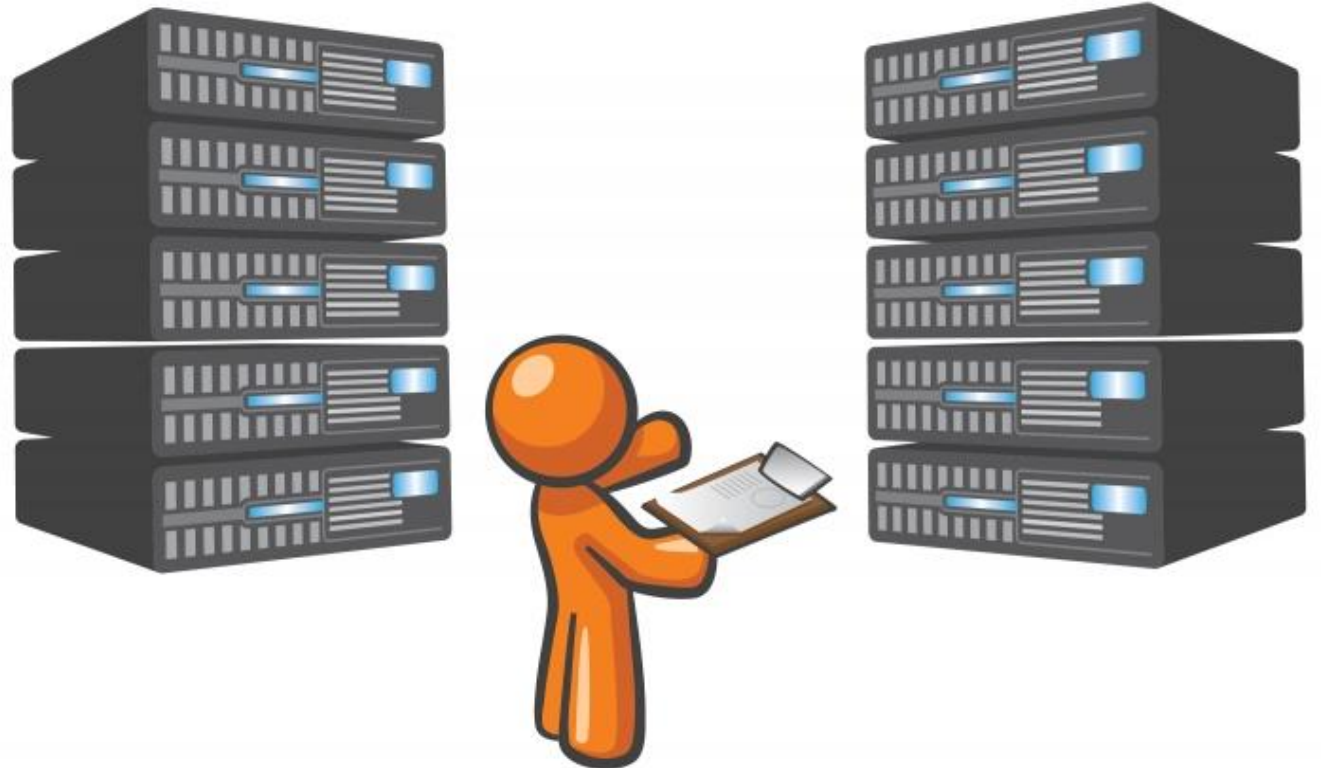


# Architectural blueprints for the Modern Data Warehouse

Data Saturday  
Denmark

25<sup>th</sup> March 2023



**ORANGEMAN**



# twoday capacity

## Fellowwind



## ORANGEMAN



# Just Blindbæk

- ▶ Self-employed BI consultant in justB
  - ▶ Trainer at Orange Man
  - ▶ Founder
    - Danish Microsoft BI Community ([MsBIP.dk](https://msbip.dk))
    - Power BI UG Denmark ([PowerBI.dk](https://powerbi.dk))
    - Power BI Cruise, Power BI Next Step and Data Platform Next Step
  - ▶ Strong focus on
    - Azure BI architecture
    - Analysis Services
    - Reporting Services
    - Power BI
- ▶ [just@justB.dk](mailto:just@justB.dk) / [blog.justB.dk](https://blog.justB.dk) / [@justblindbaek](https://twitter.com/justblindbaek) / [youtube.com/c/justblindbaek](https://youtube.com/c/justblindbaek)



**Microsoft**  
CERTIFIED  
Trainer

**Microsoft**  
CERTIFIED  
IT Professional

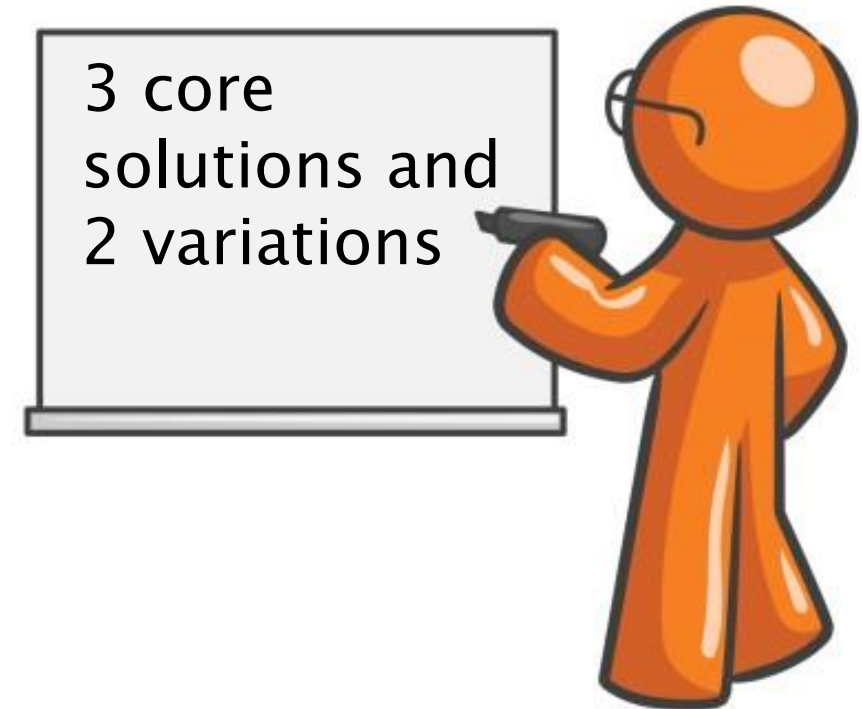
**SSAS**  
MAESTRO  
by Microsoft

**MVP** **Microsoft**  
Most Valuable  
Professional

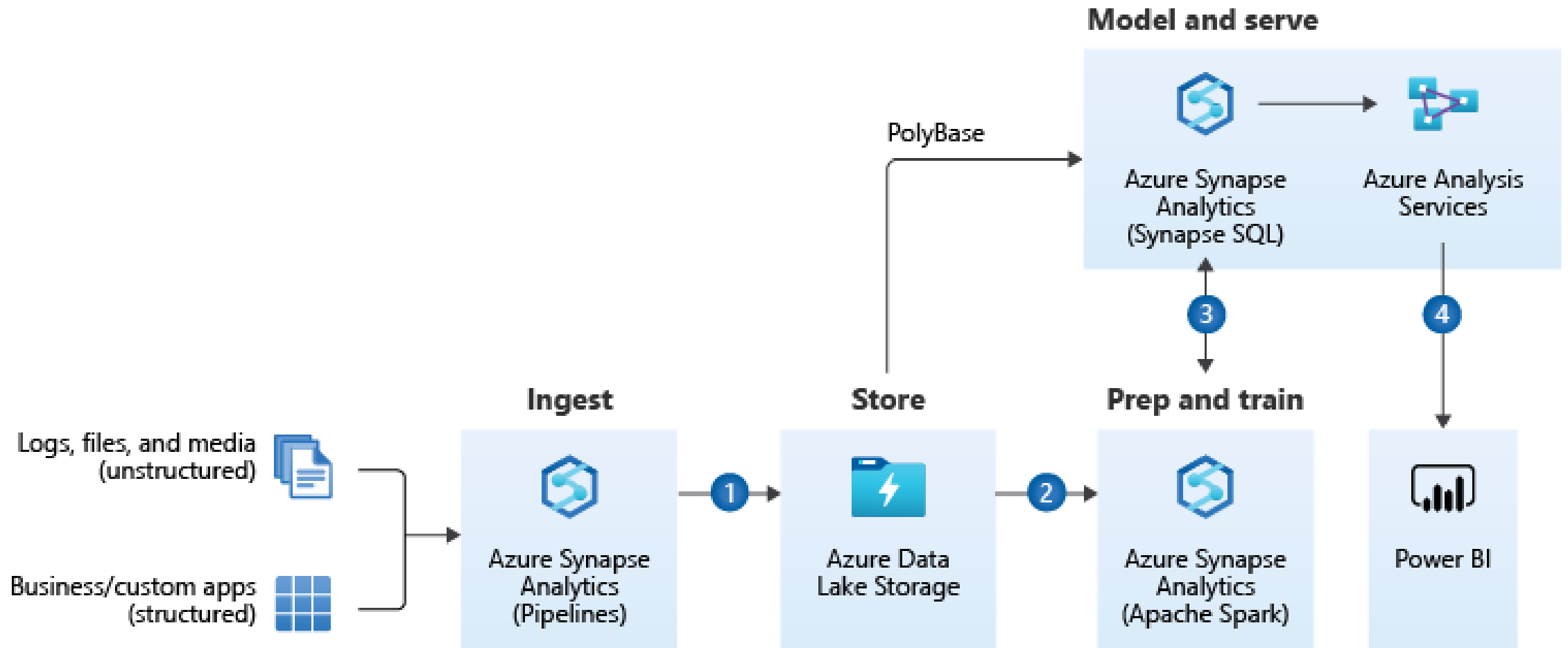
**ORANGEMAN**

# Agenda

- ▶ What is a Modern Data Warehouse?
- ▶ Traditional Modern Data Warehouse
- ▶ Data Lakehouse with Spark
- ▶ Self-service with Power BI Dataflows
- ▶ Comparison of the three solutions
- ▶ Other variations



