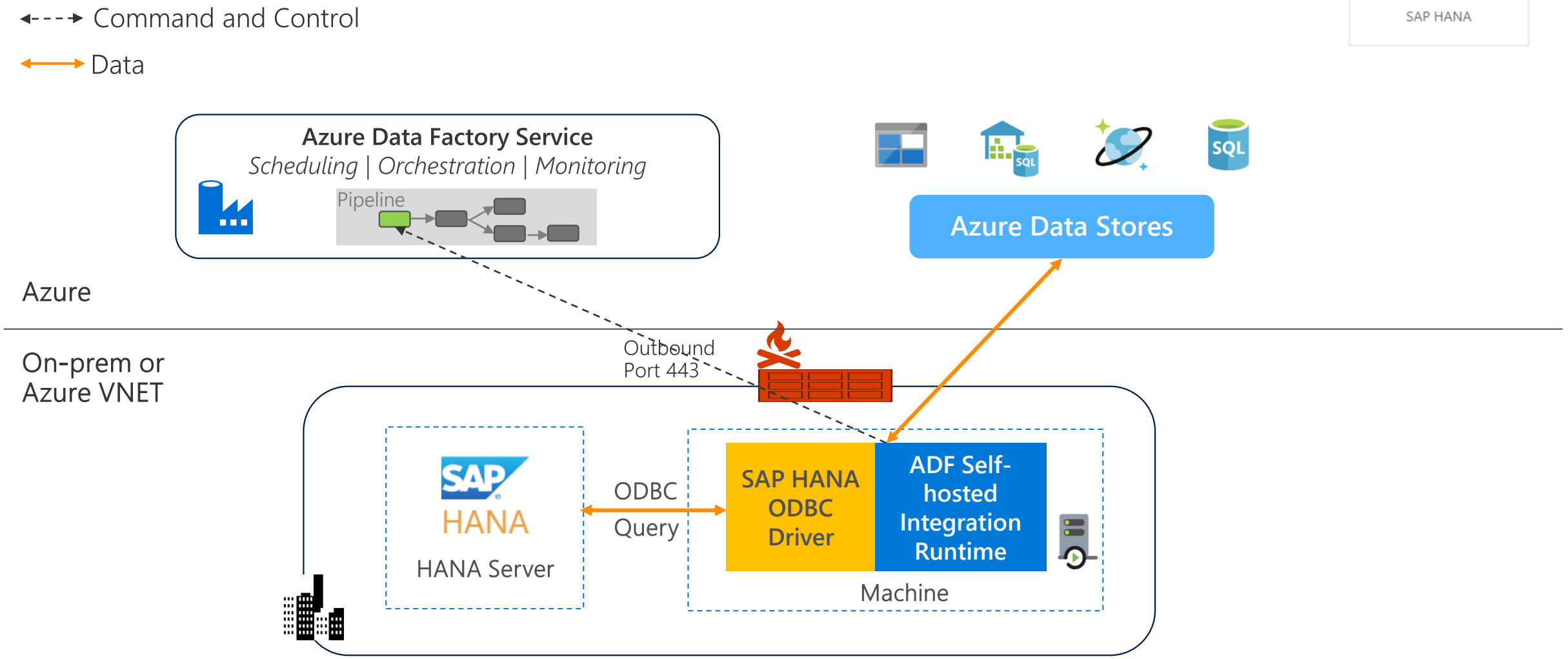
Supported versions	• All SAP HANA versions, on-prem or in the cloud
Supported SAP objects	• HANA Information Models (Analytic/Calculation views) • Row & Column Tables
Supported authentications	• **Basic** – username & password • **Windows** – Single Sign-On via Kerberos-constrained delegation
Mechanism 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



- ✓ Option to copy entire table or use custom query
- ✓ Data partition options for parallel copy to boost perf

SAP HANA Connector – How It Works

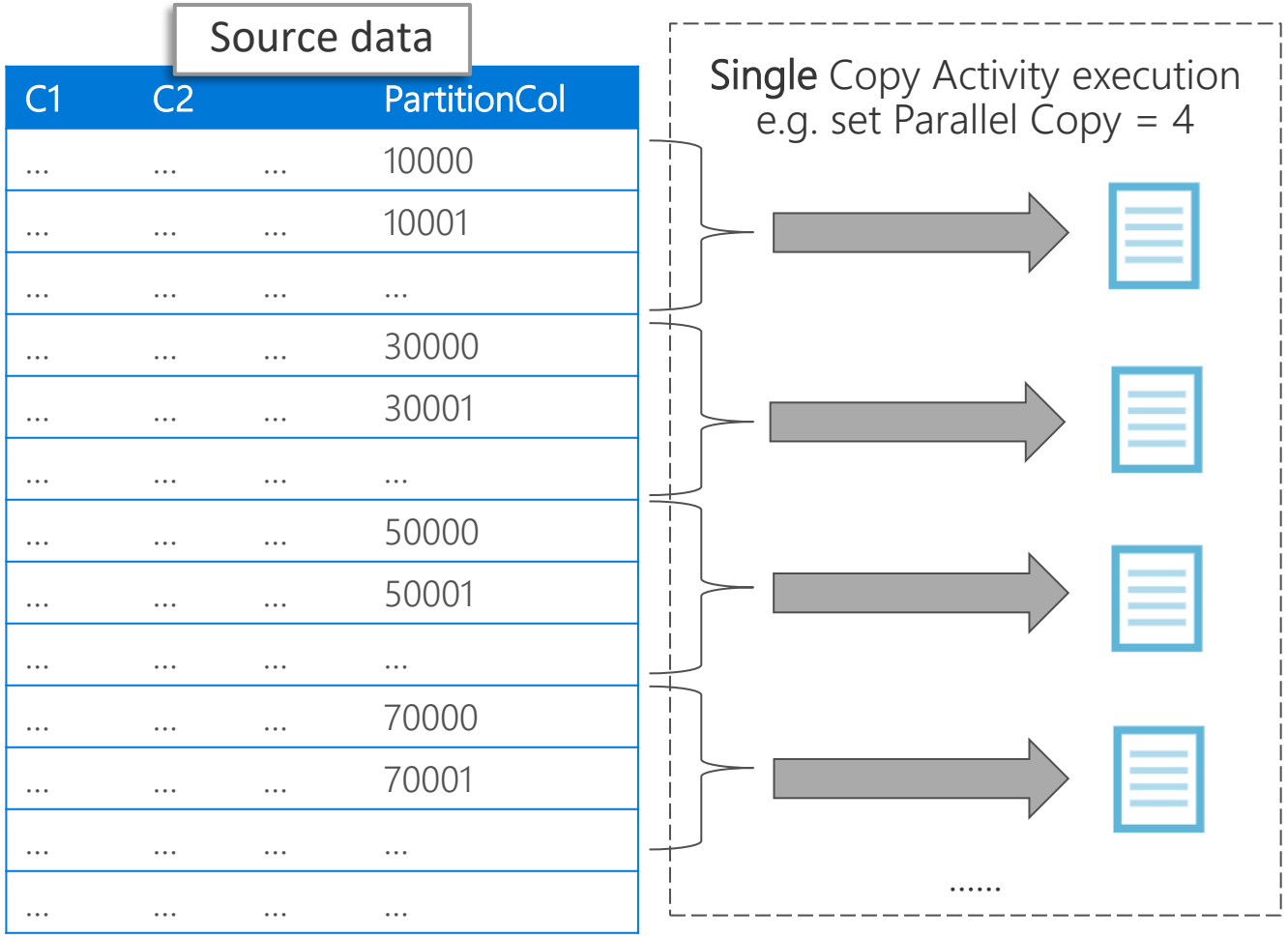


SAP HANA Connector – Built-in Parallel Loading

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"
Solution: 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')
```

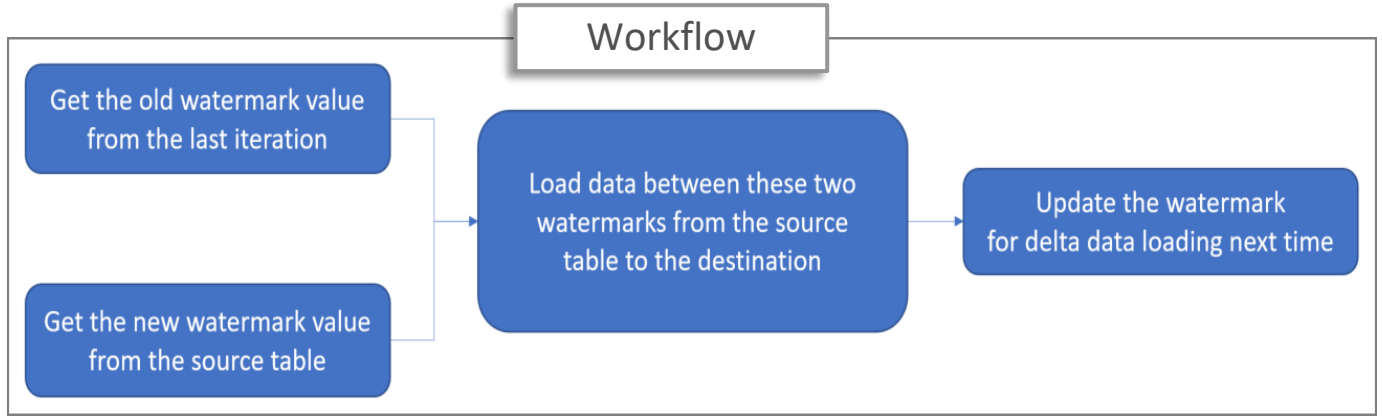
Execution start time: 2019/03/19 00:00:00 (window end time)
Delta extraction: last modified time between 2019/03/18 – 2019/03/19

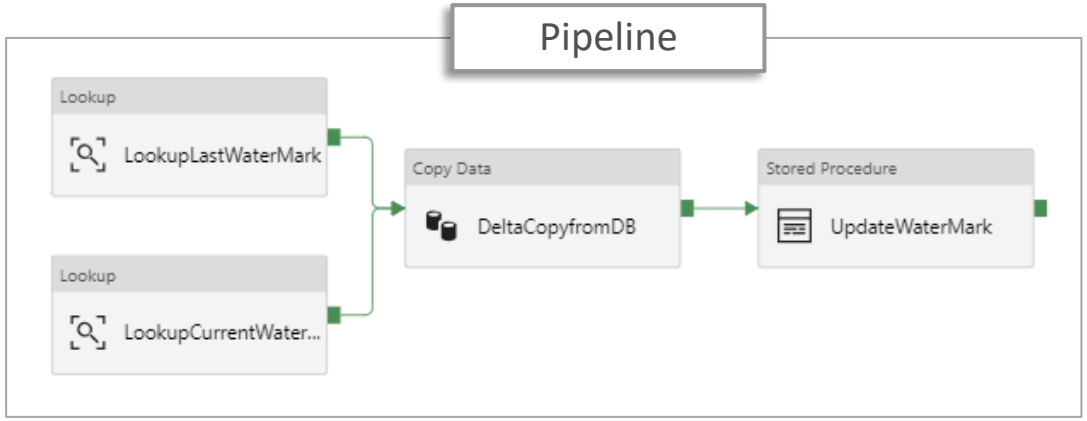
Execution start time: 2019/03/20 00:00:00 (window end time)
Delta extraction: last modified time between 2019/03/19 – 2019/03/20

SAP HANA Connector – Incremental Copy

