North Insights

SIMPLE AND CHEAP BI SOLUTION FOR SMALLER ORGANIZATIONS

Ásgeir Gunnarsson
Just Blindbæk

Thank you, partners

KOHERA

element 61
experience & expertise

lytix

bmatix
Act informed

inetum.
realdolmen
Positive digital flow

datasense

MICROPOLE
BELUX

LACO

MYRMIDONS

AKABI

Cloubis

datashift

EpicData.

Sparkle

Tabular Editor

solarwinds

u2u

ae
Adapt
and
Enable

MONIN
Database Managed Services

proximus NXT
tech. bizz. people.

tillit
data
shapers

ORDINA
Ahead of change

WHO ARE WE?

Just Blindbæk
@justblindbaek

Ásgeir Gunnarsson
@bidgeir

AGENDA

- Set the scene
  - Type of company
  - Skills available
  - Budget
- Typical patterns in DW/Analytics
- Complexity in Azure solutions
- Cost drivers
- Example of a lake based solution and why it works for the target group
- Example of a SQL based solution and why it works for the target group

TYPICAL SCENARIOS

- Smaller organizations
- No or few BI professionals
- No data engineers
- Small budget
- Few data sources
- Moderate amount of data

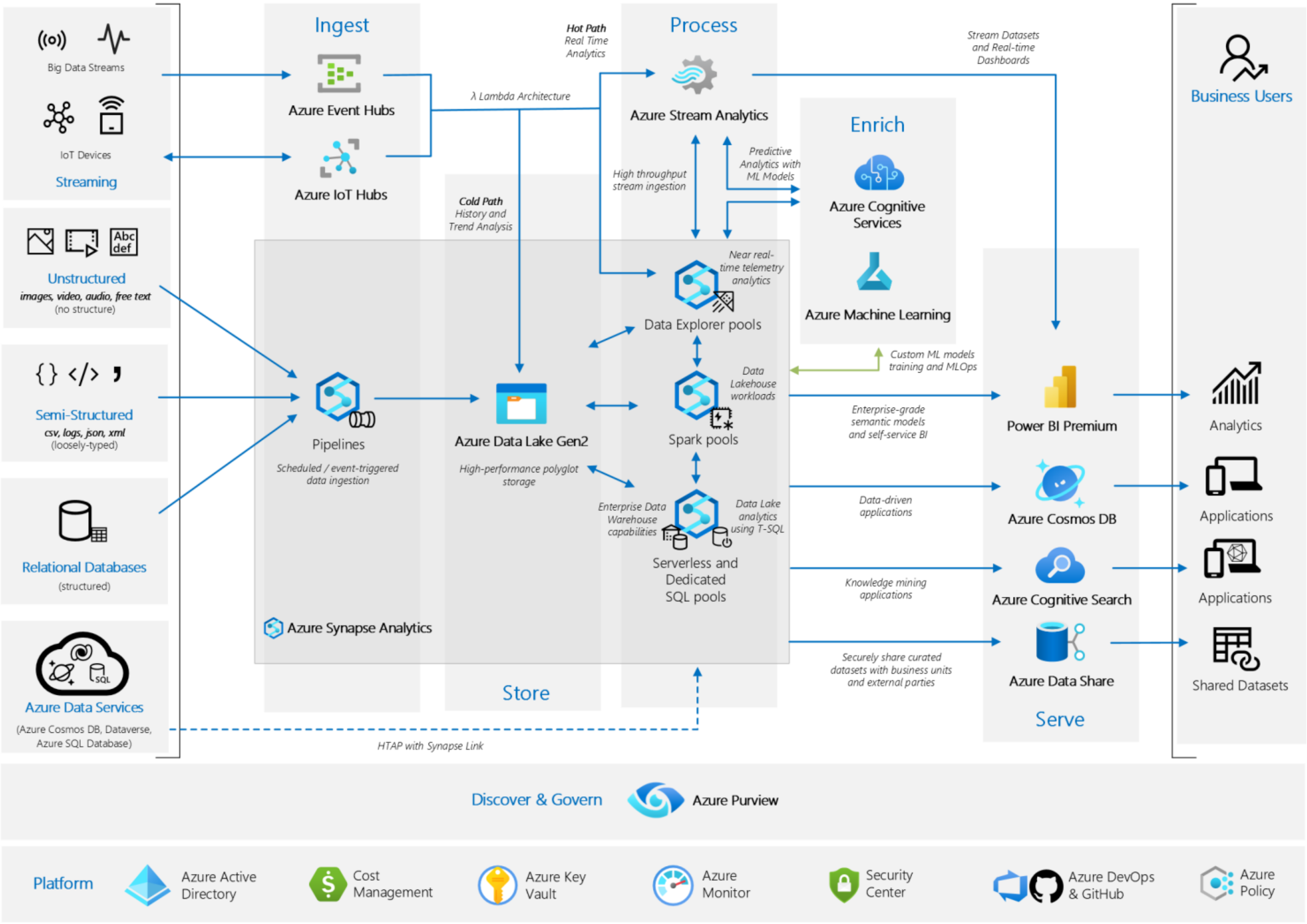
IMPACT

- Low complexity
- Maintenance friendly
- Low cost - preferably predictable

POPULAR PATTERNS

- Traditional Modern Data Warehouse
- Data Lakehouse
- Self-service with Dataflows in Power BI
- Specialized solutions
- Hybrid solutions

