SQL Data Warehouse
Key Concepts

Chris Testa-O'Neill
Features Engineer
Analytics and Data Science Team

Microsoft

Agenda

SQL DW in the context of Cortana Intelligence

Introducing SQLDW

Target workloads

Scaling up vs. Scaling out

Scaling Compute

Pause Resume

SQLDW in the CIS context

Cortana Intelligence in a Sentence:

Cortana Intelligence is a Platform and a Process to perform advanced analytics from start to finish

The Team Data Science Process

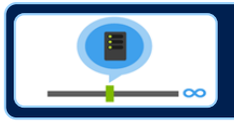
Business Understanding	• Define Objectives • Identify Data Sources
Data Acquisition and Understanding	• Ingest Data • Explore Data • Update Data
Modeling	• Feature Selection • Create and Train Model
Deployment	• Operationalize
Customer Acceptance	• Testing and Validation • Handoff • Re-train and re-score

The Cortana Intelligence Platform

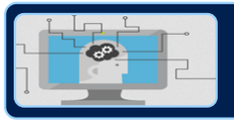
Cortana, Cognitive Services, Bot Framework

Power BI

Stream Analytics



HDInsight



Azure Machine Learning (MRS)

SQL Data Warehouse (SQL DB, Document DB)

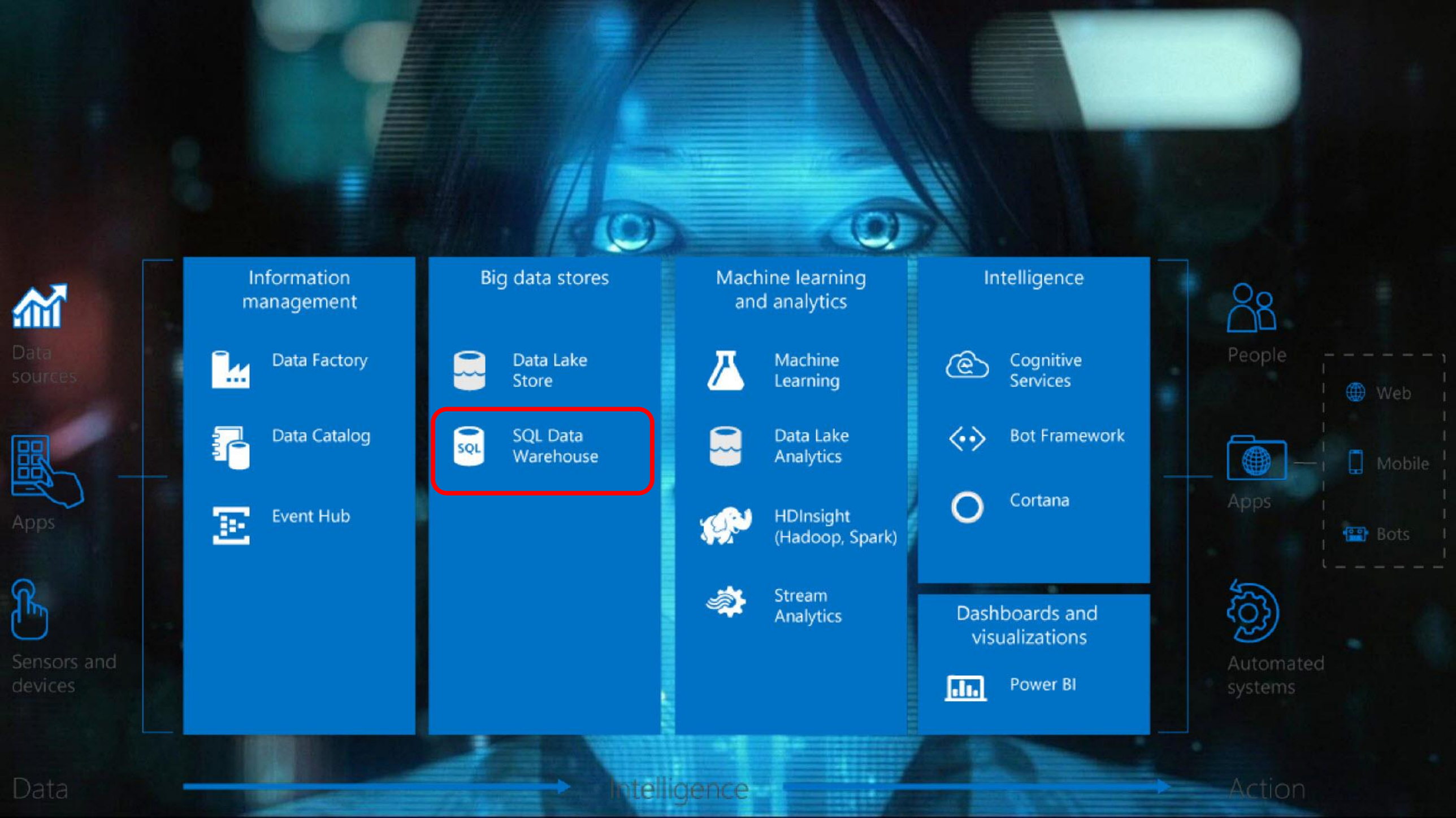
Data Lake

Event Hubs

Data Factory

Data Catalog

Microsoft Azure


Data sources
Apps
Sensors and devices
Information management
Data Factory
Data Catalog
Event Hub
Big data stores
Data Lake Store
SQL Data Warehouse
Machine learning and analytics
Machine Learning
Data Lake Analytics
HDInsight (Hadoop, Spark)
Stream Analytics
Intelligence
Cognitive Services
Bot Framework
Cortana
Dashboards and visualizations
Power BI
People
Apps
Web
Mobile
Bots
Automated systems
Data
Intelligence
Action

Introducing SQLDW

What is Azure SQL Data Warehouse

- A relational data warehouse-as-a-service (PaaS), fully managed by Microsoft.
- Industries first elastic cloud data warehouse with enterprise-grade capabilities.
- Support your smallest to your largest data storage needs while handling queries up to 100x faster.

Elastic scale & performance

Scales to petabytes of data

Massively Parallel Processing

Instant-on compute scales in seconds

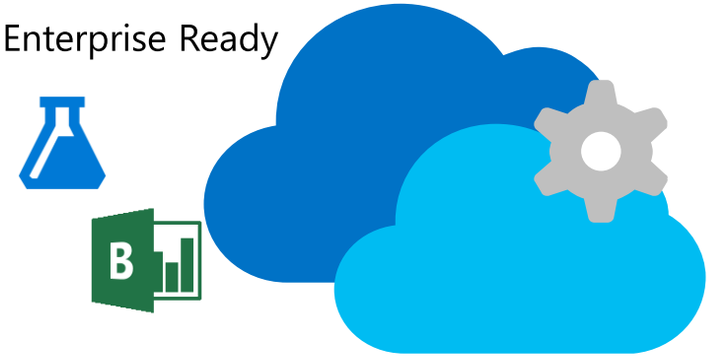
Query Relational / Non-Relational

Powered by the Cloud

Get started in minutes

Integrated with Azure ML, PowerBI & ADF

Enterprise Ready



Market Leading Price & Performance

Simple billing compute & storage

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

Bring DW to the Cloud without rewriting

What is MPP?

