

May 13th, 2023

#GlobalAzureAthens



Design cost optimized pipelines for Data factory and Azure Synapse





Antonios
Chatzipavlis

Data Solutions Consultant
Trainer



1988



v 6.0

Microsoft
CERTIFIED
Trainer

2000



60 +



2010



Founder



A community for professionals
who use the
Microsoft Data Platform

Connect Explore Learn



2010 est.

Dear Global Azure Athens 2023
sponsors,
your support made all the difference
— **thank you!**



#GlobalAzureAthens



We  Pipelines

We ❤️ Pipelines

Facts

- No upfront cost!
- We just start developing!
- We pay only for usage!

Question

How will it cost per month?

Your Estimate



Your Estimate



^ Azure Data Factory



Azure Data Factory V2 Type, Data Pipeline Service T...



Upfront: \$0.00

Monthly: \$0.00

Azure Data Factory

Region:

West Europe



Type:

Azure Data Factory V2



Service type:

Data Pipeline



This represents a single Data Factory instance

Azure Integration Runtime

^ Orchestration and Execution

\$0.00

Orchestration

0

Activity Runs (in thousands)

= \$0.00

Executions

0

Data integration unit hours



\$0.25

Per hour

= \$0.00


Learn more about [data integration units](#).

0

Pipeline activity execution
hours


\$0.005

Per hour

= \$0.00

0

External Pipeline activity
execution hours


\$0.00025

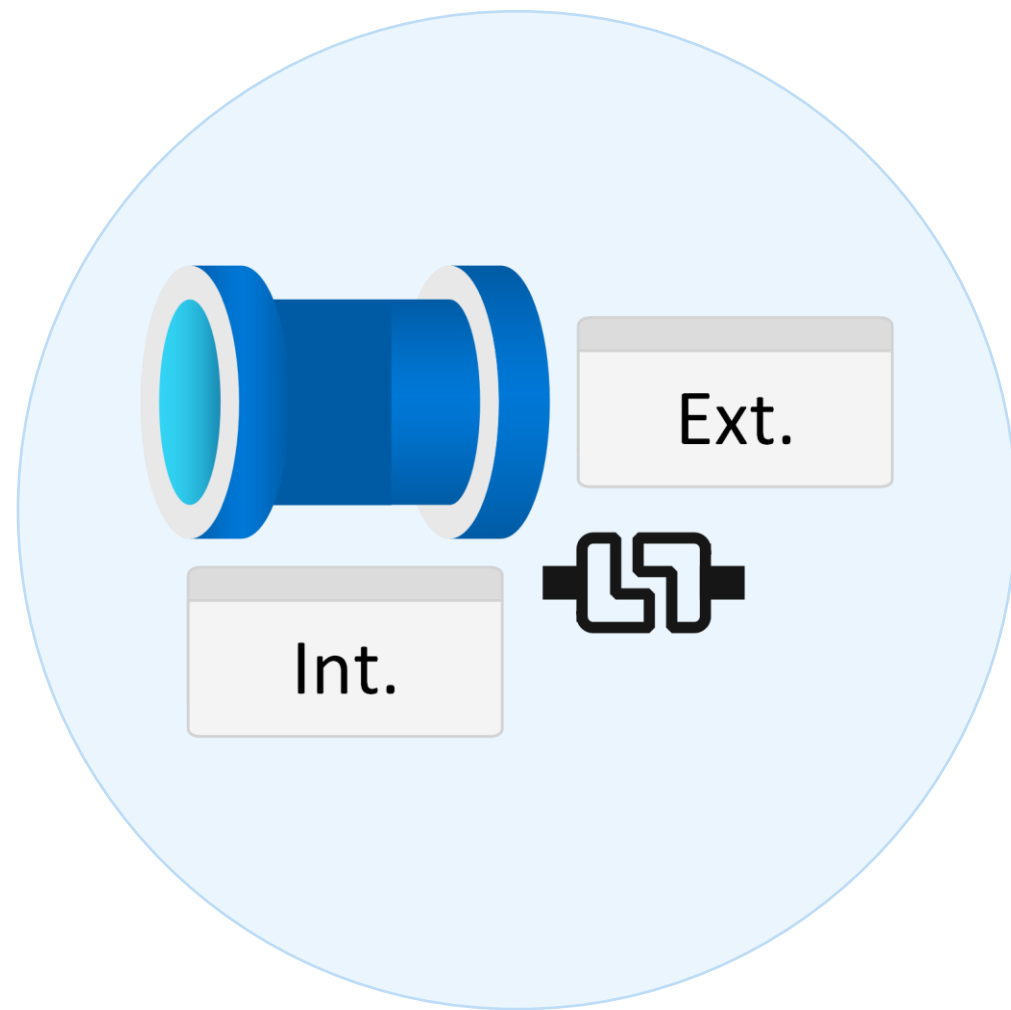
Per hour

= \$0.00

Pipeline activities execute on integration runtime. Pipeline activities include Lookup, Get Metadata, Delete, and schema operations during authoring (test



Pipeline Elements



✓ Operations

✓ Orchestration and Execution

✓ Data Flows

Operations: Azure Data Factory

Read / Write

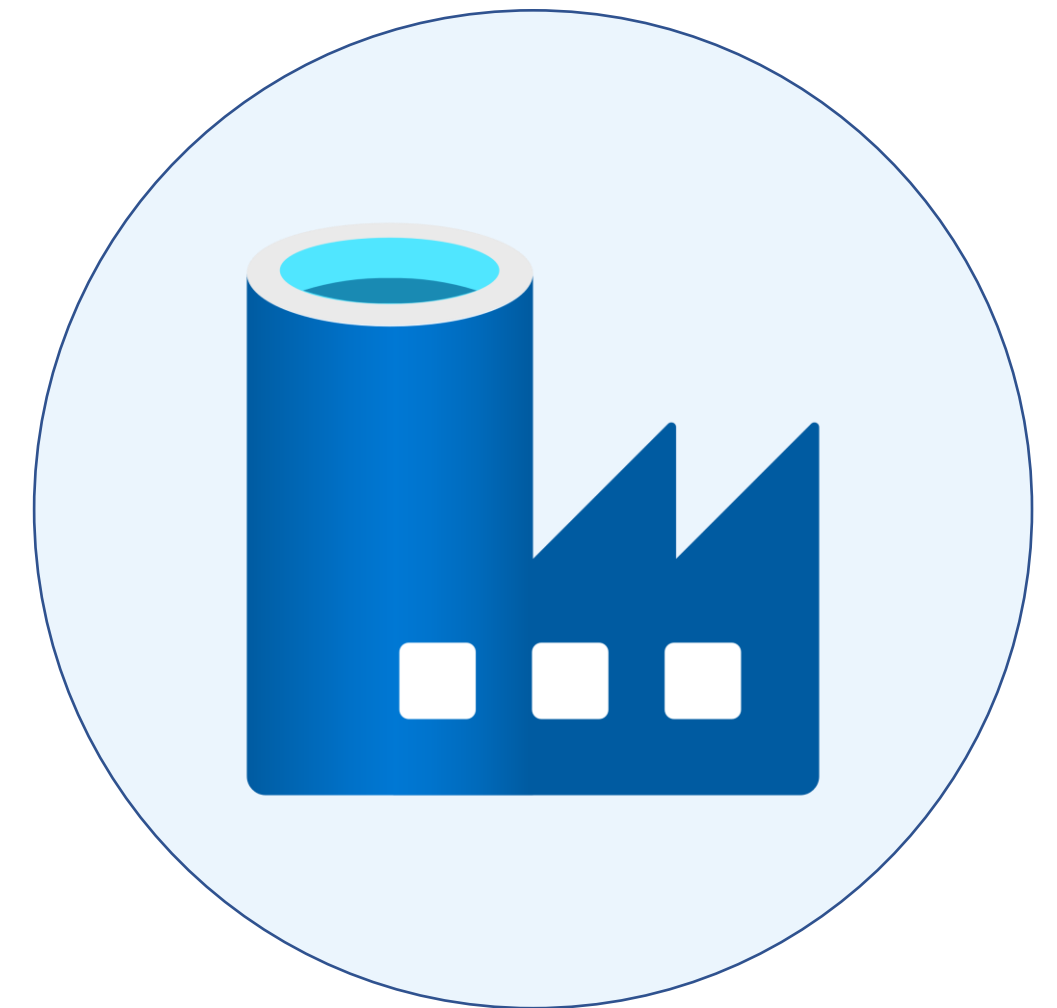
Pricing: 0,50€ per 50.000 entities modified

