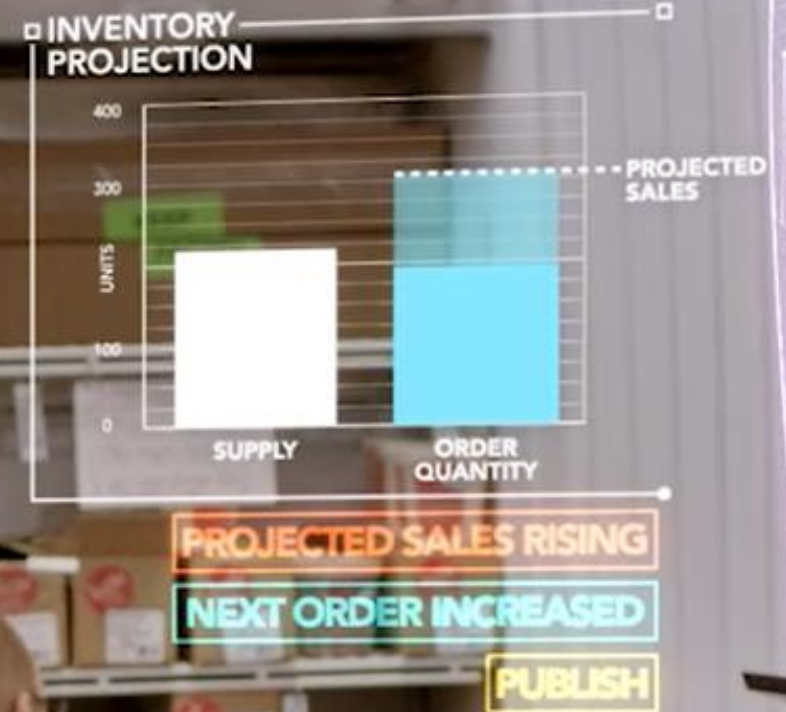


SQL Data Warehouse Key Concepts

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



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



Automated systems



Web



Mobile



Bots

Data

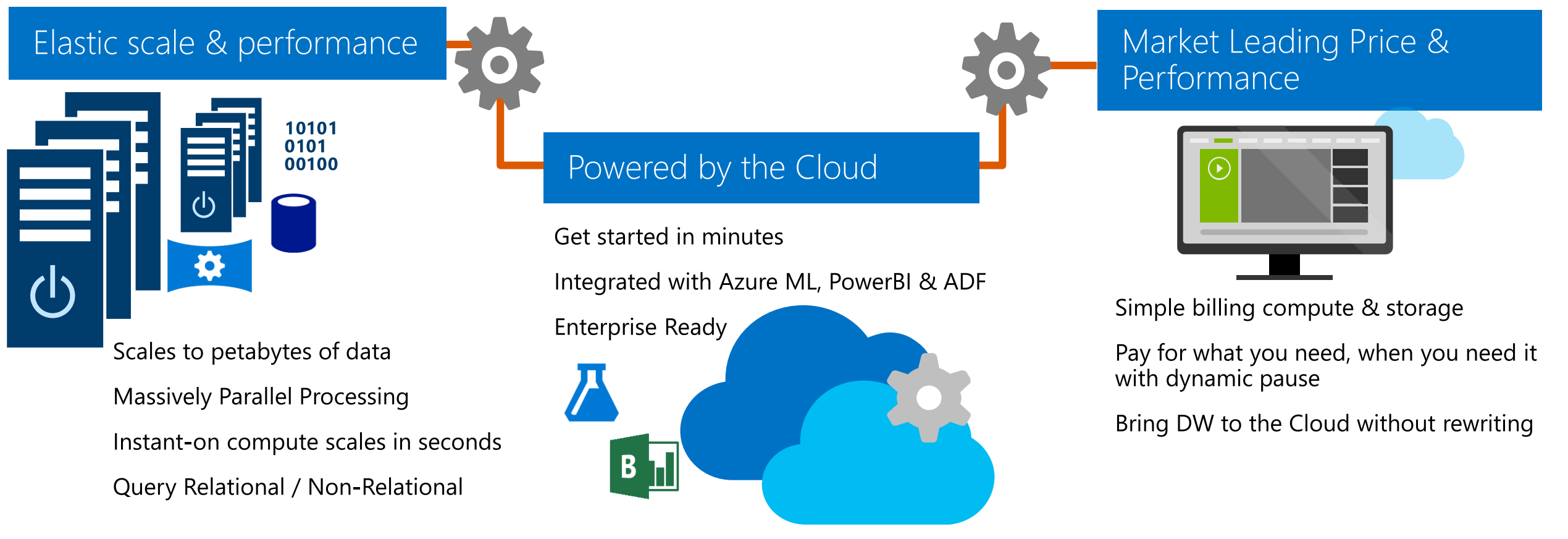
Intelligence

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.