MPP stands for “MASSIVE PARALLEL PROCESSING”

- A divide and conquer strategy
- Take one big problem & break it up & execute it individually
- Team approach “Many hands make light work”

Requires

- A method for scheduling tasks
- A communication plan to maximise efficiency
- A distribution method for exchange of goods

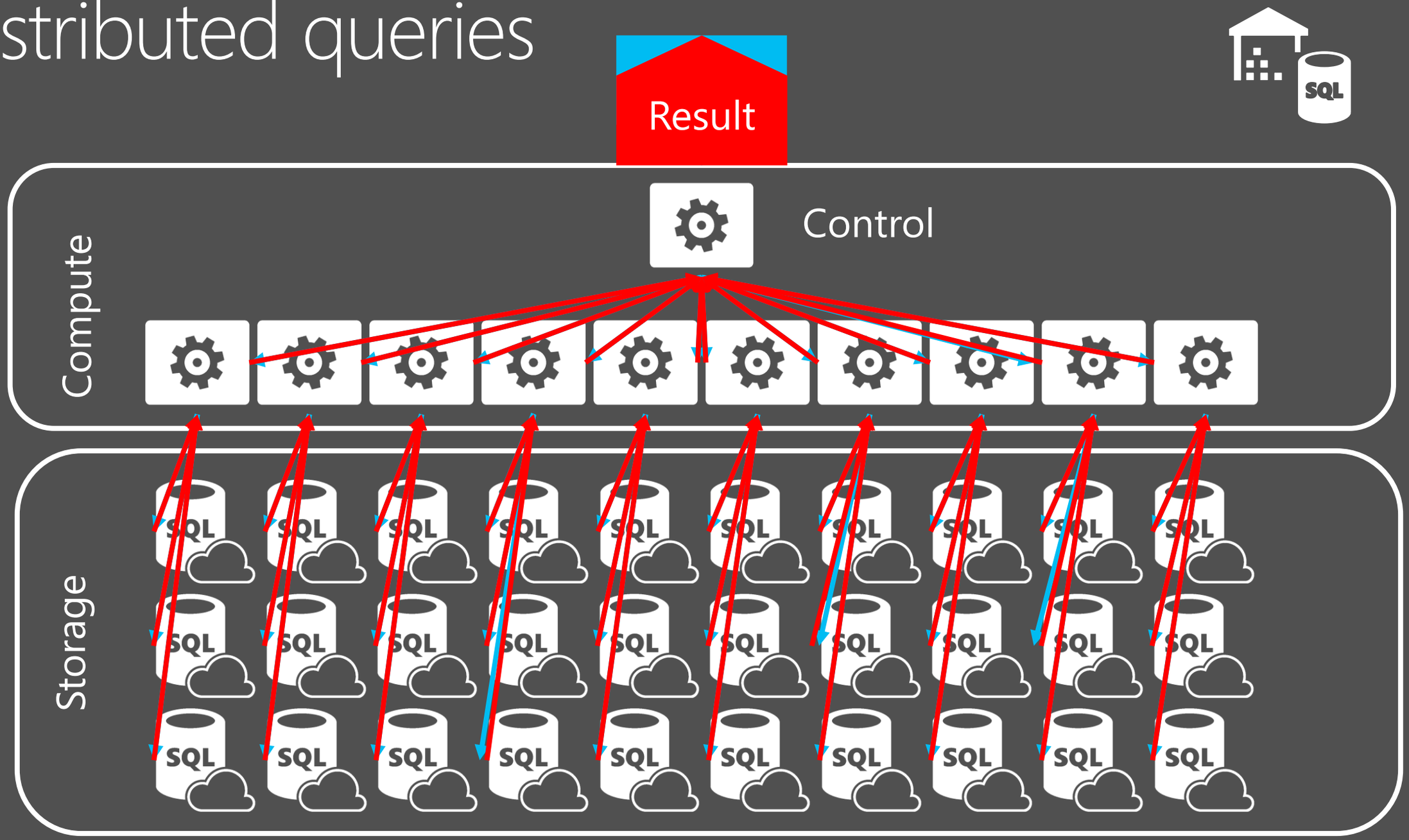
Logical Overview

Compute

