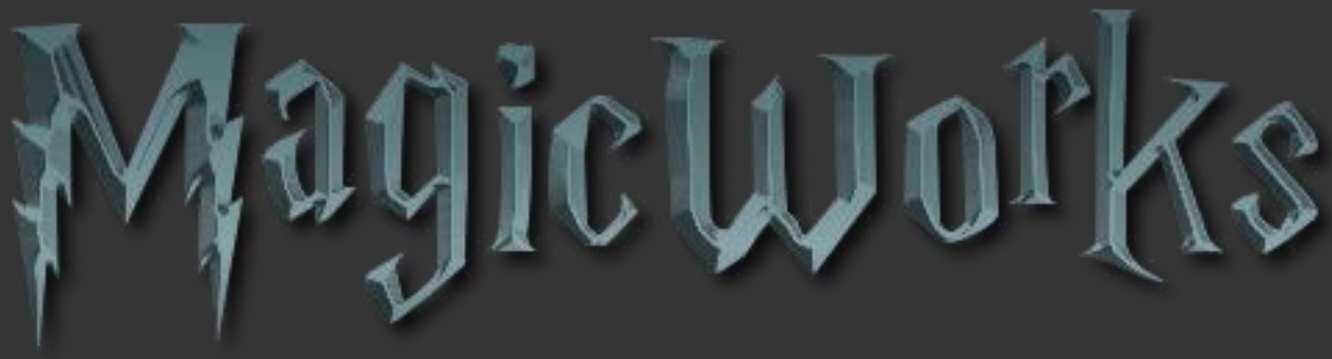
Exposing Data



MagicWorks™ now have a fully loaded Warehouse but their analytics team is growing, and more of the business wants access to data!

How can we extend the warehouse for different business scenarios?

