



# Azure Synapse Analytics

## Synapse in a Day

# Veterans Affairs

June 13<sup>th</sup>, 2022  
0900 - 1700 EST





# John Deardurff

Microsoft Customer Engineer (Global Technical Team)  
Microsoft Certified Trainer (Regional Lead)  
MVP: Data Platform (2016 – 2018)

Email: [John.Deardurff@Microsoft.com](mailto:John.Deardurff@Microsoft.com)  
Twitter: [@SQLMCT](https://twitter.com/SQLMCT)  
Website: [www.SQLMCT.com](http://www.SQLMCT.com)  
GitHub: [github.com\SQLMCT](https://github.com\SQLMCT)



# Agenda

Azure Synapse in a day is a demo driven engagement where the attendee can achieve the most in the least amount of time. It is geared to showcase all the key components of Azure Synapse Analytics.

## Azure Synapse Concepts and Terminologies (120 minutes)

- What is Azure Synapse?
- Overview of Modern Data Warehousing
- Scaling Massively Parallel Processing (MPP)
- Azure Synapse SQL Pool Architecture
- Table Distributions in SQL Pools



# Demonstrations

## Demo 01

- Create Synapse Workspace
- Overview of Synapse Studio
- Create SQL Pools

## Demo 02

- Setup Integration Runtime
- Create Synapse Pipeline
- Data Ingestion from an On-Premises Server (Single Table)

## Demo 03

- Data Ingestion from On-Premises Server (Lookup Tables)
- Structured Data to ADLS

## Demo 04

- Ingest Data using Polybase
- Ingest Data using COPY INTO
- Analyze Data in Synapse Studio

## Demo 05

- Creating Spark Pools
- Ingesting Data from a Data Lake to CosmosDB
- Reporting using Power BI

# Attendee Introductions

Name

Job/Team/Specialty

Time Zone

Experience

Expectations



# Azure Synapse Analytics

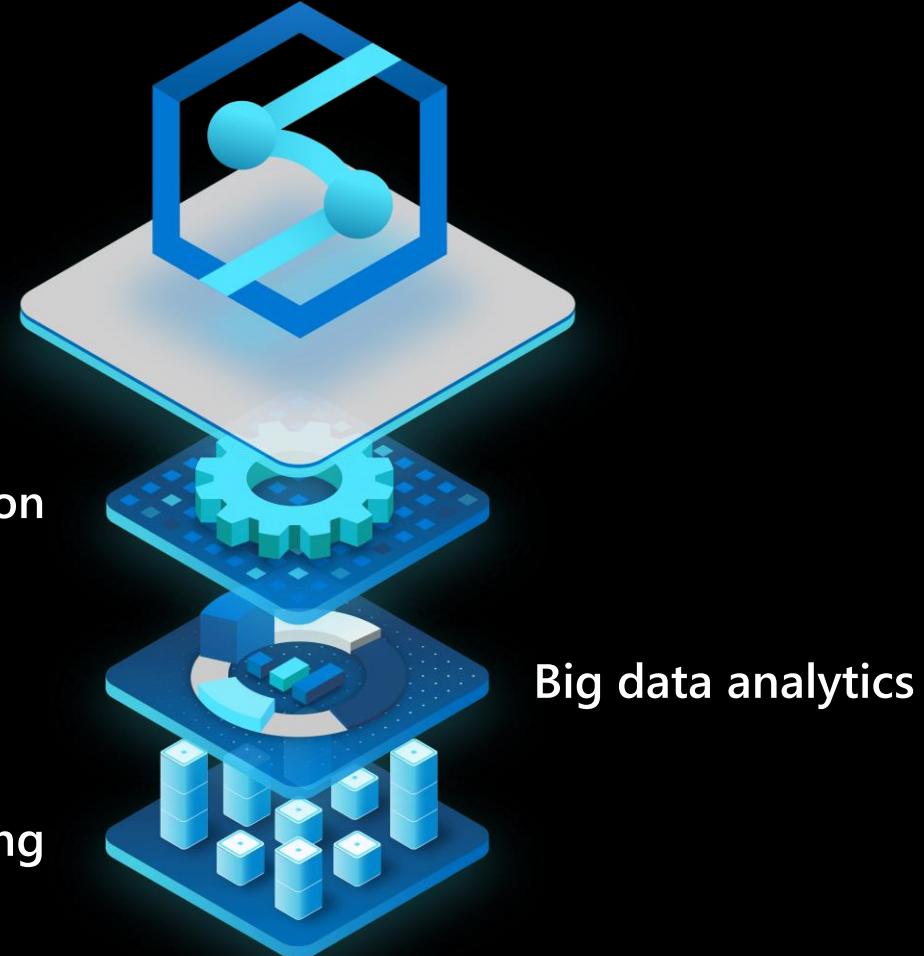
**The first unified, cloud native platform for converged analytics**

Azure Synapse is the only unified platform for analytics, blending big data, data warehousing, and data integration into a **single cloud native service** for end-to-end analytics at cloud scale.

