

iNFOTRUST

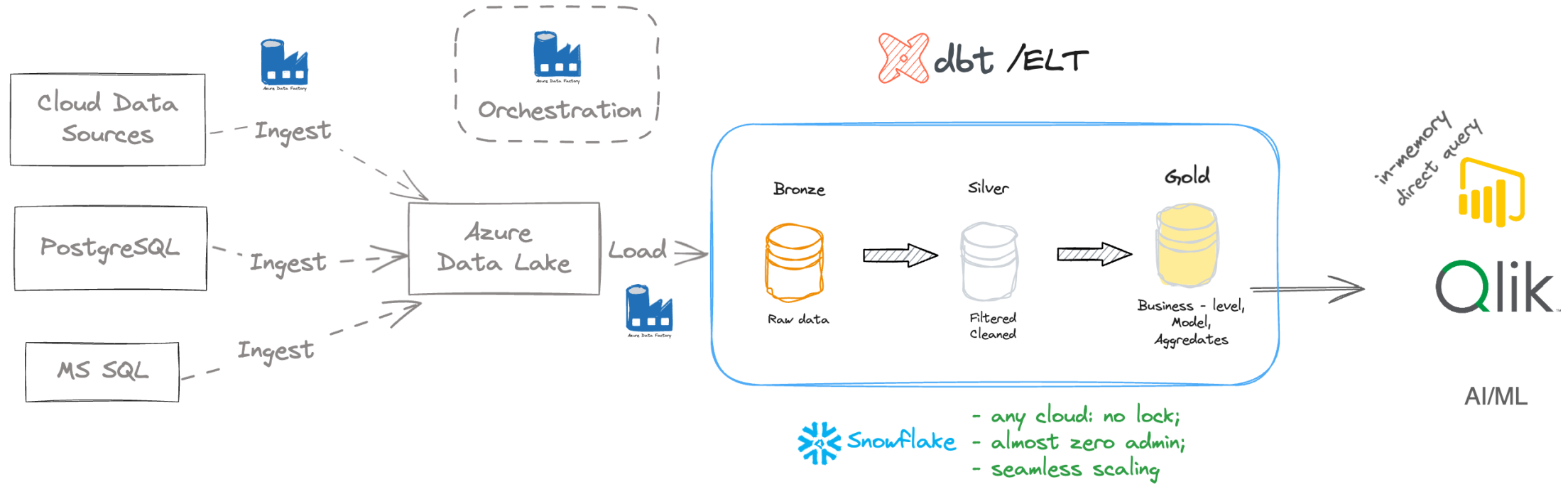


Data Ingestion into Snowflake – ADF

Simas Baranauskas
System Architect
Infotrust

September 10-12, 2024 | Snowflake Meetups, Vilnius&Riga

REFERENCE ARCHITECTURE



ORCHESTRATION

Benefits:

- Monitoring
- Complex (cycling) data loads
- Out of the box connectors
- Integration runtime for on-premise data
- Easy to use and scale (re-use pipelines)



ORCHESTRATION

DEMO

.

..

.

.

....

.....

INFOTRUST

Monitoring

[Rerun](#) [Cancel](#) [Refresh](#) [Update pipeline](#)
List
Gantt

Activity runs

Pipeline run ID 3ecda7c4-b7ba-4e48-8827-2bb1aaf17513

All status ▾

Showing 1 - 9 items

Activity name ↑↓	Activity status ↑↓	Activity type ↑↓	Run start ↑↓	Duration ↑↓
Facts_paskutinis	✔ Succeeded	Copy data	6/25/2024, 7:25:14 AM	1m 15s
RollingPlanFacts	✔ Succeeded	Copy data	6/25/2024, 7:24:55 AM	18s
GyvPlanasFacts	✔ Succeeded	Copy data	6/25/2024, 7:24:41 AM	14s
BendrasPlanasFacts	✔ Succeeded	Copy data	6/25/2024, 7:24:21 AM	19s
Kreditas2	✔ Succeeded	Copy data	6/25/2024, 7:23:38 AM	43s
Item chargai	✔ Succeeded	Copy data	6/25/2024, 7:22:32 AM	1m 5s
Kreditas	✔ Succeeded	Copy data	6/25/2024, 7:21:40 AM	51s
Facts1	✔ Succeeded	Copy data	6/25/2024, 7:16:42 AM	4m 57s

Details [Refresh](#)

Performance tuning tips:
The copy activity run spent 265 seconds waiting for source query to return data. To achieve better performance, you are suggested to check and optimize the [document](#).