Agenda

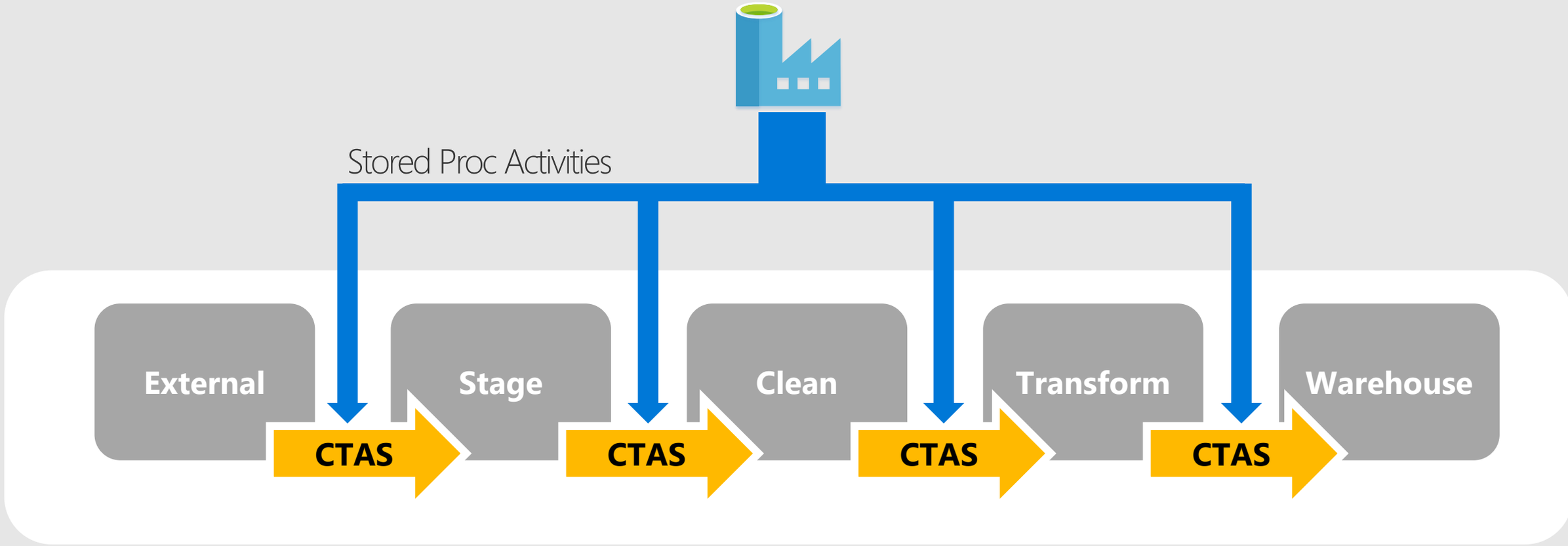
Orchestration & Automation

Introducing workload isolation

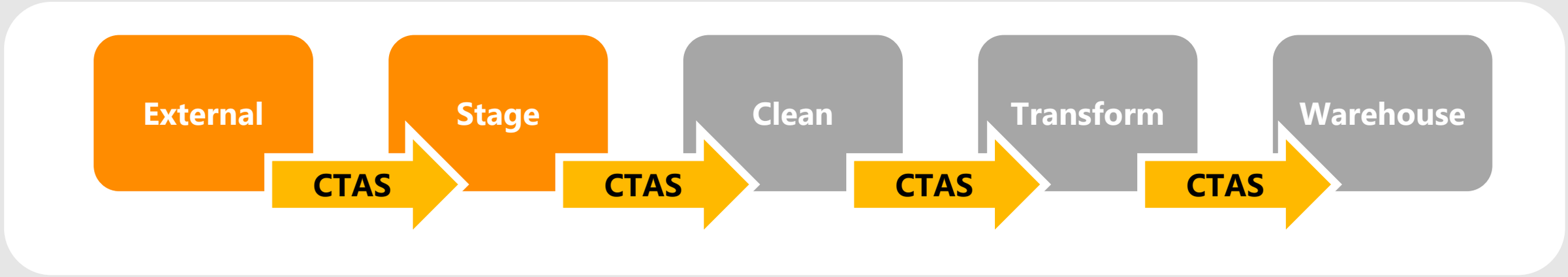
Benefits of workload isolation

Evaluating spoke options

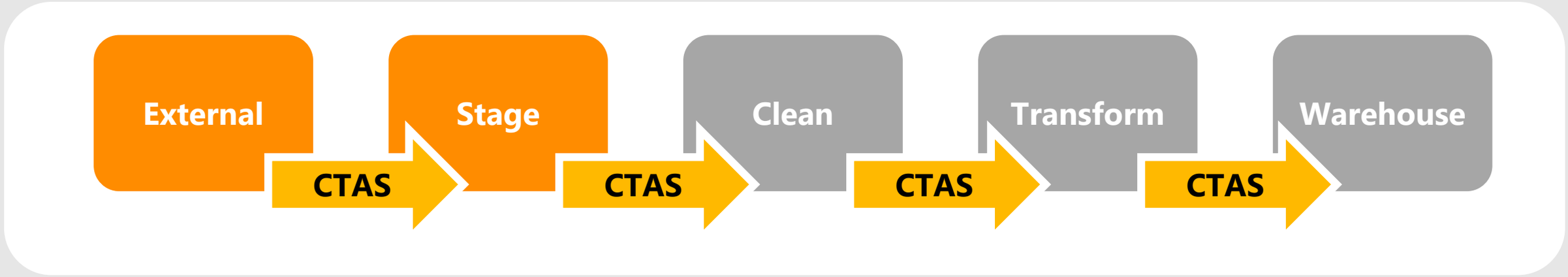
Orchestration & Automation



Data Factory Processing Pipelines

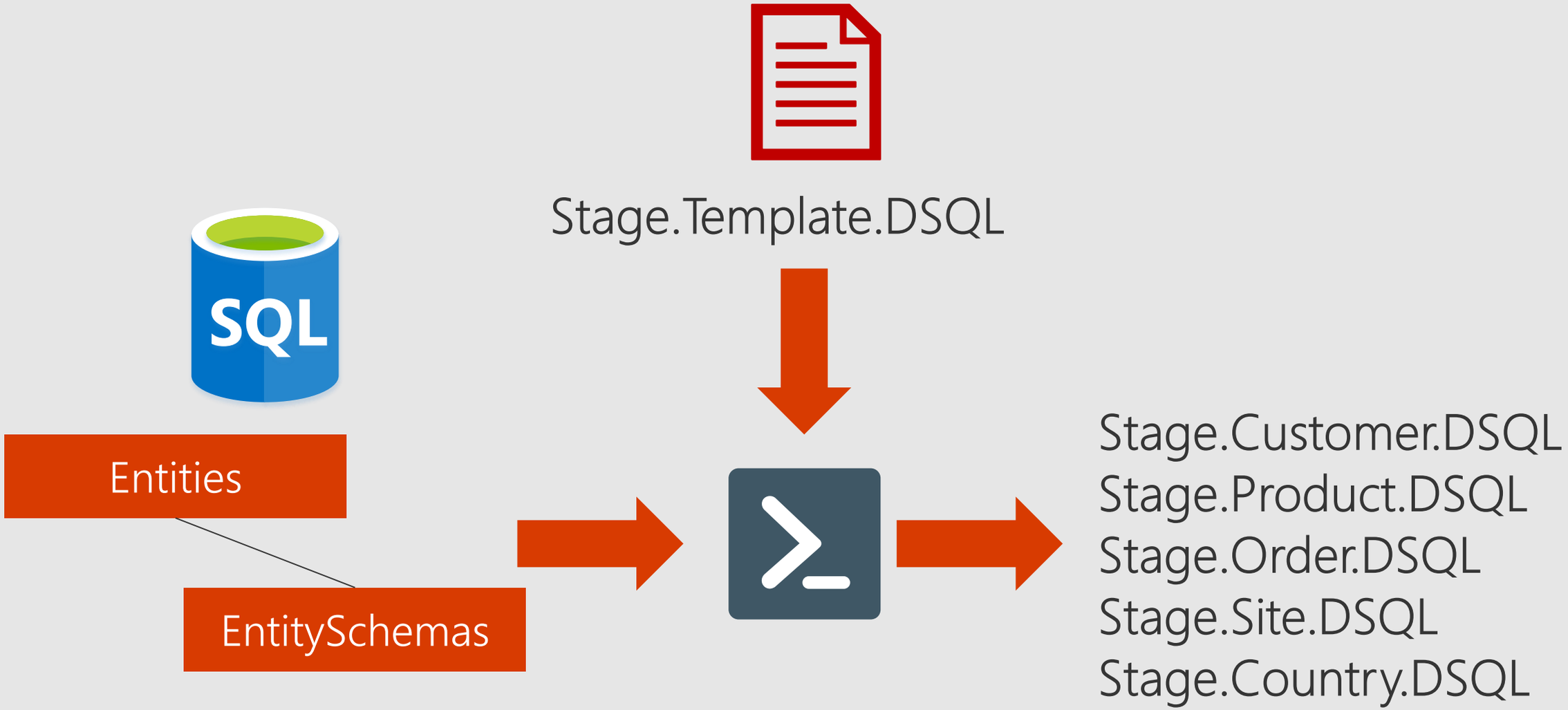



```
CREATE TABLE Staging.Customer
AS
SELECT
        Column1,
        Column2,
        Column3
FROM dbo.ExternalCustomer
```




```
CREATE TABLE Staging.%ENTITY%
AS
SELECT
        %COLUMNS%
FROM dbo.External%ENTITY%
```

Stage.Template.DSQL


Stage.Template.DSQL
SQL
Entities
EntitySchemas
Stage.Customer.DSQL
Stage.Product.DSQL
Stage.Order.DSQL
Stage.Site.DSQL
Stage.Country.DSQL

Introducing workload isolation

Drivers for workload isolation

SQLDW Strengths

Secure environment

Scalable load

Storage scale

Elastic compute

- Set based transformations
- Aggregation management
- Ad-hoc query at massive scale

SQLDW Weaknesses

Concurrency (max 32)