Data integration

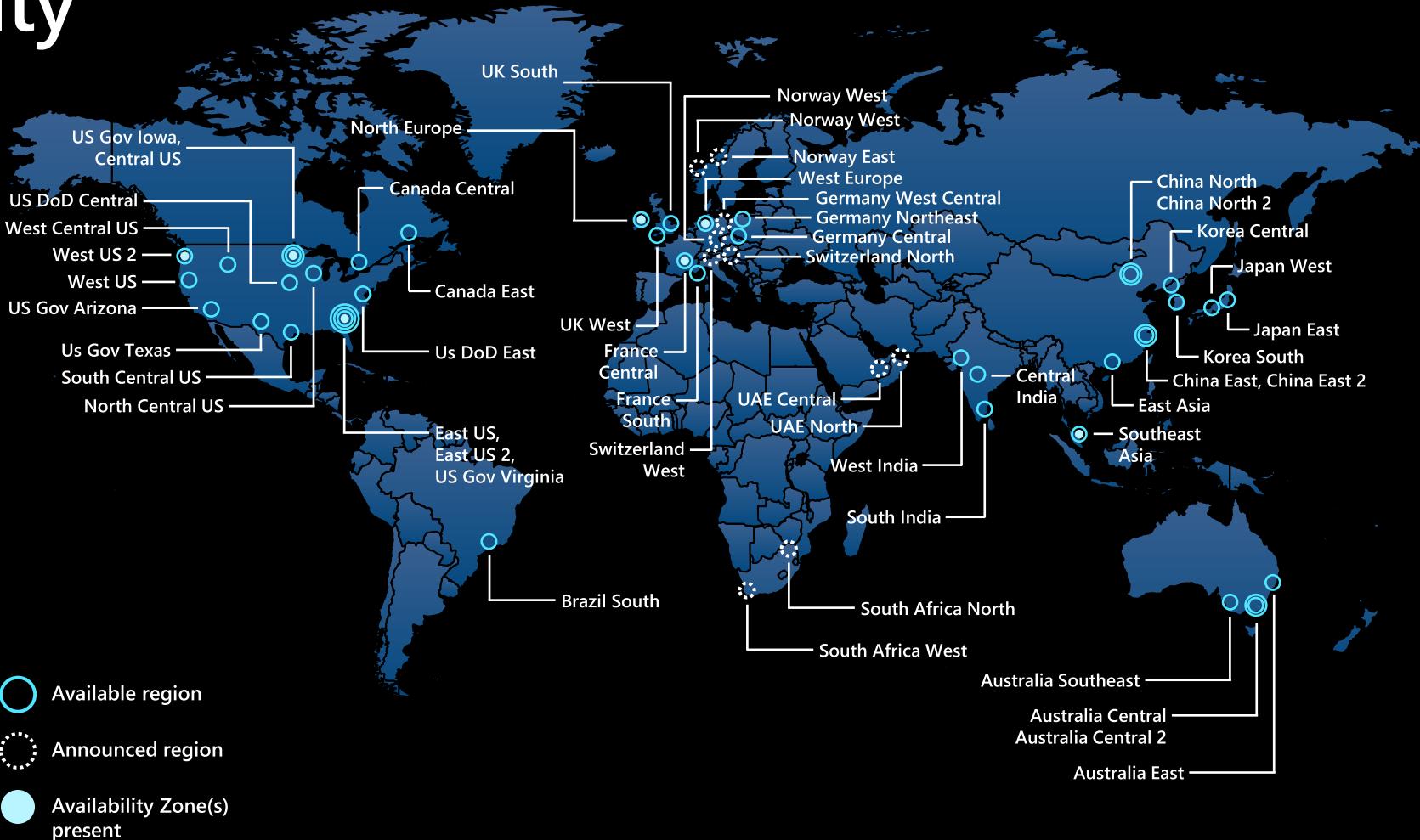
Data warehousing

Big data analytics

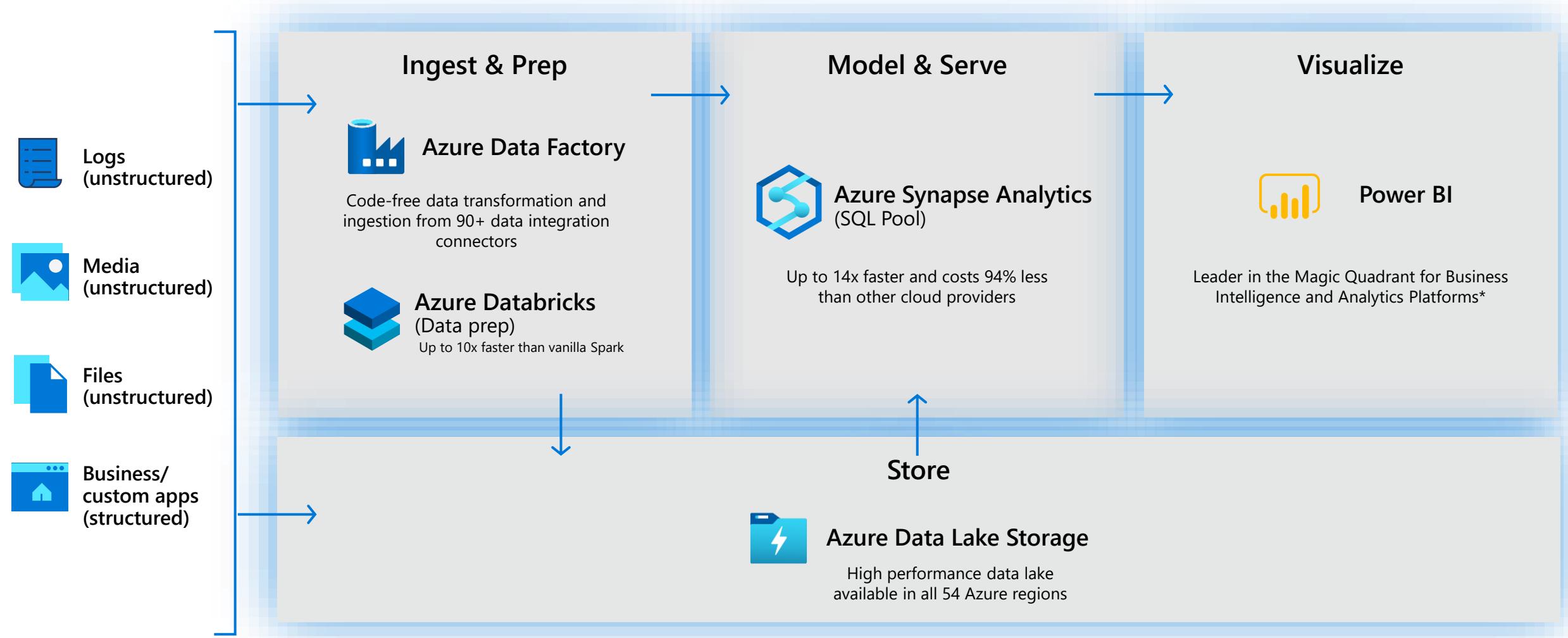


# Azure Synapse regional availability

Australia Southeast	Korea Central
Australia East	North Central US
Brazil South	North Europe
Canada Central	South Africa North
Canada East	South Central US
Central India	Southeast Asia
Central US	Switzerland North
East Asia	UK West
East US	UK South
East US 2	West Central US
France Central	West Europe
Germany West Central	West US
Japan East	West US 2
Japan West	

