



Lessons Learned:

Understanding Pipeline Pricing

in Azure Data Factory and Azure Synapse Analytics

Cathrine Wilhelmsen

DataMinutes #2 • January 21st, 2022



Session Abstract

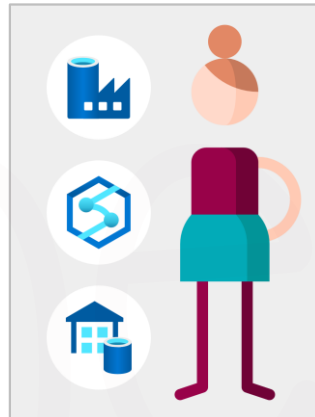
Azure Data Factory and Synapse Pipelines pricing is easy, right? No upfront costs. Pay only for what you use. It's a wonderful world for data integration developers. Just a few clicks and your solution is ready!

But what do you present to management when they ask for cost estimates? "I guess we just have to wait for the next invoice" is rarely an acceptable answer. It is definitely not an acceptable answer if that invoice turns out to be unexpectedly high.

In this session, we will look at the bits and pieces that make up the Azure Data Factory and Synapse Pipelines pricing models, how to monitor consumption and manage costs, and cover a few important gotchas to be aware of...

cathrine

WILHELMSEN



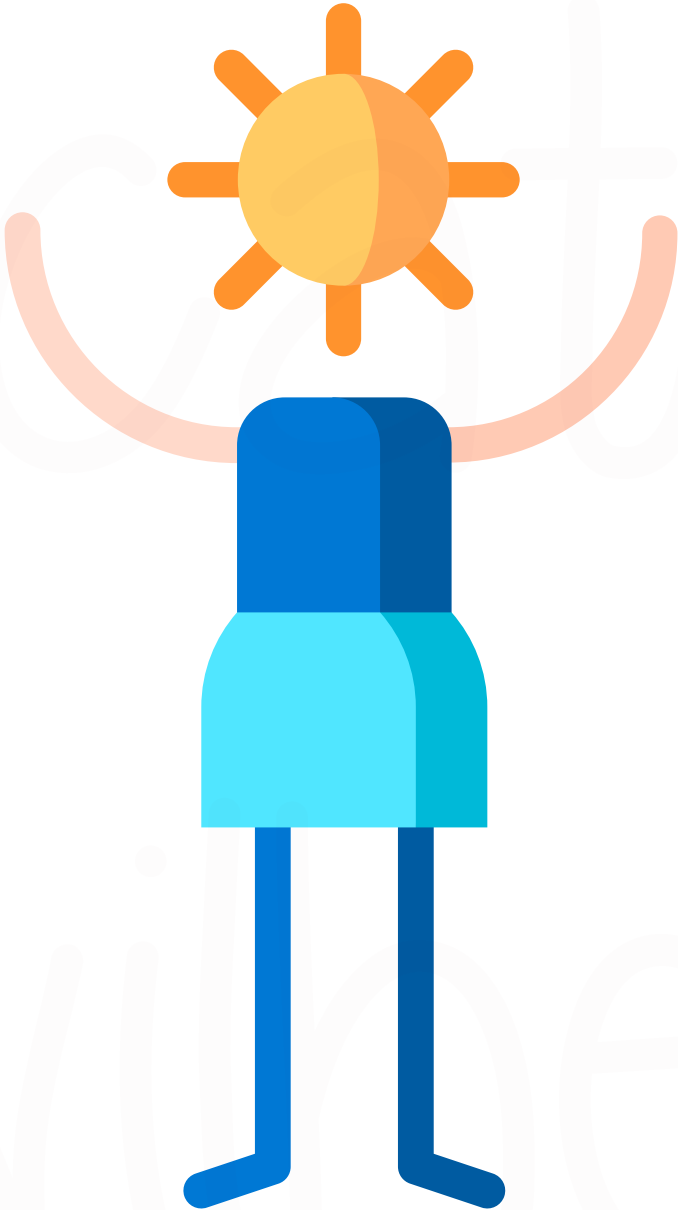
@cathrinew



cathrinew.net

Pipelines are
awesome, right?