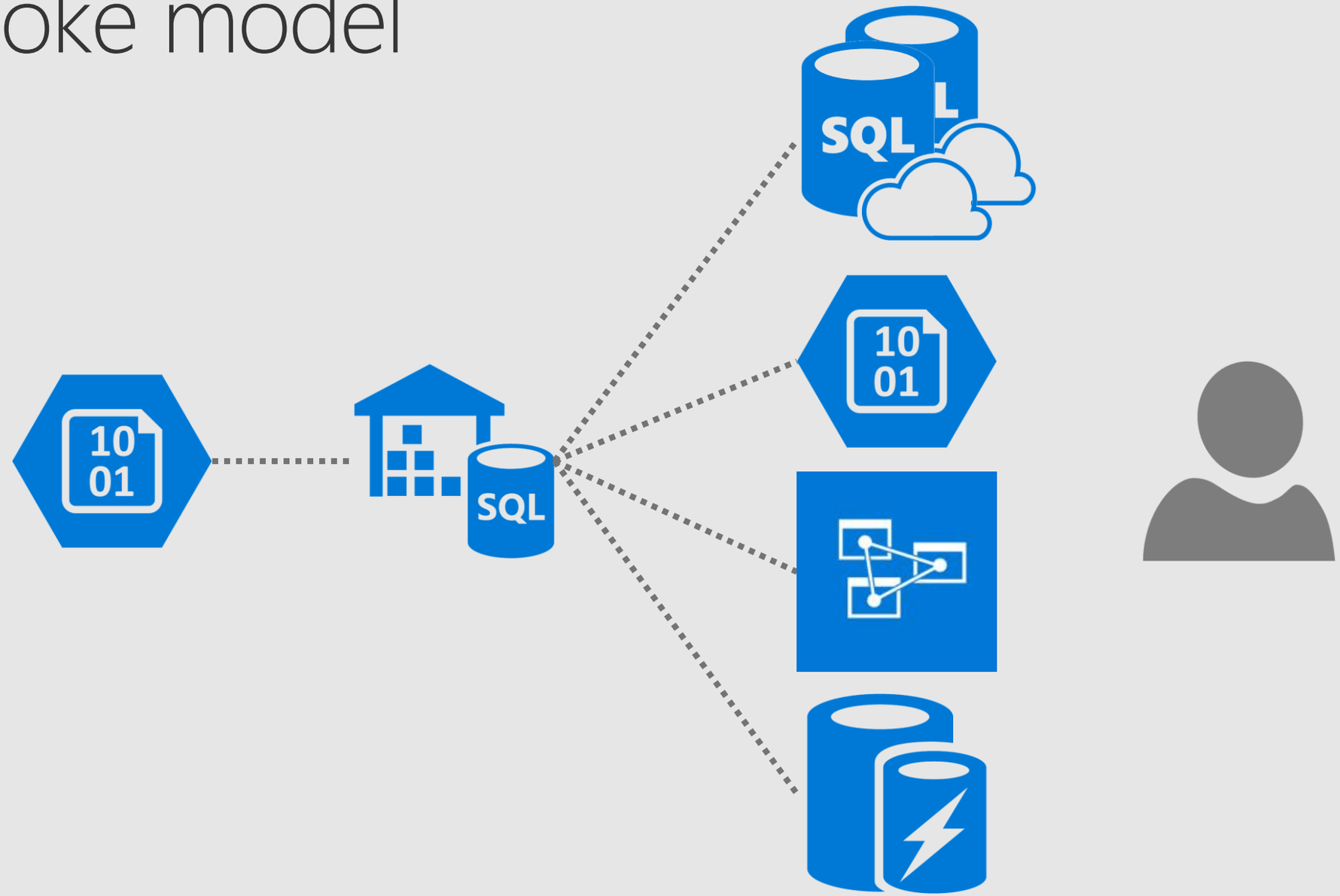
Square data only

Workload management

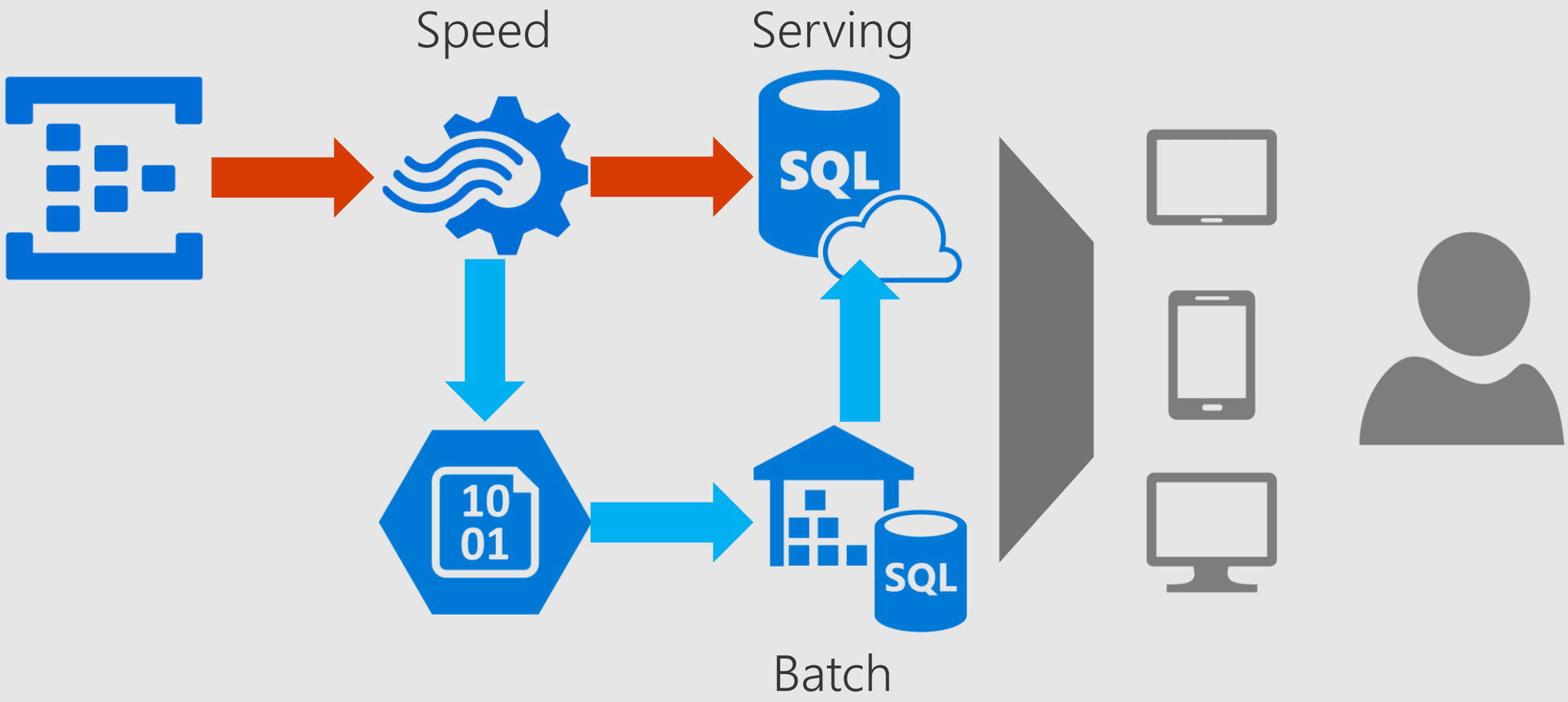
Offline re-sizing

Trickle loads

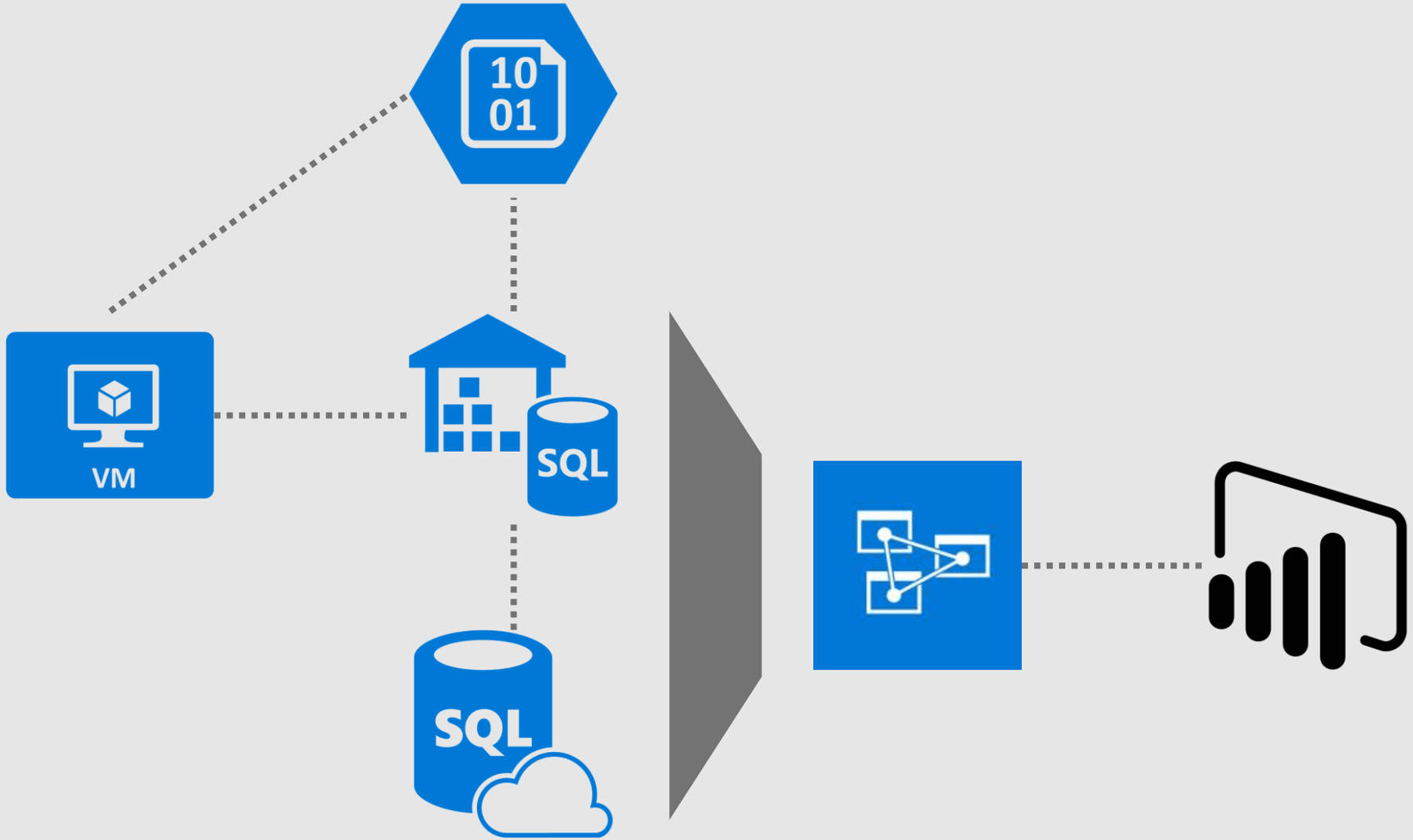
