Azure SQL Data Warehouse
Azure Analysis Services

Nakul Joshi
Cloud Solution Architect – OCP
Microsoft Singapore

Azure SQL Data Warehouse

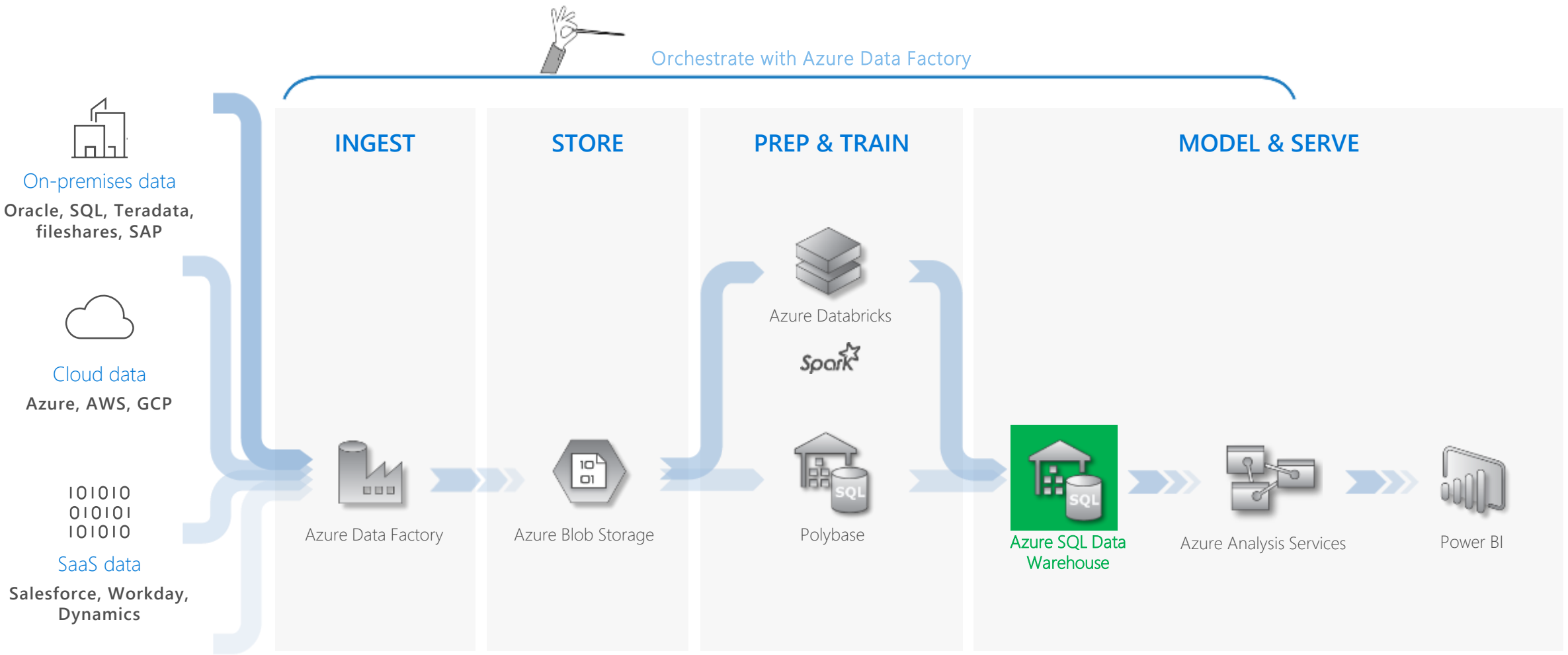
Changes in Enterprise Data Warehouse space

Organizations are changing with increasing demand to:

- Integrate with new or unstructured data
- Drive to the cloud
- Reduce or remove hardware renewal
- Reduction in support costs

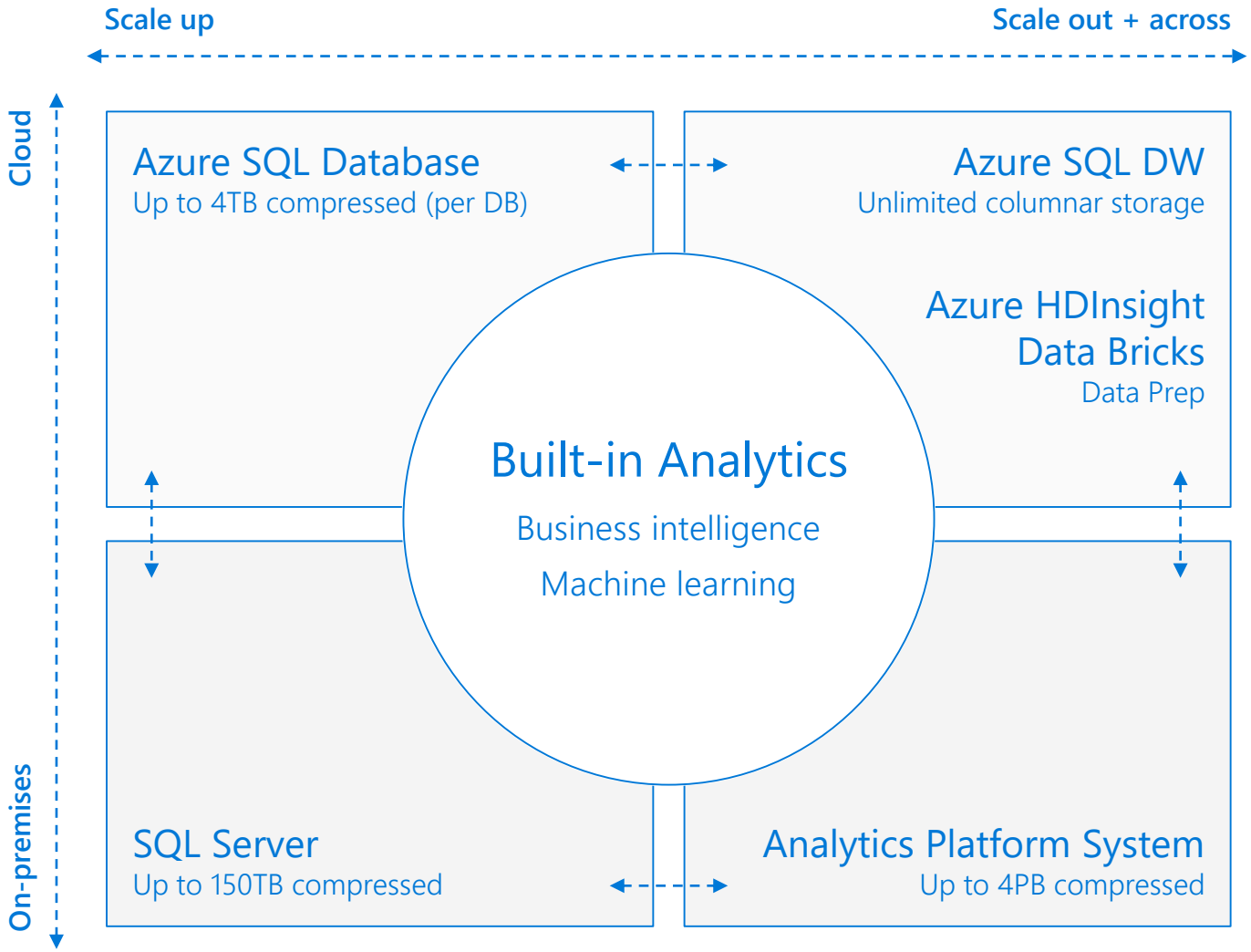
AZURE DATA WAREHOUSE

Modernize your enterprise data warehouse at scale



Microsoft Azure also supports other **Big Data** services like **Azure HDInsight**, **Azure SQL Database** and **Azure Data Lake** to allow customers to tailor the above architecture to meet their unique needs.

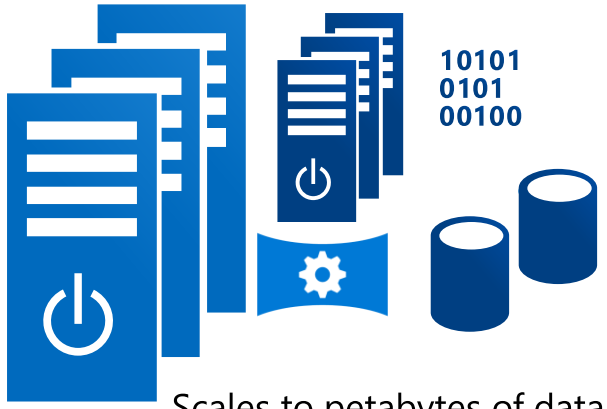
WHEN TO USE WHAT



Introducing Azure SQL Data Warehouse

A relational **platform-as-a-service**, fully managed by Microsoft.
Elastic scale cloud **data warehouse** with **proven** SQL Server capabilities.
Built for businesses of all **shapes, sizes,** and **industry**.

Elastic scale & performance



Scales to petabytes of data

Massively Parallel Processing

Instant-on compute scales in seconds

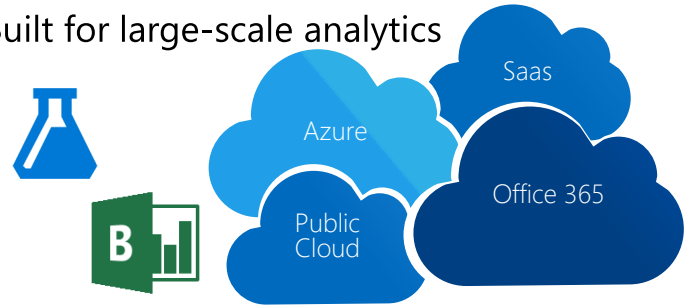
Query Relational and Non-Relational data

Relational batch processing

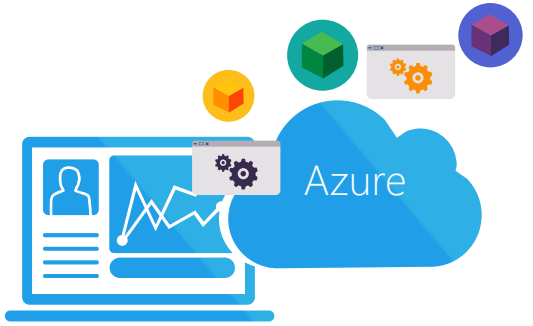
Query large datasets in minutes

Full hub-and-spoke support

Built for large-scale analytics



Market Leading Price & Performance

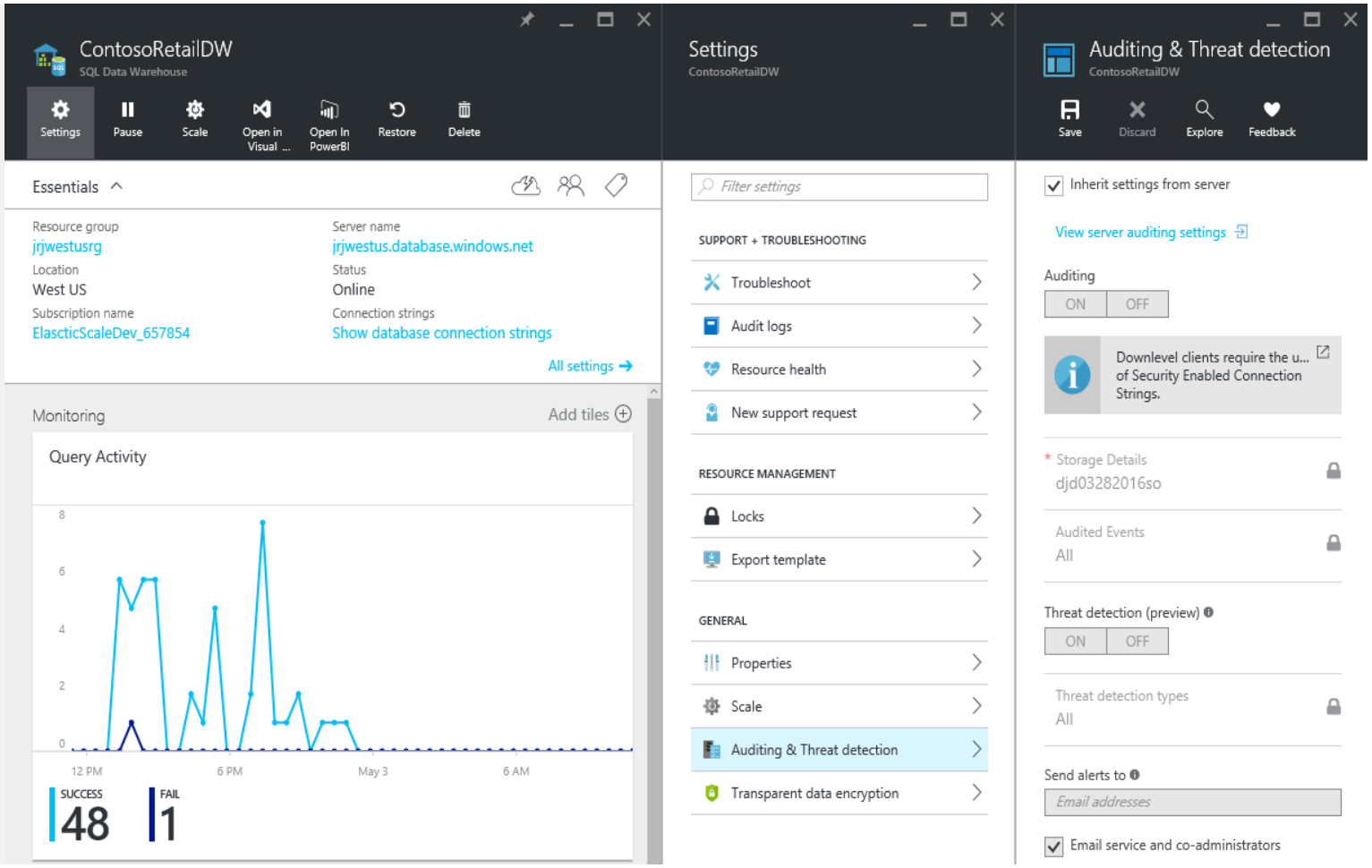


Simple billing compute & storage

Pay for what you need, when you need it with dynamic pause

A fully managed Platform-as-a-Service

- Azure cloud data warehouse service
- Elastic scale
- Separate storage and compute
- Use existing tools and skills
- Deploy and use in minutes!



AZURE SQL DATA WAREHOUSE - PERFORMANCE TIERS

Optimized for elasticity

Elastic-scale performance tier provides high performance for regular workloads and analytics

NEW!

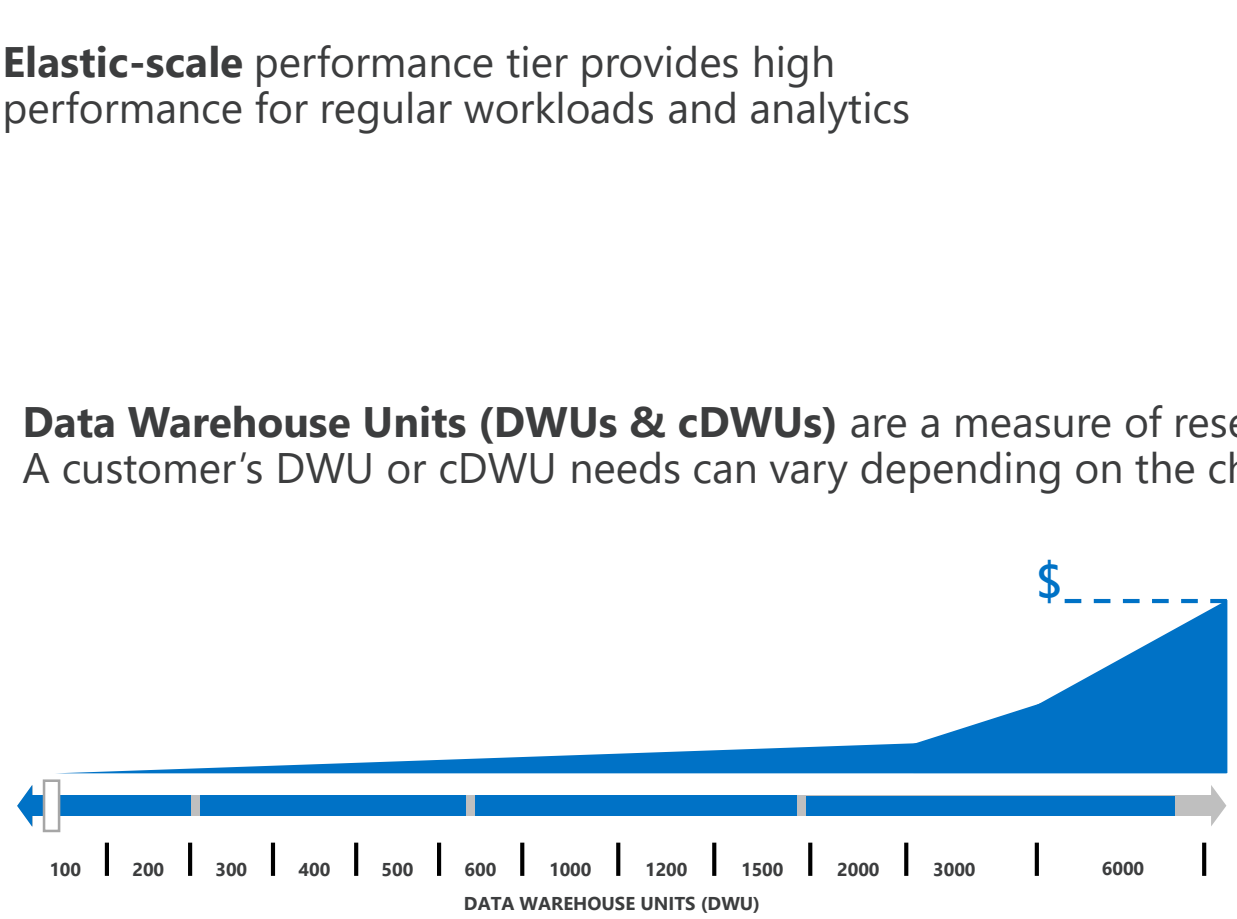
Optimized for compute

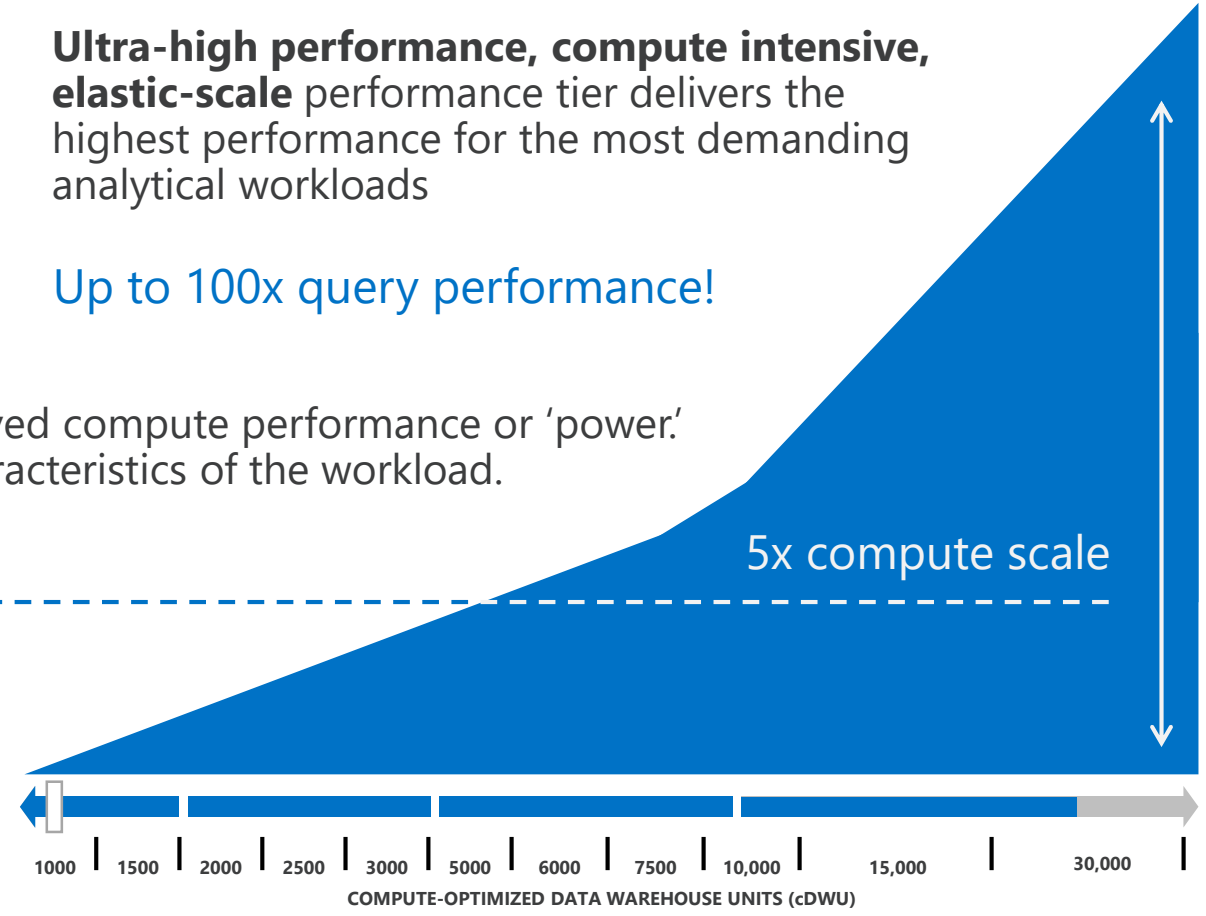
Preview available now!