# Enterprise Data Warehouse in Azure



# Enterprise Data Warehouse in Azure

## Model and serve

### Enterprise Data Warehouse

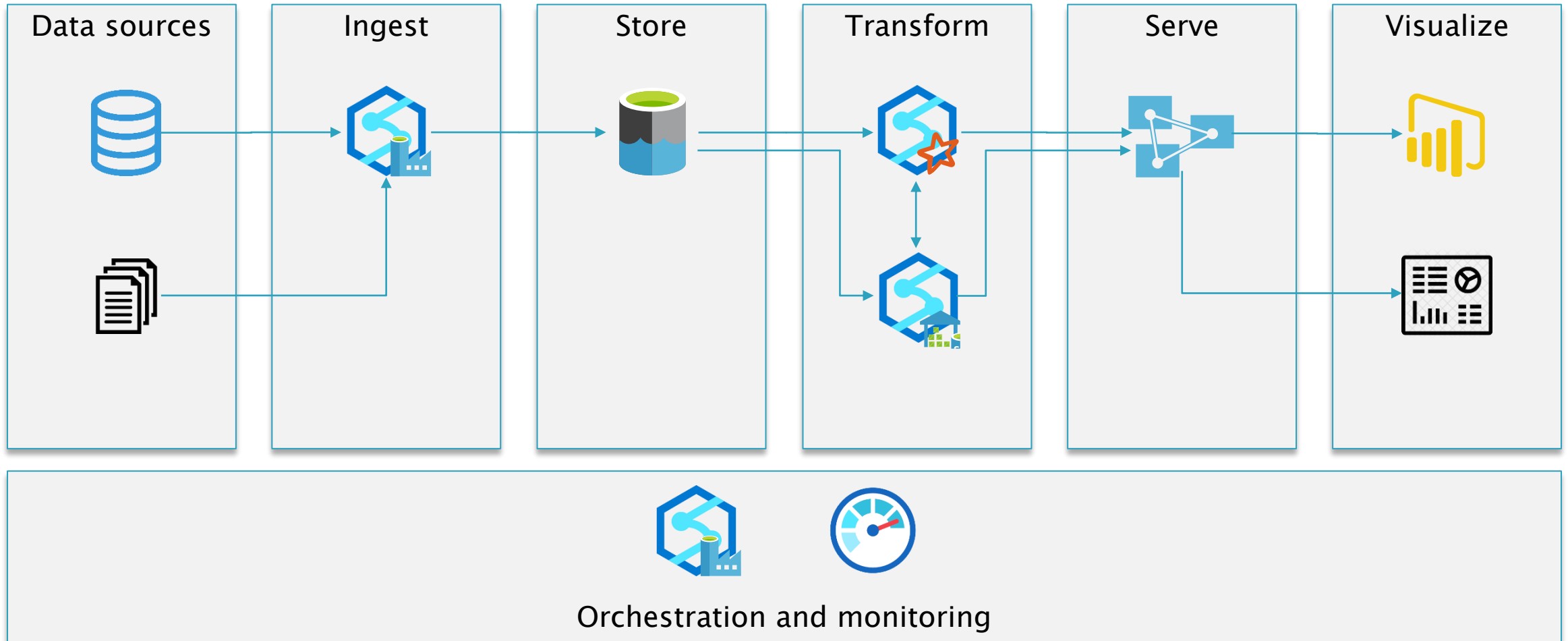
✓ Azure Synapse Analytics	Tier: Compute Optimized Gen2, Dedicated SQL Pool...	Monthly: \$1,755.90
✓ Azure Analysis Services	Developer (Hours), 5 Instance(s), 720 Hours	Monthly: \$475.20
✓ Storage Accounts	Data Lake Storage Gen2, Standard, LRS Redundancy...	Monthly: \$71.88
Estimated upfront cost		\$0.00
Estimated monthly cost		\$2,302.98

(structured) ■■■

(Pipelines)

(Apache Spark)

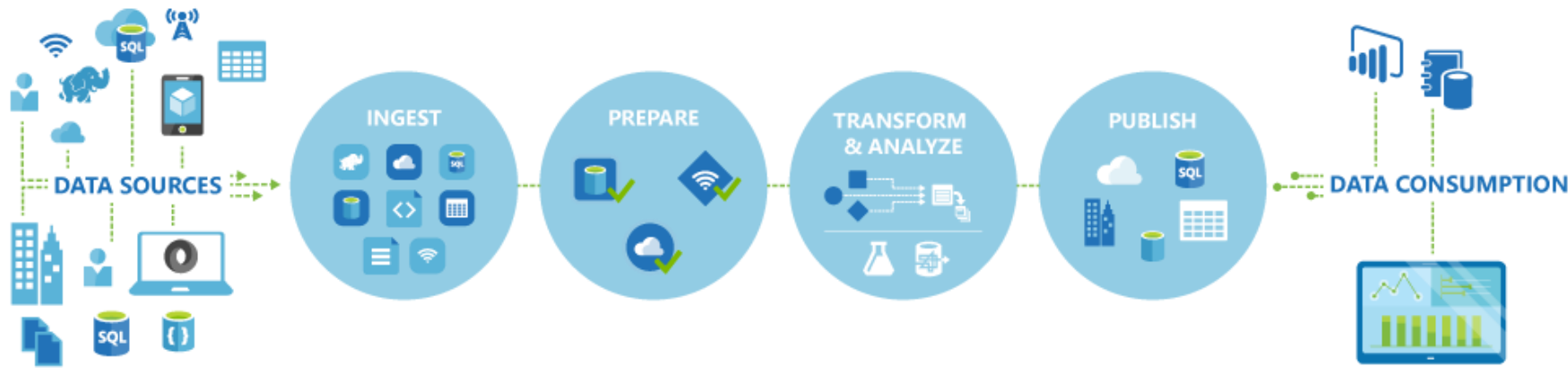
# Enterprise Data Warehouse in Azure



# Azure Data Factory in Synapse

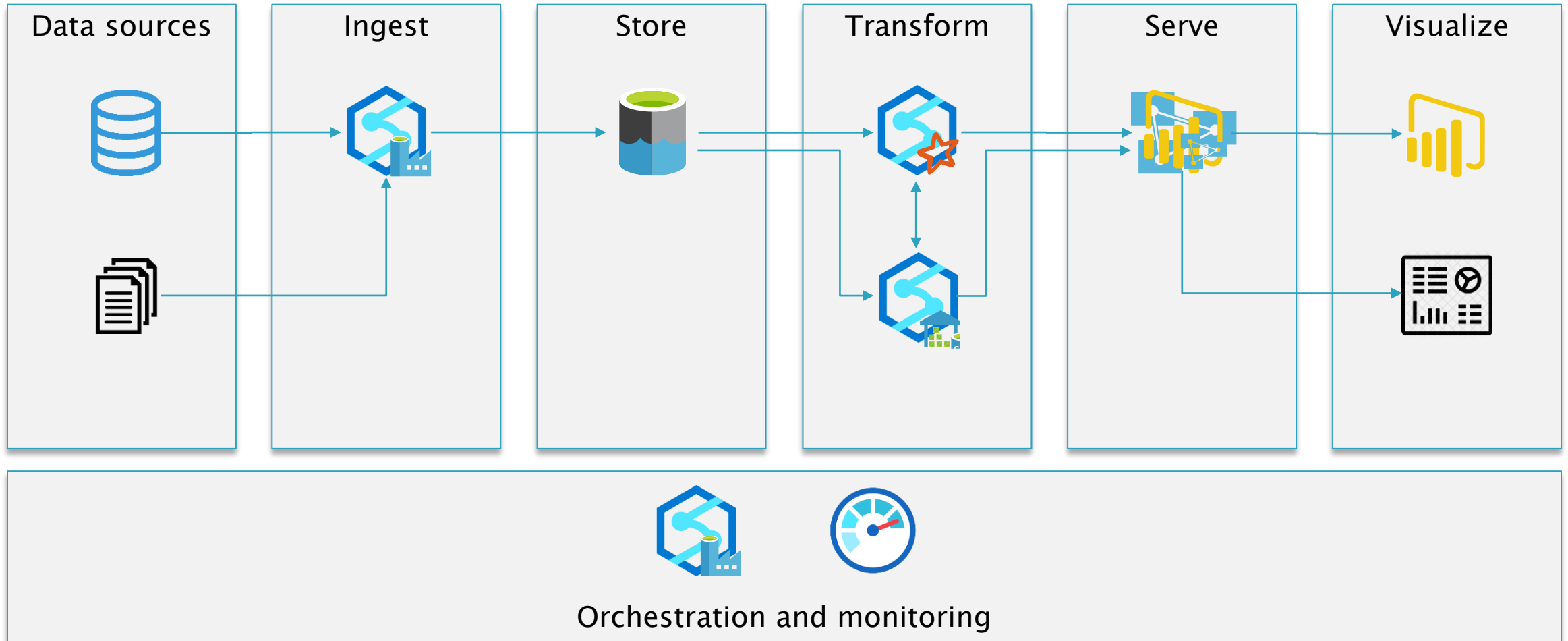


- ▶ Hybrid data integration at scale
- ▶ 90+ connectors provided to Cloud & Hybrid
- ▶ Create, plan, administer and monitor data pipelines
- ▶ Execute activities (copy, transform or orchestration)
- ▶ Parameters, control flow and triggers





