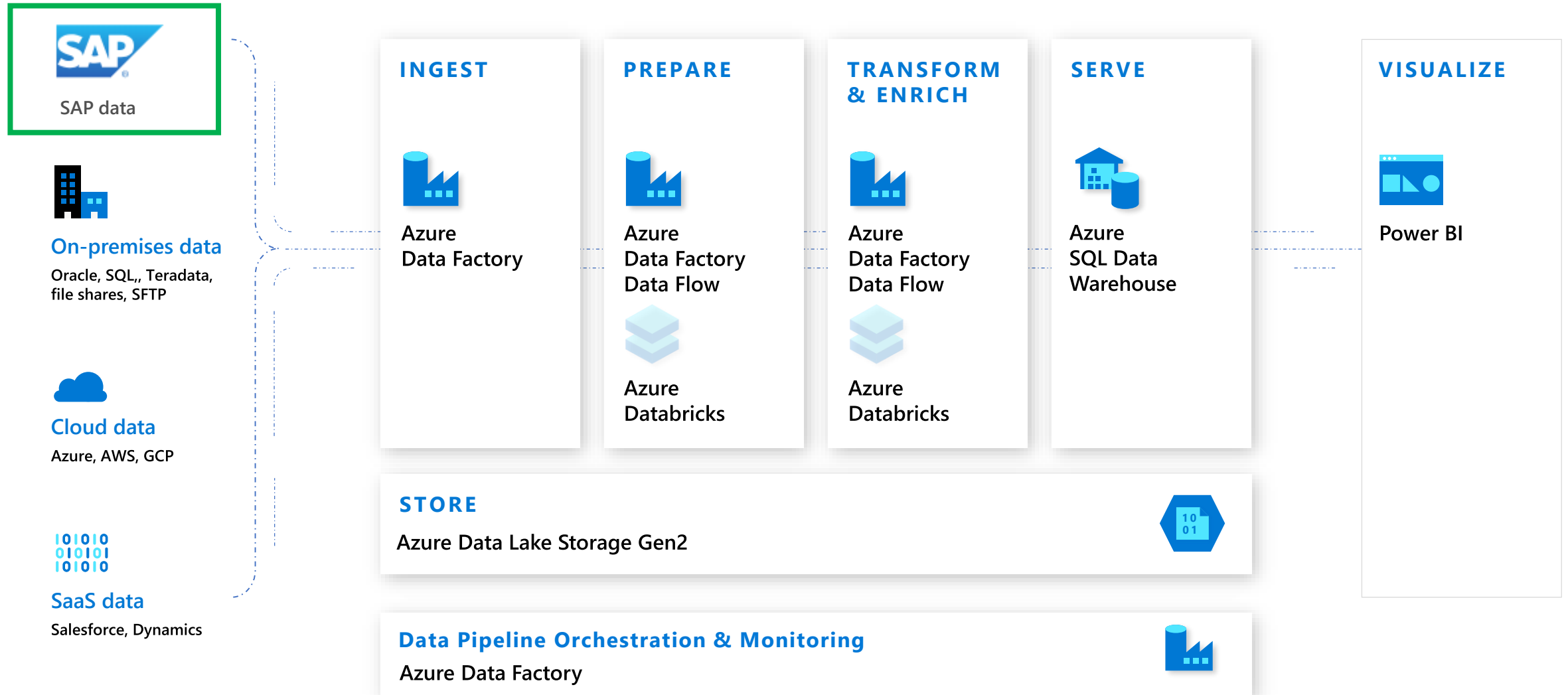


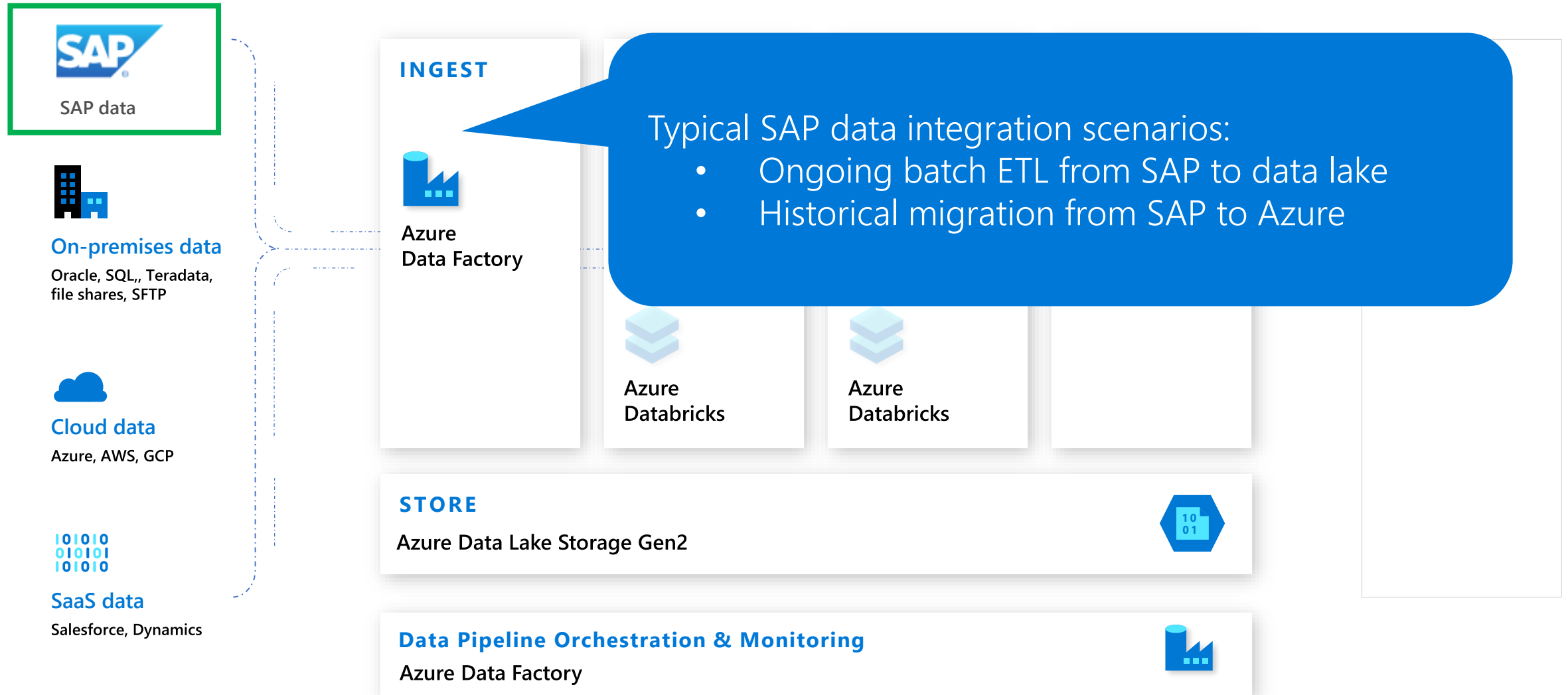
SAP Data Integration Using Azure Data Factory

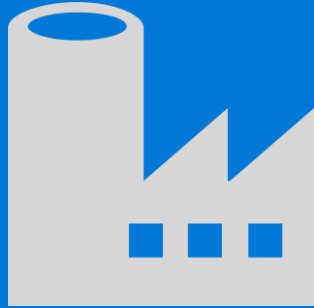
Update: Aug 30, 2019

Modern Data Warehouse



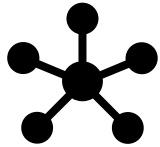
Modern Data Warehouse





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



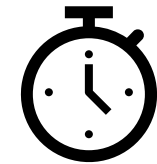
Scalable &
Cost-Effective

Serverless scalability
without infra mgmt
Pay for use



Secure &
Compliant

Certified compliance
Enterprise grade security
MSI and AKV support

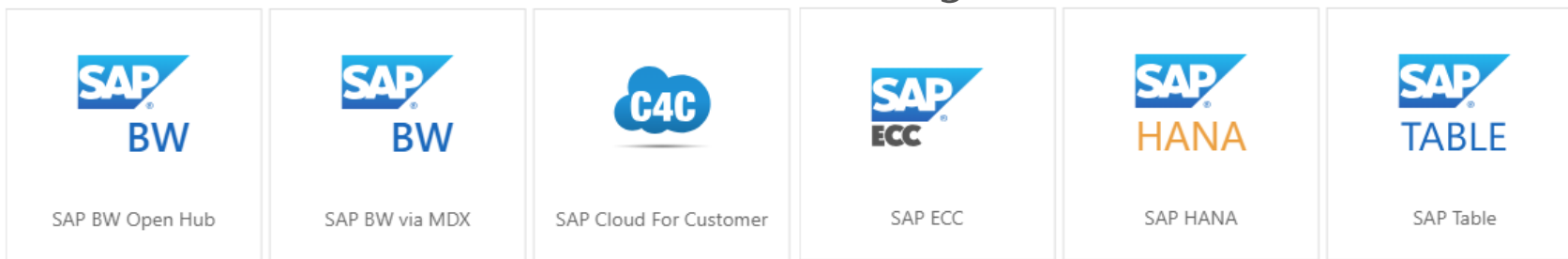


Productive

Drag & drop UI
Single-pane-of-glass
monitoring
CI/CD model

Access All Your Data (80+ connectors)

Single tool to enable data ingestion from SAP as well as other 70+ sources, and data transformation via built-in Data Flow, integration with Databricks/HDInsight/etc.



