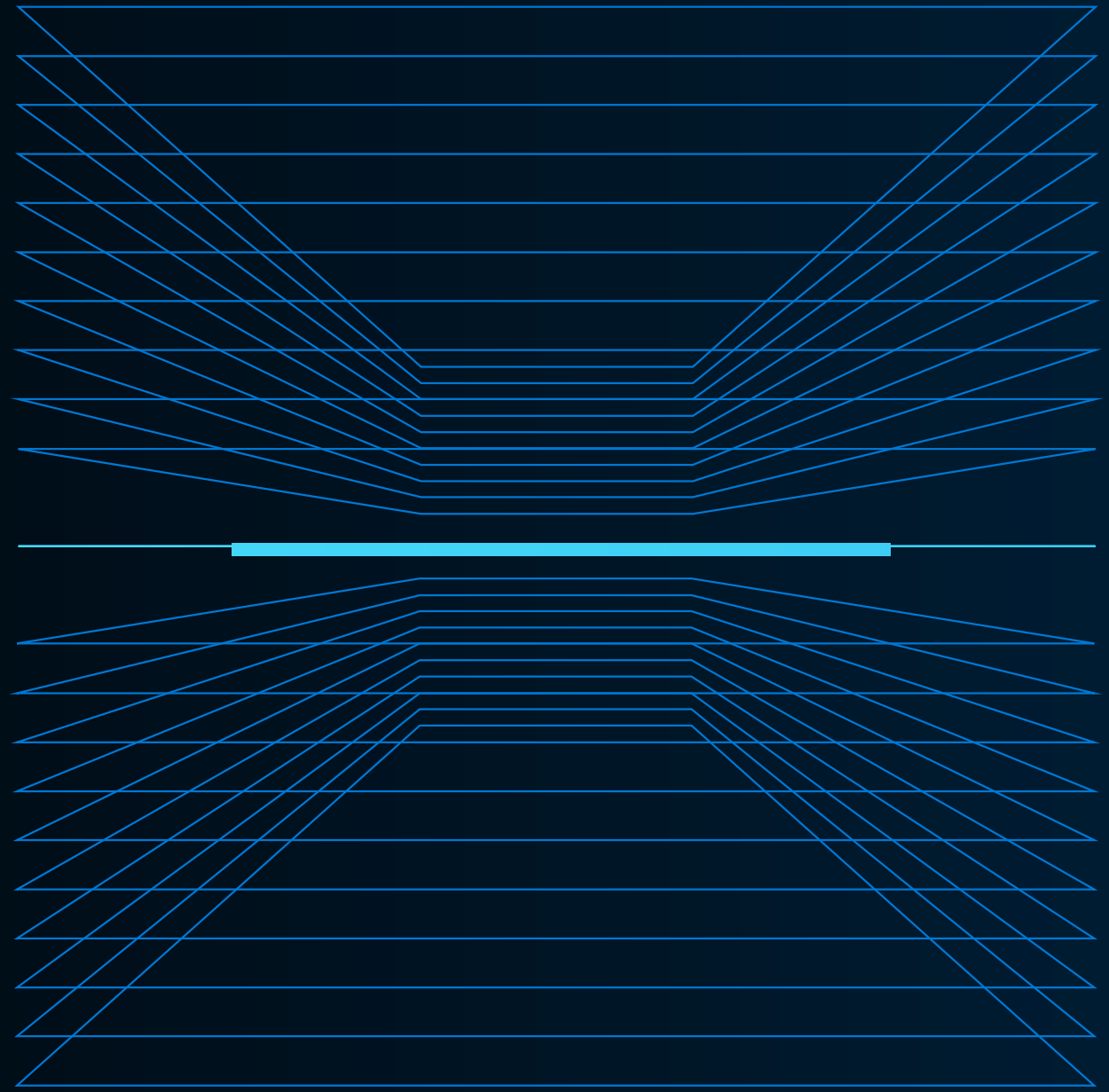
Microsoft

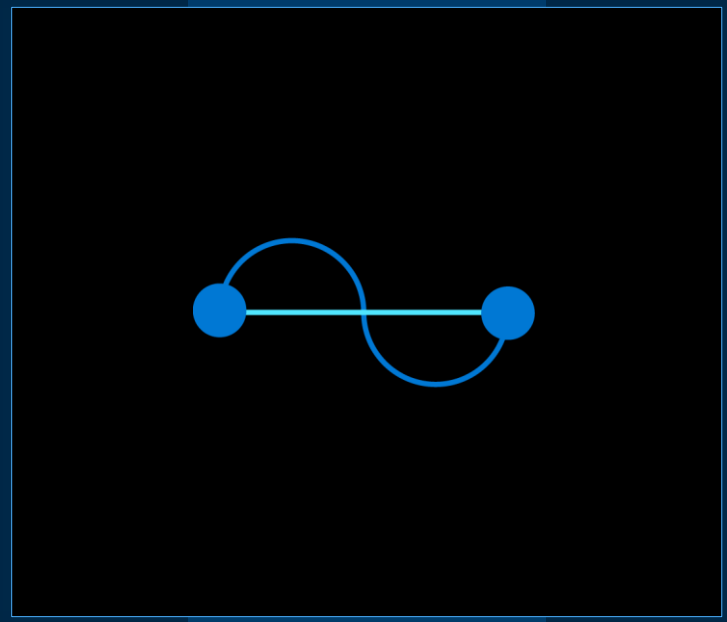
SAP Data Integration in Azure

Whitepaper



Updated: Nov 2022

SAP Data Integration



Azure offers native connectors within Data Factory/Synapse to integrate SAP Data

Generally Available

New Connector based on SAP's ODP Framework

Supports all SAP systems - ECC, S/4HANA, BW, BW/4HANA, HANA etc., irrespective of location (On-Prem, Multi-Cloud, RISE, etc.)

Which ADF/Synapse native connector do I use?

"I want to bring my SAP data into Azure "– Which source? What's your scenario?

ADF enables **data ingestion** from the following SAP sources:

Data Source\Scenario	Full Loads/Manual Delta	CDC
I have my data in **SAP HANA**	SAP HANA	SAP CDC
I have my data in **SAP BW**	SAP Table, SAP BW Open Hub, SAP BW via MDX	SAP CDC
I have my data in **SAP ECC, S/4 HANA**	SAP Table, SAP ECC	SAP CDC

SAP BW Integration

"I want to extract data from SAP BW or BW/4HANA"

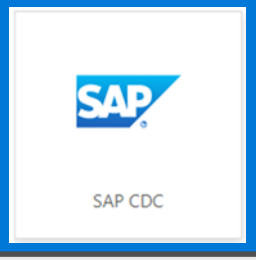
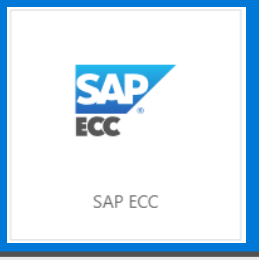
Suggested decision direction

ADF connector options	SAP CDC 	SAP Table 	SAP BW Open Hub 	SAP BW via MDX 
Objects to extract	DSO, ADSO, Other InfoProviders (InfoCube, MultiProvider, Query, etc.)	Table (Transparent, Pooled, Cluster Table) and View	DSO, InfoCube, MultiProvider, DataSource, etc	InfoCubes, QueryCubes
CDC Support	Built-in via ODP	Manual via Watermarking	Manual via Watermarking or within Open Hub	N/A
SAP side configuration	Only for DataSources/BW Extractors: Activation needed For other object types: N/A	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 configurable partitioning	Fast w/ built-in parallel loading based on OHD specific schema	Slower
Suitable workload	Large volume	Large volume	Well-thought-through workload Large volume	Exploratory workload Small volume
Support for SAP BW4/HANA	Yes	No	No	No

SAP ECC, S/4 HANA, SAP Application Integration

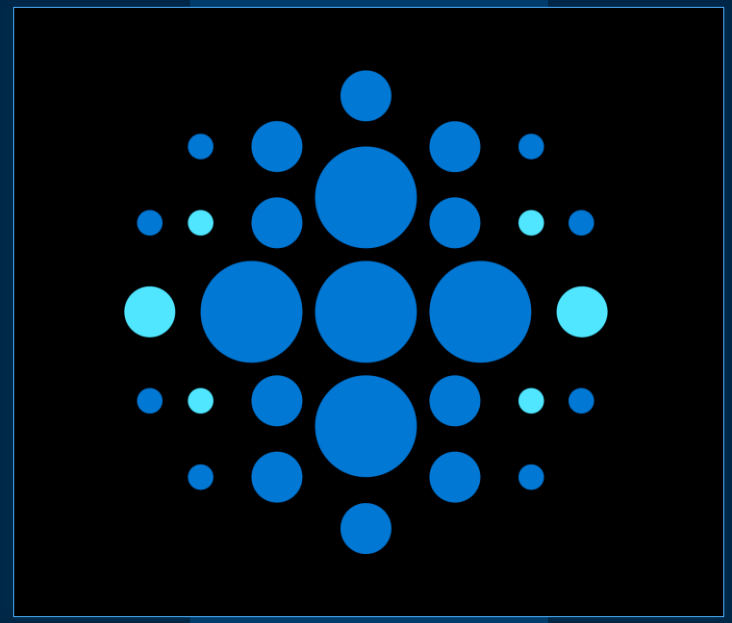
"I want to extract data from SAP ECC, S/4 HANA or other SAP applications"

Suggested decision direction

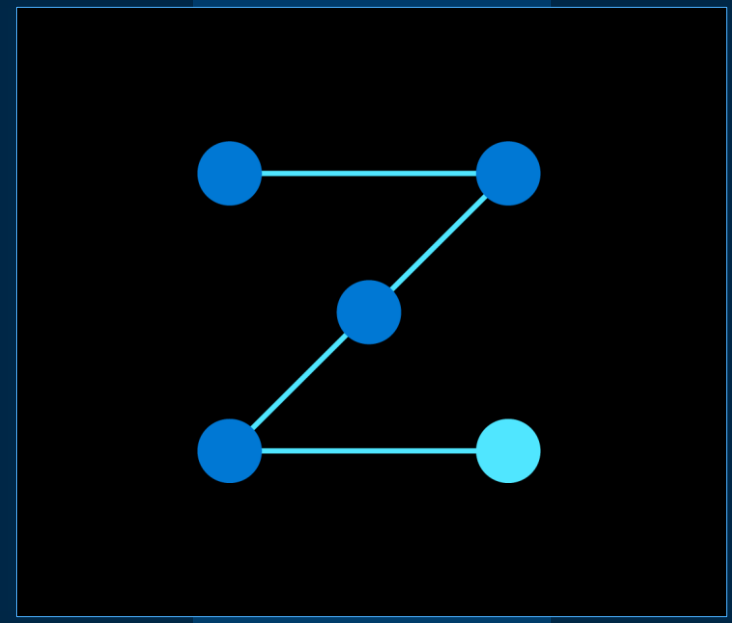
ADF connector options	SAP CDC 	SAP Table 	SAP ECC 
Objects to extract	Table (Transparent, Pooled, Cluster Table) and View, ECC Extractor, S4 CDS View	Table (Transparent, Pooled, Cluster Table) and View	OData entities exposed via SAP Gateway (BAPI, ODP)
CDC Support	Built-in via ODP	Manual via Watermarking	Possible to automate, but complicated
SAP side configuration	SLT is needed for extracting Tables N/A for Extractors, CDS Views, etc.	N/A	SAP Gateway
Performance	Fast w/ built-in parallel loading	Fast w/ built-in parallel loading	Slower
Suitable workload	Large volume	Large volume	Small volume

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

ADF/Synapse Connectors Deep-Dive



SAP CDC Connector

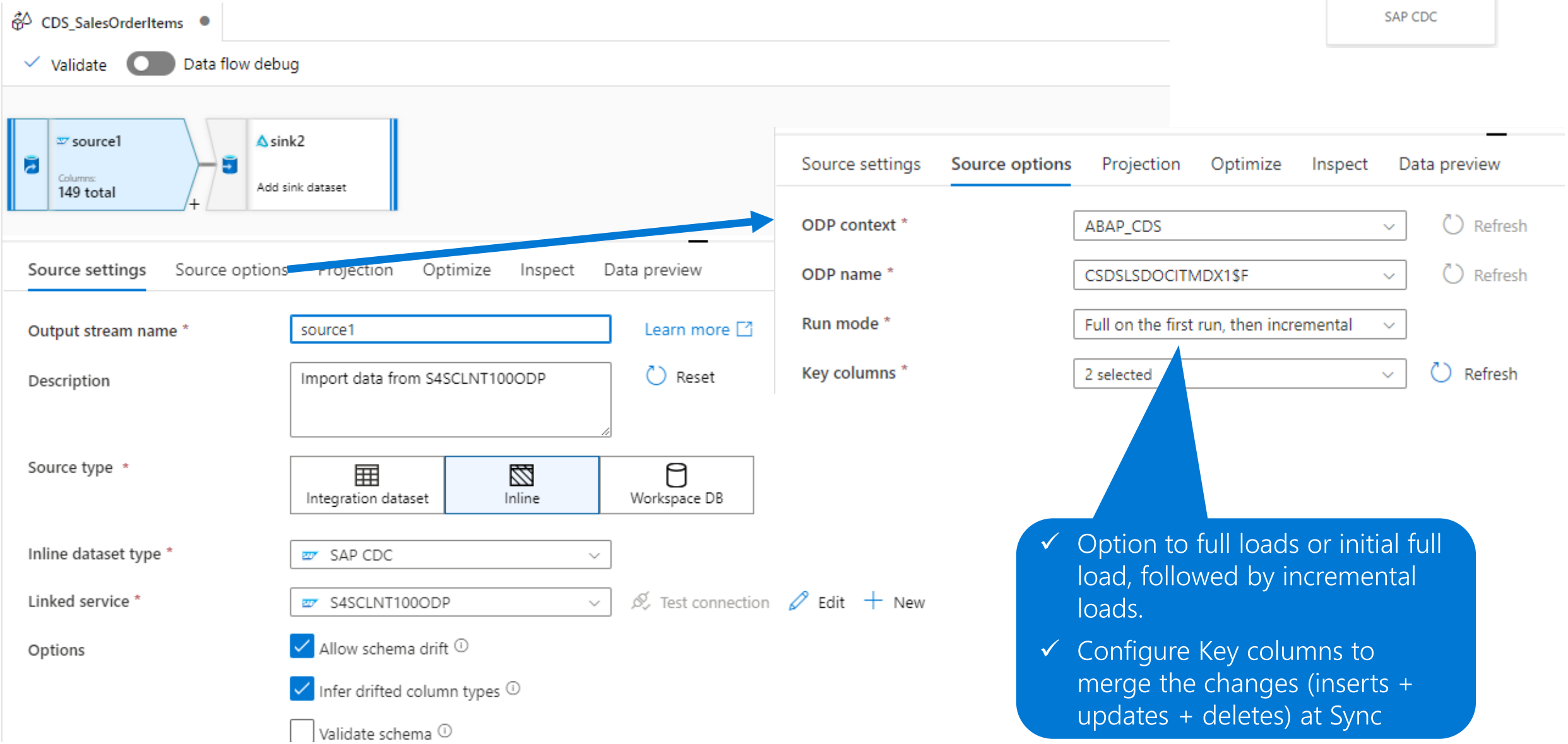


SAP CDC Connector

Suitable scenario: ingest CDC data from all **SAP Systems via ODP**

Supported versions	• **S/4 HANA, SAP ECC or other applications in Business Suite version 7.01 and above, on-prem or in the cloud** • **SAP BW/4HANA, SAP BW version 7.01 and above**, on-prem or in the cloud • **All SAP HANA versions** on-prem or in the cloud
Supported SAP objects	• Data Sources/Extractors, ABAP CDS Views, BW and BW/4HANA InfoObjects and InfoProviders, Tables via SLT as a proxy
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** • Run on ADF Self-hosted Integration Runtime • SAP side config: Data Sources/Extractors may need to be activated within ODP; SLT is needed for Tables
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 CDC Connector