[Learn more on copy performance details from here.](#)

Activity run id: 5a8e1e67-750c-479d-ac28-9d3c3afba0c2

Azure Synapse Analytics
Region: Germany West Central

Succeeded

→

Azure IR region: West Europe

Azure Data Lake Storage Gen2
Region: West Europe

Data read: ⓘ	284.885 MB	Data written: ⓘ	245.412 MB
Rows read:	484,827	Files written: ⓘ	1
Peak connections: ⓘ	1	Rows written: ⓘ	484,827
		Peak connections: ⓘ	1

Copy duration: 00:04:55

Throughput: ⓘ 12.949 MB/s

▼ Azure Synapse Analytics → Azure Data Lake Storage Gen2

Start time	6/25/2024, 7:16:43 AM		
Used DIUs ⓘ	4		
Used parallel copies ⓘ	1		
▼ Duration	00:04:55		

Details	Working duration	Total duration
✔ Queue ⓘ		00:00:06
✔ Transfer ⓘ		00:04:47
Time to first byte ⓘ	00:04:25	
Reading from source ⓘ	00:00:08	
Writing to sink ⓘ	00:00:03	

Data consistency verification ⓘ Not verified

How satisfied or dissatisfied are you with the performance of this copy activity?

★ ★ ★ ★ ★

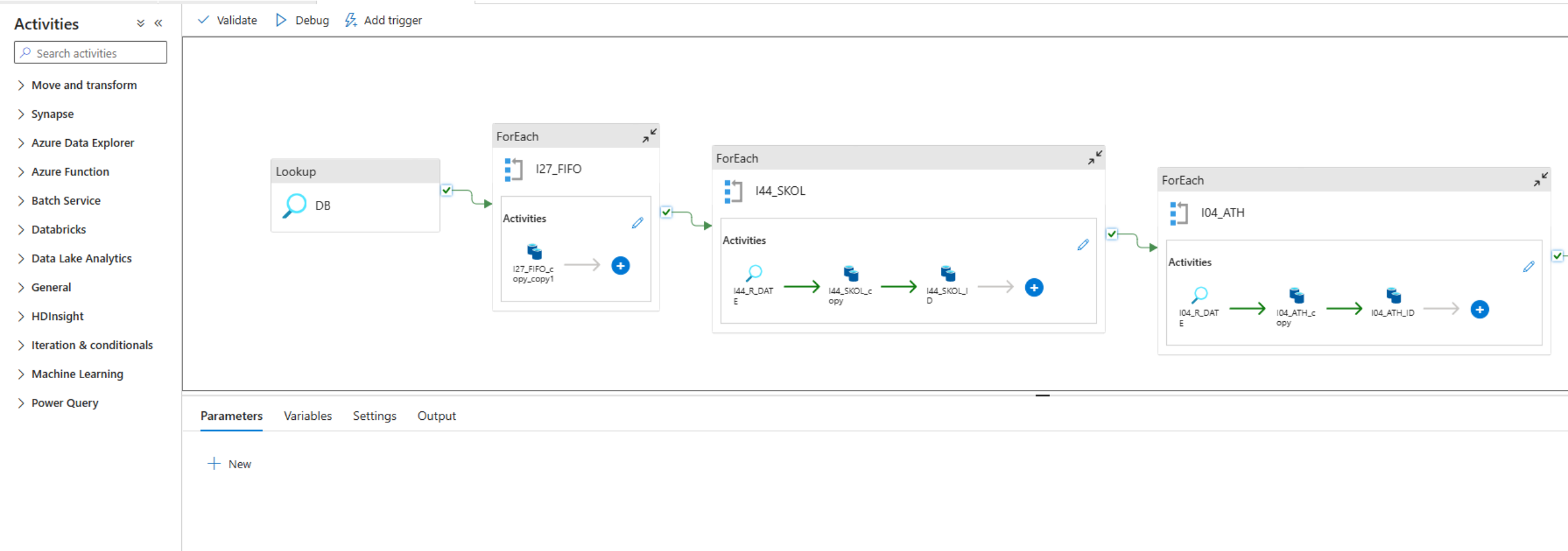
Data Factory

Benefits:

- Monitoring
- Complex (cycling) data loads
- Out of the box connectors
- Integration runtime for on-premise data
- Easy to use and scale (re-use pipelines)