Hub & Spoke model



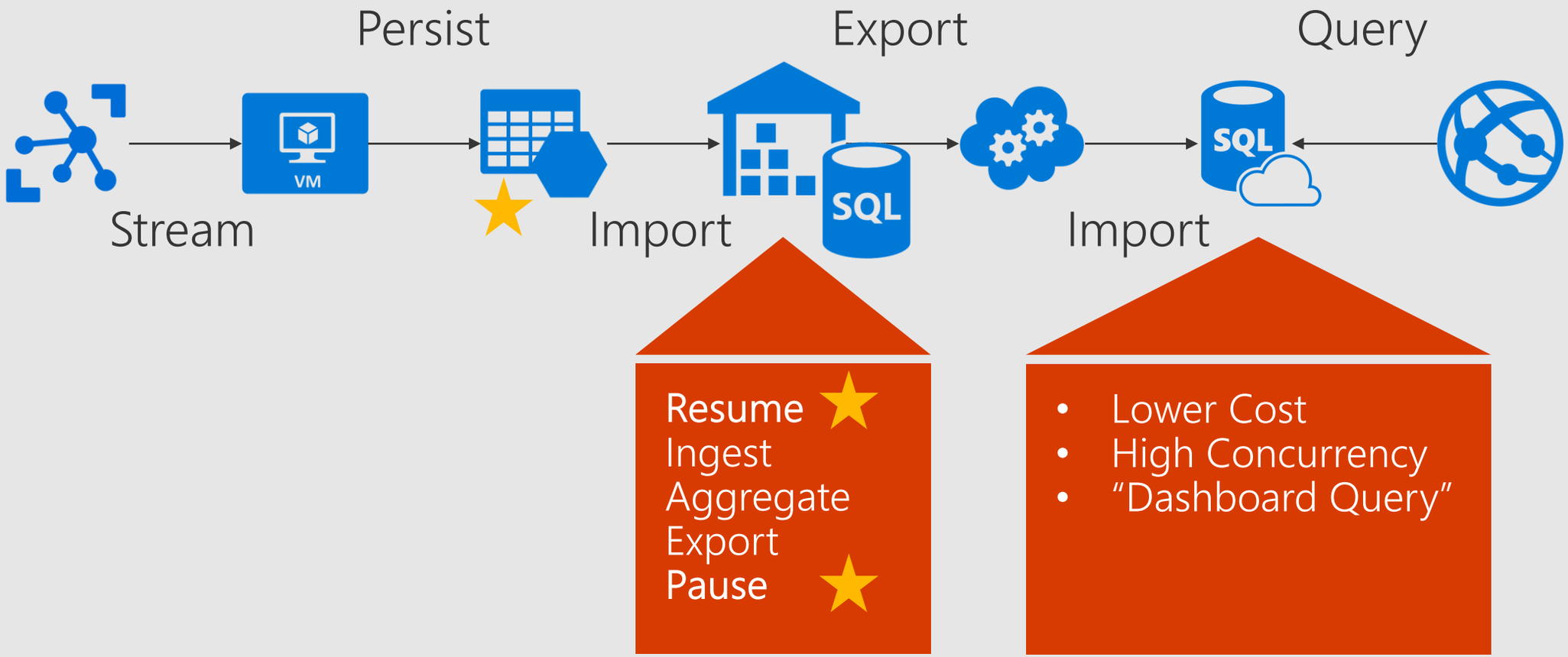
Lambda Architecture



SQL Server in the cloud

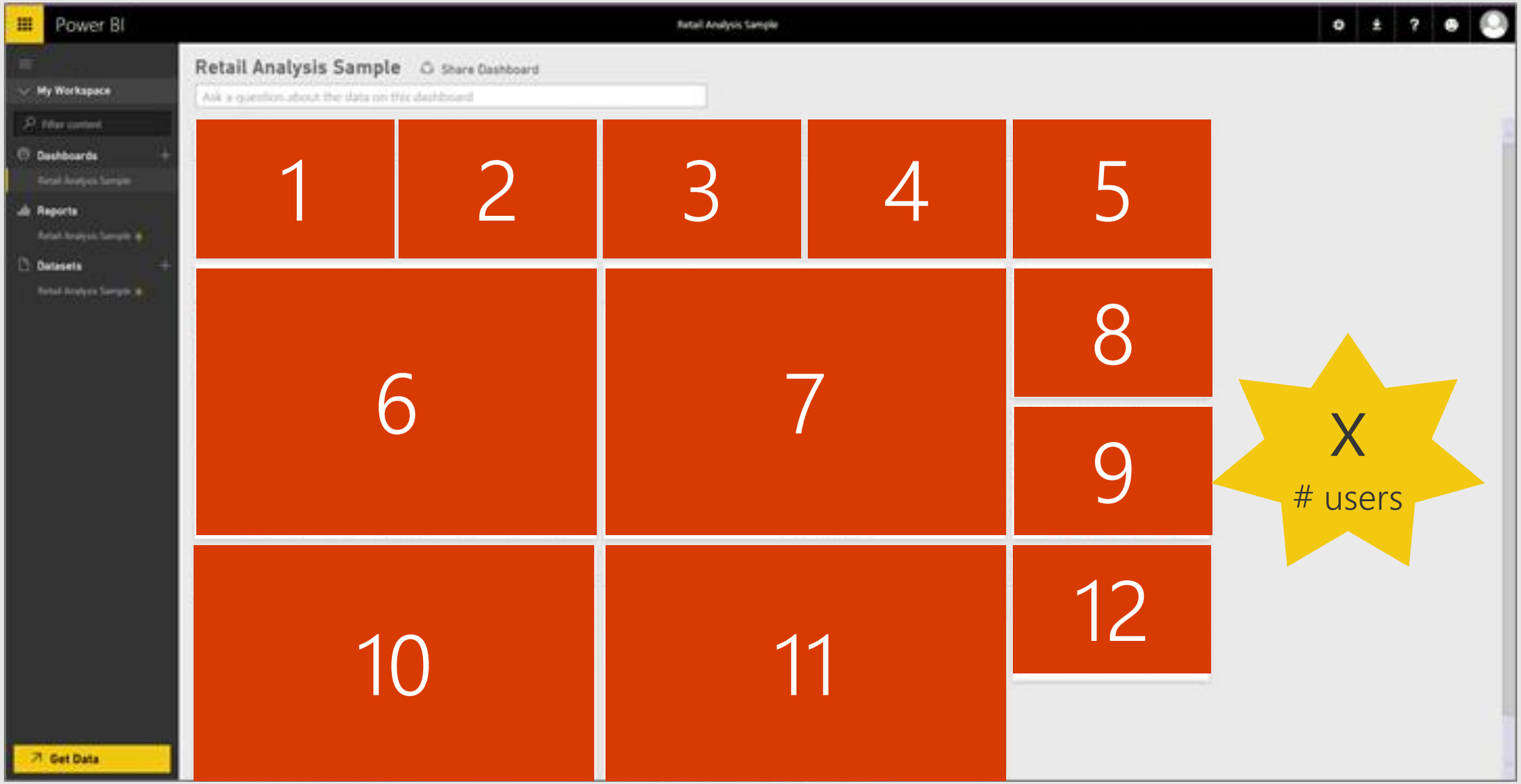


Customer scenario (Presence Orb)