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

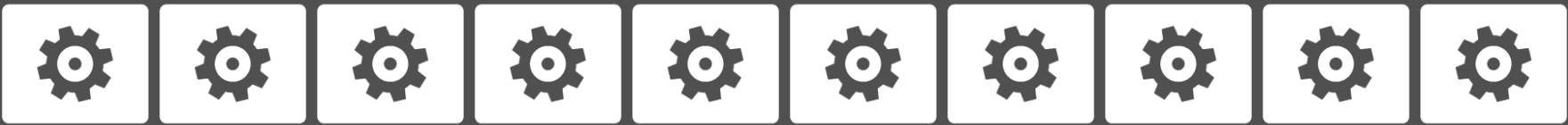
Logical Overview



Compute



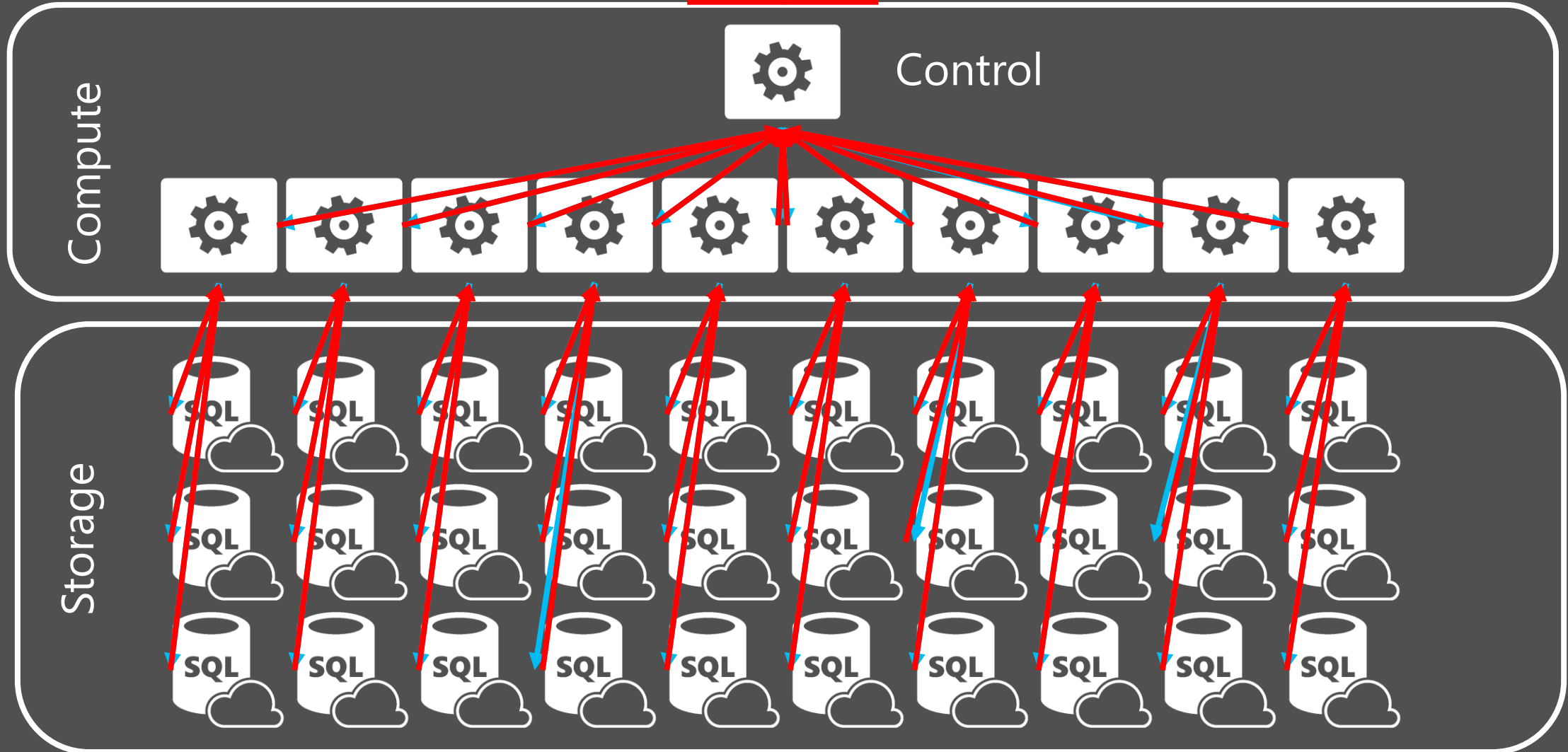
Control



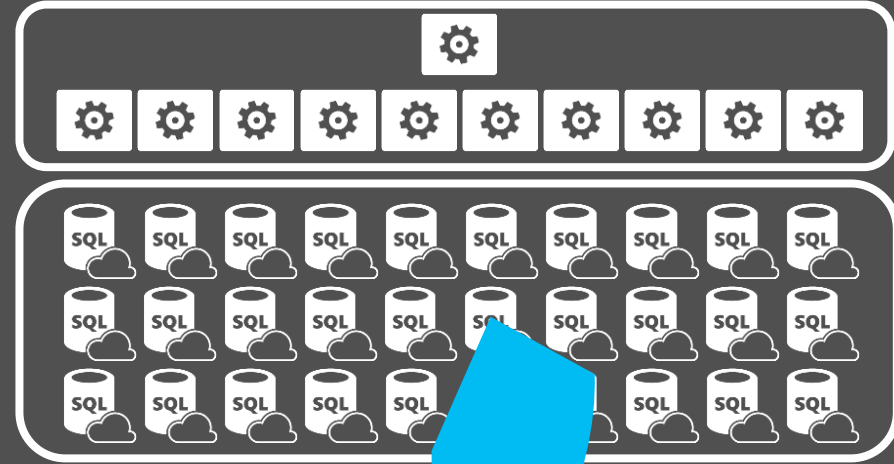
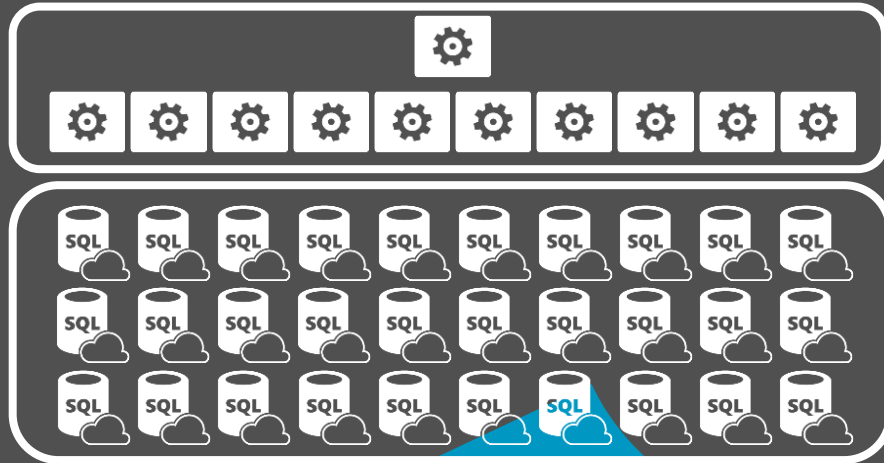
Storage