Complex (cycling) data loads

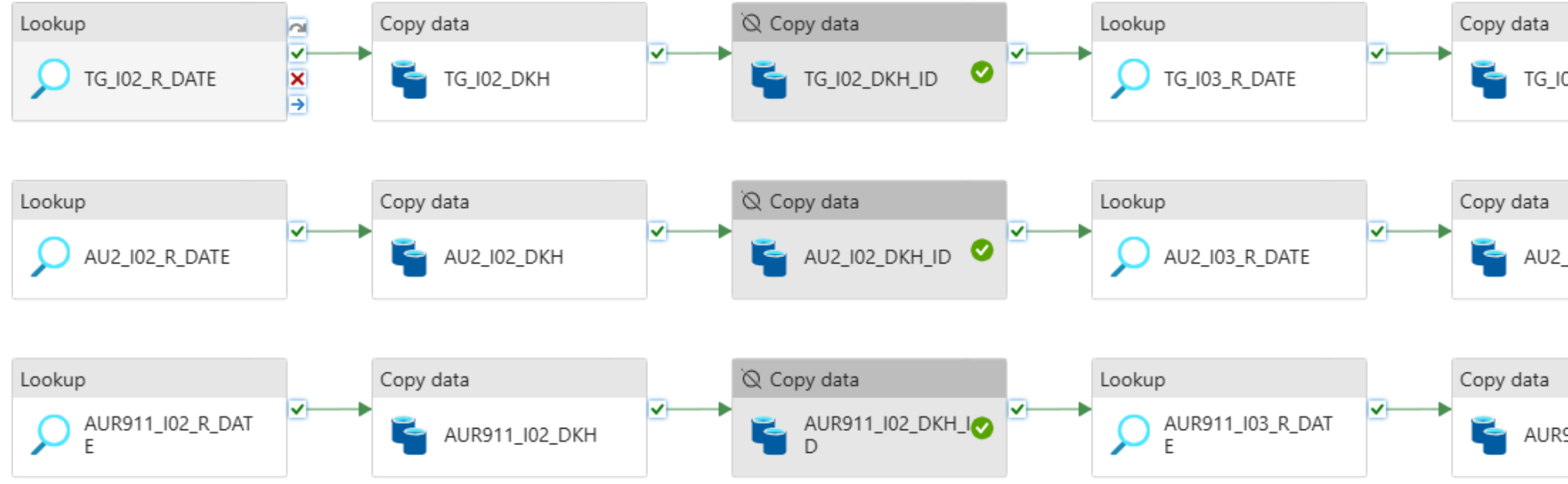


Activities

Search activities

- > Synapse
- > Move and transform
- > Azure Data Explorer
- > Azure Function
- > Batch Service
- > Databricks
- > Data Lake Analytics
- > General
- > HDInsight
- > Iteration & conditionals
- > Machine Learning

✓ Validate ▶ Debug ⚡ Add trigger



Parameters Variables Settings Output

+ New

Complex (cycling) data loads

The screenshot displays the Azure Data Factory pipeline editor for a pipeline named "Snowflake_Update". The interface includes a left-hand "Activities" pane with a search bar and a list of activity categories: Synapse, Move and transform, Azure Data Explorer, Azure Function, Batch Service, Databricks, Data Lake Analytics, General, HDInsight, Iteration & conditionals, and Machine Learning. The main workspace shows a sequence of activities:

- A "Lookup" activity named "TG_I02_Year_month" is connected to a "ForEach" loop labeled "ForEach TG_I02_DKH".
- Inside this "ForEach" loop, an "Activities" container contains a sequence: "TG_I02_DK H" (with a database icon) → "Snowflake_rivile_tmp..." (with a document icon) → a plus sign icon.
- The output of the first "ForEach" loop connects to a "Lookup" activity named "TG_I03_Year_month".
- This second "Lookup" activity connects to another "ForEach" loop labeled "ForEach TG_I03_DKD".
- Inside this second "ForEach" loop, an "Activities" container contains a sequence: "TG_I03_DK D" (with a database icon) → "Snowflake_rivile_tmp..." (with a document icon) → a plus sign icon.
- The output of the second "ForEach" loop connects to a "Lookup" activity named "TG_I04_Year_month".
- Below the main sequence, there are two more "ForEach" loops: "ForEach AU2_I02_DKH" and "ForEach ForEach", which are currently collapsed.