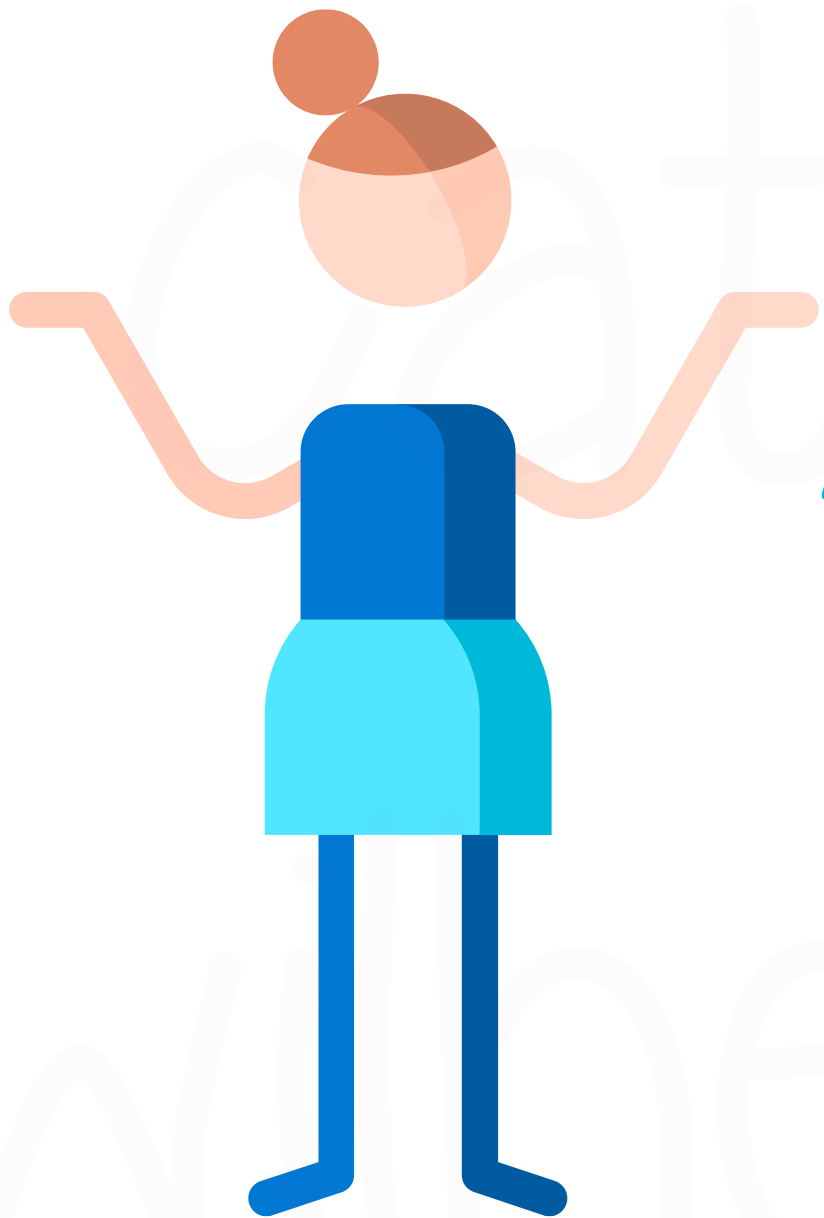




No upfront costs!
Just start developing!
Pay only for usage!

“But how much will
it cost per month?”





"I guess we just have to
wait for the invoice!"



Pipeline Pricing



Operations



Orchestration and Execution



Data Flows



Operations

(Pretty cheap, but everything has a cost)



Orchestration and Execution



Data Flows

Operations: Azure Data Factory



Read / Write: \$0.50 per 50 000 entities modified

Operations: Create, Read, Update, Delete

Monitoring: \$0.25 per 50 000 run records retrieved

Operations: Get, List

** Prices in USD from January 2022*

Operations: Azure Synapse



First 1 Million Operations: **Free!**

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

Additional Operations: **\$0.25 per 50 000 operations**

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

** Prices in USD from January 2022*



Operations



Orchestration and Execution

(Be aware of the biggest gotchas)



Data Flows

Pipeline Pricing: Integration Runtimes



Prices depend on which
Integration Runtime (IR)
you are using

Pipeline Pricing: Orchestration



Azure IR

\$1



Azure Managed VNet IR

\$1



Self-Hosted IR

\$1.50

Orchestration per 1000 runs:

Activity runs, debug runs, trigger executions

** Prices in USD from January 2022. All activities are prorated by the minute and rounded up.*

Pipeline Pricing: Orchestration



Azure IR

\$0.001



Azure Managed VNet IR

\$0.001



Self-Hosted IR

\$0.0015

Orchestration per run:

Activity runs, debug runs, trigger executions

** Prices in USD from January 2022. All activities are prorated by the minute and rounded up.*

Pipeline Pricing: Data Movement



Azure IR

\$1



Azure Managed VNet IR

\$1



Self-Hosted IR

\$1.50

Orchestration per 1000 runs:

Activity runs, debug runs, trigger executions

Data Movement per DIU-hour:

Copy Data Activity

\$0.25

\$0.25

\$0.10

** Prices in USD from January 2022. All activities are prorated by the minute and rounded up.*

Pipeline Pricing: Data Movement



Azure IR

\$1



Azure Managed VNet IR

\$1



Self-Hosted IR

\$1.50

Orchestration per 1000 runs:

Activity runs, debug runs, trigger executions

Data Movement per DIU-hour:

Copy Data Activity

\$0.25

\$0.25

\$0.10

** Prices in USD from January 2022. All activities are prorated by the minute and rounded up.*

Pipeline Pricing: Data Movement



Orchestration per 1000 runs:

Activity runs, debug runs, trigger executions

Data Movement per DIU-hour:

Copy Data Activity

The screenshot shows the Azure Data Factory interface. At the top, a pipeline is visualized with three activities: 'Lookup', 'Copy data', and 'Stored procedure'. The 'Copy data' activity is highlighted with a red circle. Below the pipeline, the 'Settings' tab is selected. A pink box highlights a warning message: "You will be charged # of used DIUs * copy duration * \$0.25/DIU-hour." Below this, the 'Data integration unit' is set to 'Auto' in a dropdown menu, with an 'Edit' button next to it. A pink dotted arrow points from the 'Data integration unit' dropdown to the pricing warning box.

** Prices in USD from January 2022. All activities are prorated by the minute and rounded up.*

Pipeline Pricing: Data Movement



Orchestration per 1000 runs:

Activity runs, debug runs, trigger executions

Data Movement per DIU-hour:

Copy Data Activity

The screenshot displays the Azure Data Factory pipeline editor. A pipeline is shown with three activities: 'Lookup', 'Copy data', and 'Stored procedure'. The 'Copy data' activity is highlighted, and its 'Settings' tab is selected. The 'Data integration unit' dropdown is set to 'Auto', which is highlighted with a red box. Below the dropdown, the text 'Auto = 4+' is displayed in red. The 'Settings' tab also shows a warning message: 'You will be charged # of used DIUs * copy duration * \$0.25/DIU-hour.'

** Prices in USD from January 2022. All activities are prorated by the minute and rounded up.*

Pipeline Pricing: Data Movement



Orchestration per 1000 runs:

Activity runs, debug runs, trigger executions

Data Movement per DIU-hour:

Copy Data Activity

Lookup → Copy data → Stored procedure

General Source Sink Mapping **Settings** User properties

i You will be charged # of used DIUs * copy duration * \$0.25/DIU-hour.

Data integration unit ⓘ

Auto

Filter...

Auto

2

4

8

16

32

Minimum = 2

* Prices in USD from January 2022. All activities are prorated by the minute and rounded up.