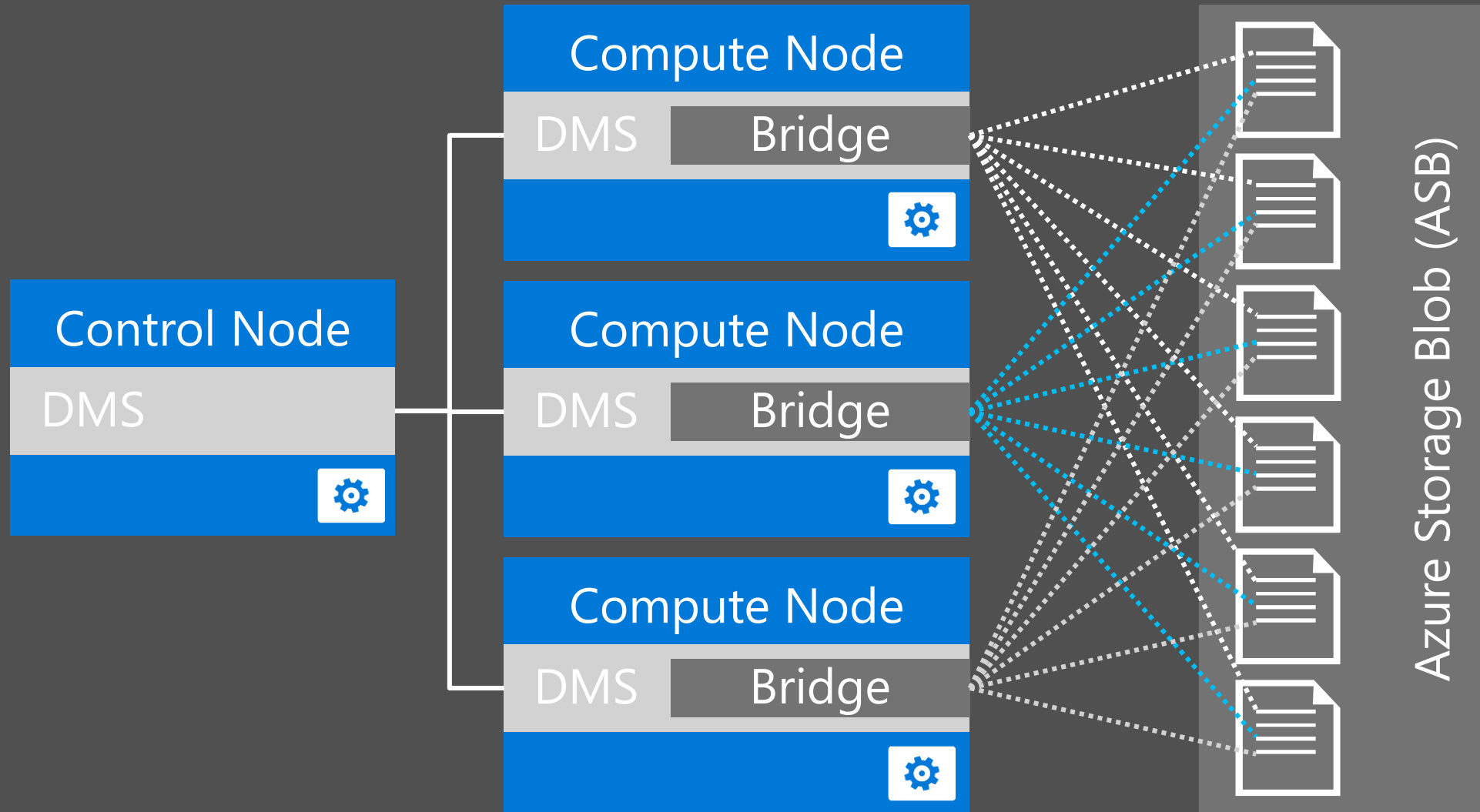
Distributed queries




Geo-redundant










Parallel Load



Fully managed PaaS

**ContosoRetailDW**
SQL Data Warehouse

 Settings  Pause  Scale  Open in Visual ...  Open In PowerBI  Restore  Delete

Essentials ^

Resource group
[jrjwestusrg](#)

Location
West US

Subscription name
[ElasticScaleDev_657854](#)

Server name
[jrjwestus.database.windows.net](#)

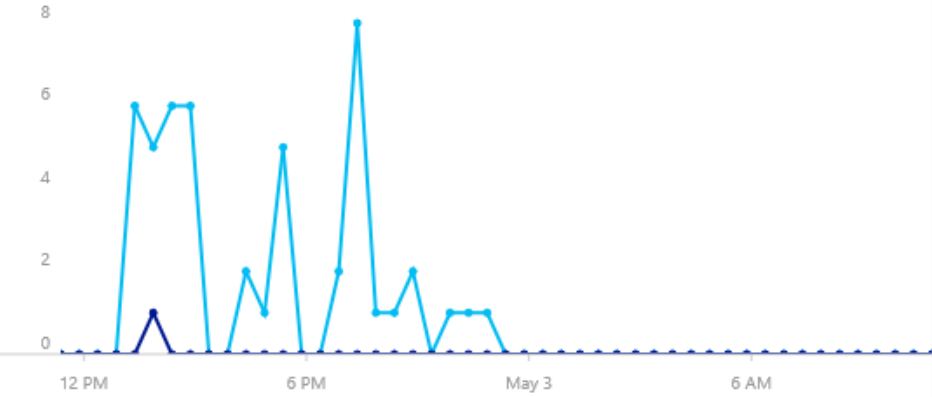
Status
Online

Connection strings
[Show database connection strings](#)

[All settings](#) →

Monitoring Add tiles +

Query Activity



SUCCESS

48

FAIL

1

Settings
ContosoRetailDW

SUPPORT + TROUBLESHOOTING

Troubleshoot

Audit logs

Resource health

New support request

RESOURCE MANAGEMENT

Locks

Export template

GENERAL

Properties

Scale

Auditing & Threat detection

Transparent data encryption

Auditing & Threat detection
ContosoRetailDW

Save

Discard

Explore

Feedback


☒ Inherit settings from server

[View server auditing settings](#)

Auditing

ON

OFF

 Downlevel clients require the u... of Security Enabled Connection Strings.

Storage Details

djd03282016so

Audited Events

All

Threat detection (preview)

ON

OFF

Threat detection types

All

Send alerts to

☒ Email service and co-administrators

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"

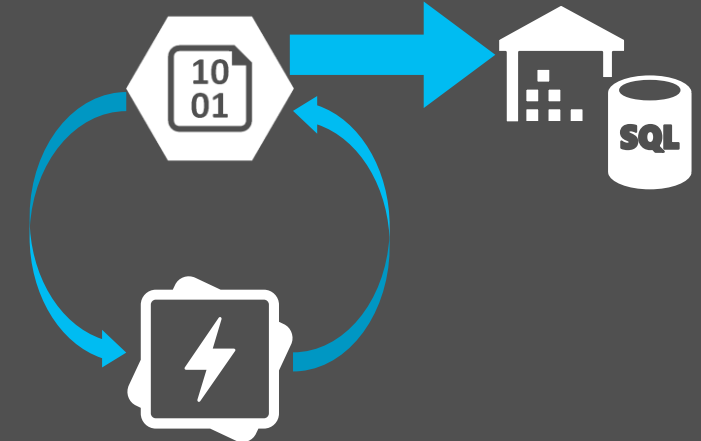
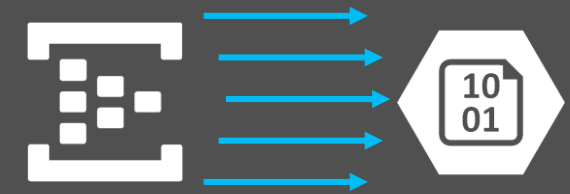
Unsuitable workloads

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:

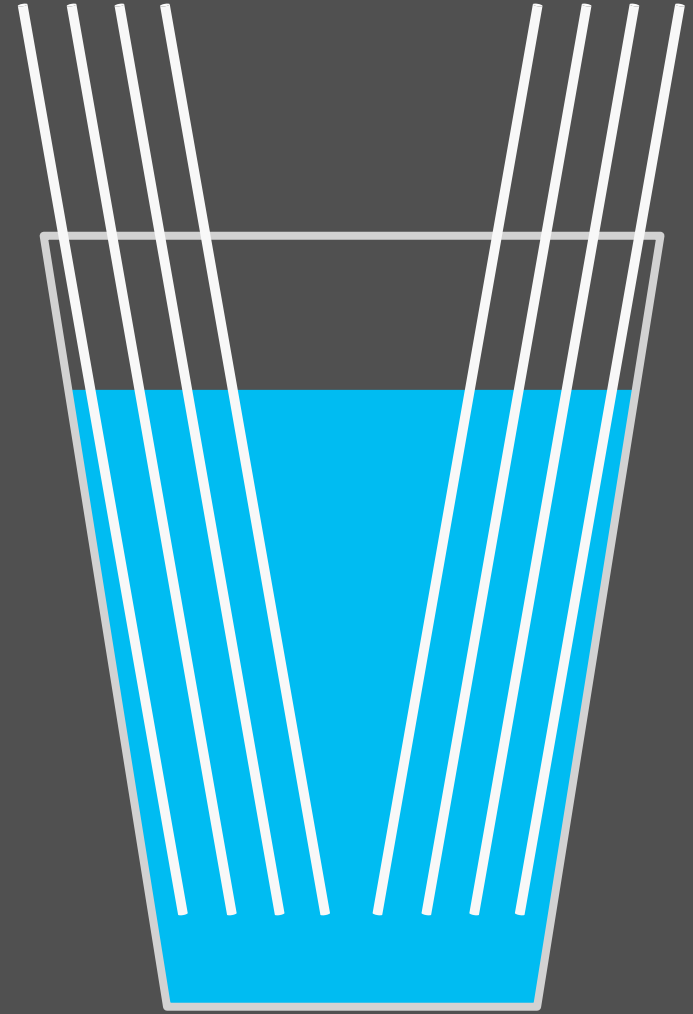
Creating a SQL DW Instance

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...



Scaling Compute (elastically)