SAP HANA

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

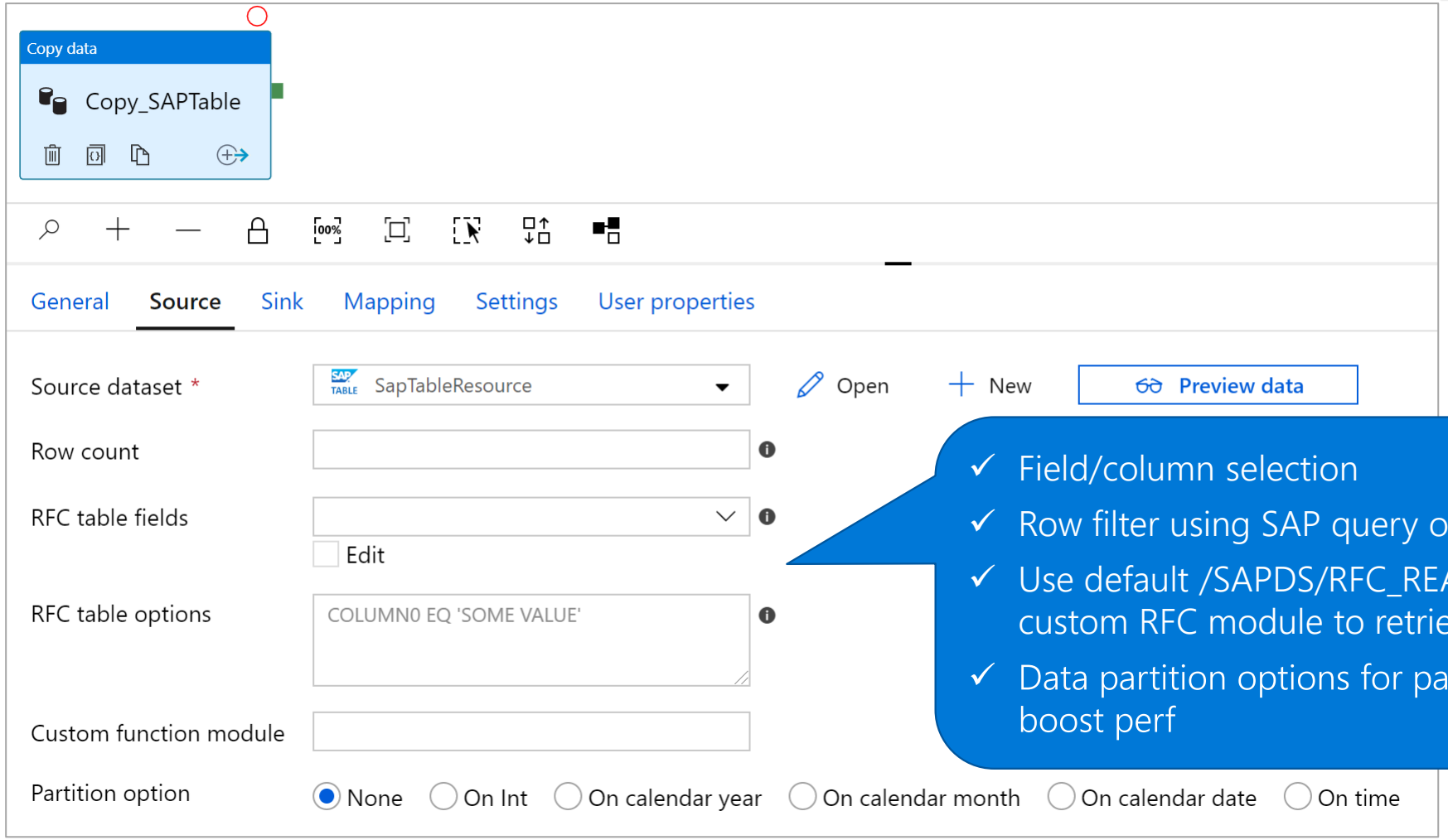
Suitable scenario: ingest data from SAP Table for SAP ECC, S/4 HANA, BW, or other application in Business Suite.

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

SAP Table



- ✓ Field/column selection
- ✓ Row filter using SAP query operators
- ✓ Use default /SAPDS/RFC_READ_TABLE2 or custom RFC module to retrieve data
- ✓ Data partition options for parallel copy to boost perf

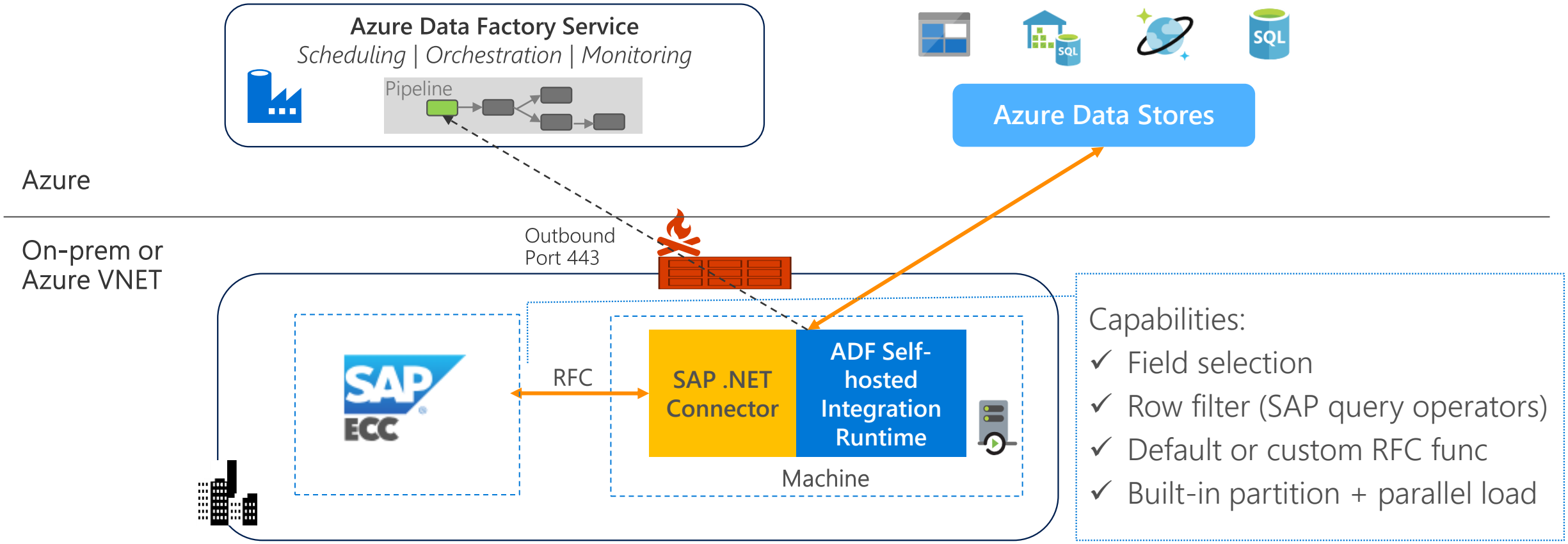
SAP Table Connector – How It Works

SAP TABLE

SAP Table

Command and Control

Data



Capabilities:

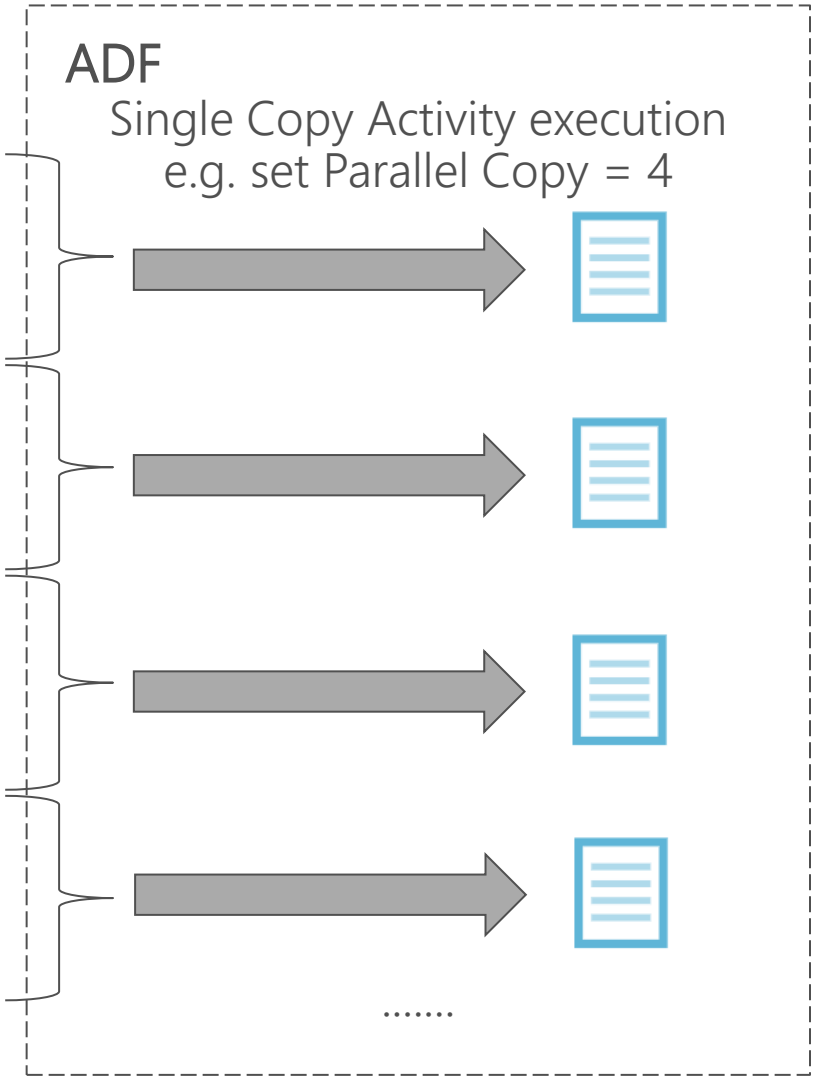
- ✓ Field selection
- ✓ Row filter (SAP query operators)
- ✓ Default or custom RFC func
- ✓ Built-in partition + parallel load

SAP Table Connector – Built-in Parallel Loading

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

SAP table

C1	C2		PartitionCol
…	…	…	201809
…	…	…	201809
…	…	…	…
…	…	…	201810
…	…	…	201810
…	…	…	…
…	…	…	201811
…	…	…	201811
…	…	…	…
…	…	…	201812
…	…	…	201812
…	…	…	…
…	…	…	…



Tips:

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.

