

The Frozen Data Lake

Deep-freeze your legacy data and
avoid cumbersome integrations





Mathias Halkjær Petersen

 www.linkedin.com/in/mhalkjaer

- Data Architect, Consultant, Trainer and Speaker



Business Intelligence Architect
JDM A/S



Business Intelligence Consultant
Universal Robots A/S



Chief Digital Officer (CDO)
Clayton Power



University of Southern Denmark
MSc in Engineering, Product Development and Innovation



Sergeant, Royal Danish Air Force
Forsvaret

Share presentation?

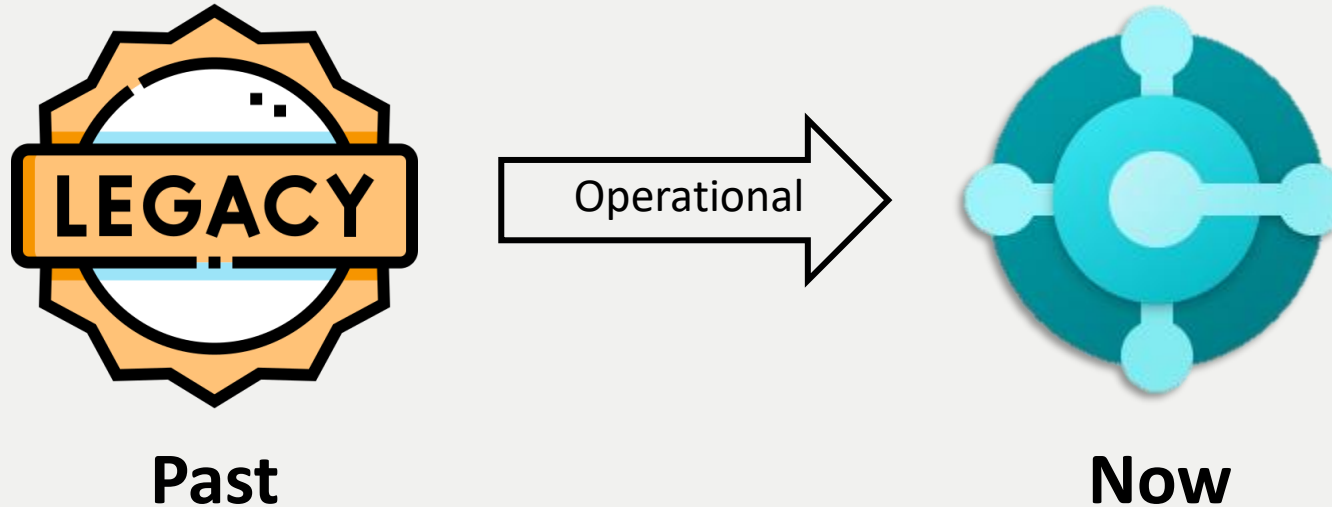
FLUX.BI



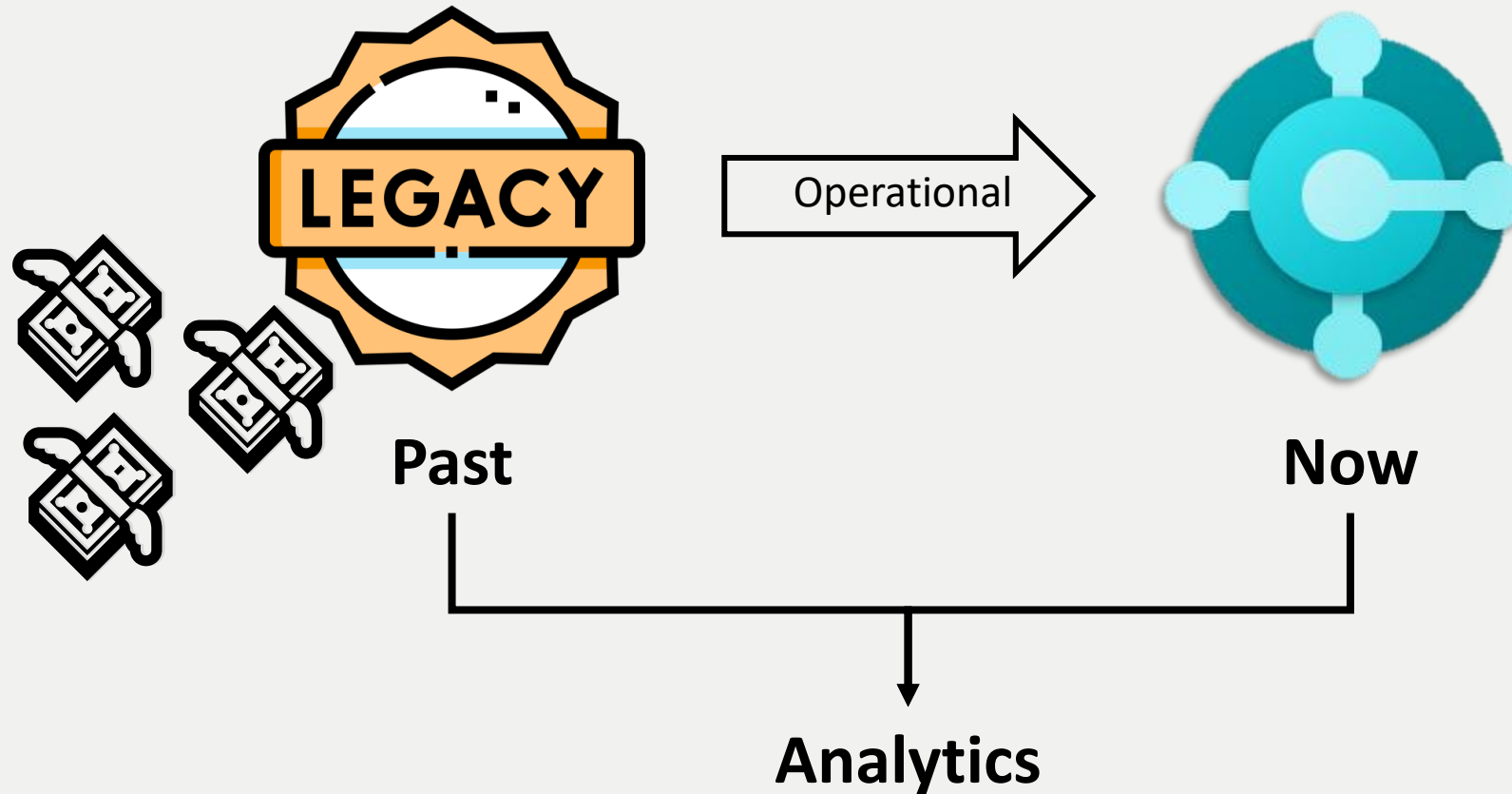
Agenda

- Why to deep-freeze data?
 - *Why not just migrate to Business Central?*
- What is Data Lakes?
- Re-using frozen data
- Unified reporting with Power BI
- Round-up
- Questions