Pipeline Pricing: Pipeline Activity



Azure IR

\$1



Azure Managed VNet IR

\$1



Self-Hosted IR

\$1.50

Orchestration per 1000 runs:

Activity runs, debug runs, trigger executions

Data Movement per DIU-hour:

Copy Data Activity

\$0.25

\$0.25

\$0.10

Pipeline Activity per hour:

Lookup, Get Metadata, Delete etc.

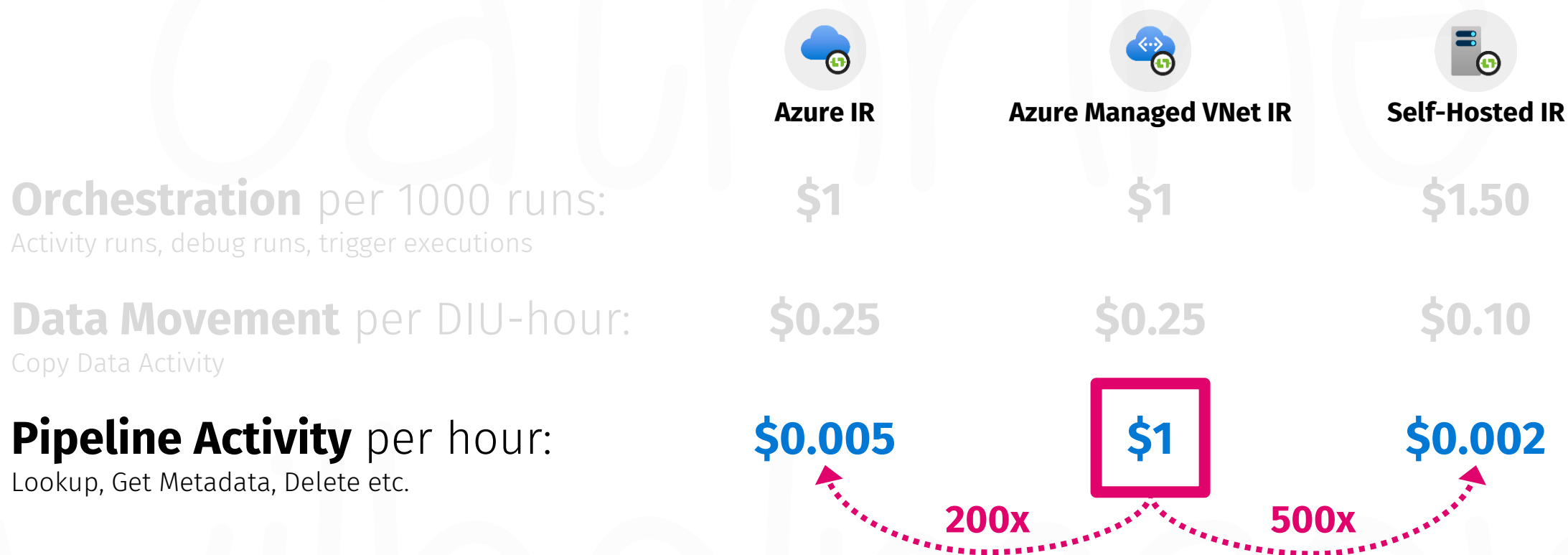
\$0.005

\$1

\$0.002

** Prices in USD from January 2022. All activities are prorated by the minute and rounded up.*

Pipeline Pricing: Pipeline Activity



* Prices in USD from January 2022. All activities are prorated by the minute and rounded up.

Pipeline Pricing: External Activity



Azure IR



Azure Managed VNet IR



Self-Hosted IR

Orchestration per 1000 runs:

Activity runs, debug runs, trigger executions

\$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

** Prices in USD from January 2022. All activities are prorated by the minute and rounded up.*

Pipeline Pricing: External Activity



	 Azure IR	 Azure Managed VNet IR	 Self-Hosted IR
Orchestration per 1000 runs: <small>Activity runs, debug runs, trigger executions</small>	\$1	\$1	\$1.50
Data Movement per DIU-hour: <small>Copy Data Activity</small>	\$0.25	\$0.25	\$0.10
Pipeline Activity per hour: <small>Lookup, Get Metadata, Delete etc.</small>	\$0.005	\$1	\$0.002
External Activity per hour: <small>Stored Procedure, Notebooks etc.</small>	\$0.00025	\$1	\$0.0001

* Prices in USD from January 2022. All activities are prorated by the minute and rounded up.

Pipeline Pricing



Azure IR

\$1



Azure Managed VNet IR

\$1



Self-Hosted IR

\$1.50

Orchestration per 1000 runs:

Activity runs, debug runs, trigger executions

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

** Prices in USD from January 2022. All activities are prorated by the minute and rounded up.*

Pipeline Pricing: Rounding Up



Azure IR



Azure Managed VNet IR



Self-Hosted IR

Orchestration per 1000 runs:

Activity runs, debug runs, trigger executions

\$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

* Prices in USD from January 2022. All activities are prorated by the minute and rounded up.

Pipeline Pricing: Rounding Up



Name	Type	Run start	Duration	Status
Stored Procedure	Stored procedure	[redacted]	00:01:06	✓ Succeeded
Copy Data	Copy data	[redacted]	00:00:11	✓ Succeeded
Lookup	Lookup	[redacted]	00:00:02	✓ Succeeded

External Activity per hour:

Stored Procedure, Notebooks etc.

\$0.00025

\$1

\$0.0001

* Prices in USD from January 2022. All activities are prorated by the minute and rounded up.

Pipeline Pricing: Rounding Up



Lookup

Lookup

Copy data

Copy Data

Stored procedure

Stored Procedure

Parameters

Variables

Settings

Output

Pipeline run ID: [redacted] [share] [refresh] [info]

Name	Type	Run start	Duration	Status
Stored Procedure	Stored procedure	[redacted]	00:01:06	✓ Success
Copy Data	Copy data	[redacted]	00:00:11	✓ Success
Lookup	Lookup	[redacted]	00:00:02	✓ Success

Debug run consumption

	Quantity	Unit
Pipeline orchestration		
Activity Runs	3	Activity runs
Pipeline execution		
Azure integration runtime		
Data movement activities	0.0167	DIU-hour
Pipeline activities	0.0167	Execution hours
External activities	0.0333	Execution hours

[Learn more](#) [Pricing calculator](#)

External Activity per hour:

Stored Procedure, Notebooks etc.