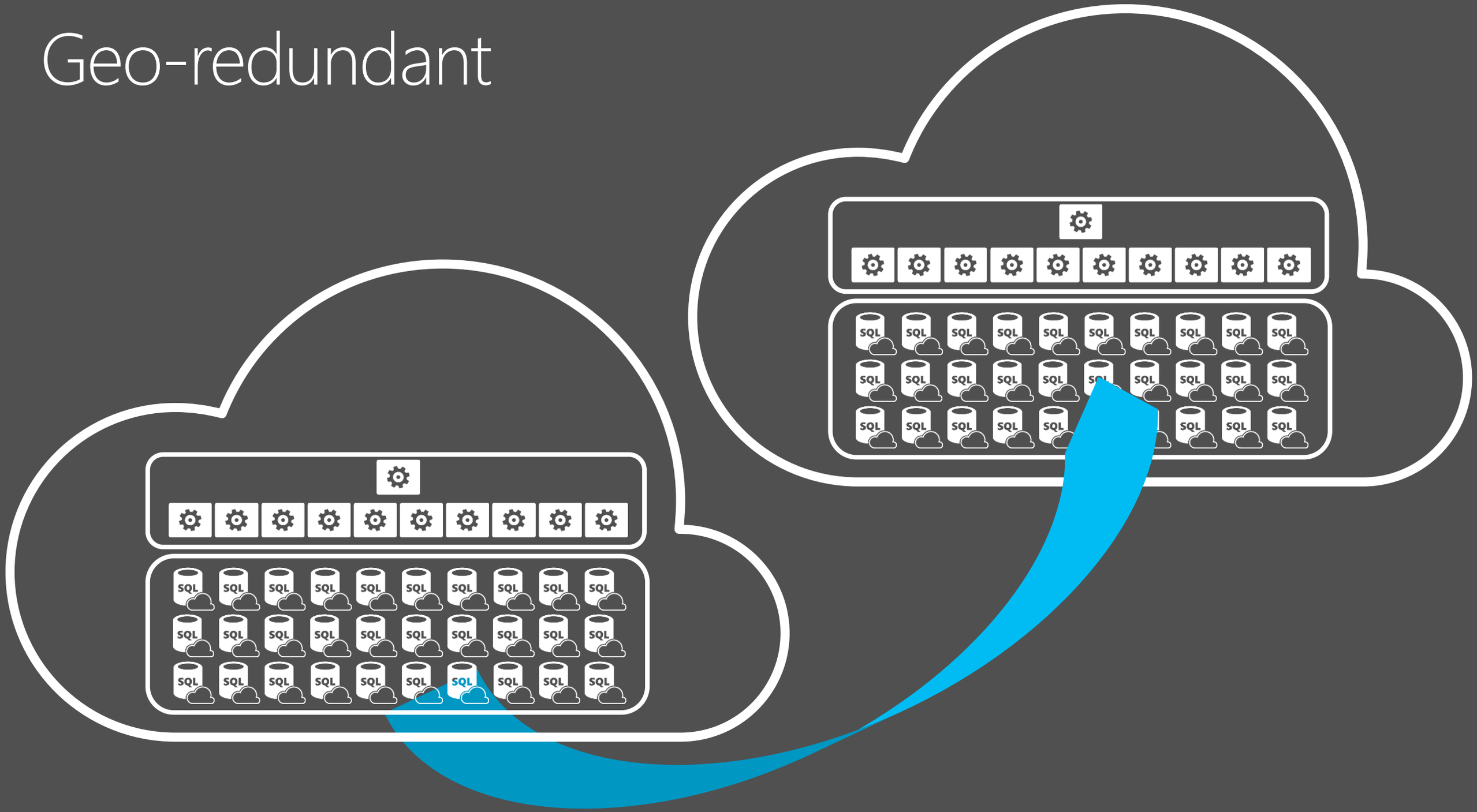
Control

Storage

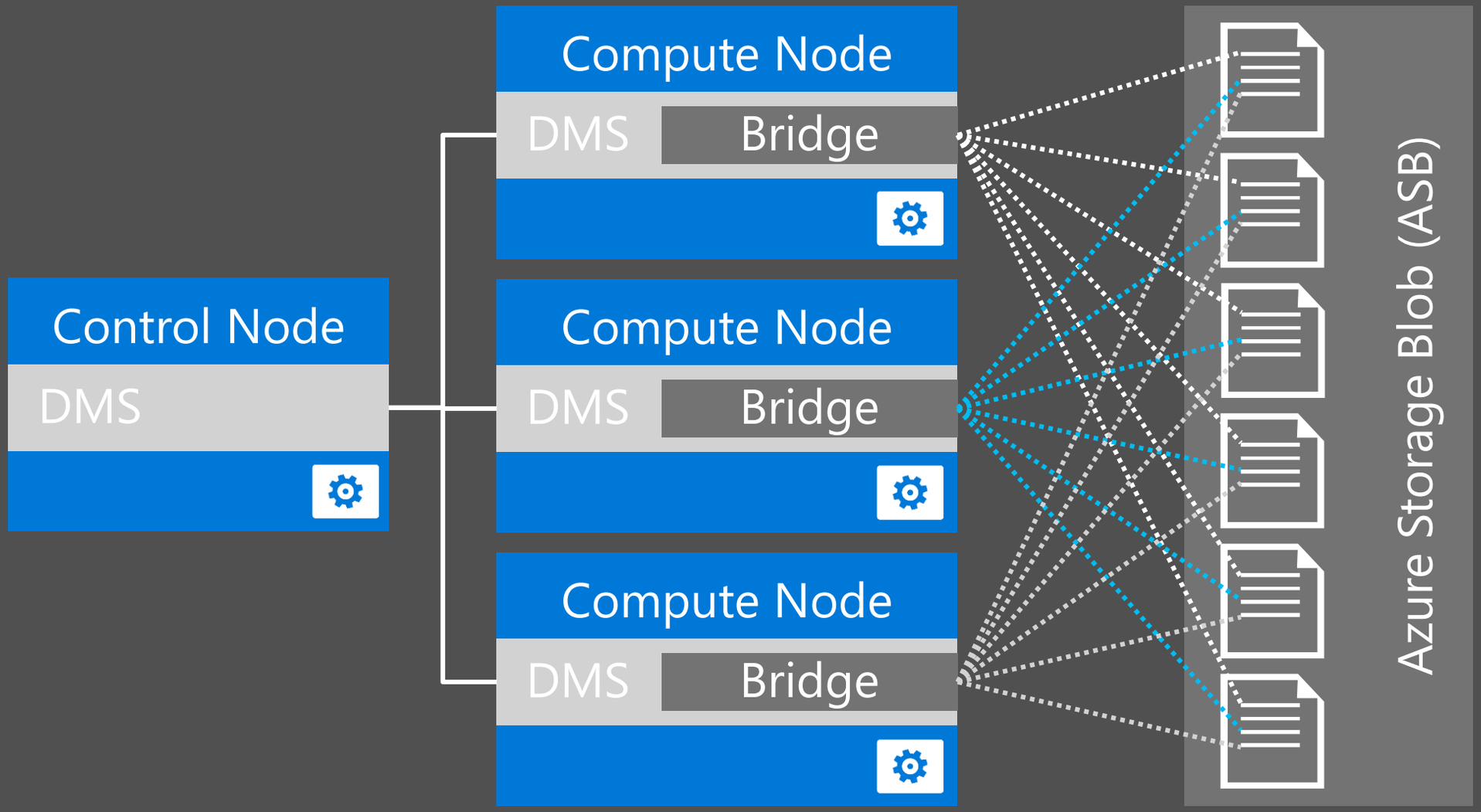
Distributed queries



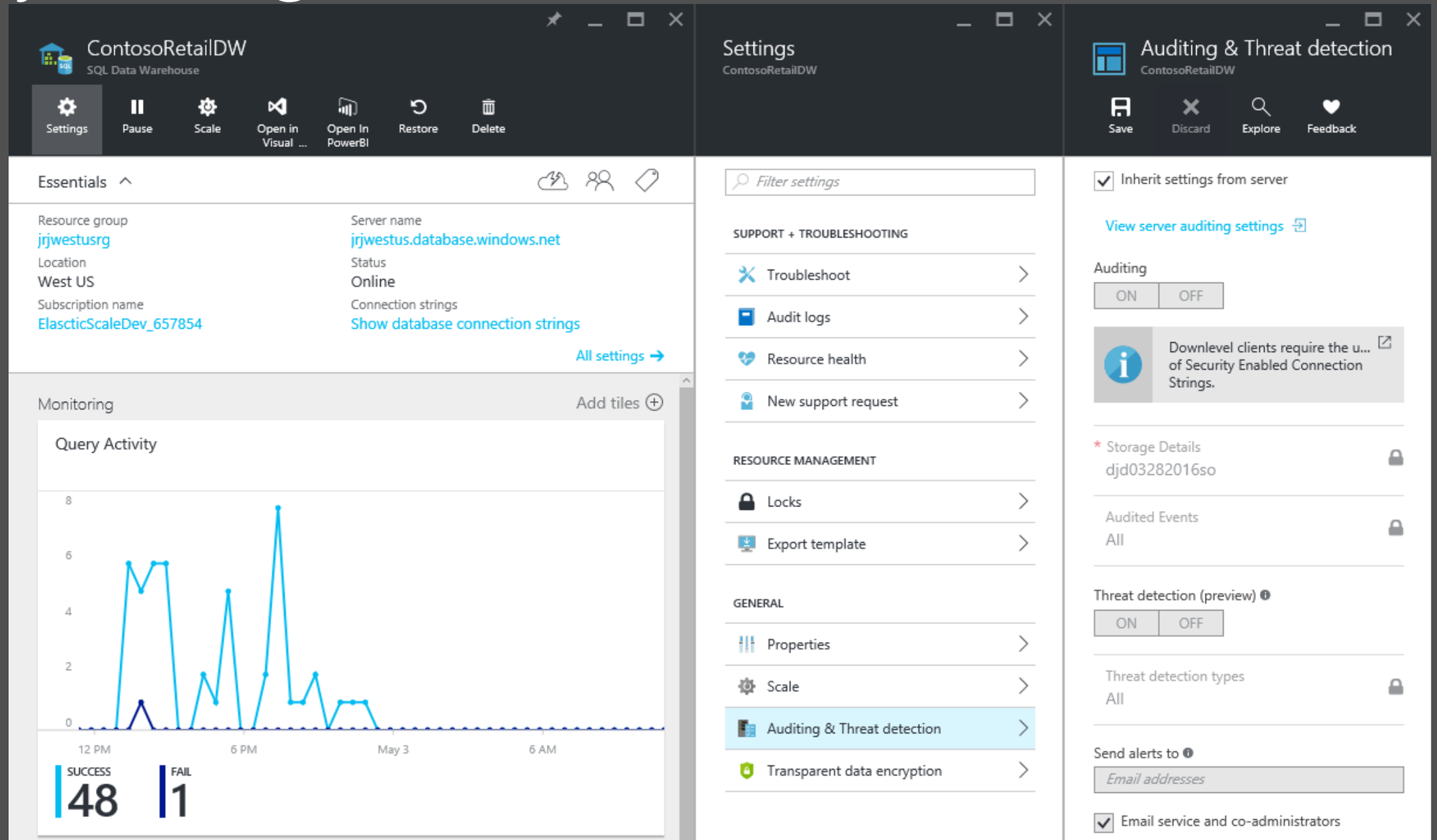
Geo-redundant



Parallel Load



Fully managed PaaS



Connectivity

Windows or Linux

ODBC

OLEDB

JDBC

ADO.NET

PHP

Target workloads

Analytical workloads

- 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

Operational workloads (OLTP)