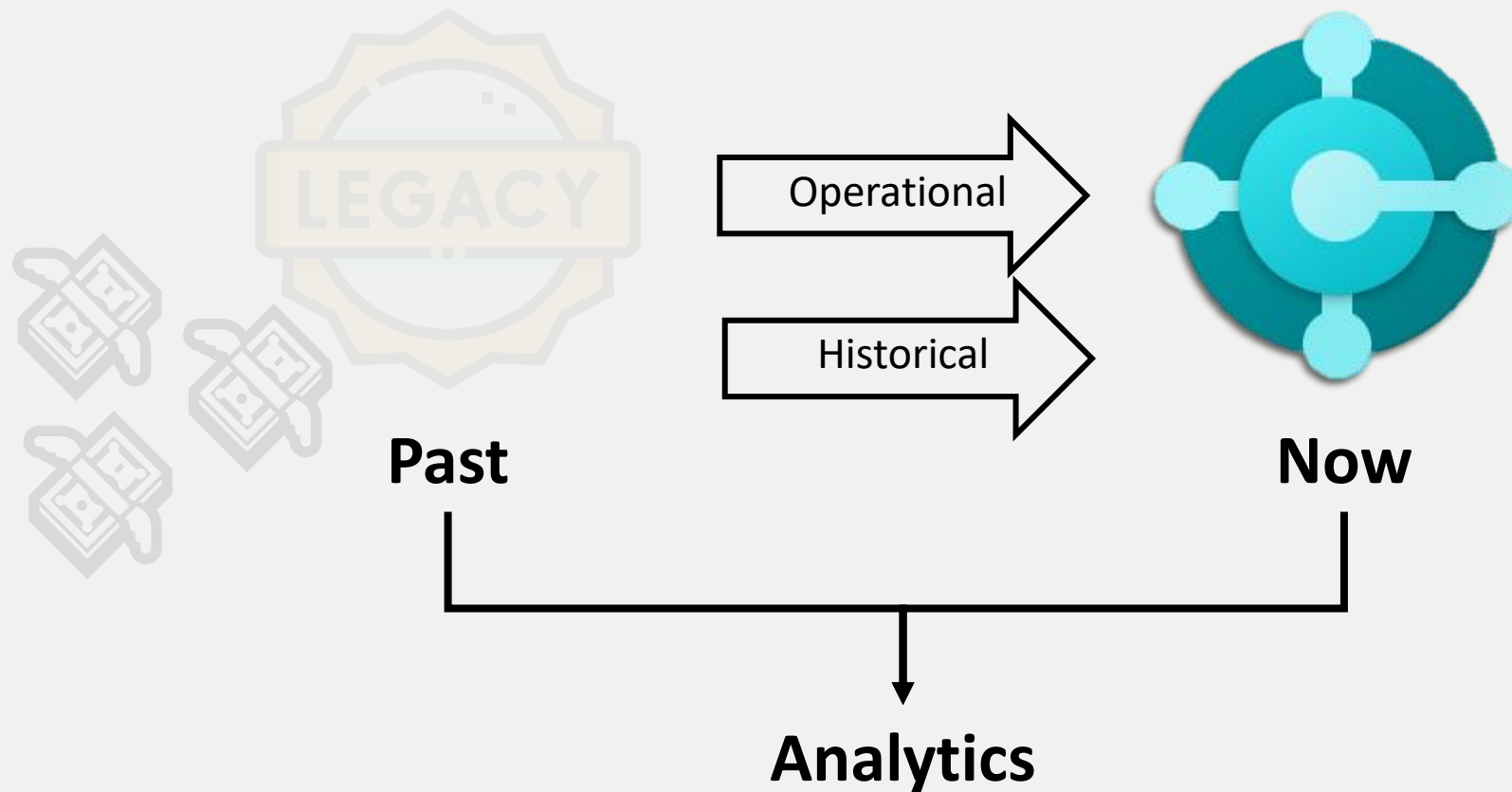
Why to deep-freeze legacy data?



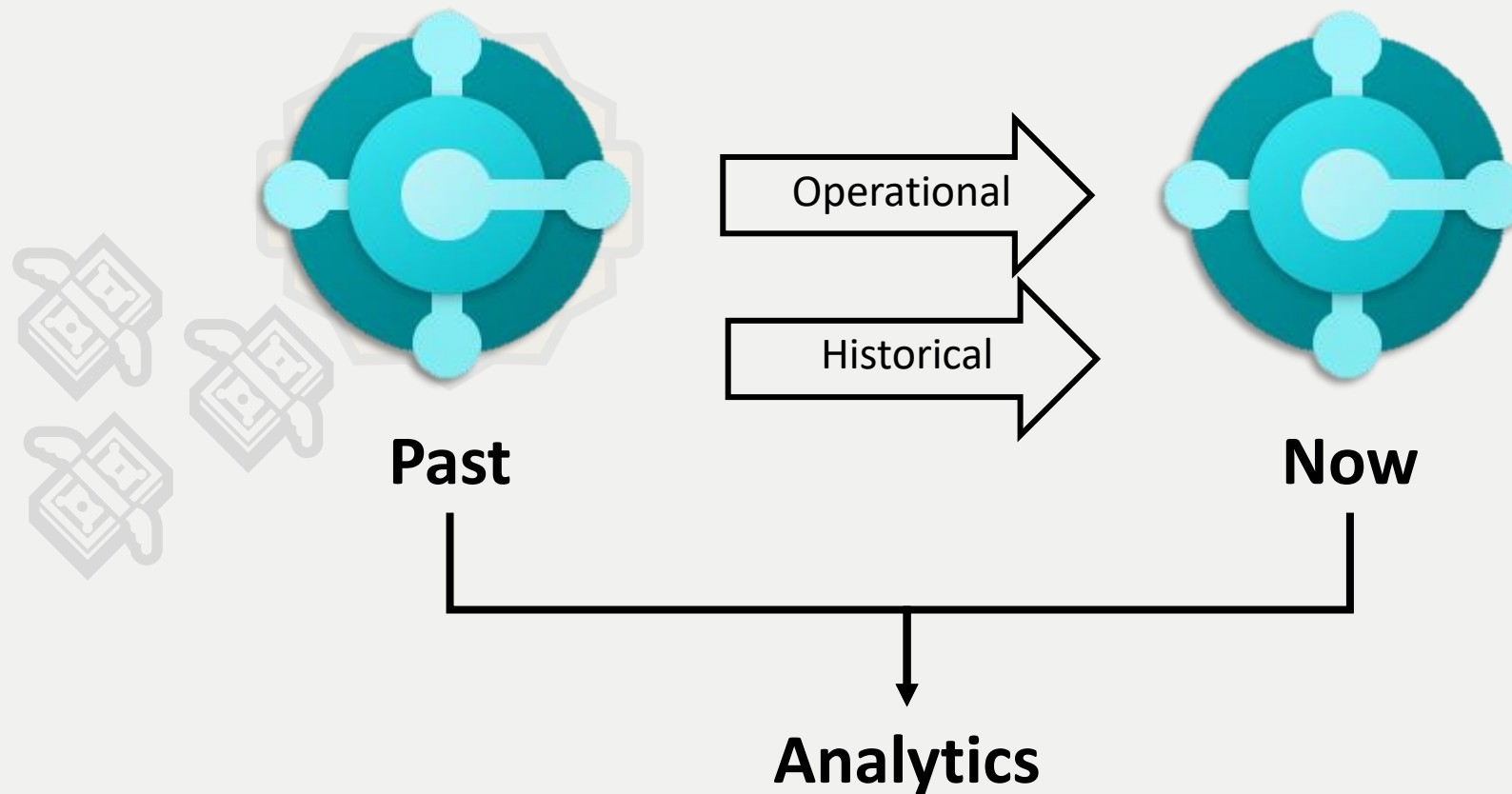
Why to deep-freeze legacy data?