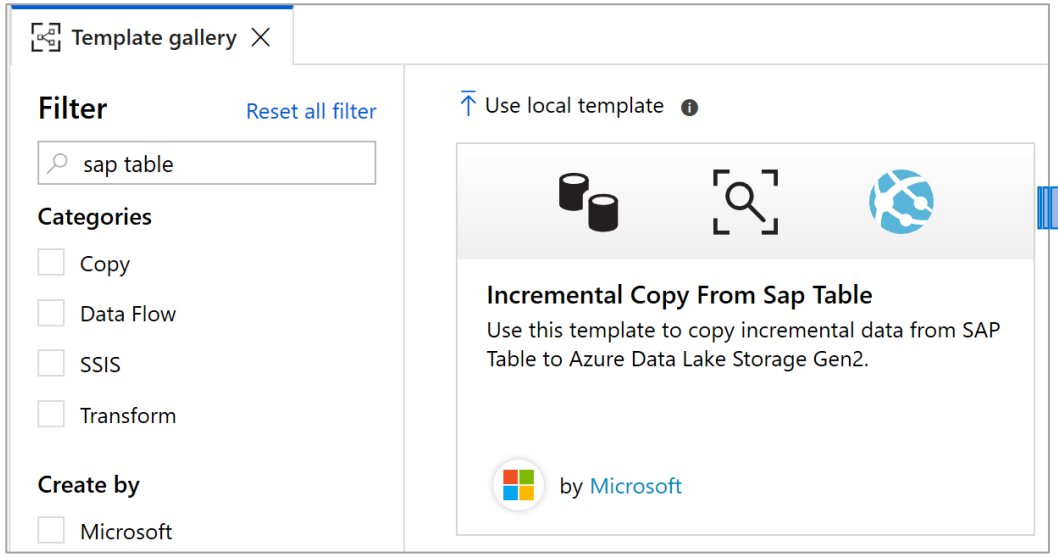
Learn more

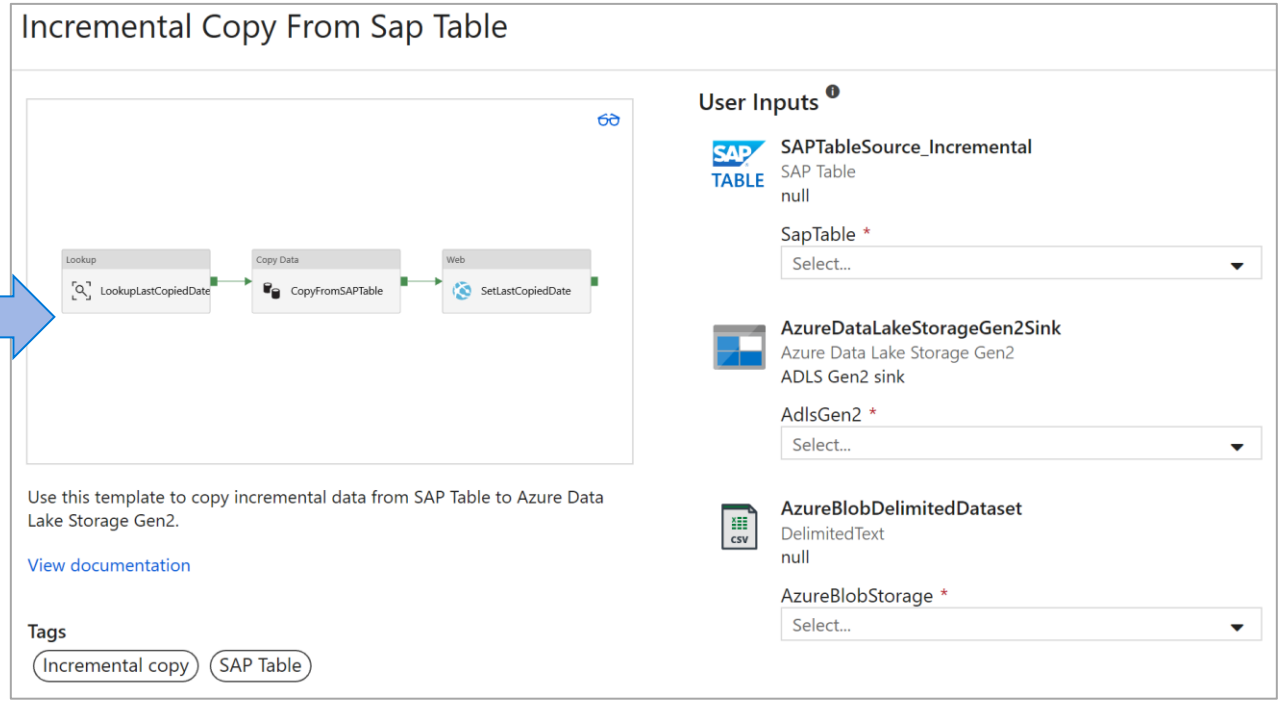
SAP Table Connector – Incremental Copy

Pattern I: "my data has timestamp column e.g. calendar date"
Solution: tumbling window trigger + dynamic query with system variables via SAP table option (filter)

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

Get started via solution template:





SAP BW Open Hub Connector

SAP BW Open Hub Connector

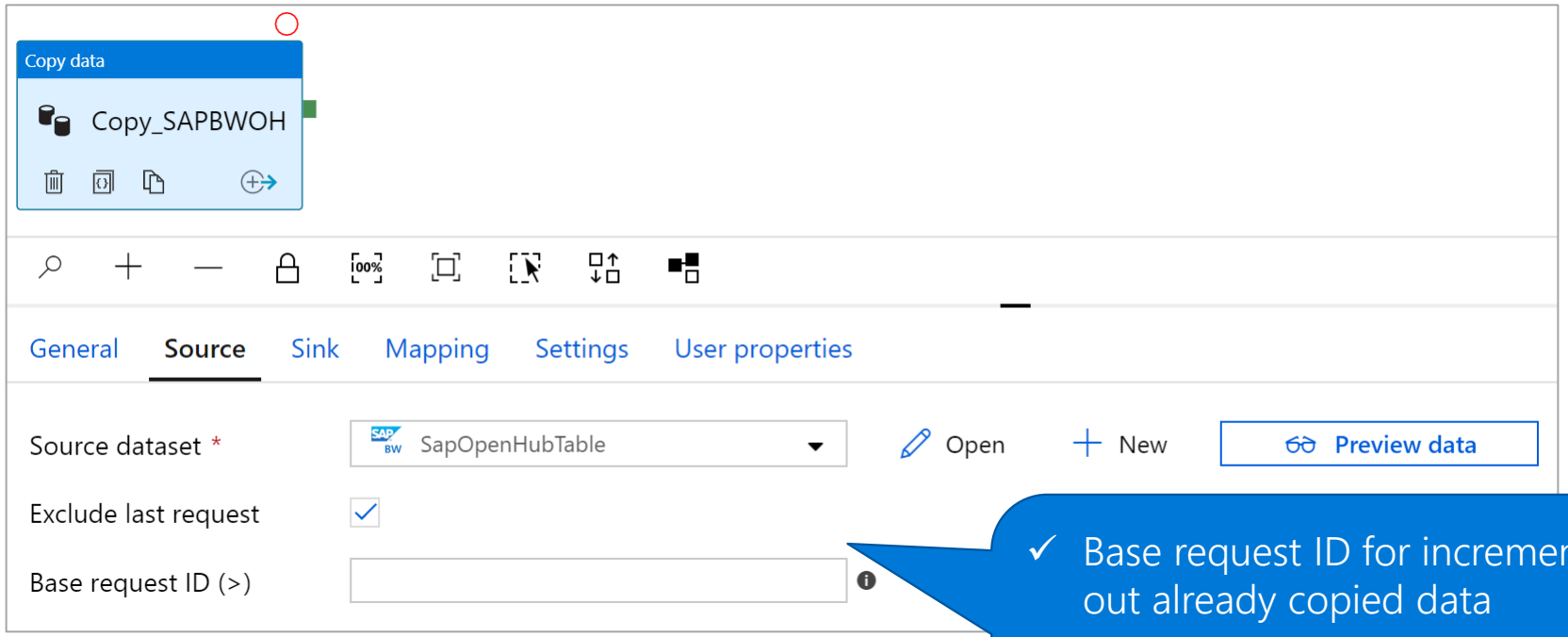
Suitable scenario: ingest data from **SAP BW** with targeted/well-thought-through workload.