\$0.00025

\$1

\$0.0001

** Prices in USD from January 2022. All activities are prorated by the minute and rounded up.*

Pipeline Pricing: Rounding Up



The screenshot shows a pipeline run with three activities: Lookup, Copy data, and Stored procedure. The 'Output' tab is selected, showing a table of activity runs. The 'Duration' column for the 'Copy Data' activity is highlighted with a red box, showing 00:00:11. A red dotted arrow points from this box to the 'Debug run consumption' table, specifically to the '0.0167' value in the 'Pipeline activities' row.

Name	Type	Run start	Duration	Status
Stored Procedure	Stored procedure		00:01:06	Success
Copy Data	Copy data		00:00:11	Success
Lookup	Lookup		00:00:02	Success

	Quantity	Unit
Pipeline orchestration		
Activity Runs	3	Activity runs
Pipeline execution		
Azure integration runtime		
Data movement activities	0.0167	DIU-hour
Pipeline activities	0.0167	Execution hours
External activities	0.0333	Execution hours

0.0167	DIU-hour
0.0167	Execution hours
0.0333	Execution hours

0.0167 hours
= 1 minute

External Activity per hour:
Stored Procedure, Notebooks etc.

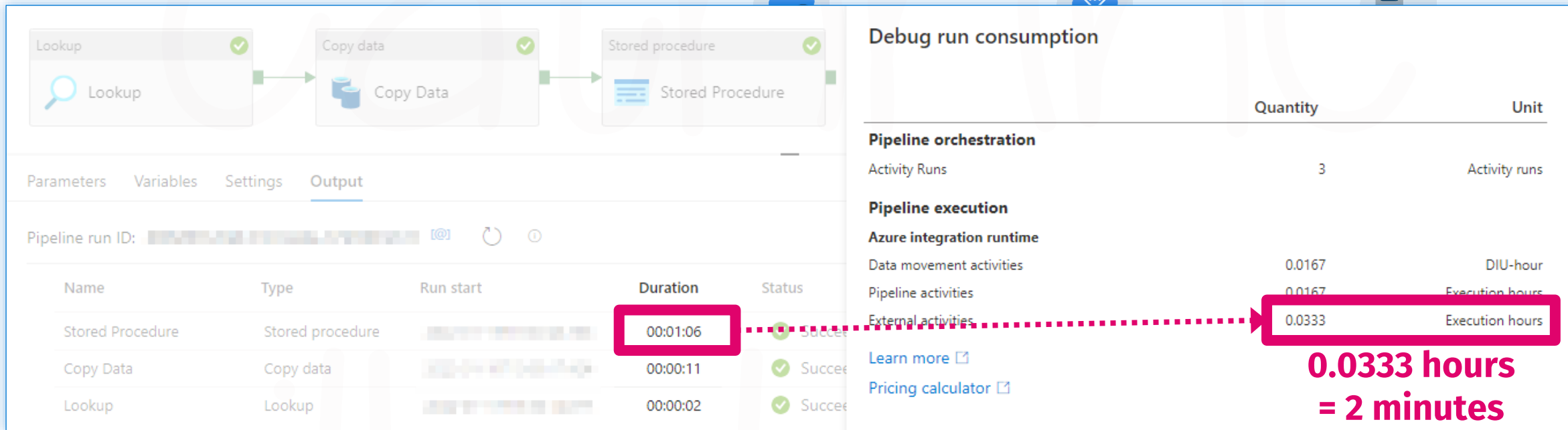
\$0.00025

\$1

\$0.0001

* Prices in USD from January 2022. All activities are prorated by the minute and rounded up.

Pipeline Pricing: Rounding Up



External Activity per hour:
Stored Procedure, Notebooks etc.

\$0.00025

\$1

\$0.0001

** Prices in USD from January 2022. All activities are prorated by the minute and rounded up.*



Operations



Orchestration and Execution



Data Flows

(Can be expensive, but worth it for the right use cases)

Data Flows



Basic: \$0.286 per vCore-hour

General Purpose

Standard: \$0.346 per vCore-hour

Memory Optimized

** Prices in USD from January 2022. All activities are prorated by the minute and rounded up. The minimum cluster size is 8 vCores.*

Data Flows: Rounding Up



Basic: \$0.286 per vCore-hour

General Purpose

Standard: \$0.346 per vCore-hour

Memory Optimized

** Prices in USD from January 2022. All activities are prorated by the minute and rounded up. The minimum cluster size is 8 vCores.*

Data Flows: Cluster Size



Basic: \$0.286 per vCore-hour

General Purpose

Standard: \$0.346 per vCore-hour

Memory Optimized

** Prices in USD from January 2022. All activities are prorated by the minute and rounded up. The minimum cluster size is 8 vCores.*

Data Flows: Cluster Size



Basic: \$0.286

General Purpose

Standard: \$0

Memory Optimized

▼ Data flow run time

Compute type *
General purpose ▼

Core count *
4 (+ 4 Driver cores) ▼

Time to live ⓘ
10 minutes ▼

☒ Quick re-use ⓘ

Billing for data flows is based upon the type of compute you select and the number of cores selected per hour. If you set a TTL, then the minimum billing time will be that amount of time. Otherwise, the time billed will be based on the execution time of your data flows and the time of your debug sessions. Note that debug sessions will incur a minimum of 60 minutes of billing time unless you switch off the debug session manually. For further details, please click [here](#) for the pricing page.

** Prices in USD from January 2022. All activities are prorated by the minute and rounded up. The minimum cluster size is 8 vCores.*

Data Flows: Cluster Size



Basic: \$0.286

General Purpose

Standard: \$0

Memory Optimized

▼ Data flow run time

Compute type *

General purpose ▼

Core count *

4 (+ 4 Driver cores) ▼

Filter...