Benefits of workload isolation

Dashboard query



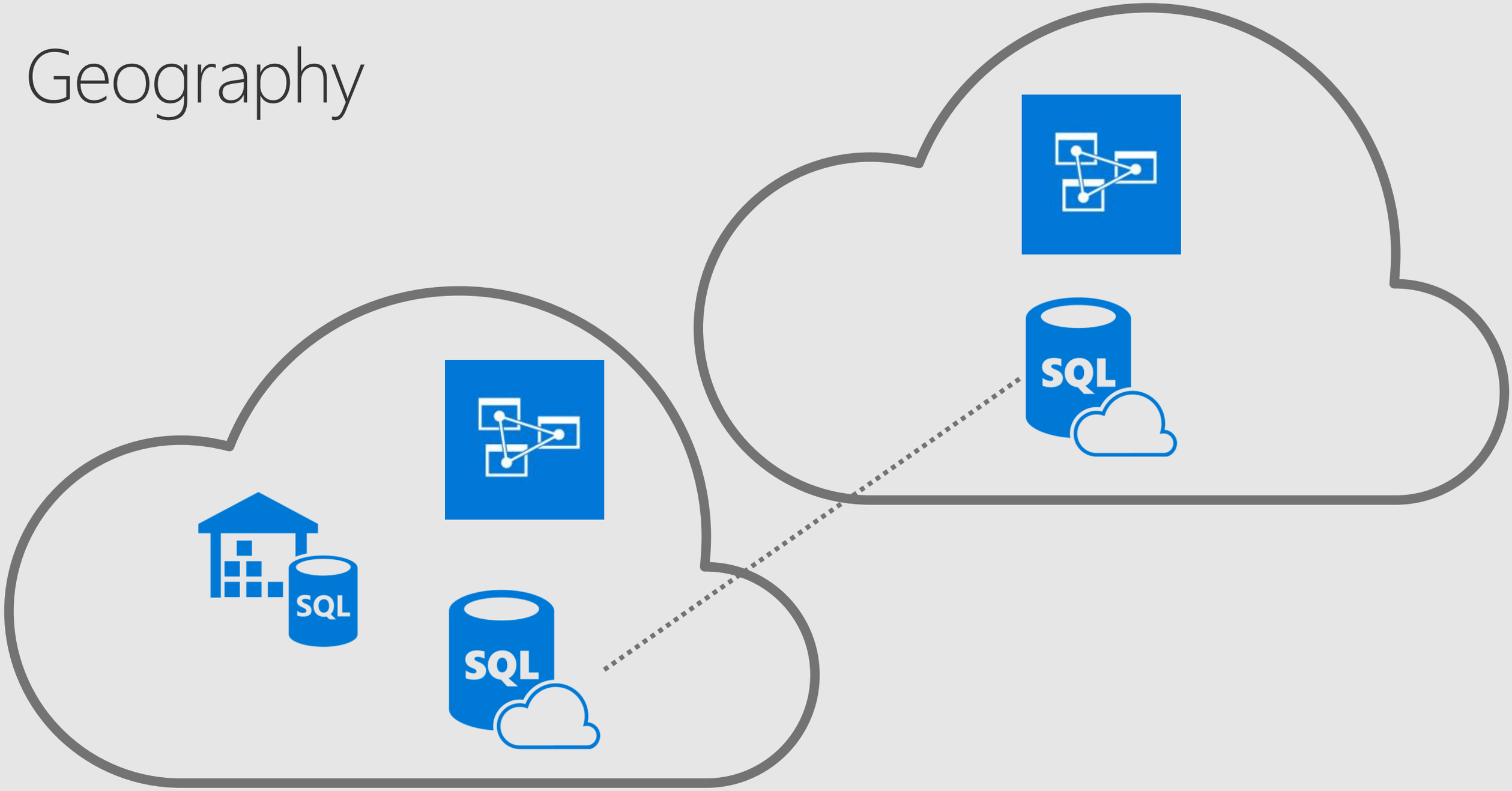
Performance

Sub-second response from cache

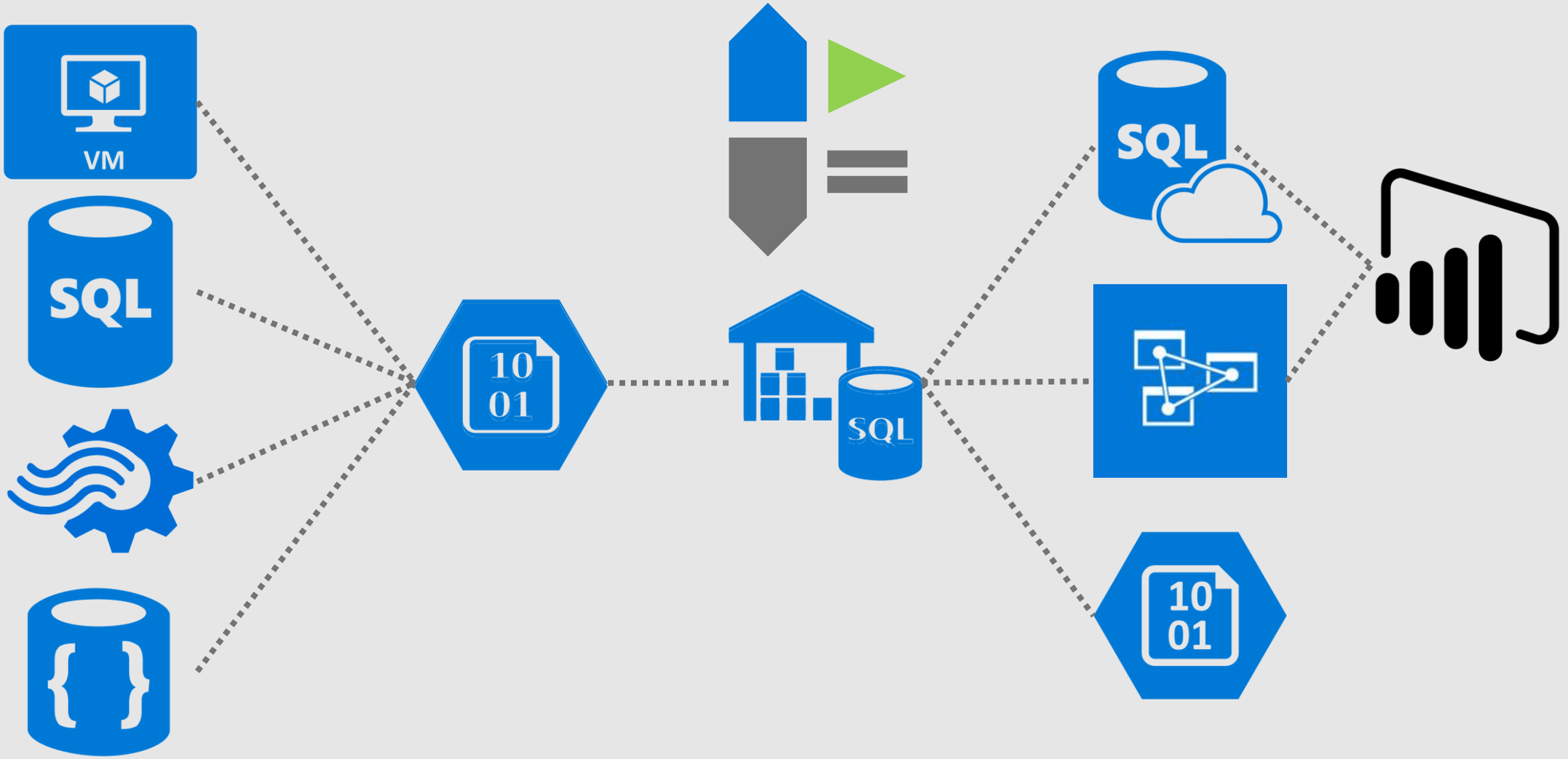
Predictable query performance

Support demanding BI interactive scenarios

Geography



Elastic Isolation



Evaluating spoke options

SQLDB use cases

Data marts

Concurrency offload

Operational reporting

Spatial models

Temporal data

SQLDB

Pro's

High concurrency

Low latency ingestion

Full T-SQL support

Operational reporting

Geo-redundancy

Familiarity

Con's

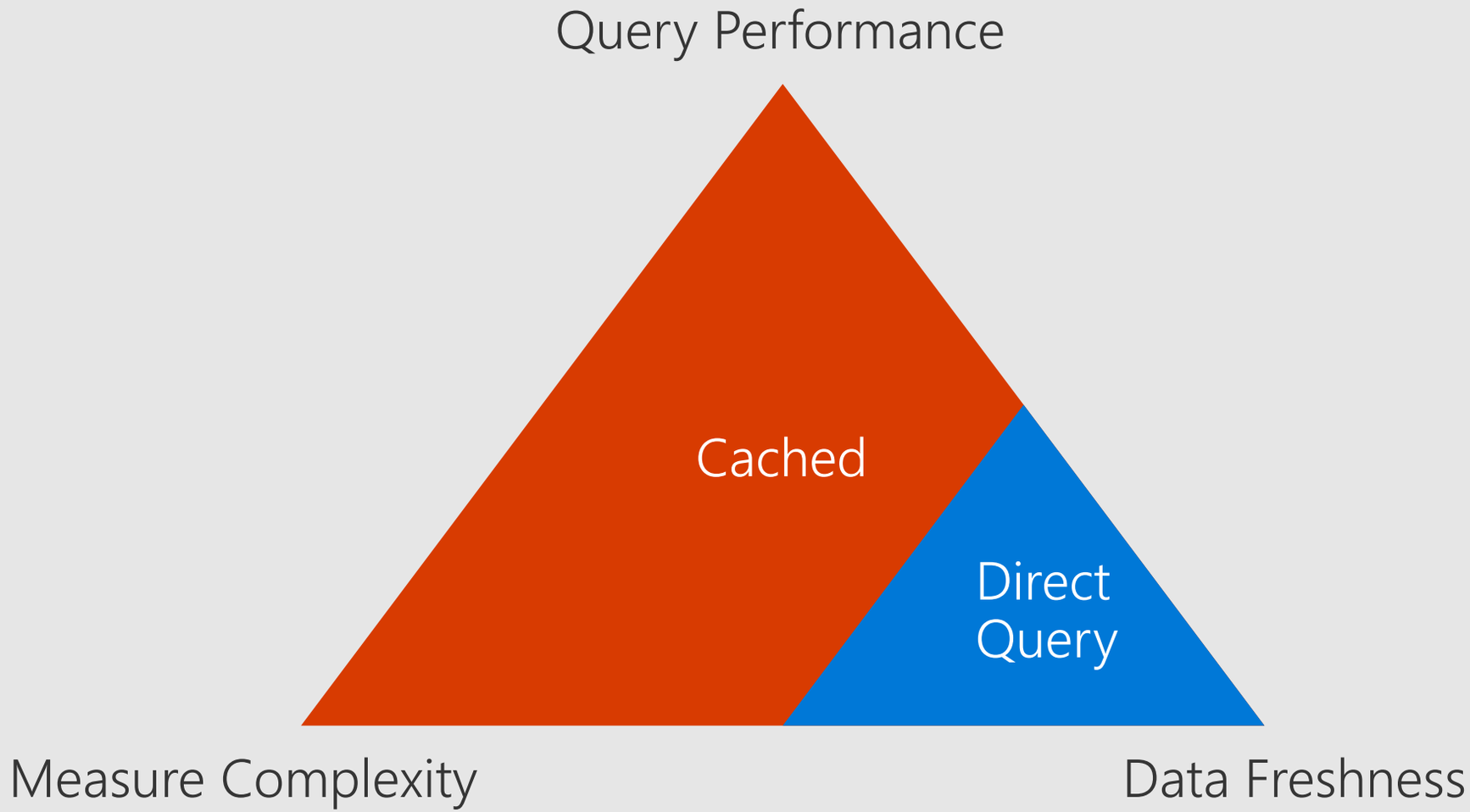
Max 4TB storage (size on disk)

Semantic model: views

No direct writeback to SQLDW

Reduced freshness on data

Analysis Services use cases



Azure Analysis Services - Cached

Pro's

Query performance

Measure complexity

Full cache

Enables SQLDW elastic scale

Con's

Reduced freshness on data

Data must fit in memory

Read only (no writeback)