Ultra-high performance, compute intensive, elastic-scale performance tier delivers the highest performance for the most demanding analytical workloads

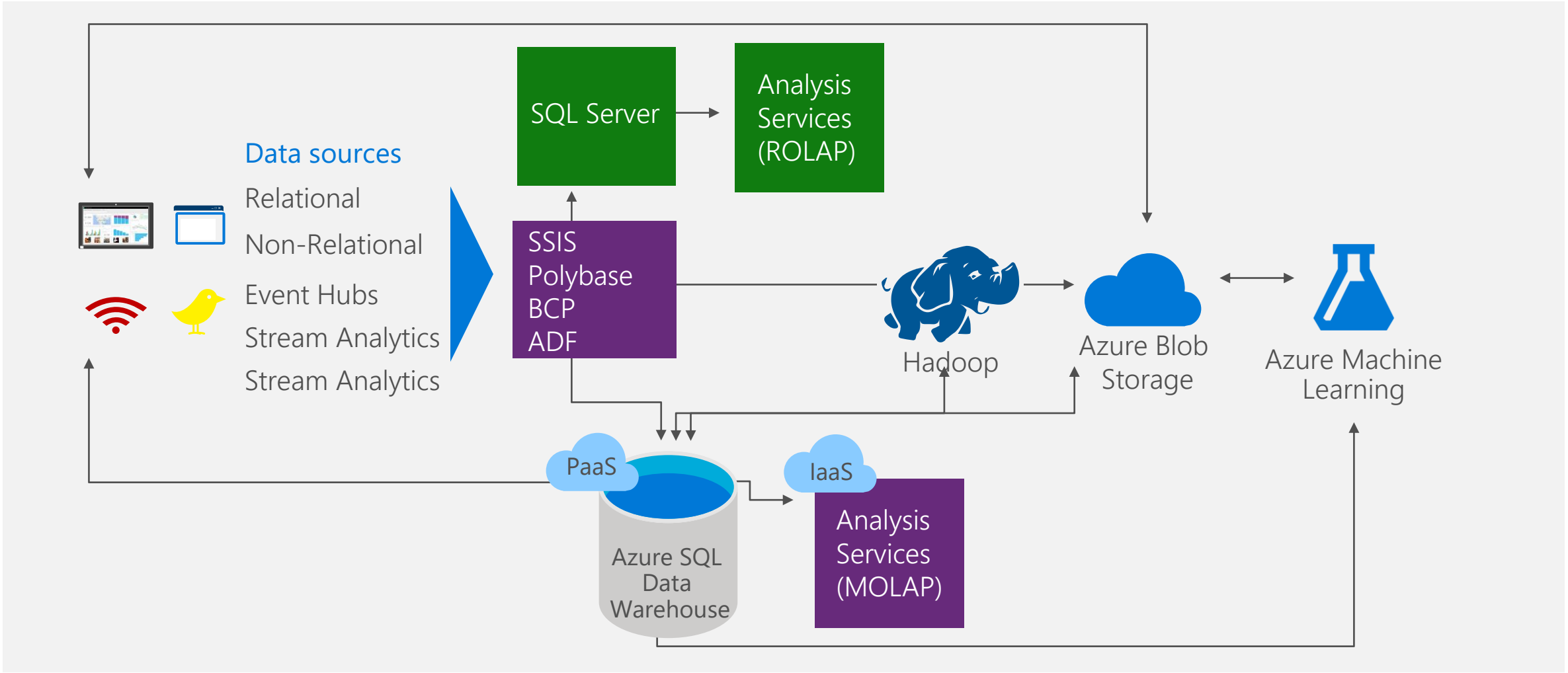
Up to 100x query performance!

Data Warehouse Units (DWUs & cDWUs) are a measure of reserved compute performance or 'power.' A customer's DWU or cDWU needs can vary depending on the characteristics of the workload.

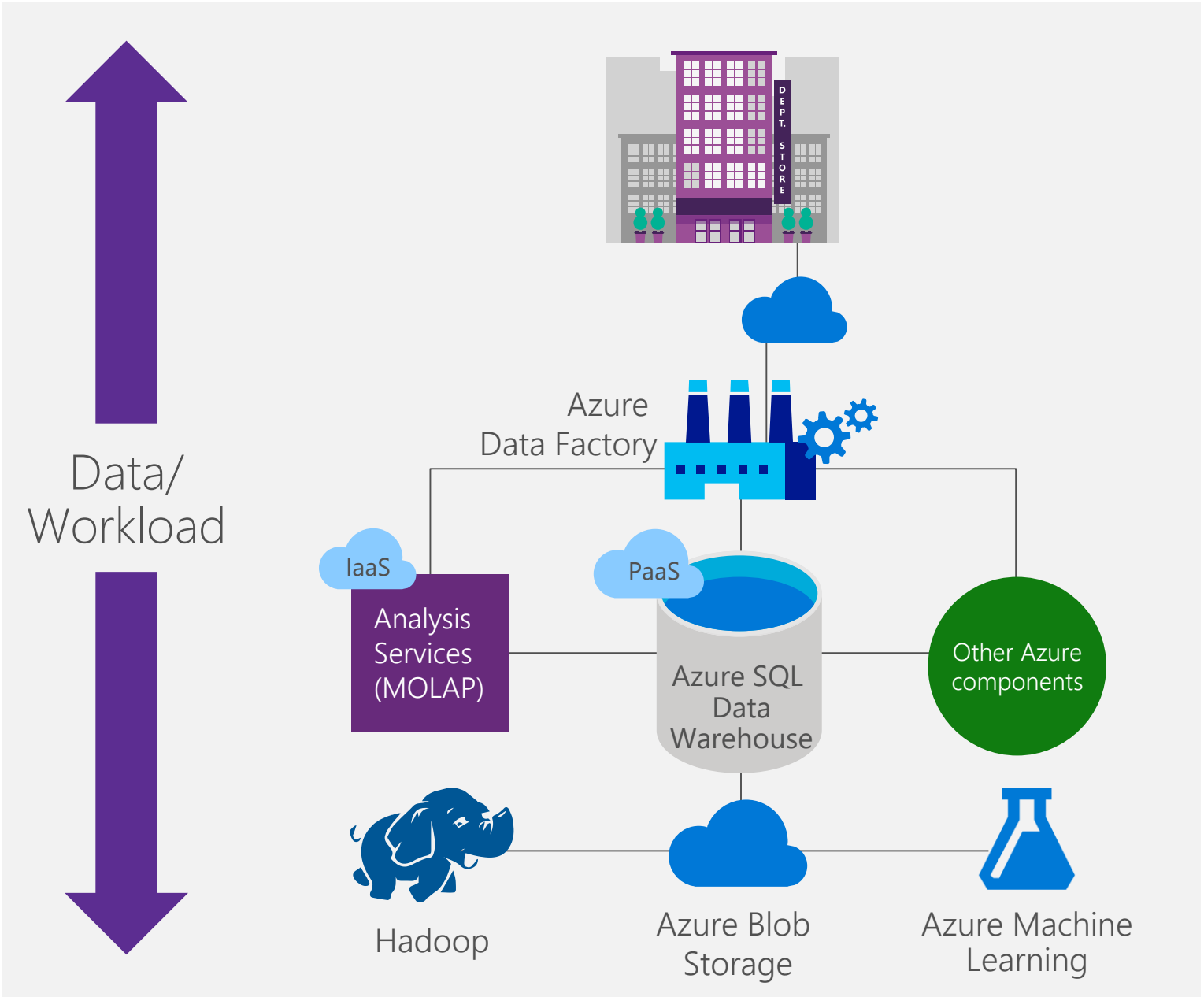




Integrates with existing processes

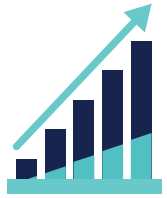


Supports data ingestion from literally anywhere…



Technical capabilities

Industry's **first** enterprise-class cloud data warehouse that can **grow, shrink, and pause** in seconds

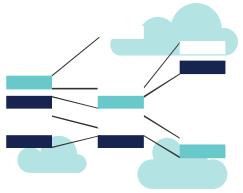


Petabyte scale data warehousing leveraging massive parallel processing

Full enterprise-class SQL Server experience

Two performance tiers designed for businesses of all sizes

Seamless compatibility with Power BI, Azure Machine Learning, HDInsight, and Azure Data Factory



Query and load big data from Hadoop, HDInsight, Data Lake and Blob Storage using Polybase

Patterns

SQL DW is good for analytical workloads. Why?

- ✓ Store large volumes of data.
- ✓ Consolidate disparate data into a single location.
- ✓ Shape, model, transform and aggregate data.
- ✓ Perform query analysis across large datasets.
- ✓ Ad-hoc reporting across large data volumes.
- ✓ All using simple SQL constructs.

"SQL on SQL"

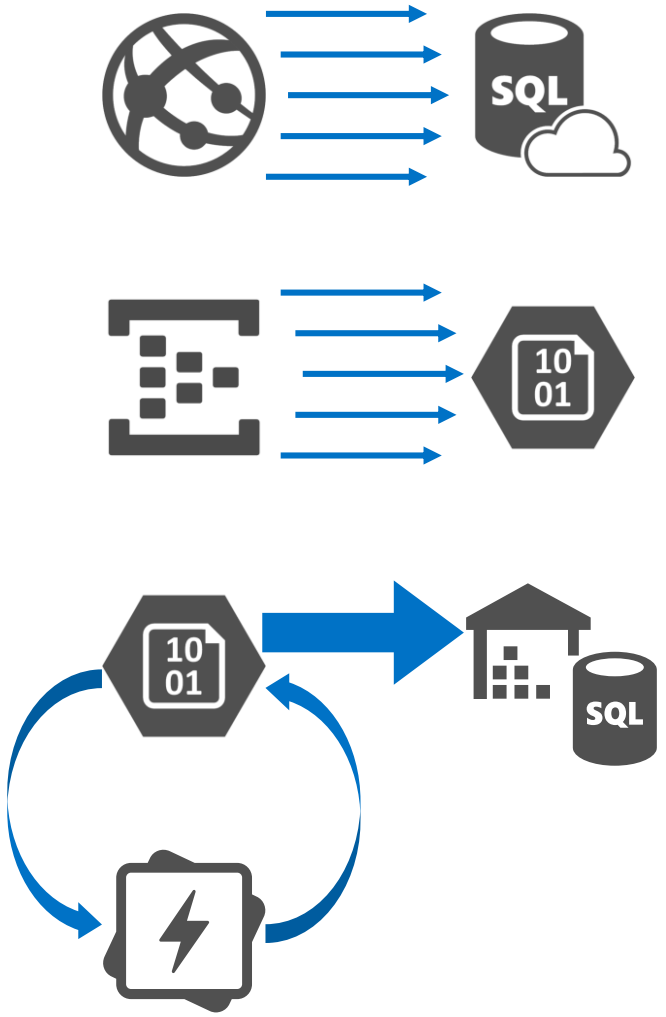
Unsuitable workloads for SQL DW

Operational workloads (OLTP)

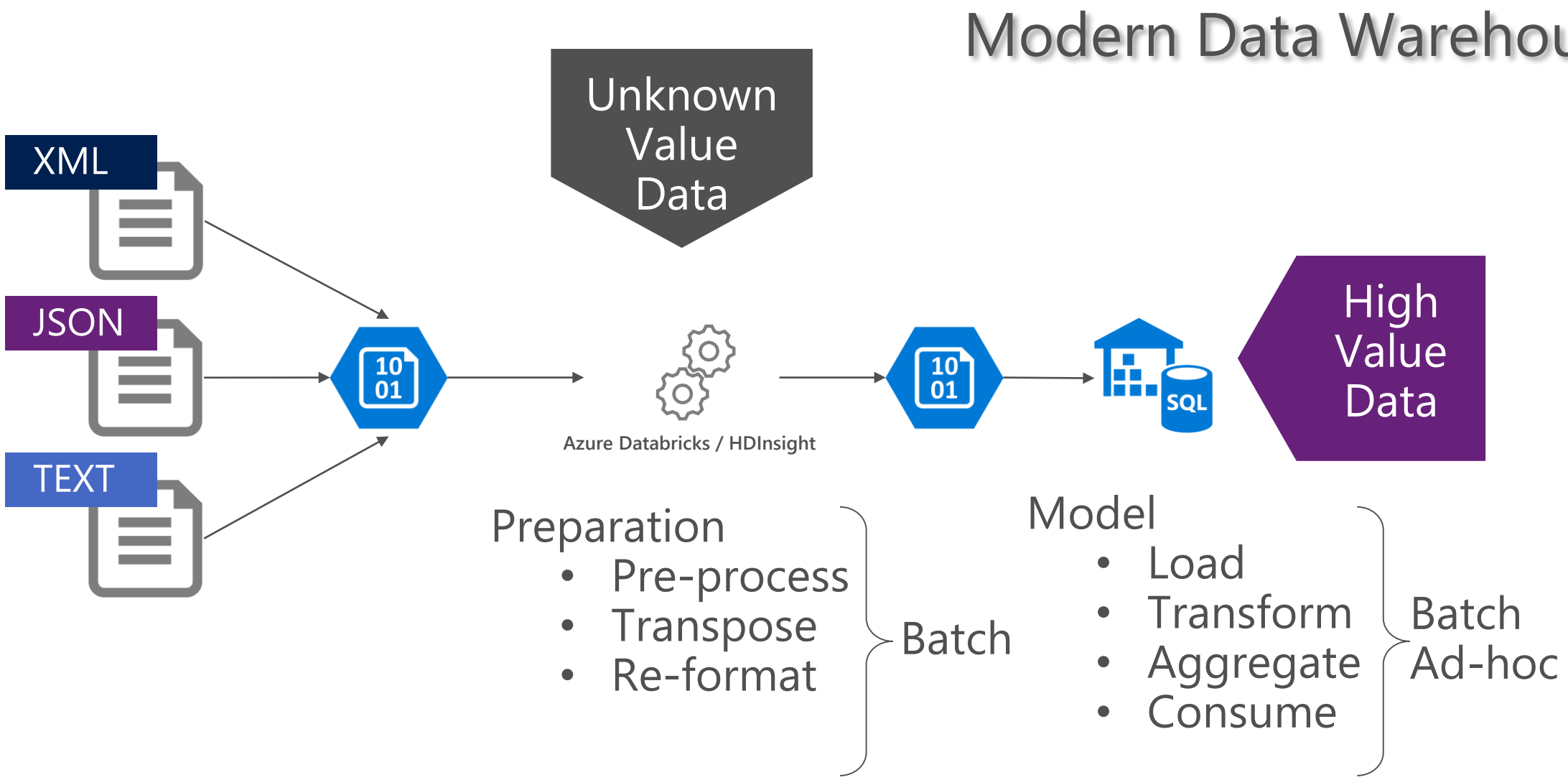
- High frequency reads and writes.
- Large numbers of singleton selects.
- High volumes of single row inserts.

Data Preparation

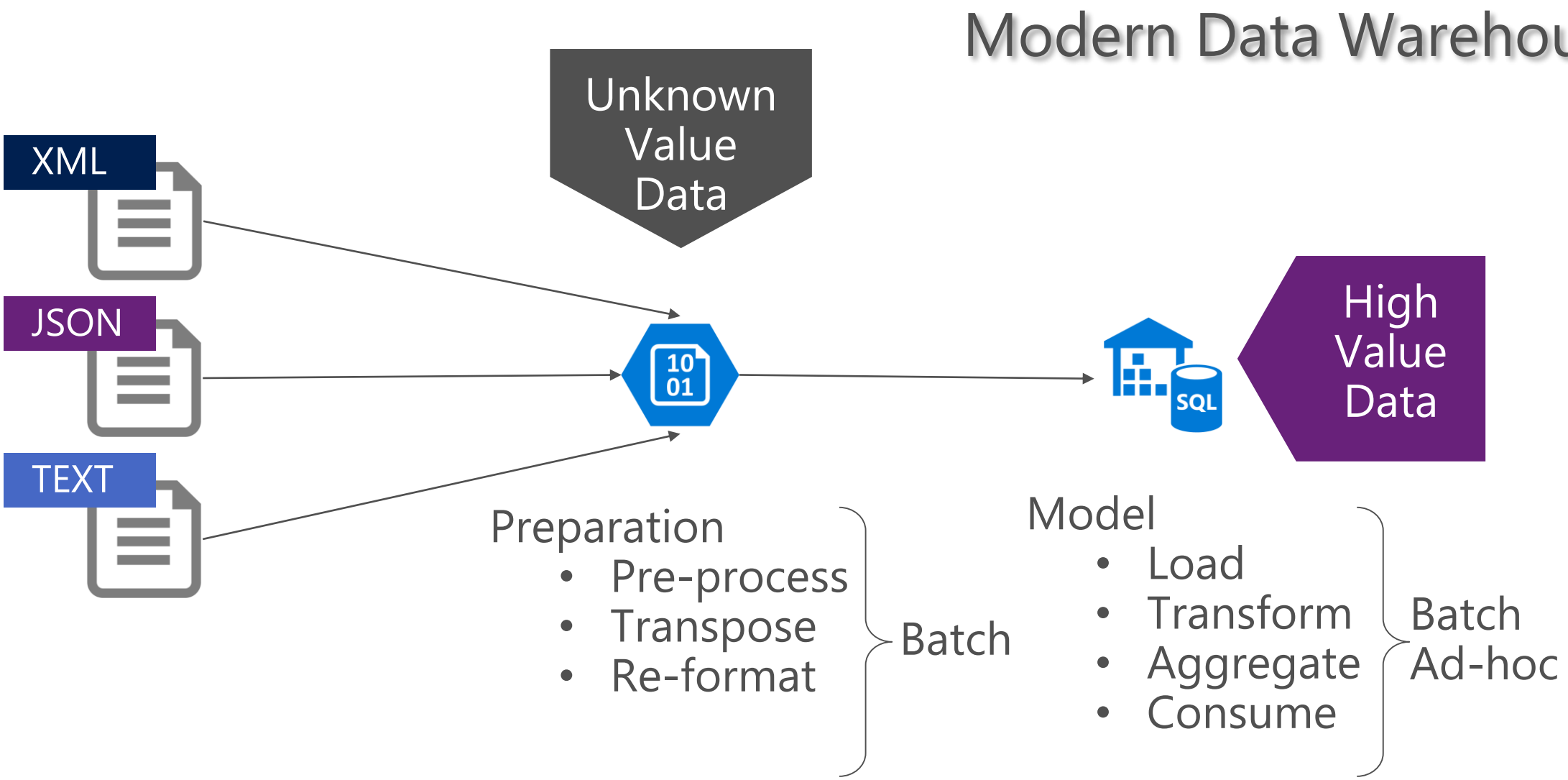
- Row by row processing needs.
- Incompatible formats (JSON, XML).