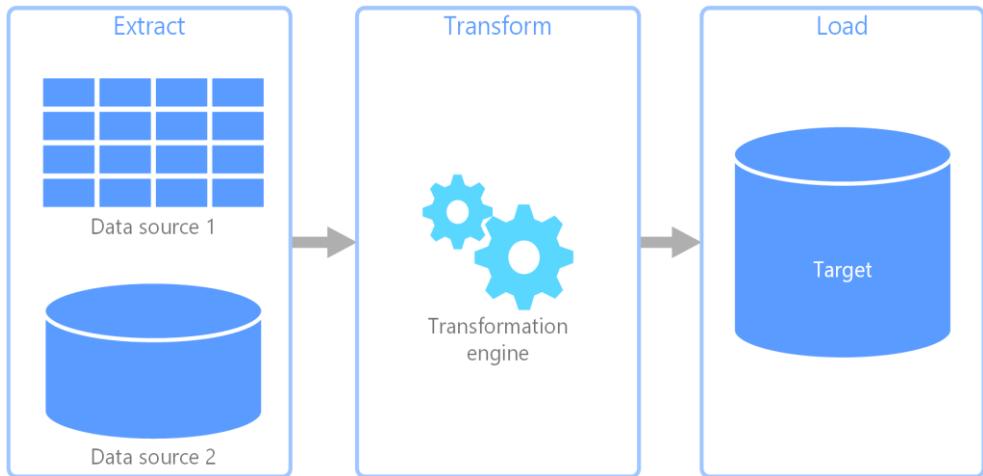


# Modern Data Warehousing Patterns



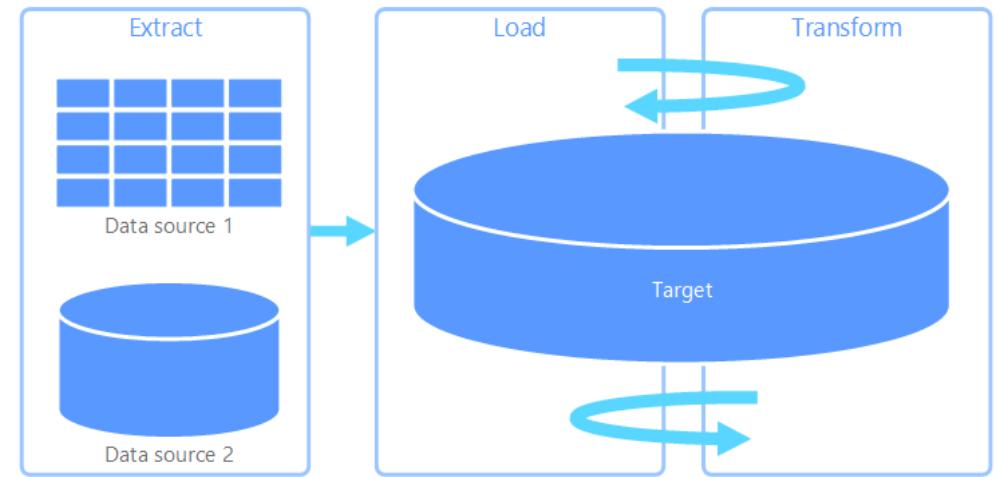
# ELT vs ETL

Traditional Data Warehouse



ETL

Modern Data Warehouse

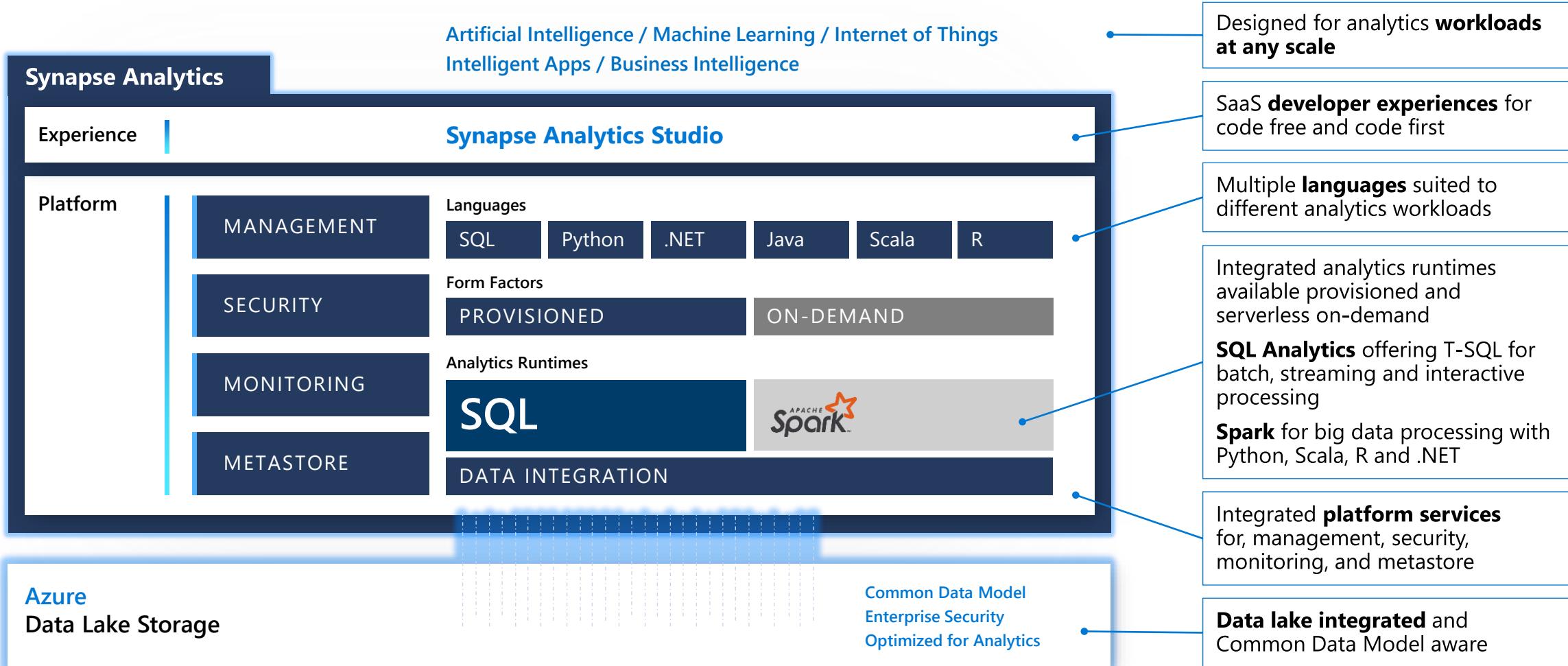


ELT