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



- ✓ Base request ID for incremental copy to filter out already copied data
- ✓ Exclude last request to avoid partial data
- ✓ Built-in parallel copy to boost perf based on OHD's specific schema

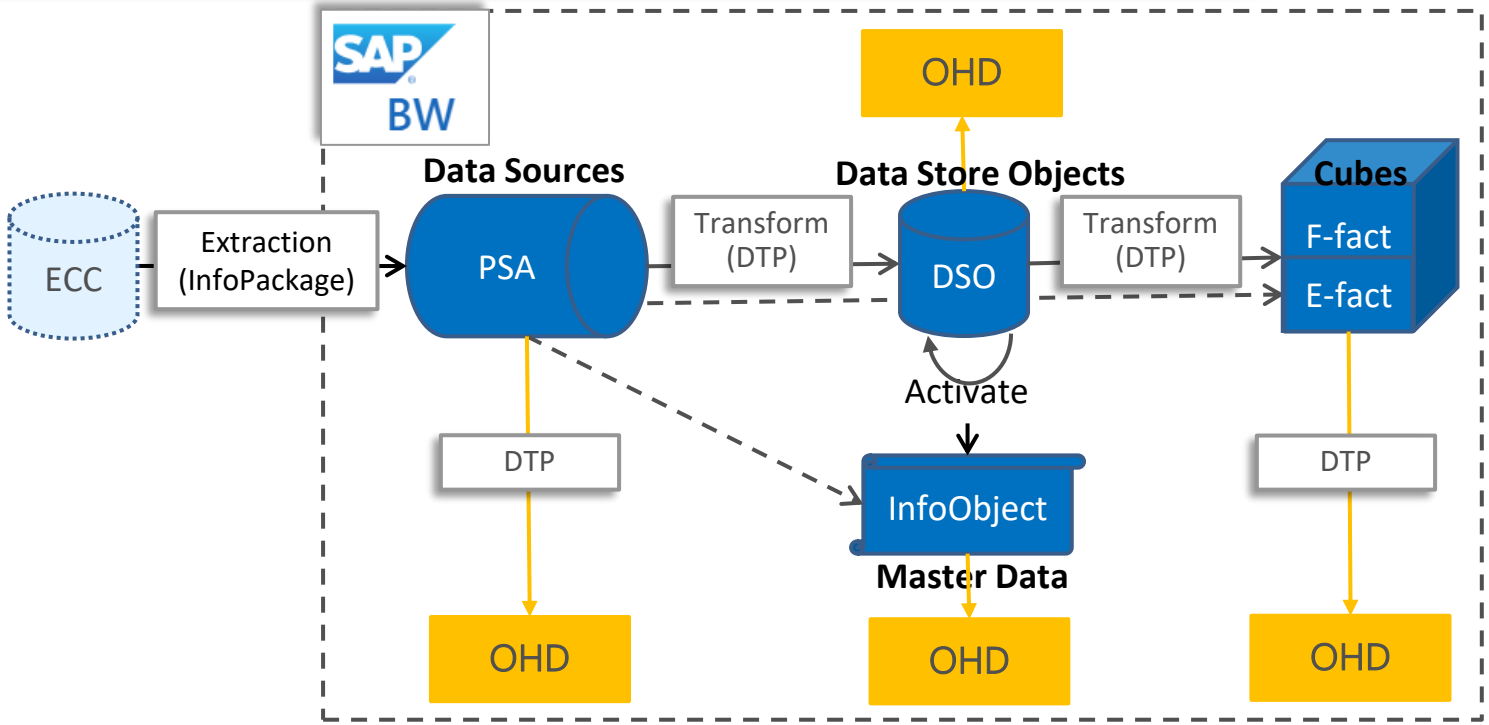
SAP BW Open Hub – How It Works

SAP BW Open Hub

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**.*

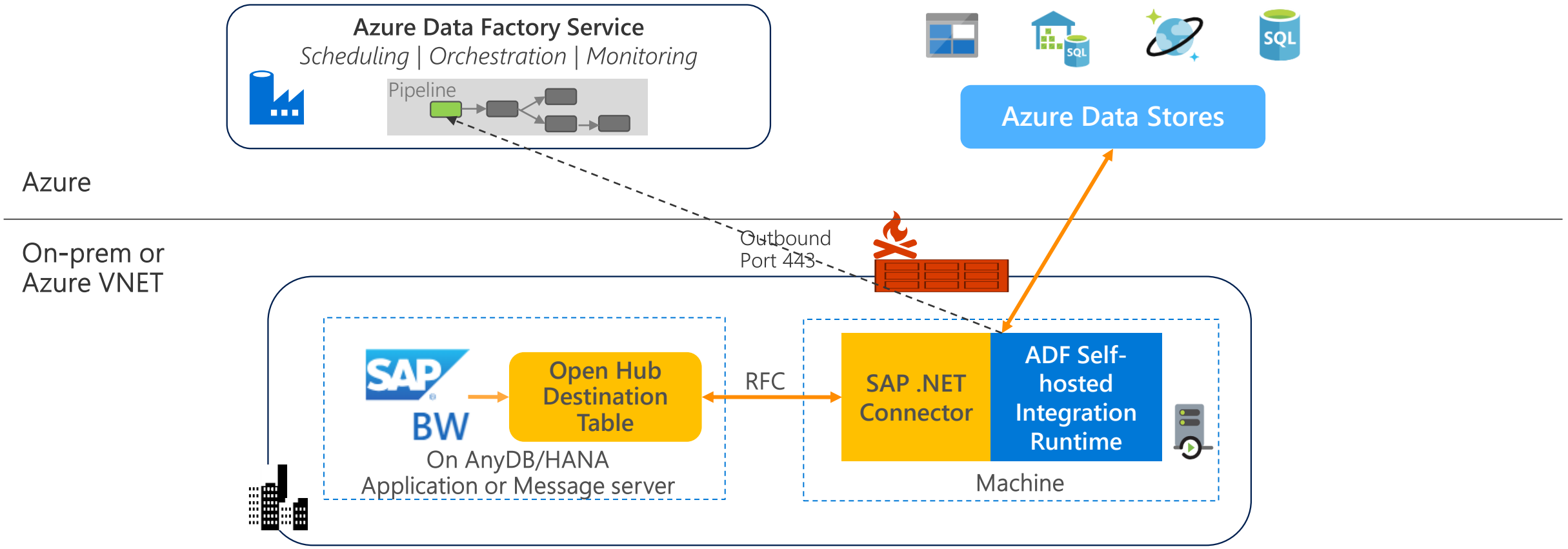


SAP BW Open Hub Connector – How It Works

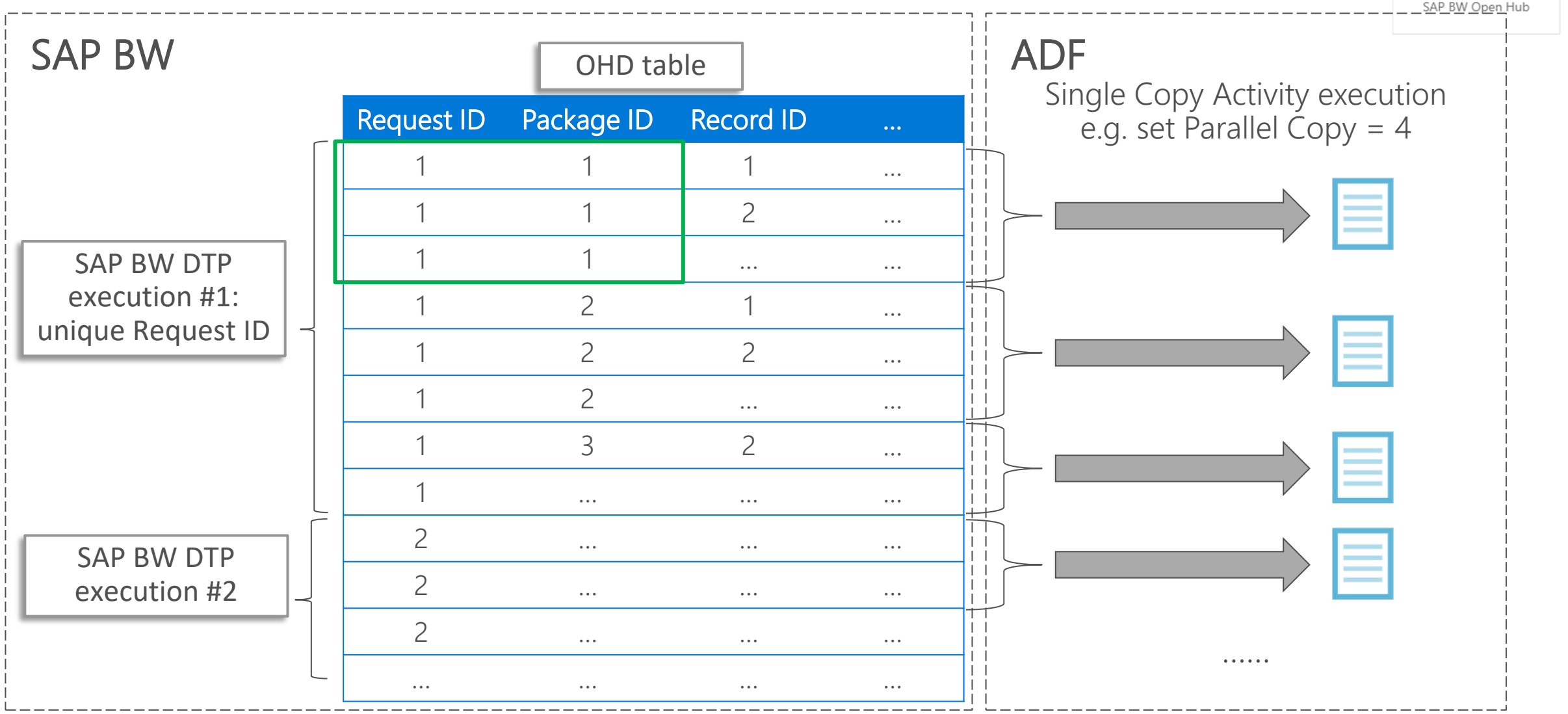
SAP BW Open Hub

Command and Control

Data



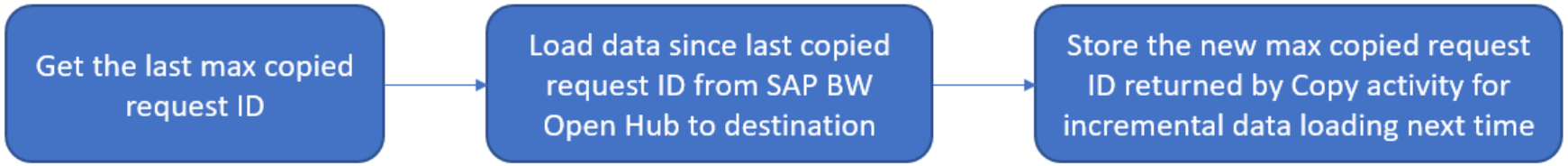
SAP BW Open Hub Connector – Built-in Parallel Loading



SAP BW Open Hub Connector – Incremental Copy

SAP BW Open Hub

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

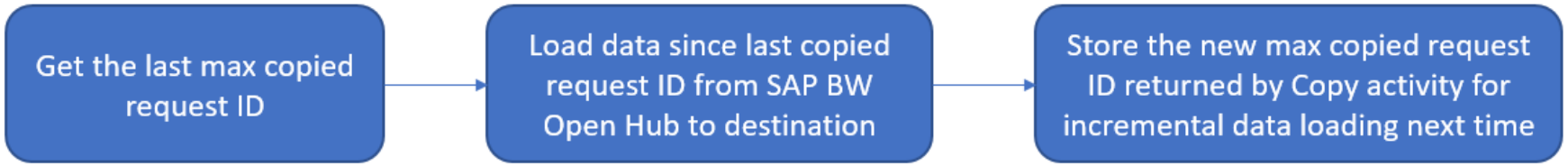


SAP OHD supports "delta" extraction mode to load incremental data into OHD table.