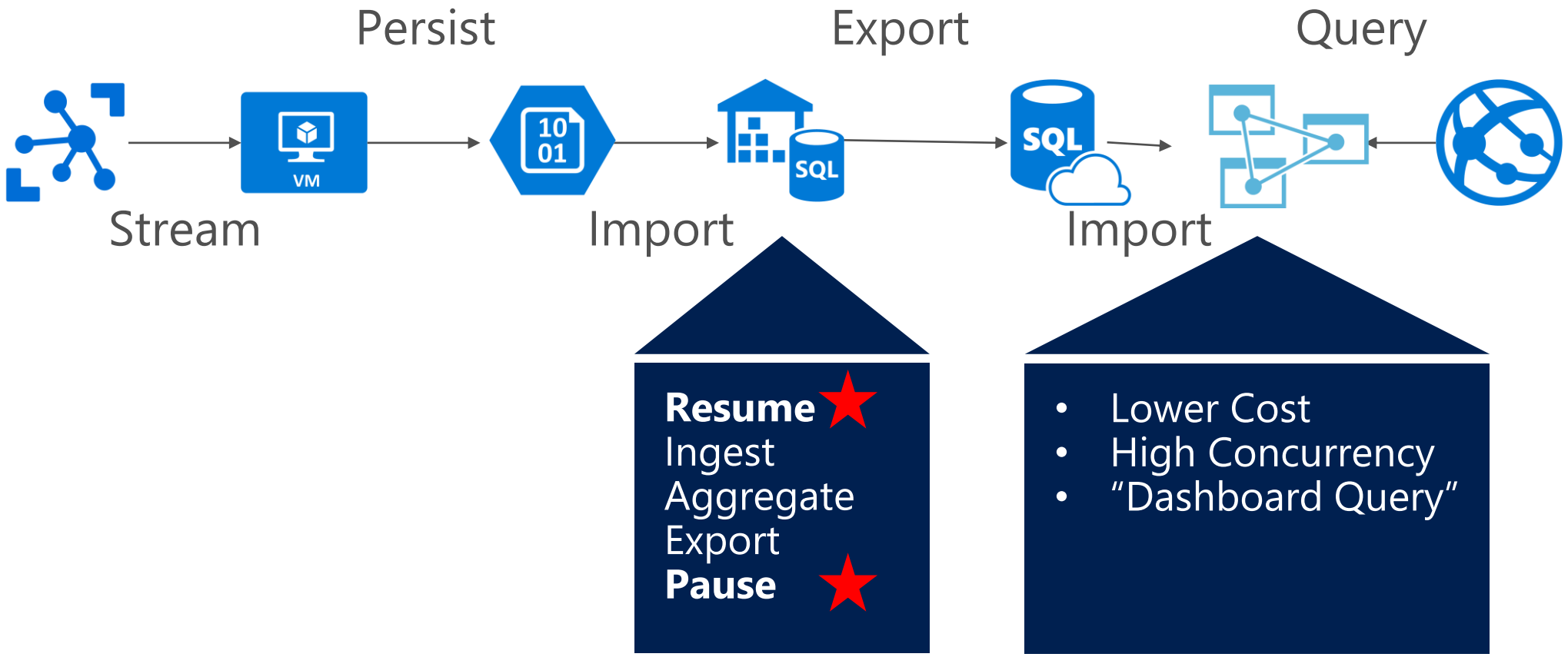
Azure Data Lake and Azure SQL Data Warehouse



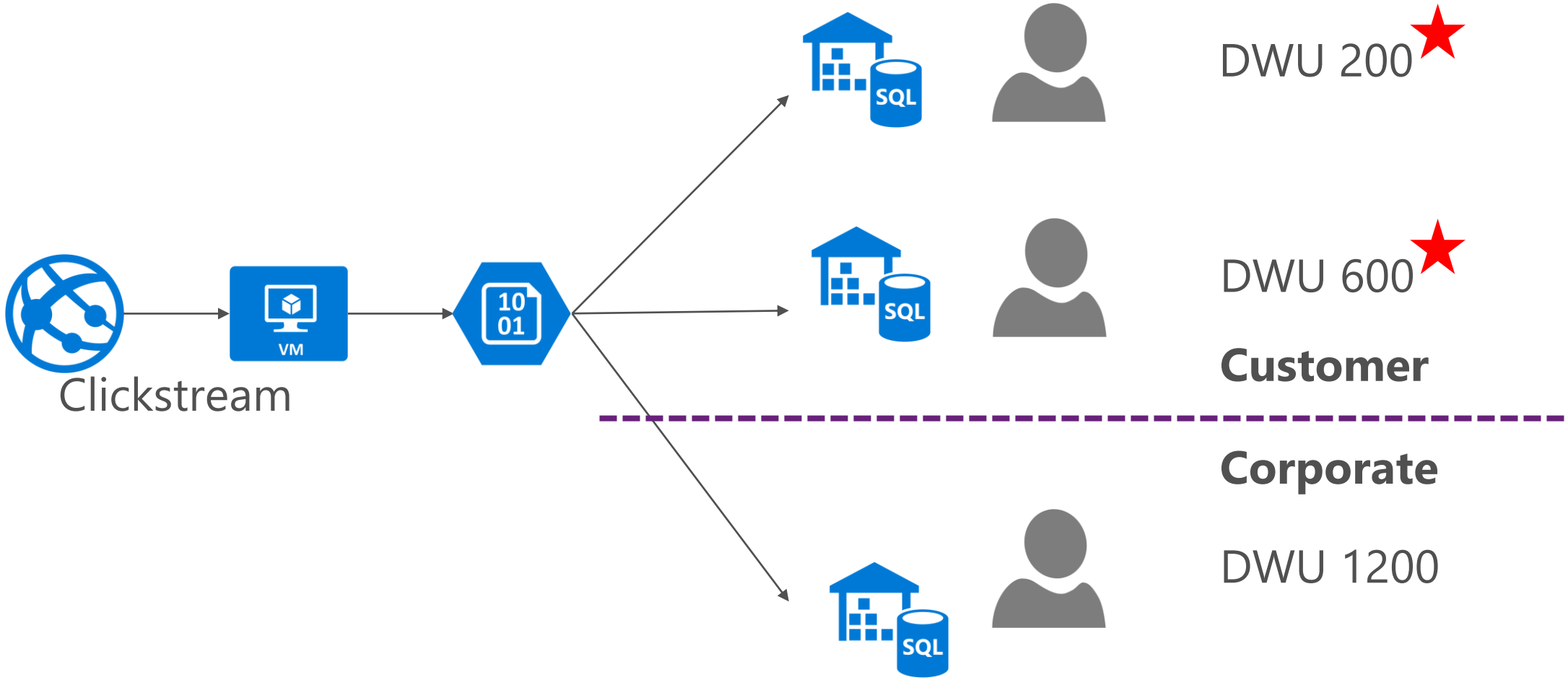
Azure Blob Storage and Azure SQL Data Warehouse