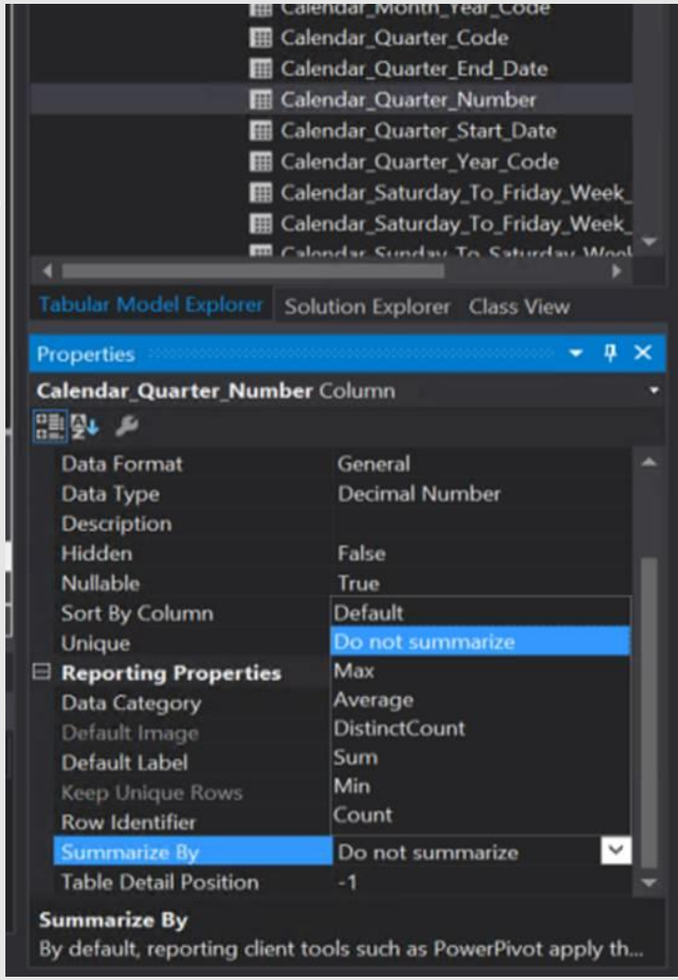
Performance tips

Column with numeric data types

- Always treated as a measure even if they aren't e.g. Year
- Set implicit measures to "Do Not Summarize"

Can also be set in Power BI

- Better to do it at the model level



Azure Analysis Services – Direct Query

Pro's

Queries source data directly

Optimal data freshness

Query at base fact level

Con's

Inconsistent query performance

Requires additional in-db design

Consumes concurrency slots

No caching

Requires SQLDW to be online

May require higher DWU

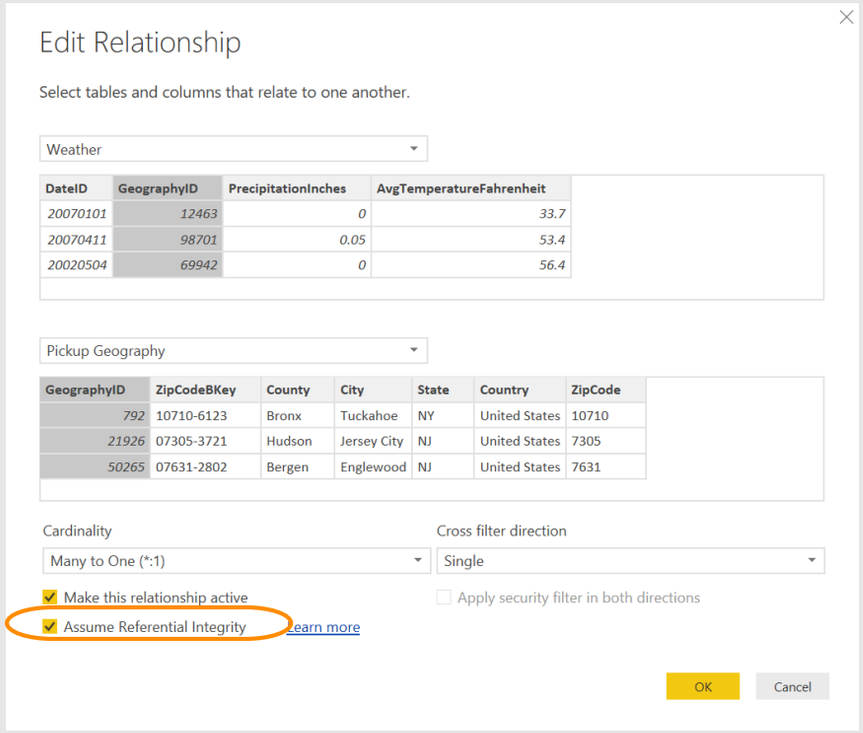
Direct Query Performance Tips

Inner Joins

Power BI : Assume Referential Integrity

Analysis Services : relyOnReferentialIntegrity: true

Power BI



Analysis Services

[*Direct Query Whitepaper for Analysis Services]

SQL VM use cases

Analysis Services

Existing multi-dimensional

Exceed Azure AS cache size

Writeback

SQL Server

>4TB of data (size on disk)