Azure (15)	Database & DW (26)		File Storage (6)	File Formats (5)	NoSQL (3)	Services & Apps (28)		Generic (4)
Blob Storage	Amazon Redshift	Oracle	Amazon S3	AVRO	Cassandra	Amazon MWS	Oracle Service Cloud	HTTP
Cosmos DB – SQL API	DB2	Phoenix	File System	Delimited Text	Couchbase	CDS for Apps	Paypal	OData
Cosmos DB – MongoDB API	Drill	PostgreSQL	FTP	JSON	MongoDB	Concur	QuickBooks	ODBC
ADLS Gen1	Google BigQuery	Presto	Google Cloud Storage	ORC		Dynamics 365	Salesforce	REST
ADLS Gen2	Greenplum	SAP BW Open Hub	HDFS	Parquet		Dynamics AX	SF Service Cloud	
Data Explorer	HBase	SAP BW MDX	SFTP			Dynamics CRM	SF Marketing Cloud	
Database for MariaDB	Hive	SAP HANA				Google AdWords	SAP C4C	
Database for MySQL	Impala	SAP Table				HubSpot	SAP ECC	
Database for PostgreSQL	Informix	Spark				Jira	ServiceNow	
File Storage	MariaDB	SQL Server				Magento	Shopify	
SQL Database	Microsoft Access	Sybase				Marketo	Square	
SQL Database MI	MySQL	Teradata				Office 365	Web Table	
SQL Data Warehouse	Netezza	Vertica				Oracle Eloqua	Xero	
Search Index						Oracle Responsys	Zoho	
Table Storage								

SAP HANA Integration

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




ADF connector:



([Connector deep-dive](#))

SAP BW Integration

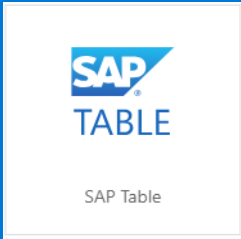
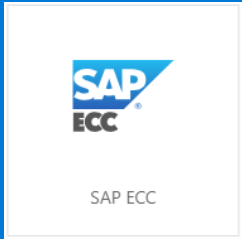
"I want to extract data from SAP BW" →

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

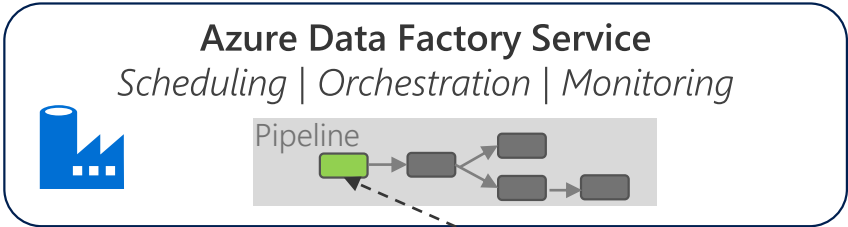
SAP HANA Connector – Supported Capabilities

Suitable scenario: ingest data from SAP HANA.

Supported versions	<ul style="list-style-type: none">• All SAP HANA versions• SAP HANA on-prem or in the cloud e.g. on Azure
Supported SAP objects	<ul style="list-style-type: none">• HANA Information Models (Analytic/Calculation views)• Row & Column Tables
Supported authentications	<ul style="list-style-type: none">• Basic – username & password• Windows – Single Sign-On via Kerberos-constrained delegation
Mechanism and prerequisites	<ul style="list-style-type: none">• Built on top of SAP's HANA ODBC driver• Run on Self-hosted Integration Runtime

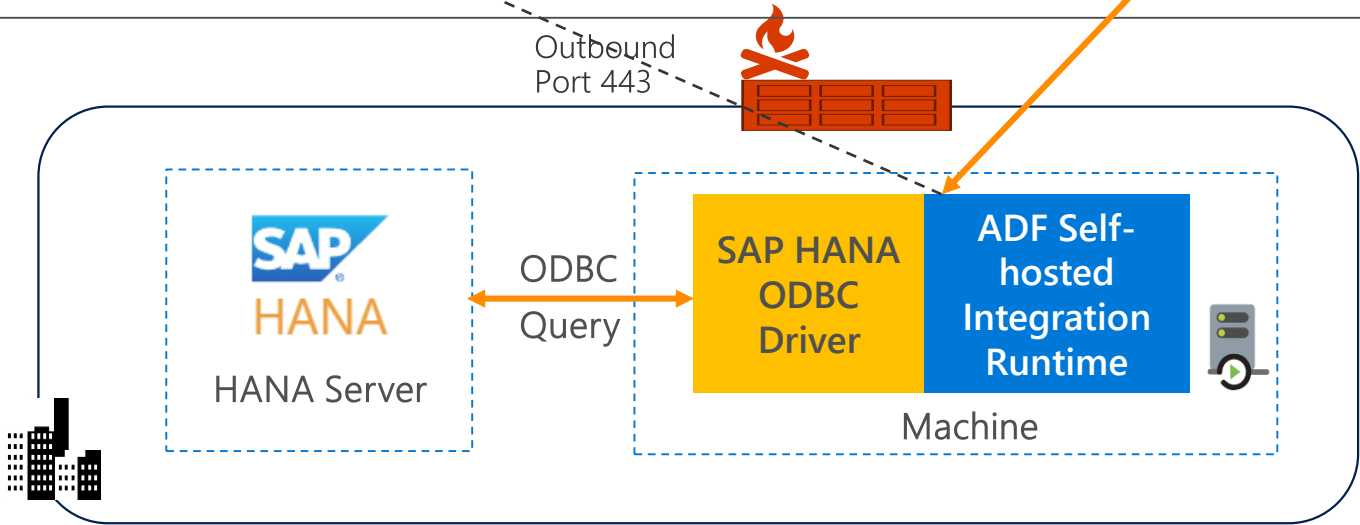
SAP HANA Connector – How It Works

←---→ Command and Control
↔ Data



Azure

On-prem or
Azure VNET



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

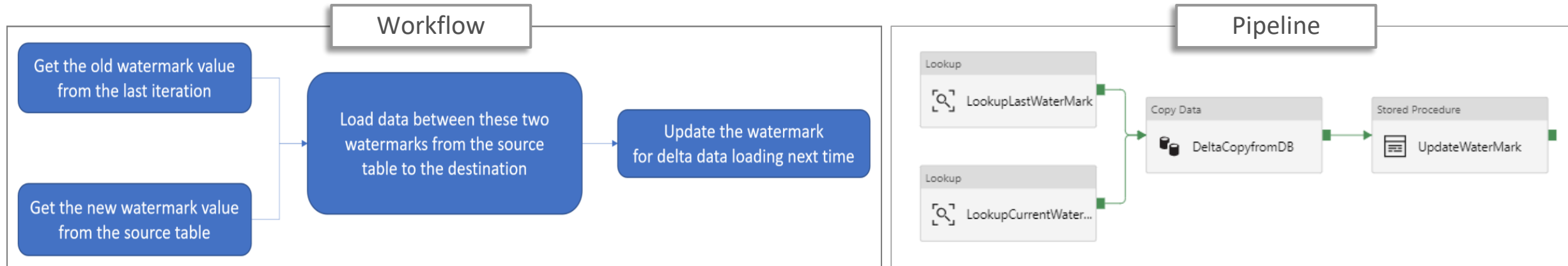
Execution start time: 2019/03/20 00:00:00 (window end time)
Delta extraction: last modified time between 2019/03/19 – 2019/03/20

SAP HANA Connector – Incremental Copy

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 HANA Connector – What's New

- Configurable connection options – all available ones in ODBC driver
- UI table/view navigation in addition to cube navigation
- Performance tuning option via batch size