Key words

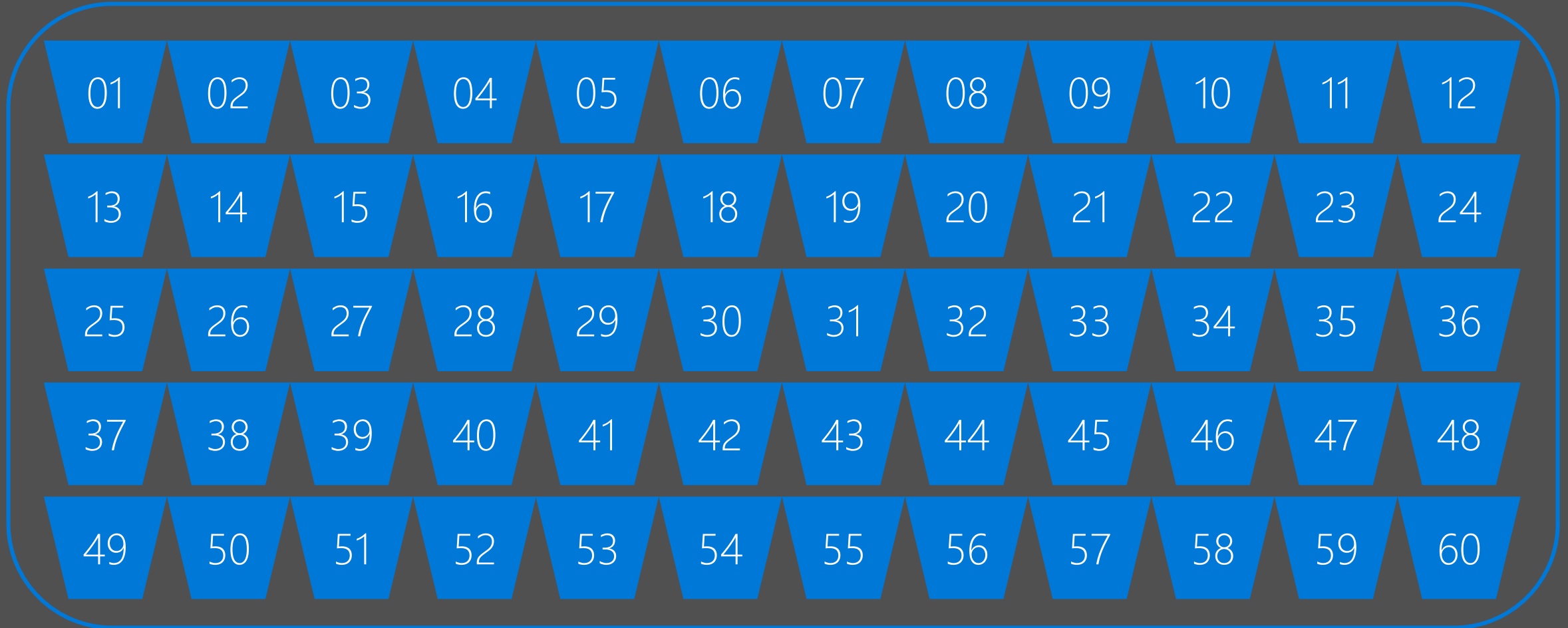
Nodes

Distributions

Service Objective

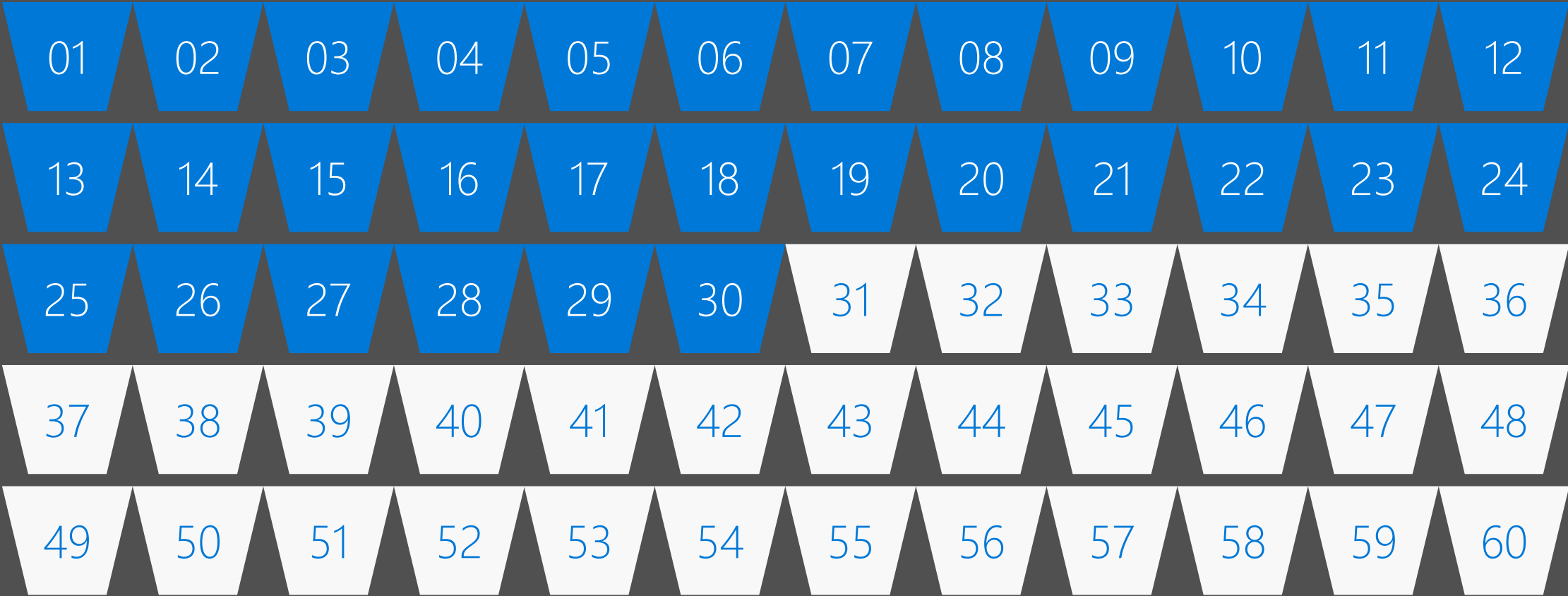
Mapping Compute in SQLDW

1 Compute Node



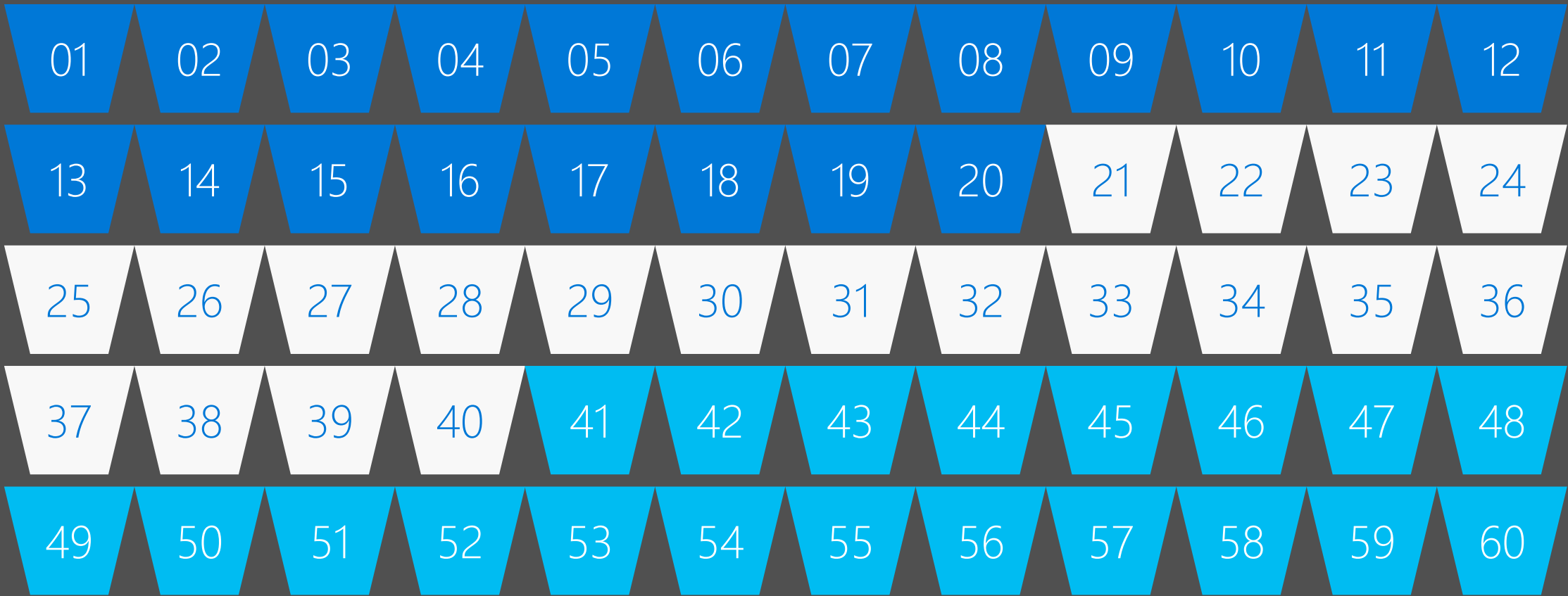
Mapping Compute in SQLDW

2 Compute Nodes



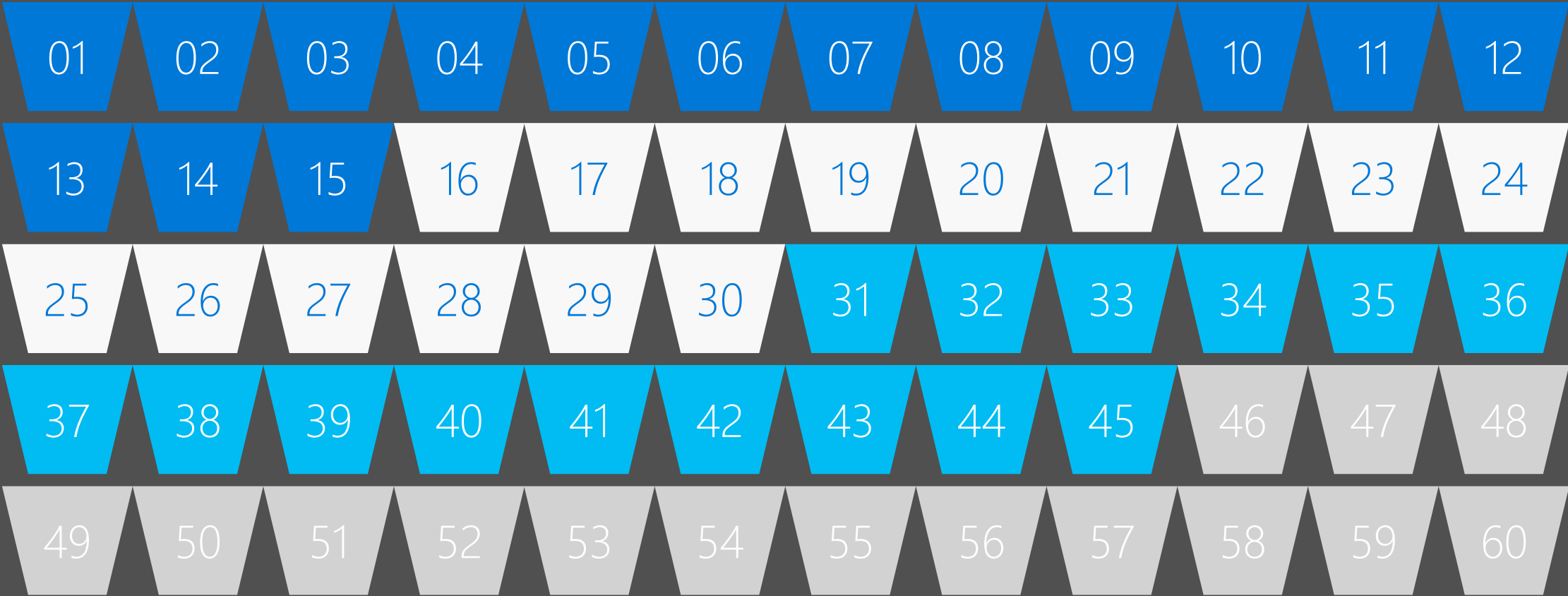
Mapping Compute in SQLDW

3 Compute Nodes



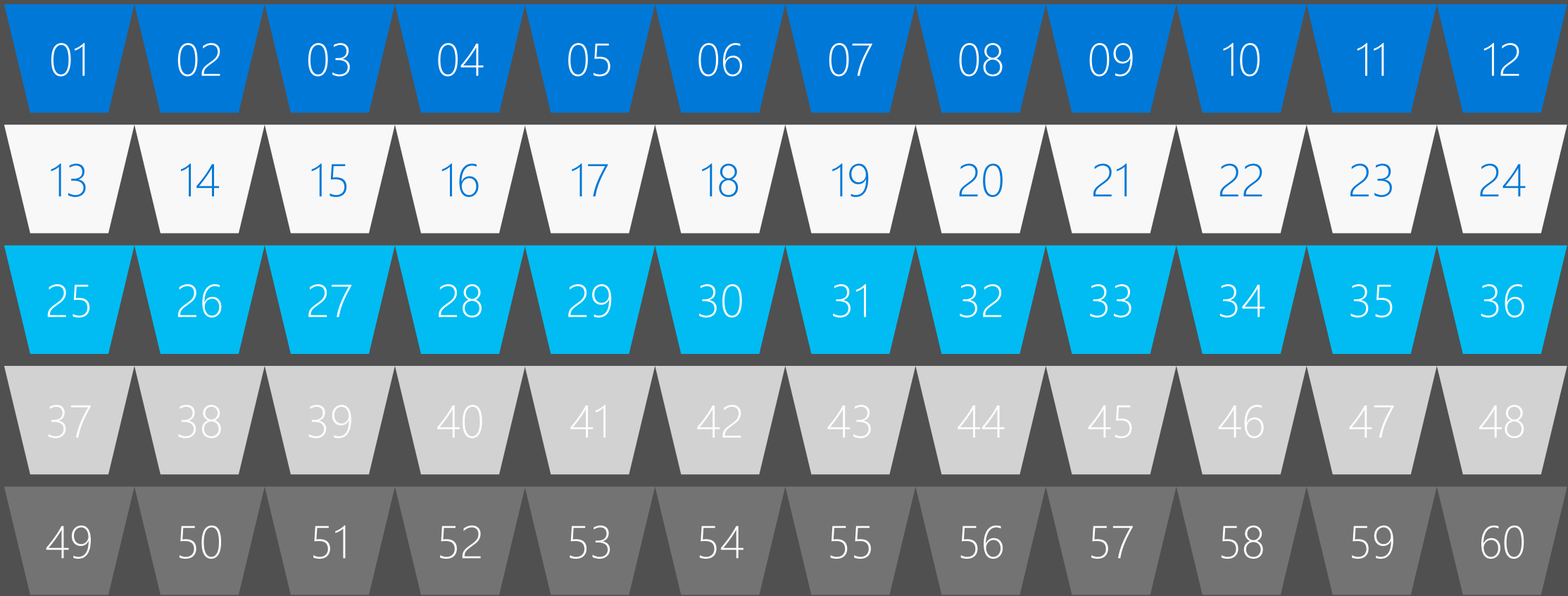
Mapping Compute in SQLDW

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

The screenshot displays the Azure Portal interface for a SQL Data Warehouse named 'JRJDW'. The top navigation bar includes icons for Settings, Pause, Scale (highlighted with a red box), Open in Visual Studio, Open In PowerBI, Restore, and Delete. The right side of the top bar shows the 'Scale' page title and 'JRJDW'.

Below the navigation bar, the 'Essentials' section is visible, showing details for the resource group, location (North Europe), and subscription name. The 'Server name' is also displayed, along with the status 'Online' and a link to 'Show database connection strings'. An 'All settings' button is located at the bottom right of the Essentials section.

On the right side, the 'Performance' section shows a slider set to 100 DWU, with the text '100 DWU @ 0.78 USD/hour' displayed below it.

Resource group: [Redacted]
Location: North Europe
Subscription name: [Redacted]

Server name: [Redacted]
Status: Online
Connection strings: [Show database connection strings](#)

[All settings →](#)