3 part names

Using SSIS

Using SQL Agent

Summary

Summary

Use AAS Cache

- Absolute performance
- Pre-aggregated data
- Complex measures

Use SQLDB

- For operational reporting
- For small data marts

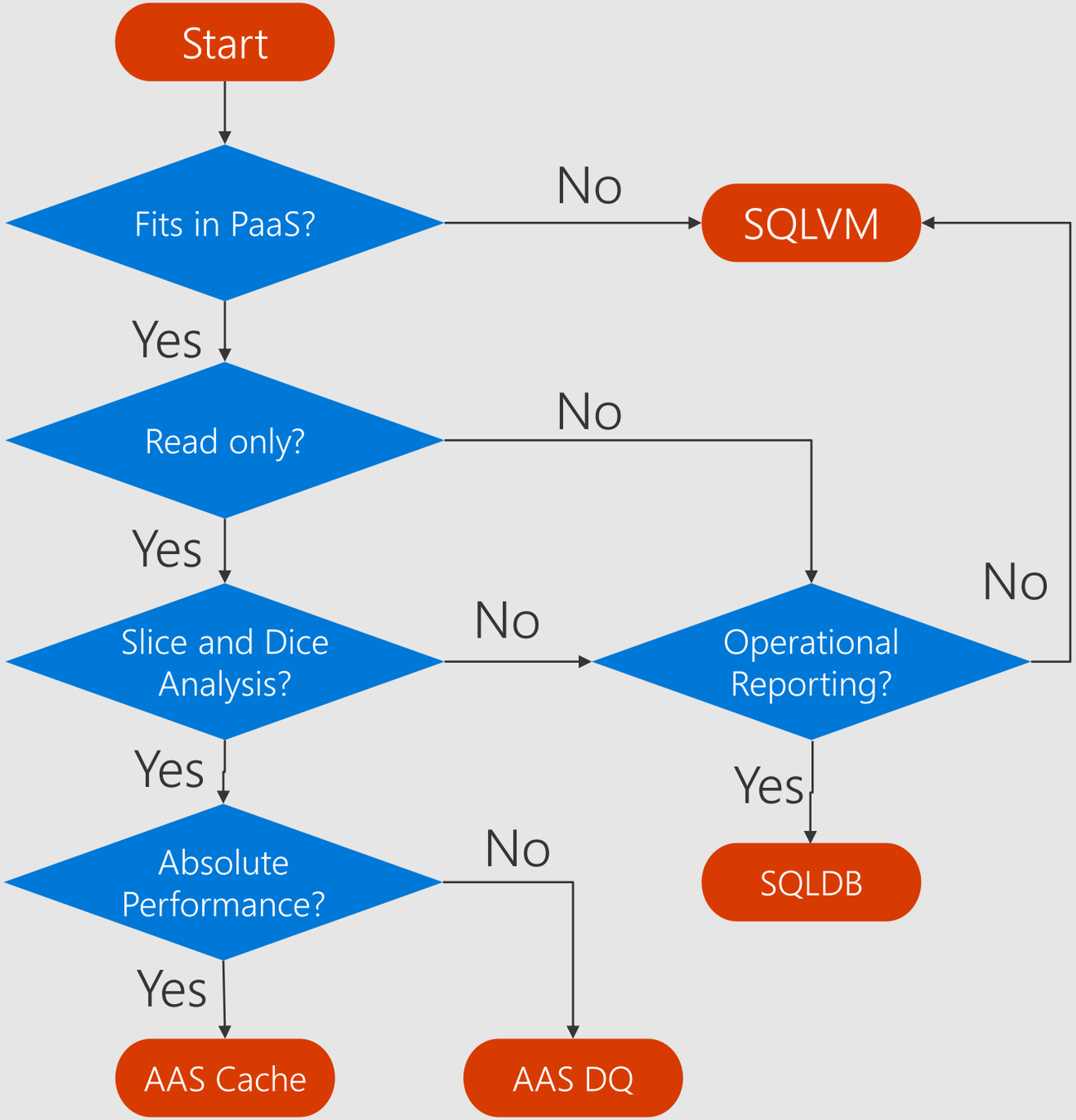
Use AAS DQ

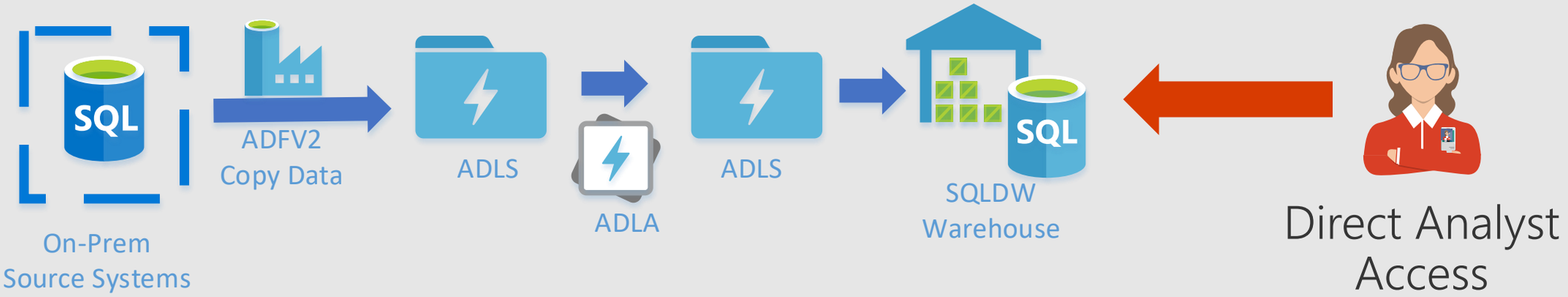
- Simple measures
- Querying latest data
- Base fact analysis

Use SQLVM

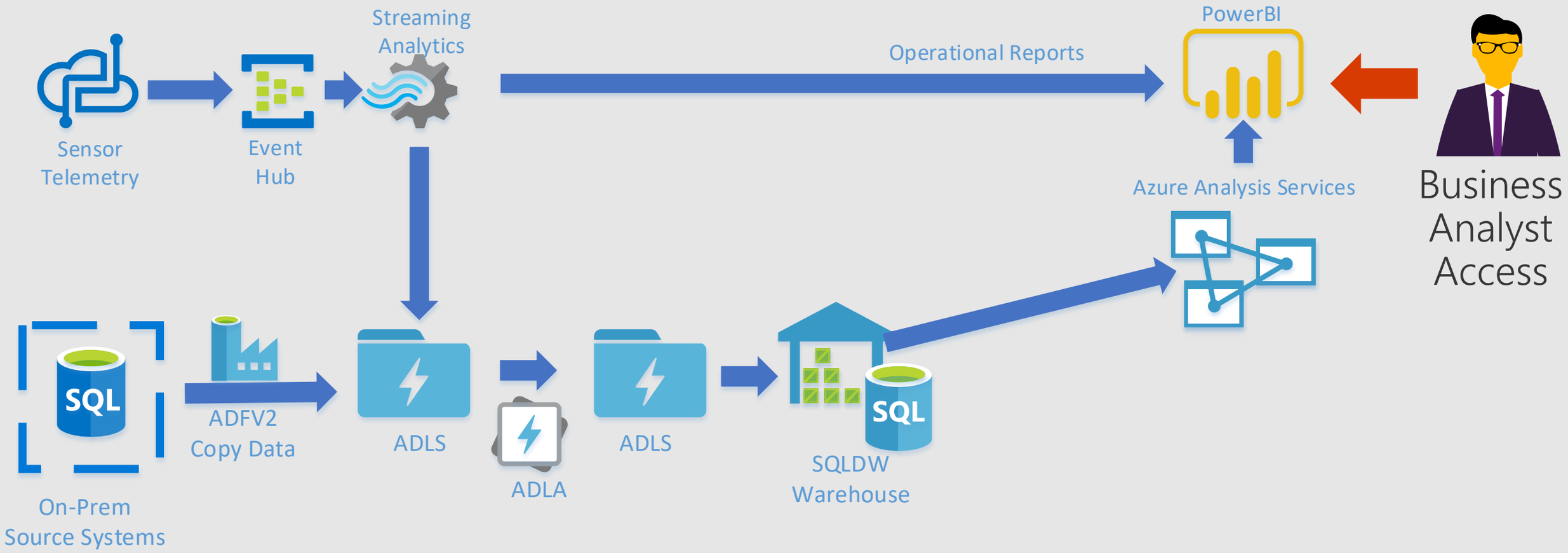
- Migrating MD models
- Leveraging existing VMs