SAP Table Connector

SAP Table Connector – Supported Capabilities

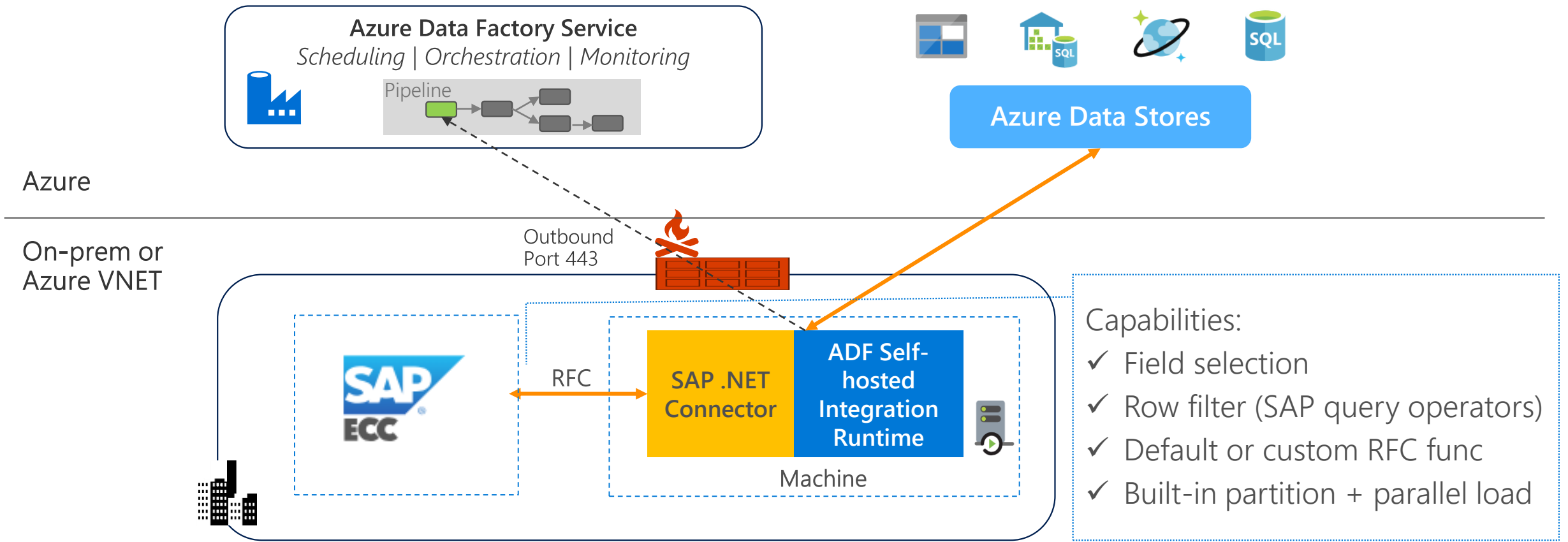
Suitable scenario: ingest data from SAP Table for SAP ECC, S/4 HANA, BW, or other application in Business Suite.

Supported versions	<ul style="list-style-type: none">• SAP ECC, BW, or other applications version 7.01 and above, on-prem or in the cloud e.g. on Azure• S/4 HANA
Supported server type	<ul style="list-style-type: none">• Connect to Application Server or Message Server
Supported SAP objects	<ul style="list-style-type: none">• SAP Transparent Table, Pooled Table, Cluster Table and View
Supported authentications	<ul style="list-style-type: none">• Basic – username & password• SNC (Secure Network Communications)
Performance	<ul style="list-style-type: none">• Built-in parallel loading option based on configurable data partitioning
Mechanism and prerequisites	<ul style="list-style-type: none">• Built on top of SAP .NET Connector 3.0, pull data via NetWeaver RFC• Run on Self-hosted Integration Runtime