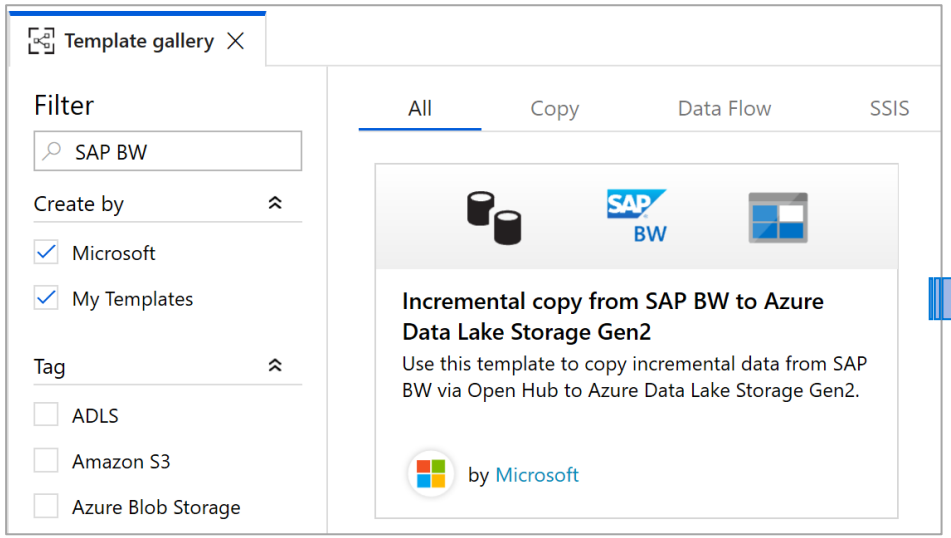
SAP BW Open Hub Connector – Incremental Copy

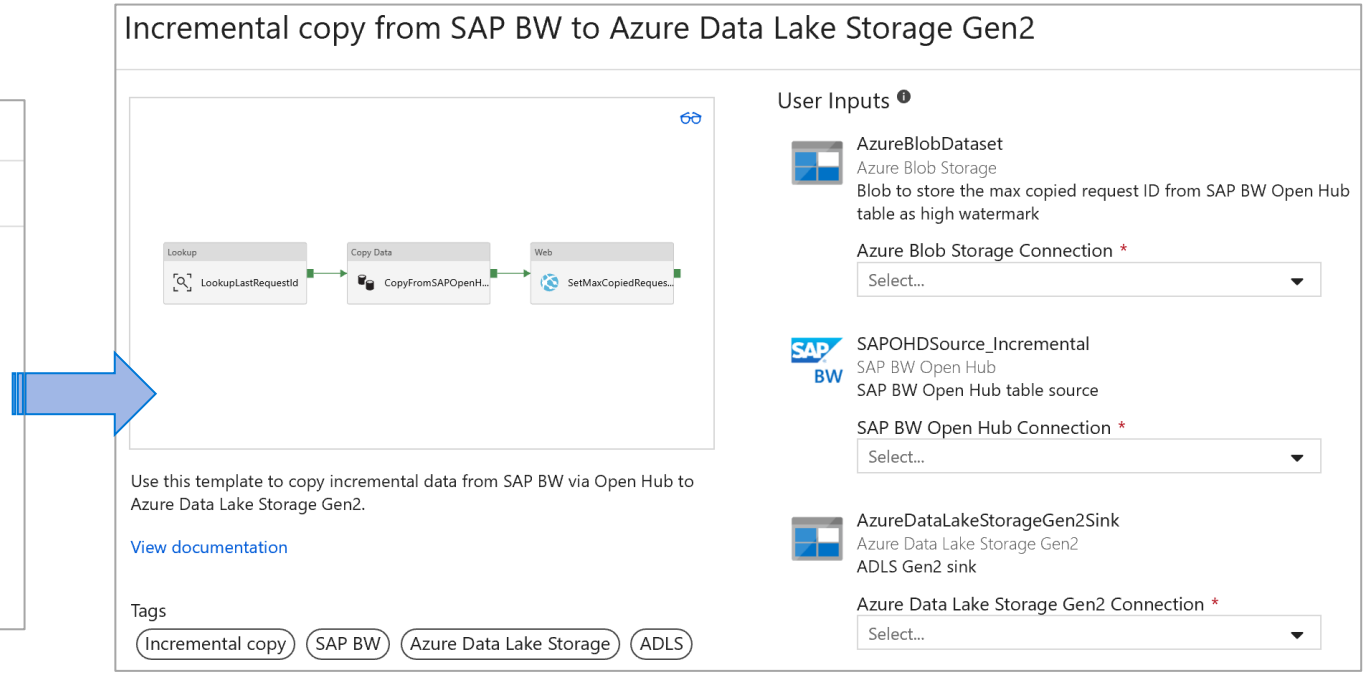
SAP BW Open Hub

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



Get started via solution template:

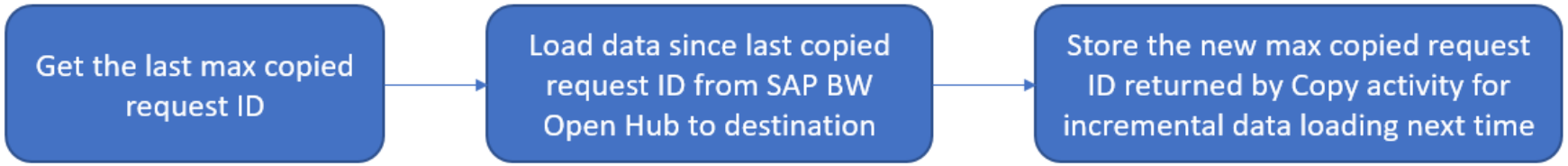




SAP BW Open Hub Connector – Incremental Copy

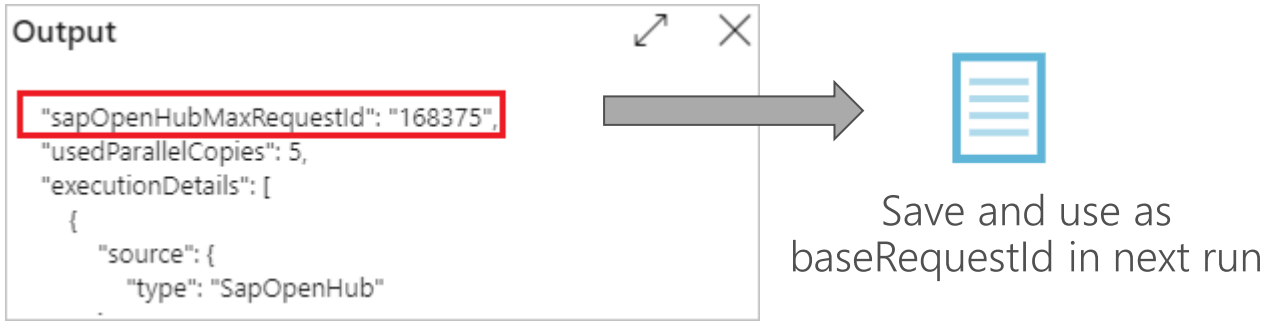
SAP BW Open Hub

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:

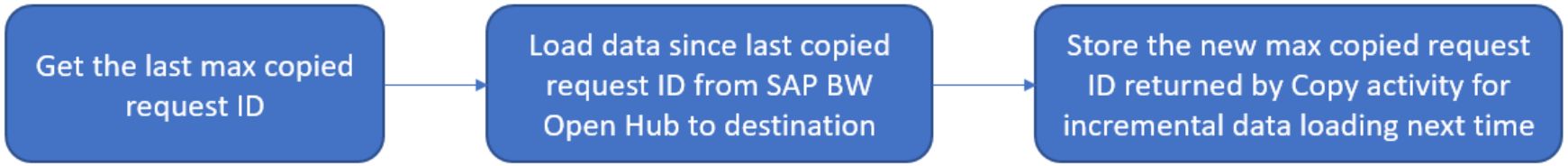


Save and use as baseRequestId in next run

SAP BW Open Hub Connector – Incremental Copy

SAP BW Open Hub

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



- 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 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

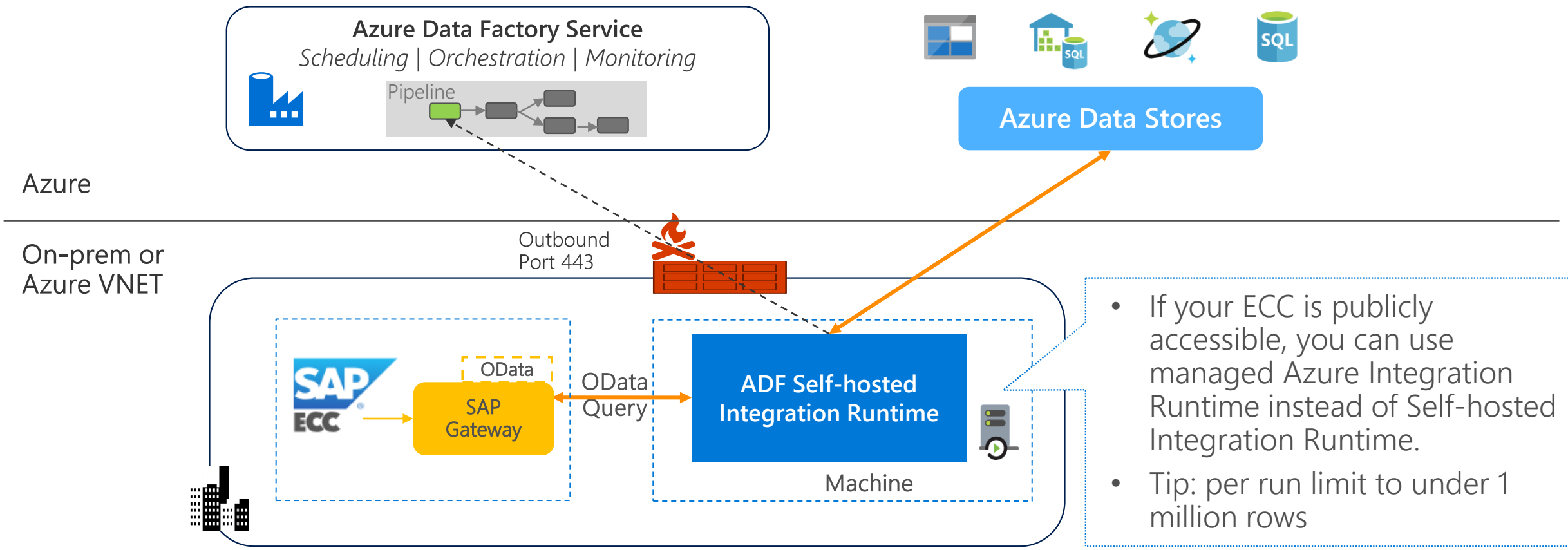