# 90+ Connectors out of the box

# Azure Synapse Analytics

Limitless analytics service with unmatched time to insight



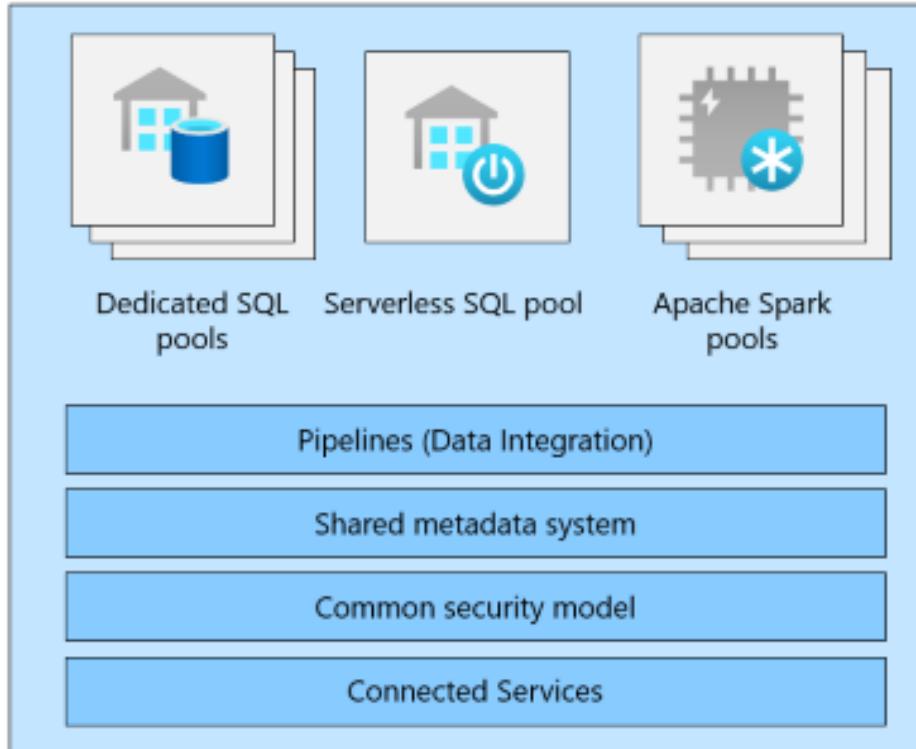
# Standalone Pools vs Workspace Pools

Dedicated SQL pool (formerly SQL DW)