Why to deep-freeze legacy data?



Why to deep-freeze legacy data?



Why to deep-freeze legacy data?

- Data has value, we shouldn't throw it away
- Customer may want period over period comparisons
- Keeping legacy system running has costs

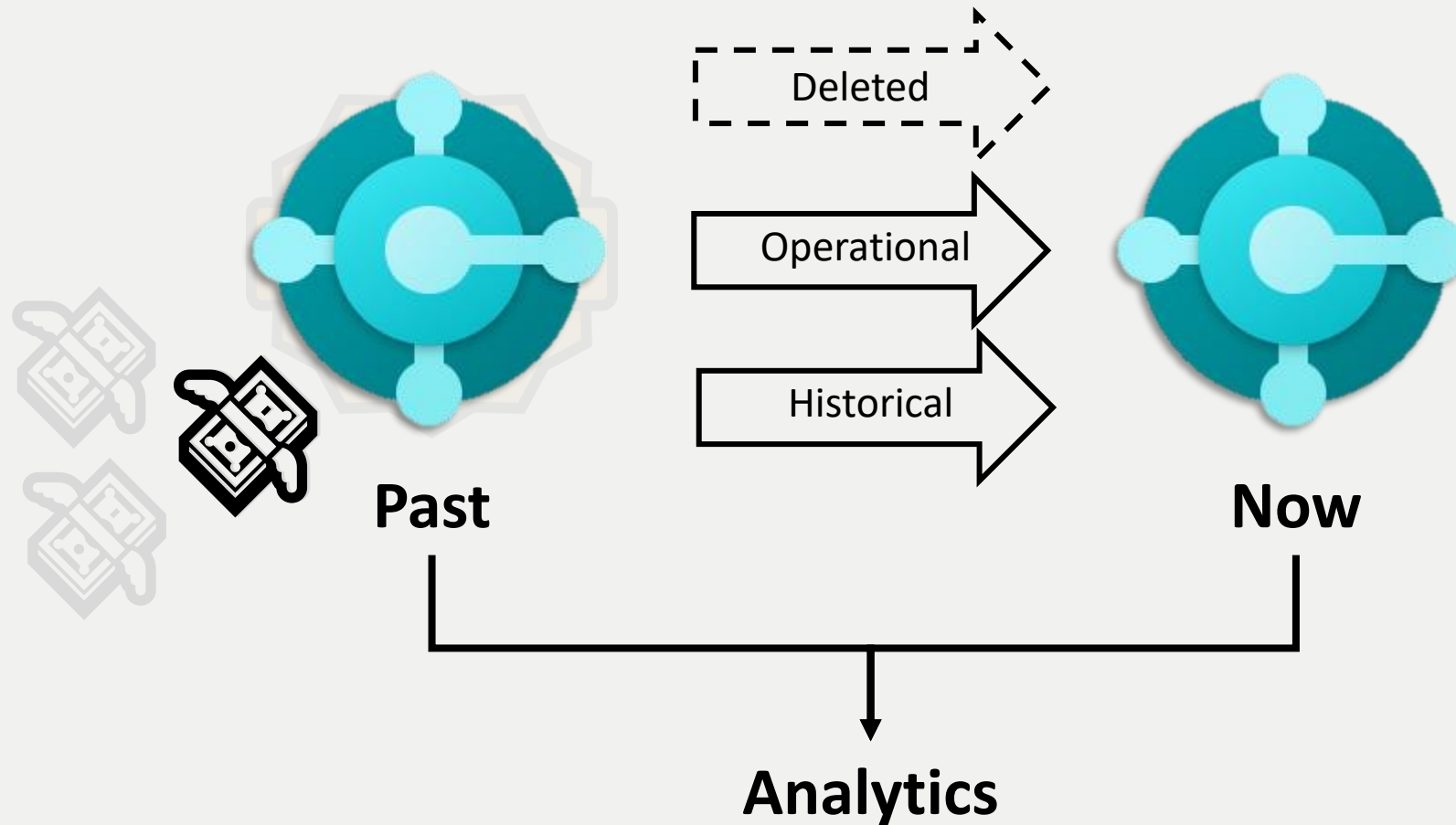
But, shouldn't we just migrate to Business Central?