At the bottom of the editor, there are tabs for "Parameters", "Variables", "Settings", and "Output". The "Parameters" tab is active, showing a "+ New" button.

ORCHESTRATION
















Benefits:

- Monitoring
- Complex (cycling) data loads
- Out of the box connectors
- Integration runtime for on-premise data
- Easy to use and scale (re-use pipelines)



Connectors

All Azure Database File Generic protocol NoSQL Services and apps

 Azure Database for MariaDB	 Azure Database for MySQL	 Azure Database for PostgreSQL
 Azure Databricks Delta Lake	 Azure File Storage	 Azure Key Vault
 Azure SQL Database	 Azure SQL Database Managed Instance	 Azure Synapse Analytics
 Azure Table Storage	 Cassandra	 Concur (Preview)
 Couchbase (Preview)	 DB2	 Dataverse (Common Data Service for Apps)

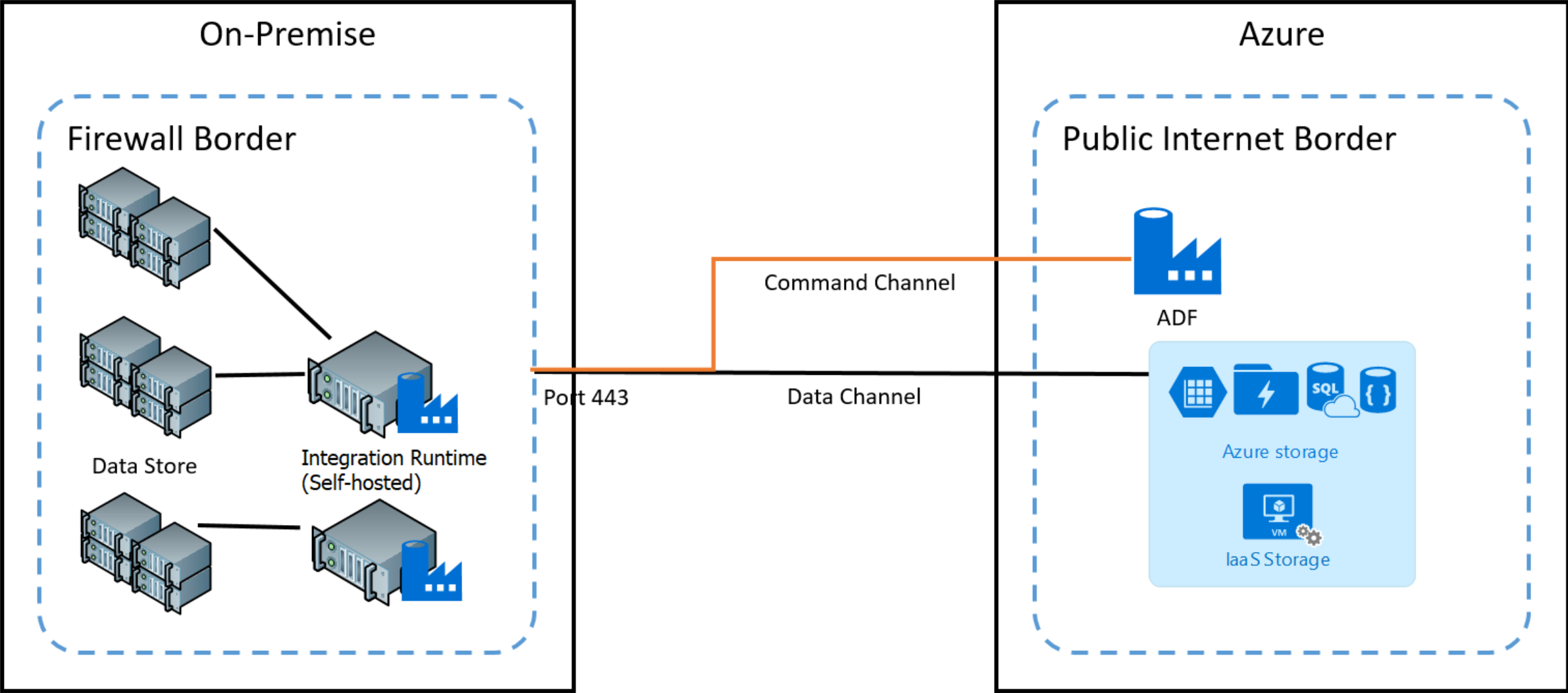
• • • •

ORCHESTRATION

Benefits:

- Monitoring
- Complex (cycling) data loads
- Out of the box connectors
- Integration runtime for on-premise data
- Easy to use and scale (re-use pipelines)





. . . .