# Enterprise Data Warehouse in Azure



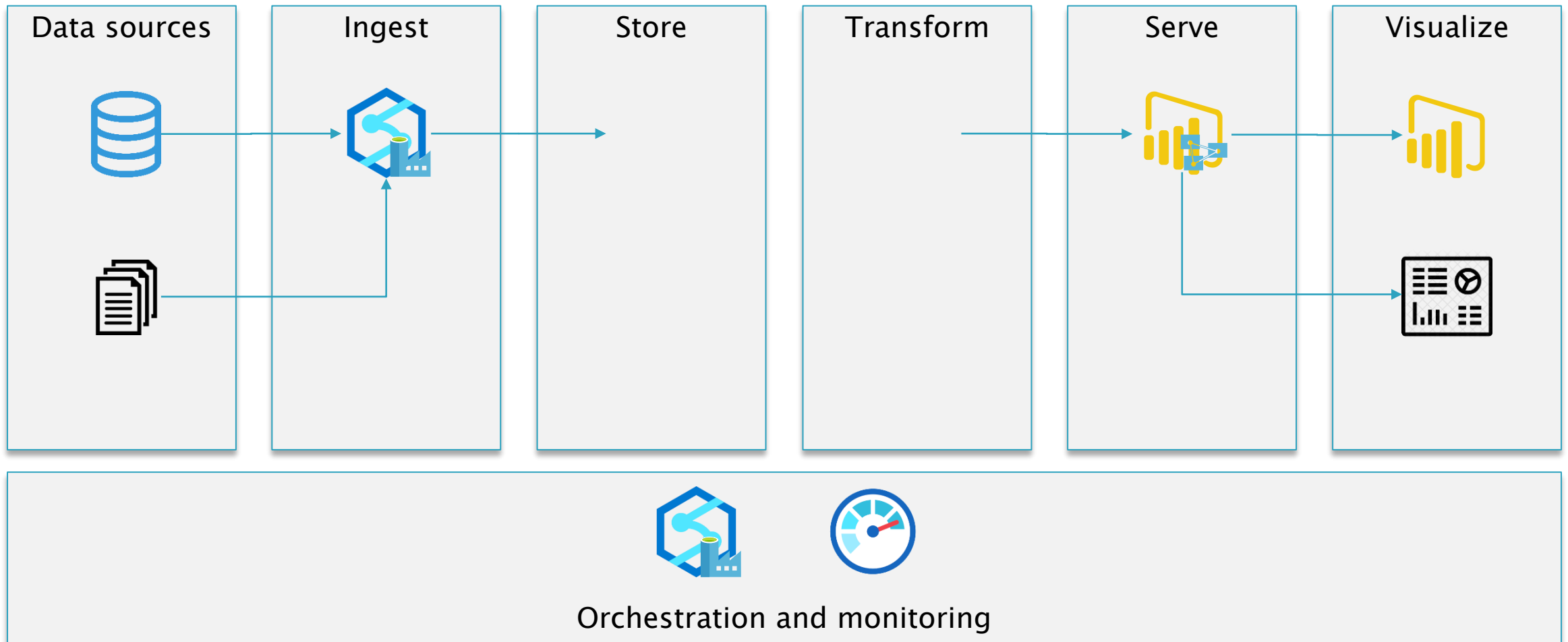
# Analysis Services in Power BI



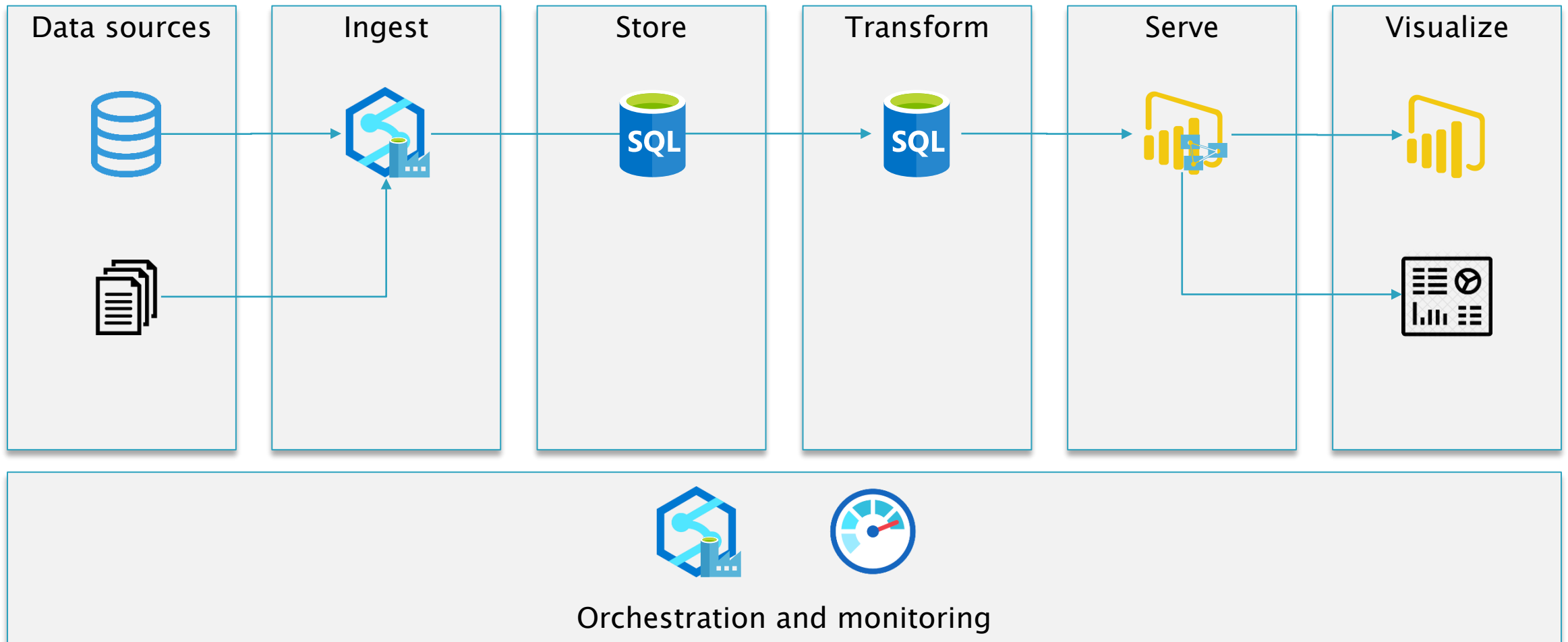
- ▶ Analysis Services engine inside
- ▶ XMLA endpoint to manage the Semantic data models
- ▶ Deploy models directly to a premium workspace from Tabular Editor and/or SQL Server Data Tools.
- ▶ Requires premium capacity when deploying the models
- ▶ License options
  - Power BI Premium capacity
  - Power BI Embedded capacity
  - Power BI Premium Per User



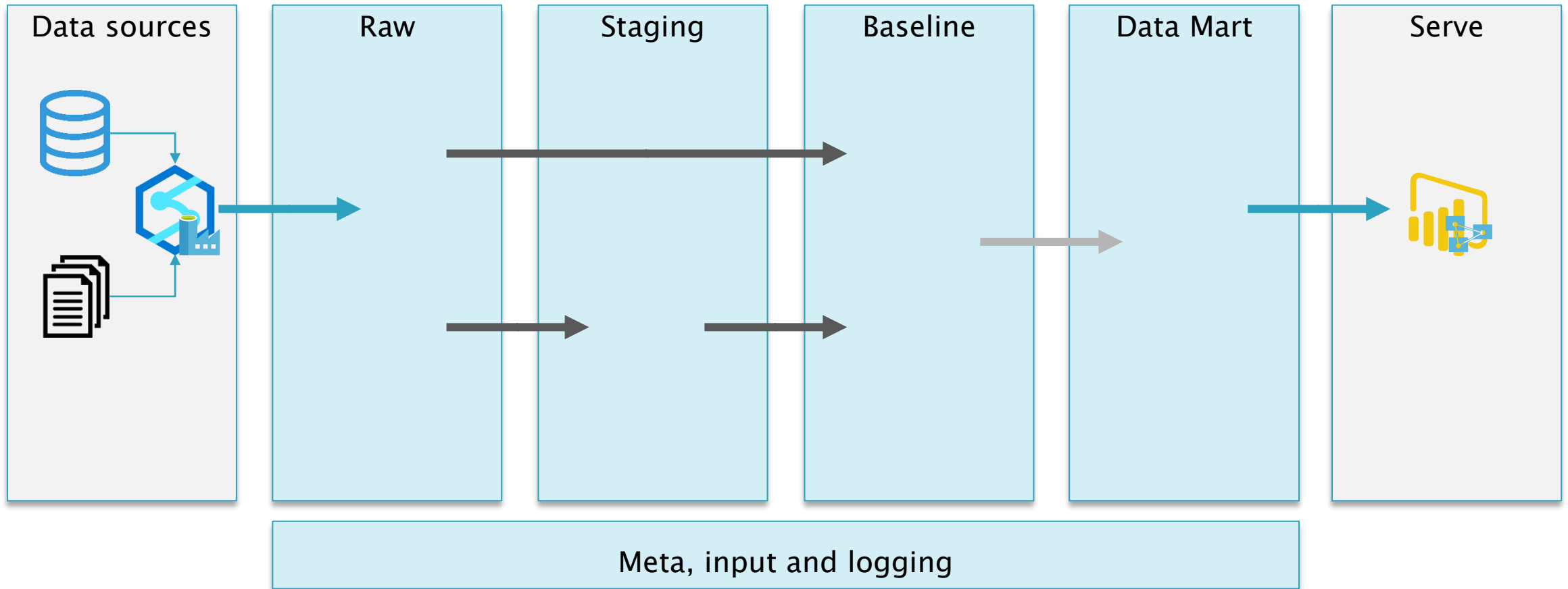
# Where to Store and Transform?



# (1) Traditional Modern Data Warehouse

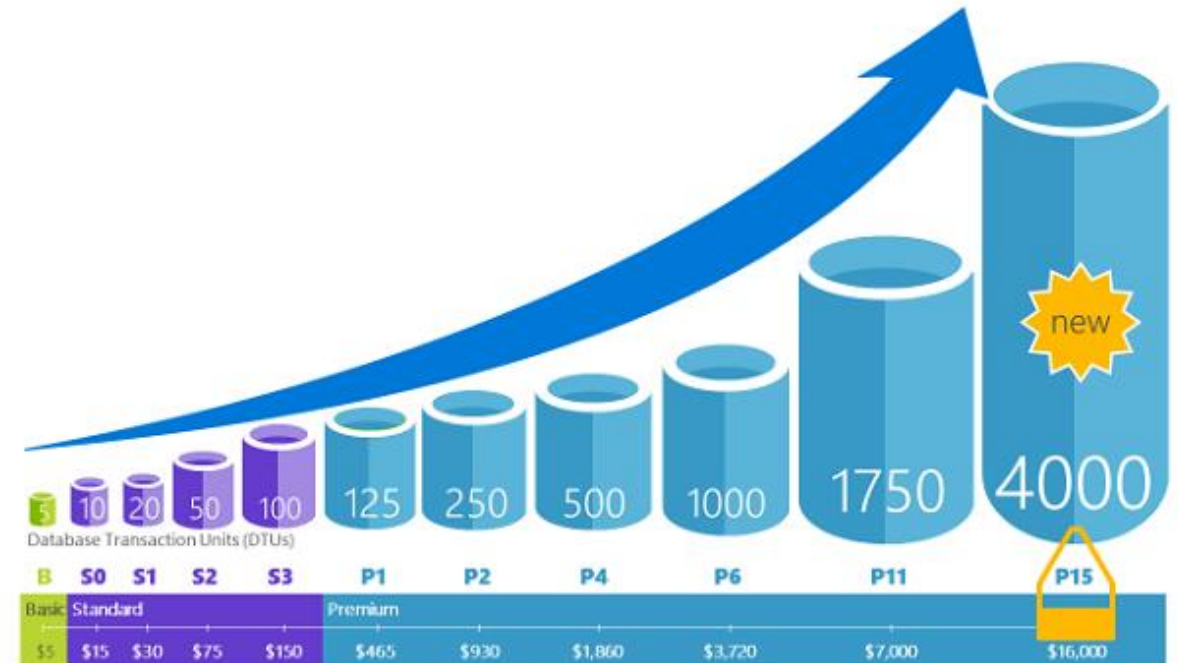
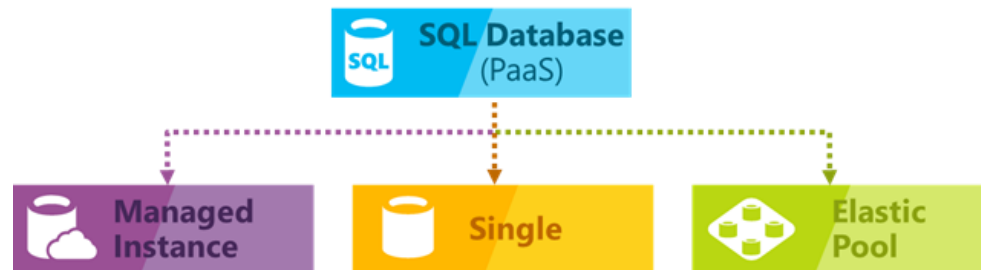