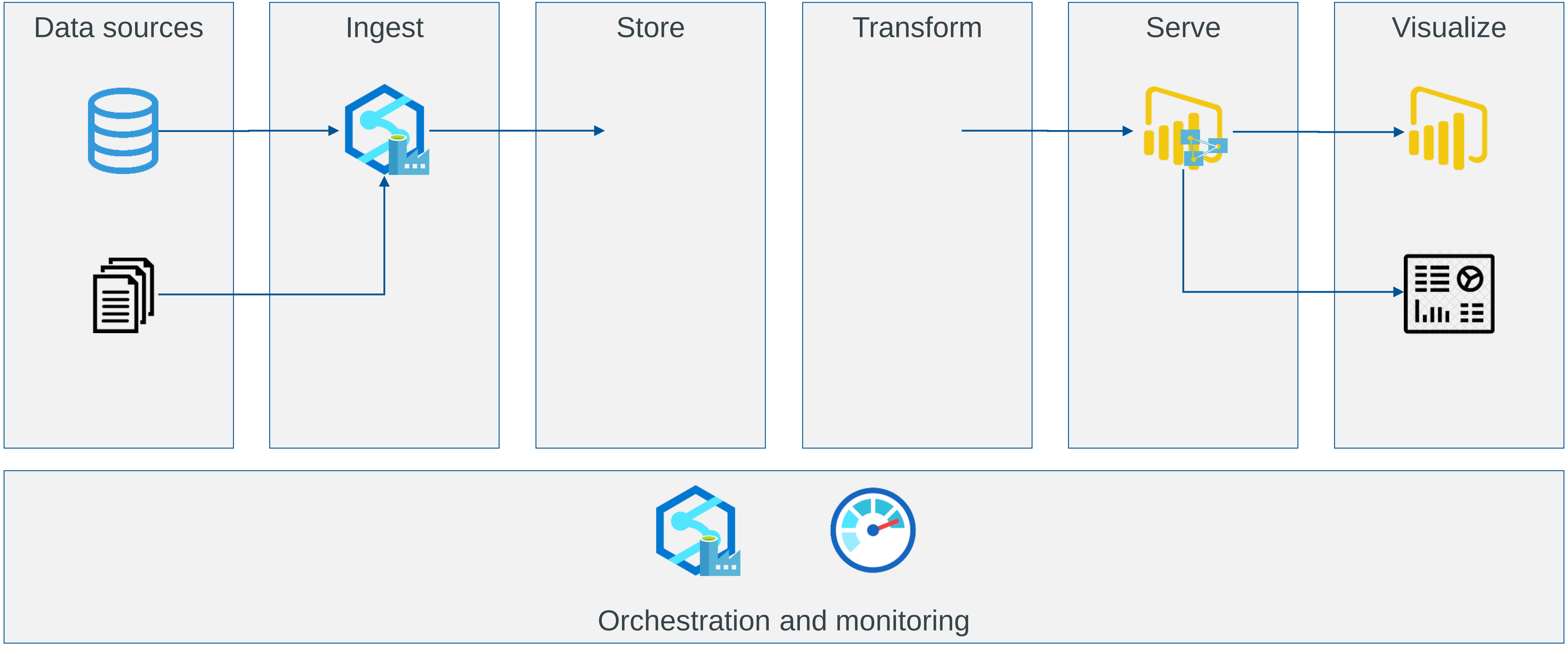
SYNAPSE ANALYTICS



ARCHITECTURES FROM MICROSOFT

- https://learn.microsoft.com/en-us/azure/architecture/example-scenario/dataplate2e/data-platform-end-to-end?tabs=portal
- No simple architectures
- No SQL based architecture