SAP CDC

- ✓ Option to full loads or initial full load, followed by incremental loads.
- ✓ Configure Key columns to merge the changes (inserts + updates + deletes) at Sync

SAP CDC Connector – How It Works

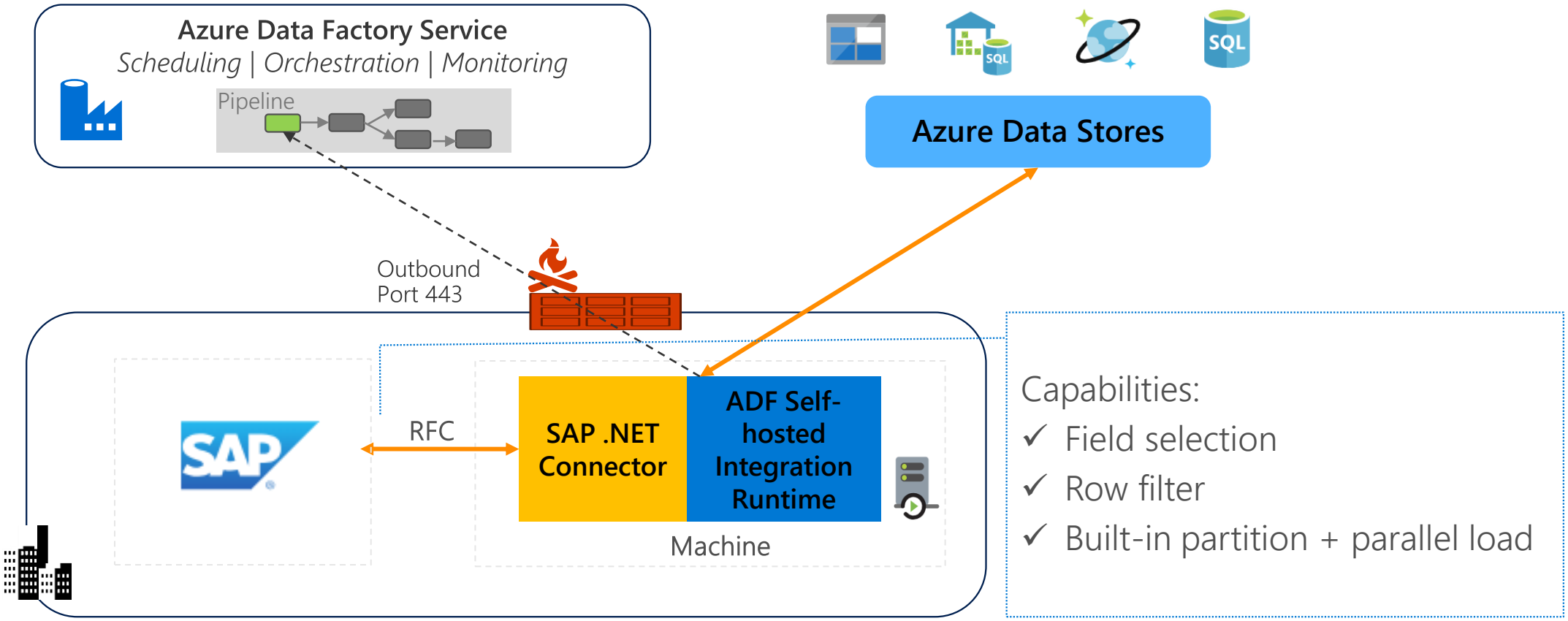
SAP CDC

Command and Control

Data

Azure

On-prem or Azure VNET

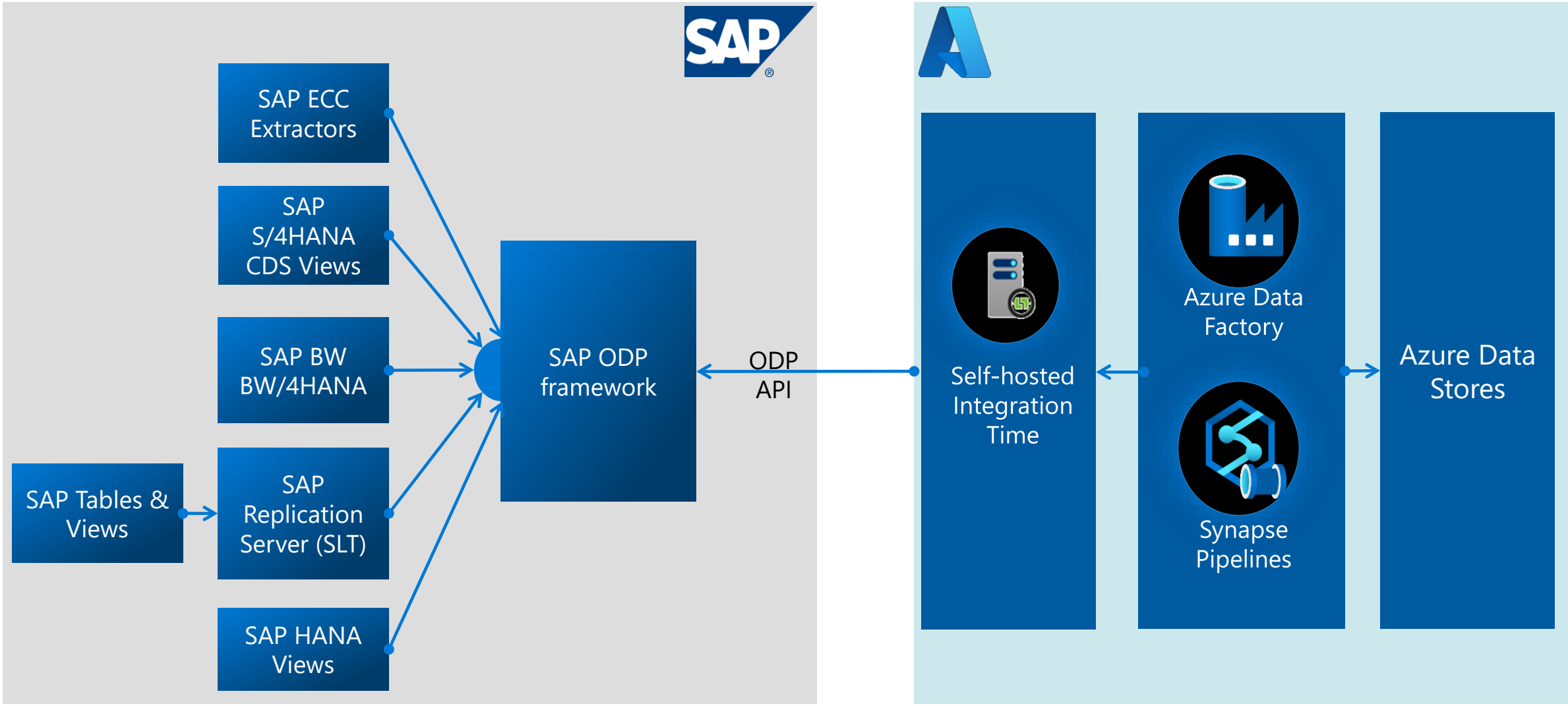


Capabilities:

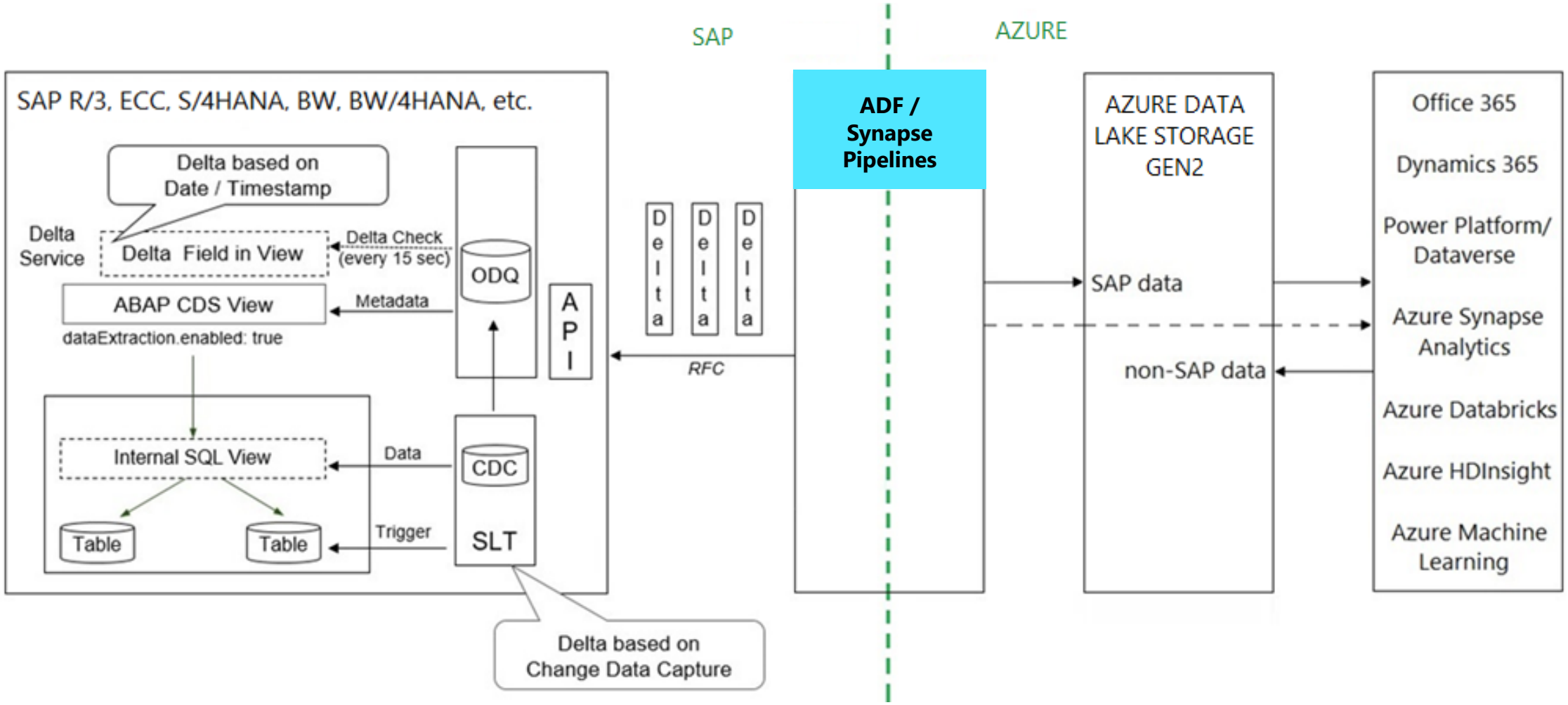
- ✓ Field selection
- ✓ Row filter
- ✓ Built-in partition + parallel load