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

Data Preparation

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

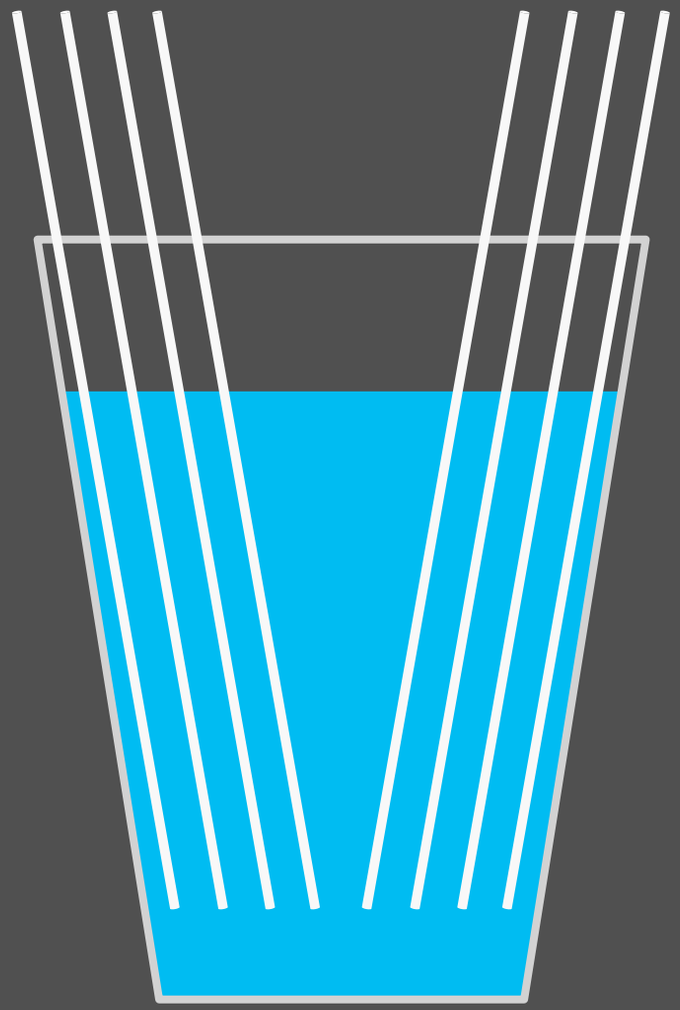
Demo:
Creating a SQL DW Instance

Scaling up vs. scaling out

Scaling up

- One bucket (motherboard)
- Contains all the water (resources)
- Drinking through straws (logical procs)
- Sometimes you only get one straw…

SMP = Scaling UP



Scaling out: The ultimate team game…

MPP
=
Scaling
OUT

Scaling out: The ultimate team drinking game…

01 

02 

03 

04 

Scaling Compute (elastically)