Operations: Create, Read, Update, Delete

Monitoring

Pricing: 0,25€ per 50.000 run records retrieved

Operations: Get, List



Operations: Azure Synapse

First 1M Operations

Pricing: Free

Operations: Create, Read, Update, Delete, Get, List

Additional Operations

Pricing: 0,25€ per 50.000 operations

Operations: Create, Read, Update, Delete, Get, List



Integration Runtimes (IR)



Azure IR



Azure Managed VNet IR



Self-Hosted IR

Prices depend on which IR you are using

Pipeline Pricing

Orchestration per 1000 runs:
Activity runs, debug runs, trigger execution



Azure IR

1€



Azure Managed VNet IR

1€



Self-Hosted IR

1,50€

Pipeline Pricing

Orchestration per runs:
Activity runs, debug runs, trigger execution



Azure IR

0,001€



Azure Managed VNet IR

0,001€



Self-Hosted IR

0,0015€

Pipeline Pricing



Azure IR

1€

0,25€



Azure Managed VNet IR

1€

0,25€



Self-Hosted IR

1,50€

0,10€

Orchestration per 1000 runs:
Activity runs, debug runs, trigger execution

Data Movement per DIU/hour:
Copy Data Activity

Pipeline Pricing

Orchestration per 1000 runs:
Activity runs, debug runs, trigger execution

Data Movement per DIU/hour:
Copy Data Activity

Copy data

Copy data1

General Source Sink Mapping **Settings** User properties

i You will be charged **# of used DIUs * copy duration * \$0.25/DIU-hour.** Local currency and separate discounting may apply per subscription type. [Learn more](#)