Data Sources supported by SAP ODP Framework

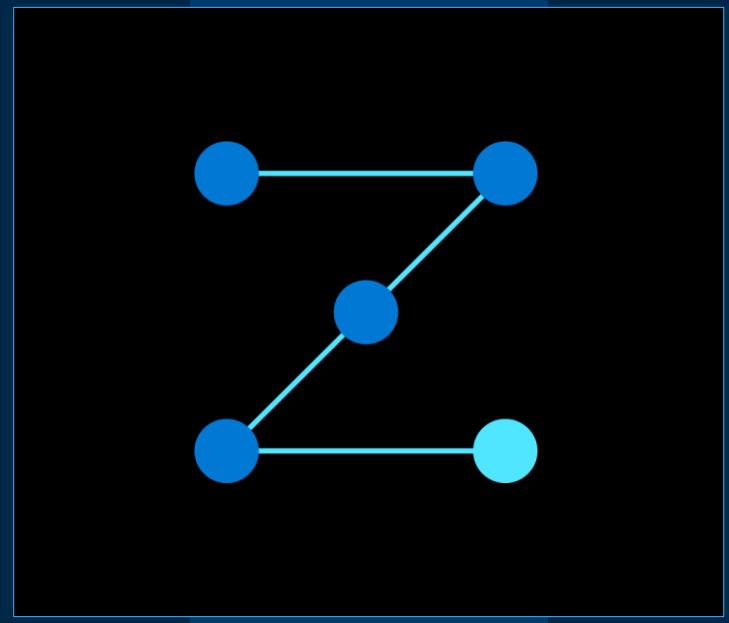
SAP CDC



SAP CDC Connector - Architecture



SAP HANA Connector

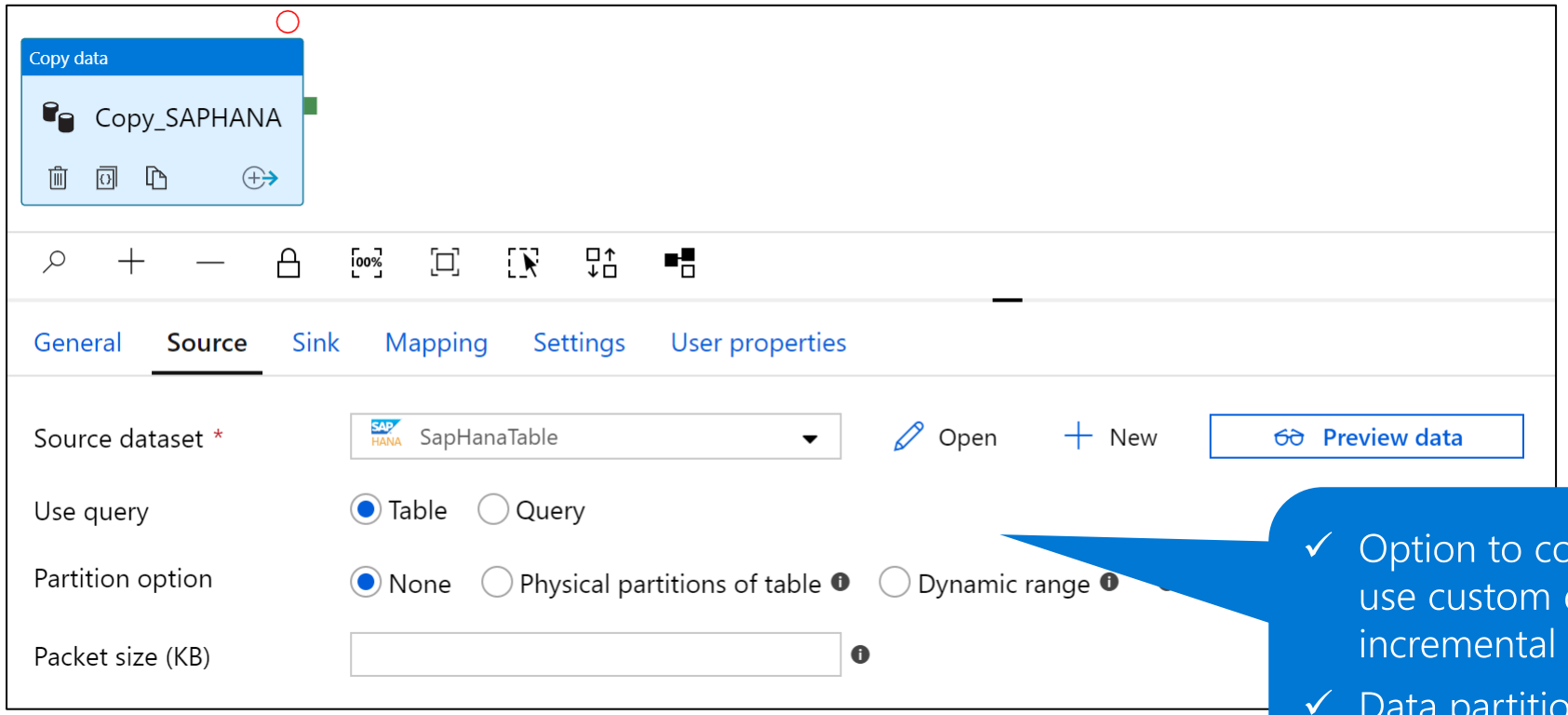


SAP HANA Connector

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

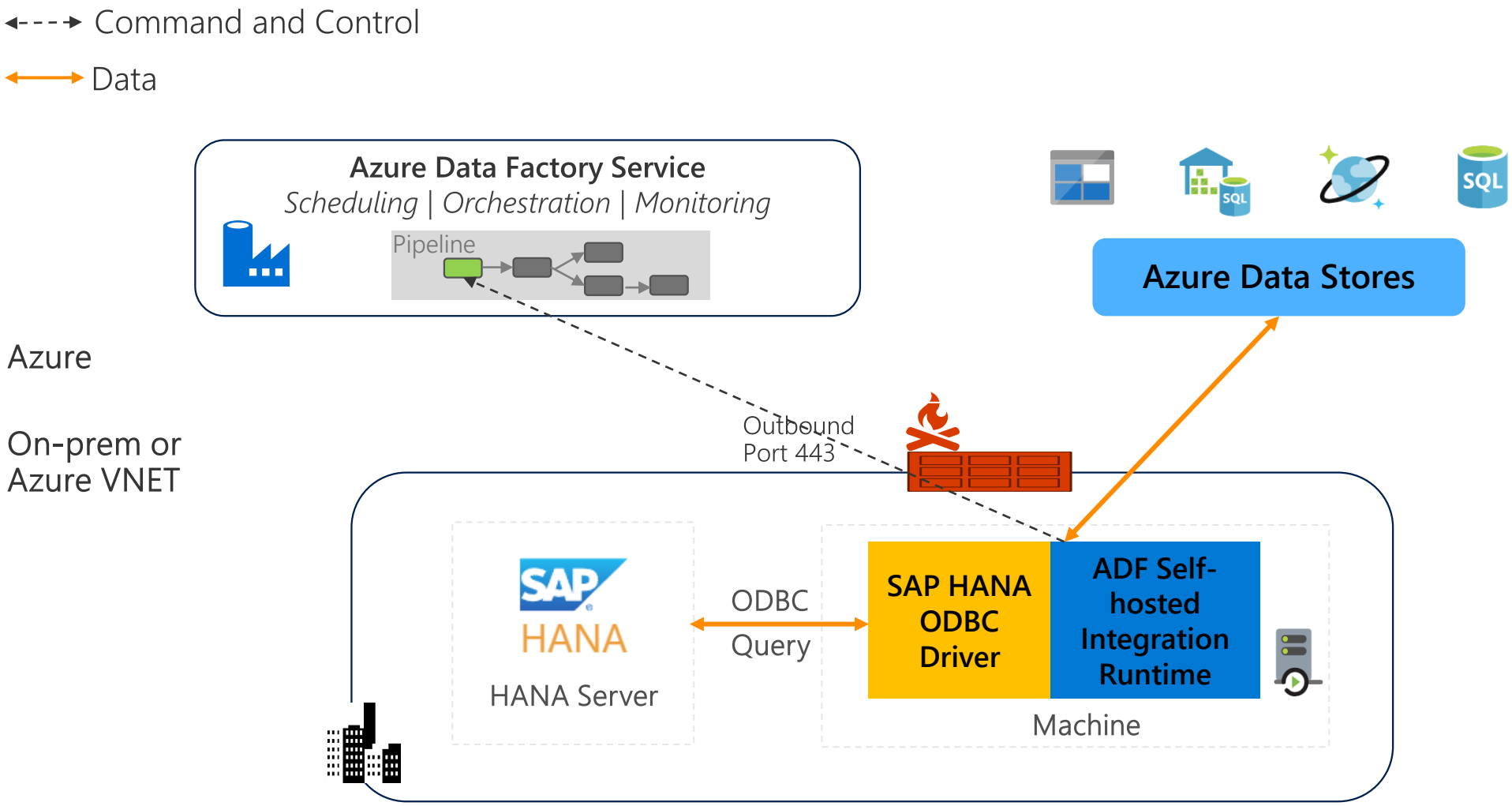
Supported versions	• **All SAP HANA versions** on-prem or in the cloud
Supported SAP objects	• HANA Information Models (Analytic/Calculation views) • Row & Column Tables
Supported authentications	• **Basic** – username & password • **Windows** – Single Sign-On via Kerberos-constrained delegation
Mechanism and prerequisites	• Built on top of **SAP's HANA ODBC driver** • Run on ADF 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 dozens millions to billion of rows** & observed several to several dozens MB/s (varies per customers' data/env.)

SAP HANA Connector



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

SAP HANA Connector – How It Works



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
…	…	…	…

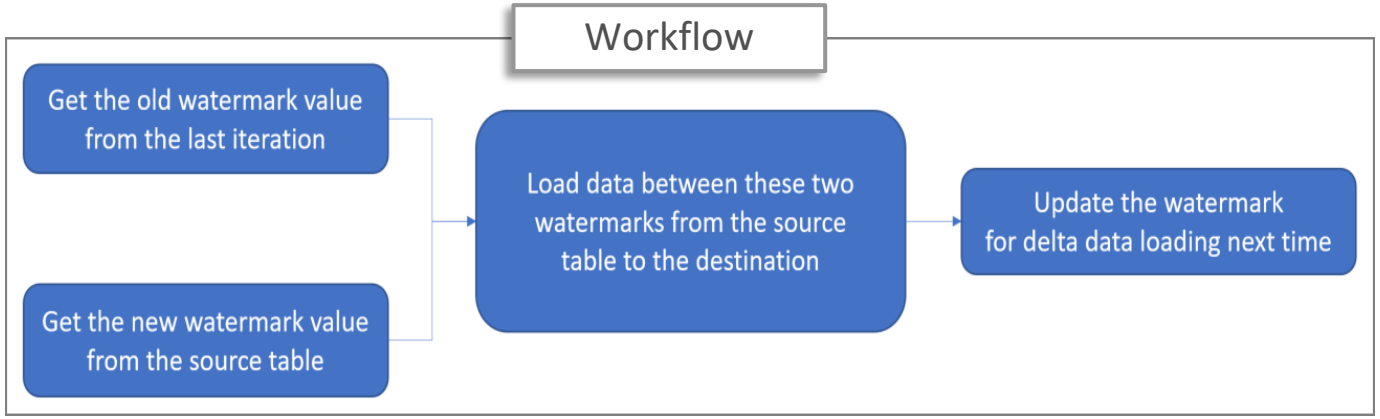
```
SELECT * FROM MyTable
WHERE LastModifiedDate >= @{formatDateTime(pipeline().parameters.windowStartTime, 'yyyy/MM/dd')
AND LastModifiedDate < @{formatDateTime(pipeline().parameters.windowEndTime, 'yyyy/MM/dd')
```

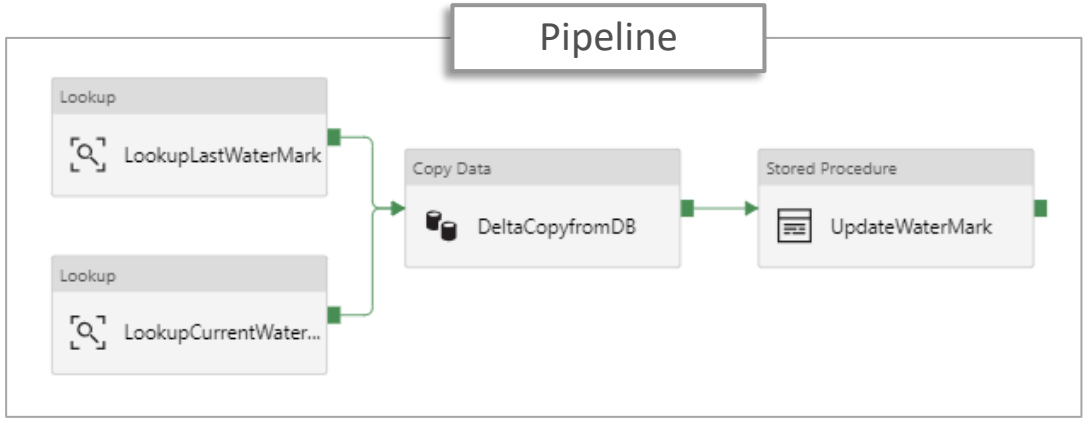
Execution start time: 2019/03/19 00:00:00 (window end time)
Delta extraction: last modified time between 2019/03/18 – 2019/03/19

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

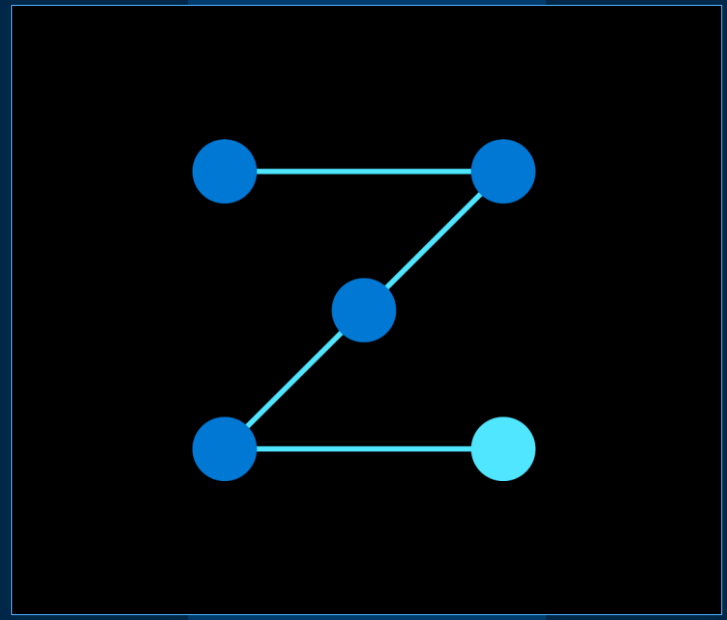
- Pattern II: "my data has an incremental column e.g. ID"
- Solution: external control table/file + high watermark.
- *Get started via solution template "Delta copy from Database".*





Pattern III: "my data is small in size as dimension data"
Solution: full copy and overwrite