4 (+ 4 Driver cores)
8 (+ 8 Driver cores)
16 (+ 16 Driver cores)
32 (+ 16 Driver cores)
64 (+ 16 Driver cores)
128 (+ 16 Driver cores)
256 (+ 16 Driver cores)

* Prices in USD from January 2022. All activities are prorated by the minute and rounded up. The minimum cluster size is 8 vCores.

Data Flows: Cluster Size



Basic: \$0.286 per vCore-hour = \$2.288 per hour

General Purpose: 4 (+4 Driver Cores)

Standard: \$0.346 per vCore-hour = \$2.786 per hour

Memory Optimized: 4 (+4 Driver Cores)

** Prices in USD from January 2022. All activities are prorated by the minute and rounded up. The minimum cluster size is 8 vCores.*

Data Flows: Cluster Size



Basic: \$0.286 per vCore-hour = \$77.792 per hour

General Purpose: 256 (+16 Driver Cores)

Standard: \$0.346 per vCore-hour = \$94.112 per hour

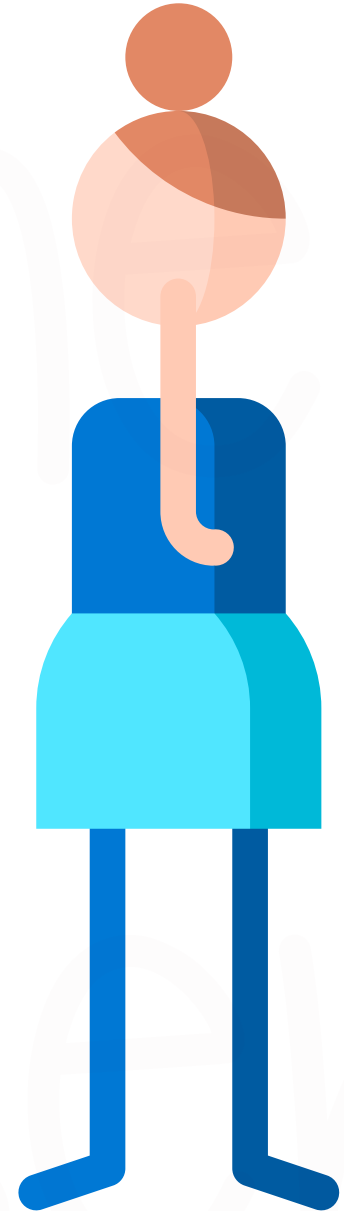
Memory Optimized: 256 (+16 Driver Cores)

** Prices in USD from January 2022. All activities are prorated by the minute and rounded up. The minimum cluster size is 8 vCores.*

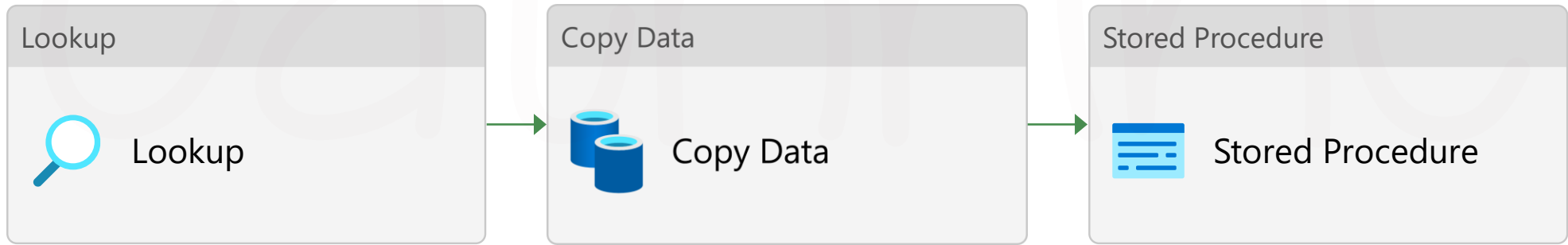


Lessons Learned

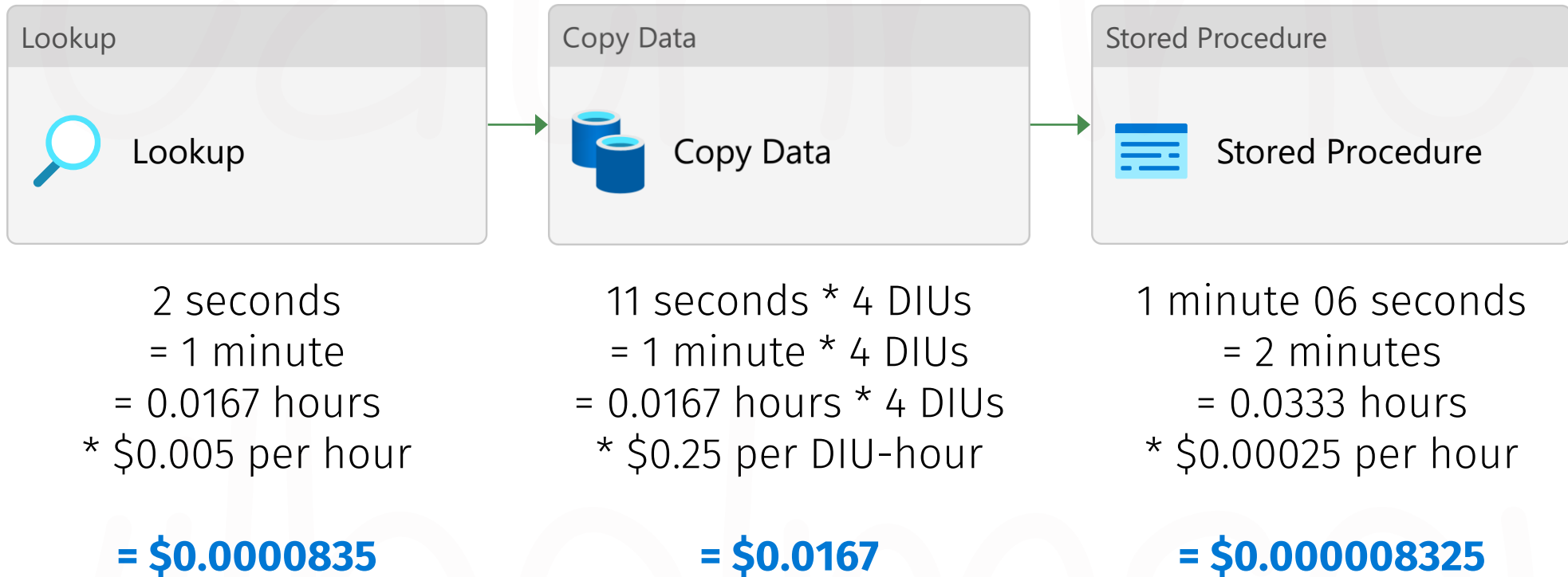
Just because you *can*...
doesn't mean you *should*...