# Transform "framework" in SQL Database

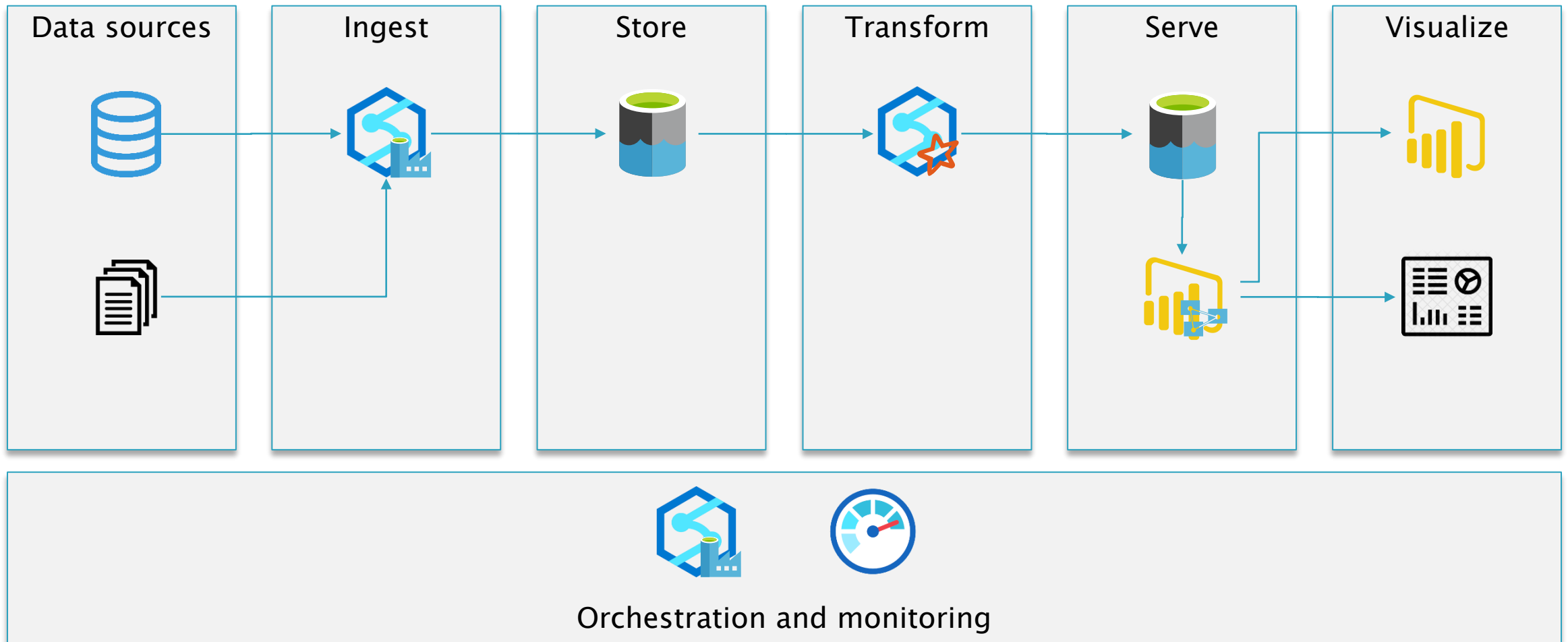


# Azure SQL Database

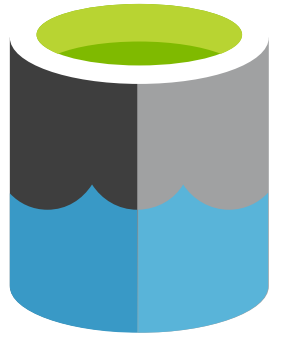
- ▶ Relational database-as-a service
- ▶ Scale to your needs
- ▶ DTU- and vCore-based models



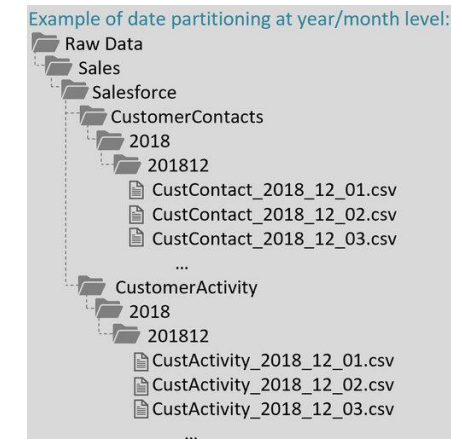
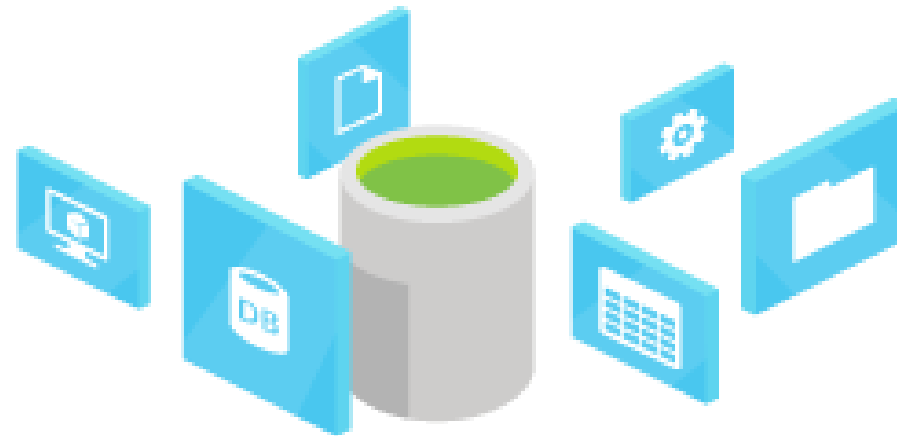
## (2) Data Lakehouse with Spark



# Azure Data Lake Store

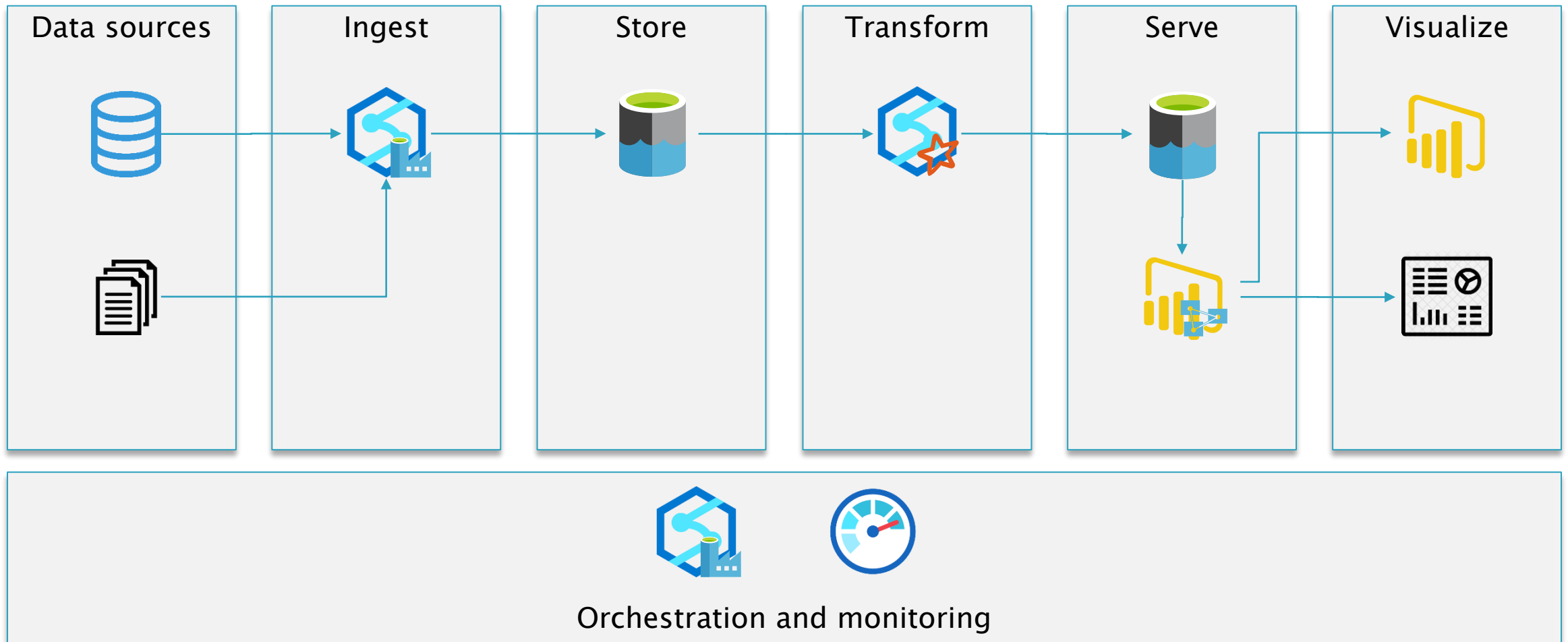


- ▶ Storage Account with “hierarchical namespace” enabled
- ▶ Unlimited cheap, cheap, cheap storage
- ▶ Can handle structured, semi-structured, and unstructured data
- ▶ Apache Hadoop file system compatible (HDFS)