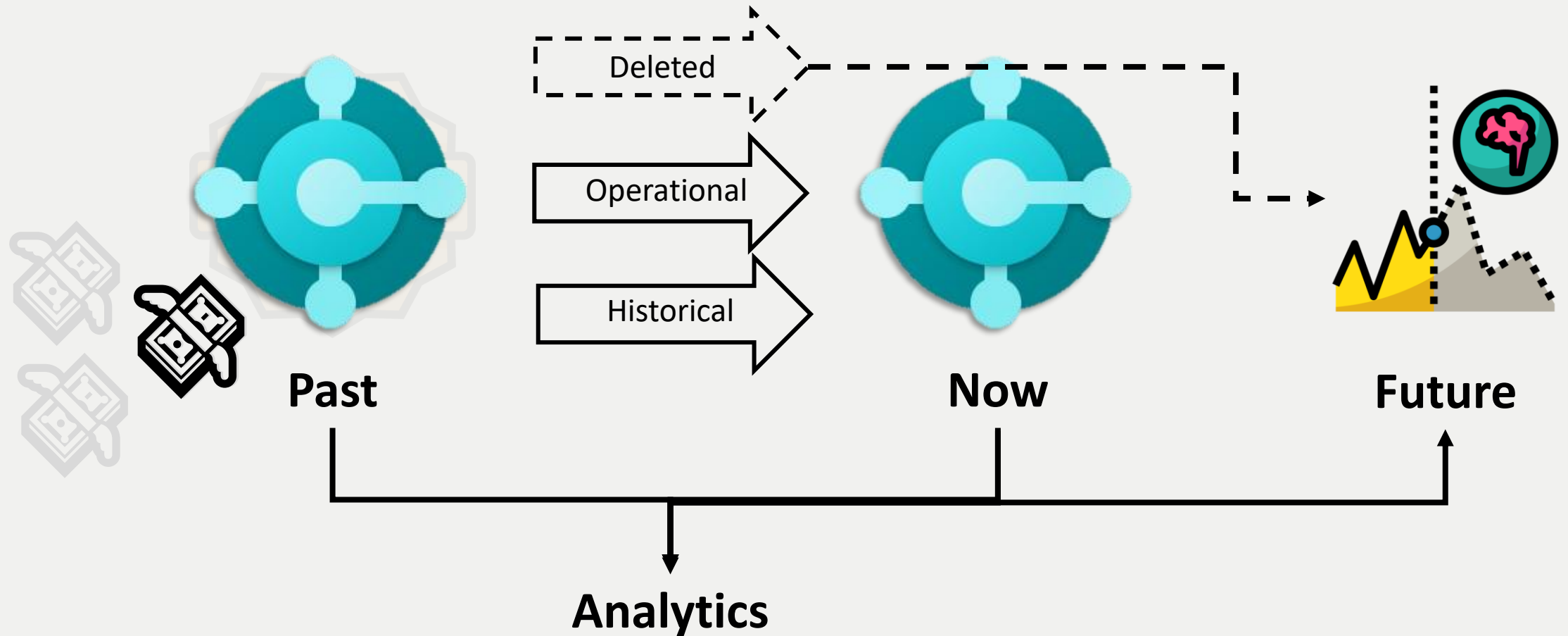
“If all you have is a hammer,
everything looks like a nail”

- A. Maslow

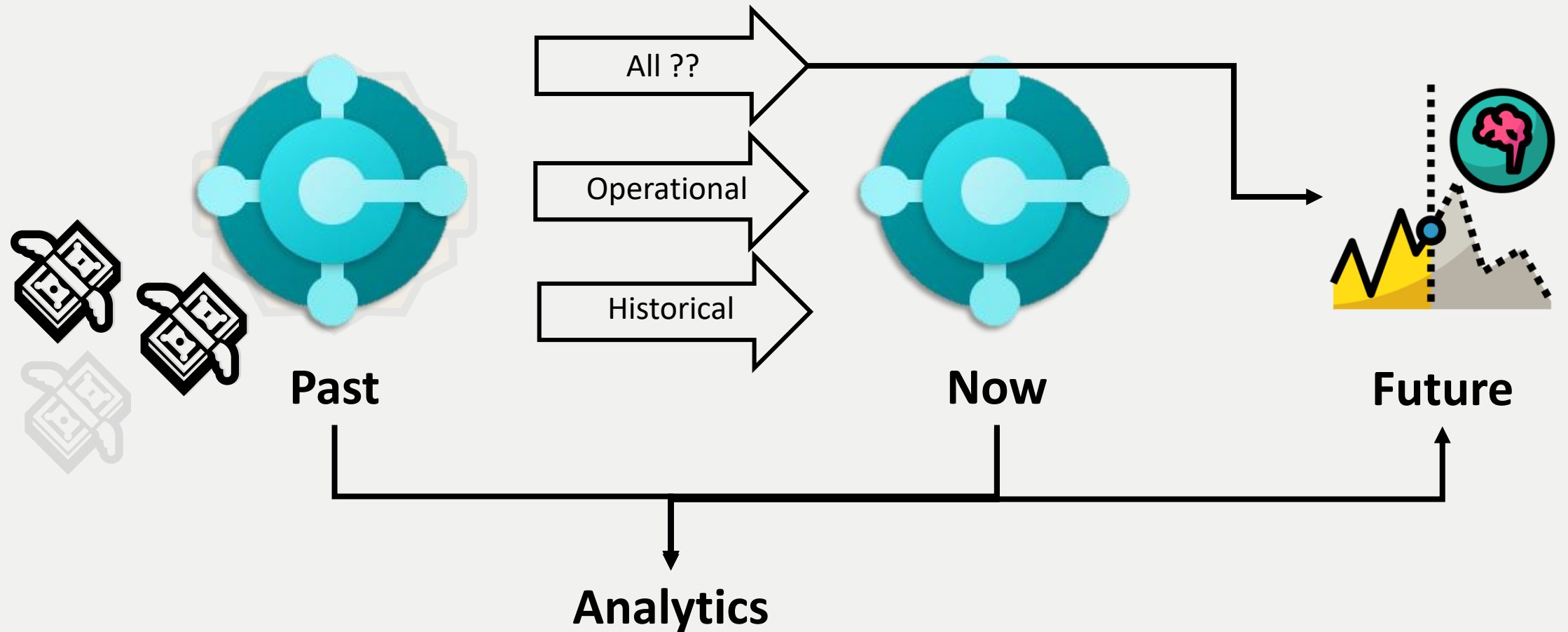
Why to deep-freeze legacy data?



Why to deep-freeze legacy data?



Why to deep-freeze legacy data?



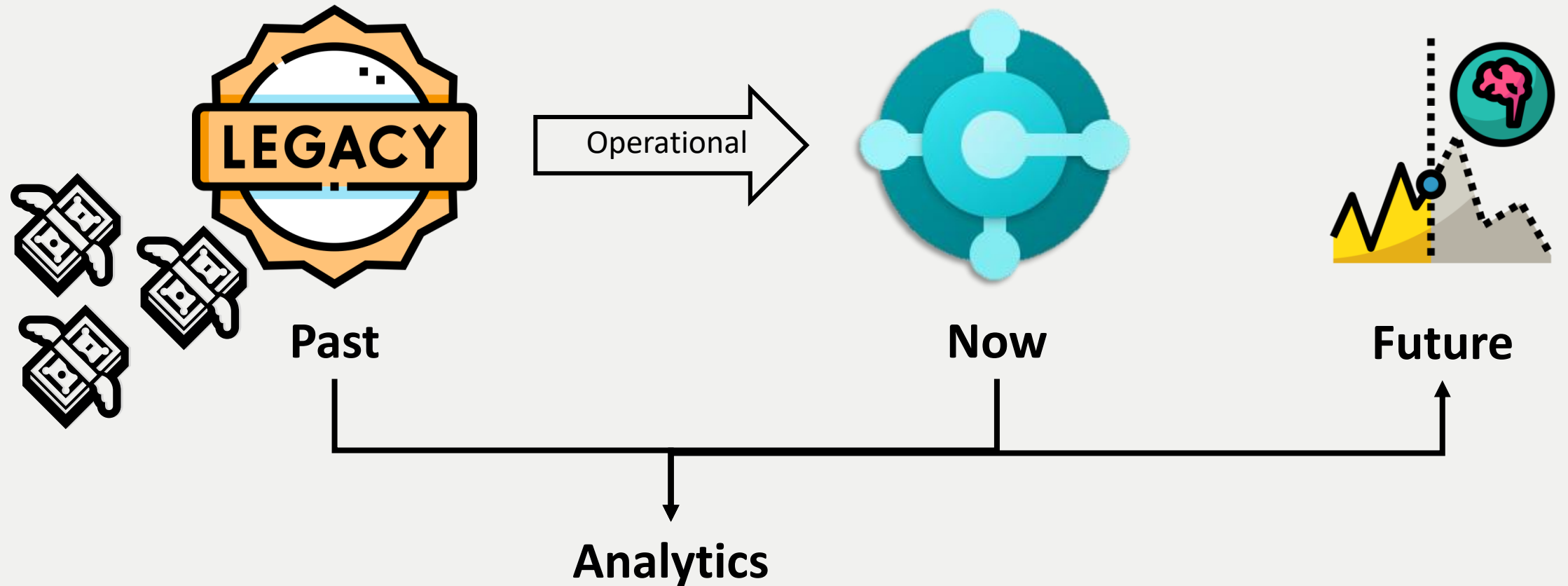
Why to deep-freeze legacy data?