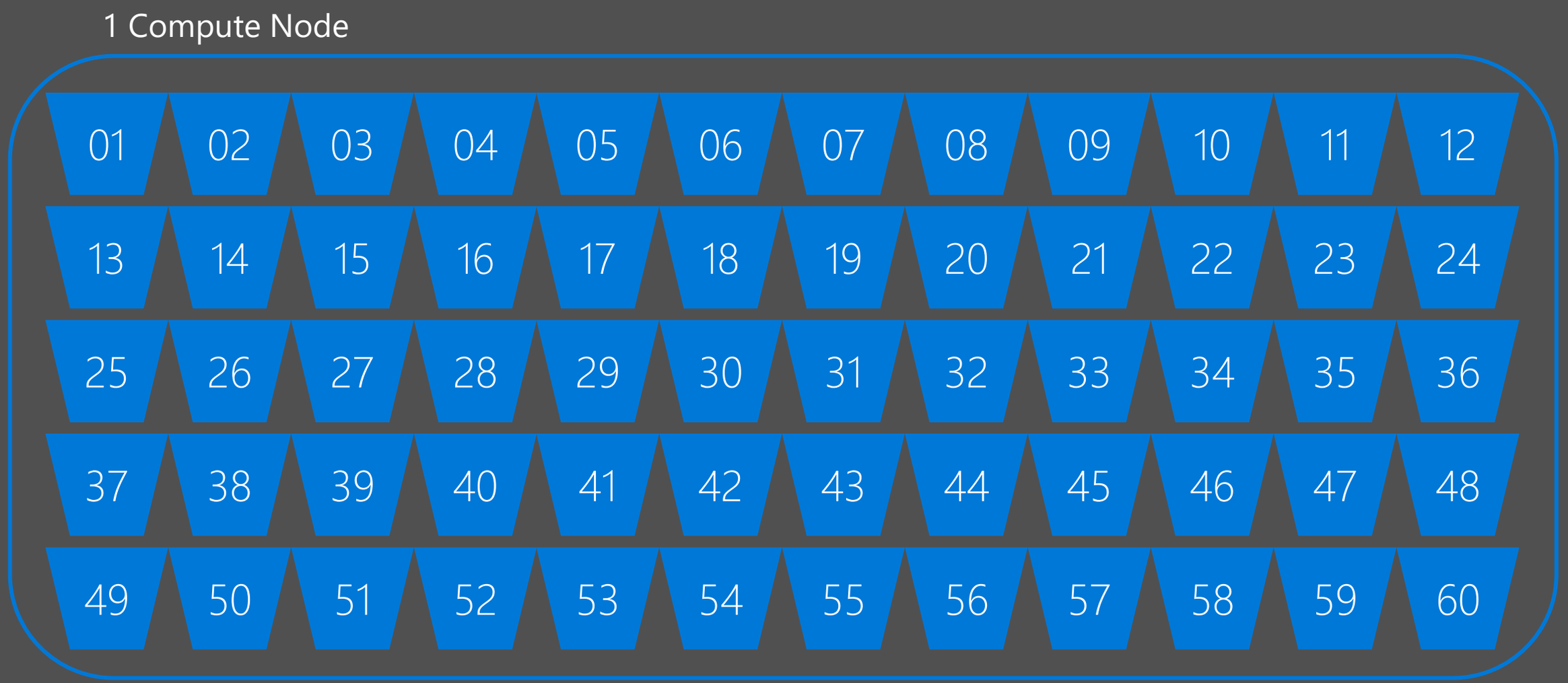
Key words

Nodes

Distributions

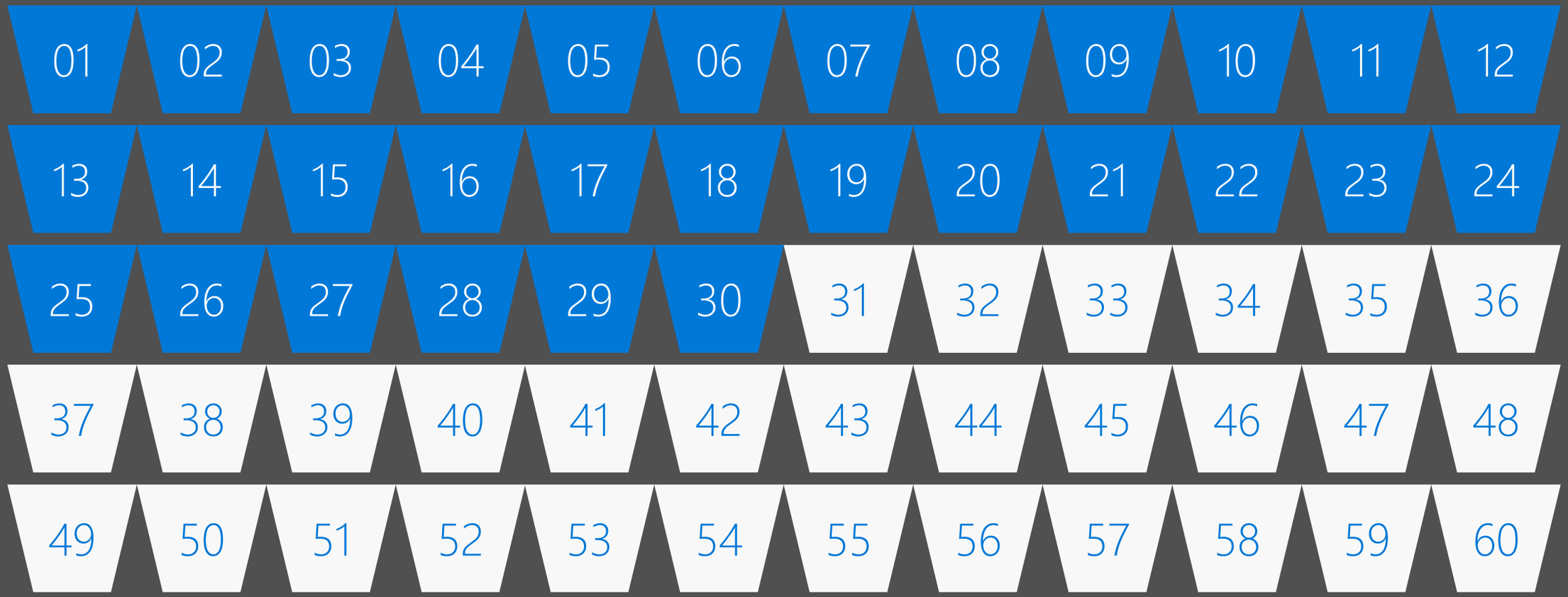
Service Objective

Mapping Compute in SQLDW



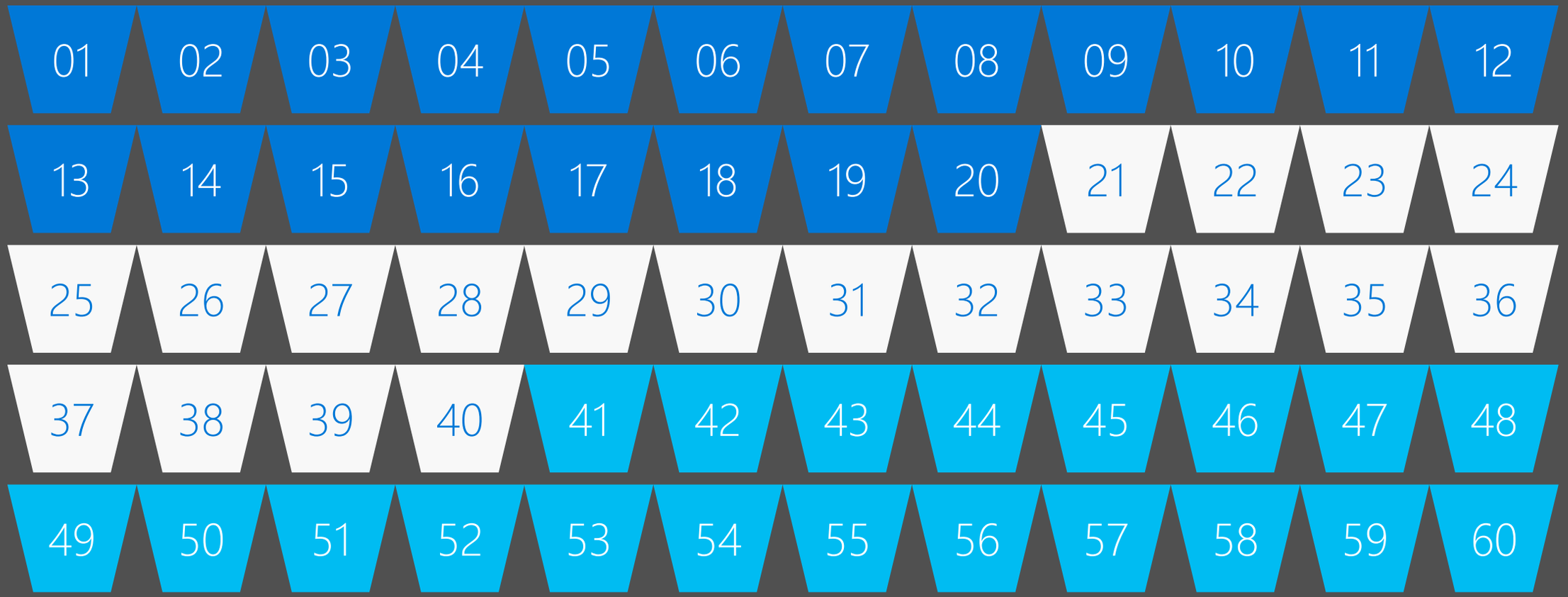
Mapping Compute in SQLDW

2 Compute Nodes



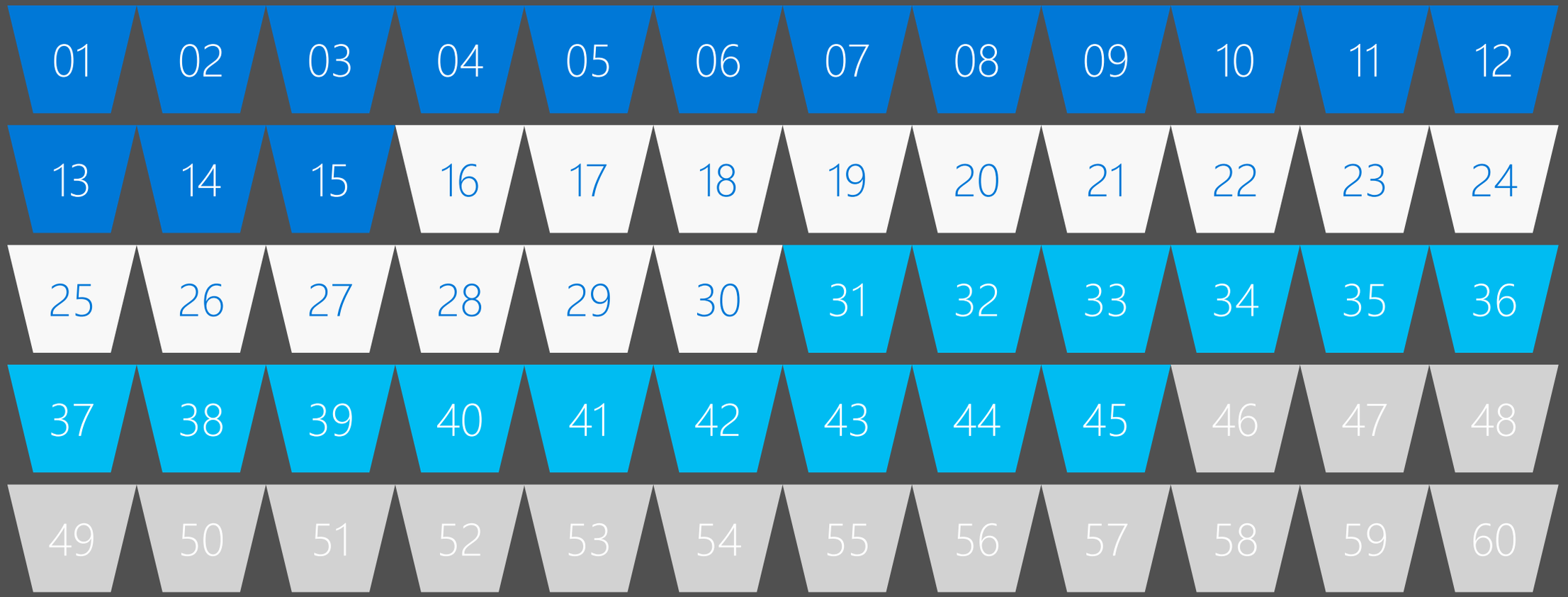
Mapping Compute in SQLDW

3 Compute Nodes



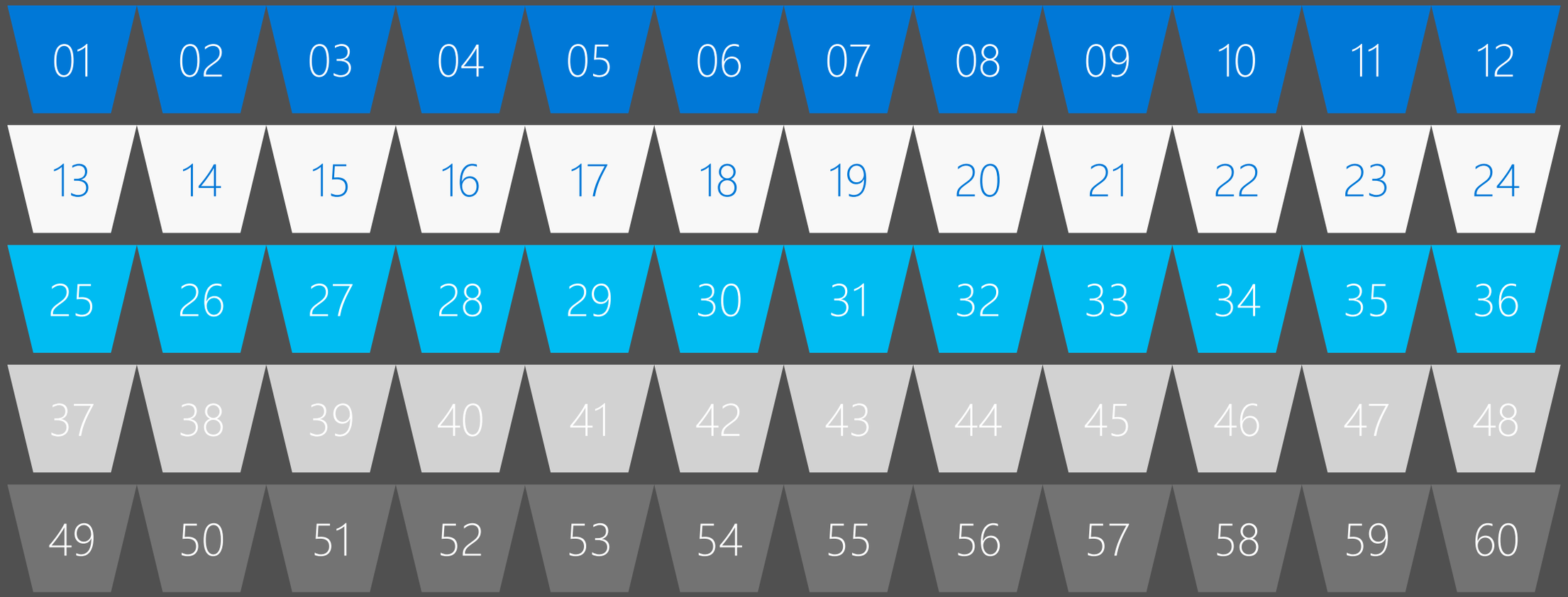
Mapping Compute in SQLDW

4 Compute Nodes



Mapping Compute in SQLDW

5 Compute Nodes



Nodes:

```
SELECT   [pdw_node_id]    AS node_id
,        [type]           AS node_type
,        [name]           AS node_name
FROM     sys.[dm_pdw_nodes]
;
```

Distributions:

```
SELECT   [distribution_id]    AS dist_id
,        [pdw_node_id]        AS node_id
,        [name]               AS dist_name
,        [position]           AS dist_position
FROM     sys.[pdw_distributions]
;
```

Service objective

```