.

ORCHESTRATION

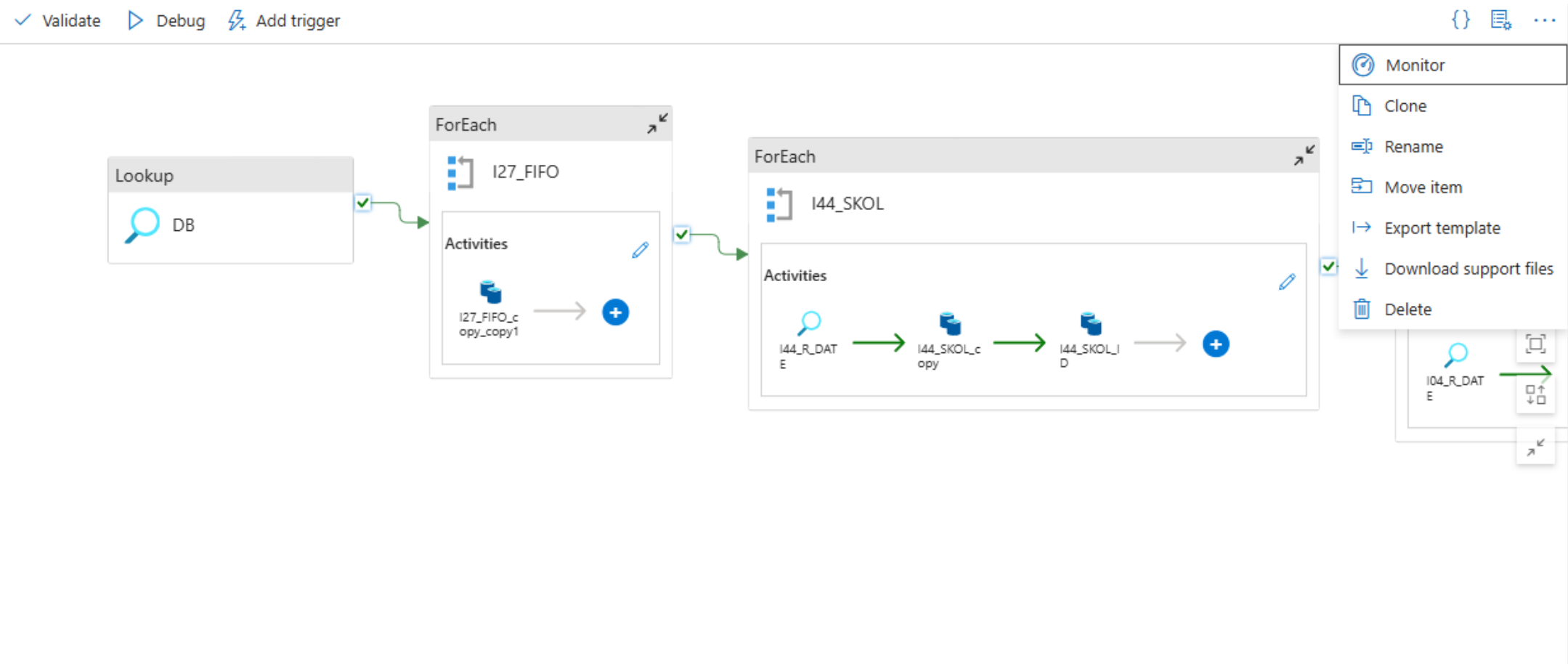
Benefits:

- Monitoring
- Complex (cycling) data loads
- Out of the box connectors
- Integration runtime for on-premise data
- Easy to use and scale (re-use pipelines)



Export as template to use in another Data Factory

- Activities** ⌵ ⌵
- Search activities
- > Move and transform
 - > Synapse
 - > Azure Data Explorer
 - > Azure Function
 - > Batch Service
 - > Databricks
 - > Data Lake Analytics
 - > General
 - > HDInsight
 - > Iteration & conditionals
 - > Machine Learning
 - > Power Query



iNFOTRUST



Thank You

Simas Baranauskas, Infotrust System Architect, s.baranauskas@theinfotrust.com