Suitable scenario: ingest data from **SAP Applications** other than SAP Table.

Supported versions	• SAP ECC version 7.0 and above • Any entities exposed by SAP ECC OData services
Supported SAP objects	• Entities exposed by SAP OData services • BAPI, ODP (DataExtractors/DataSource), etc.
Supported authentications	• Basic – user name & password
Mechanism and prerequisites	• 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



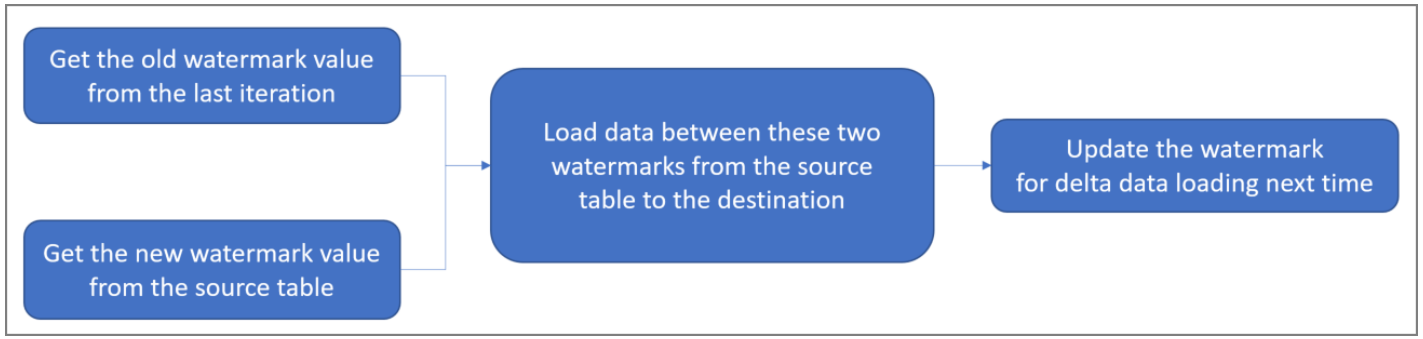
- If your ECC is publicly accessible, you can use managed Azure Integration Runtime instead of Self-hosted Integration Runtime.
- Tip: per run limit to under 1 million rows

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"
Solution: 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

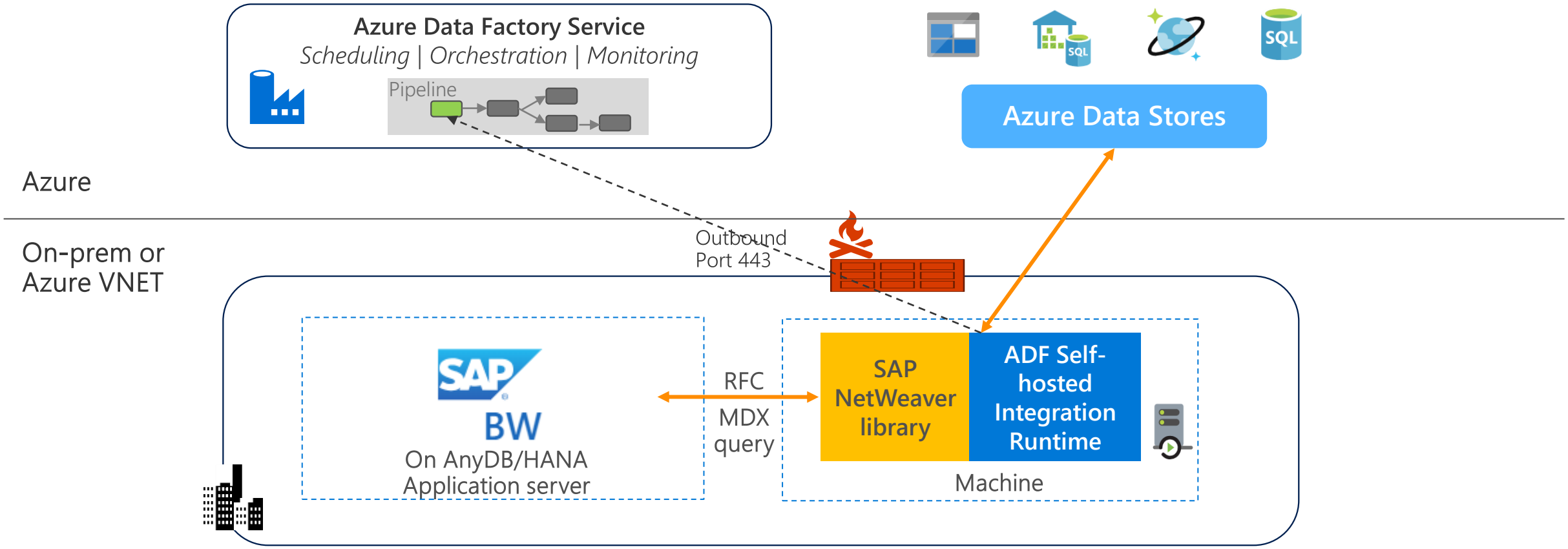
NOTE: SAP BW4/HANA is not supported now.