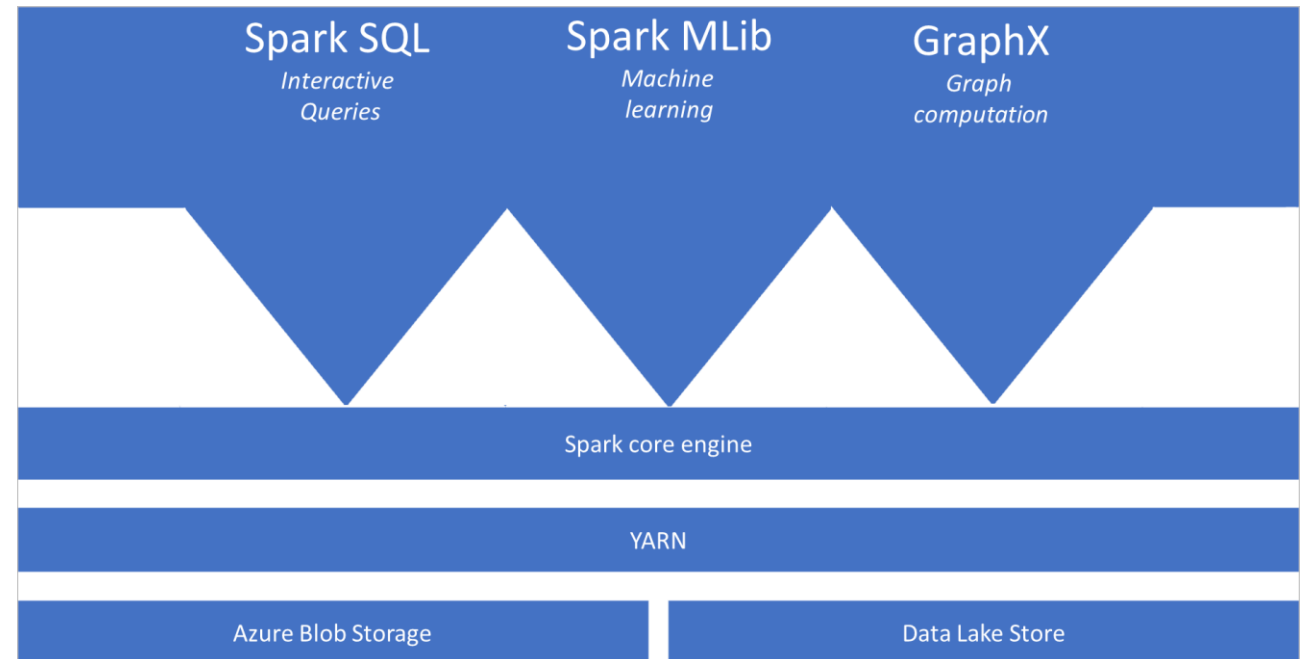
## (2) Data Lakehouse with Spark



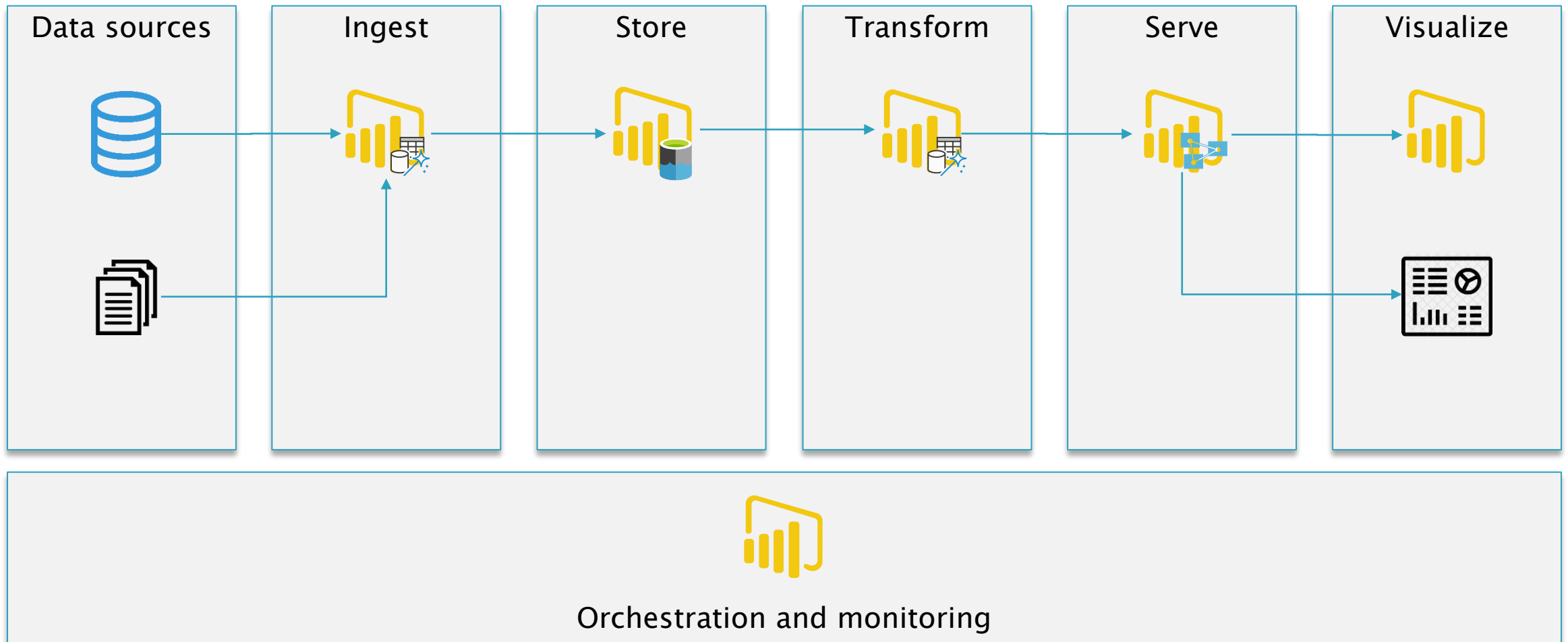
# Spark in Synapse



- ▶ Industry-standard Apache Spark in-memory engine
- ▶ The most popular open-source big data engine used for data preparation, data engineering, ETL, and machine learning
- ▶ Highly extensible with support for C#, Scala and PySpark alongside Spark SQL, GraphX, Streaming and Machine Learning Library (Mllib)
- ▶ Fast start-up and aggressive autoscaling

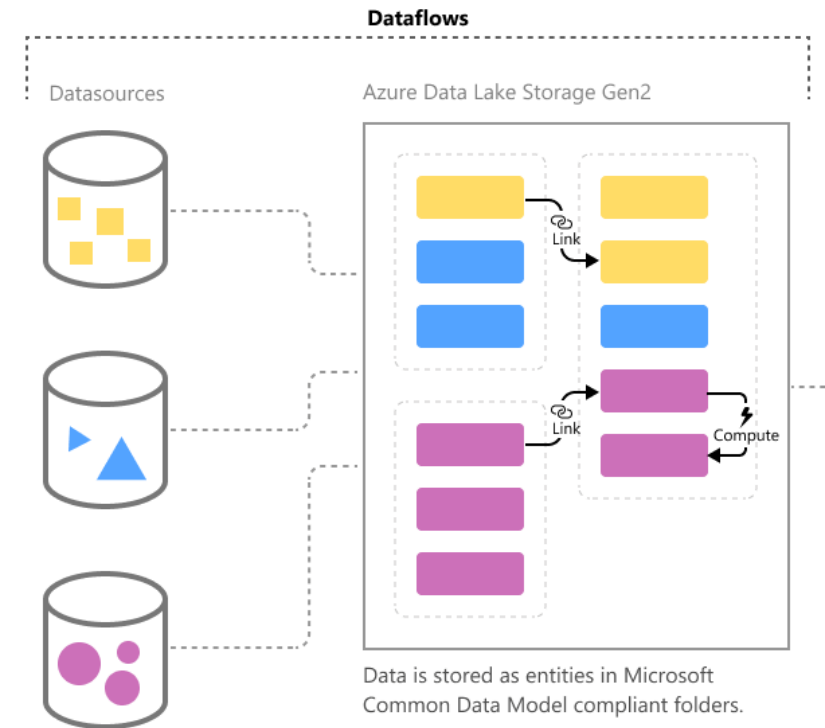


### (3) Self-service with Dataflows in Power BI



# Dataflows in Power BI

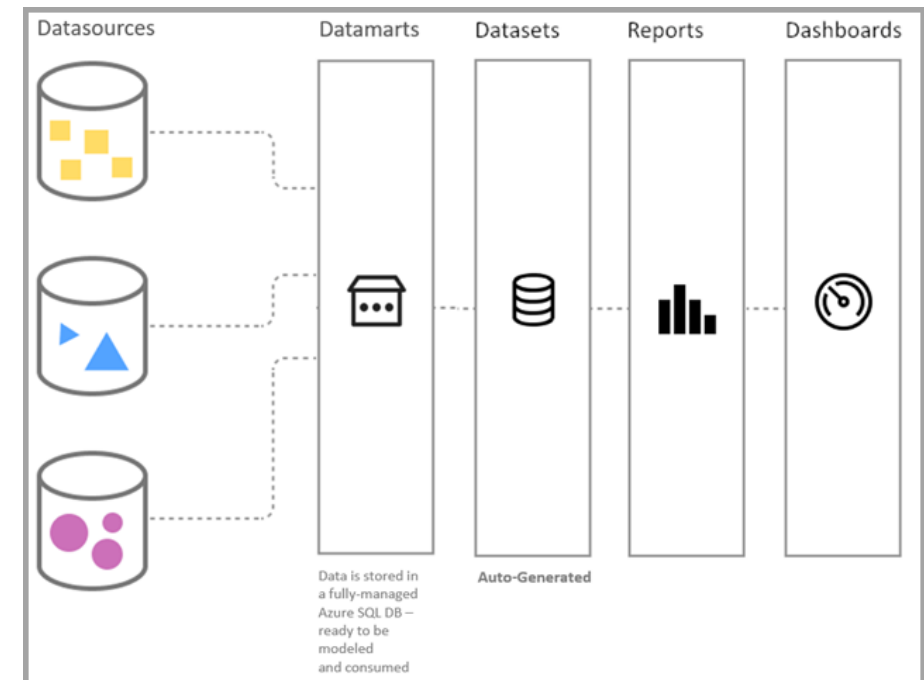
- ▶ Self-service data prep
- ▶ Power Query in the cloud
- ▶ Wide range of cloud and on-premise sources
- ▶ Link entities between dataflows
- ▶ Automatic triggers refresh processes for dependent entities in destination dataflows
- ▶ Bring your own Data Lake storage, enabling you to use other Azure services to read/write the underlying data