Dedicated SQL pool

Azure Synapse Analytics



Synapse workspace

Synapse Studio

# Azure Synapse Analytics Eco System

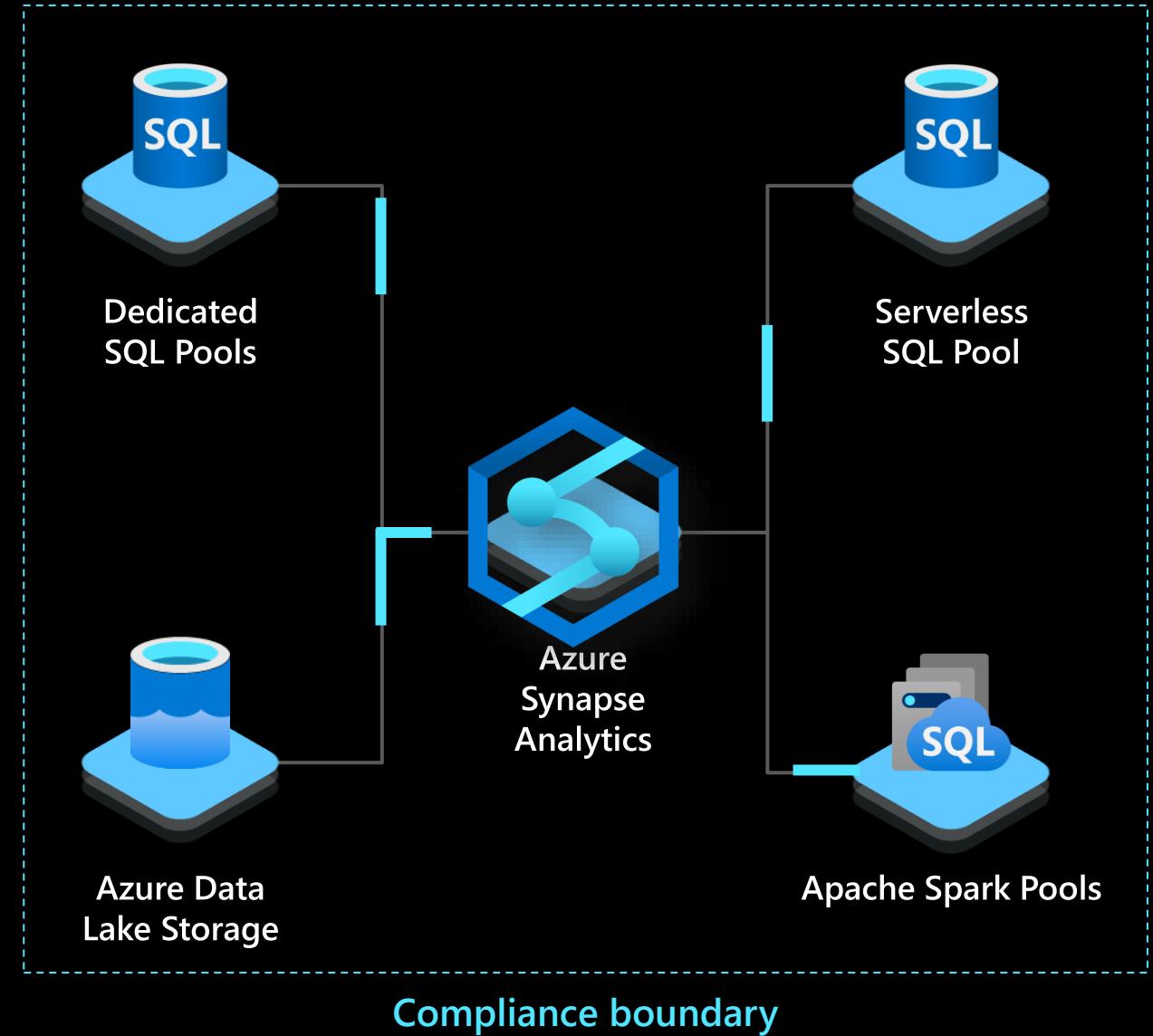
## Integrated data platform

Dedicated SQL Pools

Serverless SQL Pool

Apache Spark Pool

Azure Data Lake Storage



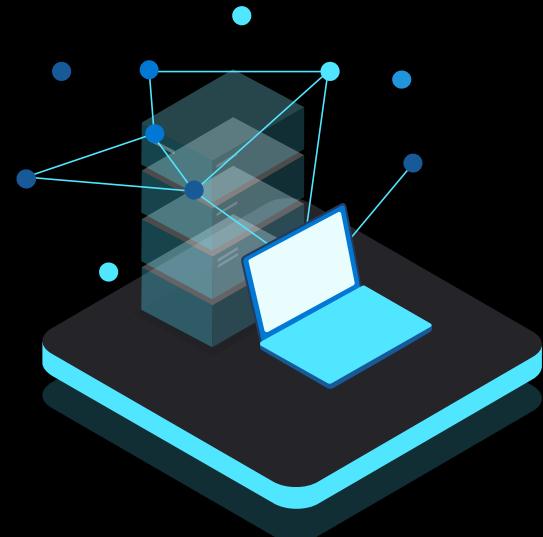
# Serverless + dedicated SQL

---

## Flexible consumption models

Serverless pay-per-query ideal for ad-hoc data lake exploration and transformation