Pattern: Compute consumption

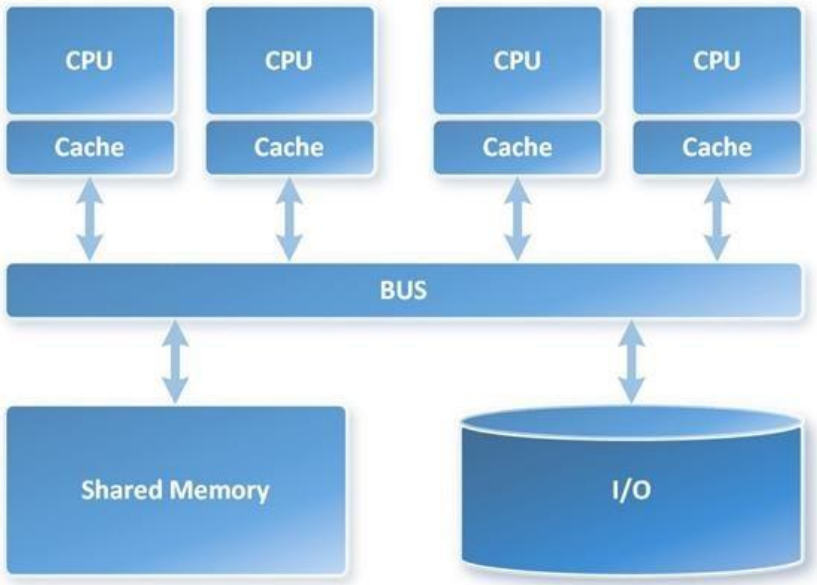


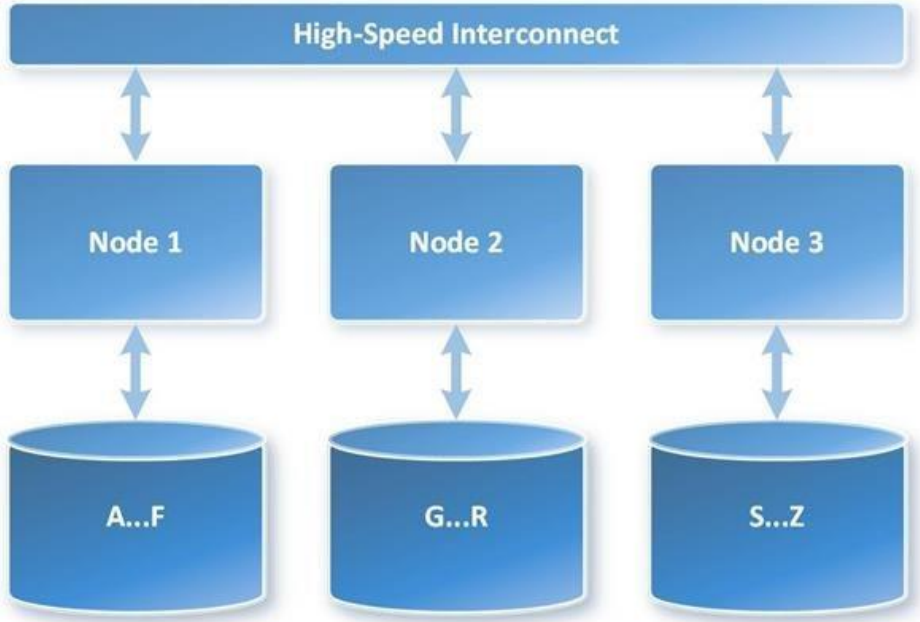
Pattern: SaaS customer isolation



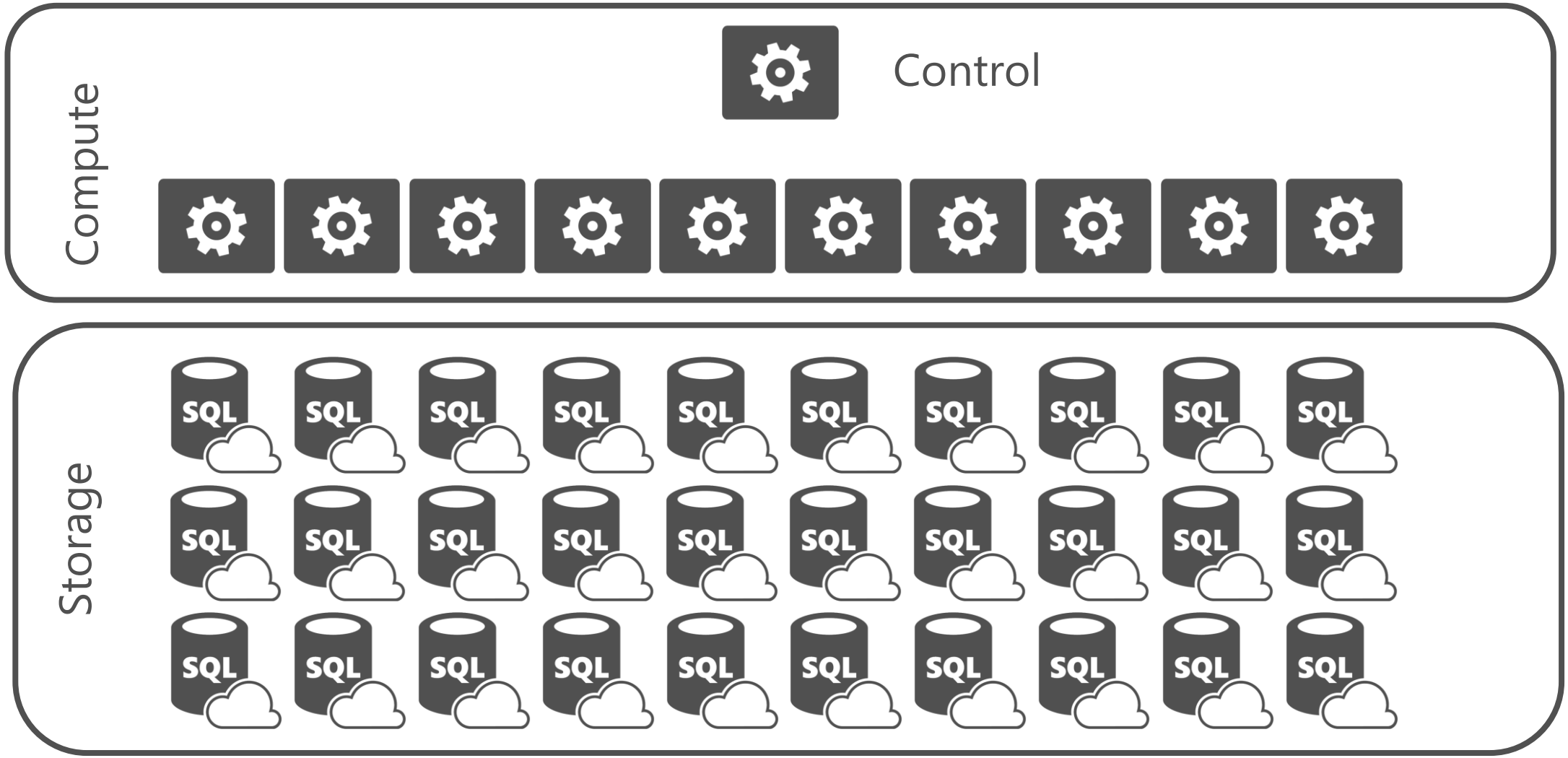
Architectural overview

SMP vs MPP


CPU
CPU
CPU
CPU
Cache
Cache
Cache
Cache
BUS
Shared Memory
I/O

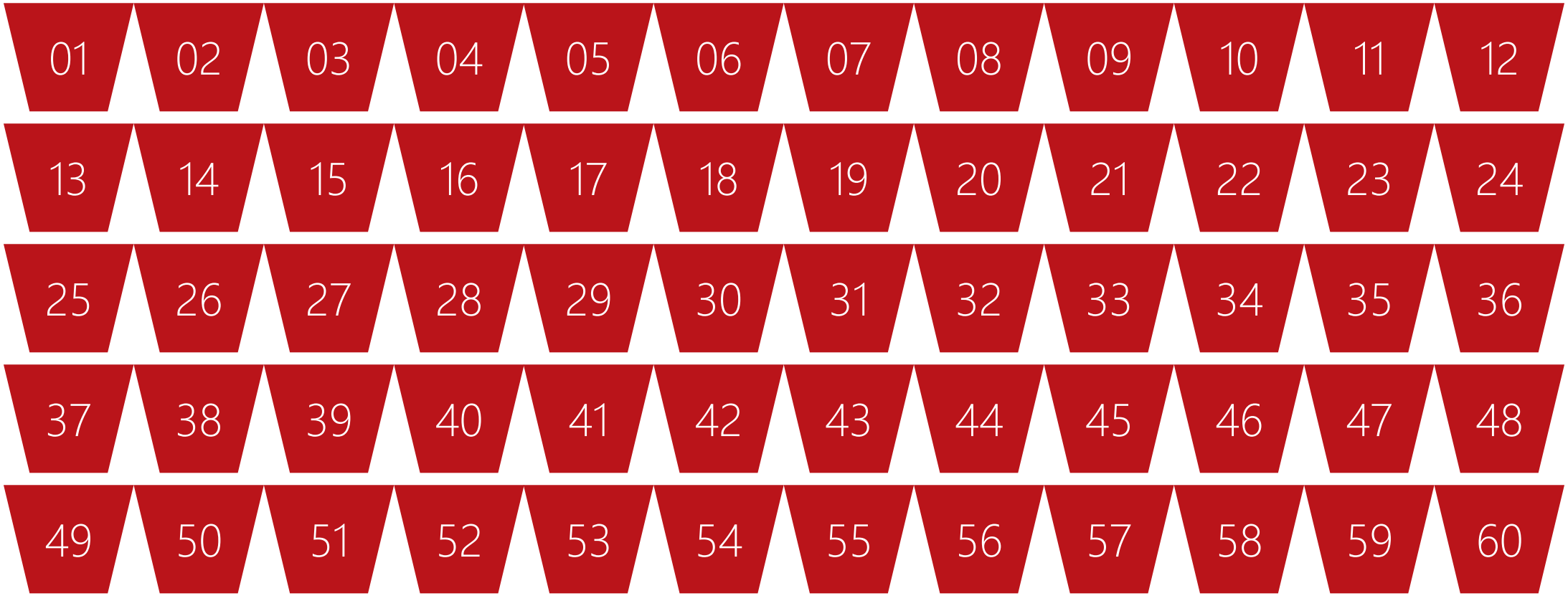

High-Speed Interconnect
Node 1
Node 2
Node 3
A...F
G...R
S...Z

Azure SQL DW - Logical overview



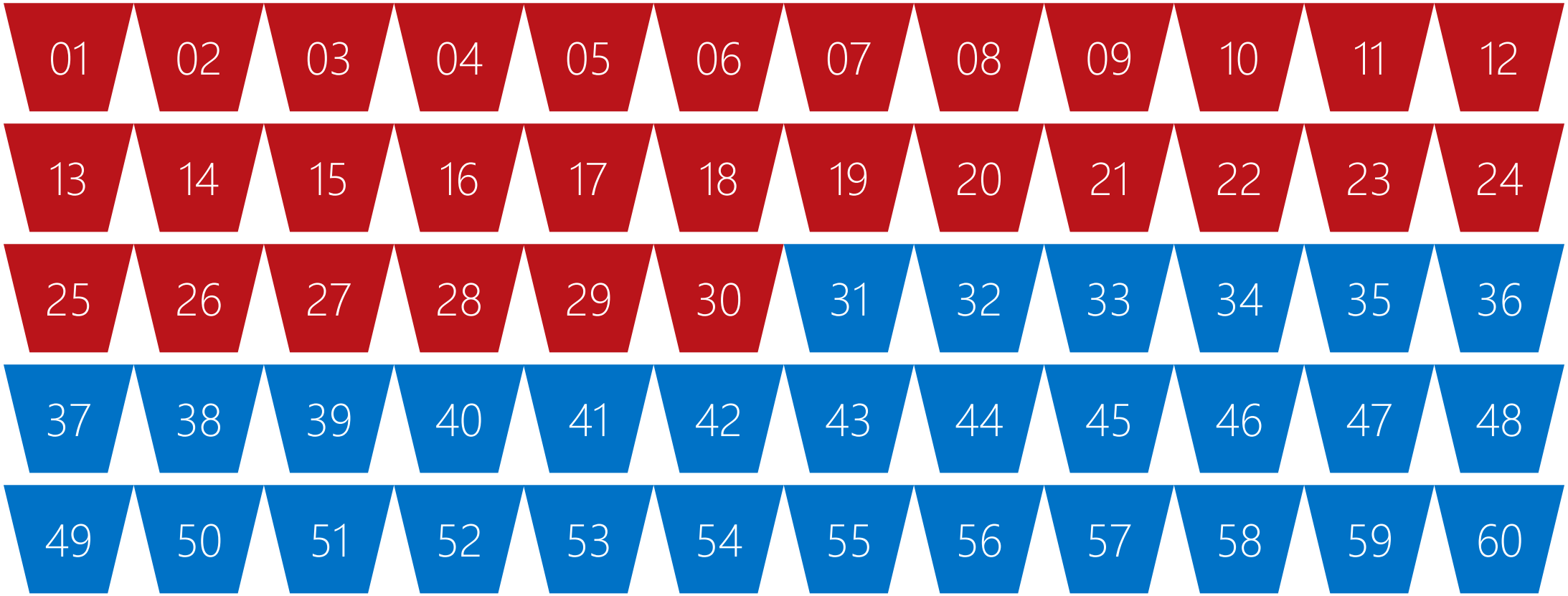
Mapping Compute in SQLDW

DW100



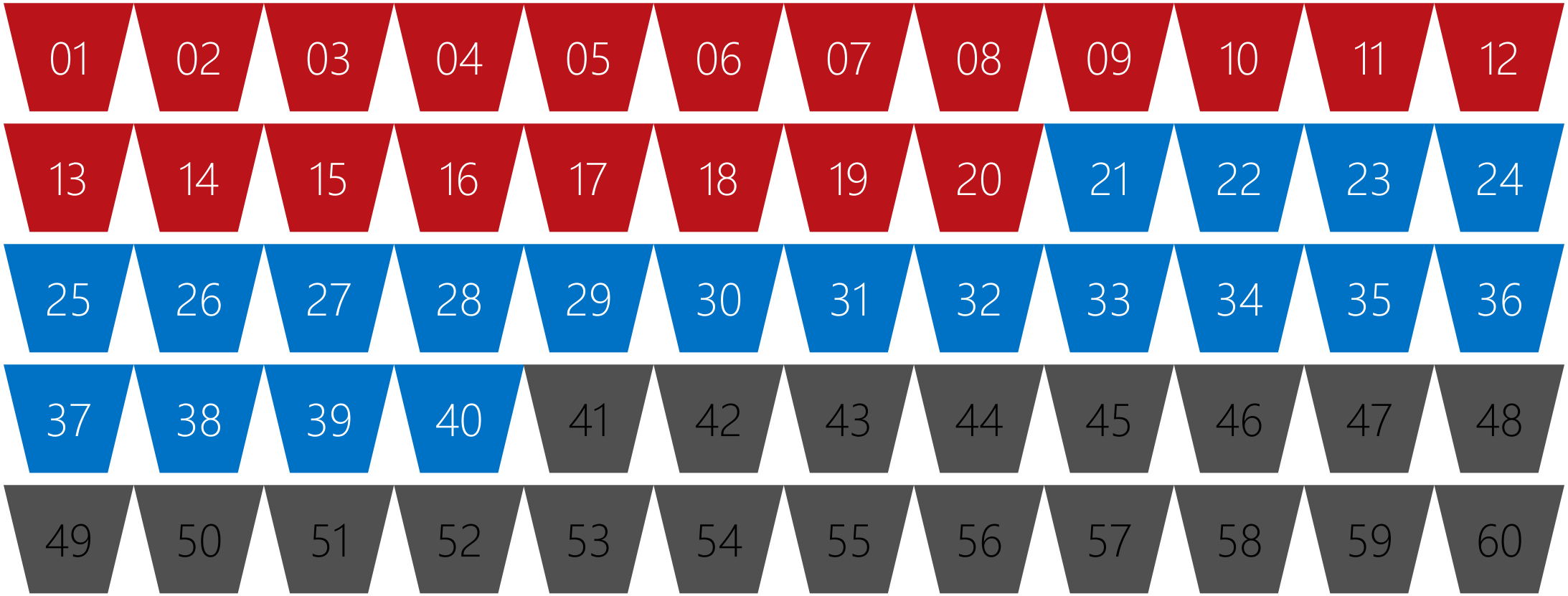
Mapping Compute in SQLDW

DW200



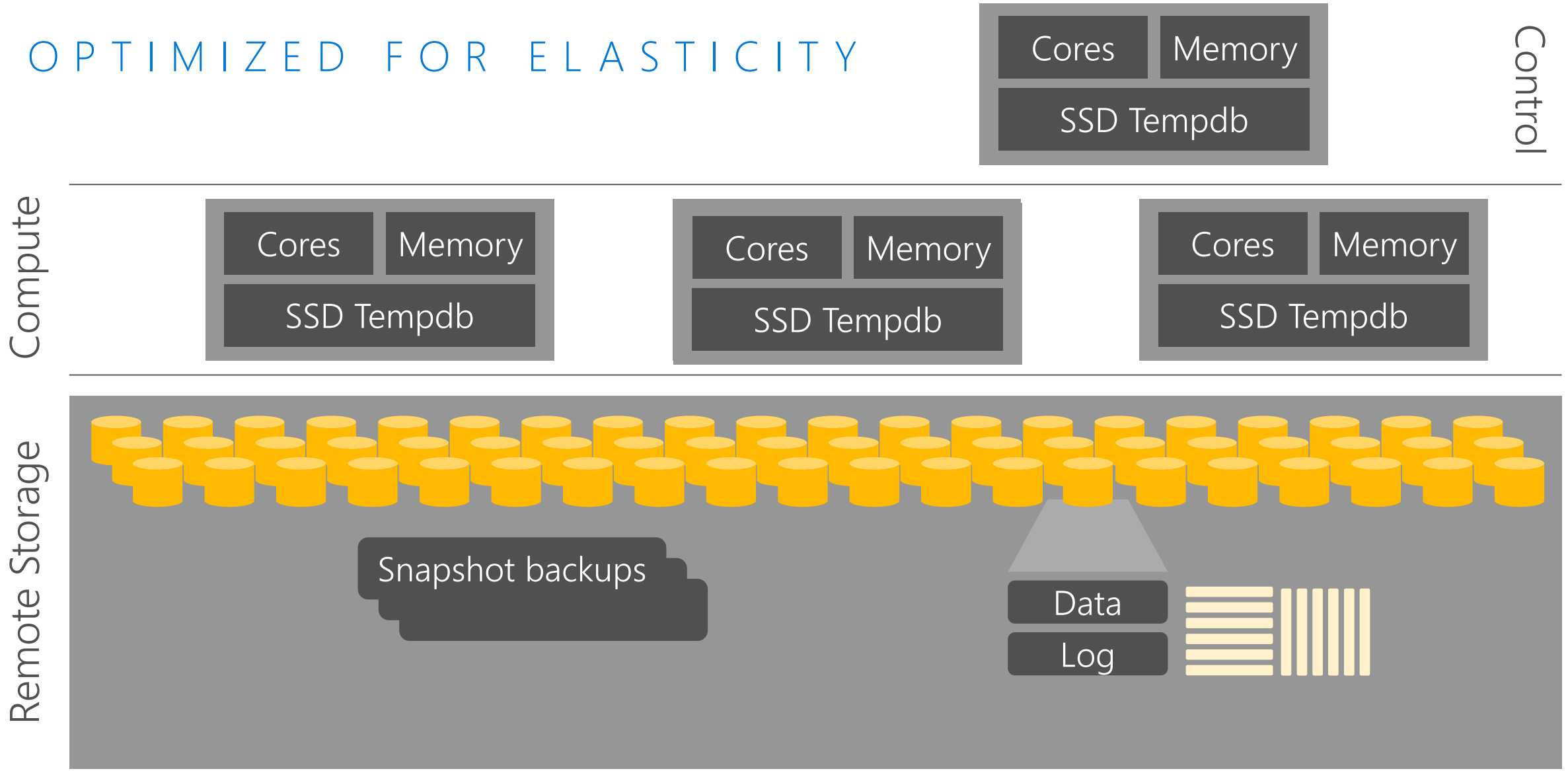
Mapping Compute in SQLDW

DW300

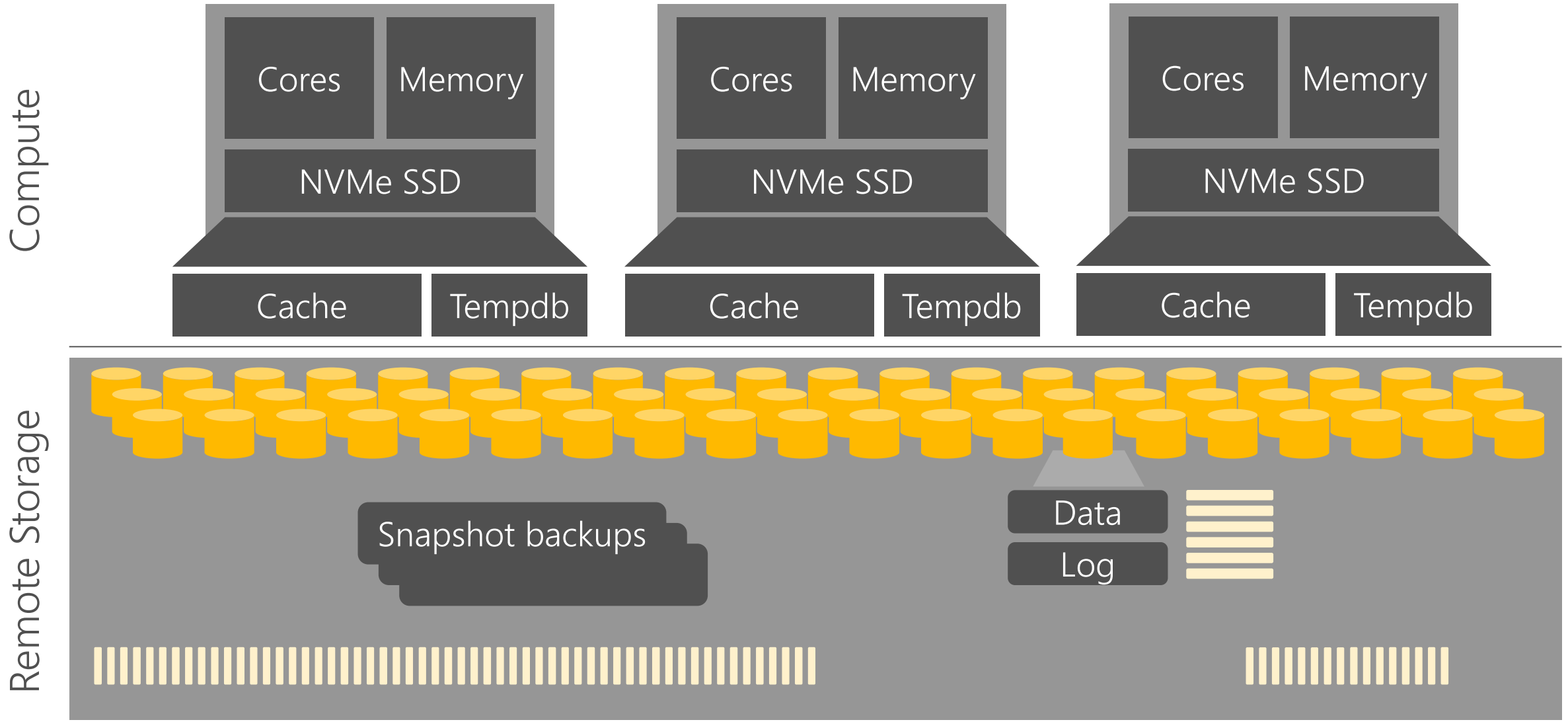


Azure SQL Data Warehouse
Optimized for Compute Performance Tier

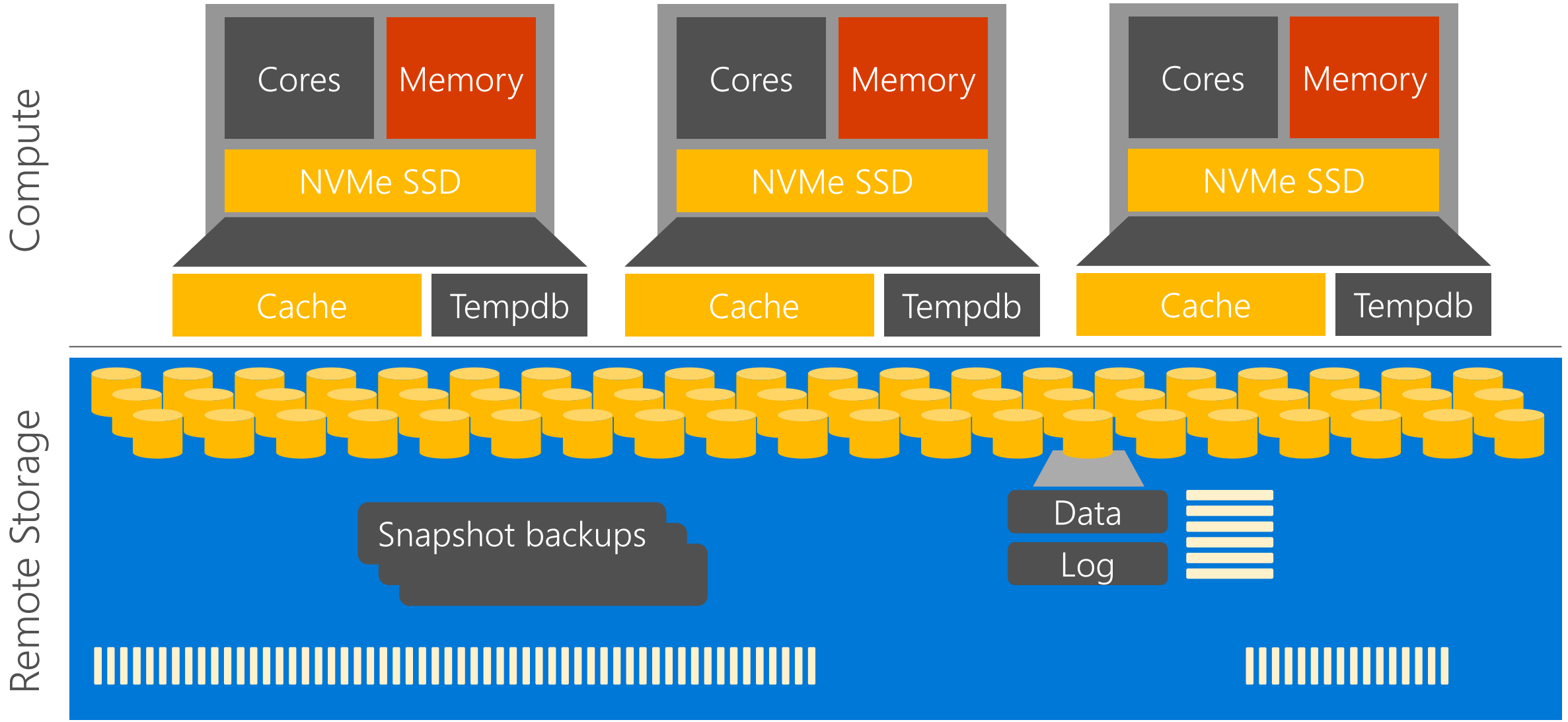
OPTIMIZED FOR ELASTICITY



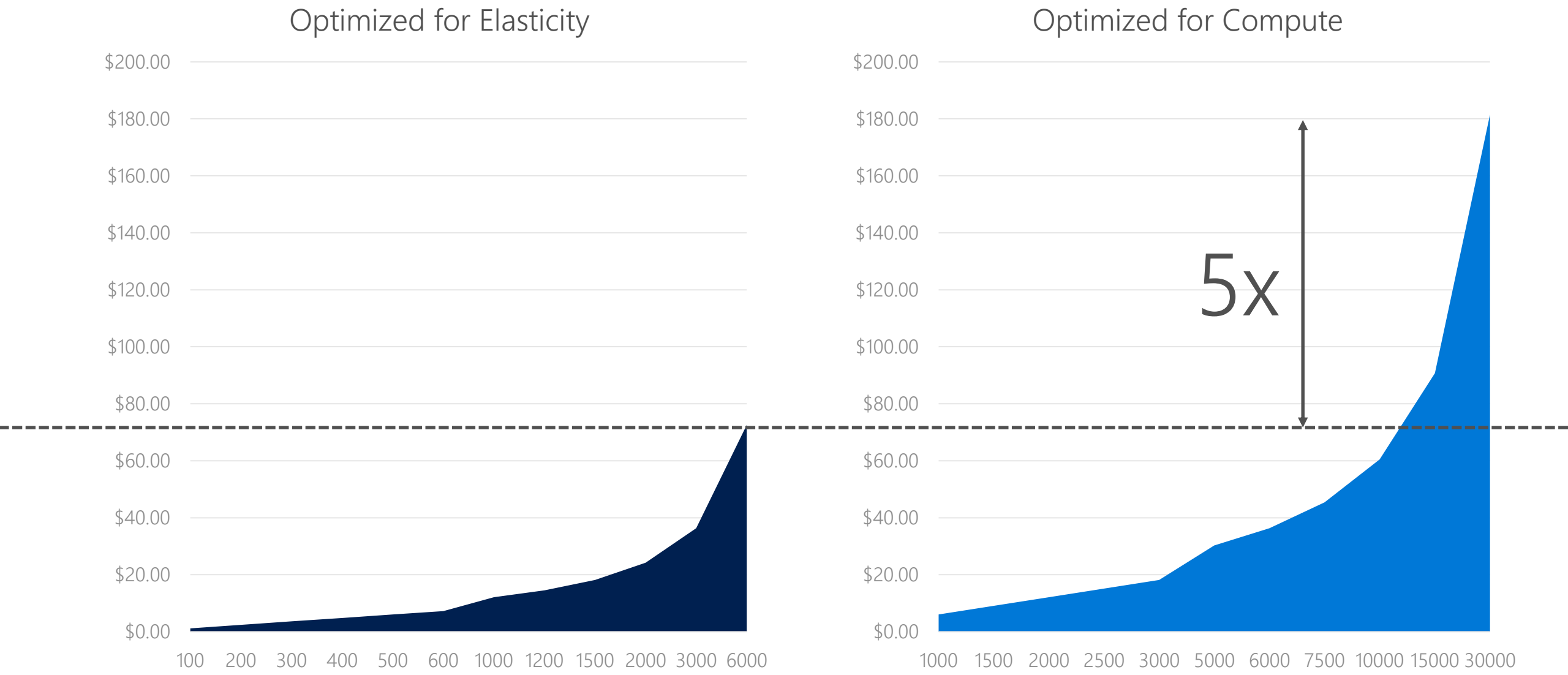
NEW! GEN2 IS OPTIMIZED FOR COMPUTE



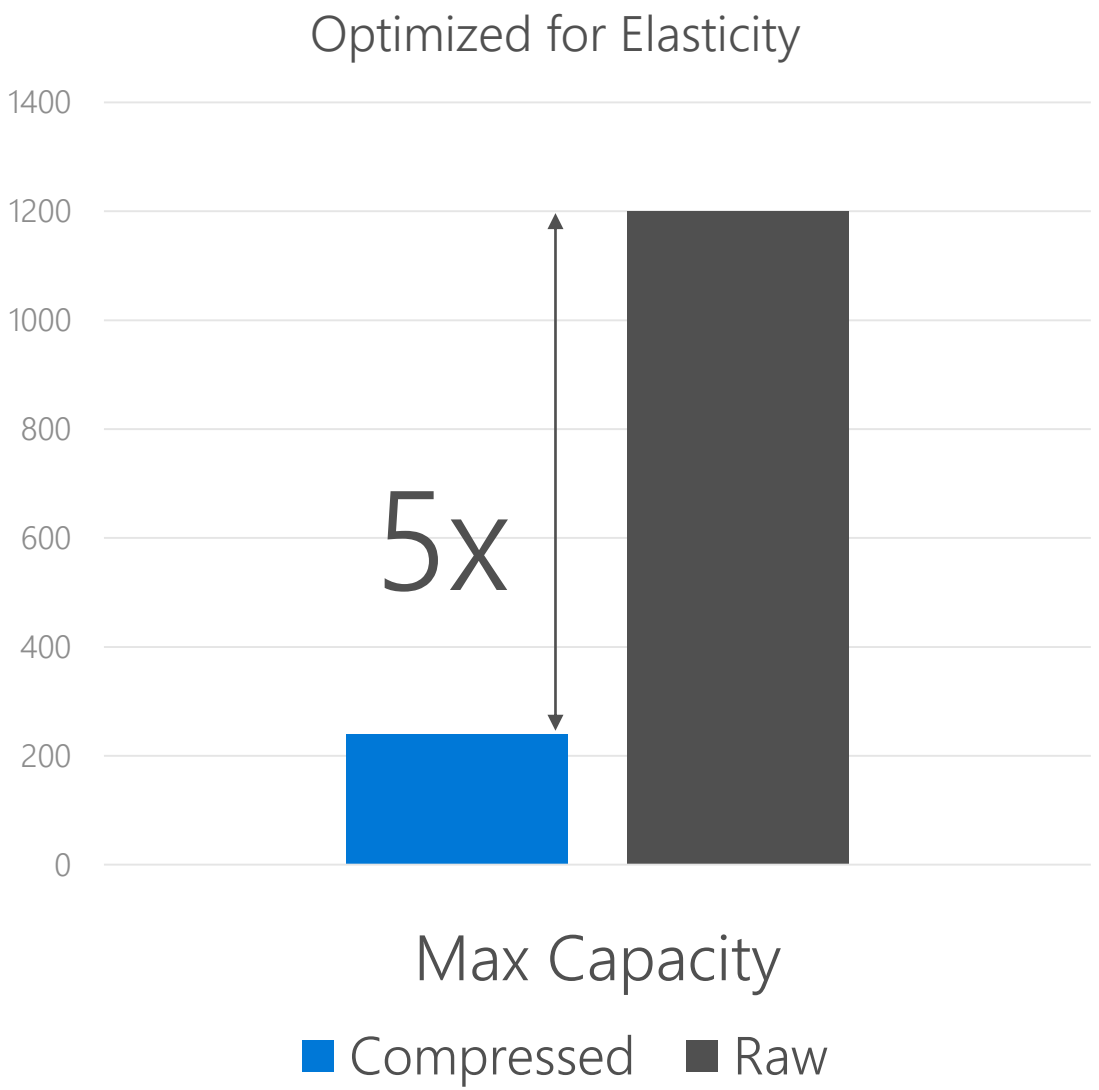
TIERED STORAGE MODEL AUTOMATES DATA TEMPERATURE