SAP Table Connector

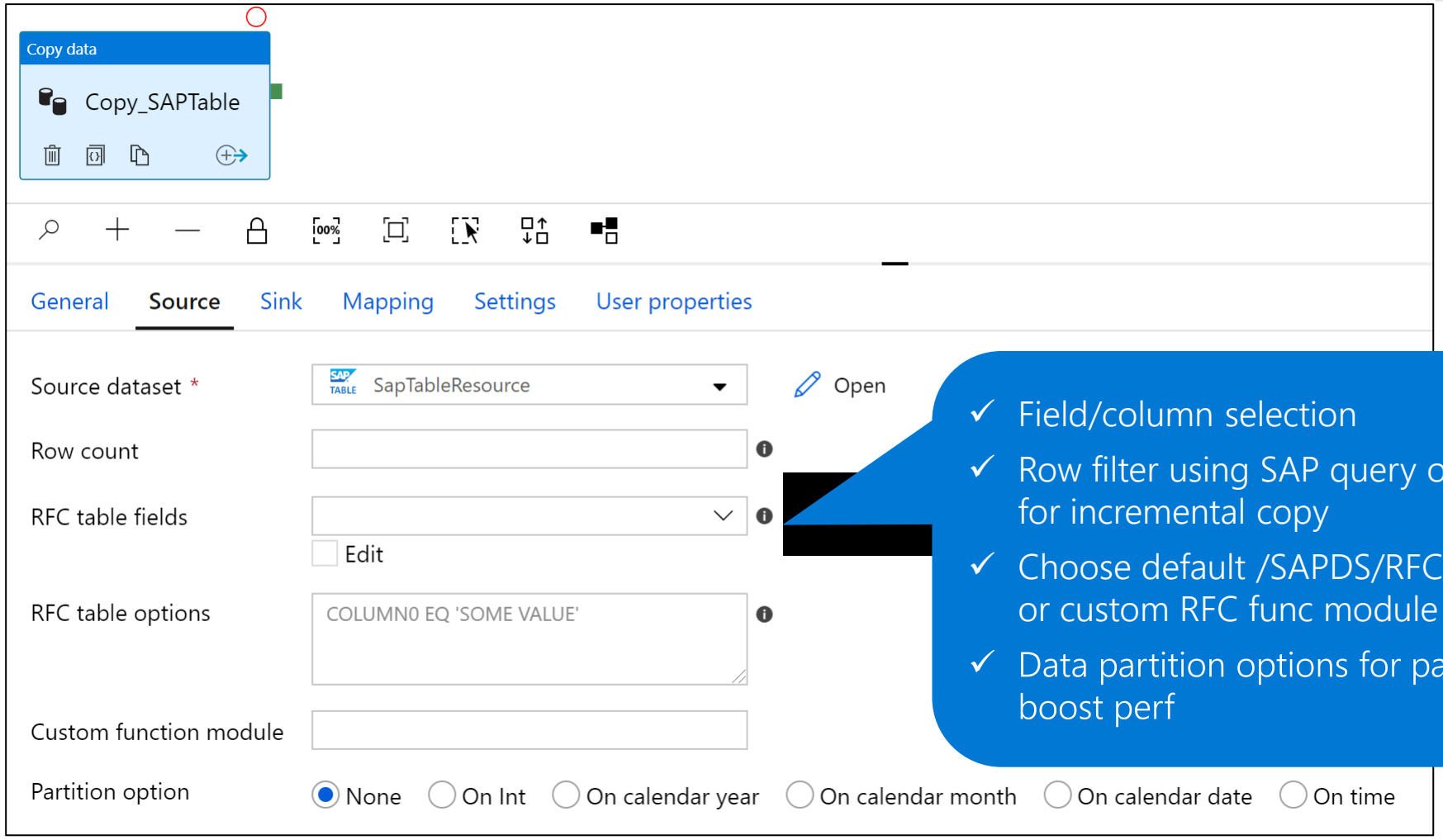


SAP Table Connector

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

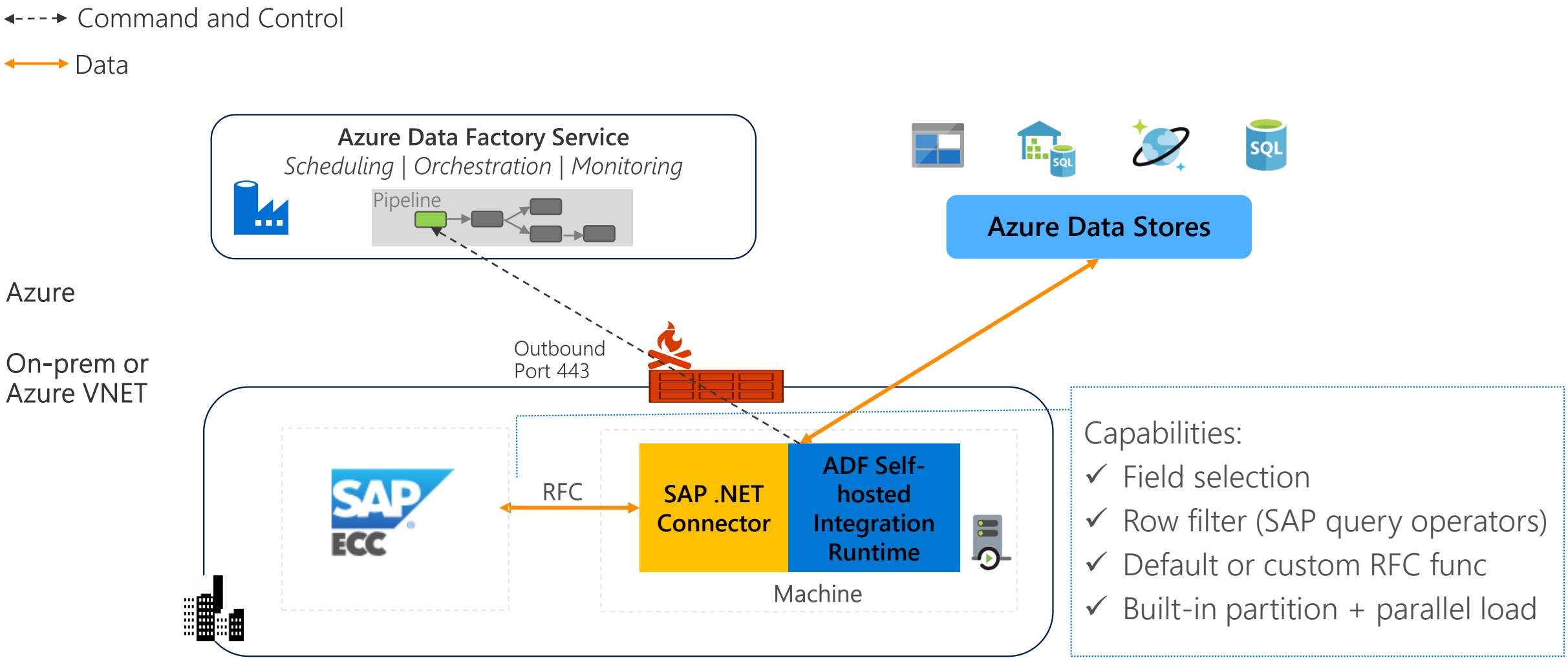
Supported versions	• **SAP ECC or other applications in Business Suite version 7.01 and above, on-prem or in the cloud** • **S/4 HANA**
Supported SAP objects	• SAP Transparent Table, Pooled Table, Cluster Table and View
Supported server type	• Connect to **Application Server** or **Message Server**
Supported 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** • Run on ADF 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



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

SAP Table Connector – How It Works



SAP Table Connector – Built-in Parallel Loading

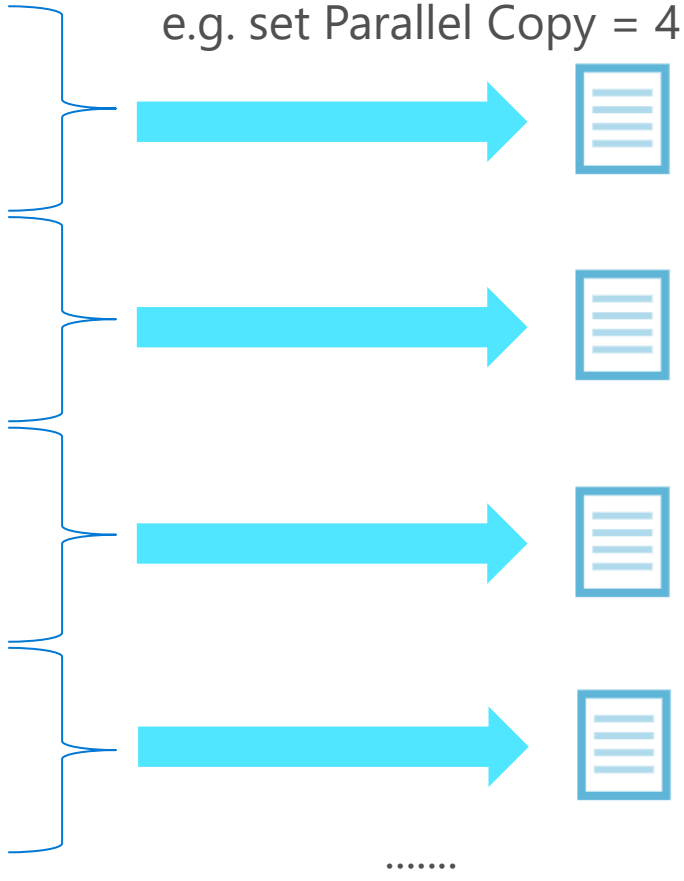
Configurable data partitioning on given column (INT, Calendar Year/Month/Date) + parallel copies

SAP table

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

ADF

Single Copy Activity execution
e.g. set Parallel Copy = 4



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.

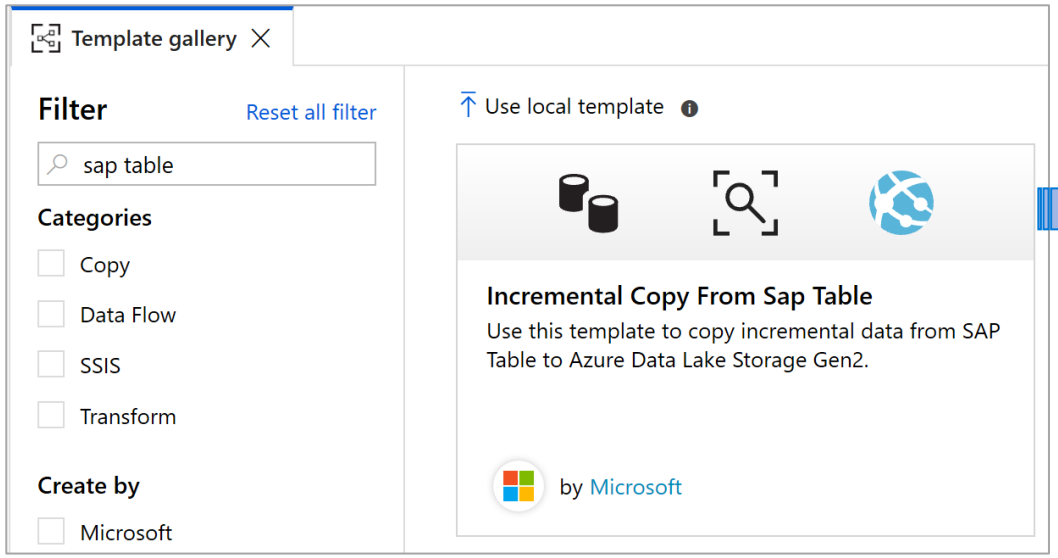
Learn more

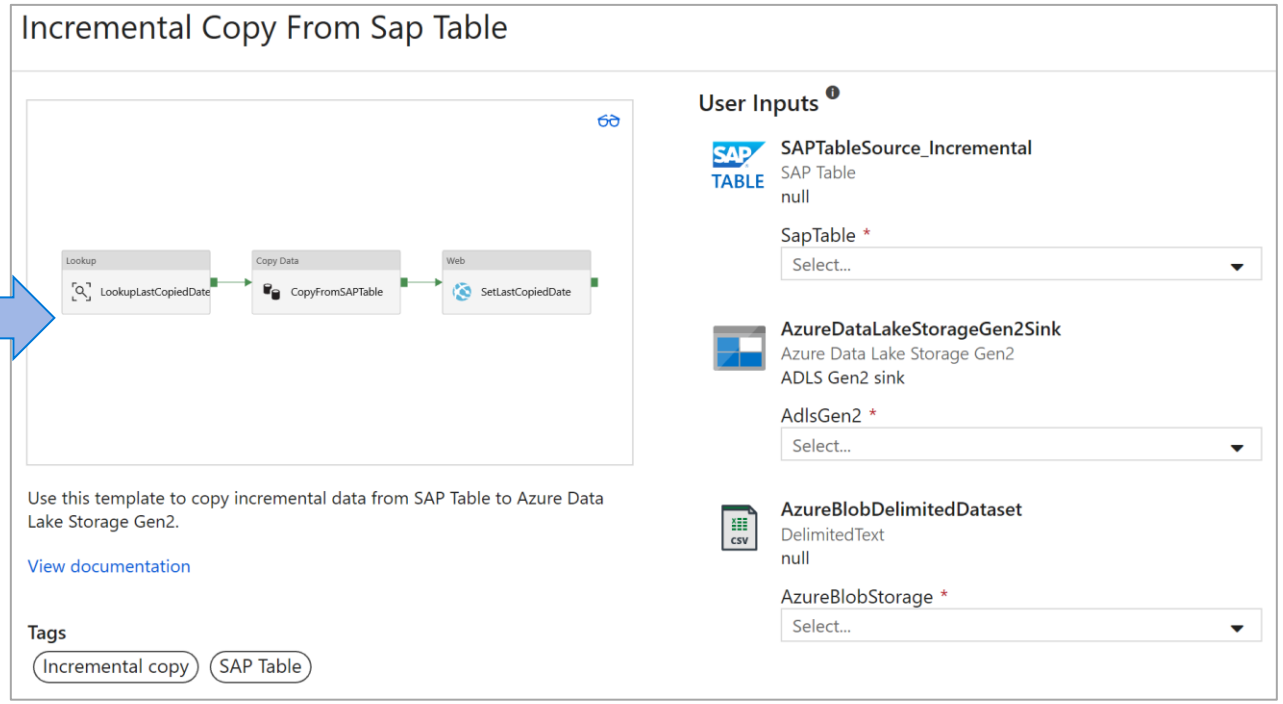
SAP Table Connector – Incremental Copy

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)

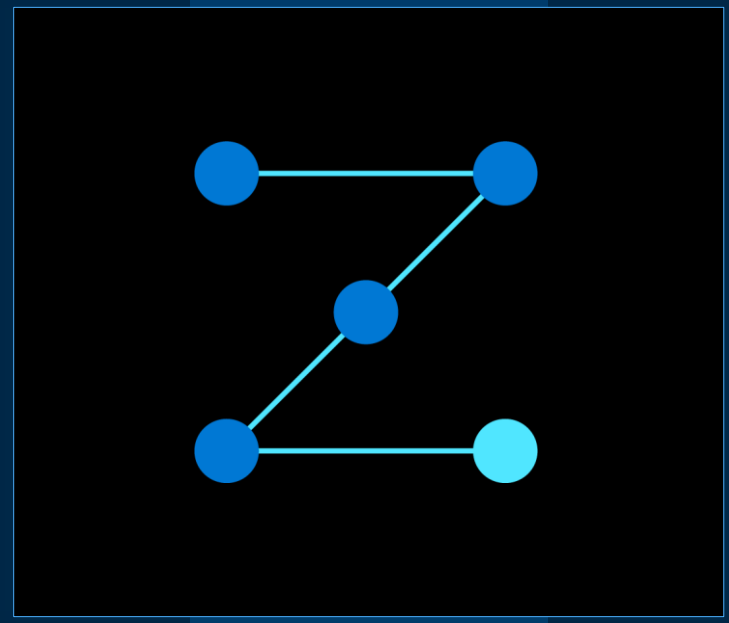
Pattern II: "my data has an incremental column e.g. id/last copied date"
Solution: external control table/file + high watermark.