Maximum data integration unit ⁱ Auto ☐ Edit

Degree of copy parallelism ⁱ ☐ Edit

Data consistency verification ⁱ ☐

Fault tolerance ⁱ ☐

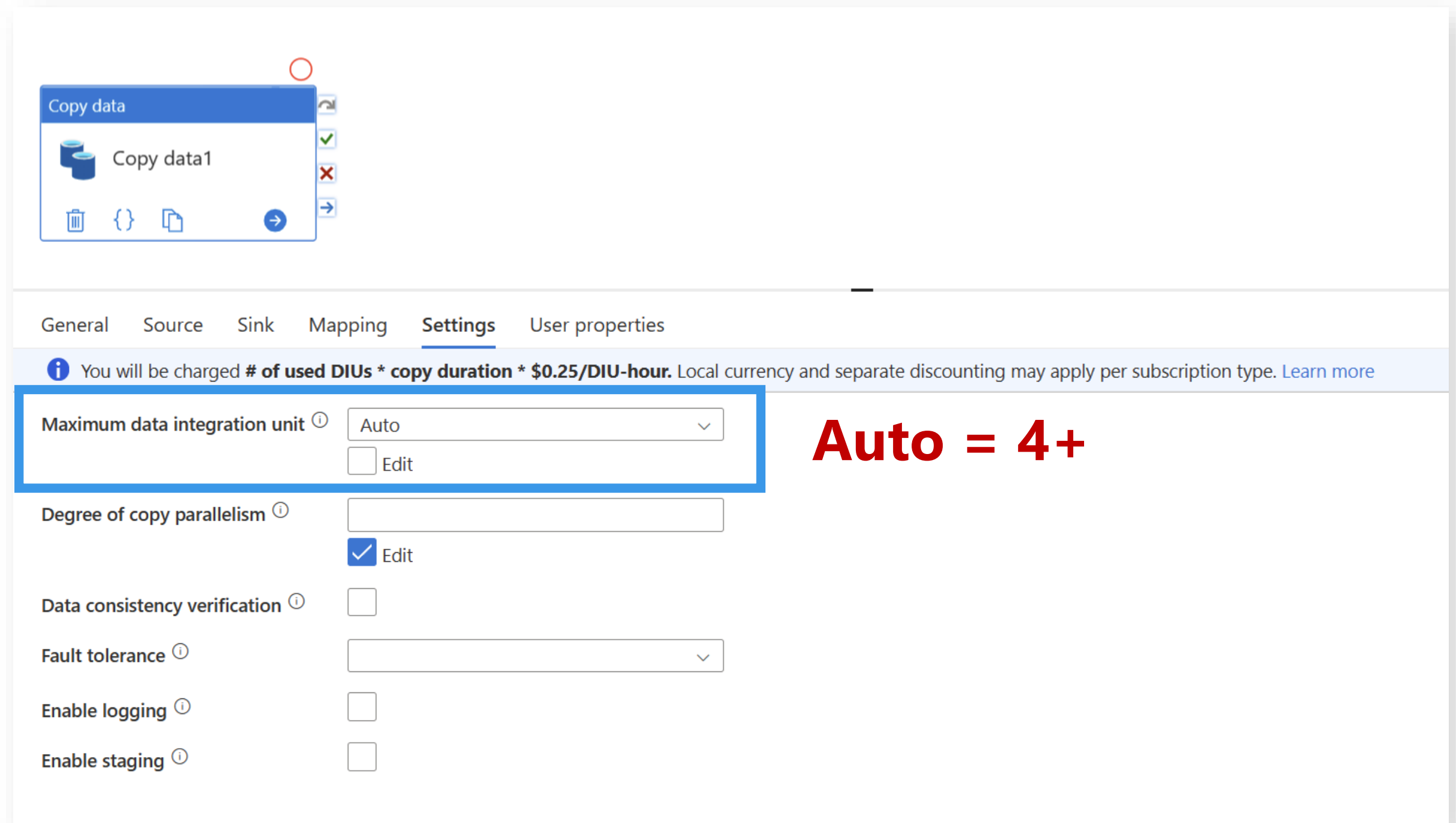
Enable logging ⁱ ☐

Enable staging ⁱ ☐

Pipeline Pricing

Orchestration per 1000 runs:
Activity runs, debug runs, trigger execution

Data Movement per DIU/hour:
Copy Data Activity



Copy data

Copy data1

General Source Sink Mapping **Settings** User properties

i You will be charged **# of used DIUs * copy duration * \$0.25/DIU-hour**. Local currency and separate discounting may apply per subscription type. [Learn more](#)

Maximum data integration unit ⁱ Auto ☐ Edit

Degree of copy parallelism ⁱ ☒ Edit

Data consistency verification ⁱ ☐

Fault tolerance ⁱ ☐

Enable logging ⁱ ☐

Enable staging ⁱ ☐

Auto = 4+

Pipeline Pricing

Orchestration per 1000 runs:
Activity runs, debug runs, trigger execution

Data Movement per DIU/hour:
[Copy Data Activity](#)

Copy data

Copy data1

General Source Sink Mapping **Settings** User properties

i You will be charged **# of used DIUs * copy duration * \$0.25/DIU-hour**. Local currency and separate discounting may apply per subscription type. [Learn more](#)

Maximum data integration unit ⓘ Auto

Degree of copy parallelism ⓘ

Data consistency verification ⓘ

Fault tolerance ⓘ

Enable logging ⓘ

Enable staging ⓘ

Filter...

Auto

2

4

8

16

32

MINIMUM

Pipeline Pricing



Azure IR



Azure Managed VNet IR



Self-Hosted IR

Orchestration per 1000 runs:
Activity runs, debug runs, trigger execution

1€

1€

1,50€

Data Movement per DIU/hour:
Copy Data Activity

0,25€

0,25€

0,10€

Pipeline Activity per hour:
Lookup, Get Metadata, Delete etc.

0,005€

1€

0,002€

Pipeline Pricing



Azure IR



Azure Managed VNet IR



Self-Hosted IR

Orchestration per 1000 runs:
Activity runs, debug runs, trigger execution

1€

1€

1,50€

Data Movement per DIU/hour:
Copy Data Activity

0,25€

0,25€

0,10€

Pipeline Activity per hour:
Lookup, Get Metadata, Delete etc.