- Data has value, we shouldn't throw it away
- Customer may want period over period comparisons
- Keeping legacy system running has costs

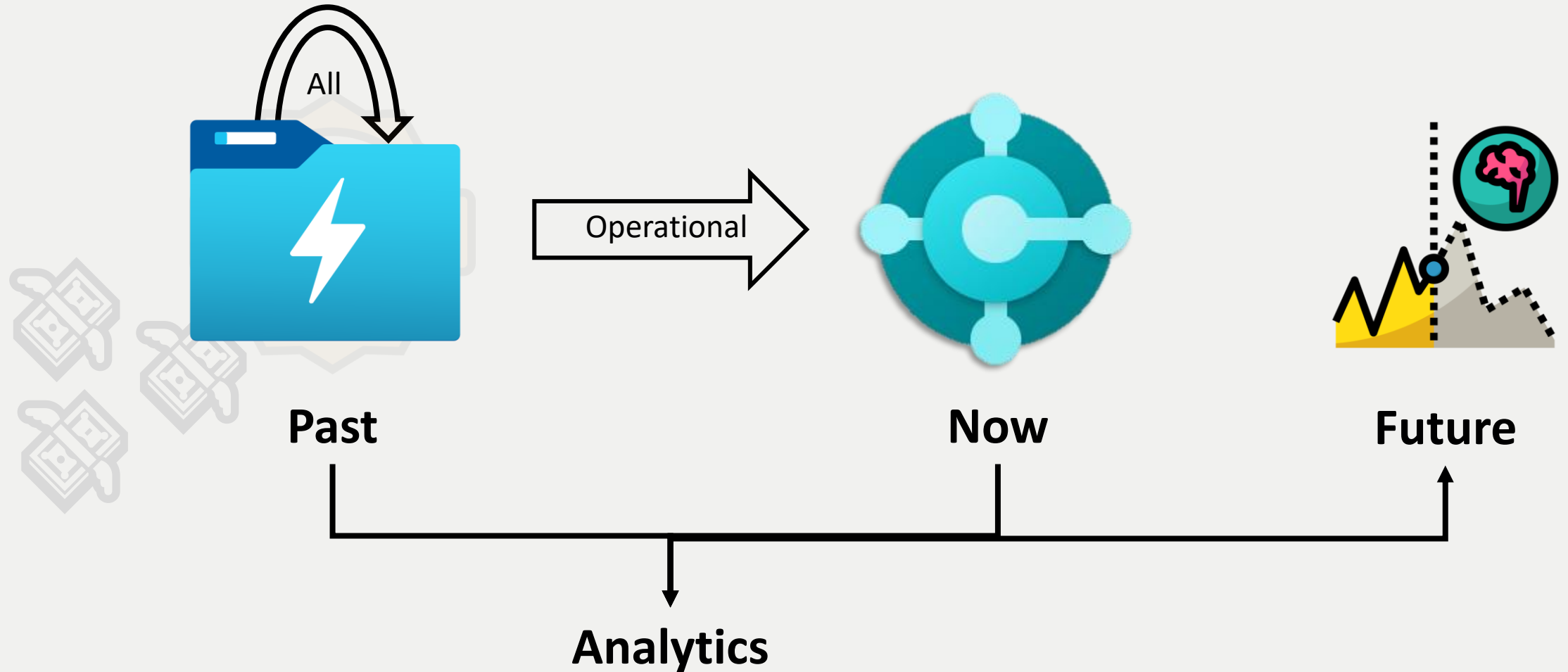
But, shouldn't we just migrate to Business Central?

- Differently structured than Business Central
- Unfeasible to keep all data – technical debt transfer
- We don't know what could be useful as ML training data
- Storage on Business Central is costly
- Business Central should mainly have operational data

Why to deep-freeze legacy data?



Why to deep-freeze legacy data?



What is a Data Lake?

- Also known as:
 - Storage account
 - Blob storage



Inexpensive file storage



Capable of big data storage

What is a Data Lake?

Pricing

Tiers:

Premium

Hot (x10 of Archive)

Cold (x5 of Archive)

Archive



What is a Data Lake?

Pricing

Tiers:

Premium

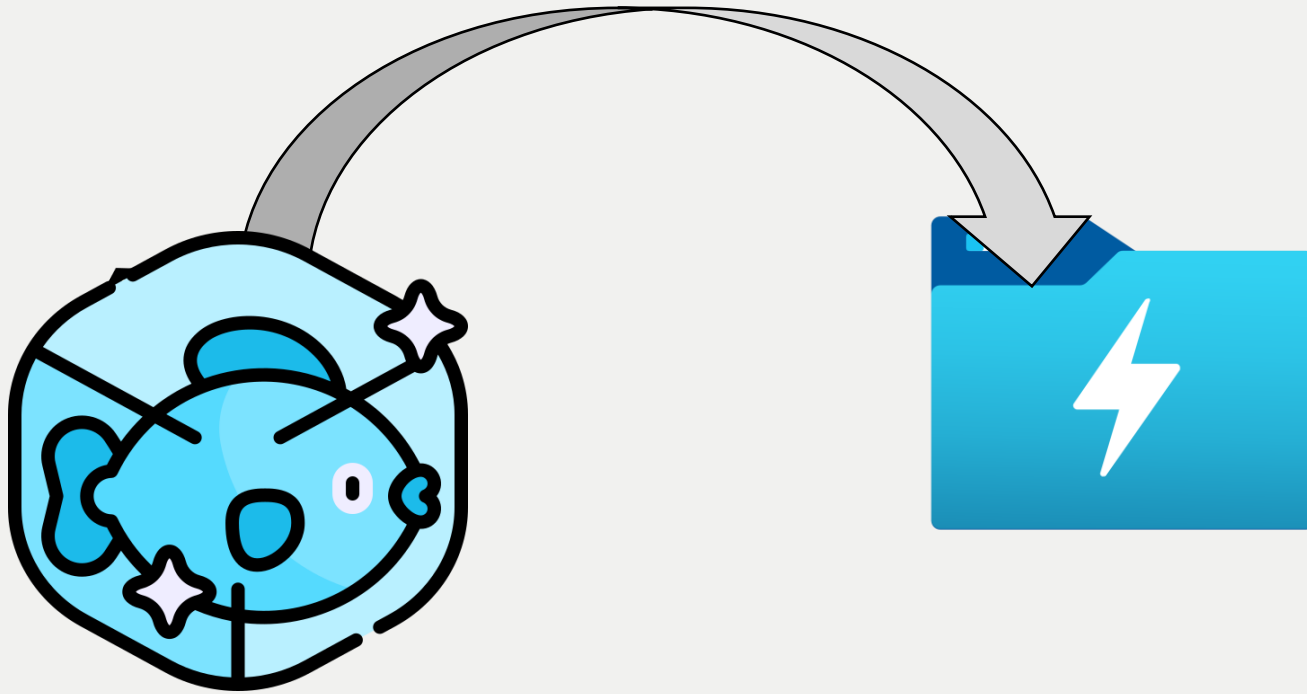
Hot (\$0.02/GB per month)