Dedicated clusters optimized mission-critical data warehouse workloads



Serverless



Dedicated

# Let's get the concepts clear

---

- Massively Parallel Processing (MPP)
- Compute Data Warehouse Units (cDWU's)
- Control Node
- Compute Nodes
- Distributions



# Massively Parallel Processing (MPP)

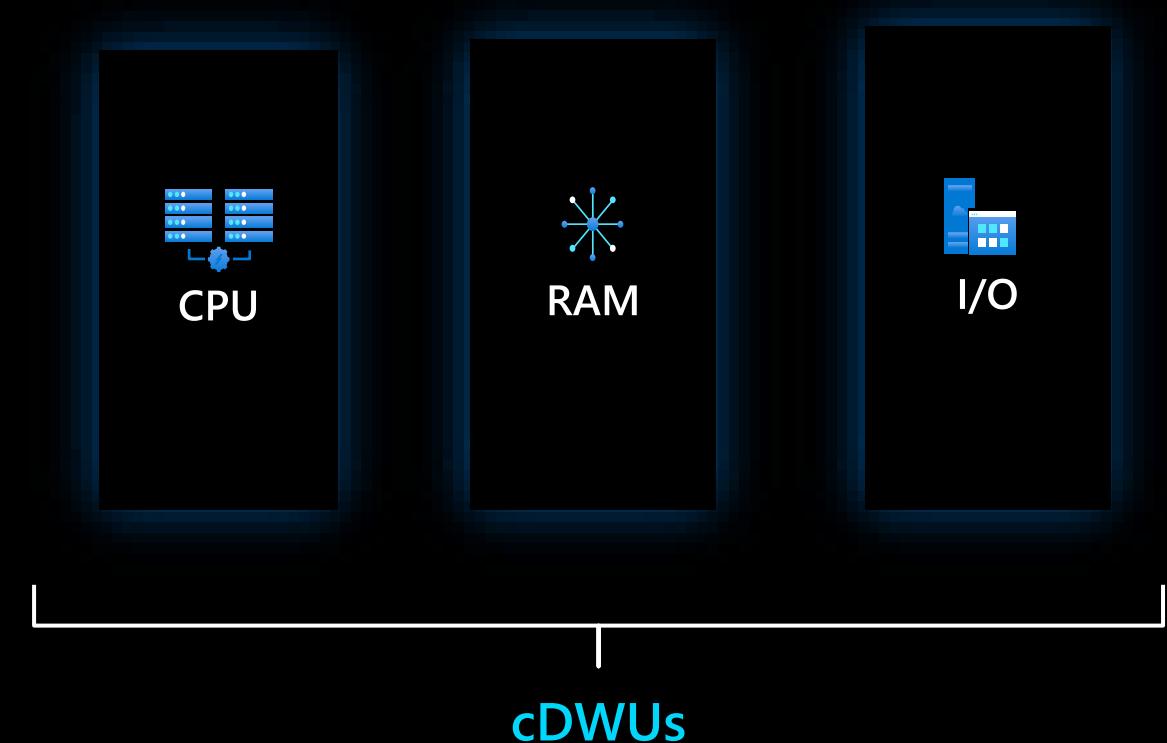
---

- “Divide and Conquer” Concept
- Uses many separate CPUs running in parallel to execute a single program
- Shared Nothing: Each CPU has its own memory and disk (scale-out)