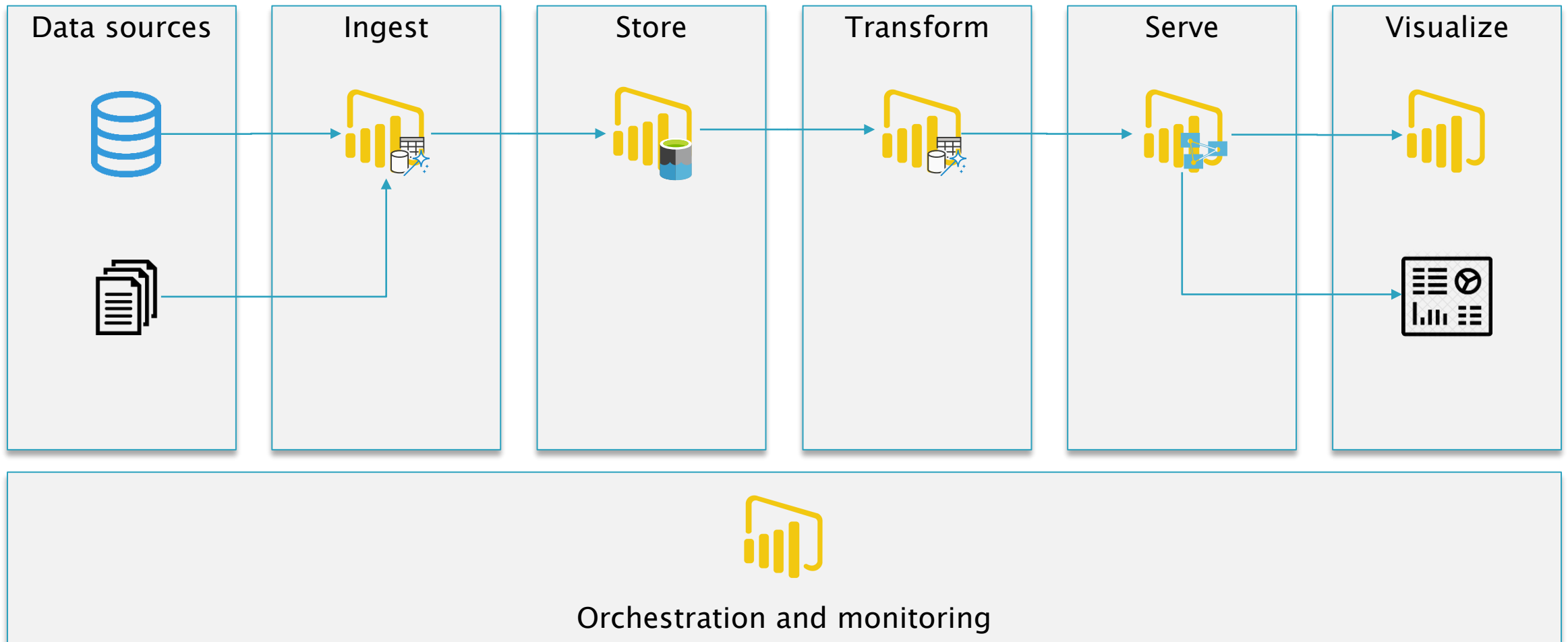
# Datamarts in Power BI



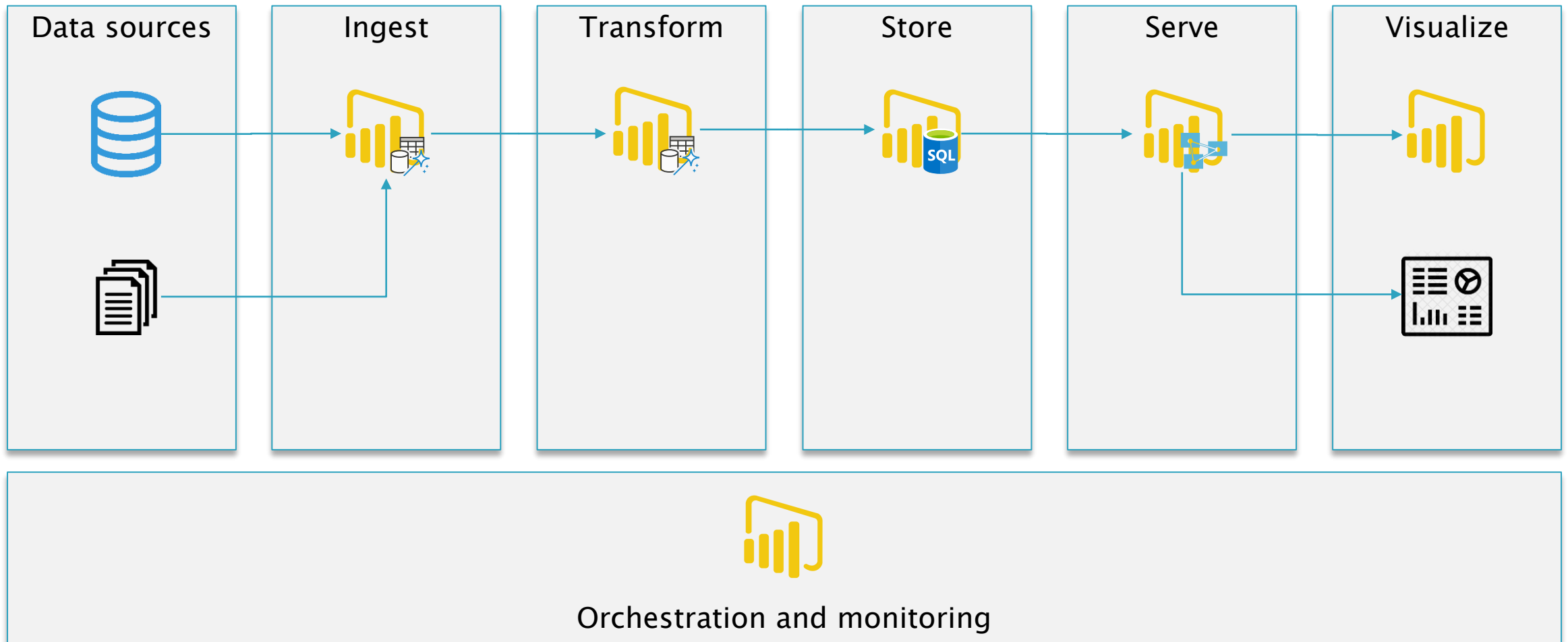
- ▶ Self-service data prep
- ▶ Power Query in the cloud
- ▶ Wide range of cloud and on-premise sources
- ▶ Fully managed Azure SQL database
- ▶ Provide data through a SQL endpoint
- ▶ For outputs that are results, sets, tables, and filtered tables of data
- ▶ Auto-generated dataset with DirectQuery



### (3) Self-service with Dataflows in Power BI

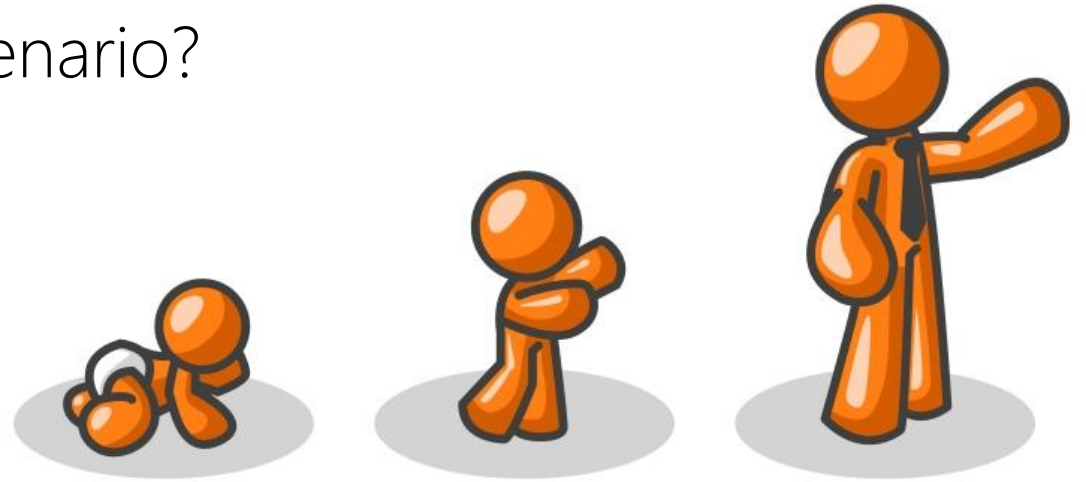


### (3) Self-service with Datamarts in Power BI



# Comparison of the solutions

- ▶ The conclusion – what's best in your senario?