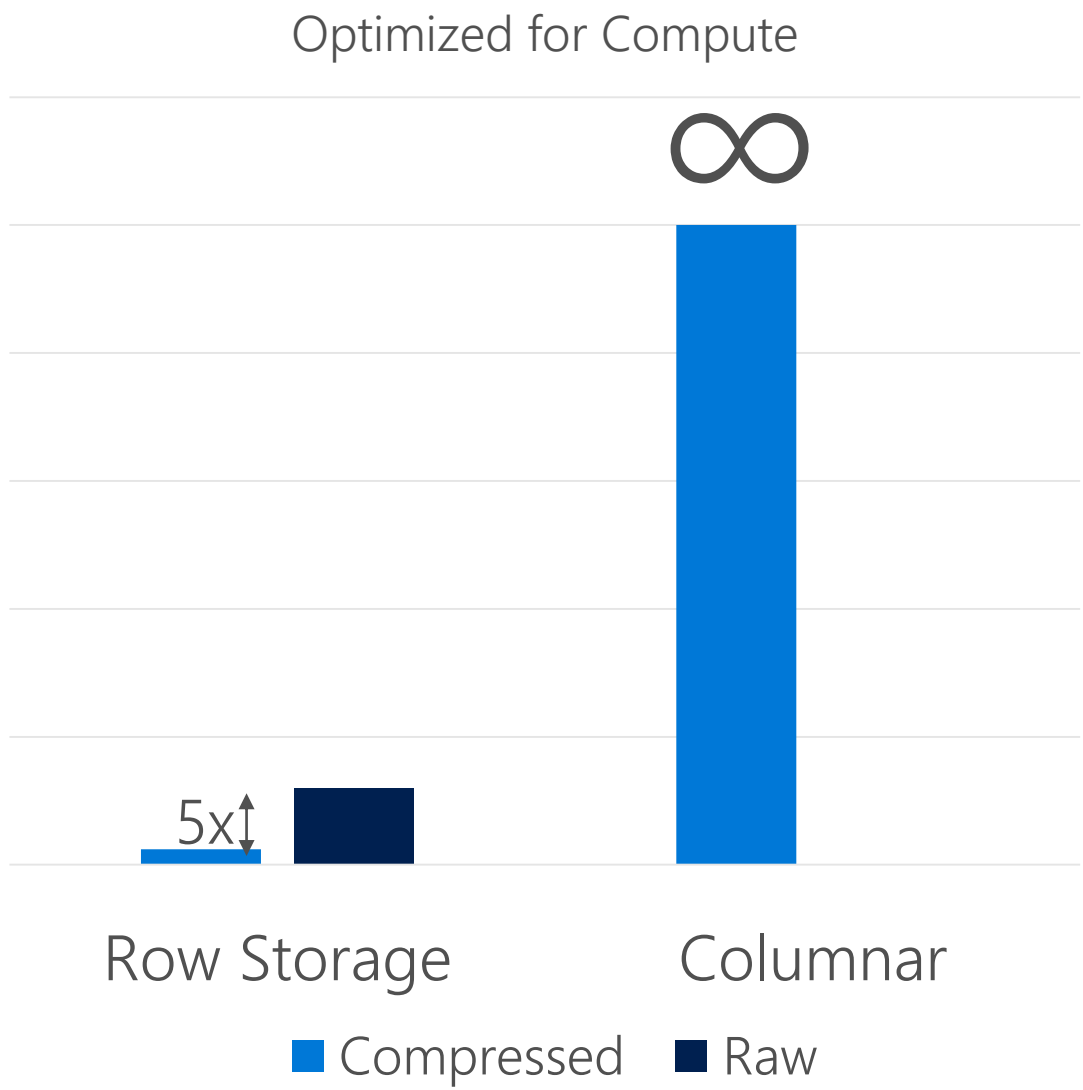


COMPUTE SCALABILITY



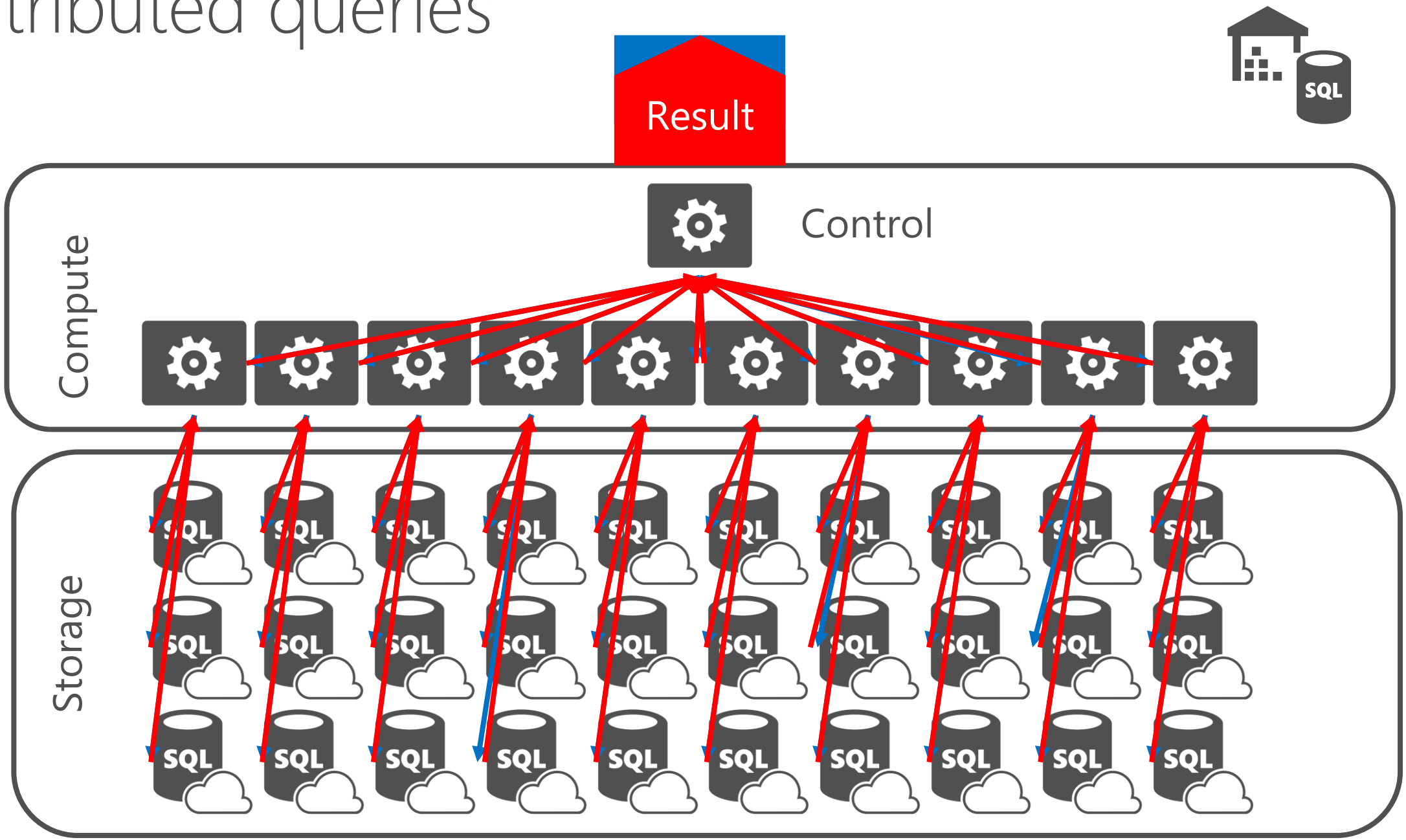
STORAGE SCALABILITY





Tables & Distribution

Distributed queries



Simple example

```sql
SELECT  COUNT_BIG(*)
FROM    dbo.[FactInternetSales]
;
```

```sql
SELECT  SUM(*)
FROM    dbo.[FactInternetSales]
;
```

Control

Compute

```sql
SELECT  COUNT_BIG(*)
FROM    dbo.[FactInternetSales]
;
```

```sql
SELECT  COUNT_BIG(*)
FROM    dbo.[FactInternetSales]
;
```

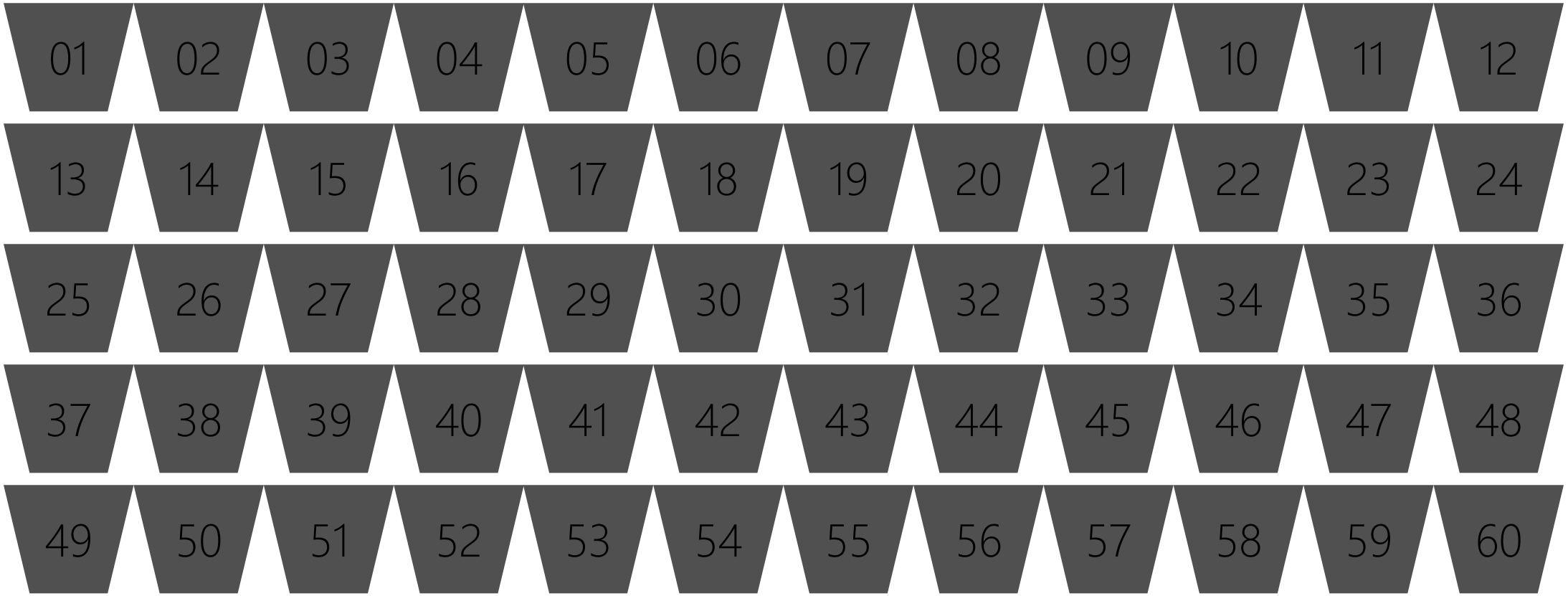
```sql
SELECT  COUNT_BIG(*)
FROM    dbo.[FactInternetSales]
;
```

```sql
SELECT  COUNT_BIG(*)
FROM    dbo.[FactInternetSales]
;
```

HASH distribution

A hash distributed table can deliver the highest query performance for joins and aggregations on large tables

HASH (02)



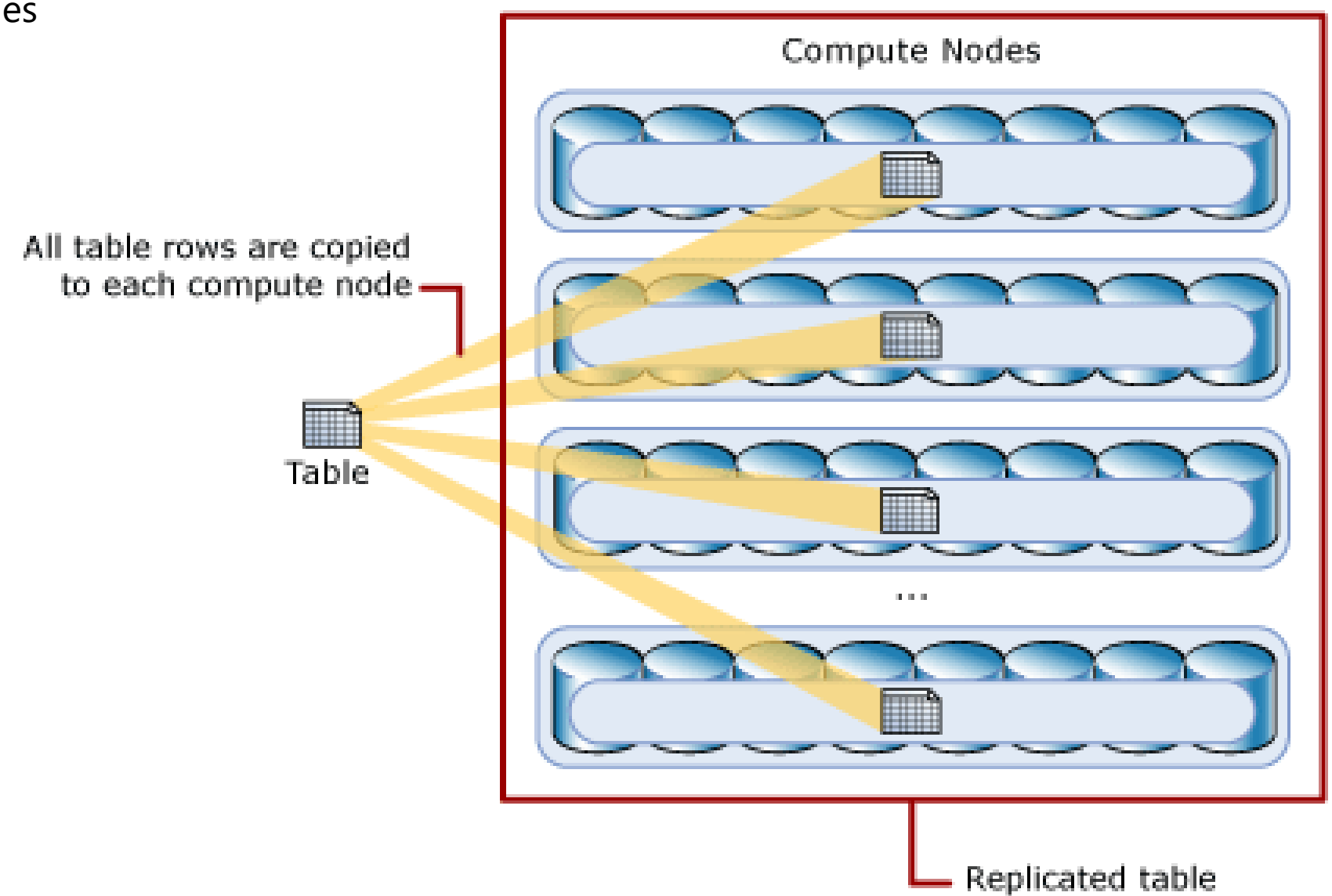
ROUND ROBIN DISTRIBUTION

A round-robin table is the simplest table to create and delivers fast performance when used as a staging table for loads.

01	02	03	04	05	06	07	08	09	10	11	12
13	14	15	16	17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32	33	34	35	36
37	38	39	40	41	42	43	44	45	46	47	48
49	50	51	52	53	54	55	56	57	58	59	60

Replicated Tables

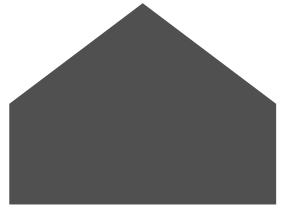
A replicated table provides the fastest query performance for small tables



Creating tables

```
CREATE TABLE [build].[FactOnlineSales]
(
    [OnlineSalesKey]        int           NOT NULL
,   [DateKey]               datetime      NOT NULL
,   [StoreKey]              int           NOT NULL
,   [ProductKey]            int           NOT NULL
,   [PromotionKey]          int           NOT NULL
,   [CurrencyKey]           int           NOT NULL
,   [CustomerKey]           int           NOT NULL
,   [SalesOrderNumber]      nvarchar(20)  NOT NULL
,   [SalesOrderLineNumber]  int               NULL
,   [SalesQuantity]         int           NOT NULL
,   [SalesAmount]           money         NOT NULL
)
WITH
(   CLUSTERED COLUMNSTORE INDEX
,   DISTRIBUTION = ROUND_ROBIN
)
;
```

```
CREATE TABLE [build].[FactOnlineSales]
(
    [OnlineSalesKey]        int           NOT NULL
,   [DateKey]               datetime      NOT NULL
,   [StoreKey]              int           NOT NULL
,   [ProductKey]            int           NOT NULL
,   [PromotionKey]          int           NOT NULL
,   [CurrencyKey]           int           NOT NULL
,   [CustomerKey]           int           NOT NULL
,   [SalesOrderNumber]      nvarchar(20)  NOT NULL
,   [SalesOrderLineNumber]  int               NULL
,   [SalesQuantity]         int           NOT NULL
,   [SalesAmount]           money         NOT NULL
)
WITH
(   CLUSTERED COLUMNSTORE INDEX
,   DISTRIBUTION = HASH([ProductKey])
)
;
```



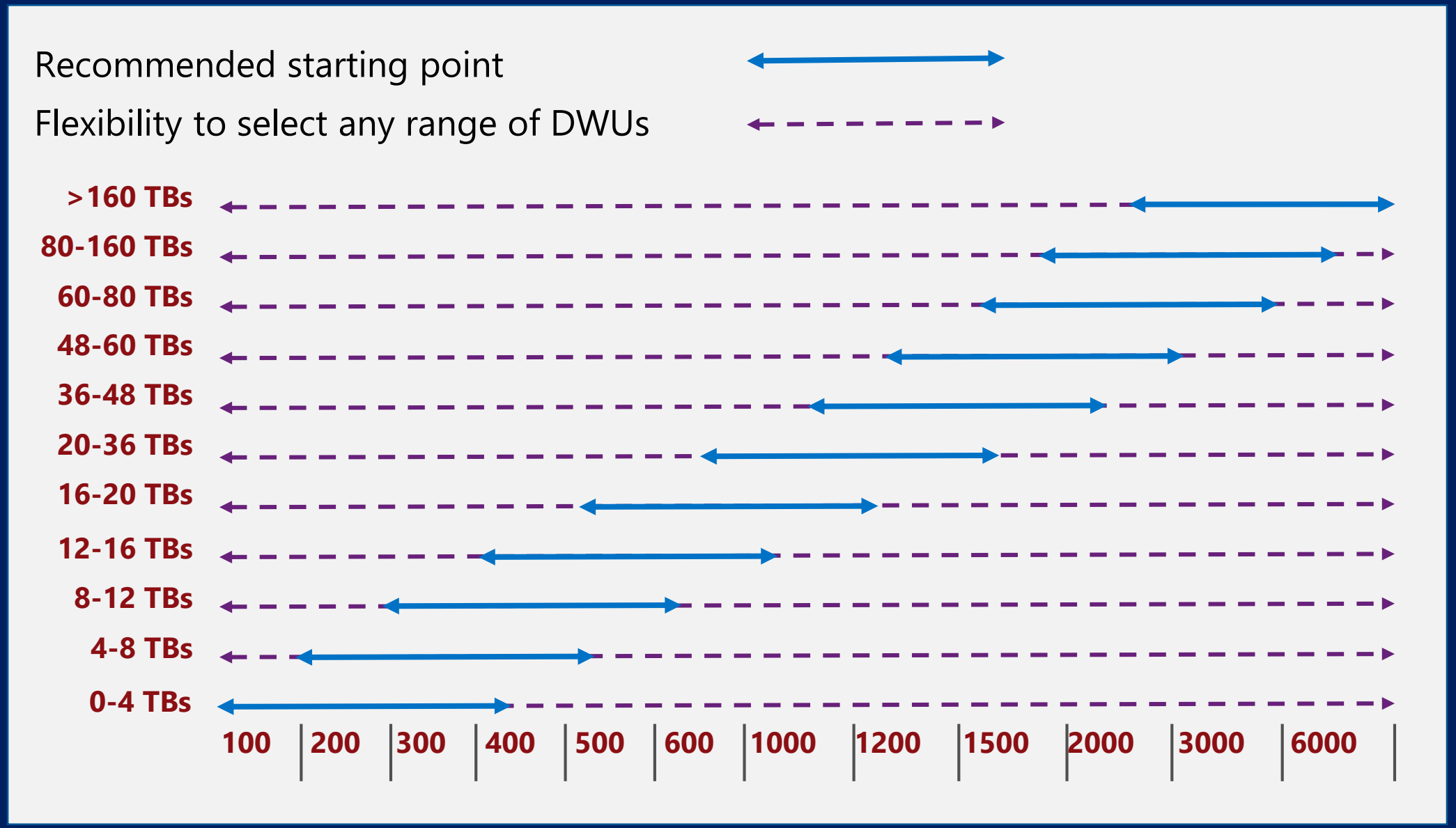
Sizing & Storage tiers

Sizing factors

Database capacity
Tempdb
Concurrency & Memory
Load
Transaction size
Memory management