# Compute Data Warehouse Units (cDWUs)

- cDWU is a combination of CPU, Memory and IO
- More cDWU you have, more the compute power
- cDWU ranges from DW100c to DW30,000c

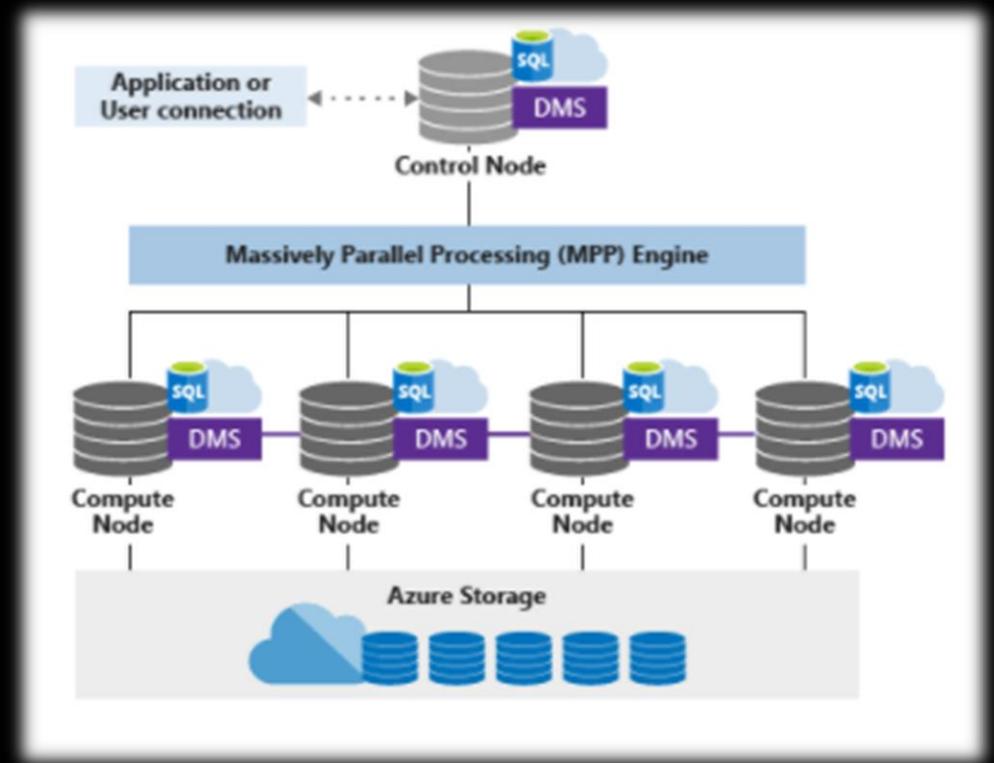


# Scaling Compute

cDWU	# of Compute Nodes	# Distributions per node
DW100c	1	60
DW200c	1	60
DW300c	1	60
DW400c	1	60
DW500c	1	60
DW1000c	2	30
DW1500c	3	20
DW2000c	4	15
DW2500c	5	12
DW3000c	6	10
DW5000c	10	6
DW6000c	12	5
DW7500c	15	4
DW10000c	20	3
DW15000c	30	2
DW30000c	60	1