SAP Table Connector – How It Works

←---→ Command and Control

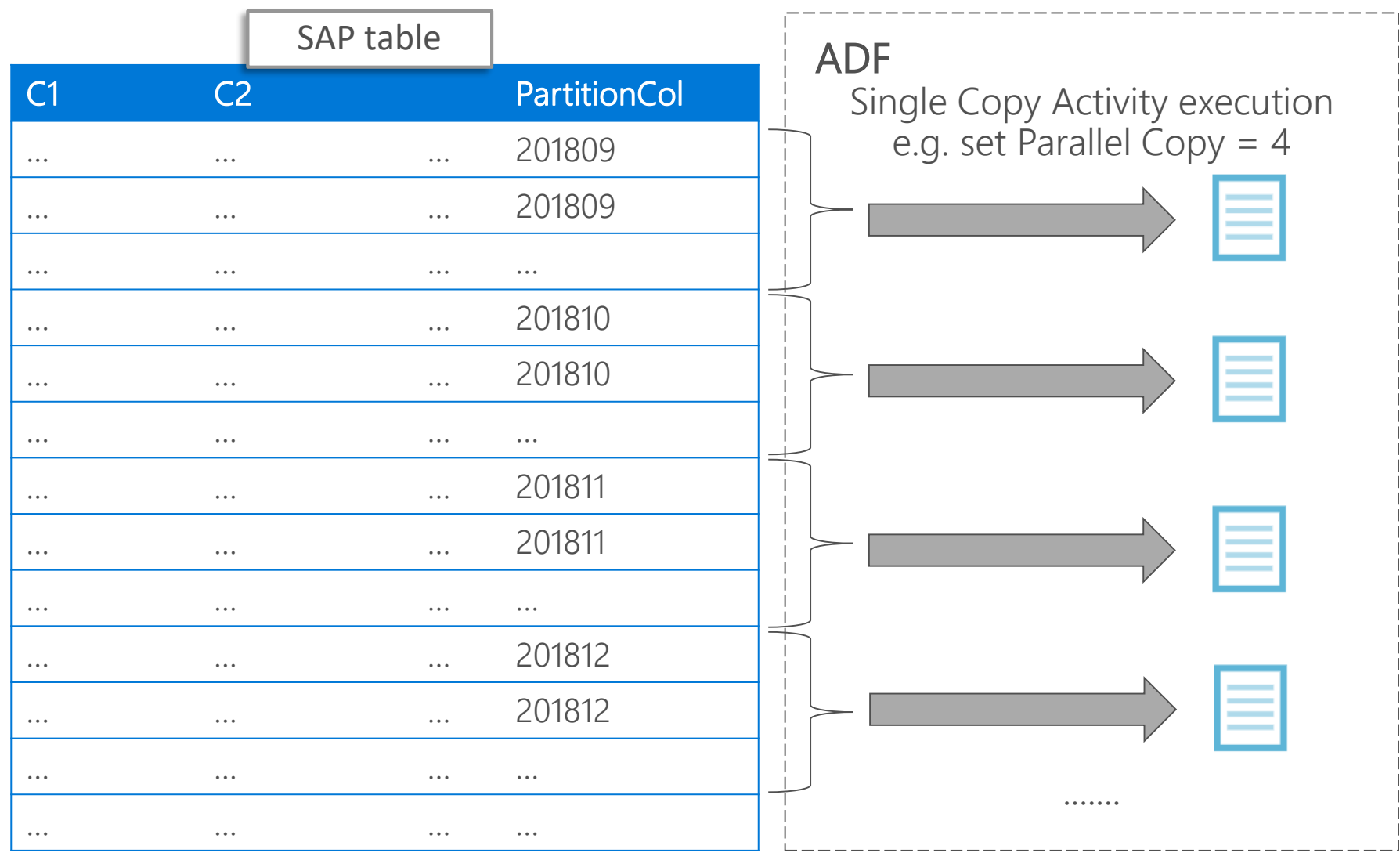
↔ Data



- 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



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

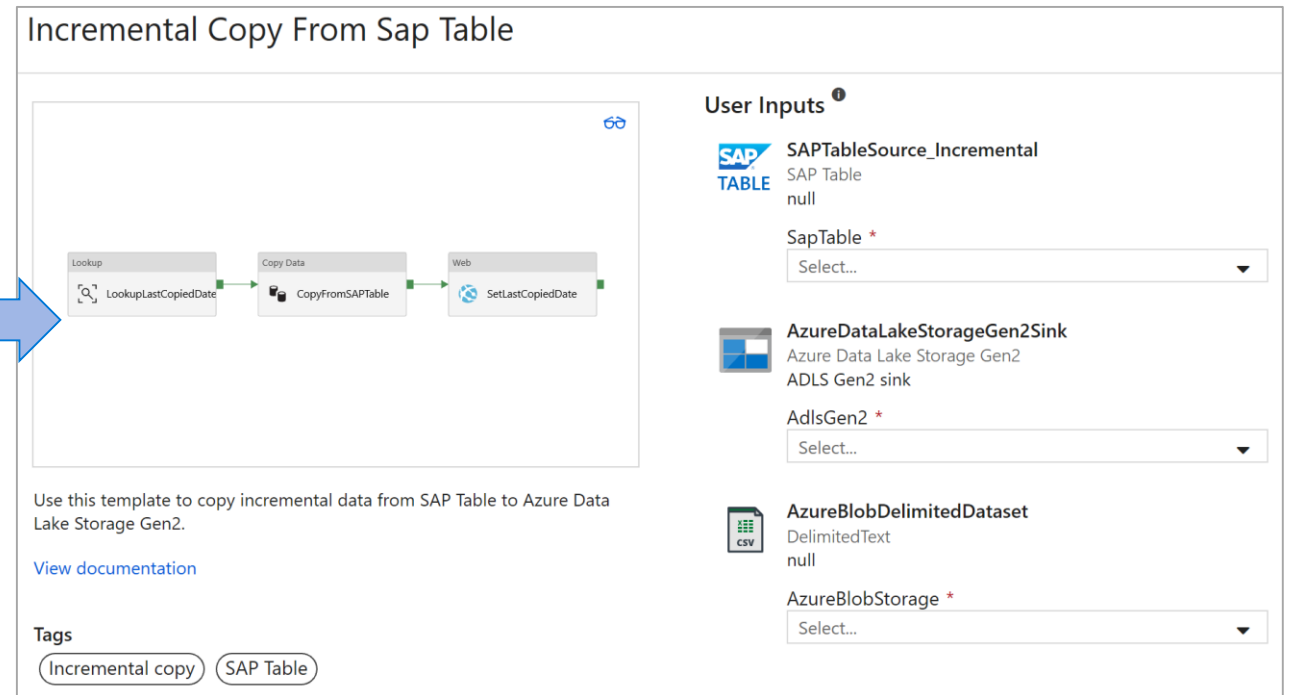
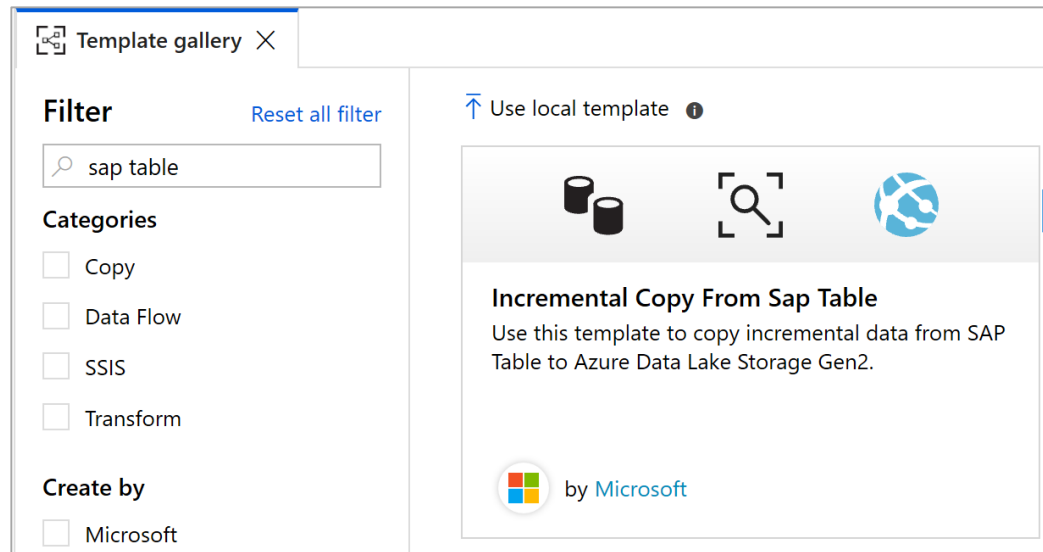
Pattern I: “my data has timestamp column e.g. calendar date”

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

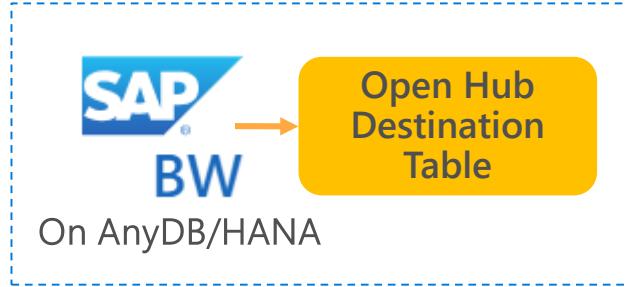
SAP BW Open Hub Connector – Supported Capabilities

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

Supported versions	<ul style="list-style-type: none">• SAP BW version 7.01 and above, on-prem or in the cloud e.g. on Azure
Supported server type	<ul style="list-style-type: none">• Connect to Application Server
Supported SAP objects	<ul style="list-style-type: none">• Open Hub Destination (OHD) local table• Underneath objects can be DSO, InfoCube, MultiProvider, DataSource etc.
Supported authentications	<ul style="list-style-type: none">• Basic – username & password
Performance	<ul style="list-style-type: none">• Built-in parallel loading option based on OHD specific schema
Mechanism and prerequisites	<ul style="list-style-type: none">• Built on top of SAP .NET Connector 3.0, pull data via NetWeaver RFC• Run on Self-hosted Integration Runtime• SAP side config: create SAP OHD in SAP BW to expose data

NOTE: SAP BW4/HANA is not supported now.

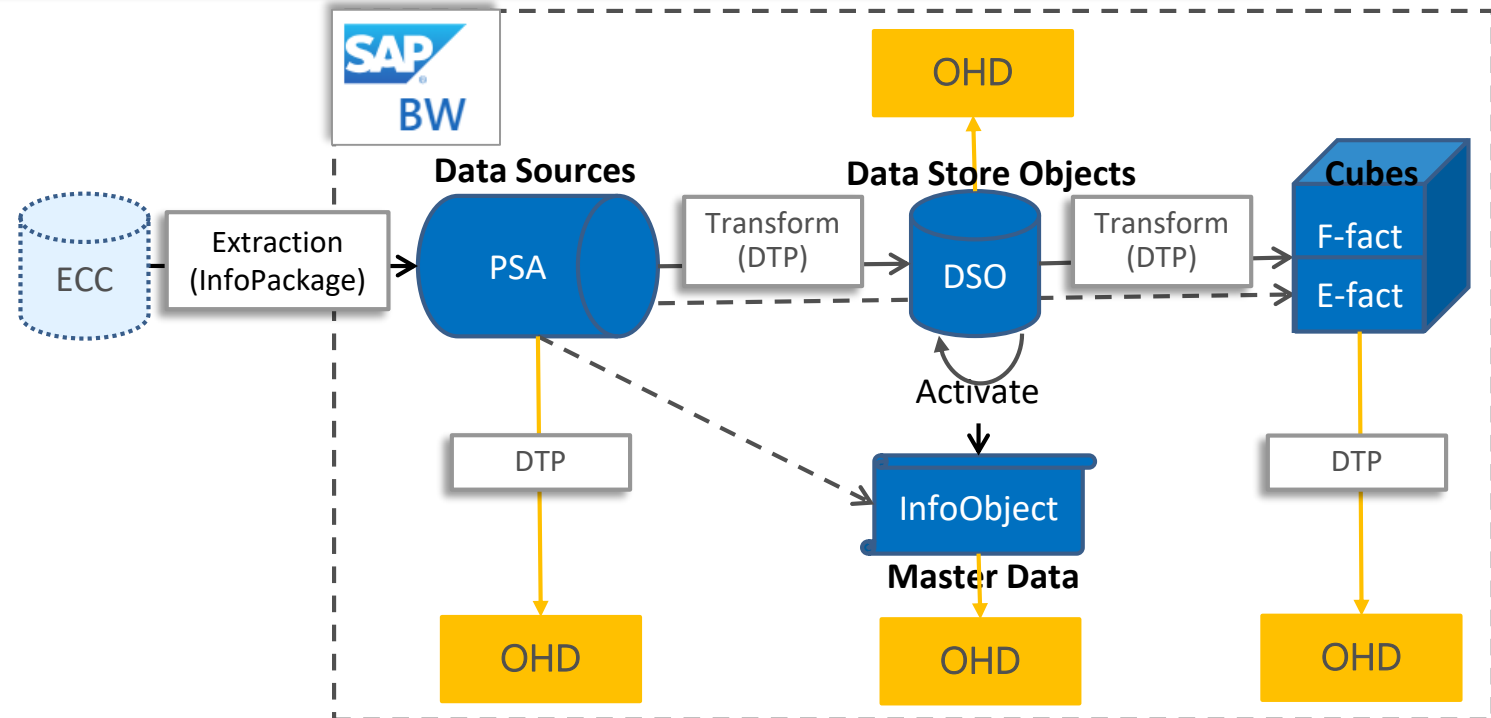
SAP BW Open Hub – How It Works



SAP Open Hub Destination (OHD):