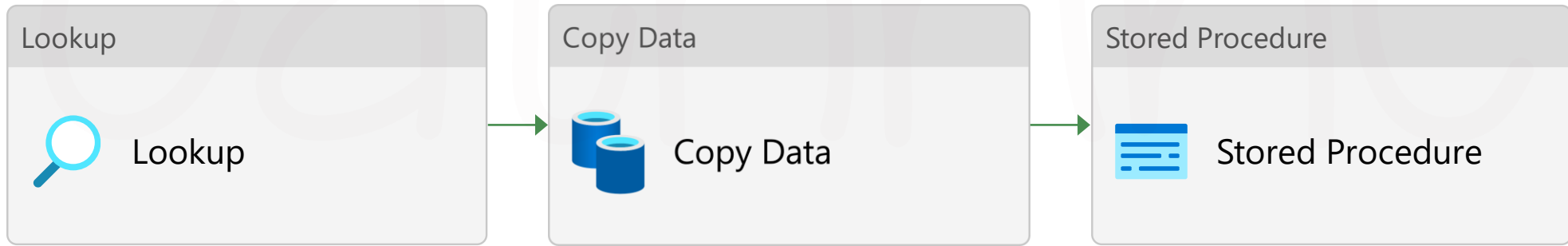
Example



Example

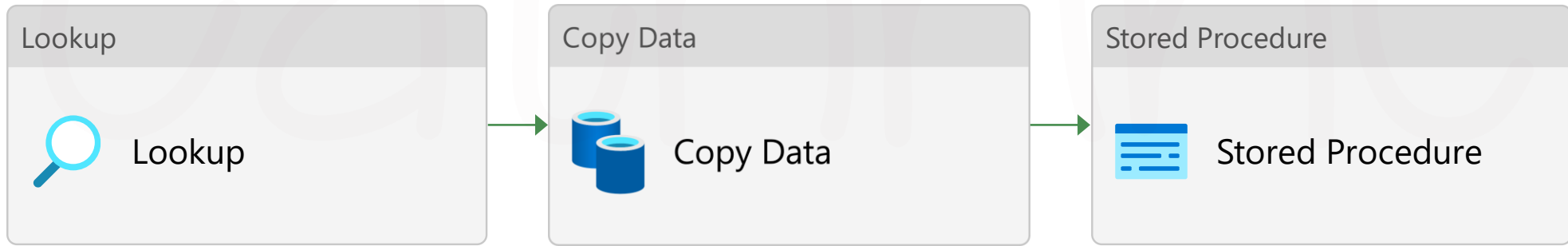


Example



Total cost per pipeline run: **\$0.053**

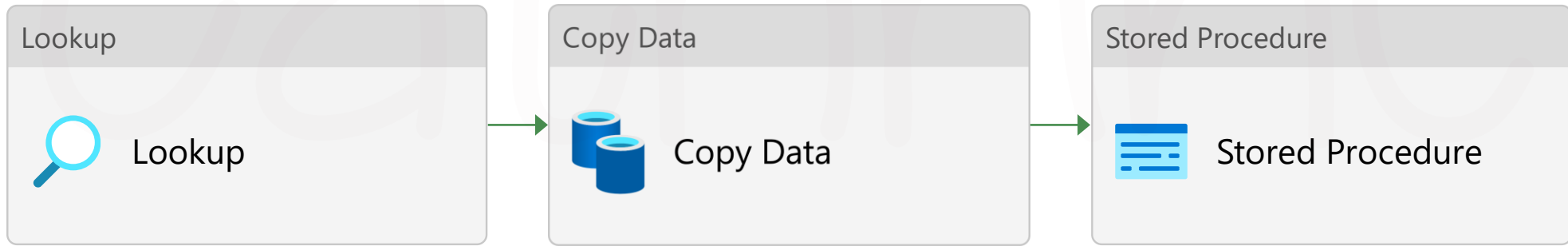
Example



Total cost per pipeline run: **\$0.053**

Run once per night for 30 days: **\$1.59**

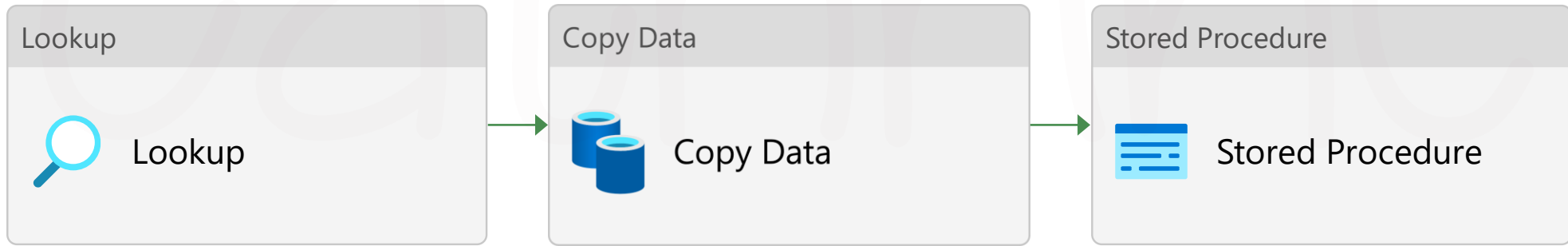
Example