SAP BW via MDX Connector – How It Works

SAP BW via MDX

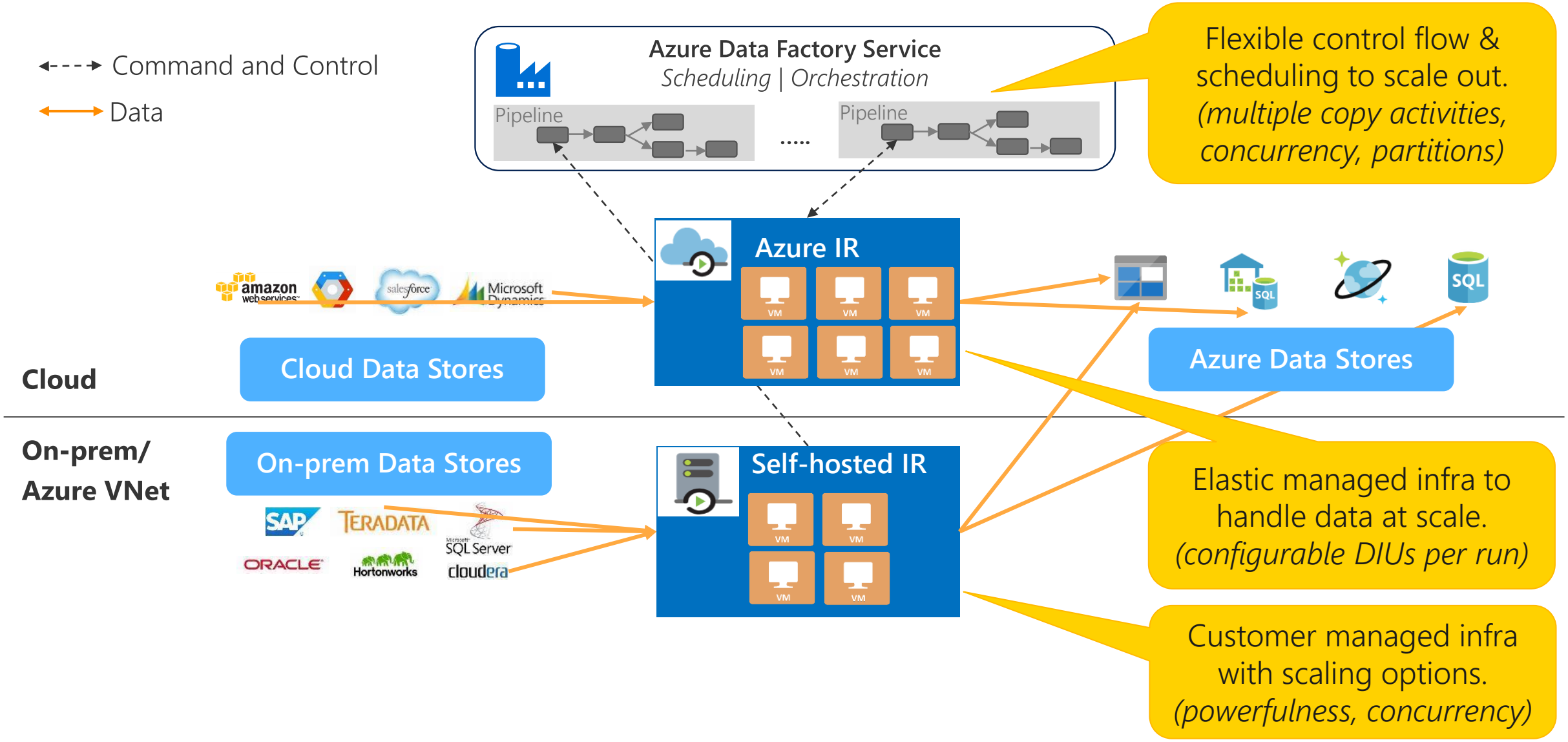
Command and Control

Data



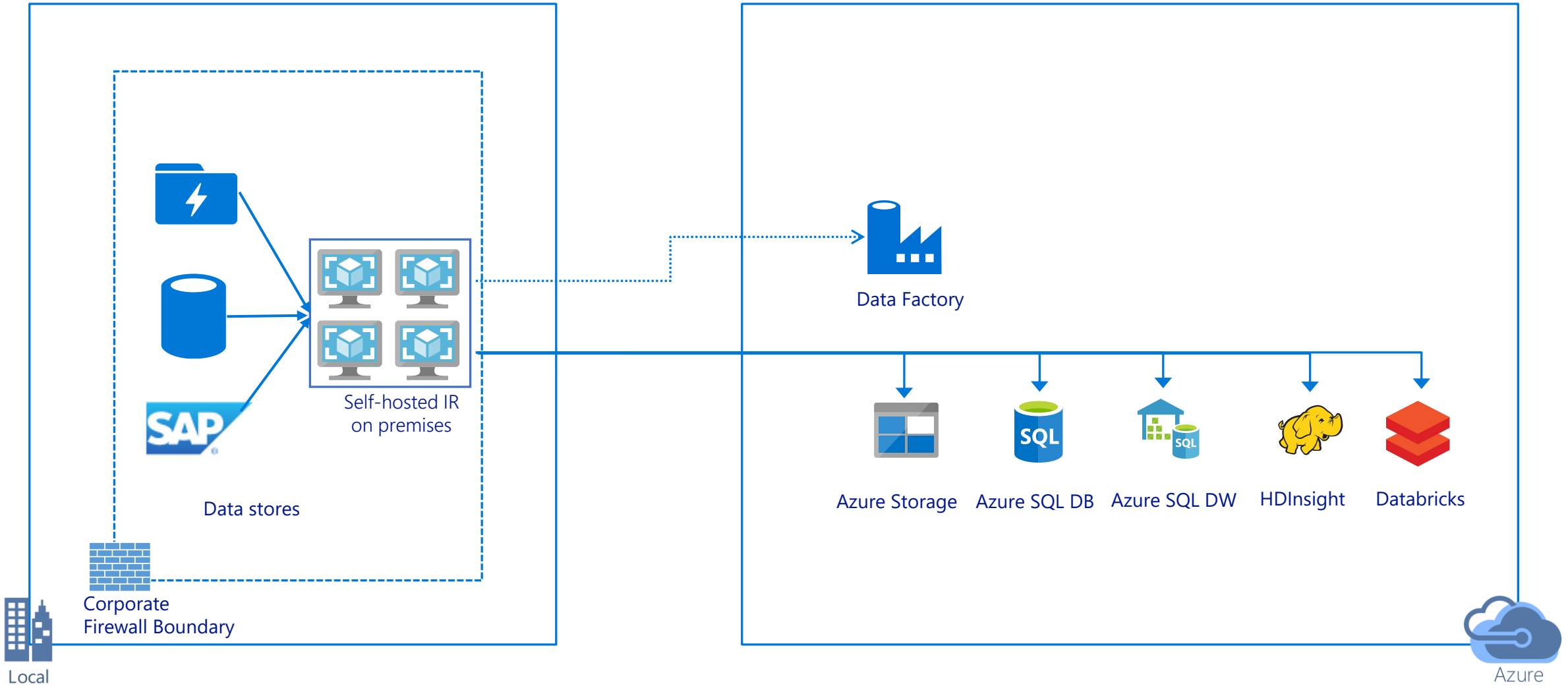
More about Azure Data Factory Copy Activity

Understand How ADF Copy Scales



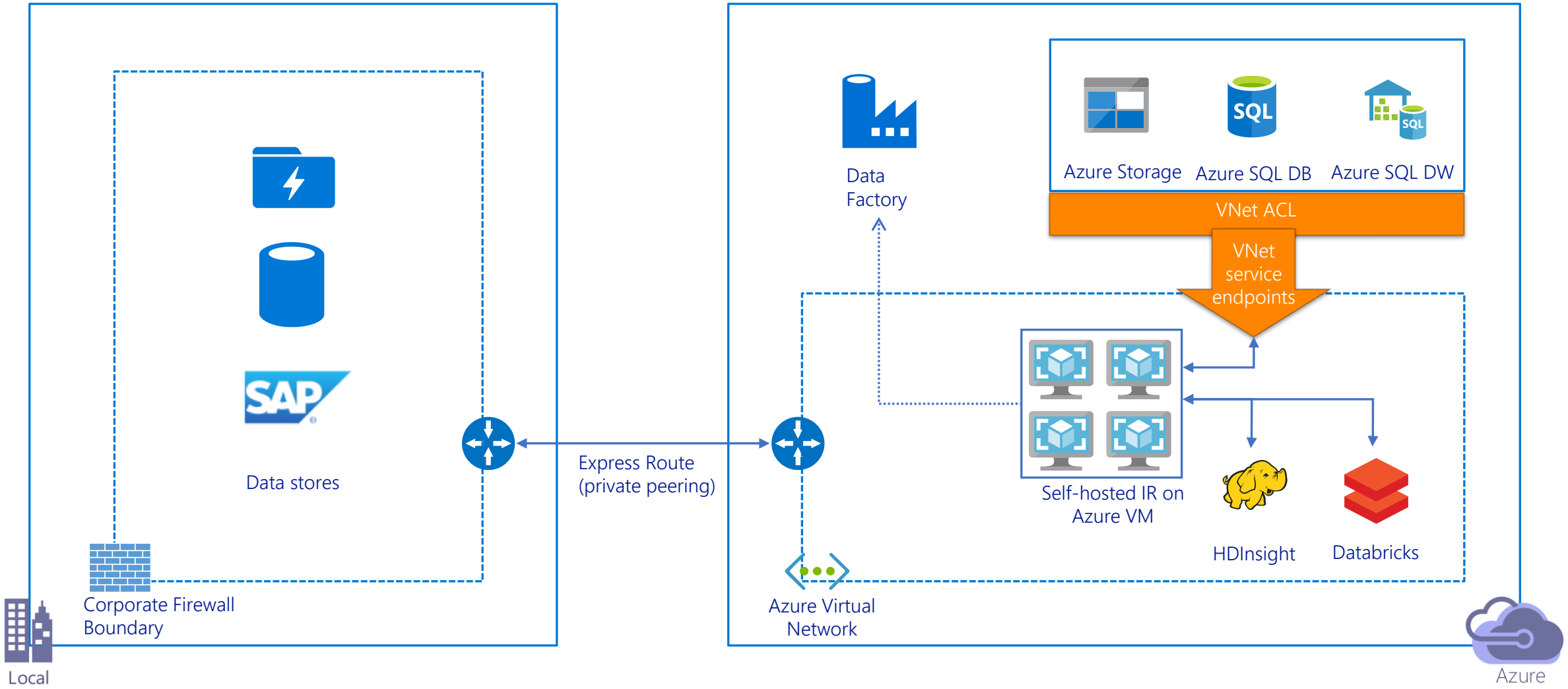
On-prem connected to Azure through public internet

Self-hosted IR deployed on premises



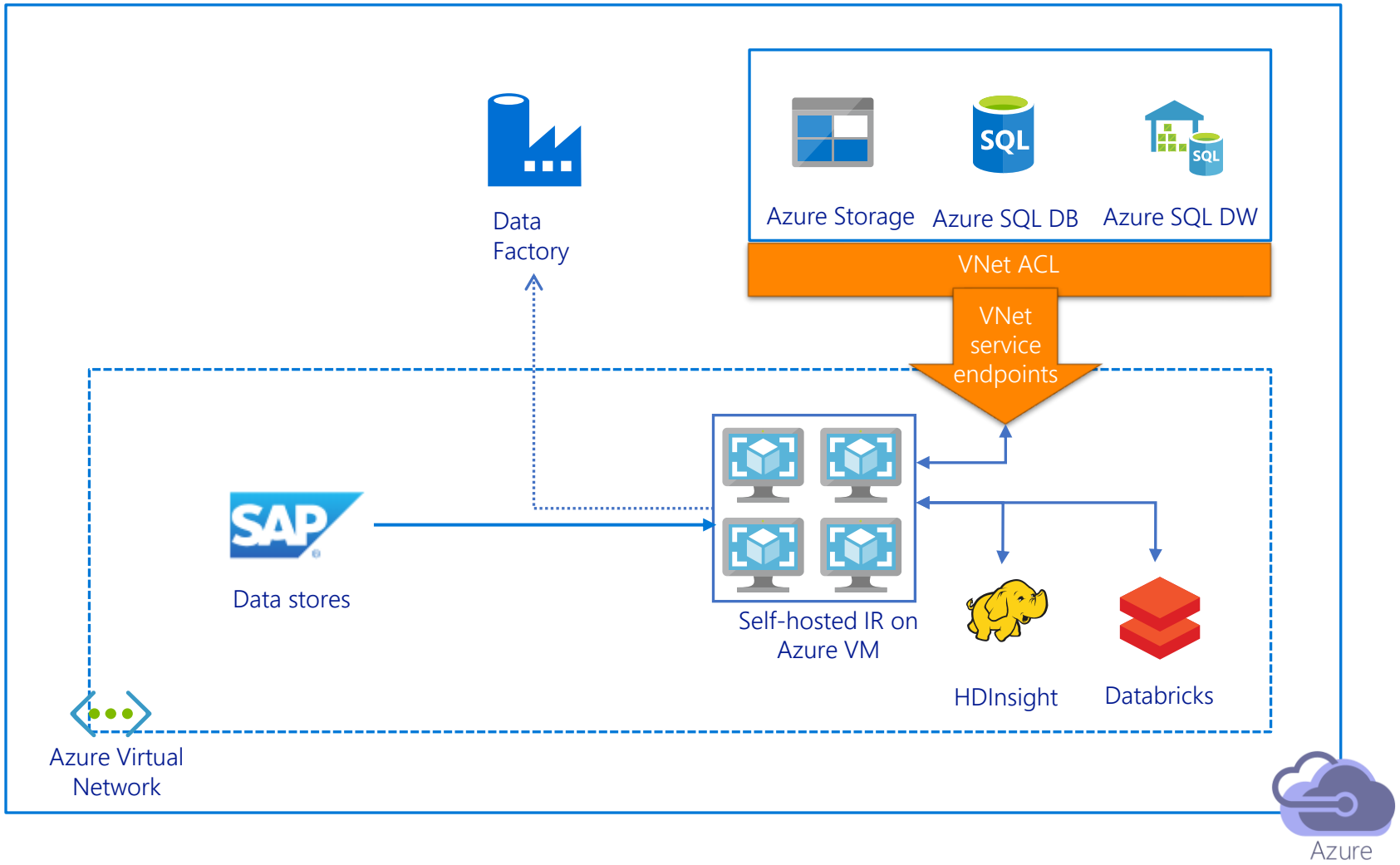