Get started via solution template:





SAP ECC Connector

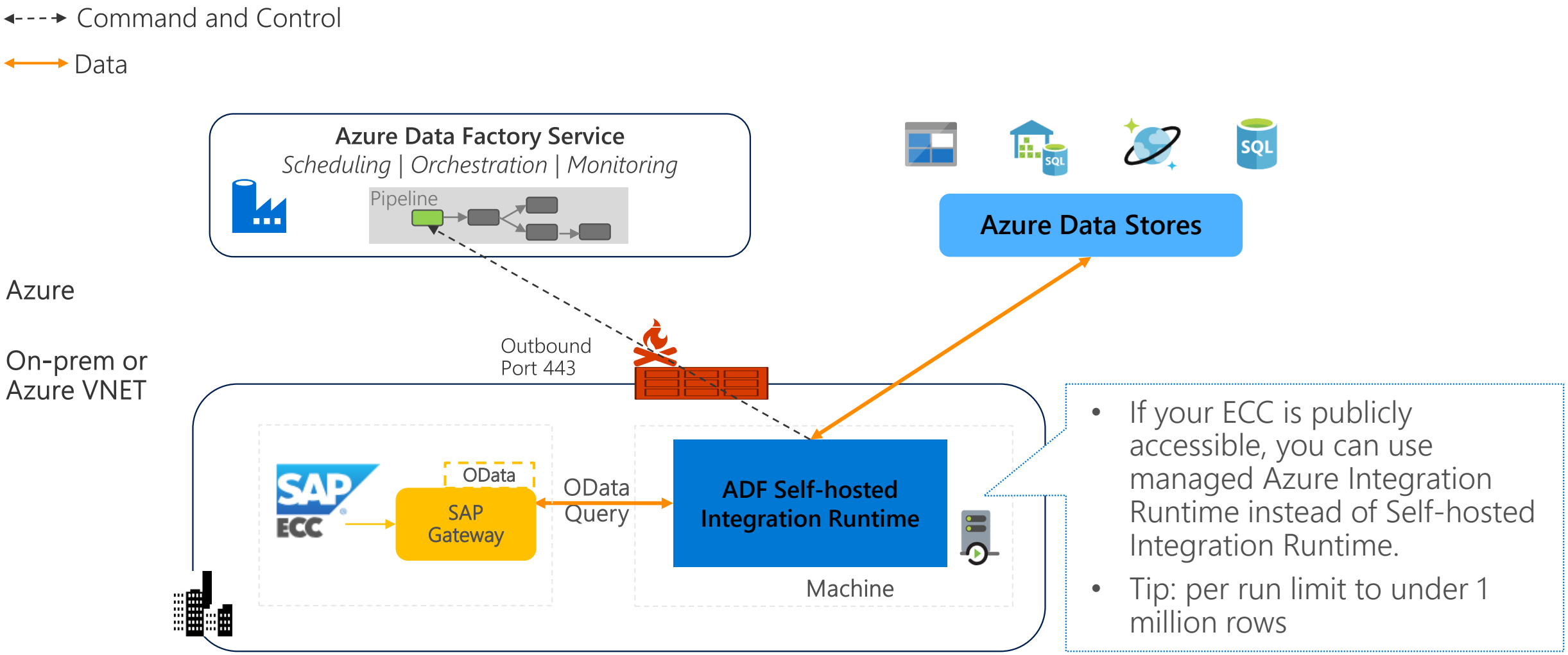


SAP ECC Connector – Supported Capabilities

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

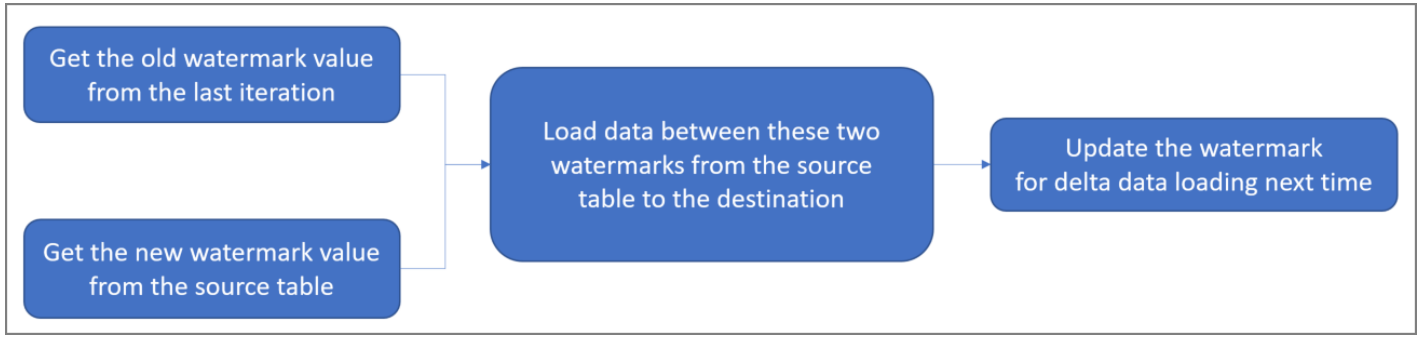


SAP ECC Connector – Incremental Copy

(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

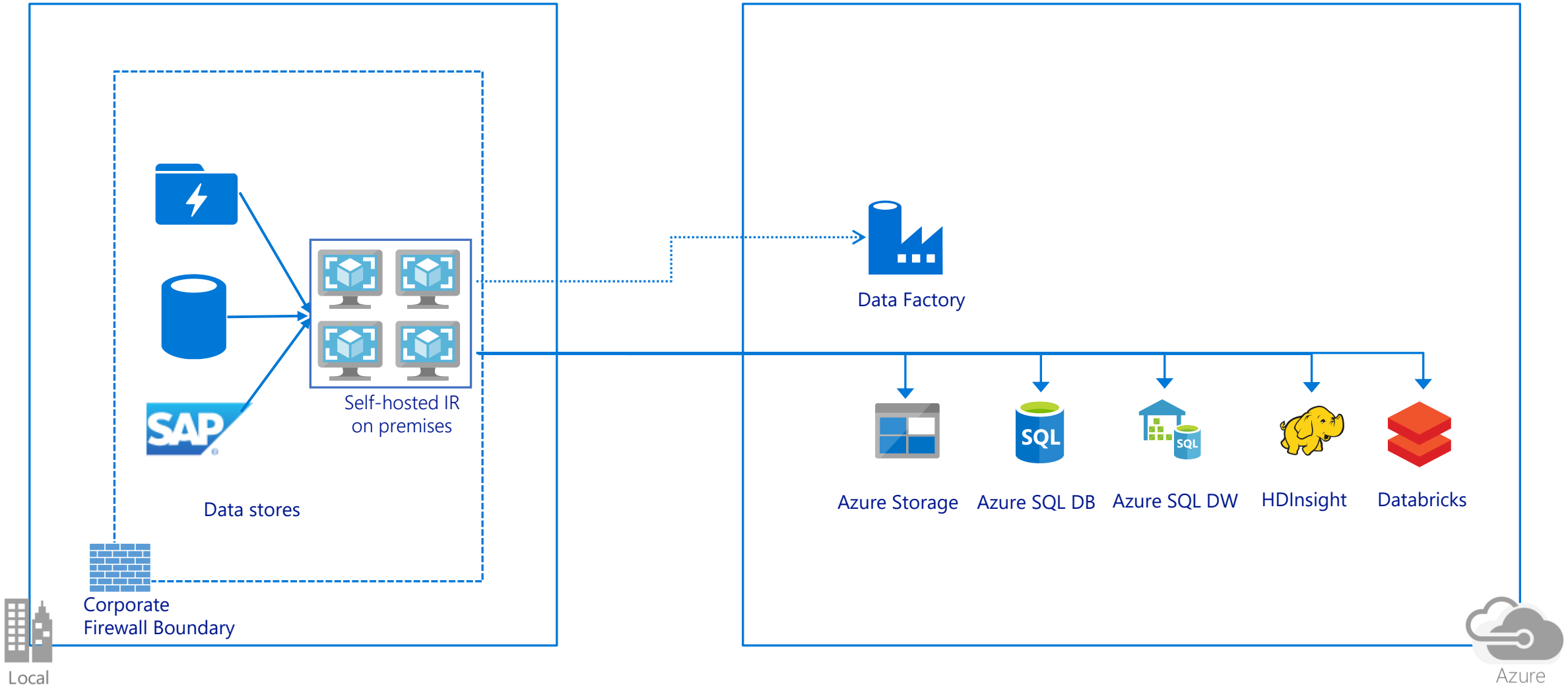
Pattern II: "my data has an incremental column e.g. ID"
Solution: external control table/file + high watermark.



Pattern III: "my data is small in size as dimension data"
Solution: full copy and overwrite

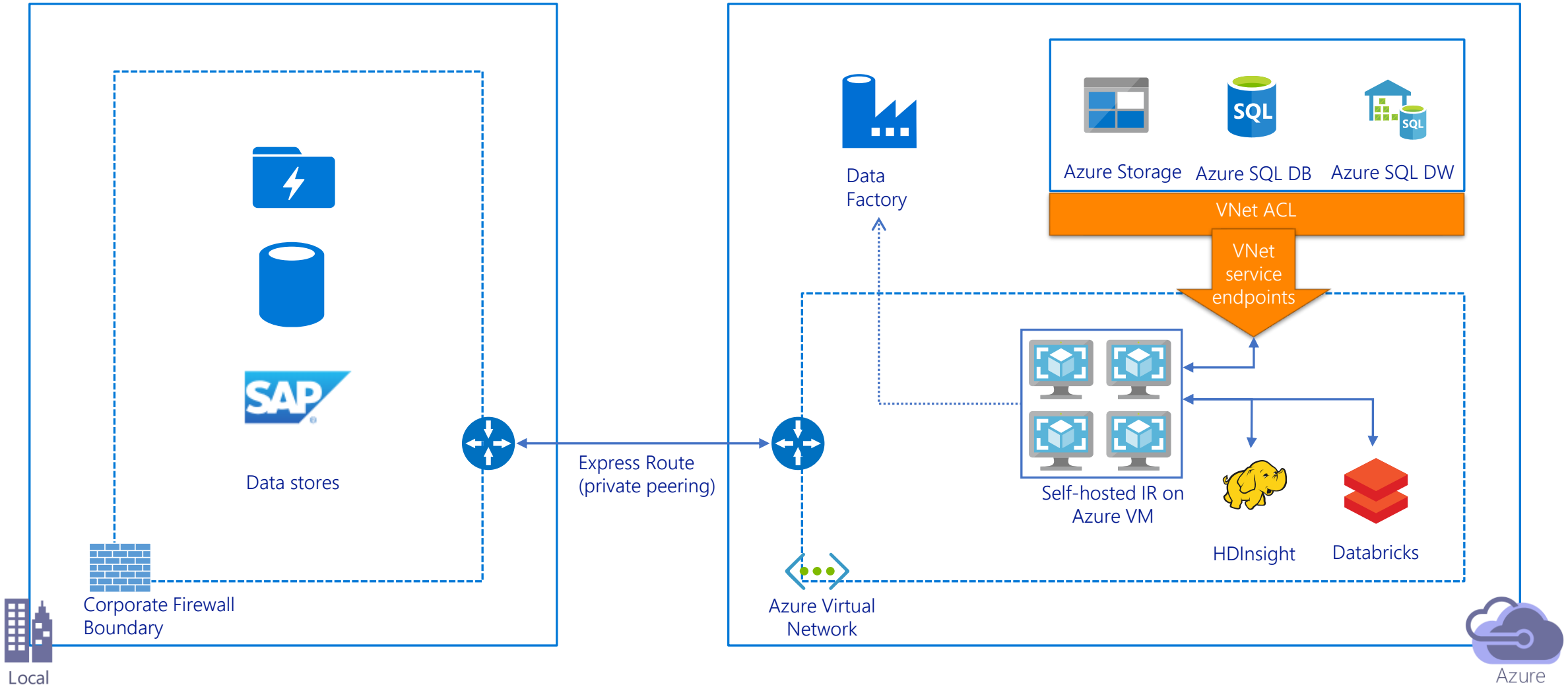
On-prem connected to Azure through public internet

Self-hosted IR deployed on premises



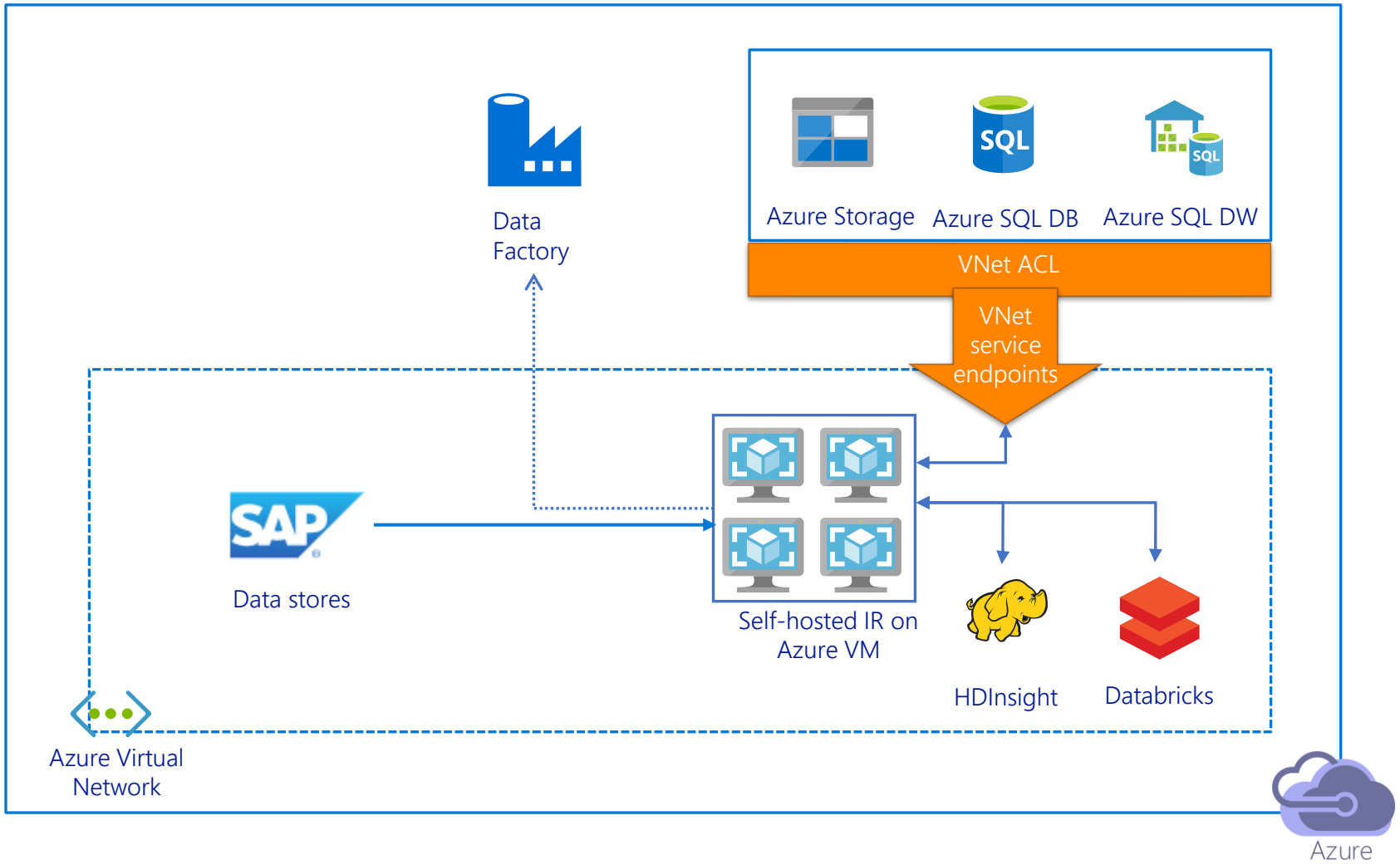
On-prem connected to Azure VNet through ExpressRoute