Total cost per pipeline run: **\$0.053**

Run for 100 tables per night for 30 days: **\$159**

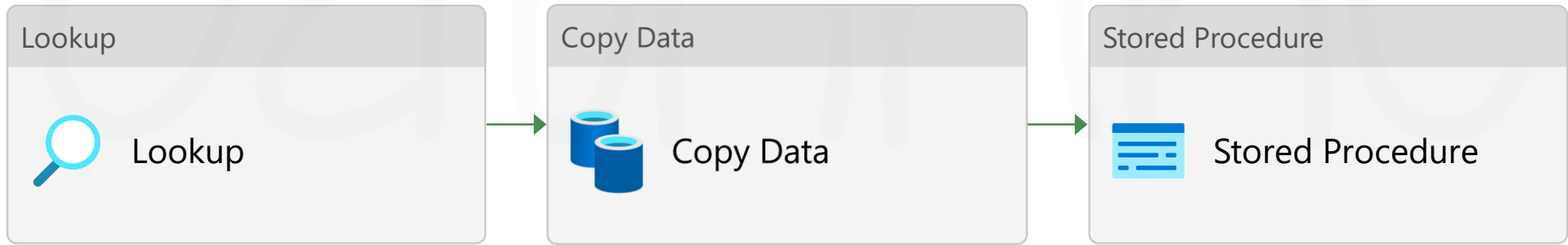
Example



Total cost per pipeline run: **\$0.053**

Run for 100 tables every *hour* for 30 days: **\$3816**

Example

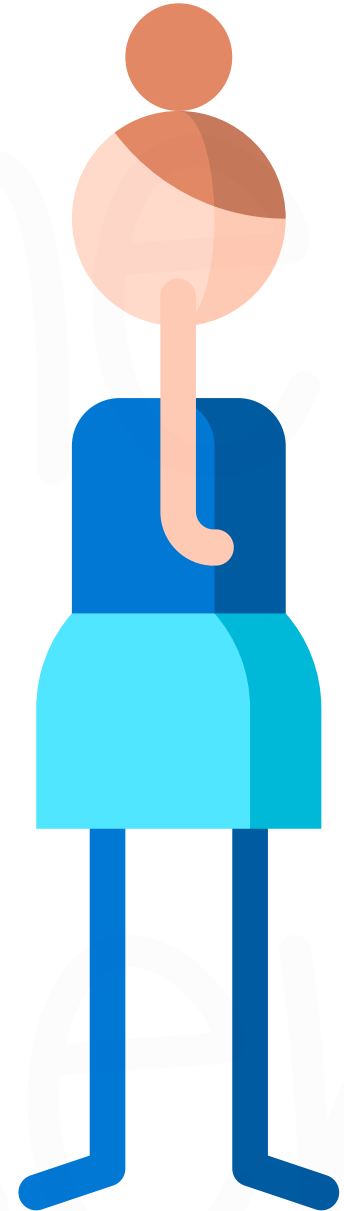


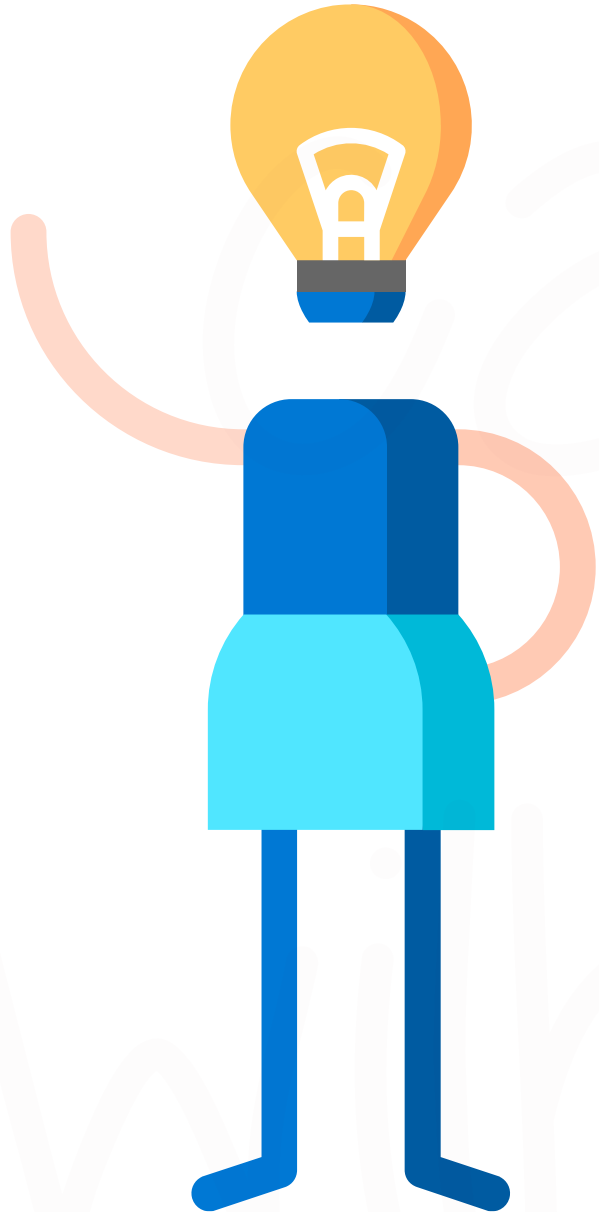
Total cost per pipeline run: **\$0.053**