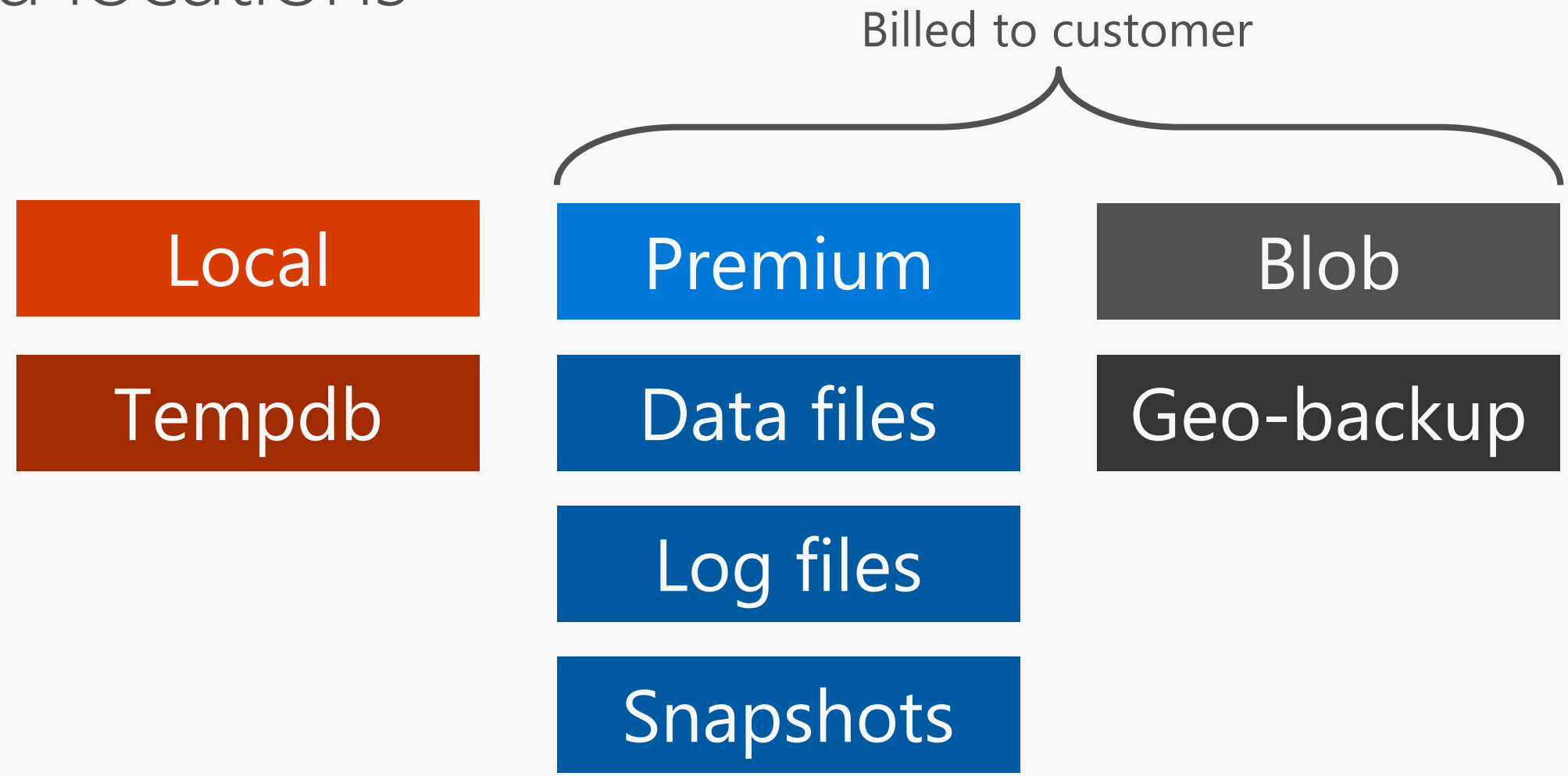
DWU

Starting point: Sizing by capacity

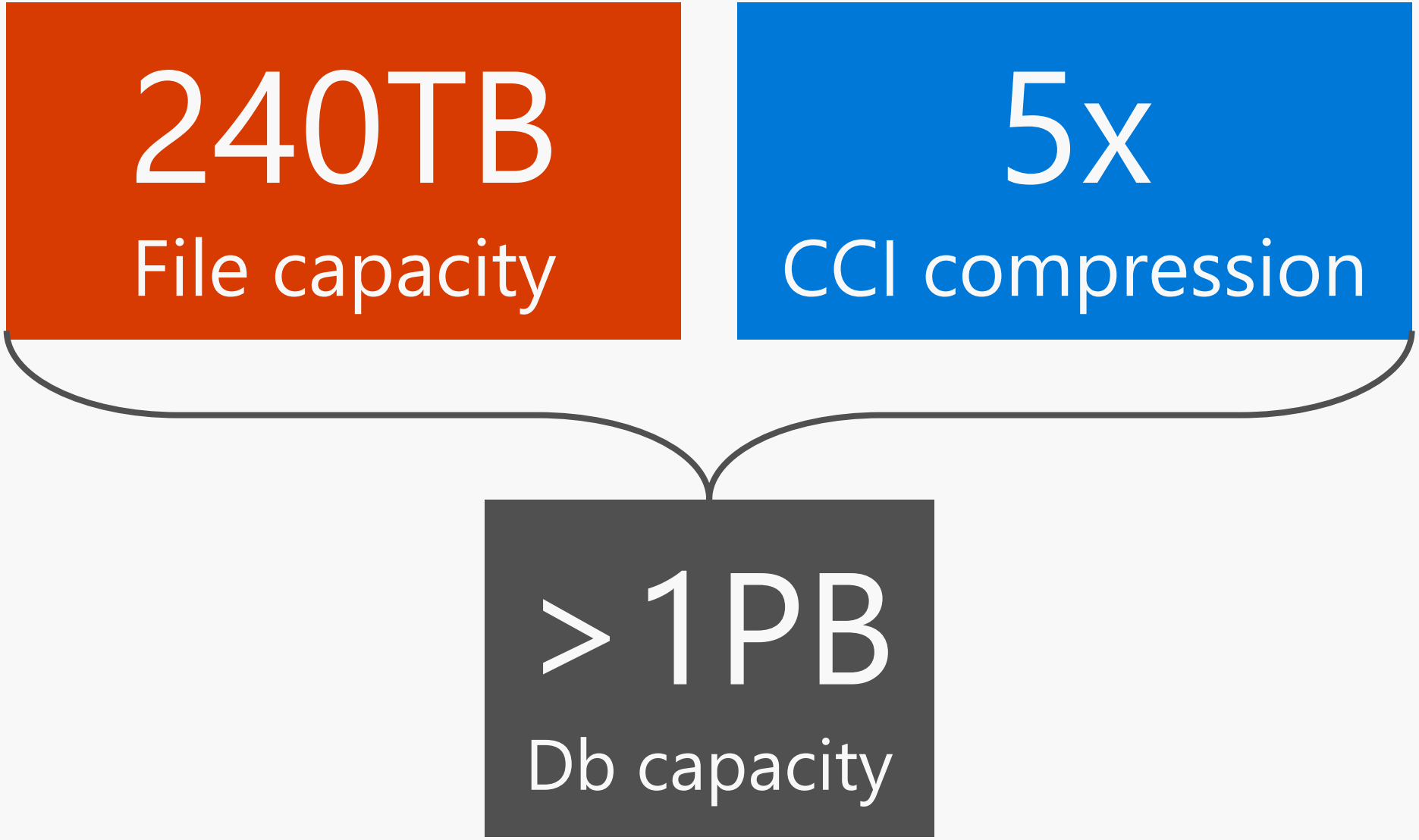


Storage tiers

Data locations

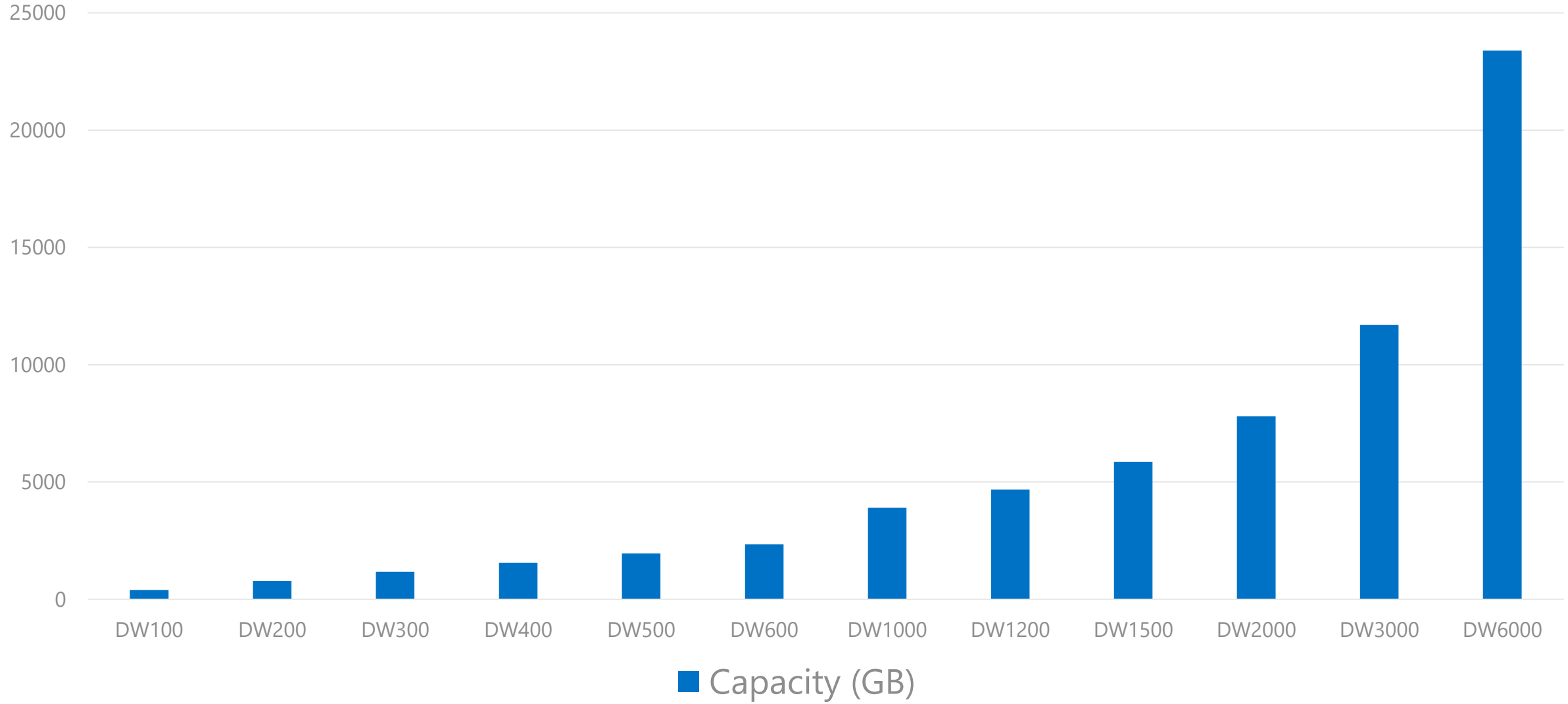


Premium Storage: Capacity limits



Local Storage: Tempdb sizing

~399GB
per DW100



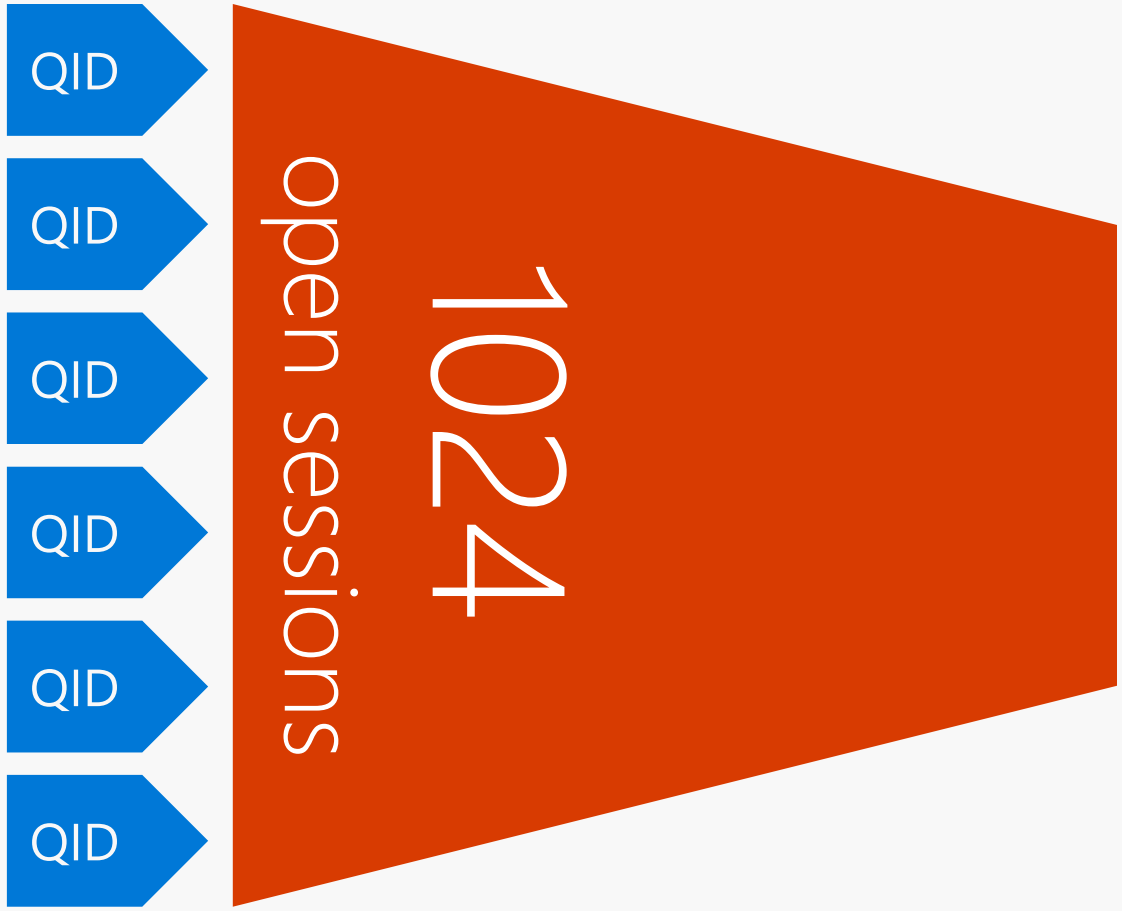
Workload Management

- The performance capacity of a query is determined by the user's resource class
  - Smaller resource classes reduce the maximum memory per query, but increase concurrency
  - Larger resource classes increases the maximum memory per query, but reduce concurrency

- There are two types of resource classes:
  - Static resources classes, which are well suited for increased concurrency on a data set size that is fixed.
  - Dynamic resource classes, which are well suited for data sets that are growing in size and increasing performance as the service level is scaled up

Refer: https://docs.microsoft.com/en-us/azure/sql-data-warehouse/memory-and-concurrency-limits

Concurrency: queries



128

active queries

Azure SQLDW Take-aways

- No CAPEX
- Low OPEX
- Provision in just 5 minutes
- Scale in seconds
- Fully parallel load
- Fully managed platform
- Time to insight measured in minutes
- Available with a one month free trial