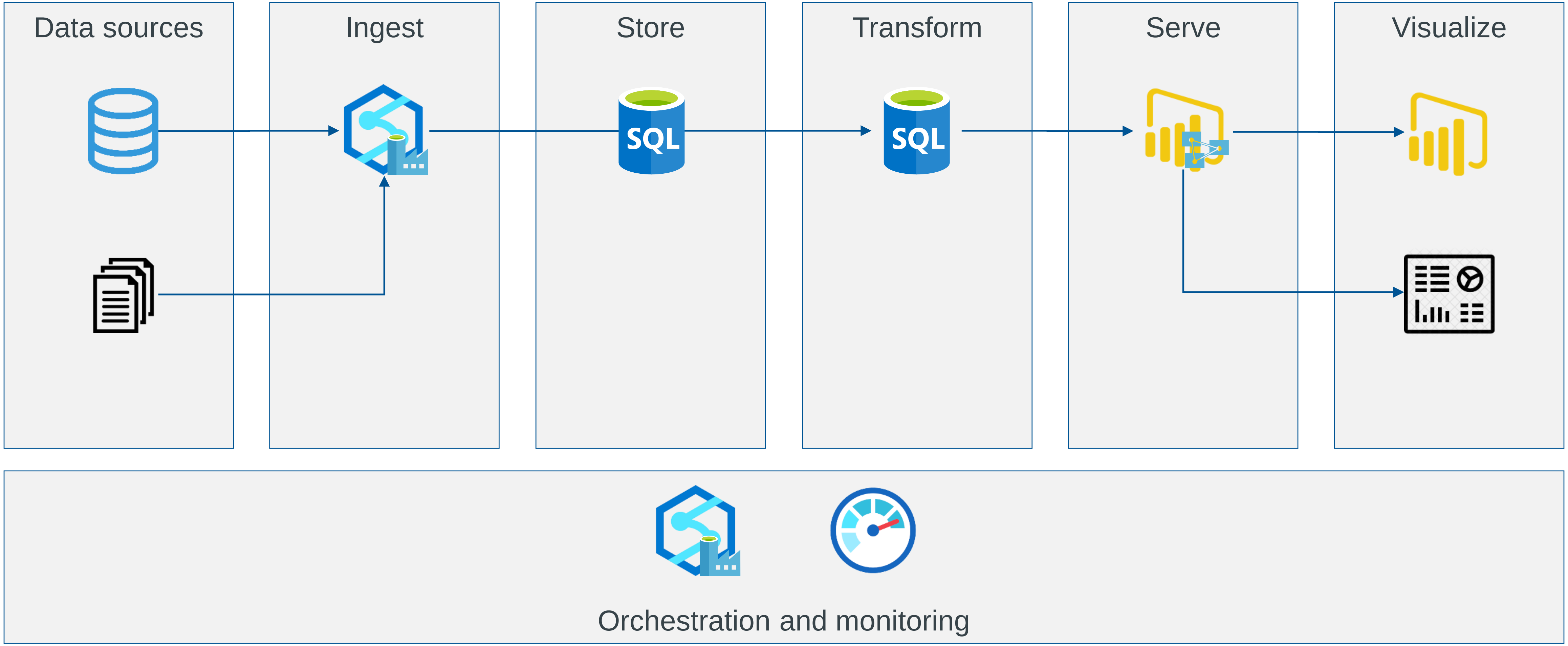
COMMON ARCHITECTURE



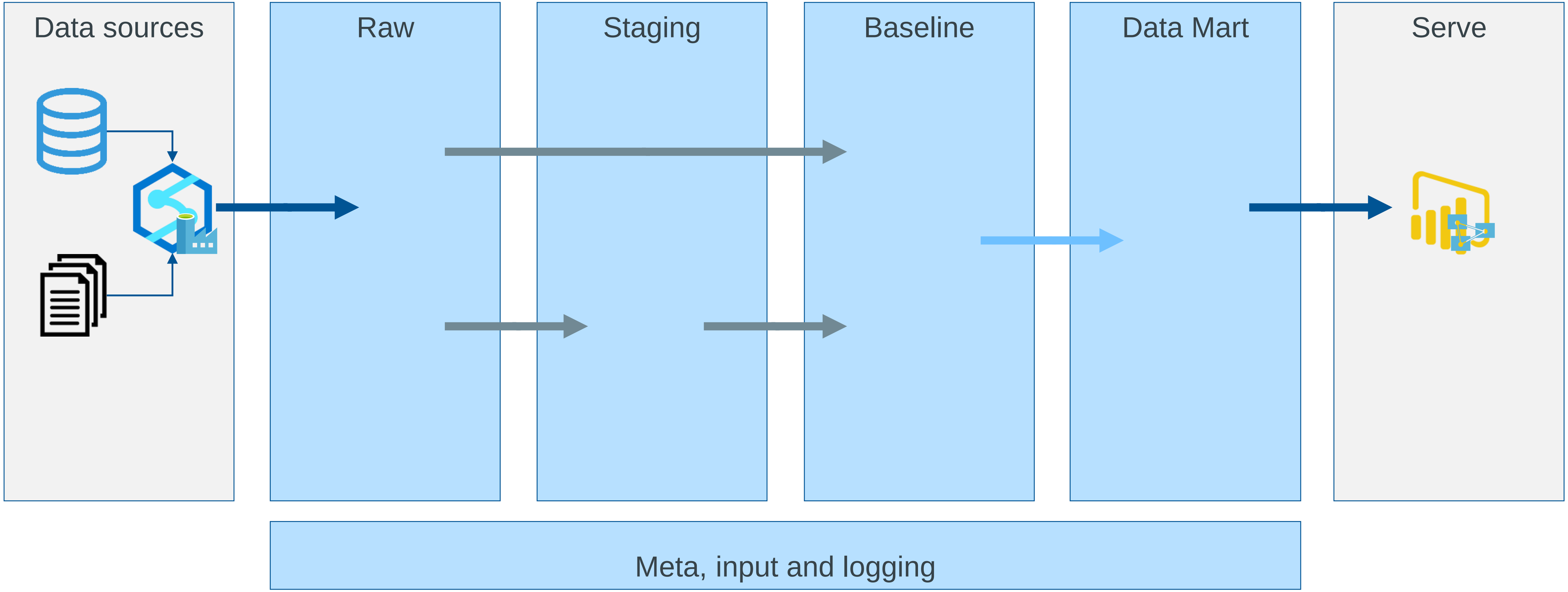
COST DRIVERS

- Compute
  - In some instances, you will pay based on compute
  - It can be fixed – Azure SQL DB
  - It can be consumption based – Synapse SQL Serverless Pools
- Storage
  - Normally must pay for storage if it´s separated from compute
  - Azure Data Lake Gen2
- Usage
  - In some instance you will pay for usage – Azure Data Factory
    - Sometimes it will be coupled with type of compute