Cold (\$0.01/GB per month)

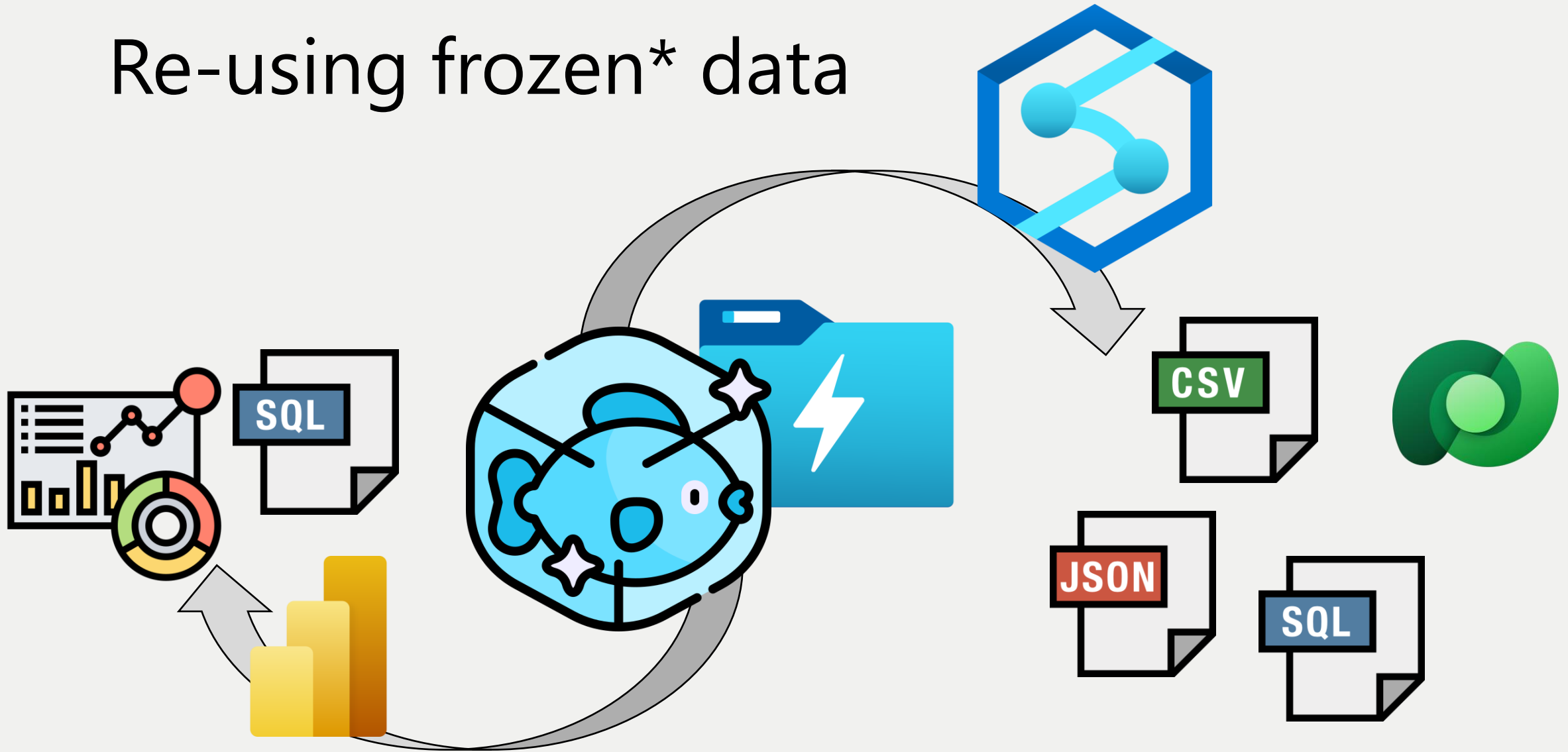
Archive (\$0.0018/GB per month)



What is a Data Lake?