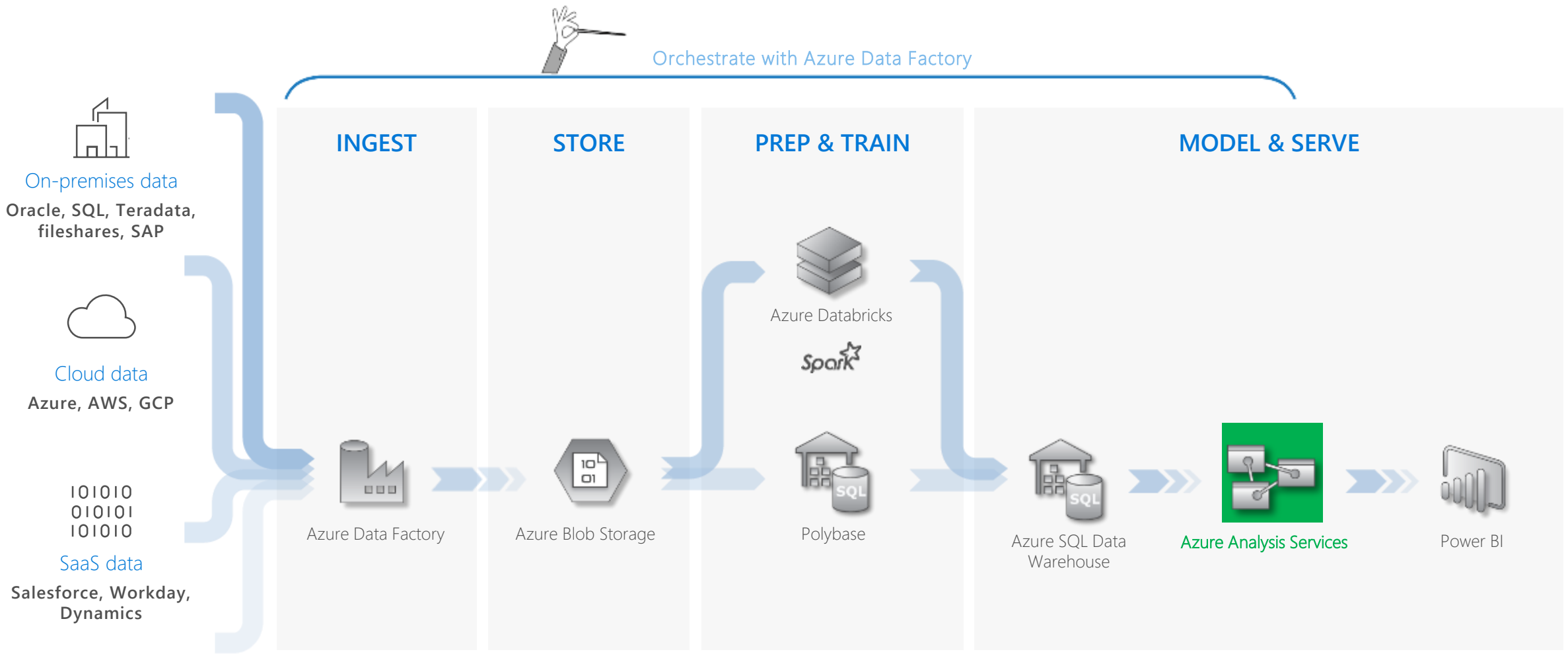
Demo

Walkthrough Portal for DW
Load the Data from Polybase

Azure Analysis Services

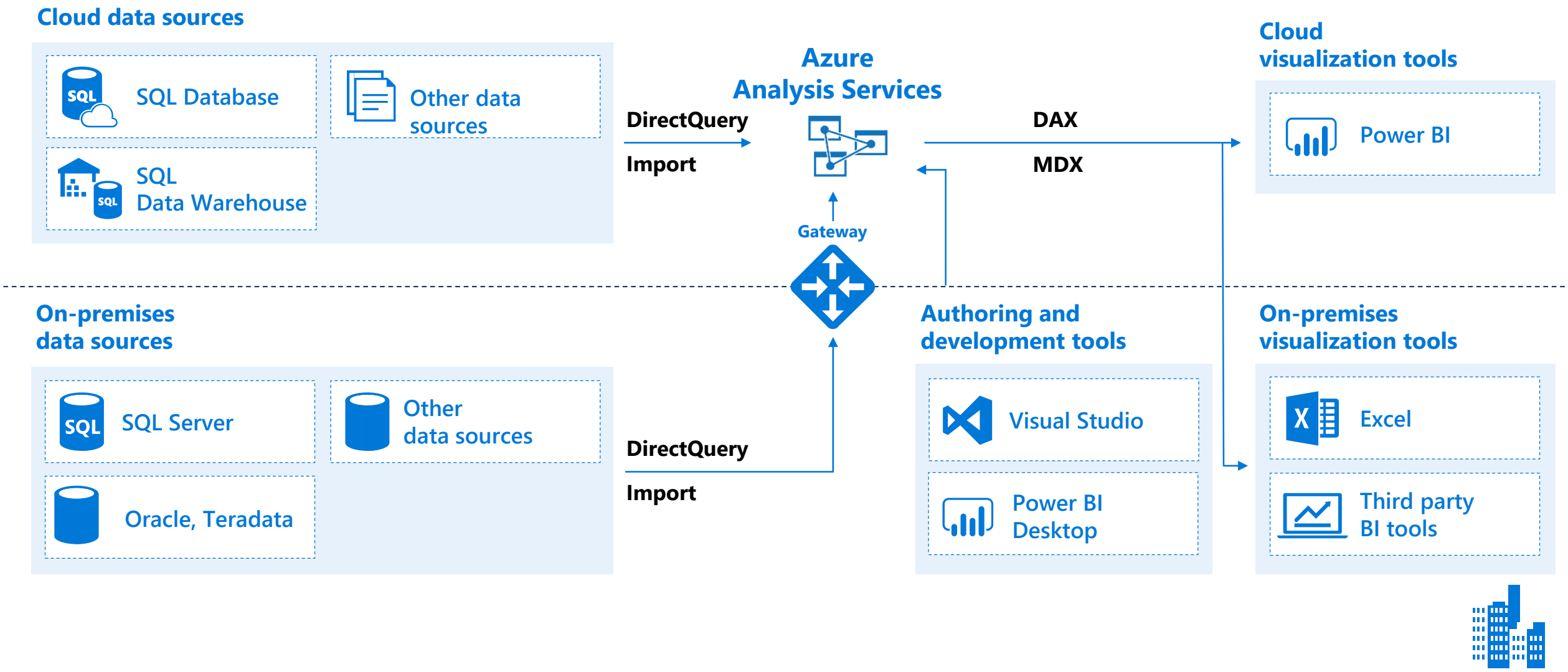
AZURE DATA WAREHOUSE

Modernize your enterprise data warehouse at scale



Microsoft Azure also supports other **Big Data** services like **Azure HDInsight**, **Azure SQL Database** and **Azure Data Lake** to allow customers to tailor the above architecture to meet their unique needs.

Azure Analysis Services Architecture



Where does Analysis Services fit?

- Analysis Services is not just about making your queries go faster – after all, we now have lots of other great scale-out/up options
  - Though it is still hard to beat at doing complex calculations quickly

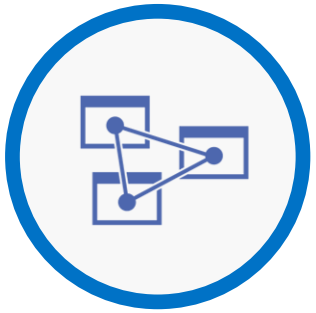
- The important thing it that it is a semantic layer:
  - Model your data once and share it with your users
  - Users can create queries by dragging and dropping in many different client tools – no need to write SQL

- Power BI is a self-service BI tool, not a replacement for Analysis Services

- Remember that Analysis Services is the engine behind the Power BI and Excel (Power Pivot) Data Models
  - Do your users want to build their own models? Can they?

Azure Analysis Services

Enterprise grade analytics engine as a service



Build rich semantic models

Transform complex data into business user friendly semantic models

Gain insights at the speed of thought

Gain instant insights with in-memory cache using your preferred visualization tools

Proven technology

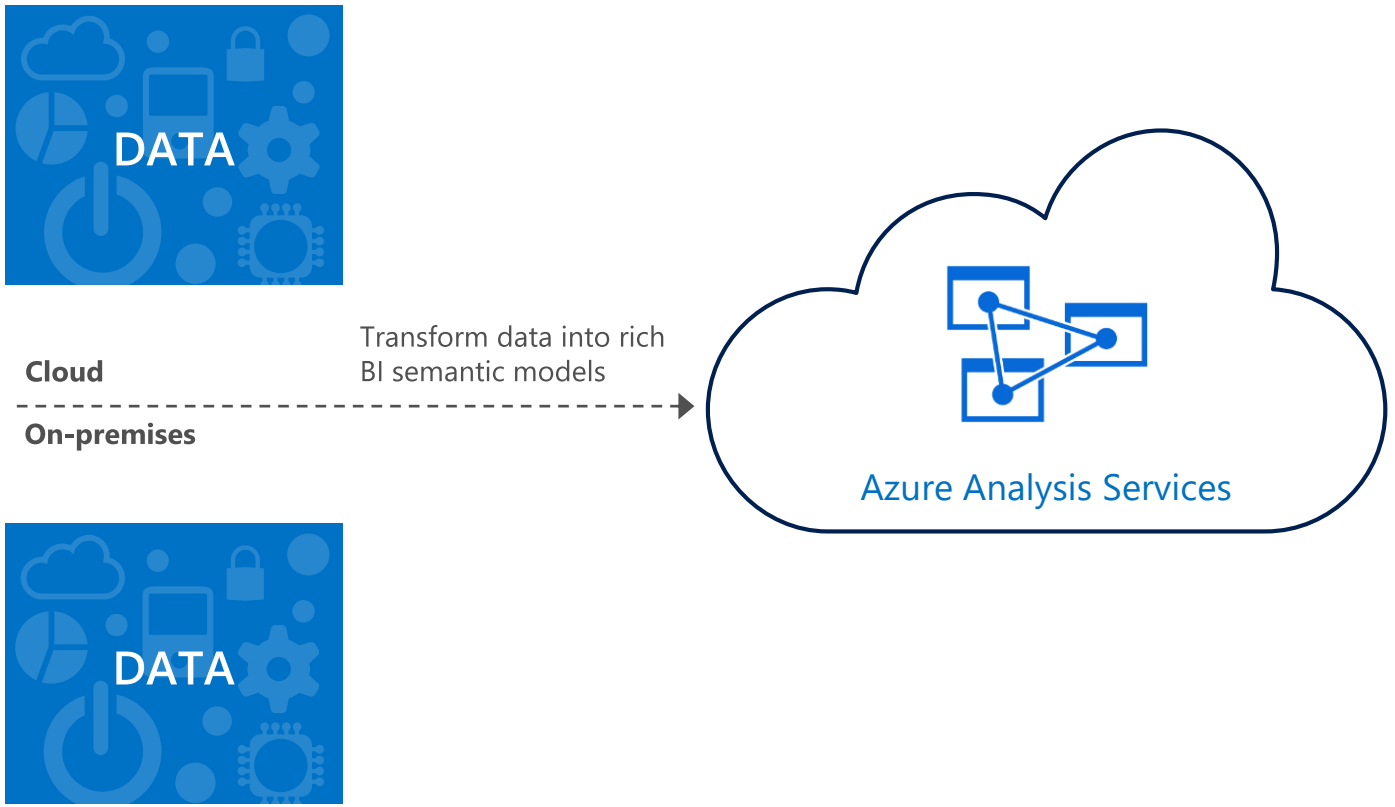
Based on powerful, proven SQL Server Analysis Services

Provision and scale with ease

Easy to deploy, scale, and manage as platform-as-a-service

Rich semantic modeling

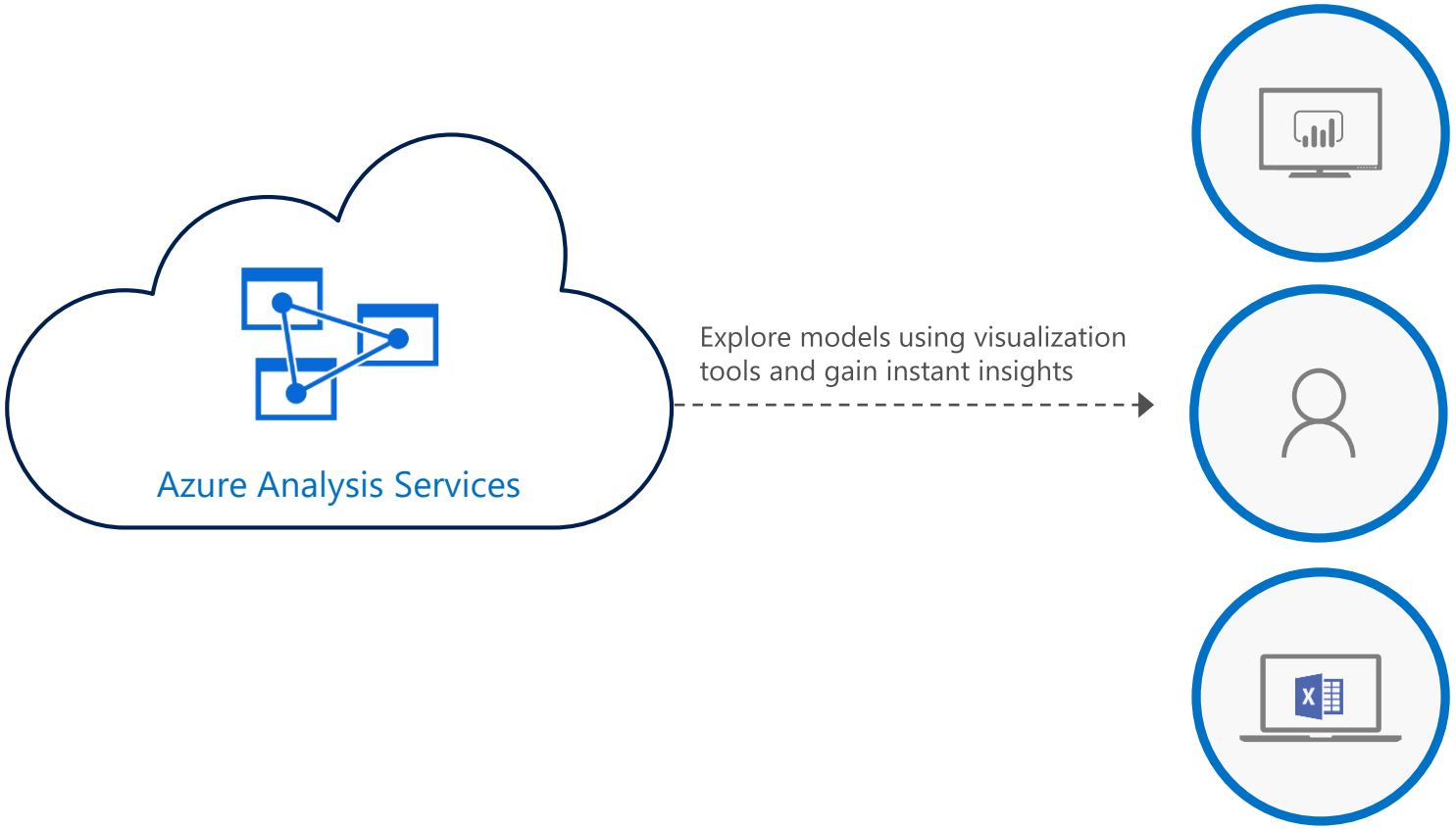
Transform complex on-premises or cloud data into insights



- Model complex data into business user friendly semantic datasets
- Combine into a single model
for one version of the truth
- Build scalable solutions over billions of rows of data
- Deliver trusted data models that business users can easily understand

Insights at the speed of thought

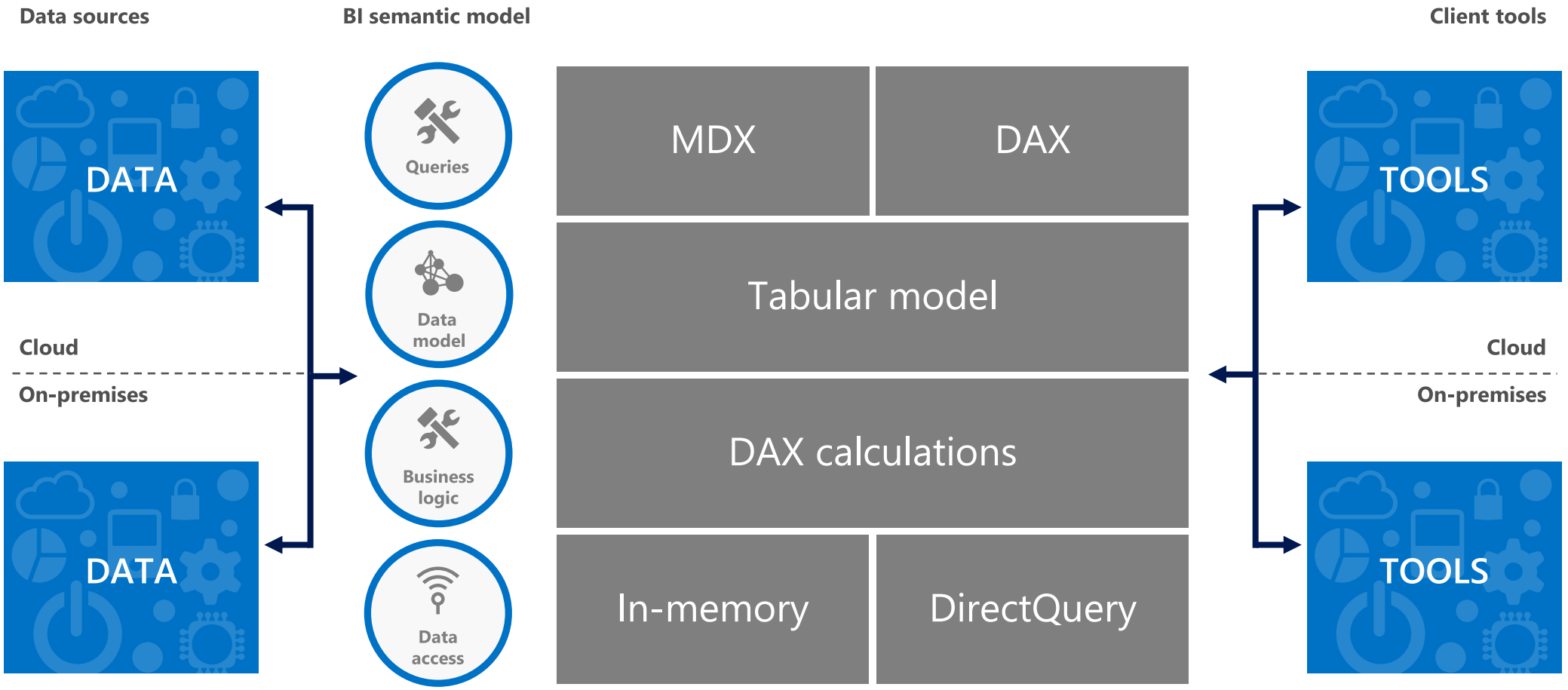
Gain instant insights using your preferred visualization tools



- Empower your business users with familiar data visualization tools like Power BI, Excel, and others
- Enable users to explore data and gain insights instantly with in-memory cache
- Extend the value of your existing data
- Access virtually any data—wherever it is

Proven analytics engine

Azure Analysis Services is based on SQL Server technology



Platform-as-a-service solution

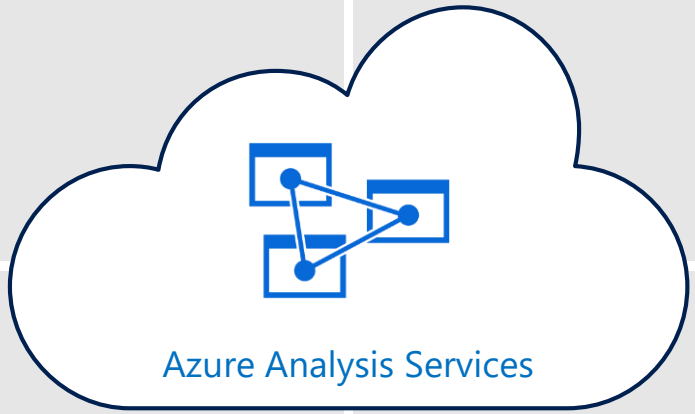
All the benefits of the cloud in your analytics engine