SELECT   db.[name]                      AS [db_name]
,        ds.[edition]                   AS [db_edition]
,        ds.[service_objective] AS [db_slo]
FROM     sys.[database_service_objectives]   AS ds
JOIN     sys.[databases]                     AS db
ON       ds.[database_id] = db.[database_id]
WHERE    ds.[edition]   = 'DataWarehouse'
;
```

Determining DWU

```
SELECT   'DW'+CONVERT(VARCHAR(15),COUNT(*)*100)
FROM     sys.[dm_pdw_nodes]
WHERE    [type] =  'COMPUTE'
;
```

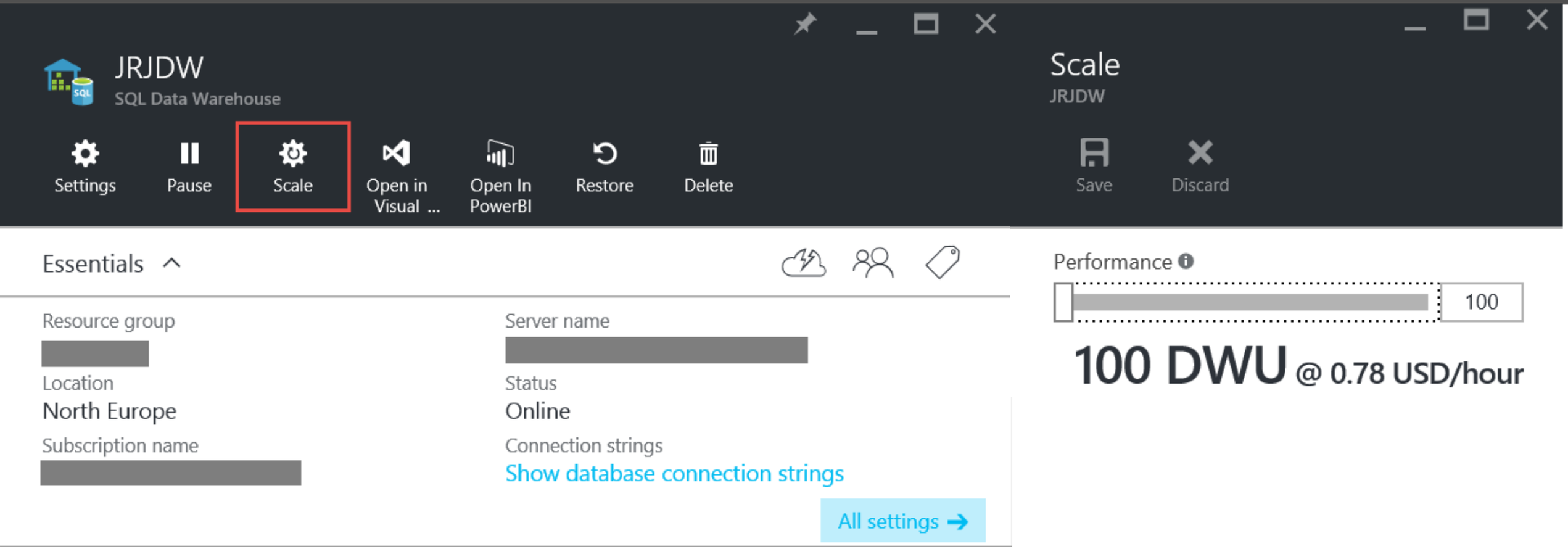
Scaling options

Azure Portal

TSQL

Powershell

Scaling with Azure Portal



Scaling with T-SQL

```
ALTER DATABASE ContosoRetailDW
MODIFY
(service_objective = 'DW100')
;
```

Scaling with PowerShell

```
Set-AzureRmSqlDatabase `
    -ResourceGroupName "RG_name" `
    -ServerName "SRV_name" `
    -DatabaseName "DB_name" `
    -RequestedServiceObjectiveName "DW100"
```