0,005€

1€

0,002€

External Activity per hour:
Stored Procedure, Notebooks etc.

0,00025€

1€

0,0001€



Pipeline Pricing



Azure IR



Azure Managed VNet IR



Self-Hosted IR

Orchestration per 1000 runs:
Activity runs, debug runs, trigger execution

1€

1€

1,50€

Data Movement per DIU/hour:
Copy Data Activity

0,25€

0,25€

0,10€

Pipeline Activity per hour:
Lookup, Get Metadata, Delete etc.

0,005€

1€

0,002€

External Activity per hour:
Stored Procedure, Notebooks etc.

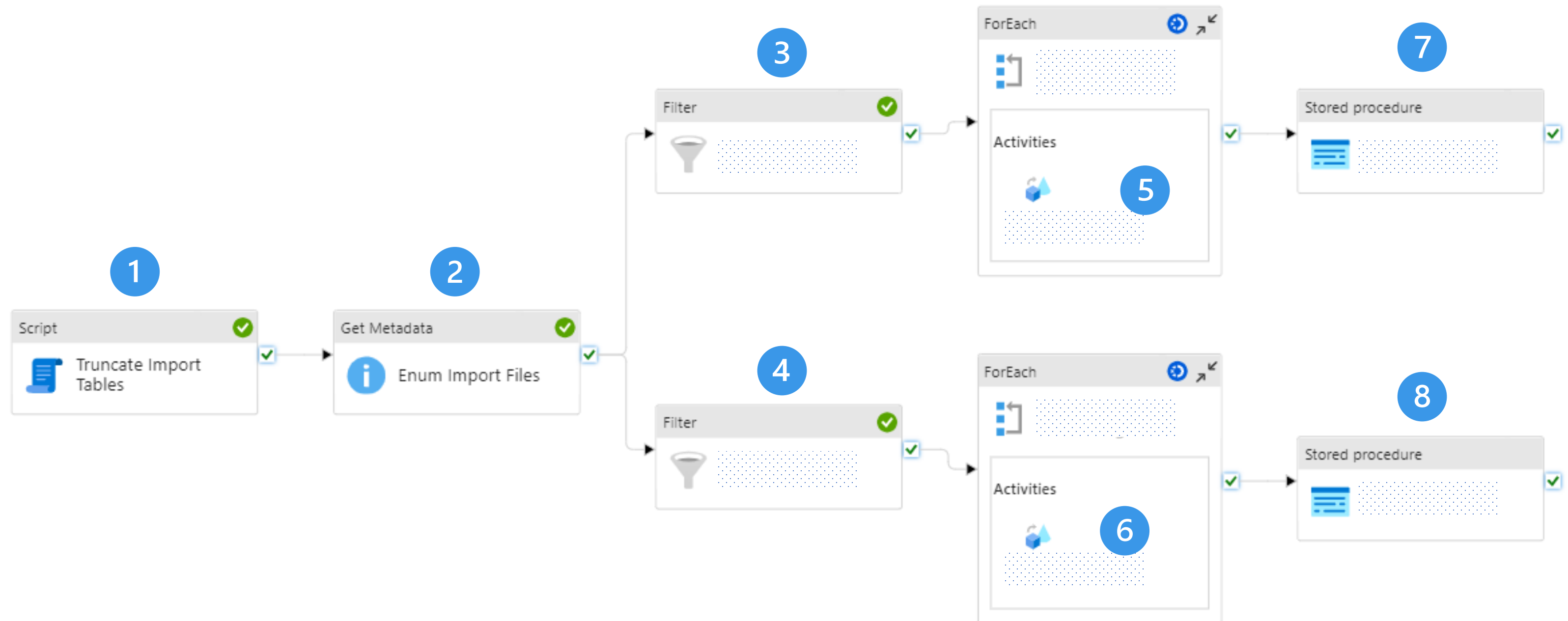
0,00025€

1€

0,0001€



Activity runs example



Activity runs example

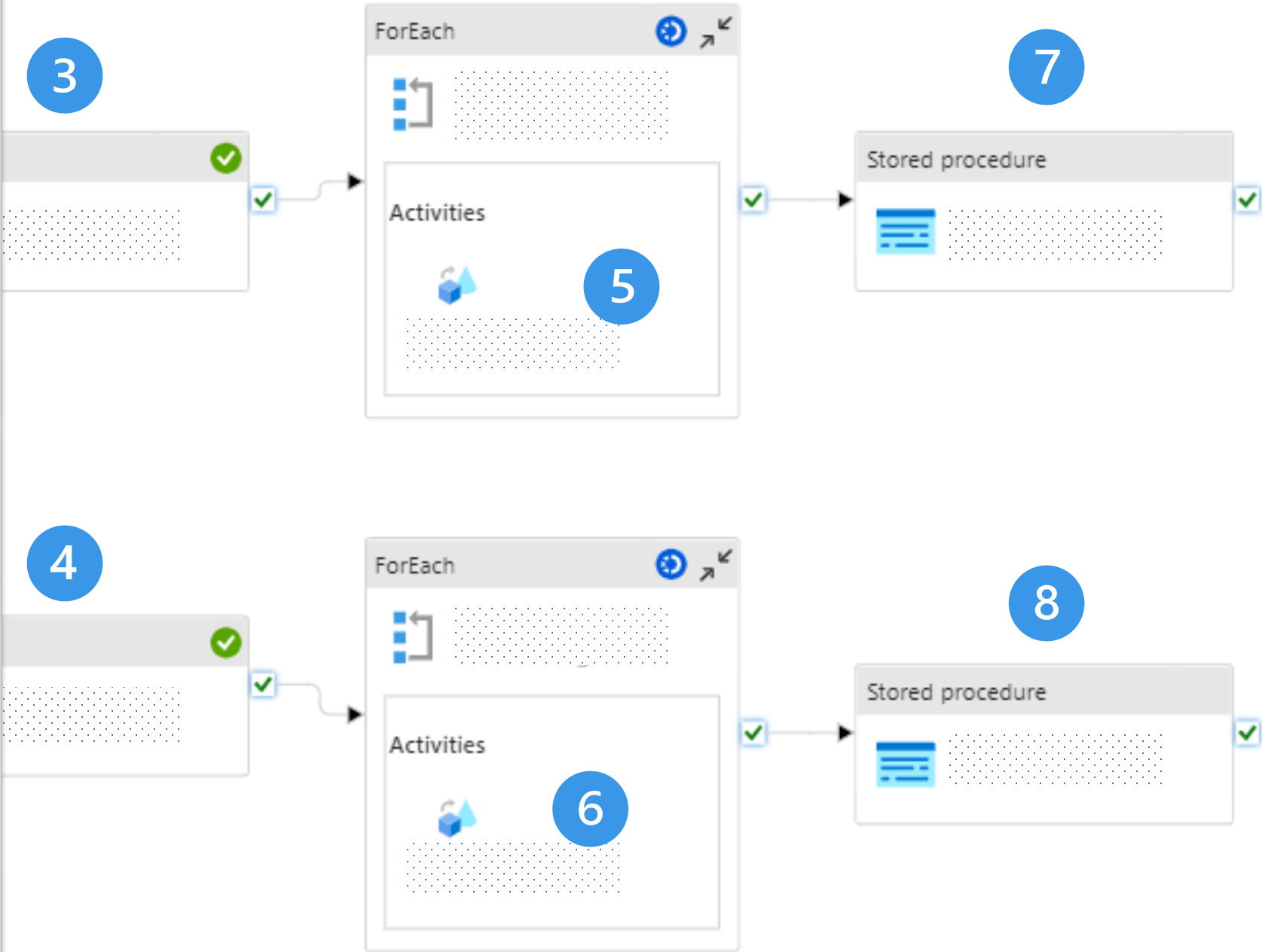
Pipeline run consumption

Name
JsonFilesProcessPipelineIv

Status
✔ Succeeded

Run ID
6a7e

	Quantity	Unit
Pipeline orchestration		
Activity runs	11	Activity runs
Pipeline execution		
Azure integration runtime		
Pipeline activities	0.0500	Execution hours
External activities	0.0333	Execution hours
Data flow		
General purpose	2.6904	vCore-hour