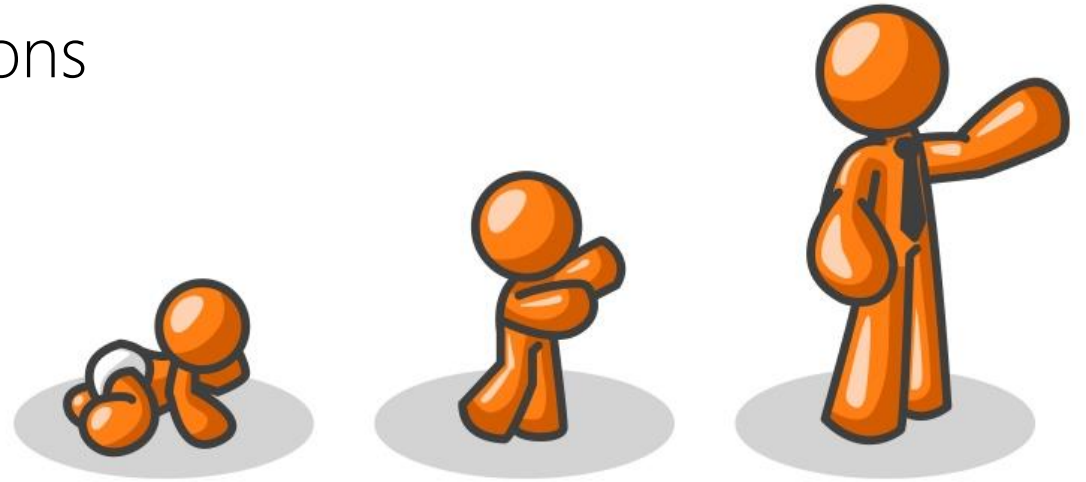


# Comparison of the solutions

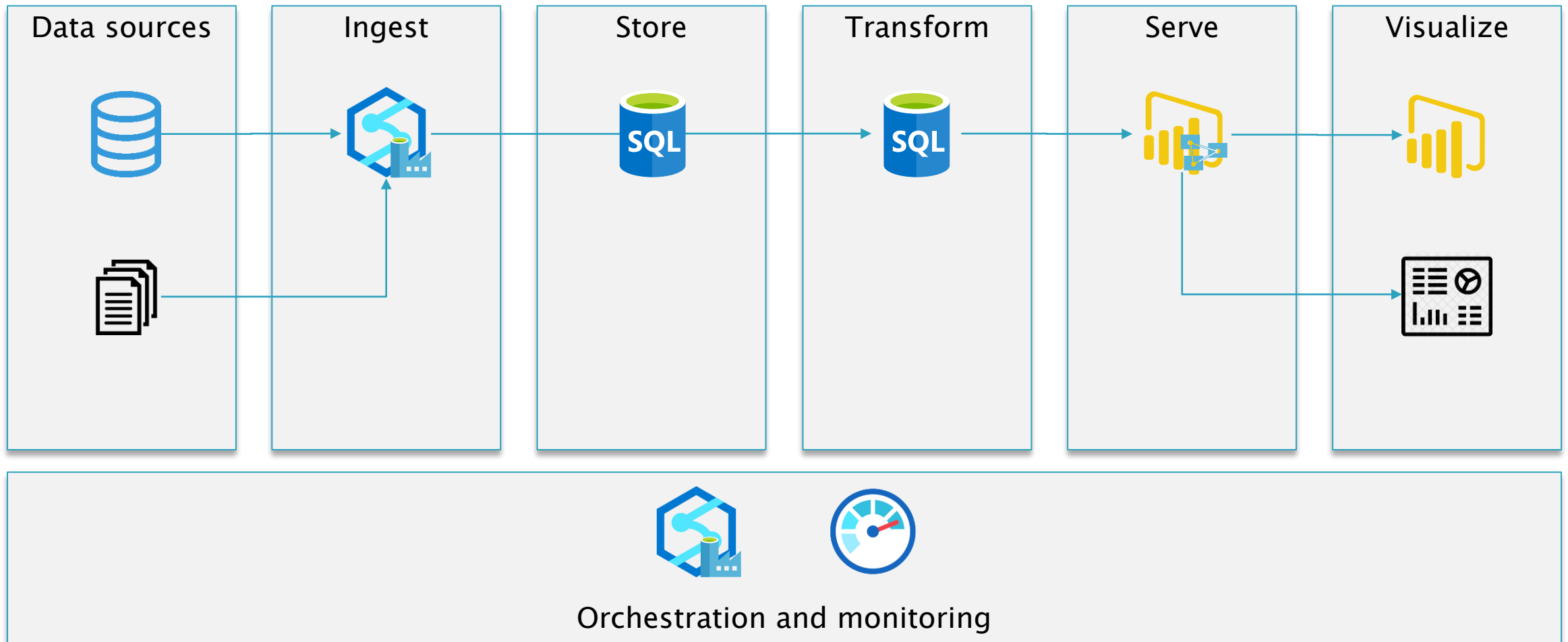
	Traditional Modern DWH	Data Lakehouse	Power BI Dataflows
Transformation language	SQL	SQL or PySpark	M or GUI
Developer persona	Pro	Pro	Citizen
Portability to other services	High	High	Low
Scalability in data volume	Medium	High	Low
Pricing model	Compute	Compute	Free?
Monthly cost	Low	Low	Very low
Extensibility with AI, ML etc.	Low	High	Low
Data processing	Schema on write	Schema on read	Schema on write
DevOps possibilities	High	High	Very low

# Other variations

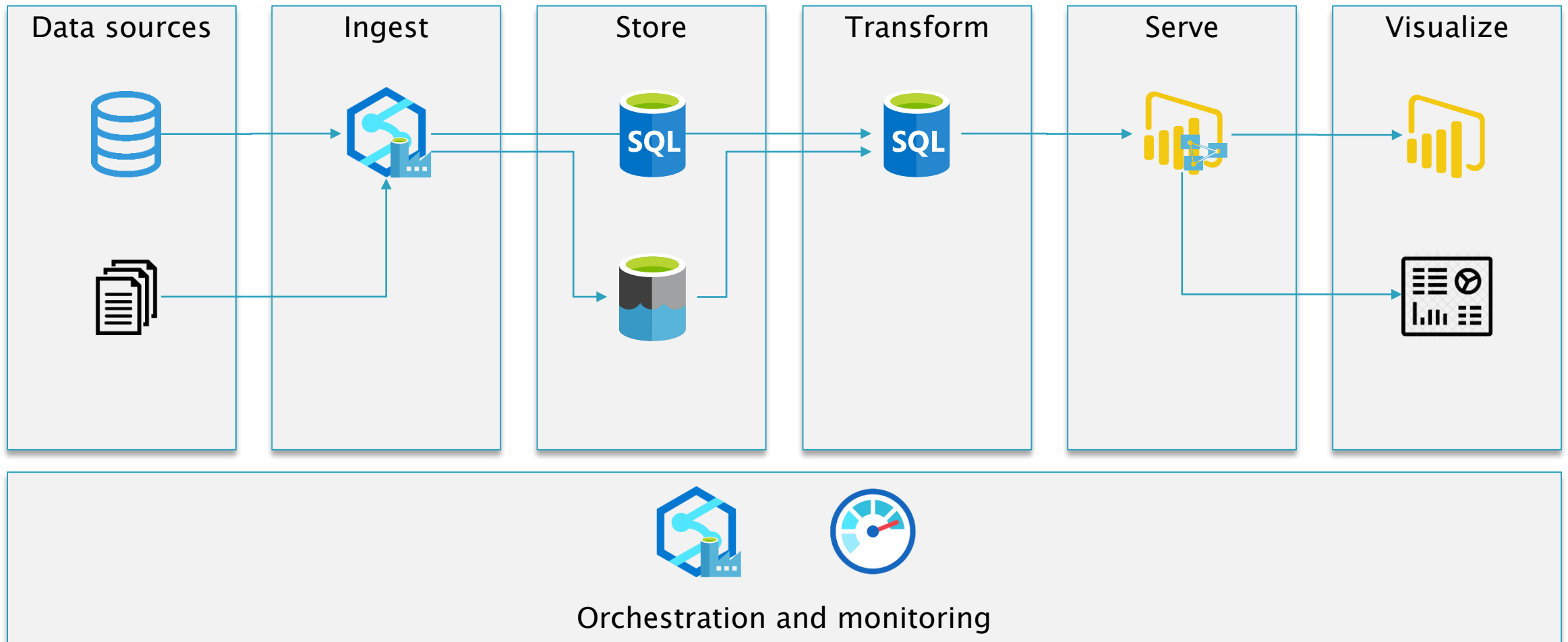
- ▶ Mix'n'match and combining the solutions



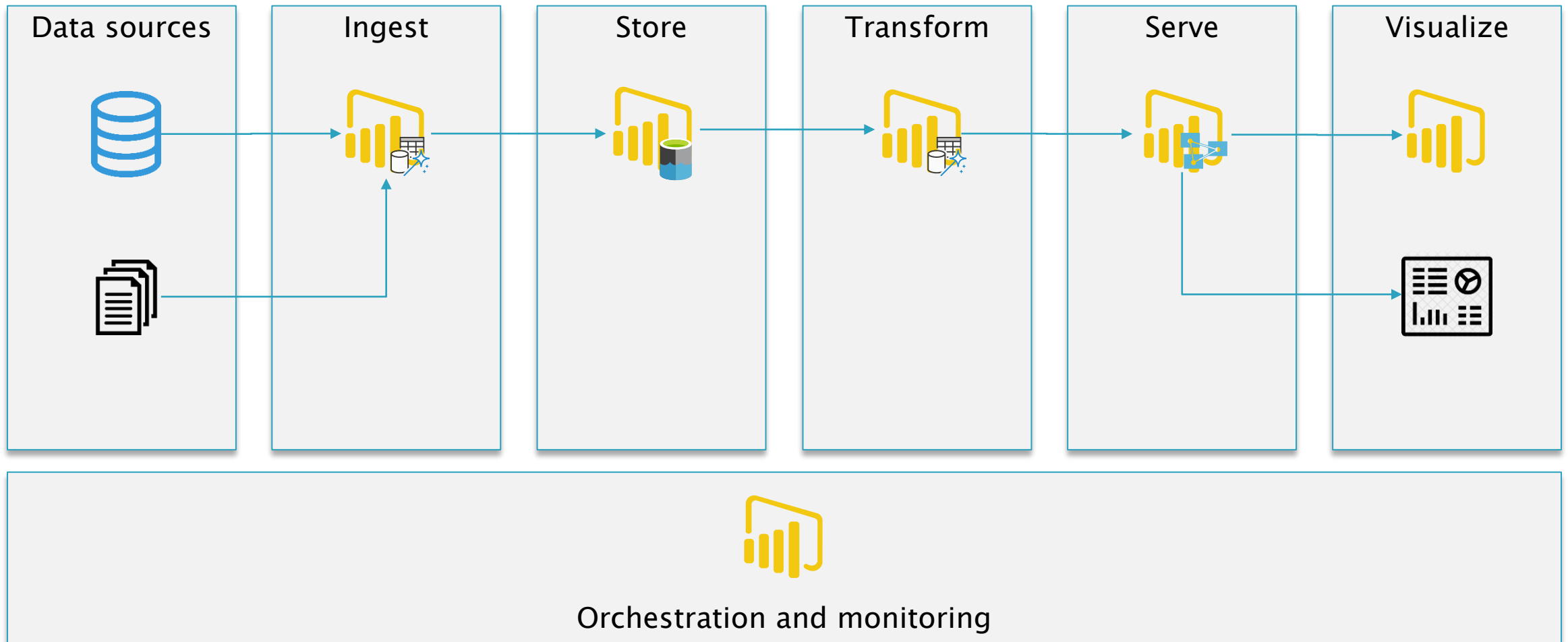
# Traditional Modern Data Warehouse



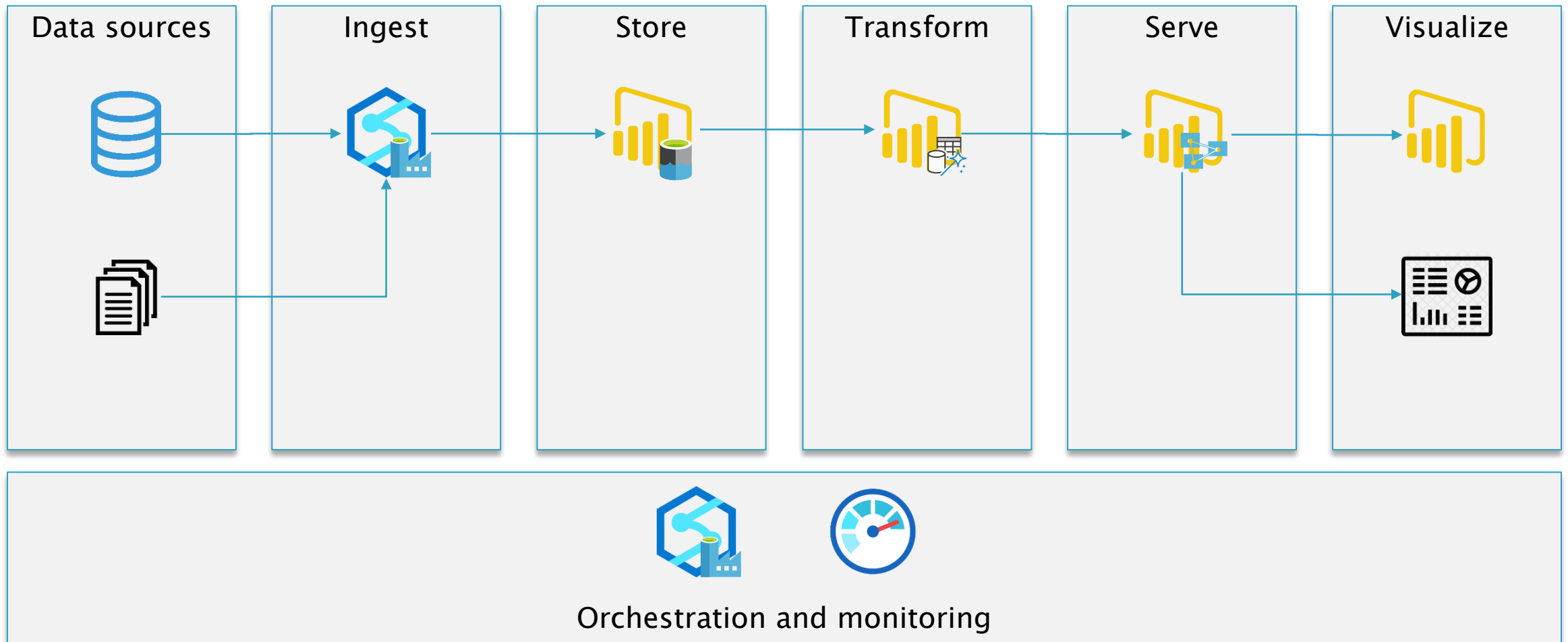
# Traditional Modern Data Warehouse with Lake