Self-hosted IR deployed on Azure VM

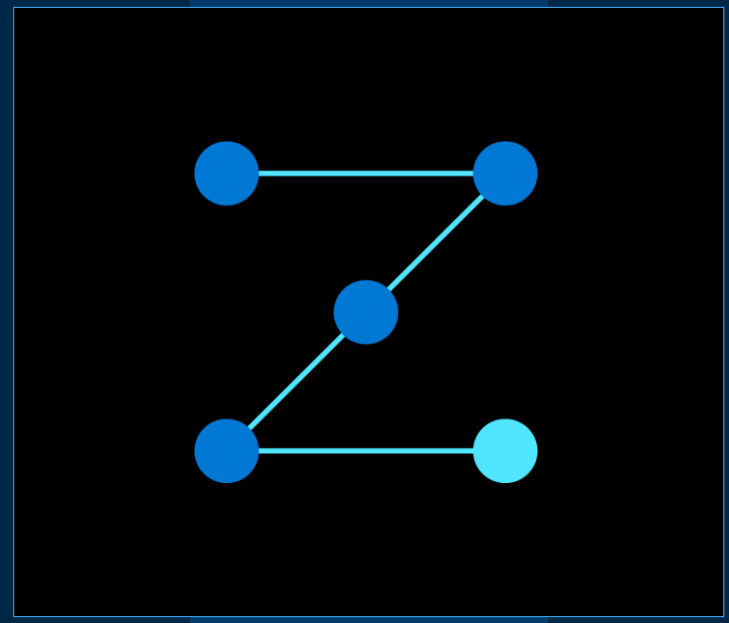


SAP on Azure

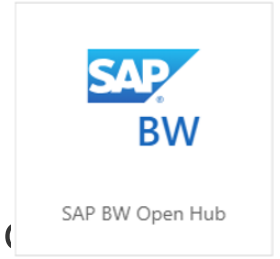
Self-hosted IR deployed on Azure VM



SAP BW Open Hub Connector



SAP BW Open Hub Connector



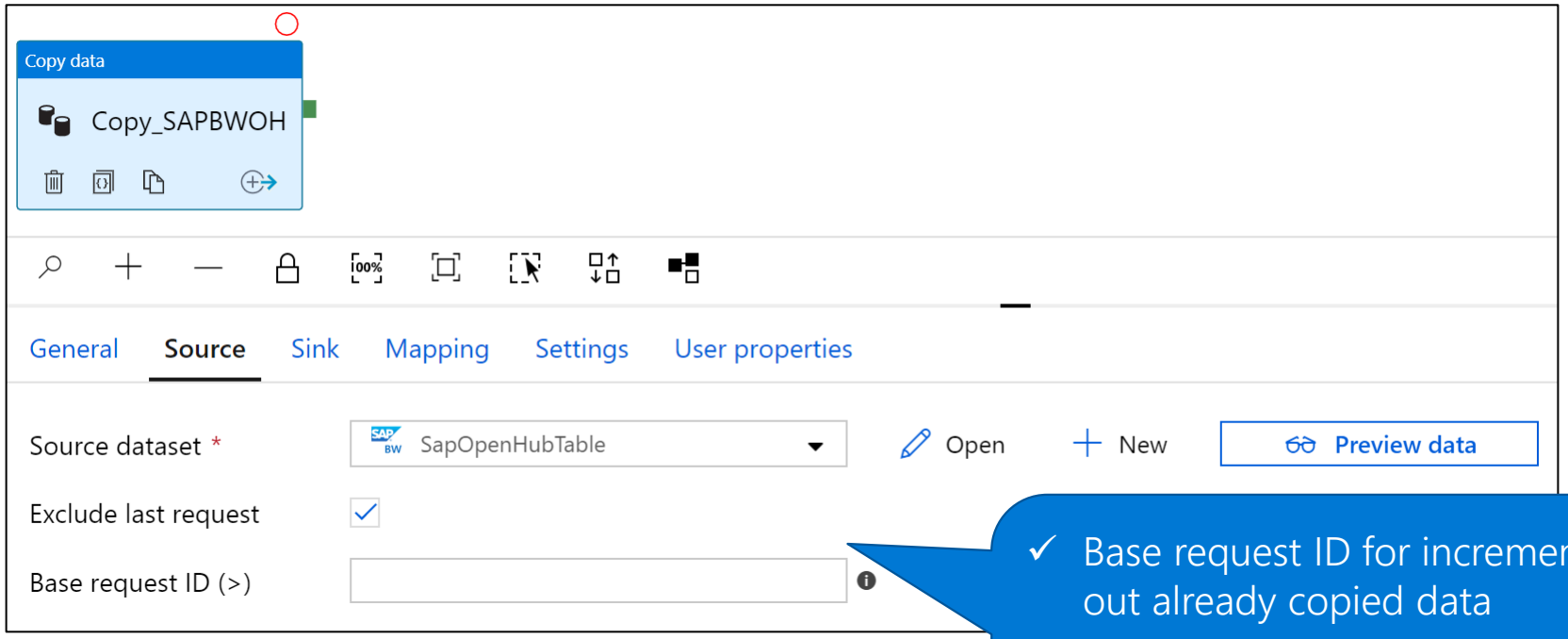
Suitable scenario: ingest data from **SAP BW** with targeted/well-thought-through workl

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**
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. 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



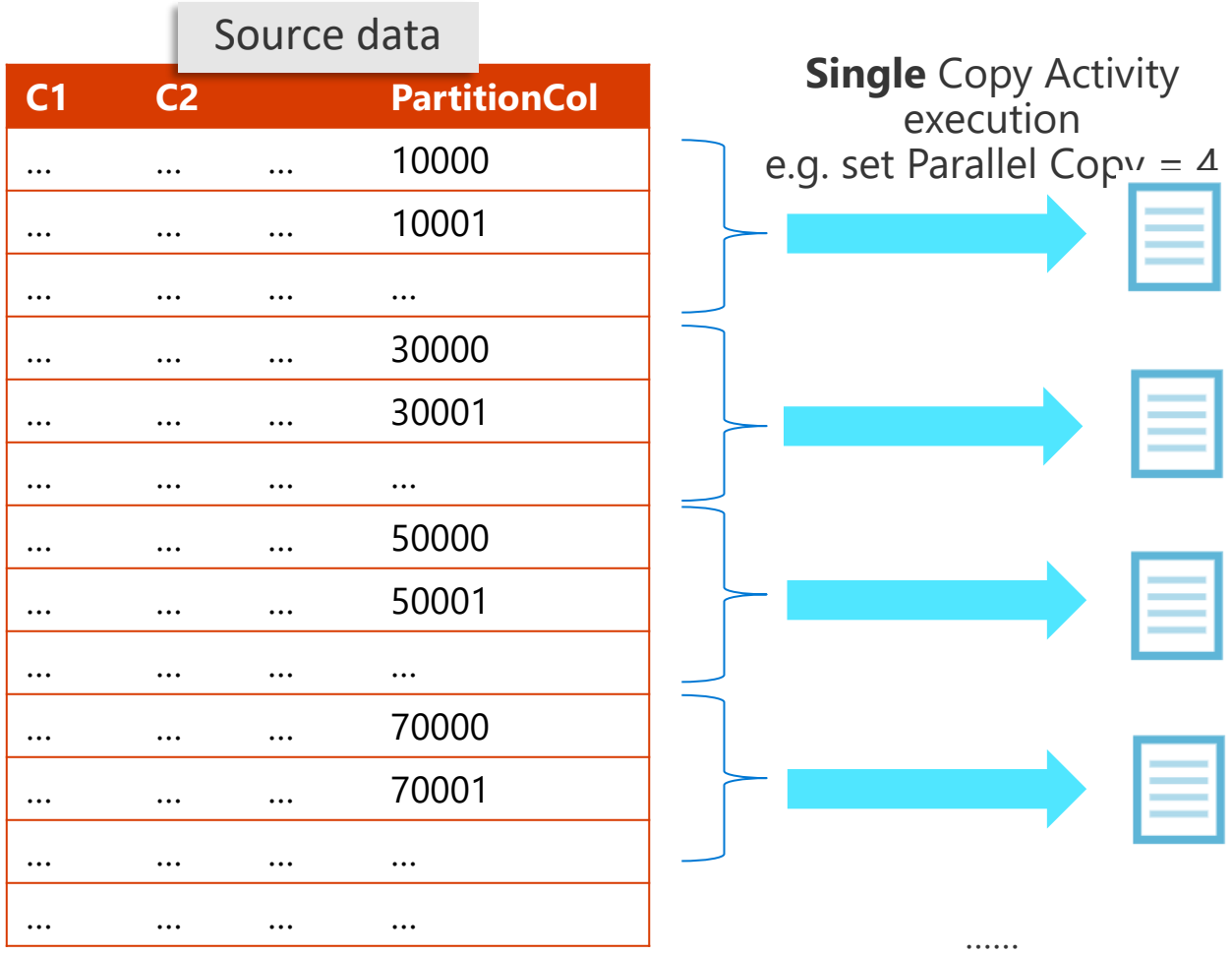
- ✓ Base request ID for incremental copy to filter out already copied data
- ✓ Exclude last request to avoid partial data
- ✓ Built-in parallel copy to boost perf based on OHD's specific schema

Copy Data Performantly w/ Built-in Parallelism

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

Out-of-box optimization for SAP HANA, SAP Table, SAP BW via Open Hub:

- Built-in **parallel copy by partitions** to boost performance for large table migration/ingestion.
- Options of dynamic range partition and native partition mechanism per data store.