Get started quickly

Spin up a new server in seconds without managing the infrastructure

Provide secured access

From virtually anywhere



Access data when you need it

99.9% availability

Scale up, down, and pause

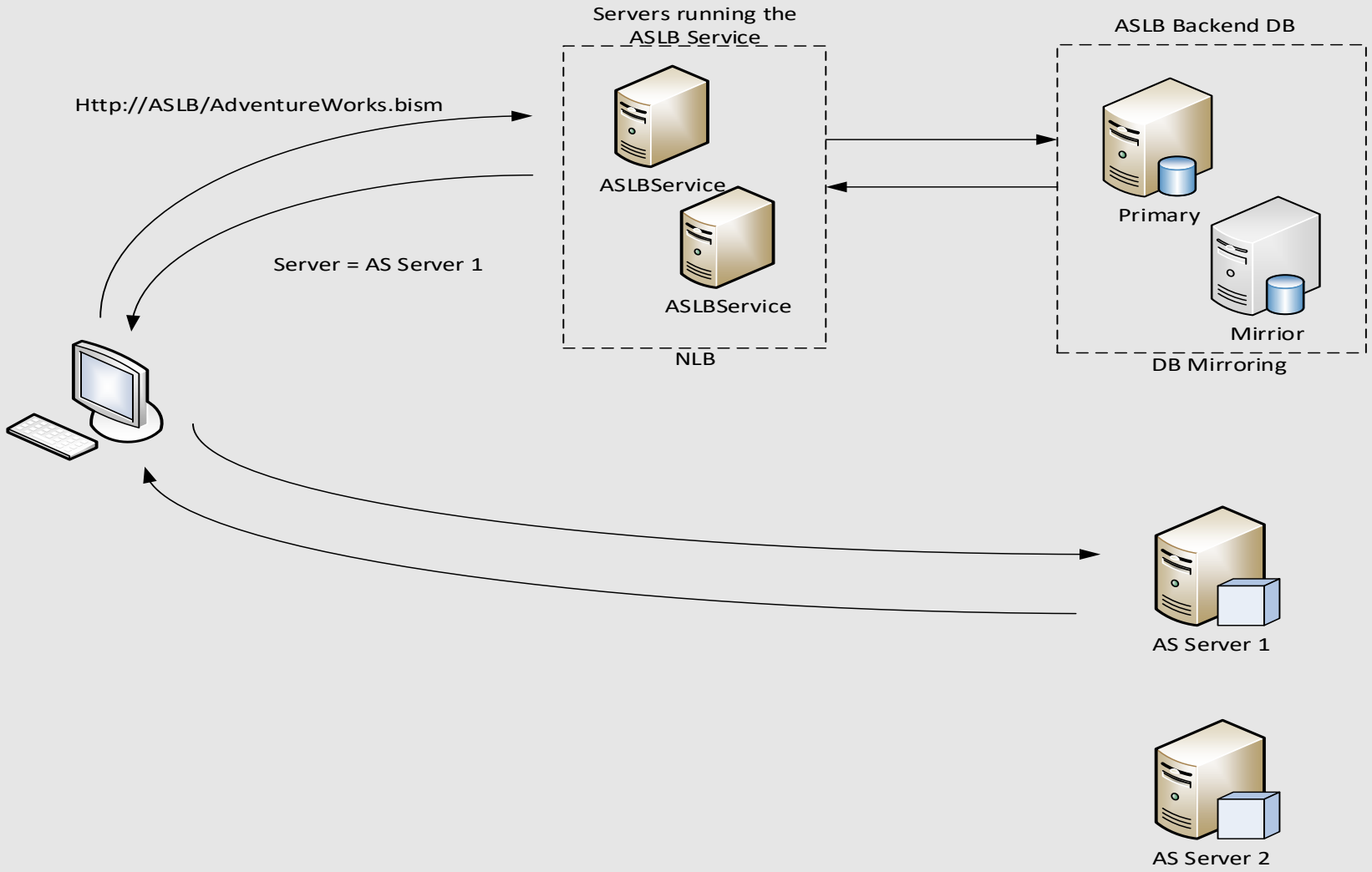
Only pay for what you need

Rely on Microsoft's experience running trusted enterprise cloud services

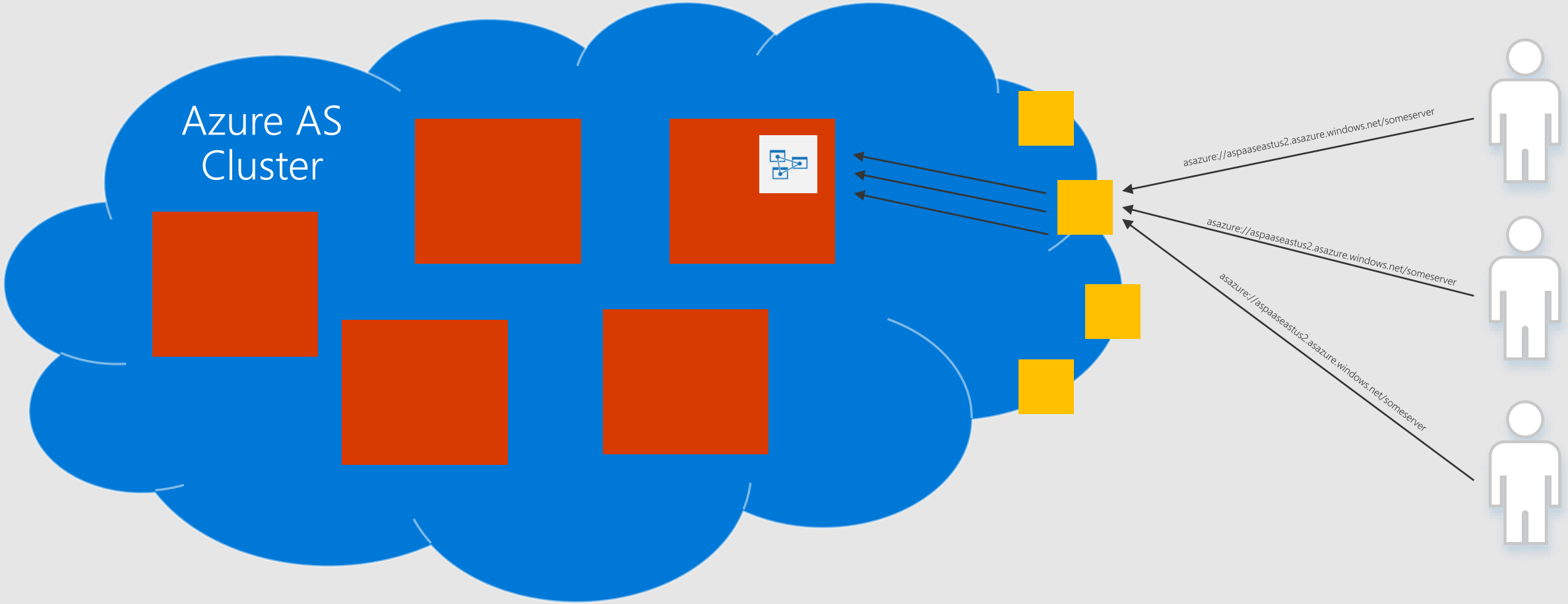
Azure Analysis Services capabilities at a glance

- Fully managed Platform-as-a-Service
- 99.9% uptime SLA
- Scale out
- Elastic scale up/down
- Pause & resume
- Up to 400 GB memory per server
- SSAS management tool compatibility: SSMS, SQL Profiler, Deployment Wizard …
- Azure Active Directory
  - **Azure B2B support**
  - **Application service principals**
- Backup/restore
- Unified Gateway
- Firewall

Scaling queries - before



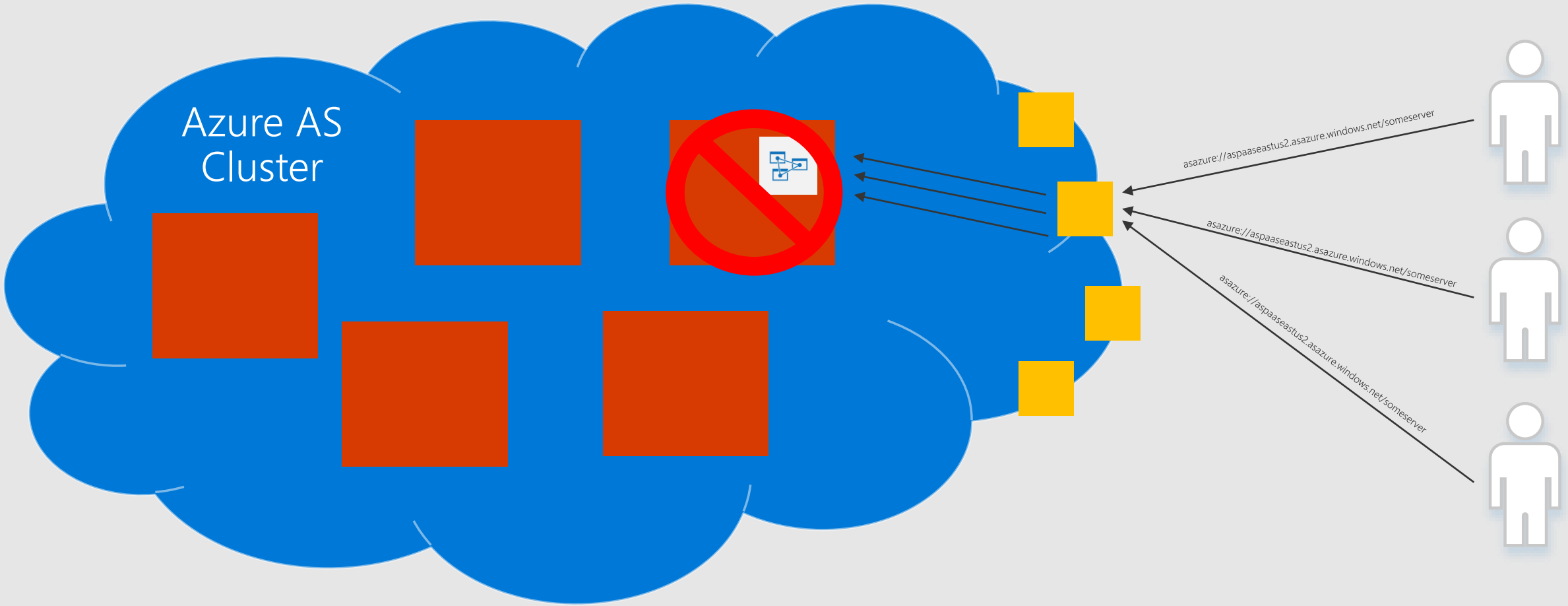
Scale out



Frontend node

Backend node

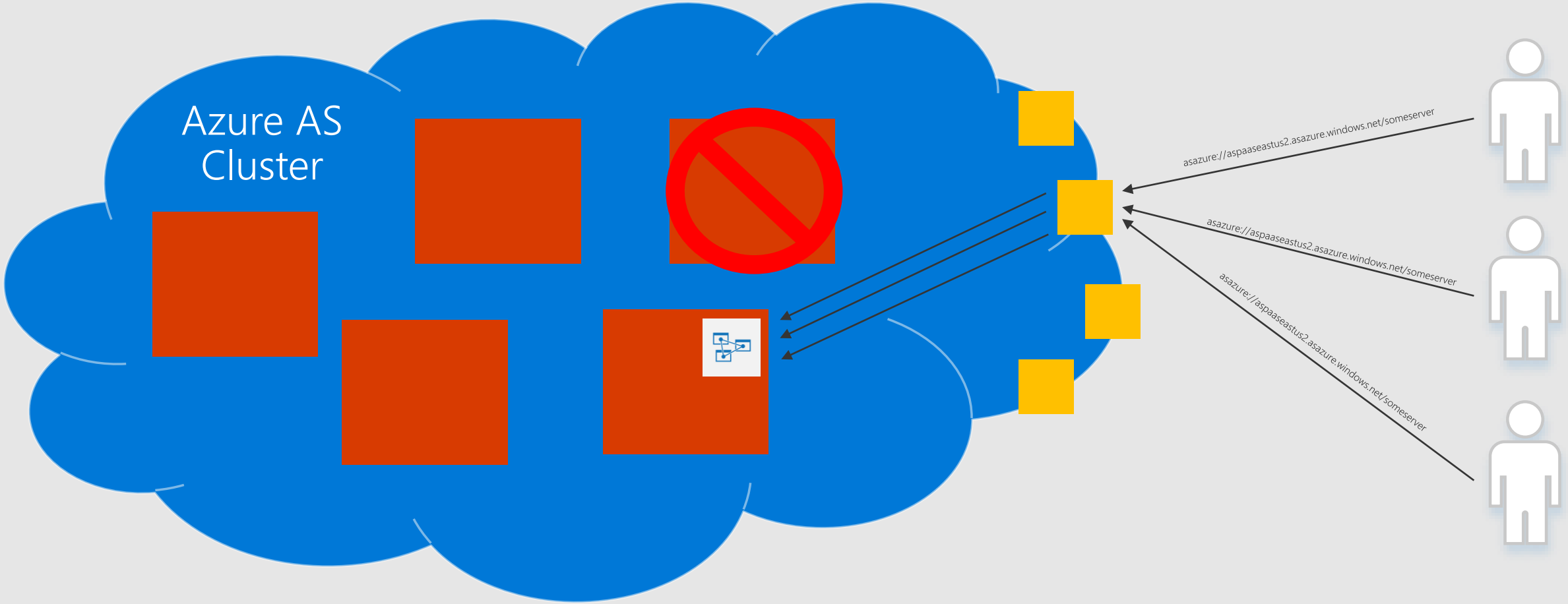
Scale out



Frontend node

Backend node

Scale out

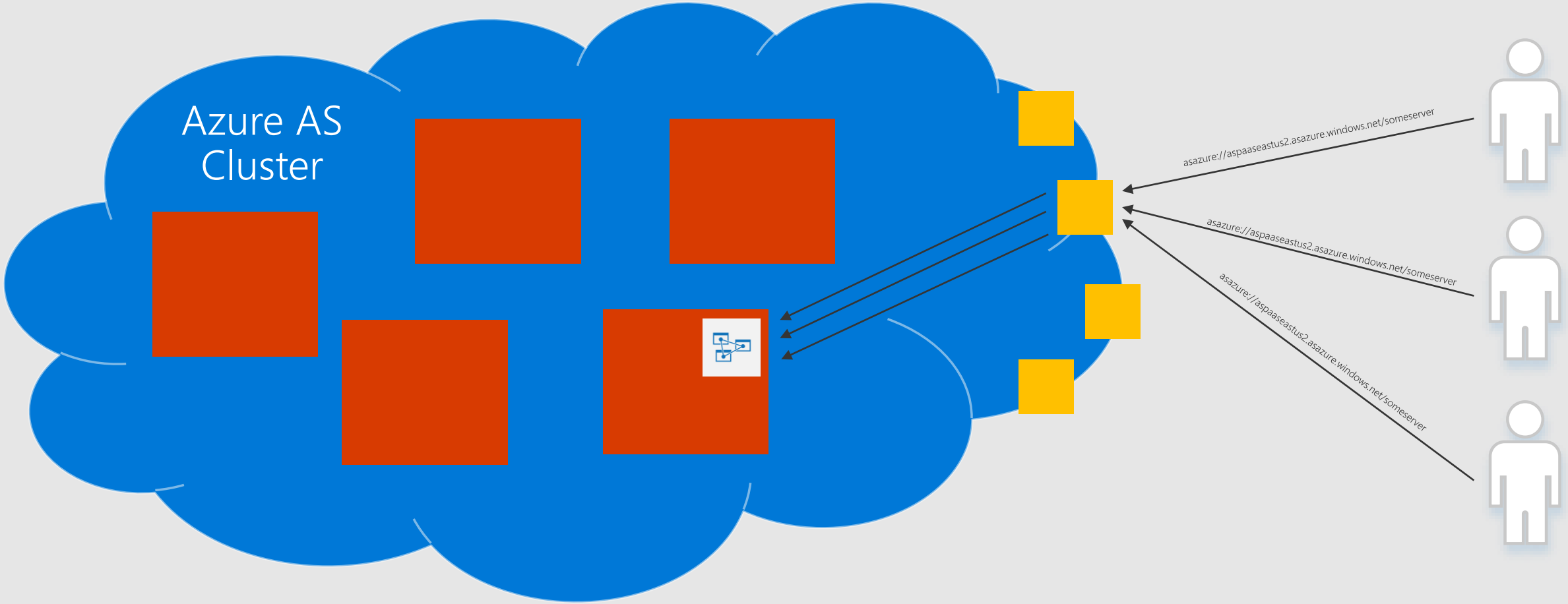


Frontend node

Backend node

Scale out

Number of query replicas: 0

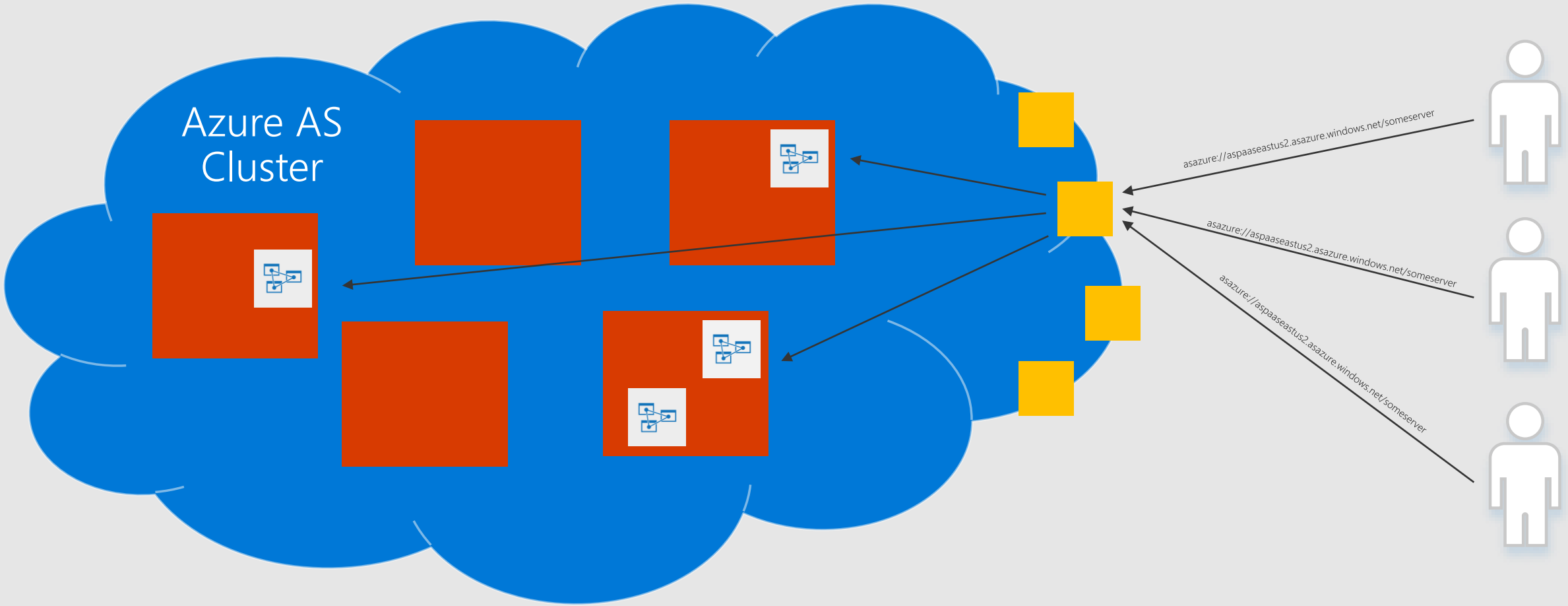


Frontend node

Backend node

Scale out

Number of query replicas: 3

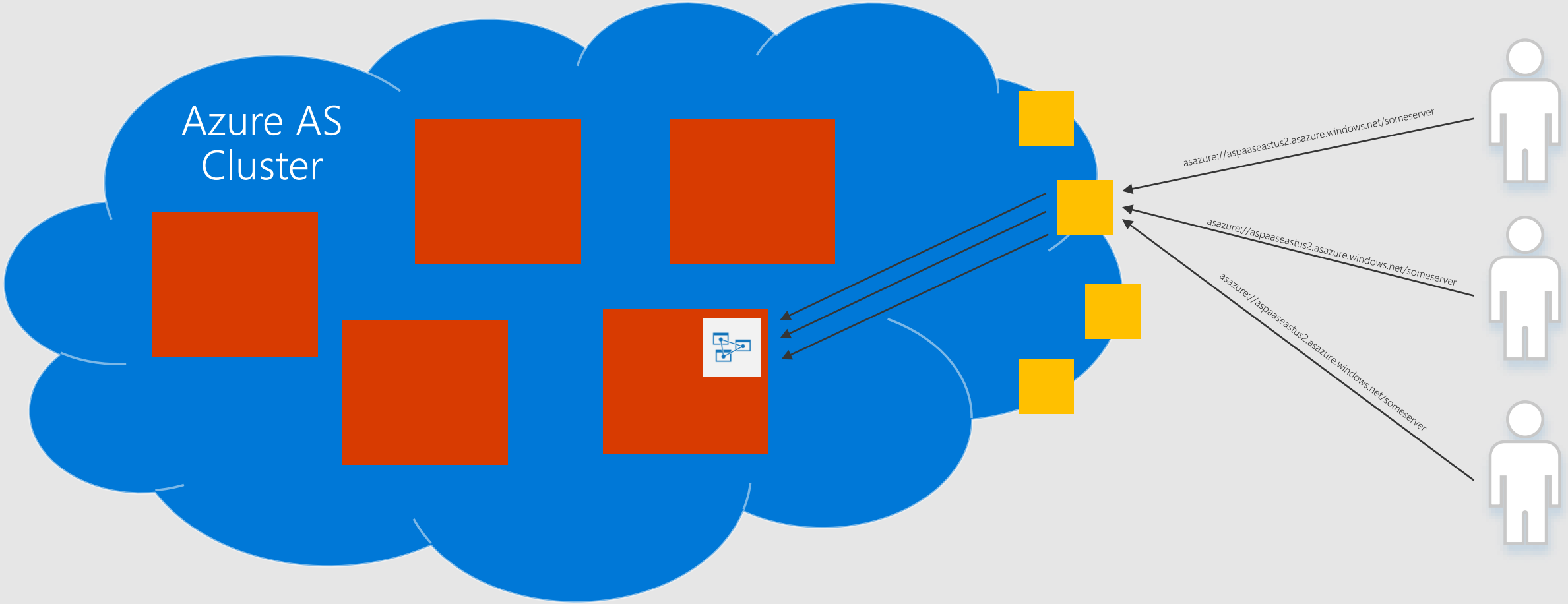


Frontend node

Backend node

Scale out

Number of query replicas: 0



Frontend node

Backend node

Performance levels

LEVEL	QPUS	MEMORY (GB)	SLA	PRICE
B1	40	10	99.9	$319.92/mo; $0.43/hr
B2	80	20	99.9	$639.84/mo; $0.86/hr
S0	40	10	99.9	$602.64/mo; $0.81/hr
S1	100	25	99.9	$1,510.32/mo; $2.03/hr
S2	200	50	99.9	$3,020.64/mo; $4.06/hr
S4	400	100	99.9	$6,033.84/mo; $8.11/hr
S8	320	200	99.9	$7,722.72/mo; $10.38/hr
S9	640	400	99.9	$15,445.44/mo; $20.76/hr
Developer	20	3	None	$98.21/mo; $0.132/hr

B – basic level can't have perspectives, multiple partitions and DirectQuery mode

Query processing units (QPUs)

- Unit of measure in Azure Analysis Services
- Based on a set of typical analytical queries and processing commands run on Azure and determines how many transactions are completed per second under fully loaded conditions
- 20 QPUs is about 1 pretty fast core

Demo