Performance ⓘ
100 DWU @ 0.78 USD/hour

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

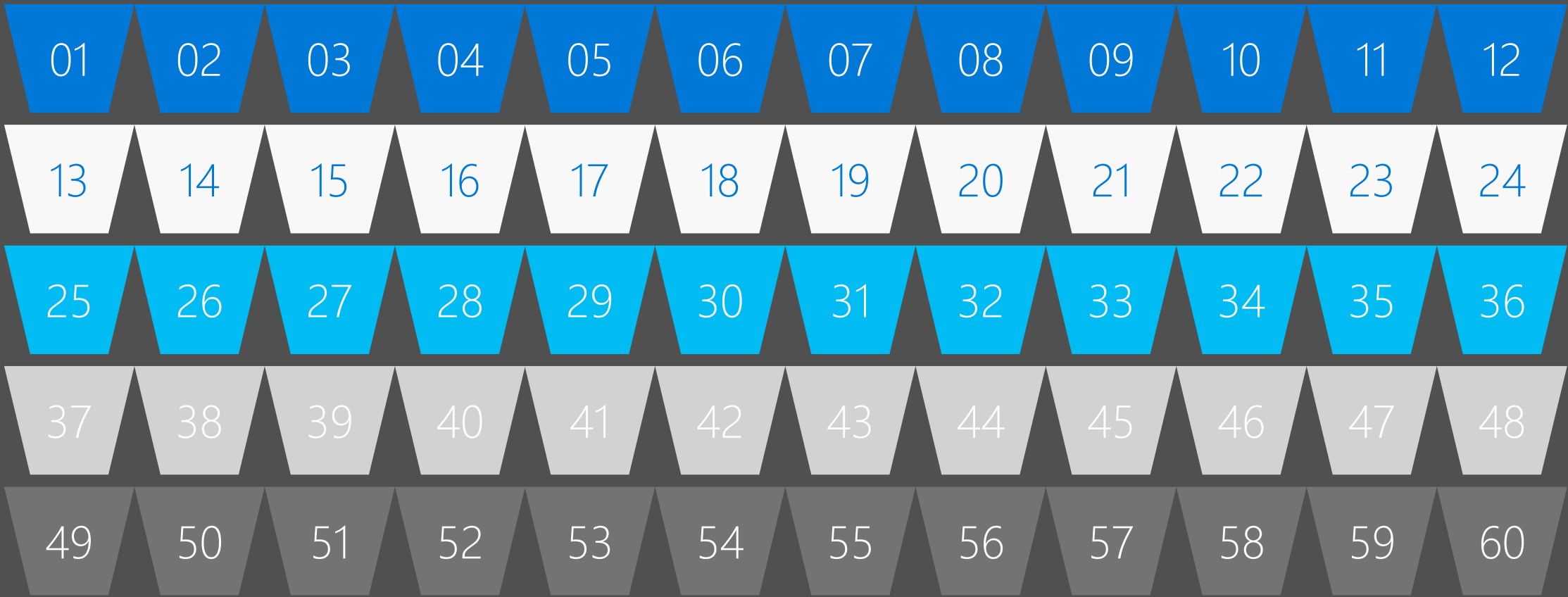
Select-AzureRmSubscription
-SubscriptionName "SUB_name"

Demo: Scaling

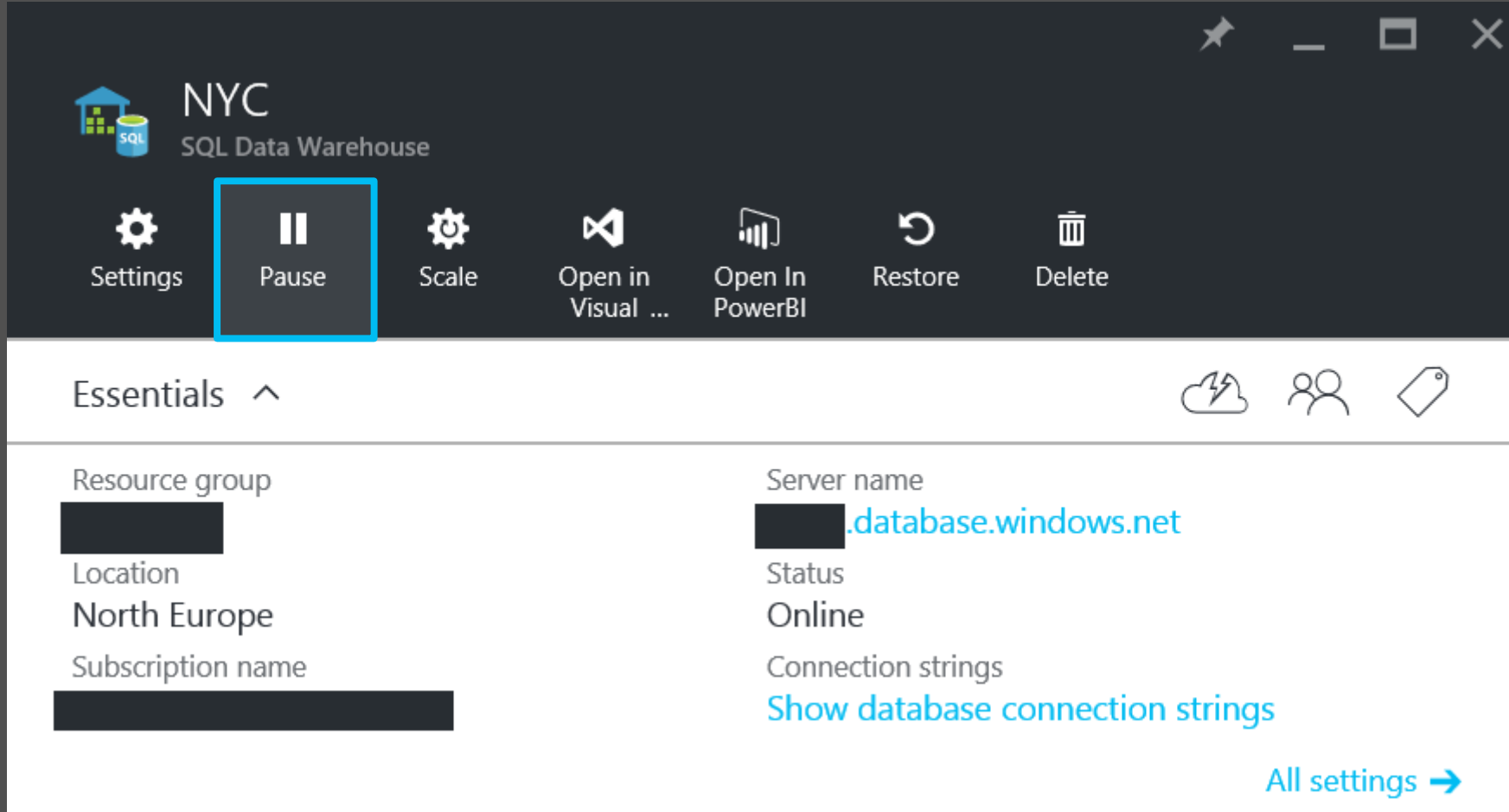
Pause and Resume

Pausing compute in SQLDW

5 Compute Nodes



Pause with the portal



The screenshot shows the Azure portal interface for a SQL Data Warehouse resource named 'NYC'. The top navigation bar includes icons for Settings, Pause (highlighted with a red box), Scale, Open in Visual Studio, Open In PowerBI, Restore, and Delete. Below the navigation bar, the 'Essentials' section displays key information about the resource group, location, subscription name, server name, status, and connection strings. The 'Pause' button is the primary action for stopping the data warehouse.