Decision tree

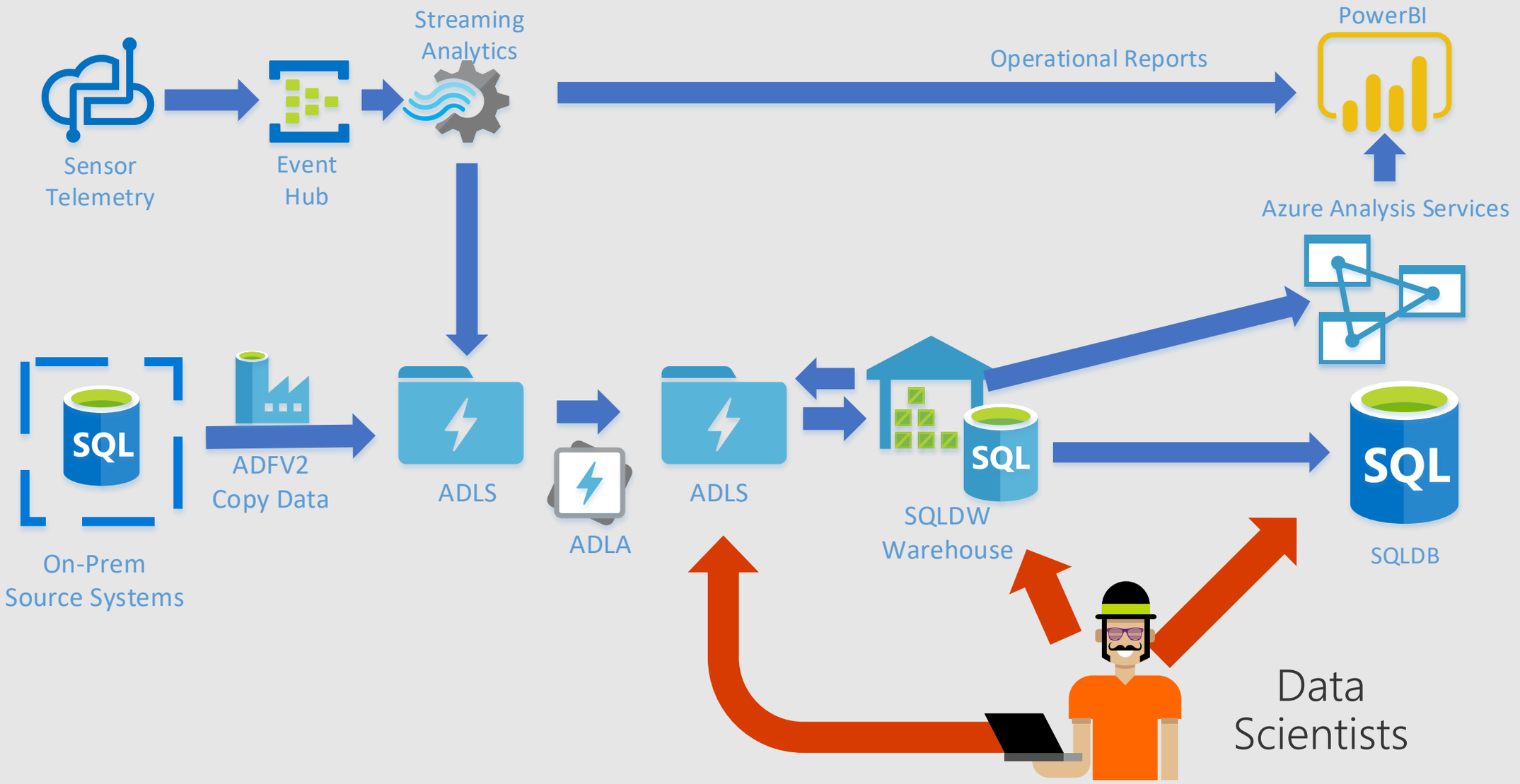




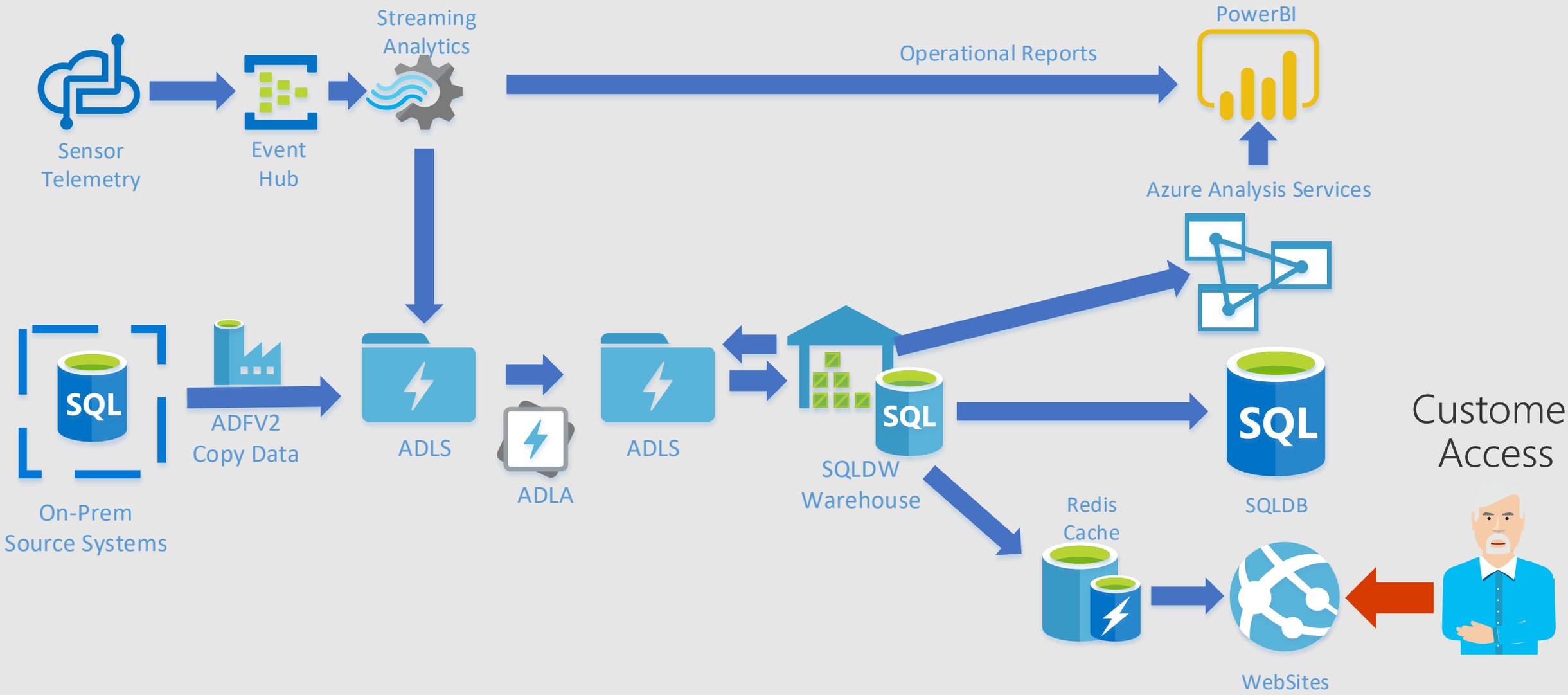
The Modern Warehouse



The Modern Warehouse



The Modern Warehouse



The Modern Warehouse