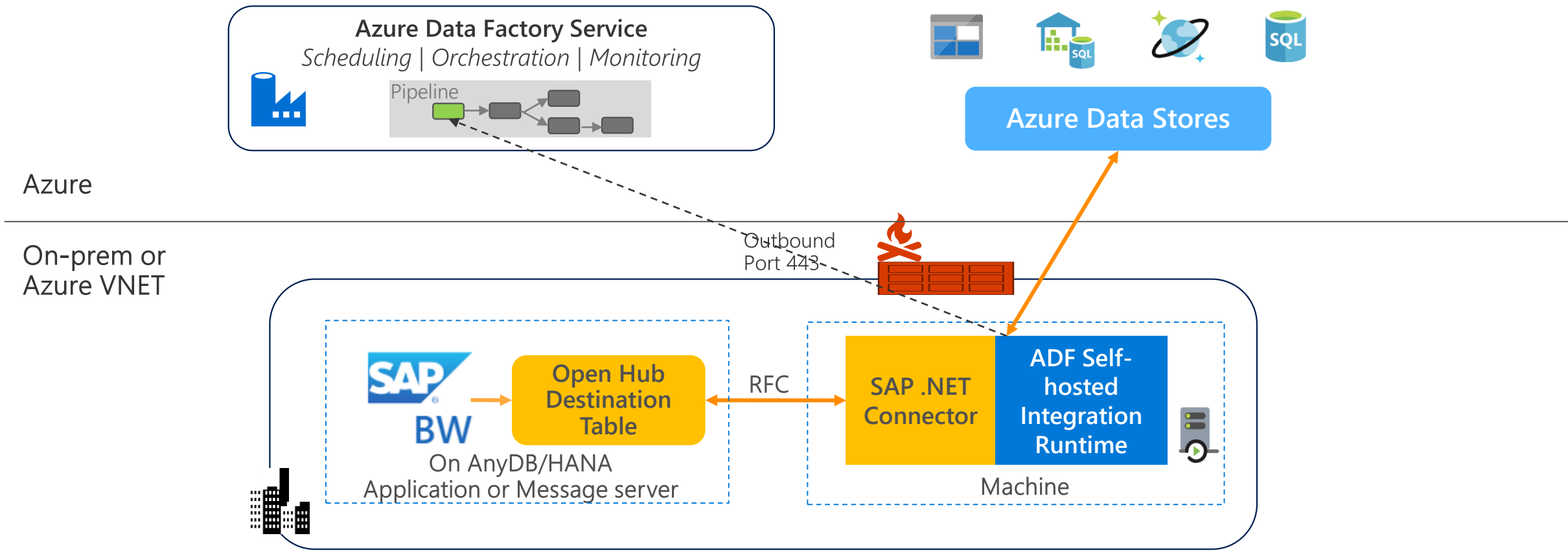
- **What is OHD:** defines the target to which the data is relayed.
- **Supported data:** any objects supported by SAP Data Transfer Process (DTP) can be used as open hub data sources.
- **OHD types:** database tables (local or remote) and flat files.

The connector support OHD local table in BW.

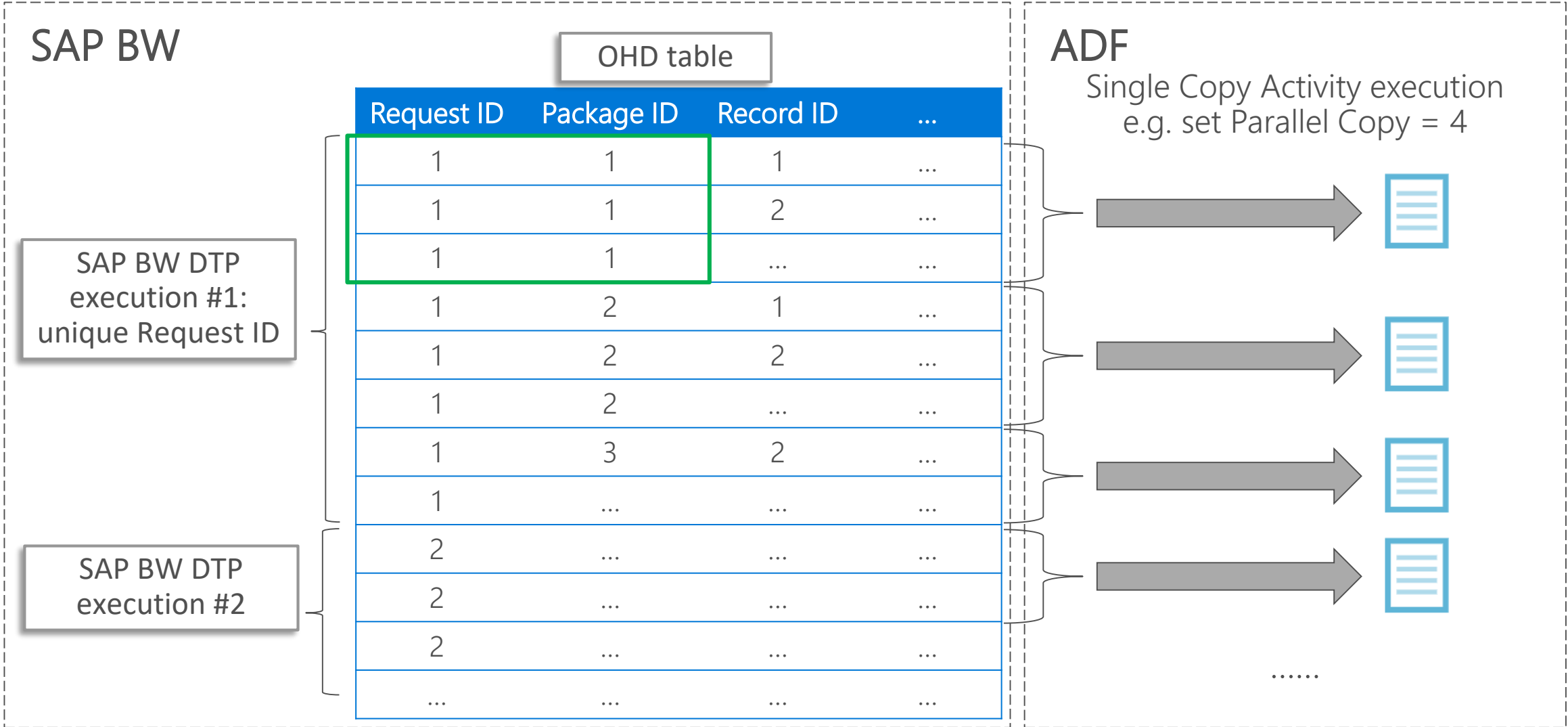


SAP BW Open Hub Connector – How It Works

←---→ Command and Control
↔ Data

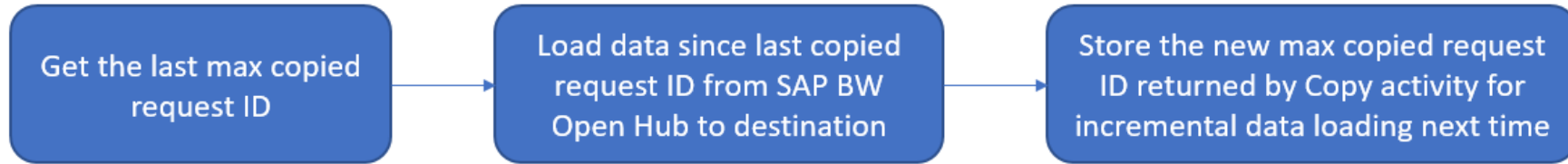


SAP BW Open Hub Connector – Built-in Parallel Loading



SAP BW Open Hub Connector – Incremental Copy

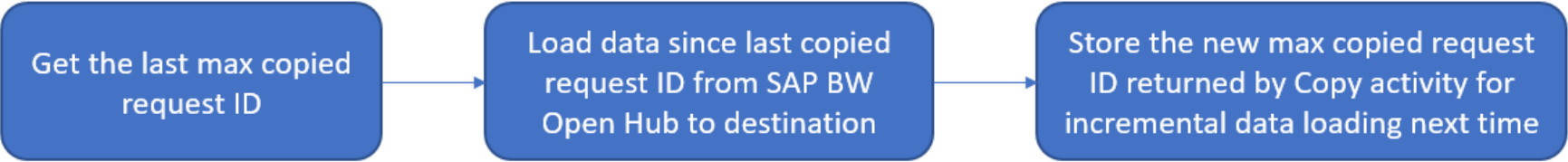
Solution: external control table/file + high watermark (max copied request ID).



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

SAP BW Open Hub Connector – Incremental Copy

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



Get started via solution template:

Template gallery

Filter

SAP BW

Create by

☒ Microsoft

☒ My Templates

Tag

☐ ADLS

☐ Amazon S3

☐ Azure Blob Storage

AllCopyData FlowSSIS

Incremental copy from SAP BW to Azure Data Lake Storage Gen2

Use this template to copy incremental data from SAP BW via Open Hub to Azure Data Lake Storage Gen2.

by Microsoft

Incremental copy from SAP BW to Azure Data Lake Storage Gen2

Use this template to copy incremental data from SAP BW via Open Hub to Azure Data Lake Storage Gen2.

[View documentation](#)

Tags

Incremental copySAP BWAzure Data Lake StorageADLS

User Inputs

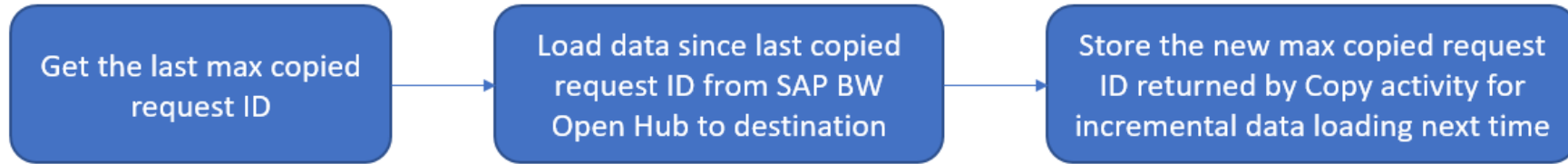
AzureBlobDataset
Azure Blob Storage
Blob to store the max copied request ID from SAP BW Open Hub table as high watermark
Azure Blob Storage Connection *
Select...