Run for 100 tables every *five minutes* for 30 days: **\$45792**

Again

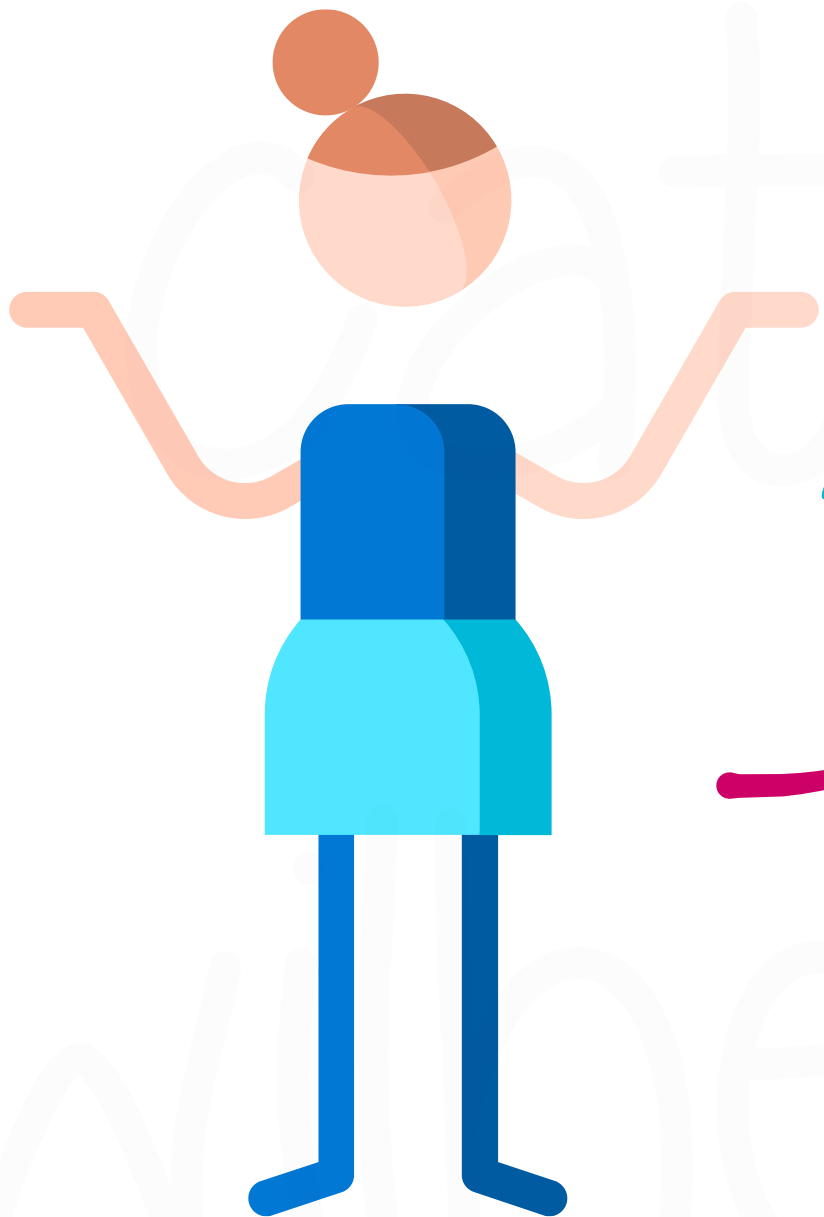
Just because you *can*...
doesn't mean you *should*...





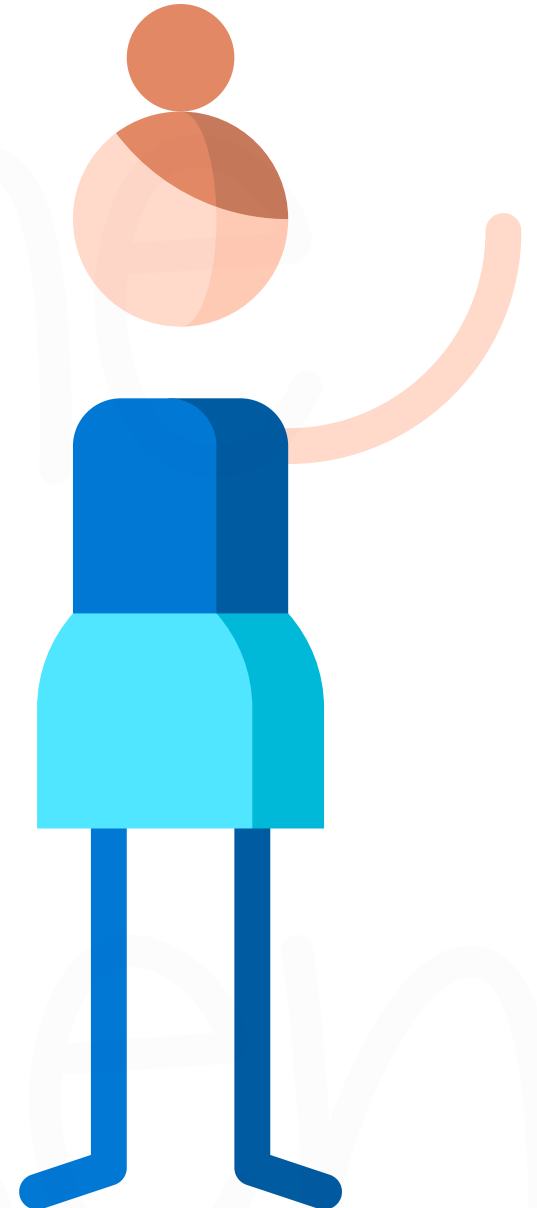
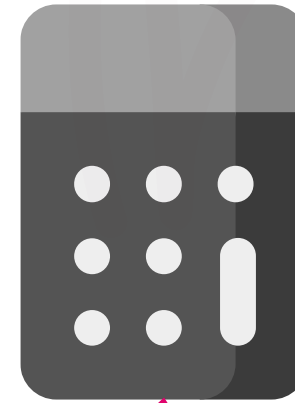
Everything has a price

We need to understand the pricing model to create solutions optimized for performance *and* cost



~~"I guess we just have to
wait for the invoice!"~~

Here's our estimate!



azure.microsoft.com/pricing/details/data-factory
azure.microsoft.com/pricing/details/synapse-analytics

Thank you!



hi@cathrinew.net



@cathrinew



cathrinew.net