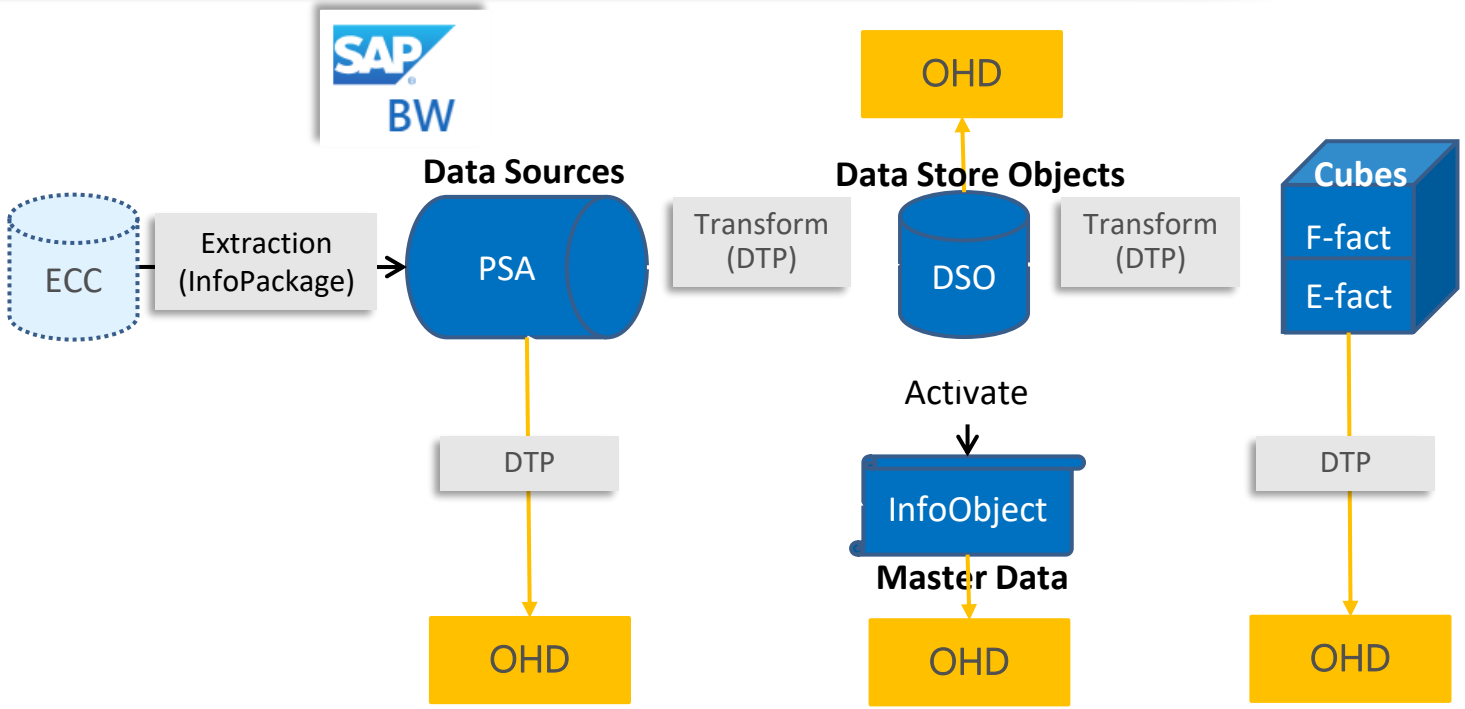


SAP BW Open Hub – How It Works

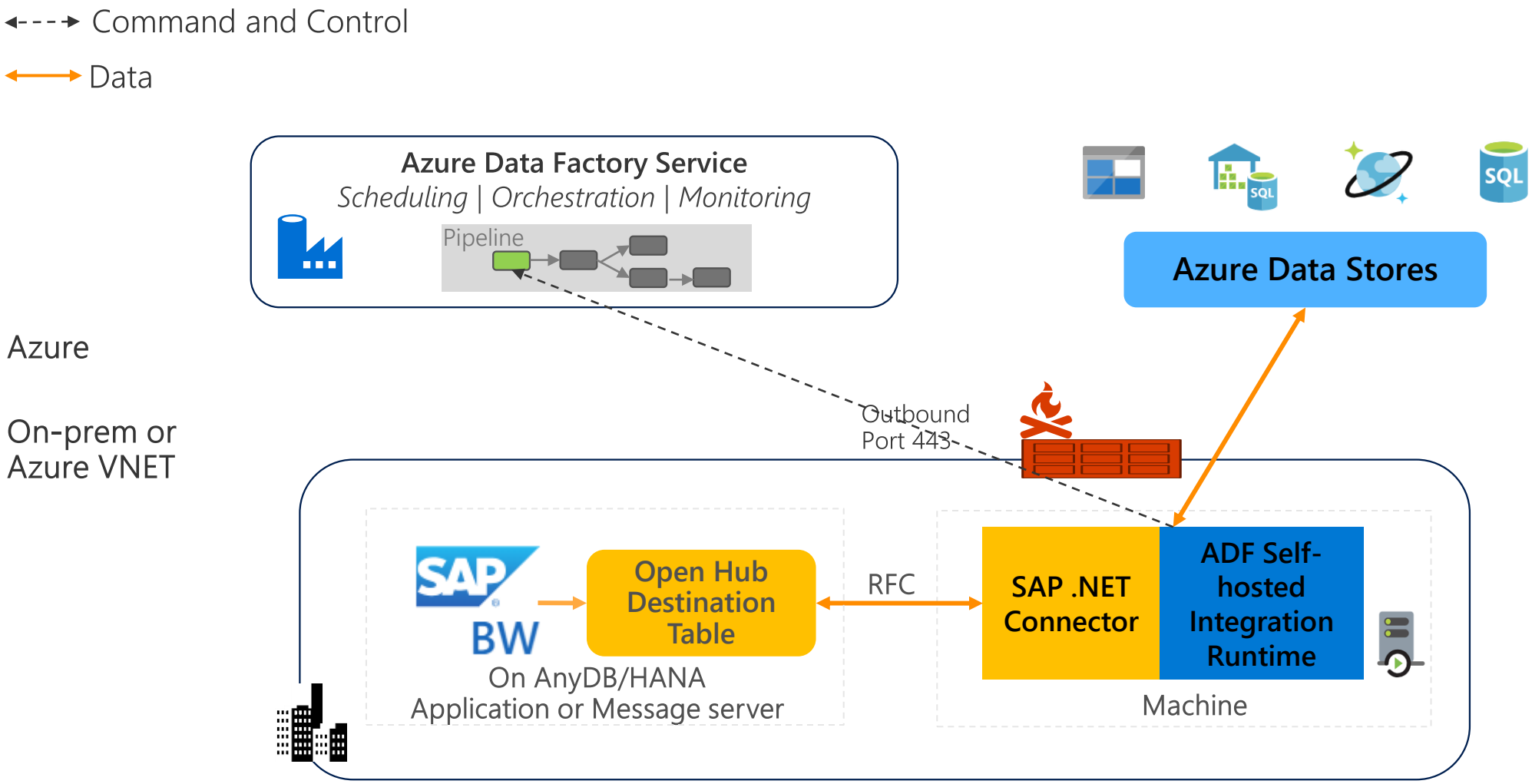
SAP Open Hub Destination (OHD):

- 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.

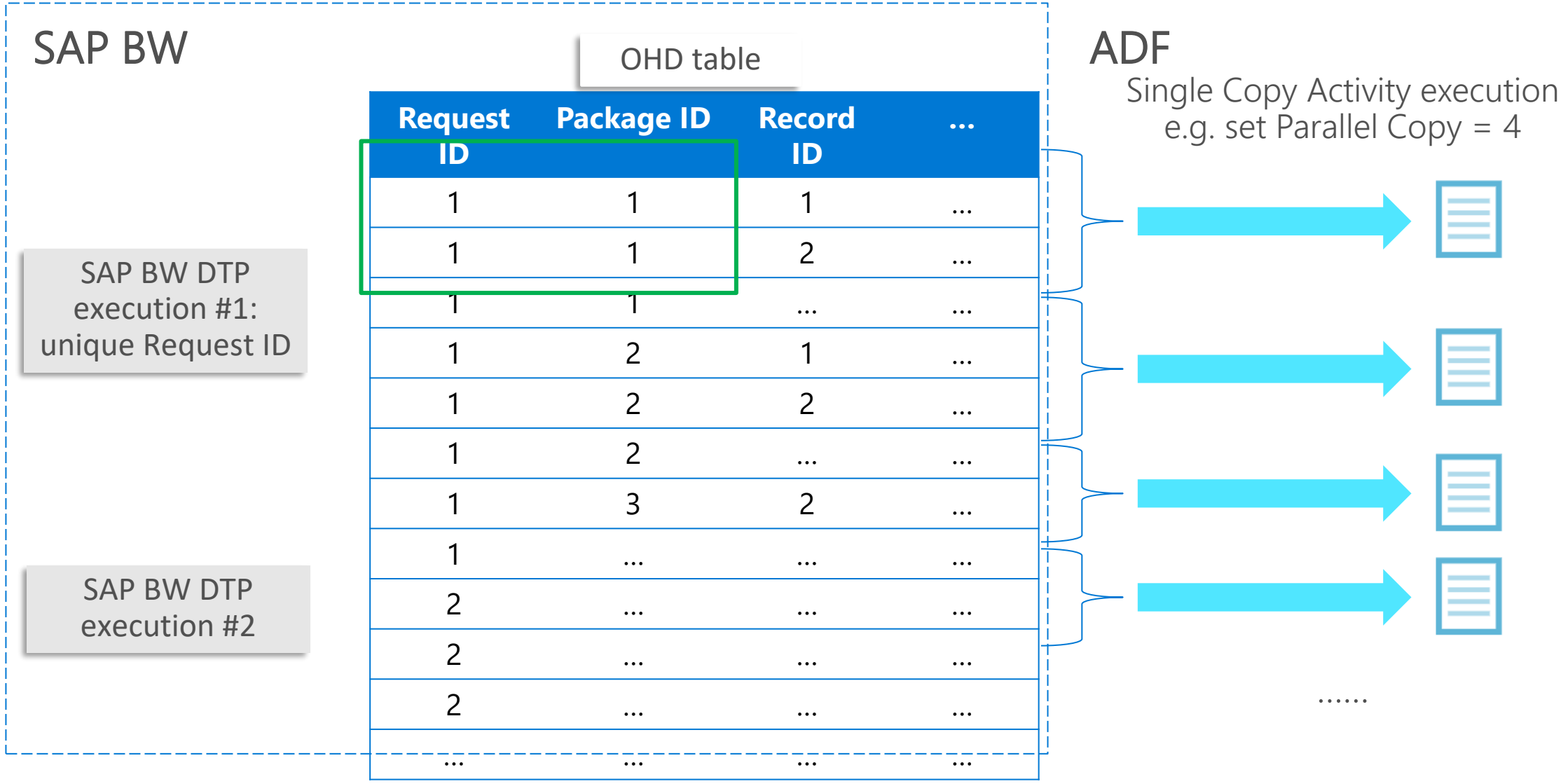
*The connector support OHD **local table in BW**.*



SAP BW Open Hub Connector – How It Works

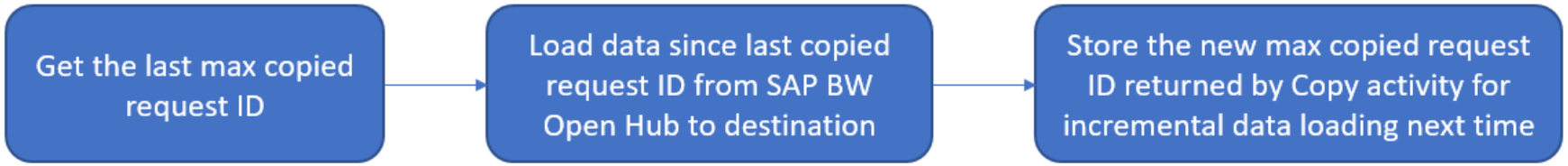


SAP BW Open Hub Connector – Built-in Parallel Loading



SAP BW Open Hub Connector – Incremental Copy

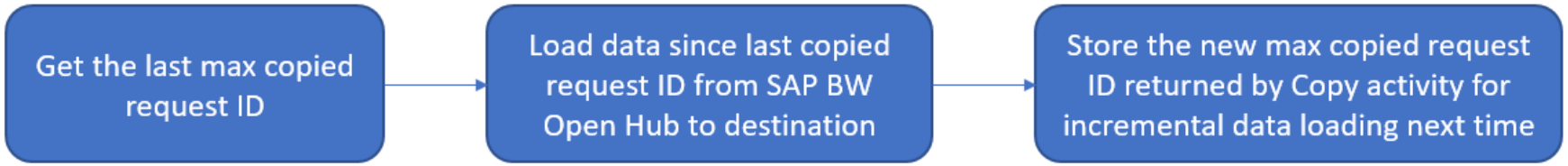
Solution: external control table/file + high watermark (max copied request ID).



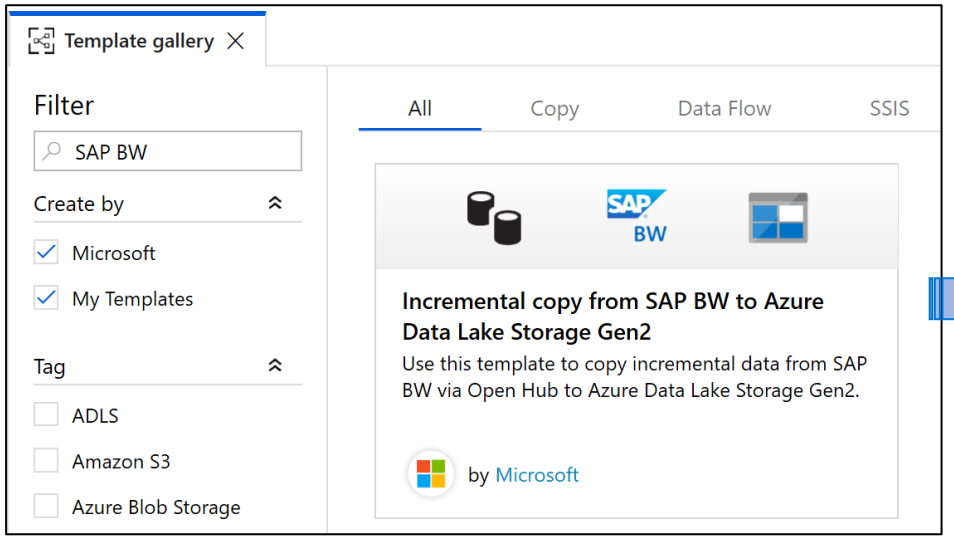
SAP OHD supports “delta” extraction mode to load incremental data into OHD table.

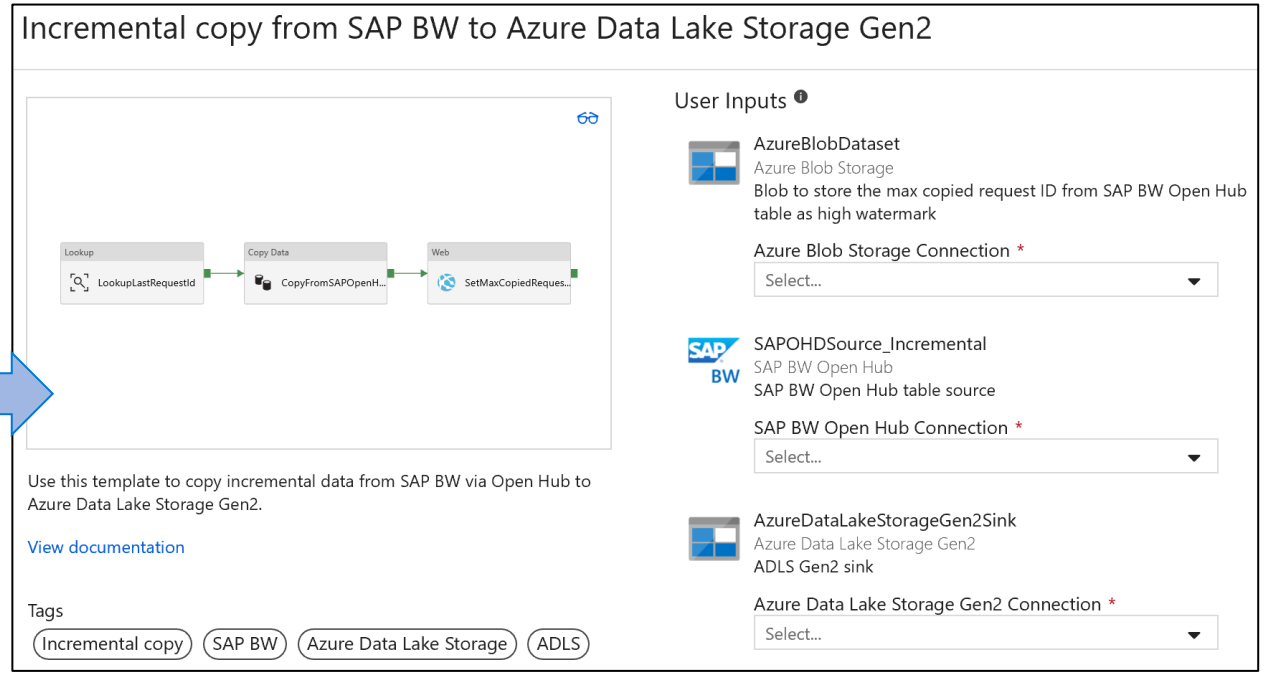
SAP BW Open Hub Connector – Incremental Copy

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



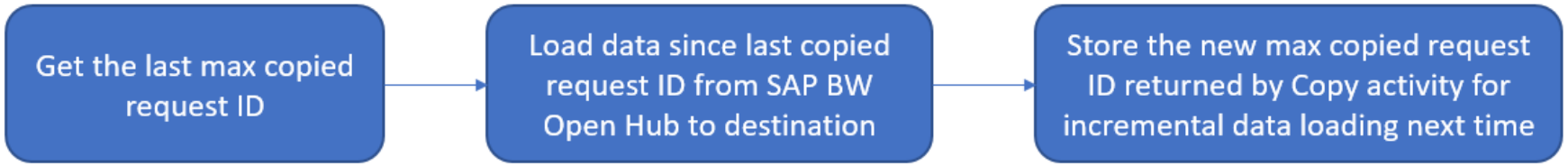
Get started via solution template:





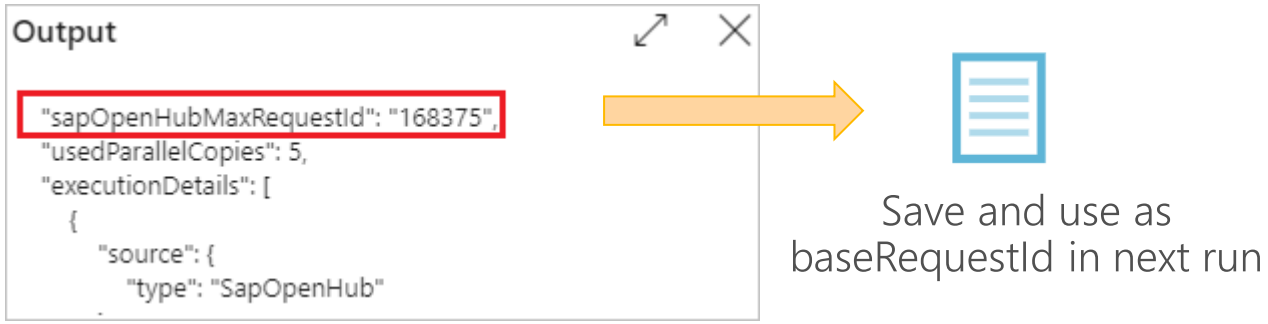
SAP BW Open Hub Connector – Incremental Copy

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



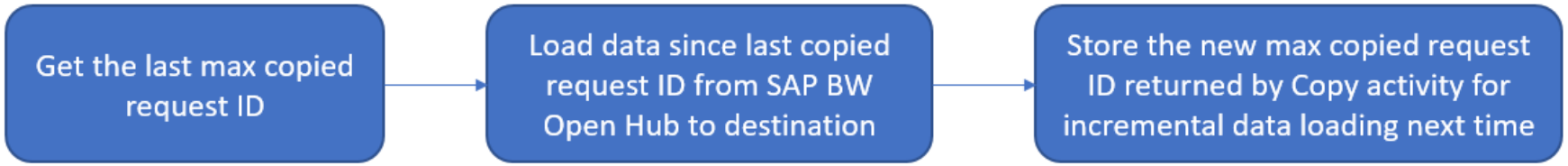
- **baseRequestId:** The ID of request for delta loading. Once it is set, only data with requestId **larger than** the value of this property will be retrieved.

Copy activity output:



SAP BW Open Hub Connector – Incremental Copy

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



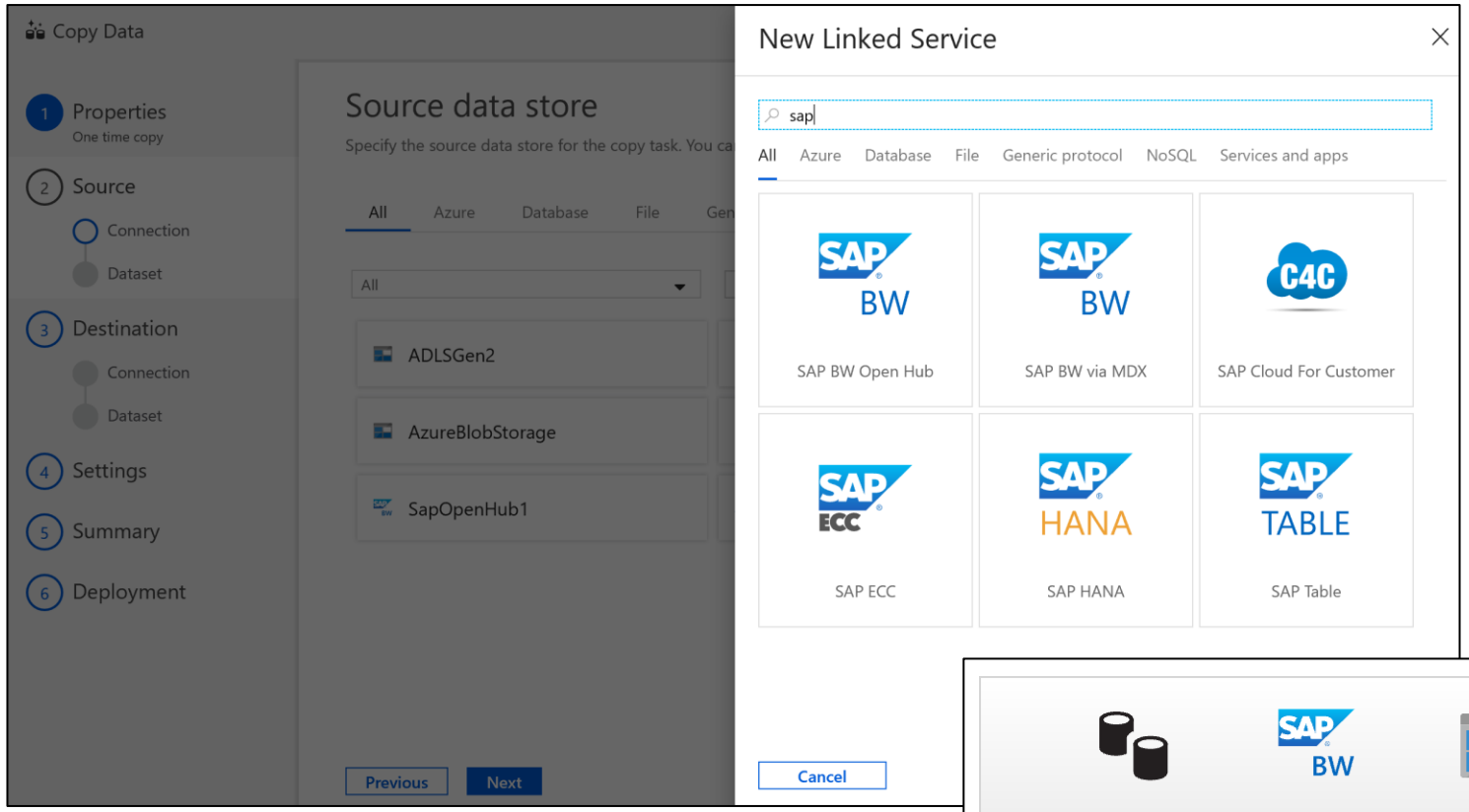
- **excludeLastRequestId**: Whether to exclude the records of the last request. Default value is true.

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

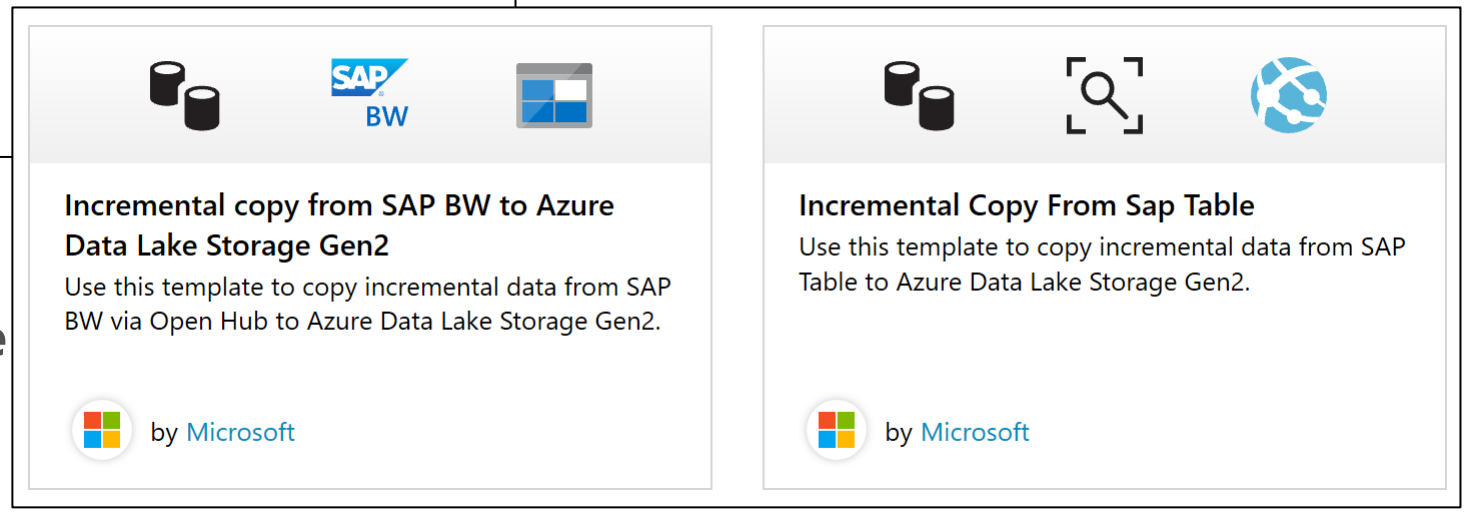
Include Last request ID:
- Applicable if Copy is always invoked after DTP is done

Get Started

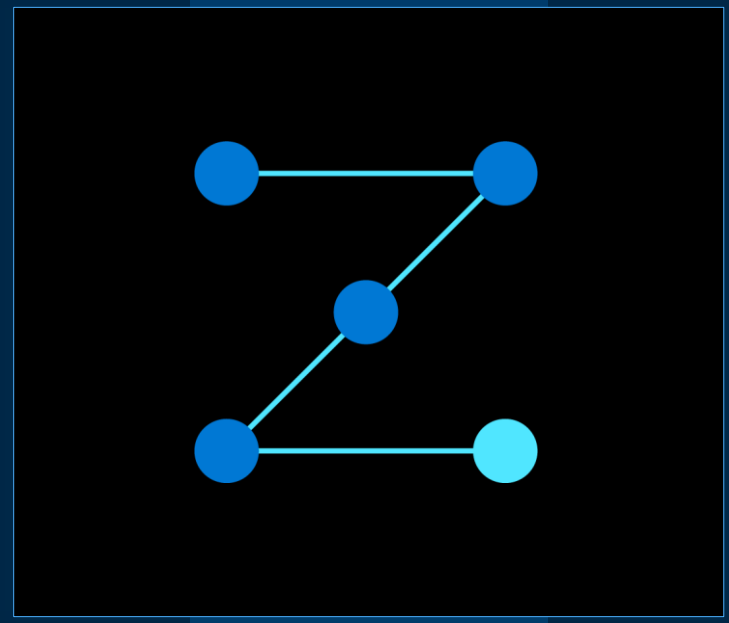


Copy Data Tool

Solution Template



SAP BW via MDX Connector



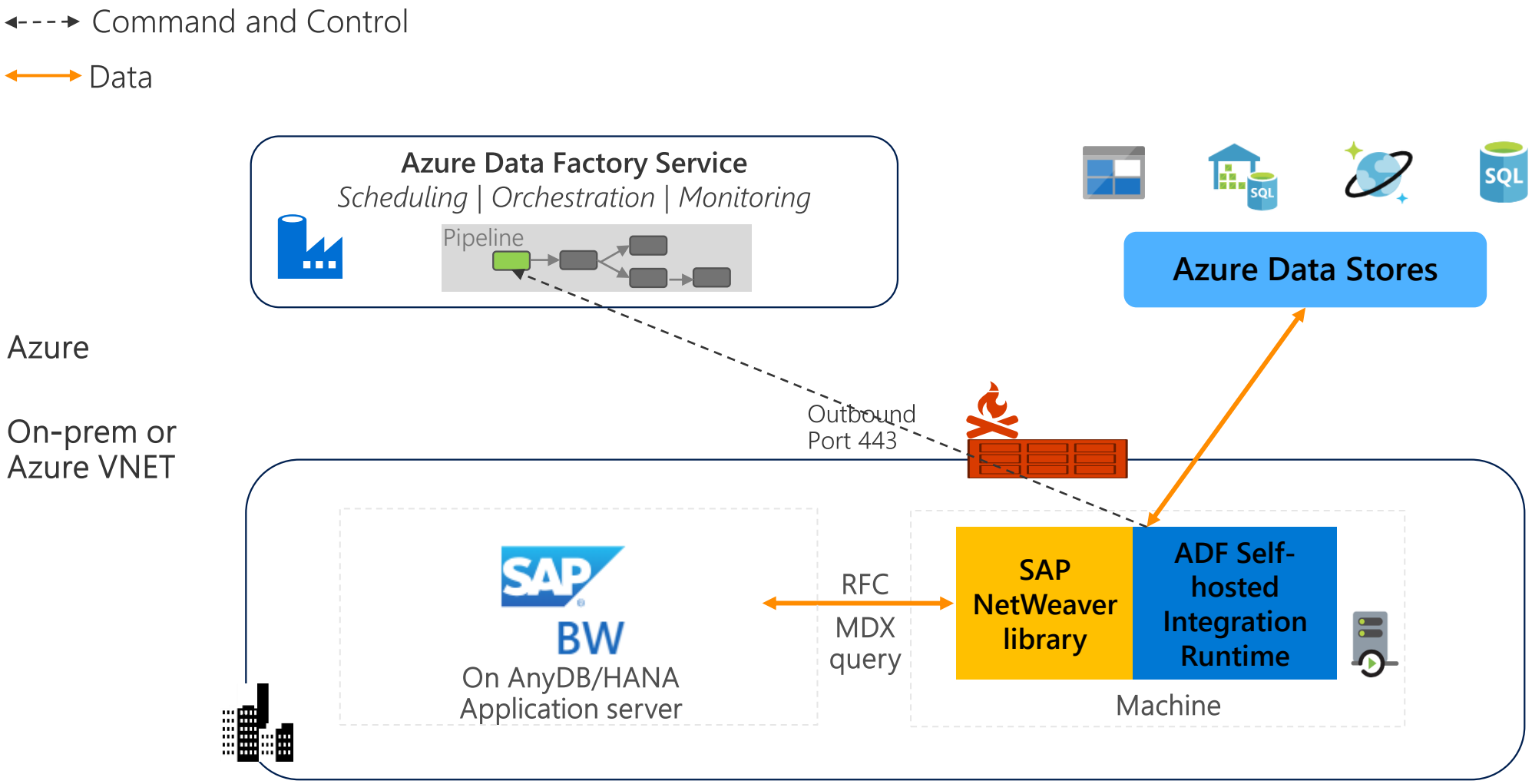
SAP BW via MDX Connector – Supported Capabilities

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

Supported versions	• **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



Microsoft

Thank you