Re-using frozen* data

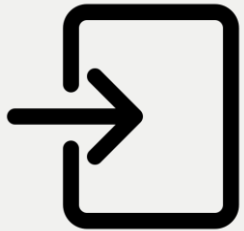


**Immutable*

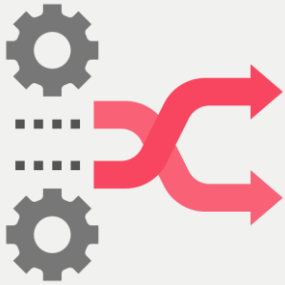
Unified reporting with Power BI



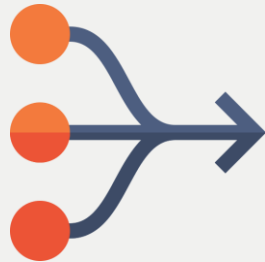
Unified reporting with Power BI



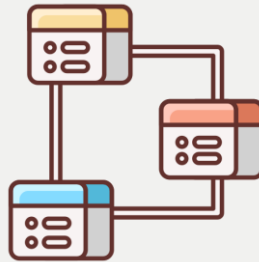
Load



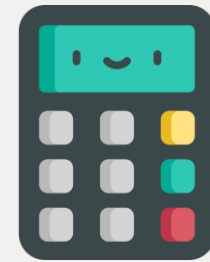
Transform



Merge



Model



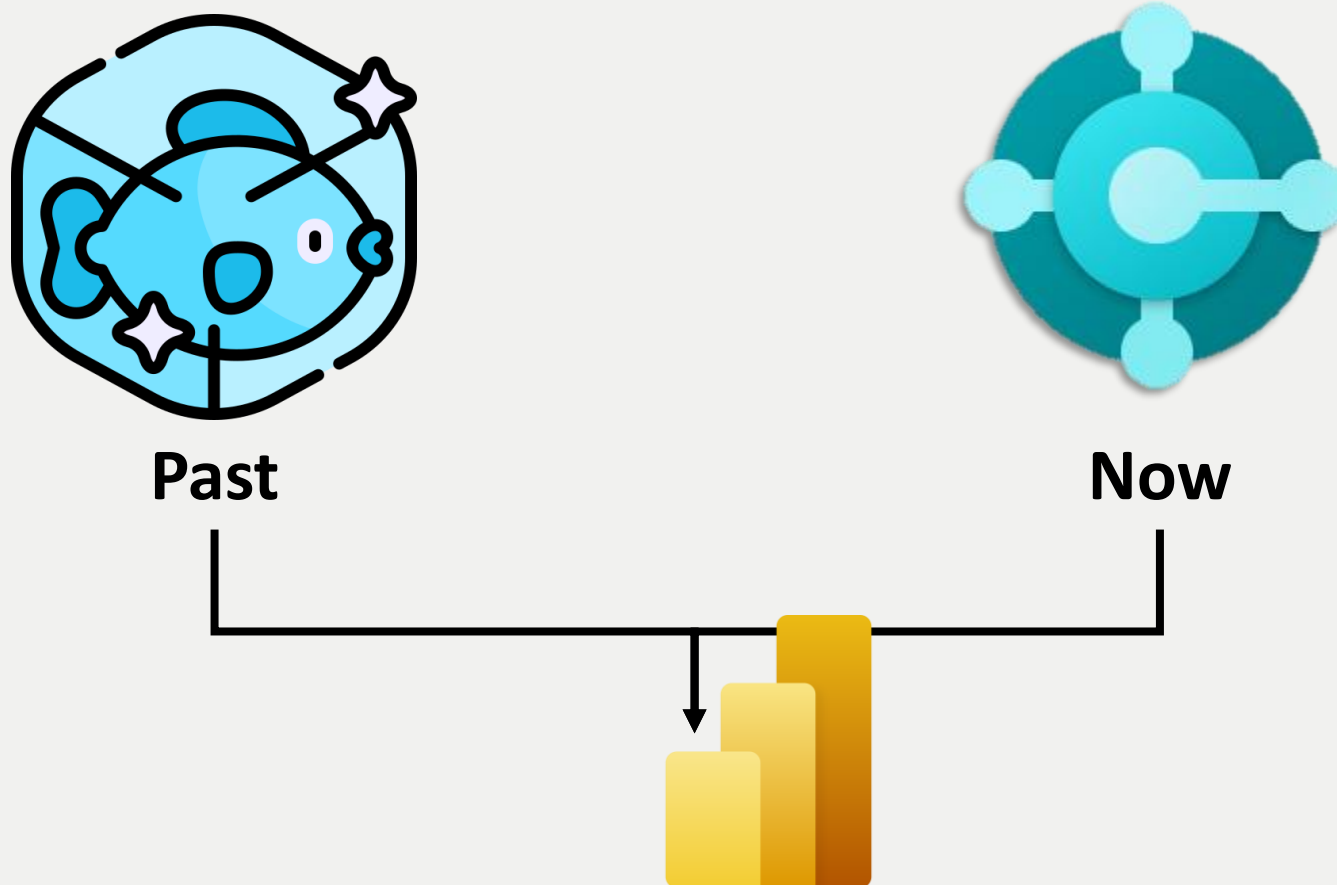
Calculate



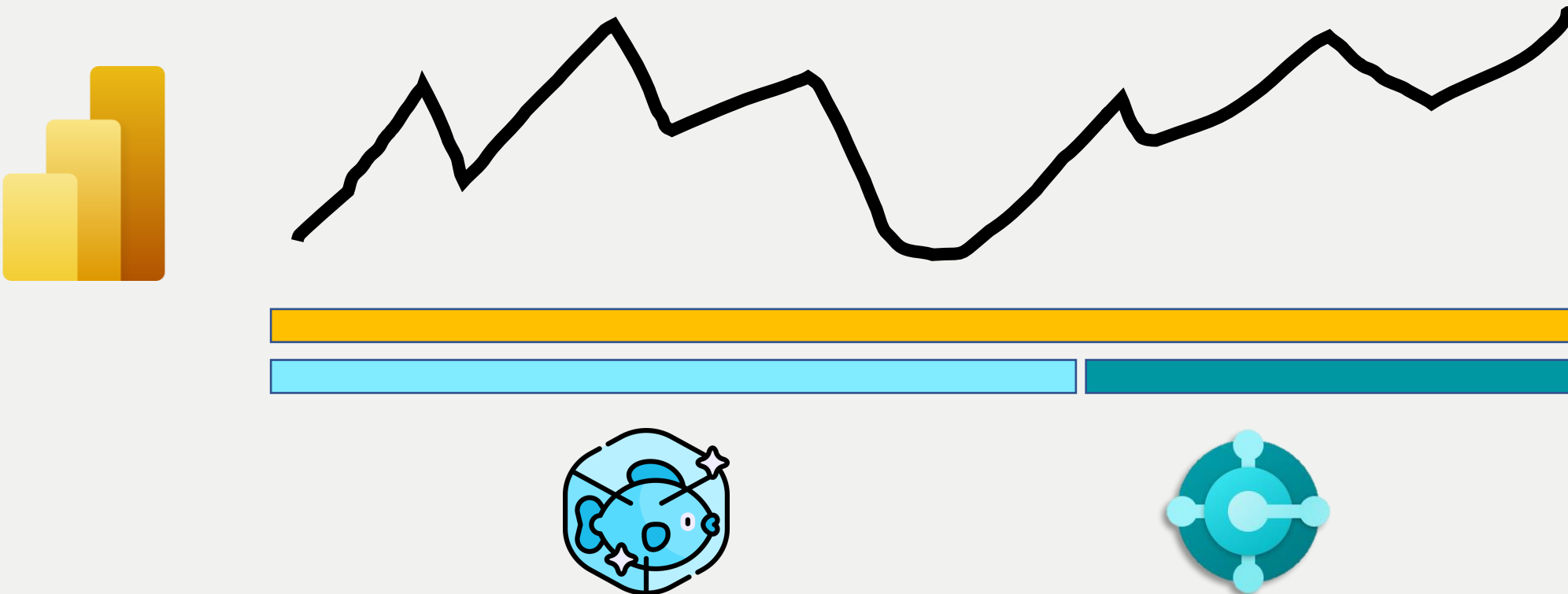
Dashboard
(Report)