Rounding Up

Activity runs

Pipeline run ID 6a70fa7e

All status ▾ List ▾

Showing 1 - 11 items

Activity name ↑↓	Status ↑↓	Activity type ↑↓	Run start ↑↓	Duration ↑↓	Log
	✓ Succeeded	Stored procedure		00:00:33	
	✓ Succeeded	Stored procedure		00:00:12	
	✓ Succeeded	Data flow		00:07:06	
	✓ Succeeded	Data flow		00:06:31	
	✓ Succeeded	Data flow		00:06:55	
	✓ Succeeded	ForEach		00:07:10	
	✓ Succeeded	ForEach		00:06:35	
	✓ Succeeded	Filter		00:00:01	
	✓ Succeeded	Filter		00:00:01	
	✓ Succeeded	Get Metadata		00:00:02	
	✓ Succeeded	Script		00:01:13	

Consumption example



5 seconds

= 1 minute

= 0,0167 hours

= 0,0167 hours *
0,005 / hours

= **0,0000835**

22 seconds

= 1 minute

= 0,0167 hours

= 0,0167 hours * 4
DIUs * 0,24 DIU/h

= **0,0167**

1 min 10 seconds

= 2 minutes

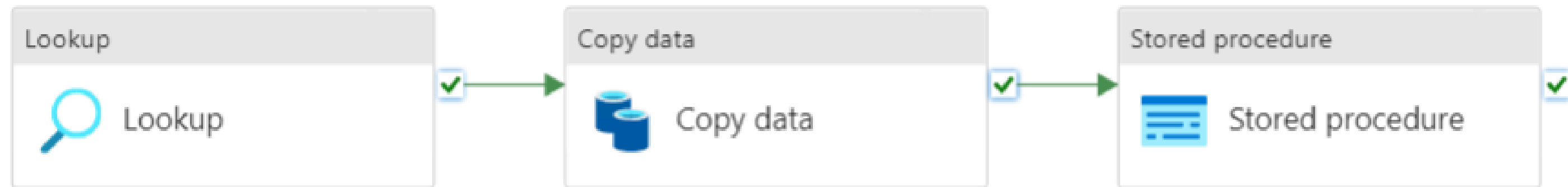
= 0,0333 hours

= 0,0333 hours *
0,00025 / h

= **0,000008325**

Total cost per pipeline run = 0,053

Consumption example



5 seconds

= 1 minute

= 0,0167 hours

= 0,0167 hours *
0,005 / hours

= **0,0000835**

22 seconds

= 1 minute

= 0.0167 hours

= 0.0167 hours * 4
DIUs * 0,24 DIU/h

= **0,0167**

1 min 10 seconds

= 2 minutes

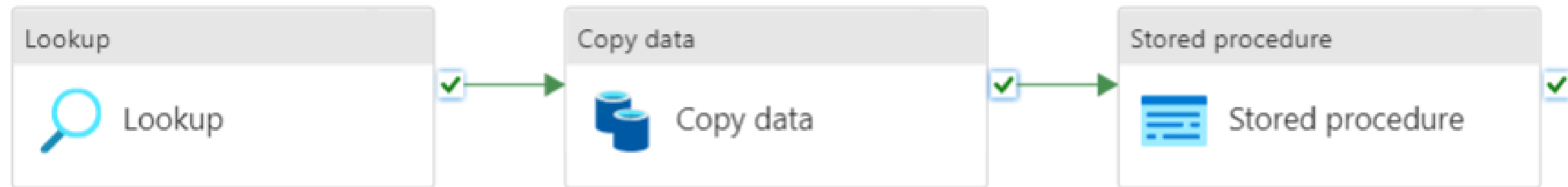
= 0,0333 hours

= 0,0333 hours *
0,00025 / h

= **0,000008325**

Run once per night x 30 days = 1,59

Consumption example



5 seconds

= 1 minute

= 0,0167 hours

= 0,0167 hours *
0,005 / hours

= **0,0000835**

22 seconds

= 1 minute

= 0.0167 hours

= 0.0167 hours * 4
DIUs * 0,24 DIU/h

= **0,0167**

1 min 10 seconds

= 2 minutes

= 0,0333 hours

= 0,0333 hours *
0,00025 / h

= **0,000008325**

Run for 100 tables per night x 30 days = 159

Consumption example



5 seconds

= 1 minute

= 0,0167 hours

= 0,0167 hours *
0,005 / hours

= **0,0000835**

22 seconds

= 1 minute

= 0.0167 hours

= 0.0167 hours * 4
DIUs * 0,24 DIU/h

= **0,0167**

1 min 10 seconds

= 2 minutes

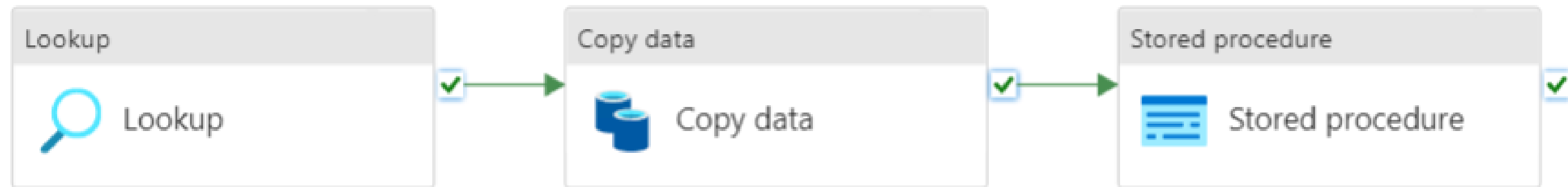
= 0,0333 hours

= 0,0333 hours *
0,00025 / h

= **0,000008325**