SAPOHDSrc_Incremental
SAP BW Open Hub
SAP BW Open Hub table source
SAP BW Open Hub Connection *
Select...

AzureDataLakeStorageGen2Sink
Azure Data Lake Storage Gen2
ADLS Gen2 sink
Azure Data Lake Storage Gen2 Connection *
Select...

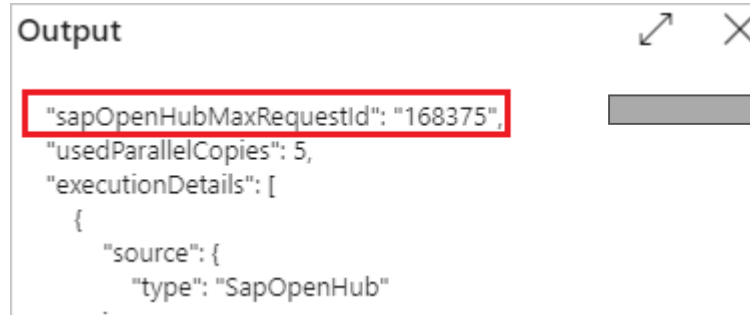
SAP BW Open Hub Connector – Incremental Copy

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.

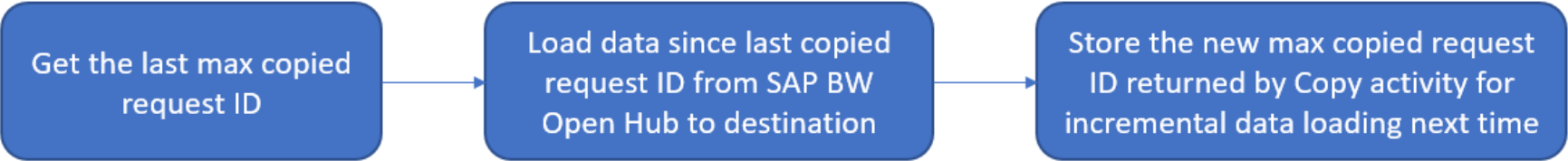
Copy activity output:



Save and use as
baseRequestId in next run

SAP BW Open Hub Connector – Incremental Copy

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 BW via MDX Connector

SAP BW via MDX Connector – Supported Capabilities

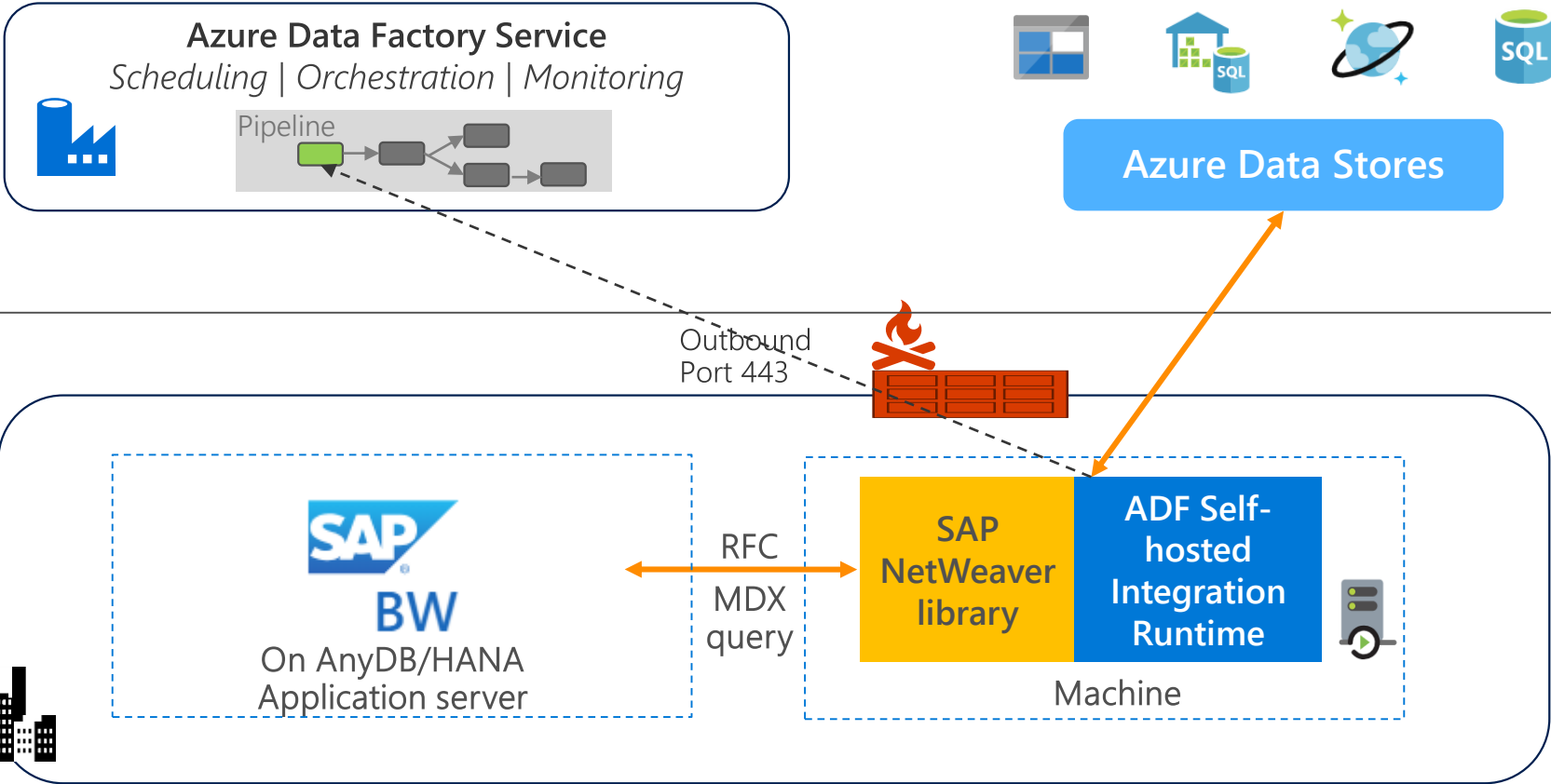
Suitable scenario: ingest data from SAP BW, with exploratory use case.

Supported versions	<ul style="list-style-type: none">• SAP BW version 7.x, on-prem or in the cloud e.g. on Azure
Supported server type	<ul style="list-style-type: none">• Connect to Application Server
Supported SAP objects	<ul style="list-style-type: none">• InfoCubes and QueryCubes (including BEx queries)
Supported authentications	<ul style="list-style-type: none">• Basic – username & password
Mechanism and prerequisites	<ul style="list-style-type: none">• 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

←---→ Command and Control
↔ Data



SAP ECC Connector

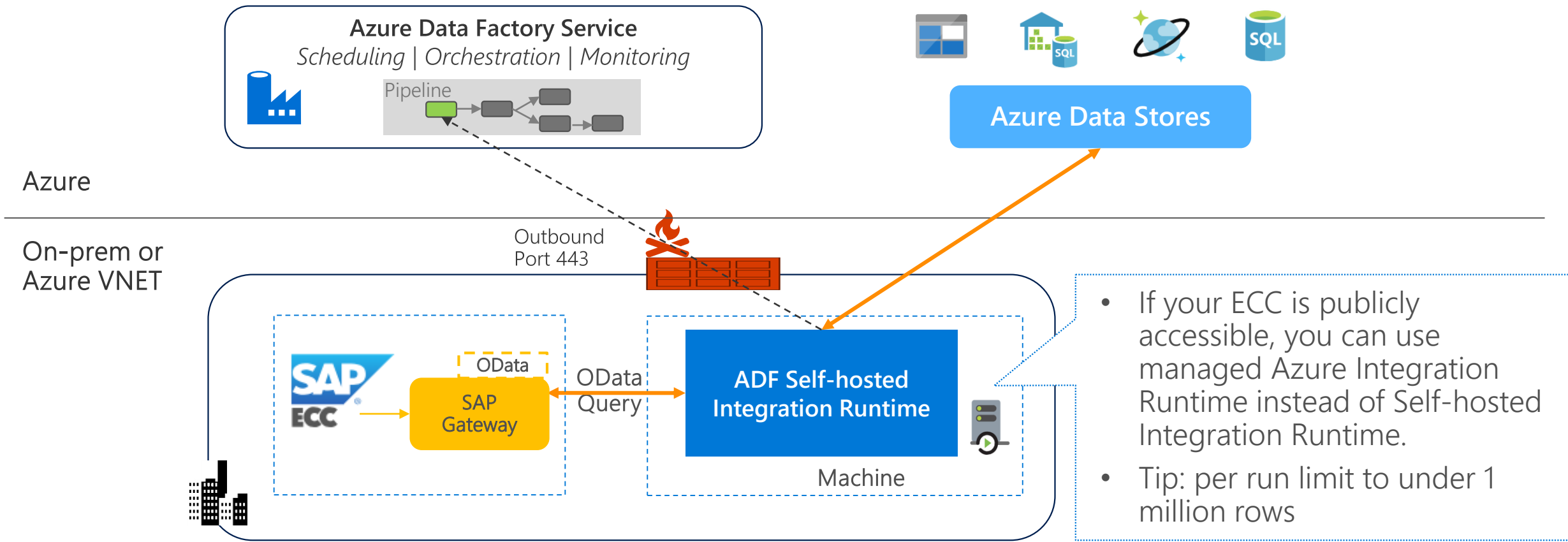
SAP ECC Connector – Supported Capabilities