NYC
SQL Data Warehouse

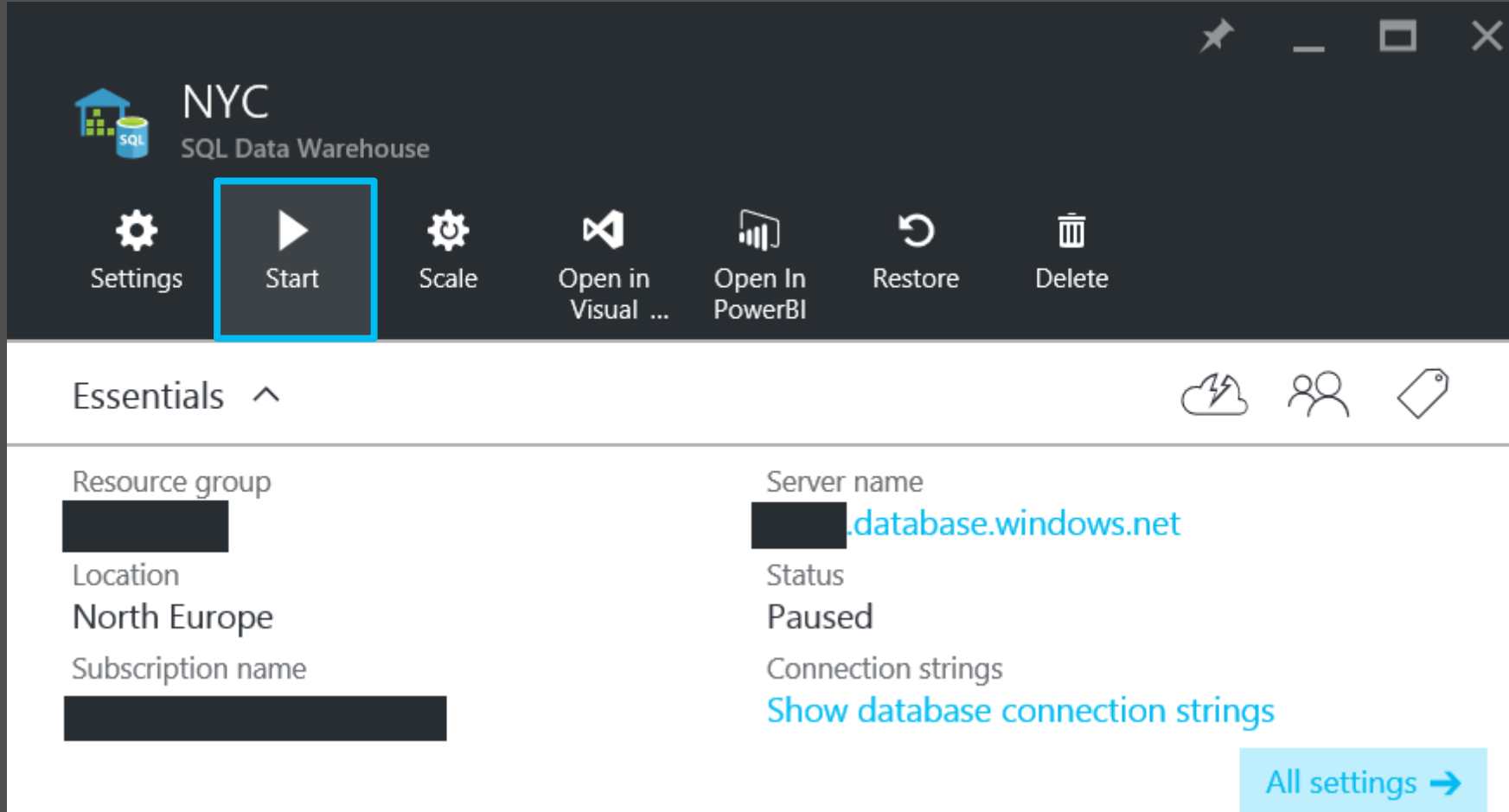
Settings **Pause** Scale Open in Visual ... Open In PowerBI Restore Delete

Essentials ^

Resource group	Server name
[REDACTED]	[REDACTED].database.windows.net
Location	Status
North Europe	Online
Subscription name	Connection strings
[REDACTED]	Show database connection strings

[All settings](#) →

Resuming in the portal



The screenshot displays the Azure portal interface for a SQL Data Warehouse named 'NYC'. The top navigation bar includes icons for Settings, Start (highlighted with a red box), Scale, Open in Visual Studio, Open In PowerBI, Restore, and Delete. Below the navigation bar, the 'Essentials' section shows the following details:

Resource group	Server name
[REDACTED]	[REDACTED].database.windows.net
Location	Status
North Europe	Paused
Subscription name	Connection strings
[REDACTED]	Show database connection strings

An 'All settings' button with a right arrow is located at the bottom right of the Essentials section.

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


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

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

[LearnAnalytics@MS](#) [Training](#) [Certifications](#) [Training Partners](#)


Start Learning Today

Dive into Webinars, On-Demand Videos, and Classroom Training to quickly master big data and advanced analytics techniques with Microsoft.



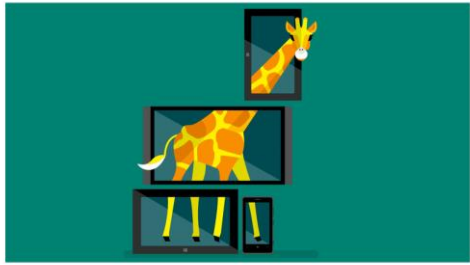
Blog: Data Science 101
Explore resources for learning data science with Ryan Swanstrom

[Learn more](#)



Cortana Intelligence Corner
Helping you navigate the world of the Cortana Intelligence Suite


[Learn more](#)




Blog: Backyard Data Science
Buck Woody's non-traditional route to learn data science

[Learn more](#)


Find out how Cortana Intelligence is helping your industry




Healthcare



Retail



Manufacturing



Banking

Course Documentation

SQLW301 - Microsoft Azure SQL Data Warehouse

This material covers using and managing the Azure SQL Data Warehouse.

The Azure SQL Data Warehouse (**Course Materials**)

Primary 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.



Information in this document, including URL and other Internet Web site references, is subject to change without notice. Unless otherwise noted, the companies, organizations, products, domain names, e-mail addresses, logos, people, places, and events depicted herein are fictitious, and no association with any real company, organization, product, domain name, e-mail address, logo, person, place, or event is intended or should be inferred. Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Microsoft Corporation.

For more information, see **Microsoft Copyright Permissions** at <http://www.microsoft.com/permission>

Microsoft may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Microsoft, the furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

The Microsoft company name and Microsoft products mentioned herein may be either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries. The names of actual companies and products mentioned herein may be the trademarks of their respective owners.

This document reflects current views and assumptions as of the date of development and is subject to change. Actual and future results and trends may differ materially from any forward-looking statements. Microsoft assumes no responsibility for errors or omissions in the materials.

THIS DOCUMENT IS FOR INFORMATIONAL AND TRAINING PURPOSES ONLY AND IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, WHETHER EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, AND NON-INFRINGEMENT.