TRADITIONAL MODERN DWH

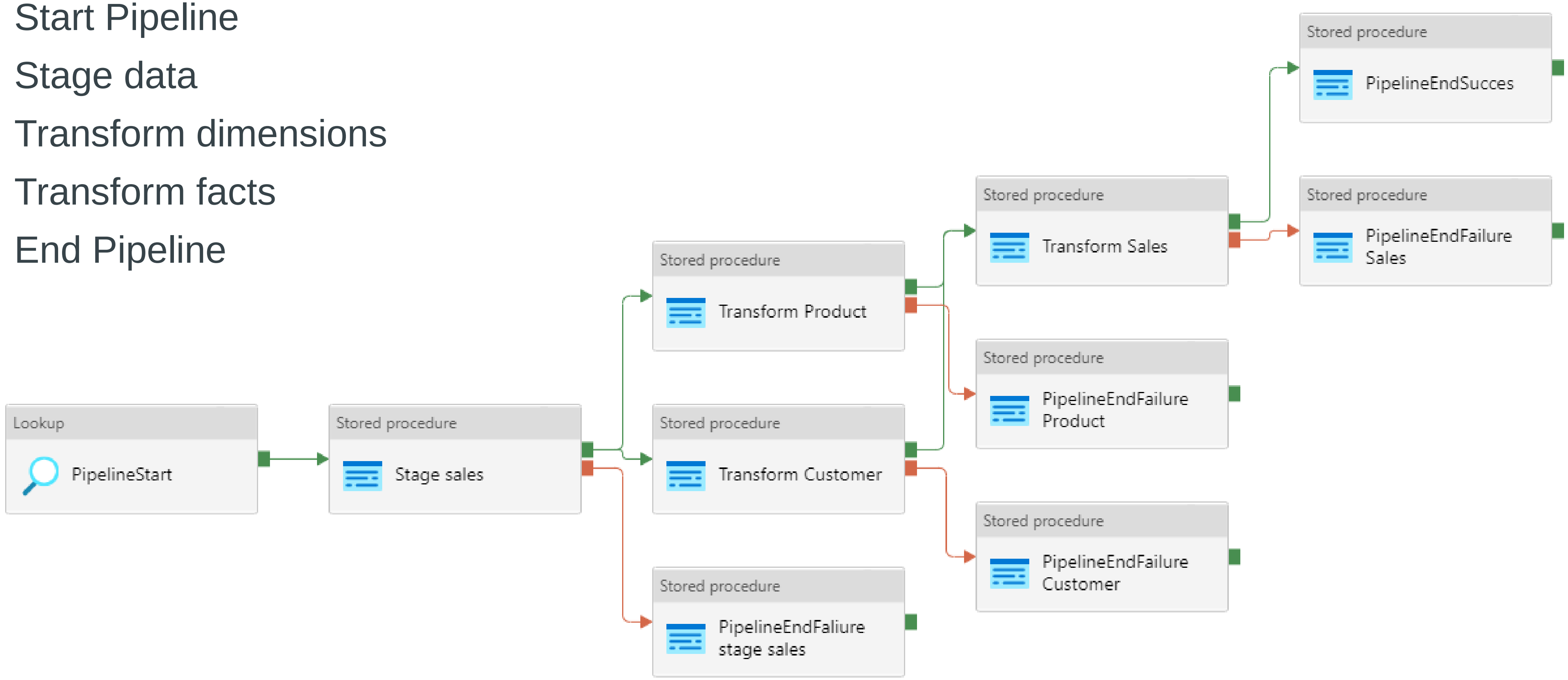


TRANSFORM ”FRAMEWORK” IN SQL

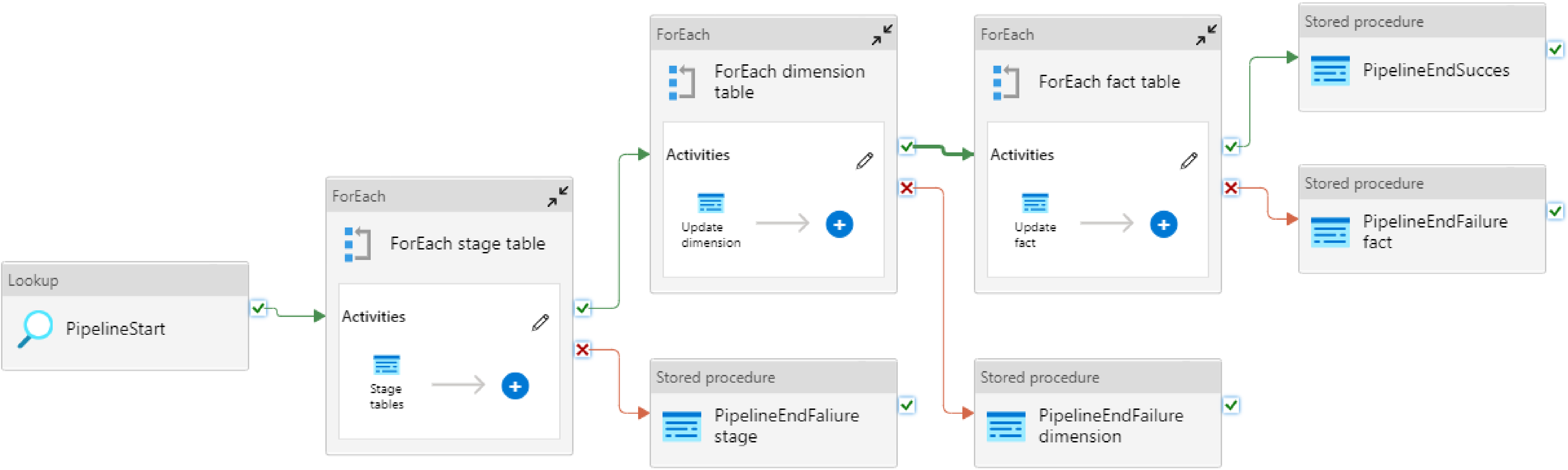


SIMPLE TRANSFORM PIPELINE

- Start Pipeline
- Stage data
- Transform dimensions
- Transform facts
- End Pipeline



DYNAMIC TRANSFORM PIPELINE



TRADITIONAL MODERN DATA WAREHOUSE

PRICING ESTIMATE

Service type	Description	Estimated Cost
Azure SQL Database	Single Database, DTU Purchase Model, Standard Tier, S0: 10 DTUs, 250 GB included storage per DB, 1 Database(s) x 640 Hours	**$ 12.90**
Azure SQL Database	Single Database, DTU Purchase Model, Standard Tier, S6: 400 DTUs, 250 GB included storage per DB, 1 Database(s) x 90 Hours	**$ 72.59**
Azure Data Factory	Azure Data Factory V2 Type, Data Pipeline Service Type, Azure Integration Runtime: 1 Activity Run(s), 30 Data movement unit(s), 90 Pipeline activities	**$ 9.72**
Monthly Total		**$ 95.21**