Sign in to Azure RM

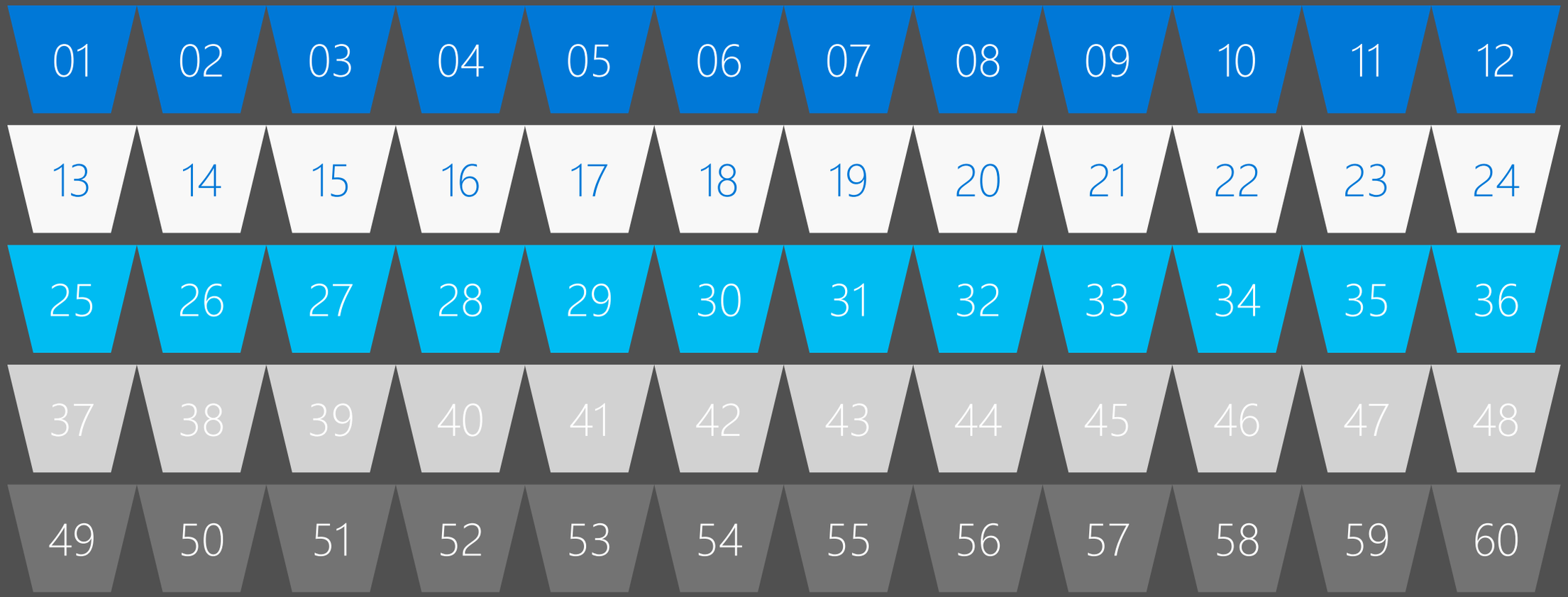
```
Login-AzureRmAccount
Get-AzureRmSubscription
Select-AzureRmSubscription `
-SubscriptionName "SUB_name"
```