FLUXBI

Unified reporting with Power BI



Unified reporting with Power BI



Unified reporting with Power BI



Unified reporting with Power BI



Sales ID	Cust. ID	Amount
21010	00001	\$15000
21011	00003	\$2500
...
21015	00002	\$7775

Cust. ID	Name
00001	Thingies
00002	Corp. Inc
00003	Yoohoo

Sales ID	Cust. ID	Amount
SO-22001	A24	\$11250
SO-22002	A16	\$9550
SO-22003	A24	\$450
SO-22004	A23	\$1000

Cust. ID	Name
...	
A16	Thingies ApS
...	
A23	Corporate Inc
A24	Yoohoo

Unified reporting with Power BI



Sales ID	Cust. ID	Amount
21010	00001	\$15000
21011	00003	\$2500
...
21015	00002	\$7775

Sales ID	Cust. ID	Amount
SO-22001	A24	\$11250
SO-22002	A16	\$9550
SO-22003	A24	\$450
SO-22004	A23	\$1000

Cust. ID	Name
00001	Thingies
00002	Corp. Inc
00003	Yoohoo

Cust. ID	Name	LegacyID
...		
A16	Thingies ApS	00001
...		
A23	Corporate Inc	00002
A24	Yoohoo	00003

Unified reporting with Power BI



Sales ID	Cust. ID	Amount
21010	00001	\$15000
21011	00003	\$2500
...
21015	00002	\$7775

Sales ID	Cust. ID	Amount
SO-22001	A24	\$11250
SO-22002	A16	\$9550
SO-22003	A24	\$450
SO-22004	A23	\$1000

Cust. ID	Name
00001	Thingies
00002	Corp. Inc
00003	Yoohoo

95%

87%

100%

Fuzzy
merging

Cust. ID	Name
...	
A16	Thingies ApS
...	
A23	Corporate Inc
A24	Yoohoo

Unified reporting with Power BI



Sales ID	Cust. ID	Amount
21010	00001	\$15000
21011	00003	\$2500
...
21015	00002	\$7775

Cust. ID	Name
00001	Thingies
00002	Corp. Inc
00003	<u>Yoohoo</u>

Sales ID	Cust. ID	Amount
SO-22001	A24	\$11250
SO-22002	A16	\$9550
SO-22003	A24	\$450
SO-22004	A23	\$1000

Cust. ID	Name
...	
A16	Thingies <u>ApS</u>
...	
A23	Corporate Inc
A24	<u>Yoohoo</u>

LegacyID	CustID
0001	A16
0002	A23
0003	A24
0004	A32
0005	A32

Other use-cases?

- Reduce costly storage
- Reduce load on source system
- Avoid API limits

Additional resources

Business Central – Azure Data Lake Sync

- <https://github.com/microsoft/bc2adls>



Overall Equipment Effectiveness



69.8% !

PY: 69.4% (+0.6109%)

Total Produced



4,244,699 !

PY: 4,118,712 (+3.0589%)

Finished Goods



3,930,063 !

PY: 3,801,500 (+3.3819%)

Scraps



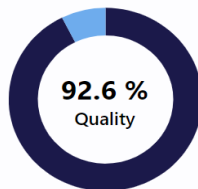
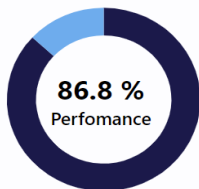
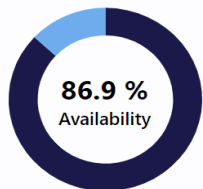
317,788 ✓

PY: 318,846 (-0.3318%)

Year

2021

Overall Equipment Effectiveness



Operating Time

Ideal Operating Time



302

LY: 283 (+6.7138%)

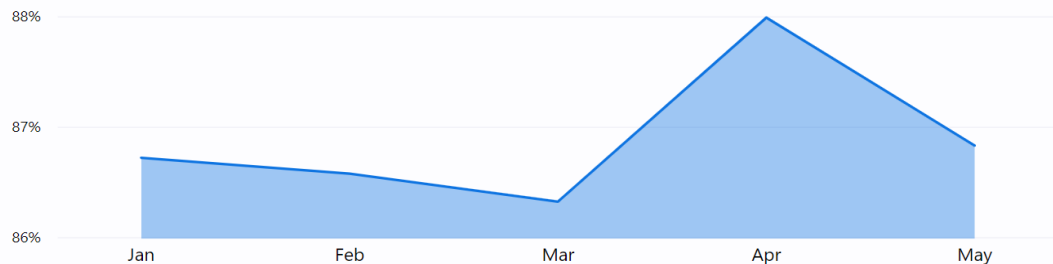
Net Operating Time



246

LY: 36 (+583.3333%)

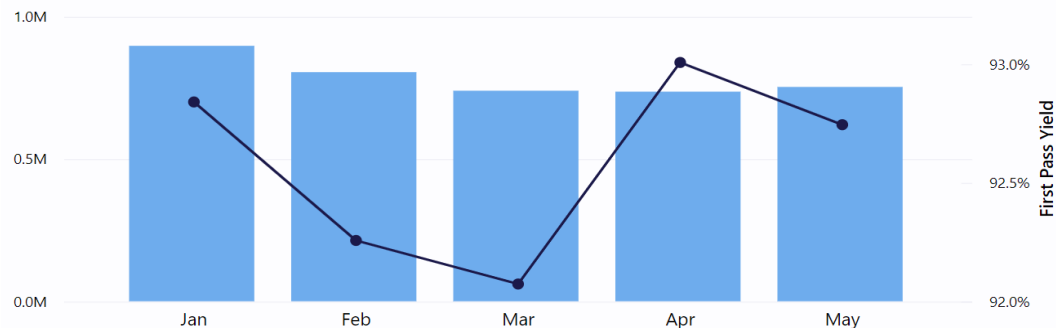
Availability % by Month



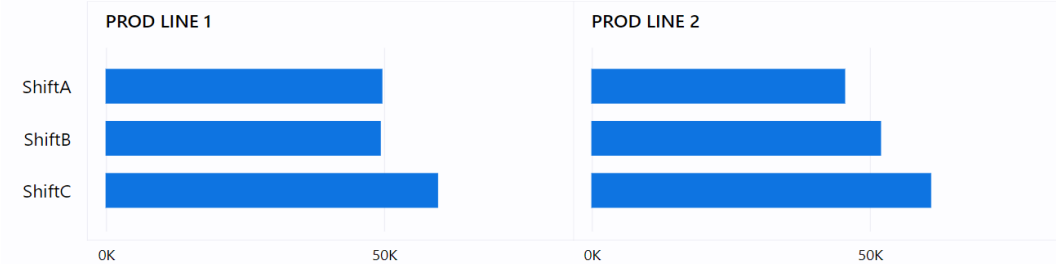
Production

Finished Goods VS First Pass Yield

Finished Goods First Pass Yield



Reject Count by Product



Questions



Thank you