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

Supported versions	<ul style="list-style-type: none">• SAP ECC version 7.0 and above• Any entities exposed by SAP ECC OData services
Supported SAP objects	<ul style="list-style-type: none">• Entities exposed by SAP OData services• BAPI, ODP (DataExtractors/DataSource), etc.
Supported authentications	<ul style="list-style-type: none">• Basic – user name & password
Mechanism and prerequisites	<ul style="list-style-type: none">• 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



SAP ECC Connector – Incremental Copy

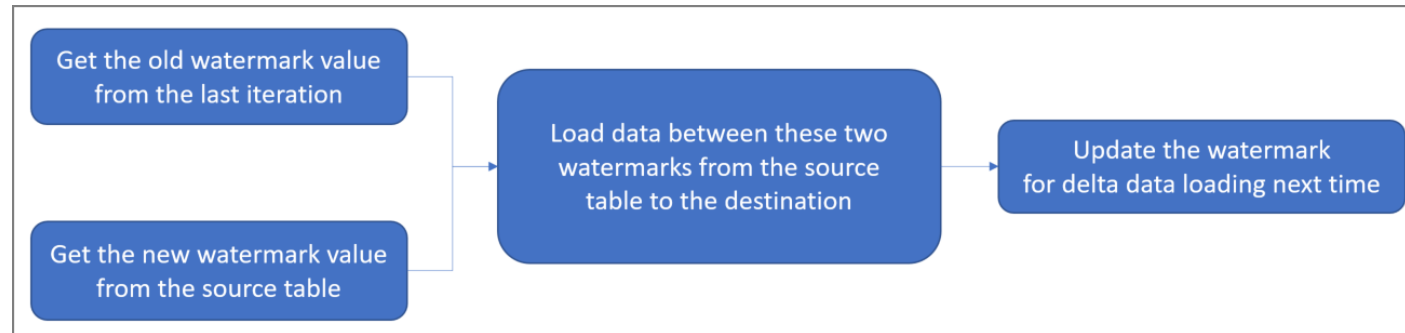
(in general, same as HANA in earlier slides)

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.

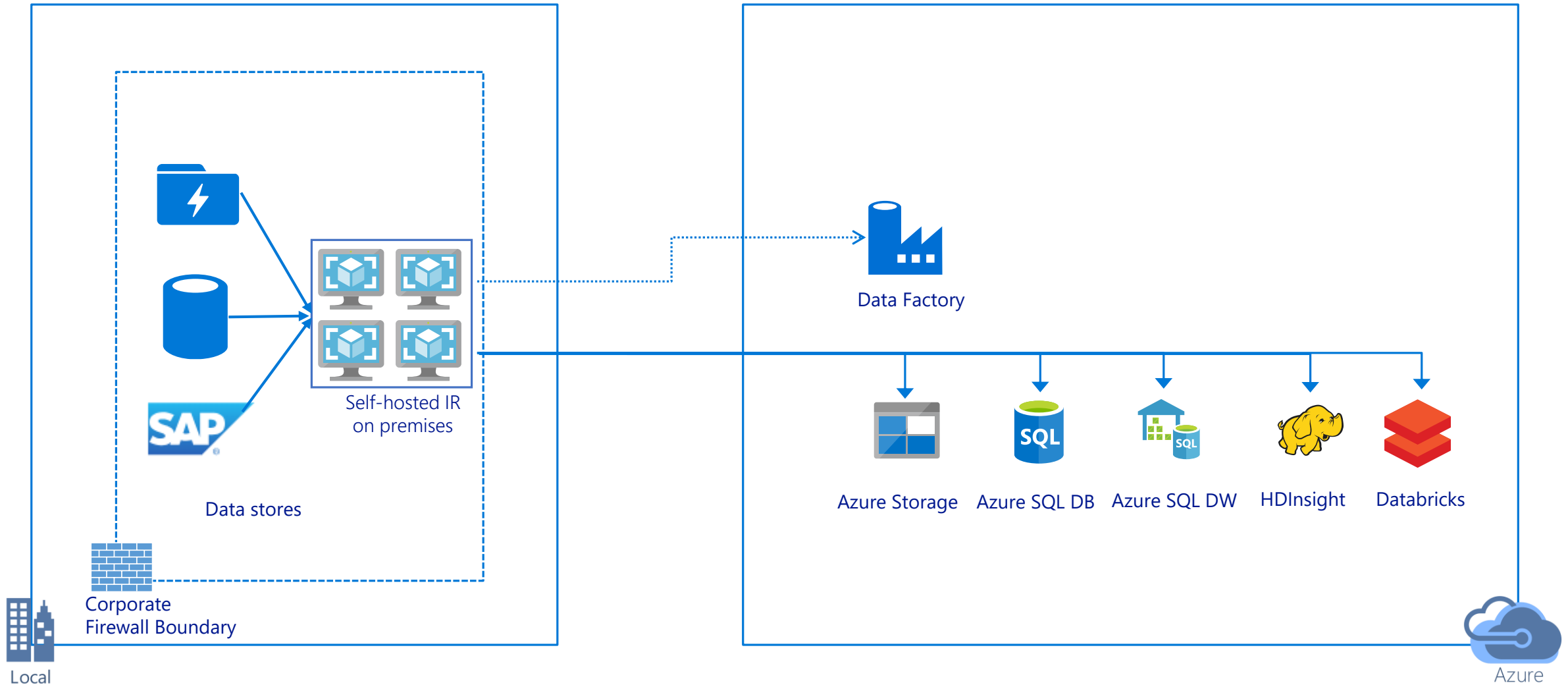


Pattern III: “my data is small in size as dimension data”

Solution: full copy and overwrite

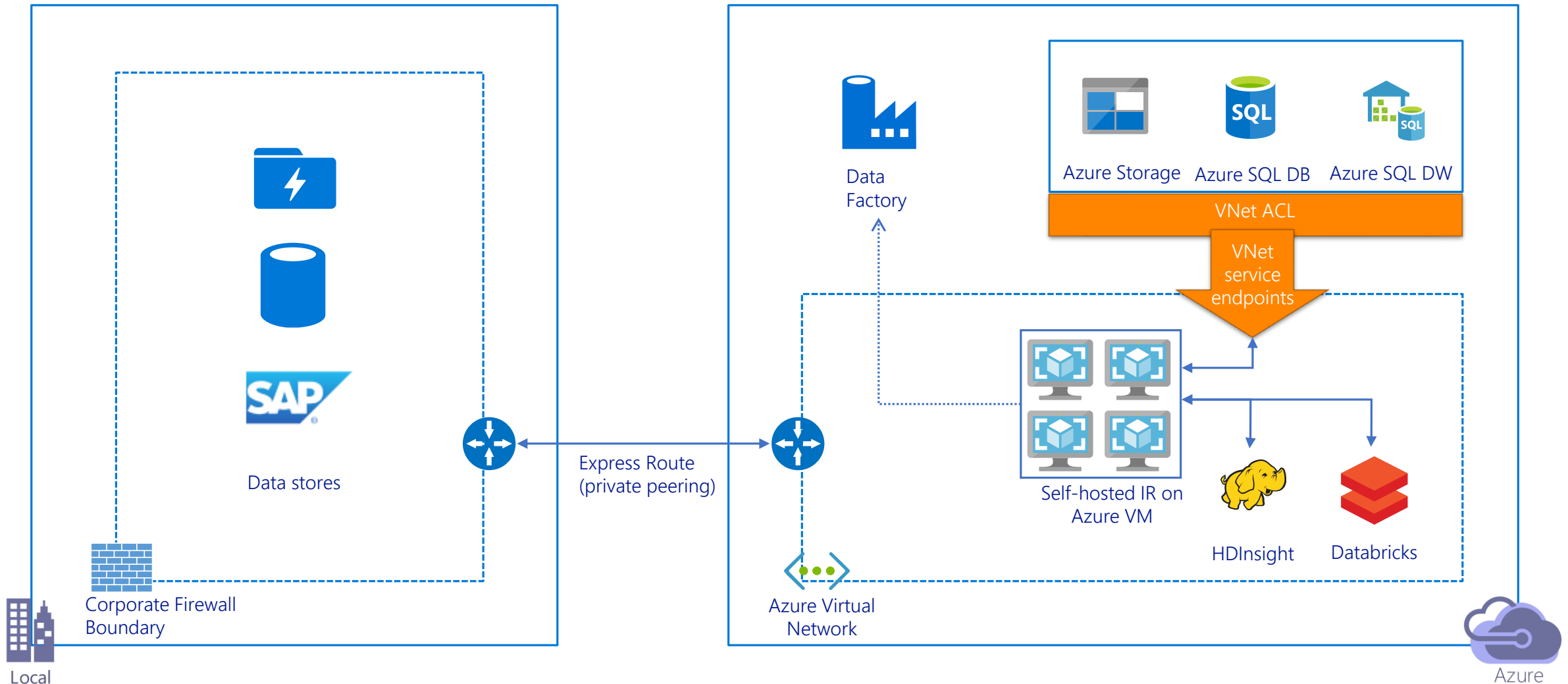
On-prem connected to Azure through public internet