Demo:
Scaling

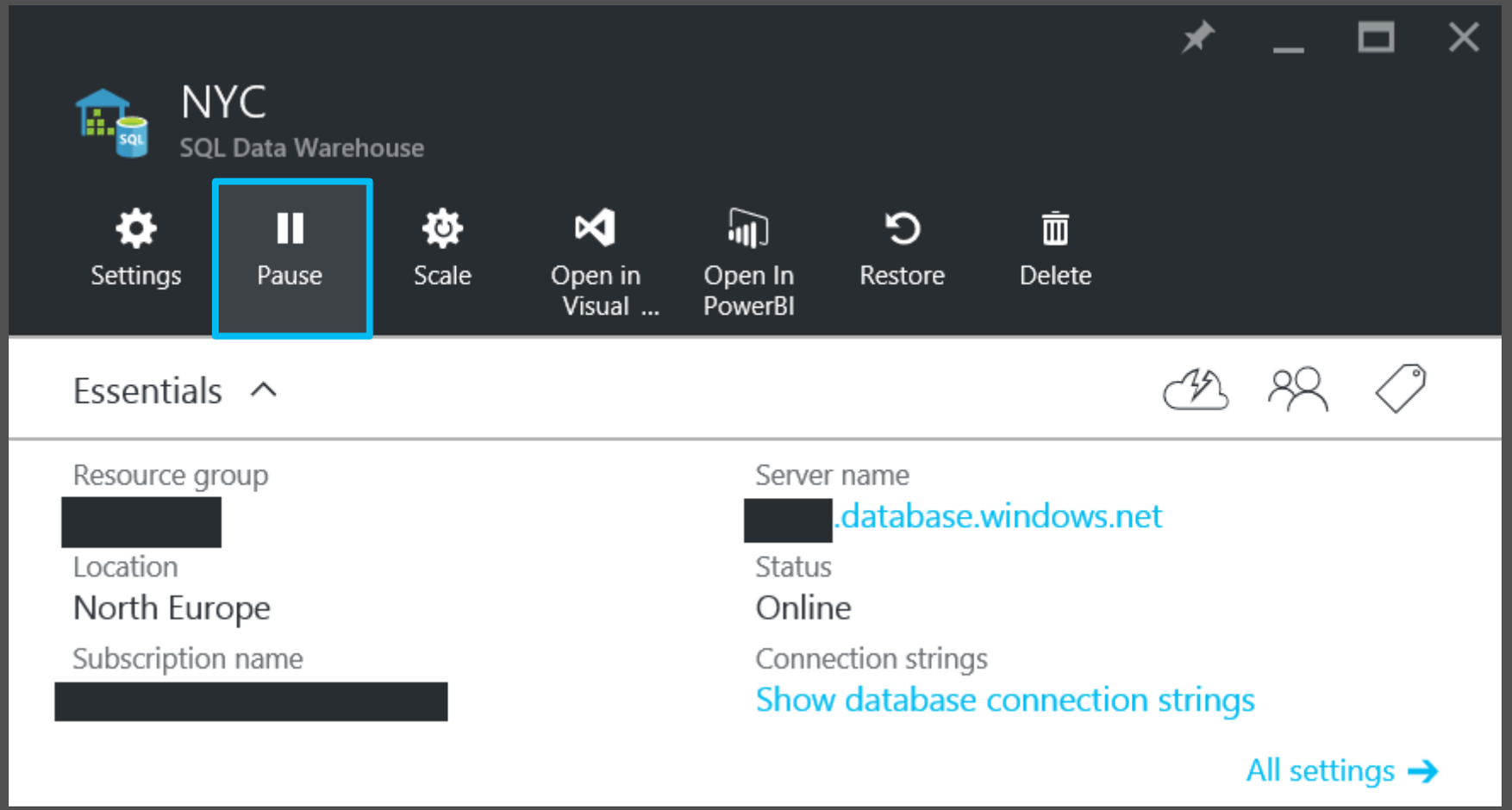
Pause and Resume

Pausing compute in SQLDW

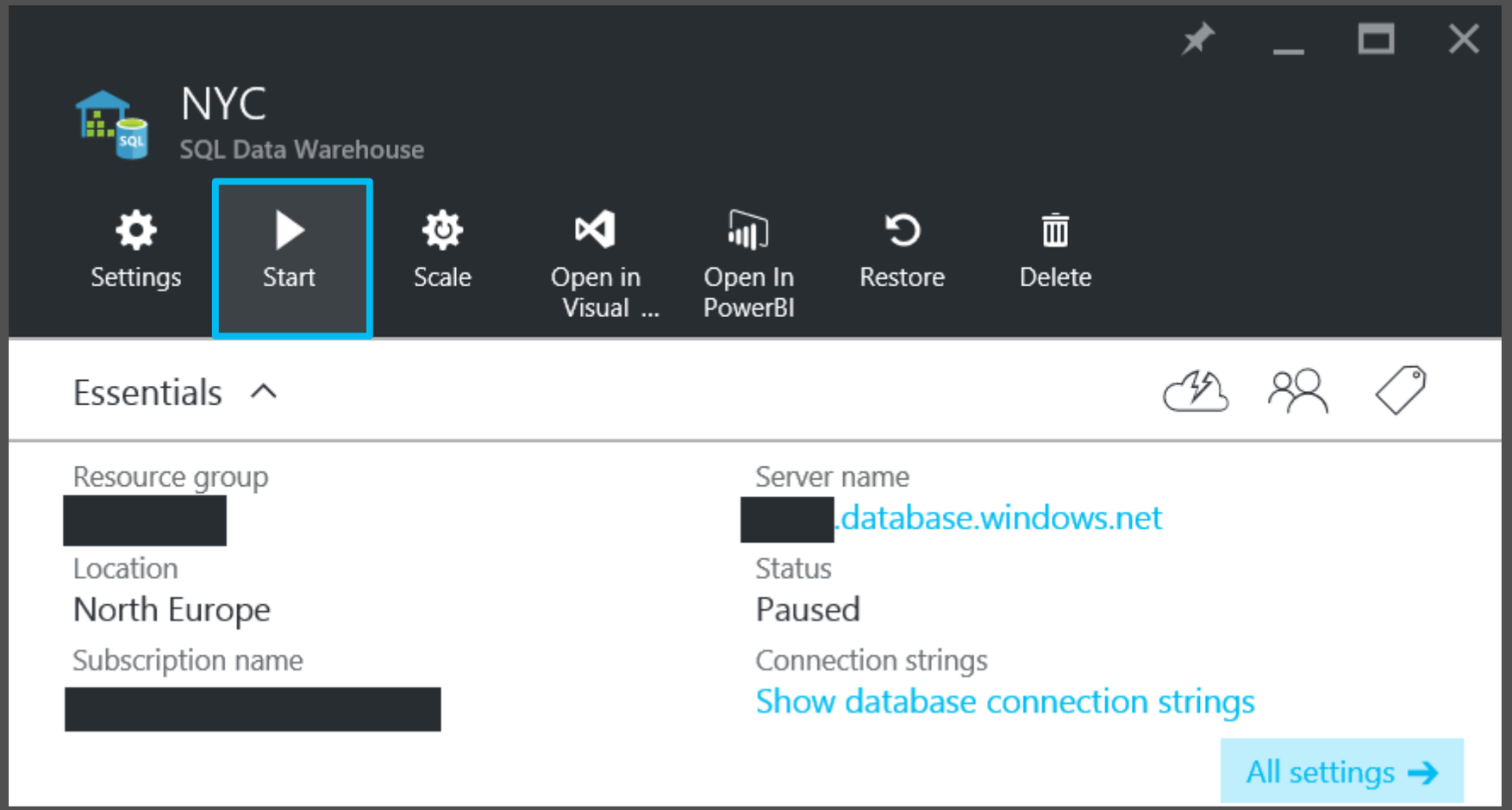
5 Compute Nodes



Pause with the portal



Resuming in the portal



Pause with PowerShell

```
Suspend-AzureRmSqlDatabase `
    -ResourceGroupName "RG_name" `
    -ServerName "SRV_name" `
    -DatabaseName "DB_name"
```

Resume with PowerShell

```
Resume-AzureRmSqlDatabase `
    -ResourceGroupName "RG_name" `
    -ServerName "SRV_name" `
    -DatabaseName "DB_name"
```

Demo:
Pause & Resume

Lab:

Creating Azure SQL Data Warehouse instance

Microsoft Azure

Lab review

1. Which category of the Cortana Intelligence Suite does SQL Data Warehouse fall under?
2. Name two key benefits to using an Azure SQL Data Warehouse over an on-premise data warehouse using SQL Server?
3. What are the advantages of using PowerShell over the Azure Portal to create a SQL Data Warehouse?
4. What is the maximum size of DWU that a SQL Data Warehouse can scale to?
5. When shouldn't you use Azure SQL Data Warehouse?

Summary

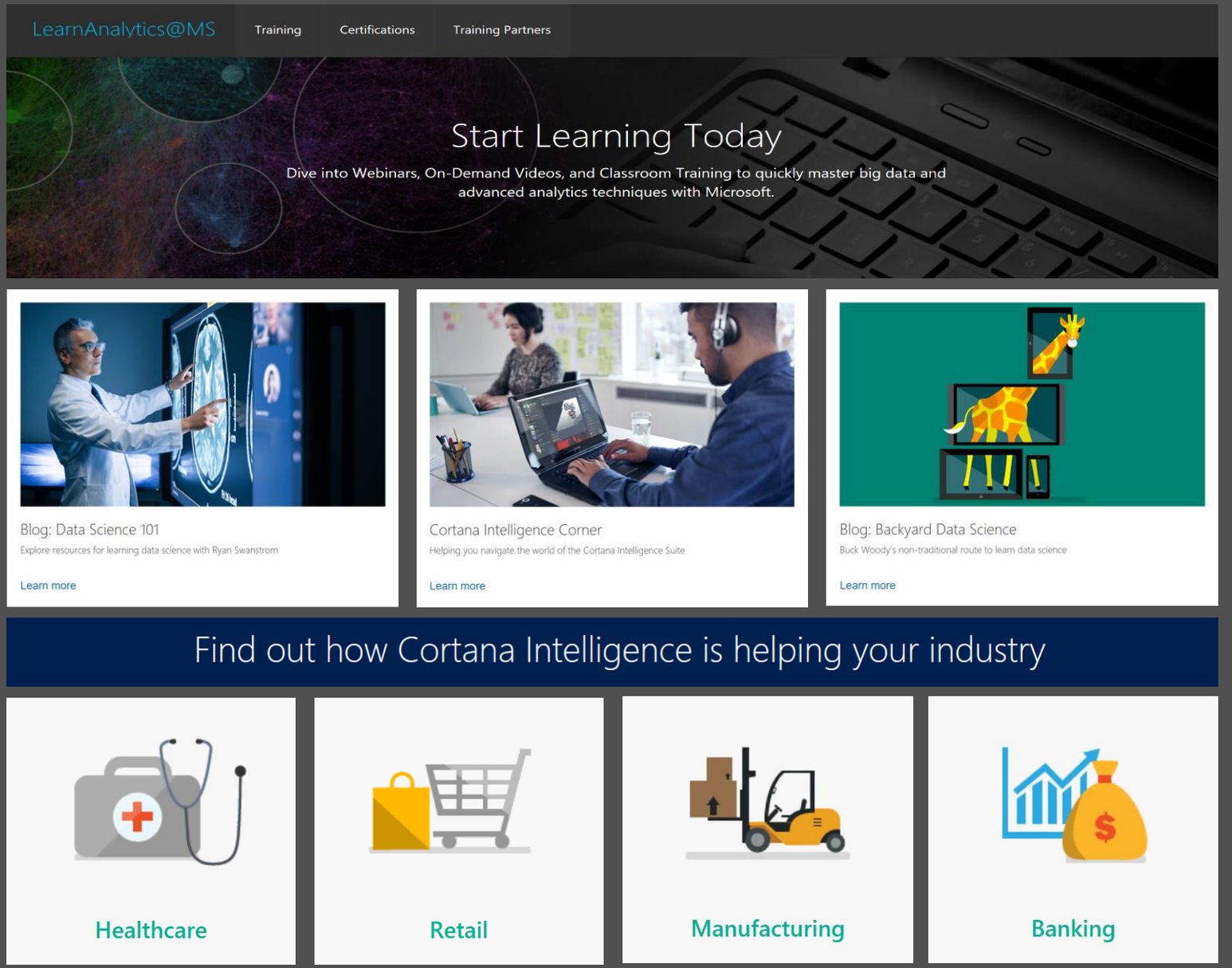
Summary

Scale-out distributed query engine
De-coupled storage from compute
Fully managed
Completely elastic
Platform as a Service (PaaS)
Petabyte scale
Leveraging cloud ecosystem
Broad range of connectivity options

Microsoft

There are more learning options as shown in the links on the right, including:

- Online training
- Videos
- Instructor Led training
- Blogs
- Cortana Intelligence Gallery



Course Documentation

Accessing the course materials

1. Click on the picture on the left.
2. Sign in with your Live ID.
3. Look for the SQLW301 item.
4. Click on the course materials link.

Microsoft