On-prem connected to Azure VNet through ExpressRoute

Self-hosted IR deployed on Azure VM

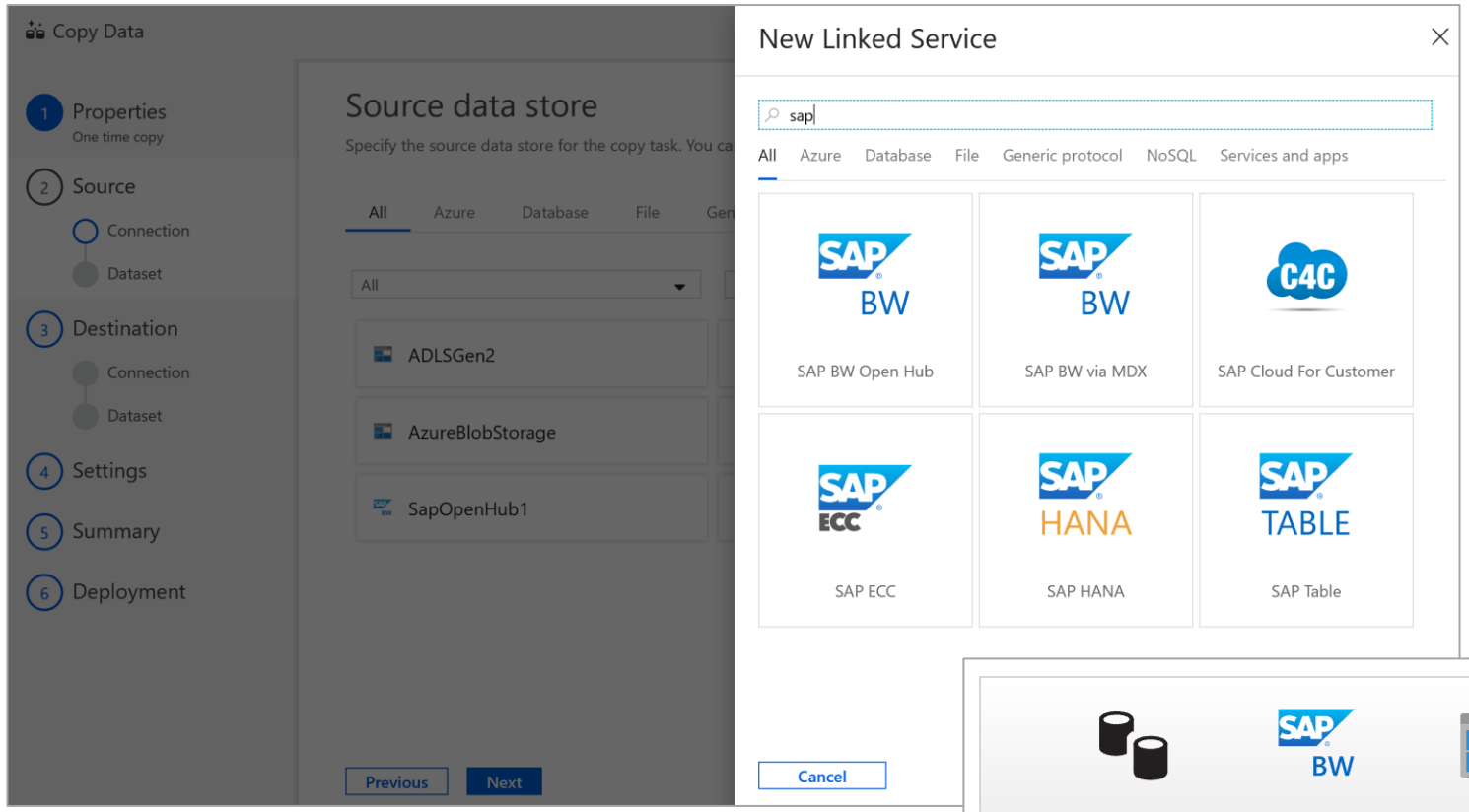


SAP on Azure

Self-hosted IR deployed on Azure VM

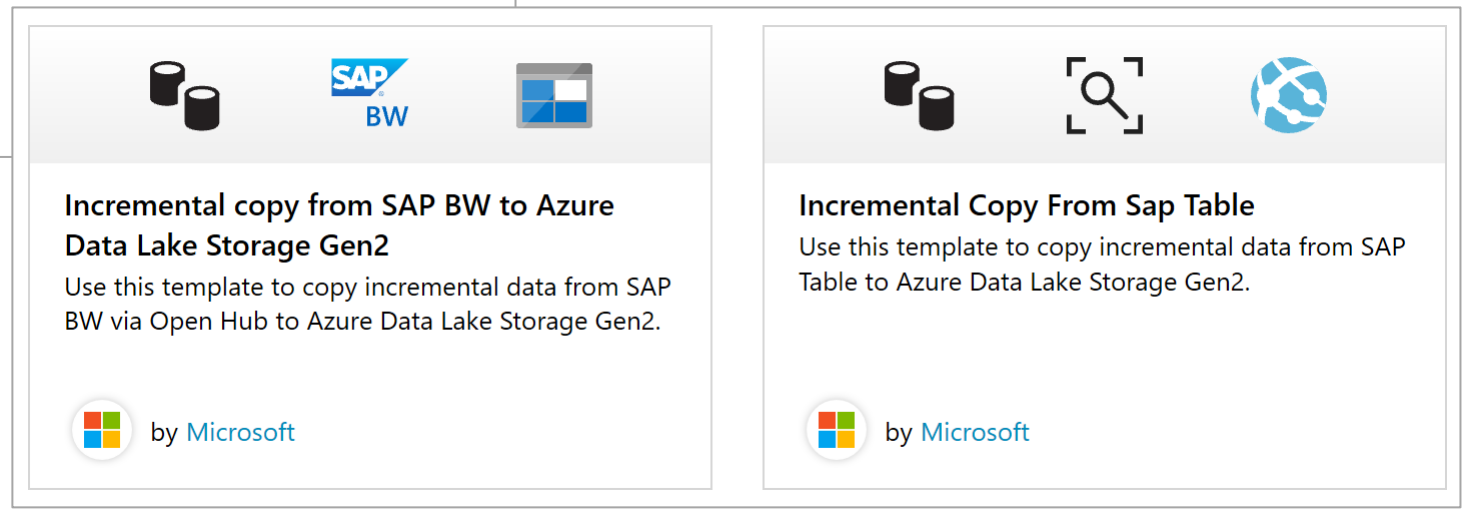


Get Started



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/reckitt-benckiser-consumer-goods-power-bi • Newell: https://customers.microsoft.com/story/newell-brands-consumer-goods-azure