Self-hosted IR deployed on premises



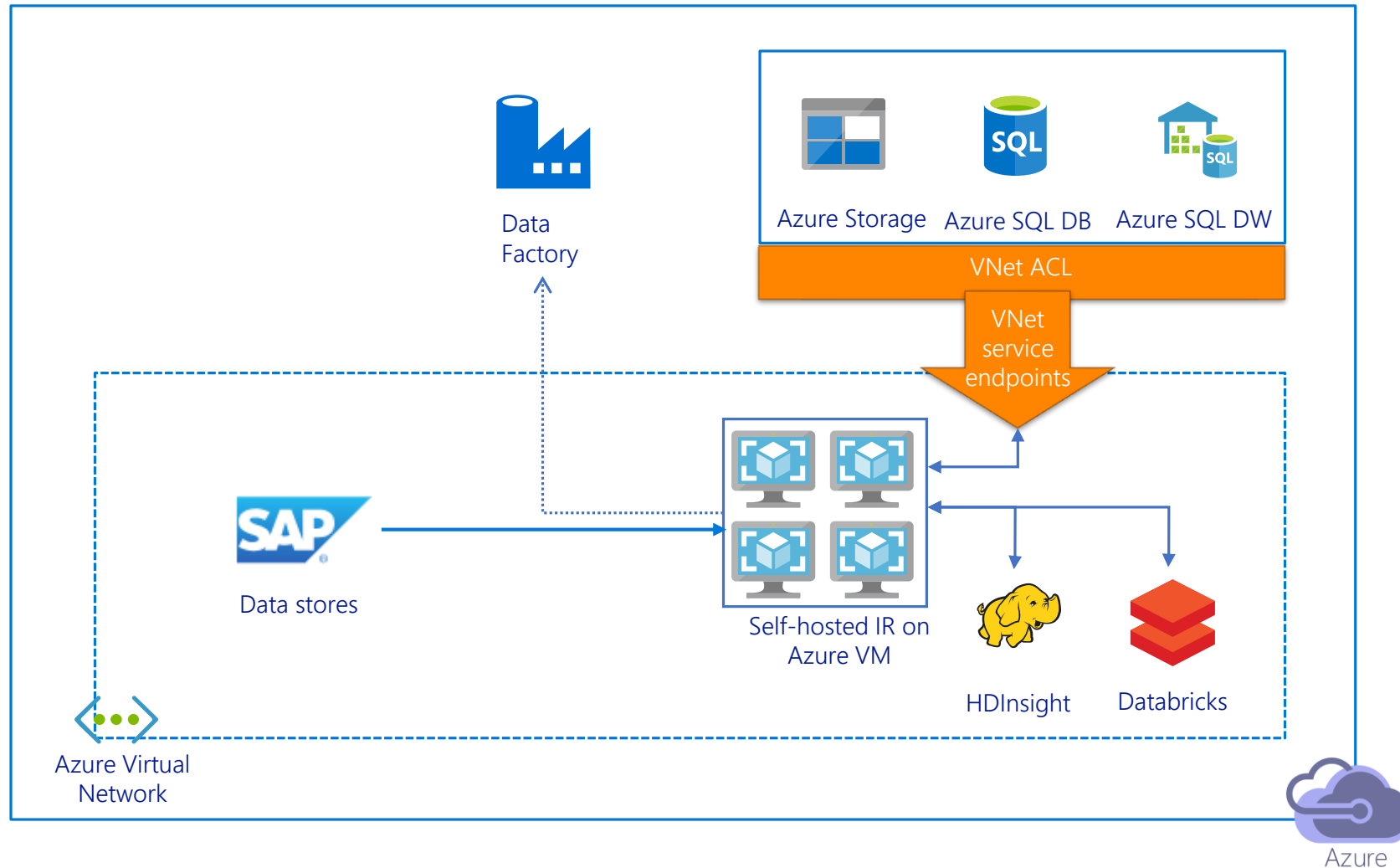
On-prem connected to Azure VNet through ExpressRoute

Self-hosted IR deployed on Azure VM

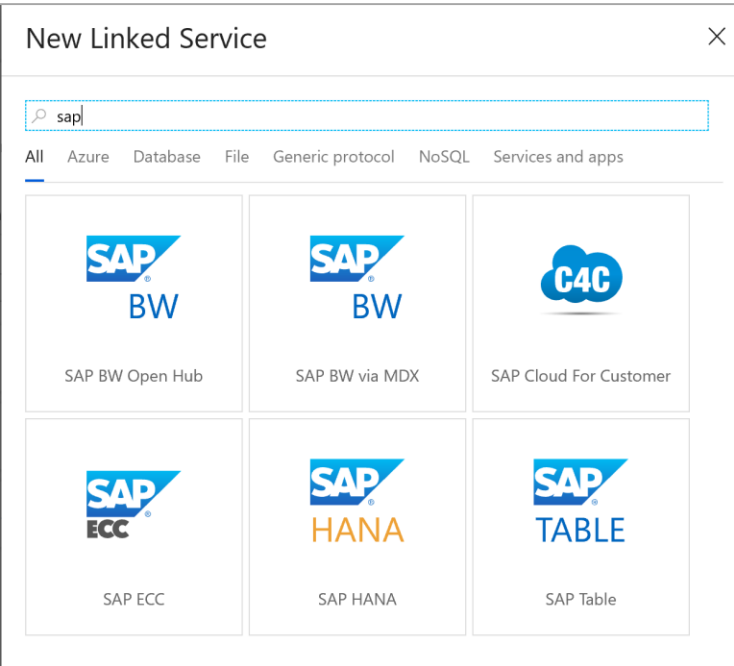
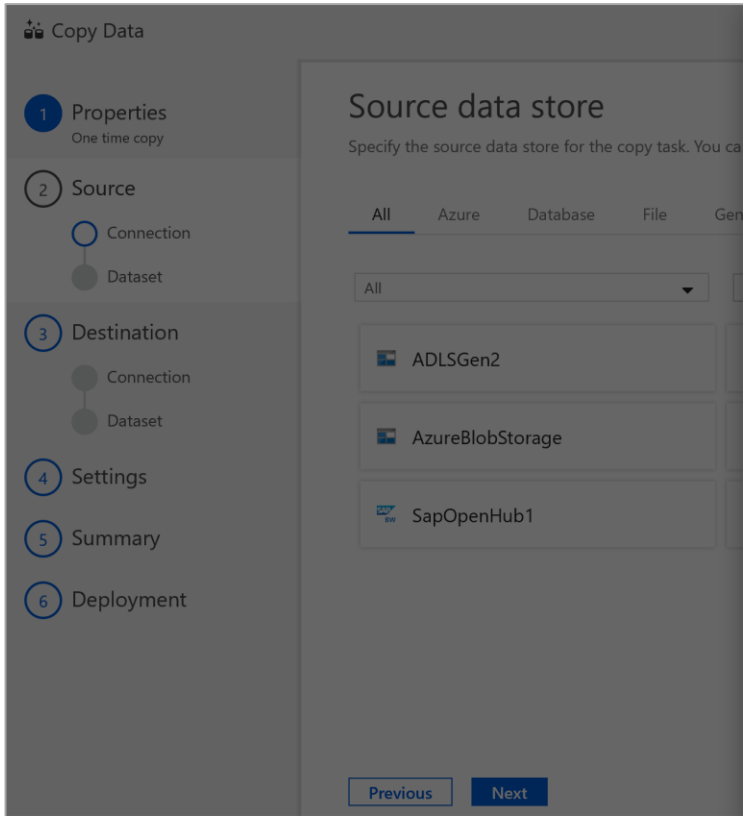


SAP on Azure

Self-hosted IR deployed on Azure VM






Get Started







Copy Data Tool

Solution Template




Incremental copy from SAP BW to Azure Data Lake Storage Gen2
Use this template to copy incremental data from SAP BW via Open Hub to Azure Data Lake Storage Gen2.

 by Microsoft



Incremental Copy From Sap Table
Use this template to copy incremental data from SAP Table to Azure Data Lake Storage Gen2.

 by Microsoft

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	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