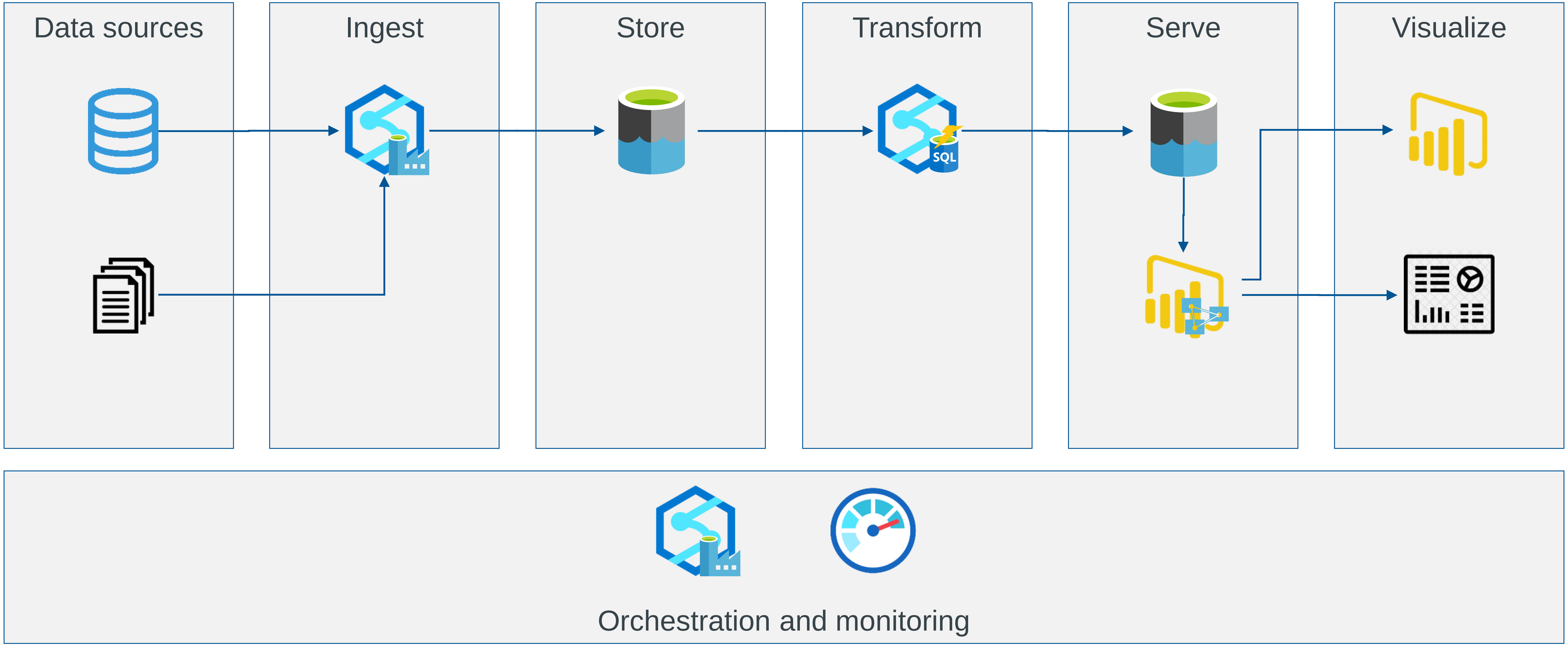
TRADITIONAL MODERN DATA WAREHOUSE

DIFFERENT ASPECTS

Advantages	Disadvantages
Simple	Pro code transformations with SQL
Modern	
Inexpensive	
Predictable performance	