Run for 100 tables every hour x 30 days = 3.816

Consumption example



5 seconds

= 1 minute

= 0,0167 hours

= 0,0167 hours *
0,005 / hours

= **0,0000835**

22 seconds

= 1 minute

= 0.0167 hours

= 0.0167 hours * 4
DIUs * 0,24 DIU/h

= **0,0167**

1 min 10 seconds

= 2 minutes

= 0,0333 hours

= 0,0333 hours *
0,00025 / h

= **0,000008325**

Run for 100 tables every 5 min x 30 days = 45.792

Data Flows

Basic - General Purpose

Pricing: 0,286€ per vCore/hour

Standard – Memory Optimized

Pricing: 0,346€ per vCore/hour



Data Flows – Cluster Size

Basic

General Purpose 4 (+4 Drive cores)

Pricing: $0,286\text{€} \times 8 = 2,228\text{€} / \text{hour}$

General Purpose 256 (+16 Drive cores)

Pricing: $0,286\text{€} \times 272 = 77,792\text{€} / \text{hour}$

Standard

Memory Optimized 4 (+4 Drive cores)

Pricing: $0,346\text{€} \times 8 = 2,786\text{€} / \text{hour}$

Memory Optimized 256 (+16 Drive cores)

Pricing: $0,346\text{€} \times 8 = 94,112\text{€} / \text{hour}$

The screenshot shows the 'Compute size' configuration for an Azure Databricks cluster. The 'Compute size' is set to 'Custom'. Under the 'Advanced' section, the 'Compute type' is 'Basic (General Purpose)'. The 'Core count' dropdown menu is open, showing a list of options: '4 (+ 4 Driver cores)', '8 (+ 8 Driver cores)', '16 (+ 16 Driver cores)', '32 (+ 16 Driver cores)', '64 (+ 16 Driver cores)', and '128 (+ 16 Driver cores)'. The '4 (+ 4 Driver cores)' option is currently selected.

Consumption example

An input dataset on Azure Storage

An output dataset on Azure Storage

Triggered every hour for 8 hours / day



Consumption

Run pipeline

= 2 activities x 8 execution per day x 30 days = 480 activity runs

rounded up since the calculator allow increments of 1000

Data flow takes 10 min per execution

= 10 m / 60 * 8 executions * 30 days = 40 hours

Case calculation



Thank you!





Please evaluate !



<https://bit.ly/GA23Evaluation>

#GlobalAzureAthens

A big **thank you** to our sponsors!