# Dataflows in Power BI

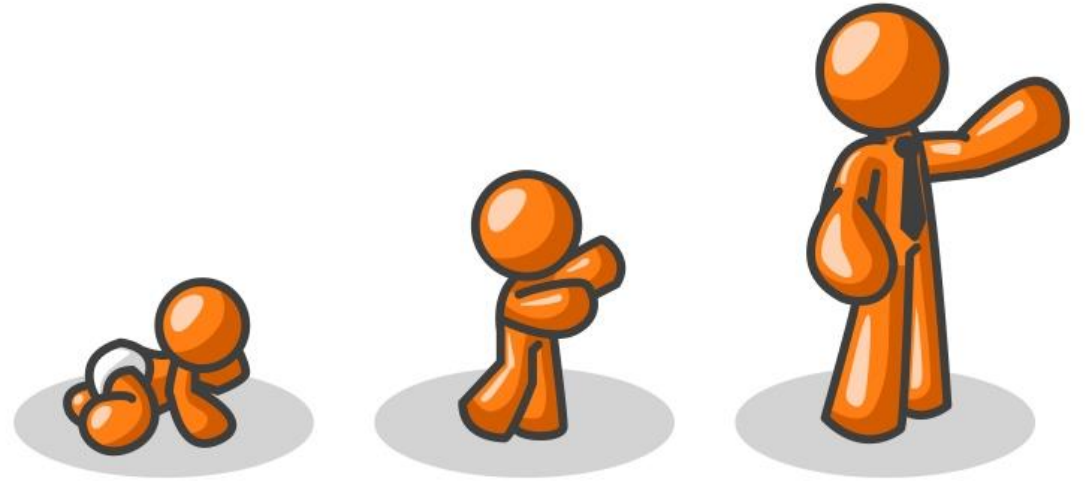


# Dataflows in Power BI with Data Factory



# Pricing

- ▶ Estimate



**ORANGEMAN**

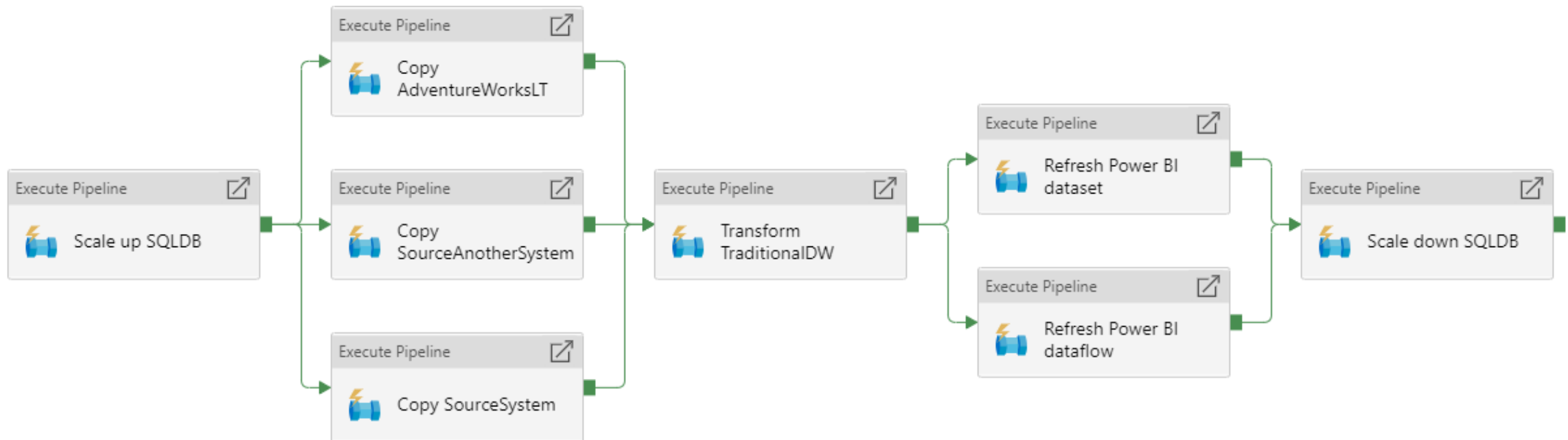
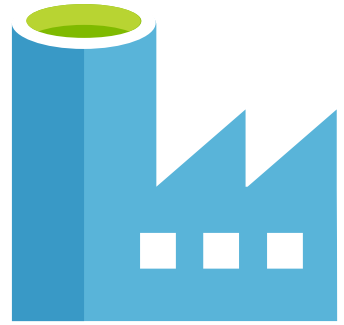
# Pricing estimate: Traditional Modern DW

Service type	Description	Estimated Cost
Storage Account	Data Lake Storage Gen2, Standard, LRS Redundancy, Hot Access Tier, Hierarchical Namespace File Structure, 100 GB Capacity	\$ 3.71
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
Power BI Embedded	Power BI Embedded, Node type: A1, 3GB Memory, 10 Hours	\$ 10.08
Azure Monitor	Log analytics:	\$ 1.60
	Monthly Total	\$ 110.60



# Orchestration

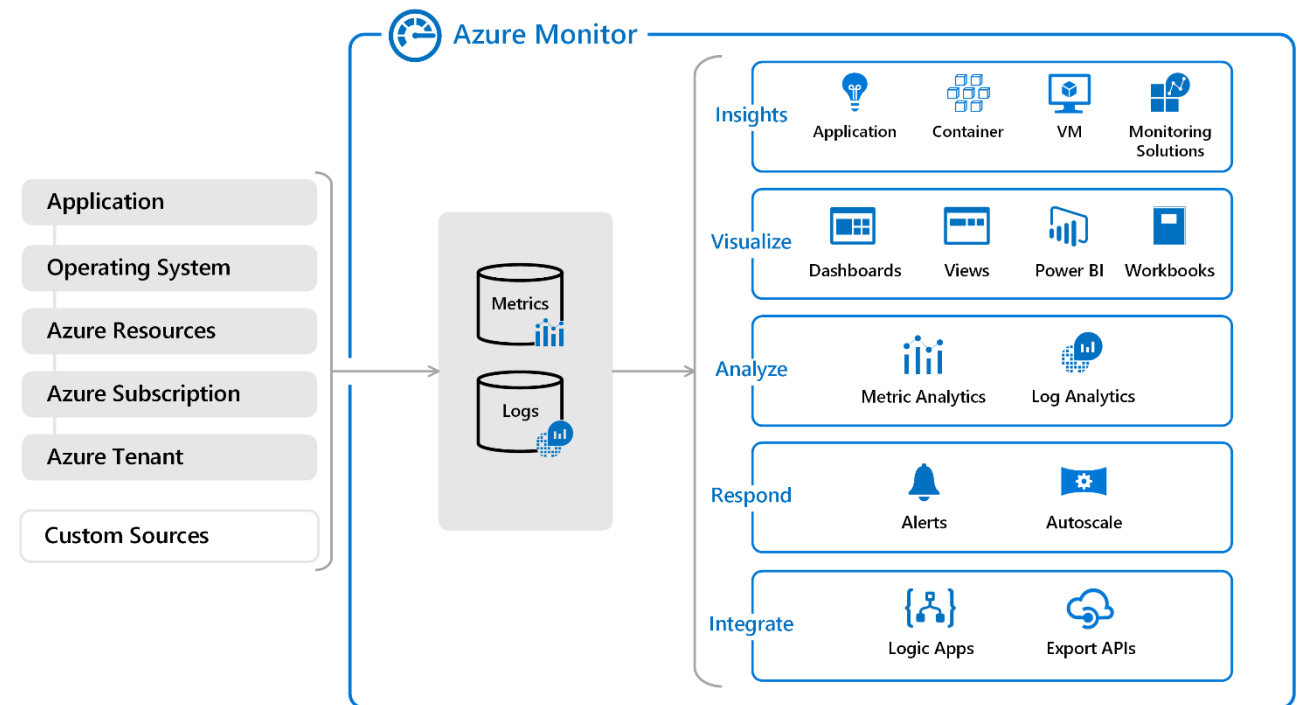
- ▶ Master Pipeline in Azure Data Factory



# Monitoring with Azure Monitor



- ▶ Gather diagnostic from Azure Services
- ▶ Analyze with Log Analytics
- ▶ Setup alerts
- ▶ Export with Power Query



# Usefull ressources

- ▶ Enterprise Data Warehouse: [https://docs.microsoft.com/en-us/azure/architecture/solution-ideas/articles/enterprise-data-](https://docs.microsoft.com/en-us/azure/architecture/solution-ideas/articles/enterprise-data-warehouse)
- ▶ Traditional Modern Data Warehouse on GitHub :  
<https://github.com/justBlindbaek/TraditionalModernDW>
- ▶ Pricing estimate:  
<https://azure.com/e/9d56929b959546b384594e1074a1f506>



# Key takeaways

- ▶ Cheap is an option – or actually, there are multiple options!
- ▶ Current skillset play a big role – or are you ready to change?
- ▶ It's easy to get started with a Proof-of-Concept



# Just Blindbæk

- ▶ Self-employed BI consultant in justB
  - ▶ Trainer at Orange Man
  - ▶ Founder
    - Danish Microsoft BI Community ([MsBIP.dk](https://msbip.dk))
    - Power BI UG Denmark ([PowerBI.dk](https://powerbi.dk))
    - Power BI Cruise, Power BI Next Step and Data Platform Next Step
  - ▶ Strong focus on
    - Azure BI architecture
    - Analysis Services
    - Reporting Services
    - Power BI
- ▶ [just@justB.dk](mailto:just@justB.dk) / [blog.justB.dk](https://blog.justB.dk) / [@justblindbaek](https://twitter.com/justblindbaek) / [youtube.com/c/justblindbaek](https://youtube.com/c/justblindbaek)



**Microsoft**  
CERTIFIED  
Trainer

**Microsoft**  
CERTIFIED  
IT Professional

**SSAS**  
MAESTRO  
by Microsoft

**MVP** **Microsoft**  
Most Valuable  
Professional

**ORANGEMAN**

# Questions



ORANGEMAN



# twoday capacity

## Fellowwind



ORANGEMAN