NorthInsights

DATA LAKEHOUSE PATTERN



DATA LAKEHOUSE PATTERN

DIFFERENT ASPECTS

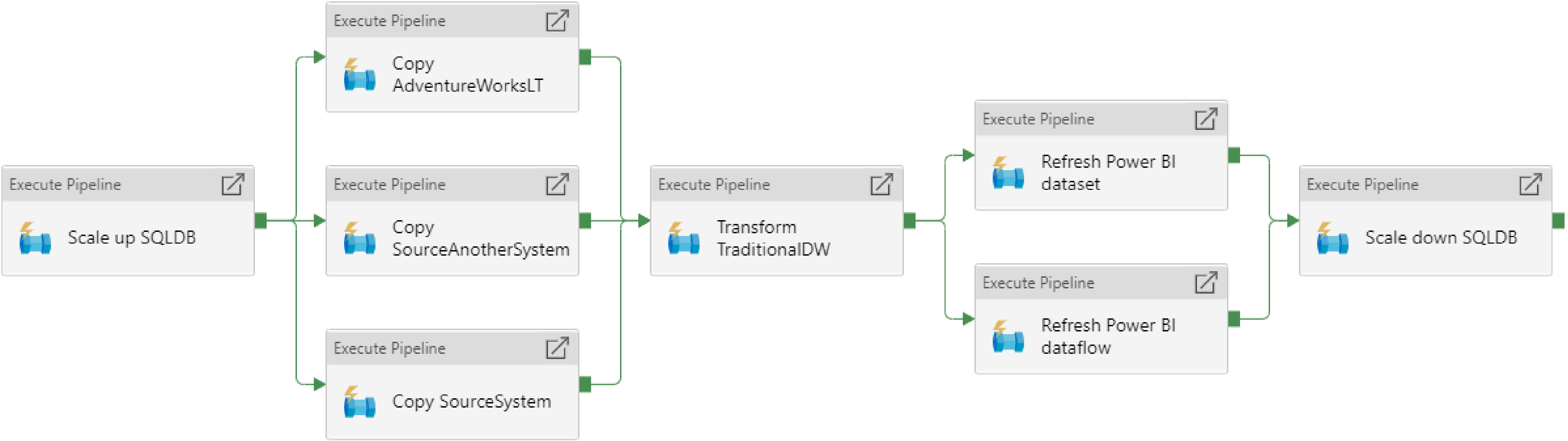
Advantages	Disadvantages
Simple	GUI tool to write transformations add cost – otherwise it´s code based
Modern	Can be tricky to calculate cost
Inexpensive	Unpredictable performance

DATA LAKEHOUSE PATTERN

PRICING ESTIMATE

Service type	Description	Estimated Cost
Storage Account	Data Lake Storage Gen2, Standard, LRS Redundancy, Hot Access Tier, Hierarchical Namespace File Structure, 100 GB Capacity	**$ 1.96**
Synapse SQL Serverless	You only pay for executed queries and the pricing is based on the amount of data processed by each query. Metadata-only queries (DDL statements) do not incur a cost. Queries will incur a minimum charge of 10 MB and each query will be rounded up to the nearest 1 MB of data processed.	**$ 5.00**
Azure Data Factory	Azure Data Factory V2 Type, Data Pipeline Service Type, Azure Integration Runtime: 1 Activity Run(s), 30 Data movement unit(s), 90 Pipeline activities	**$ 16.45**
Monthly Total		**$ 23.41**

ORCHESTRATION WITH PIPELINE



COMPARISON OF THE SOLUTIONS

	Traditional Modern DWH	Data Lakehouse
Transformation language	SQL	SQL
Portability to other services	High	High
Scalability in data volume	Medium	High
Pricing model	Predictable	Unpredictable
Monthly cost	Low	Low
Ad-hoc querying	Excluded	Included
Extensibility with AI, ML etc.	Low	High
Data processing	Schema on write	Schema on read
Compute and storage	Combined	Separate

KEY CONSIDERATIONS

- Keep it simple
  - Don´t overly do metadata driven pipelines
  - Put the logic in views or stored procedures
- Minimize cost
  - Consider weather keeping all history in raw is necessary
  - Delta load data if possible – this will add to complexity if something goes wrong
  - Consider closely which components to use in Synapse Pipelines

Session Feedback

https://bit.ly/dMC2023_SessionFeedback