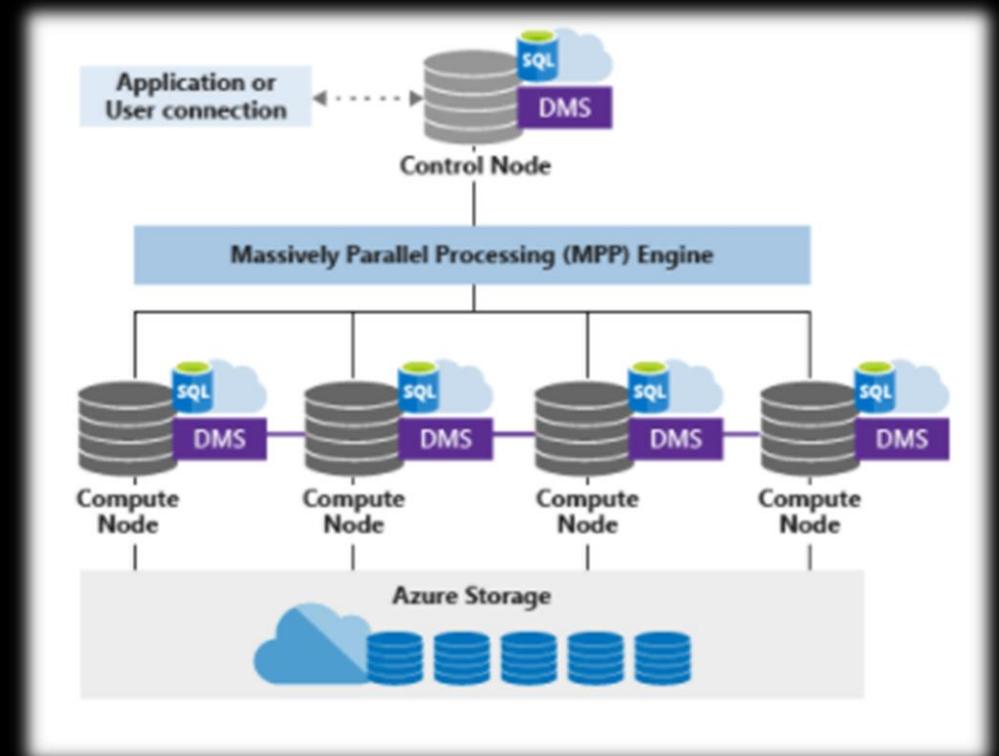
# Control Node

- The brain of the MPP architecture
- It is the front end to that interacts with applications and connections
- The distributed query engine runs on the control node
- Responsible for optimizing and coordinating queries

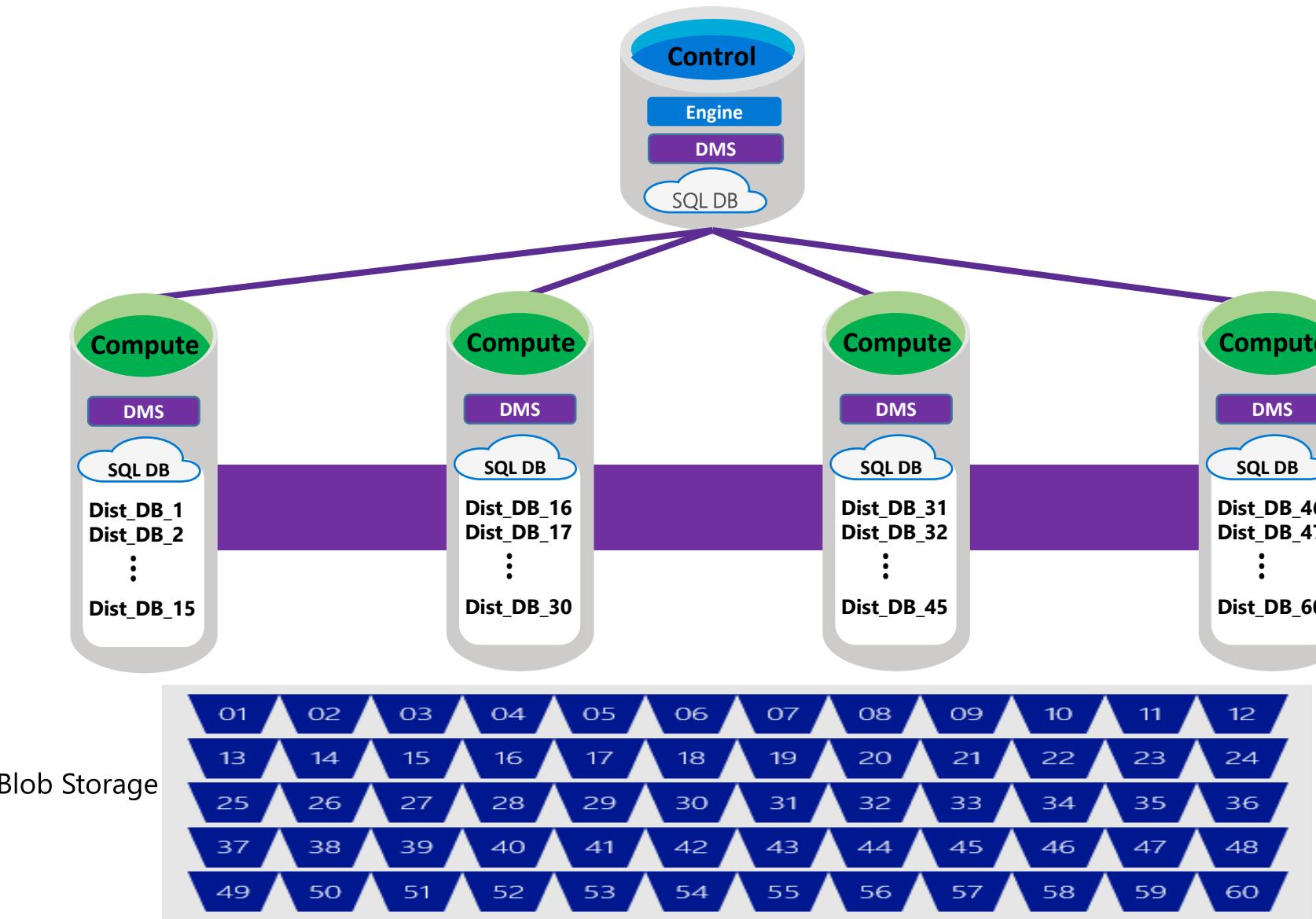


# Compute Nodes

- The brawn of the MPP architecture
- Provides computational power
- Number of compute nodes can range from 1 to 60



# Synapse SQL MPP Architecture



## Control

Connection and tool endpoint.  
Coordinates compute activity.

## Compute

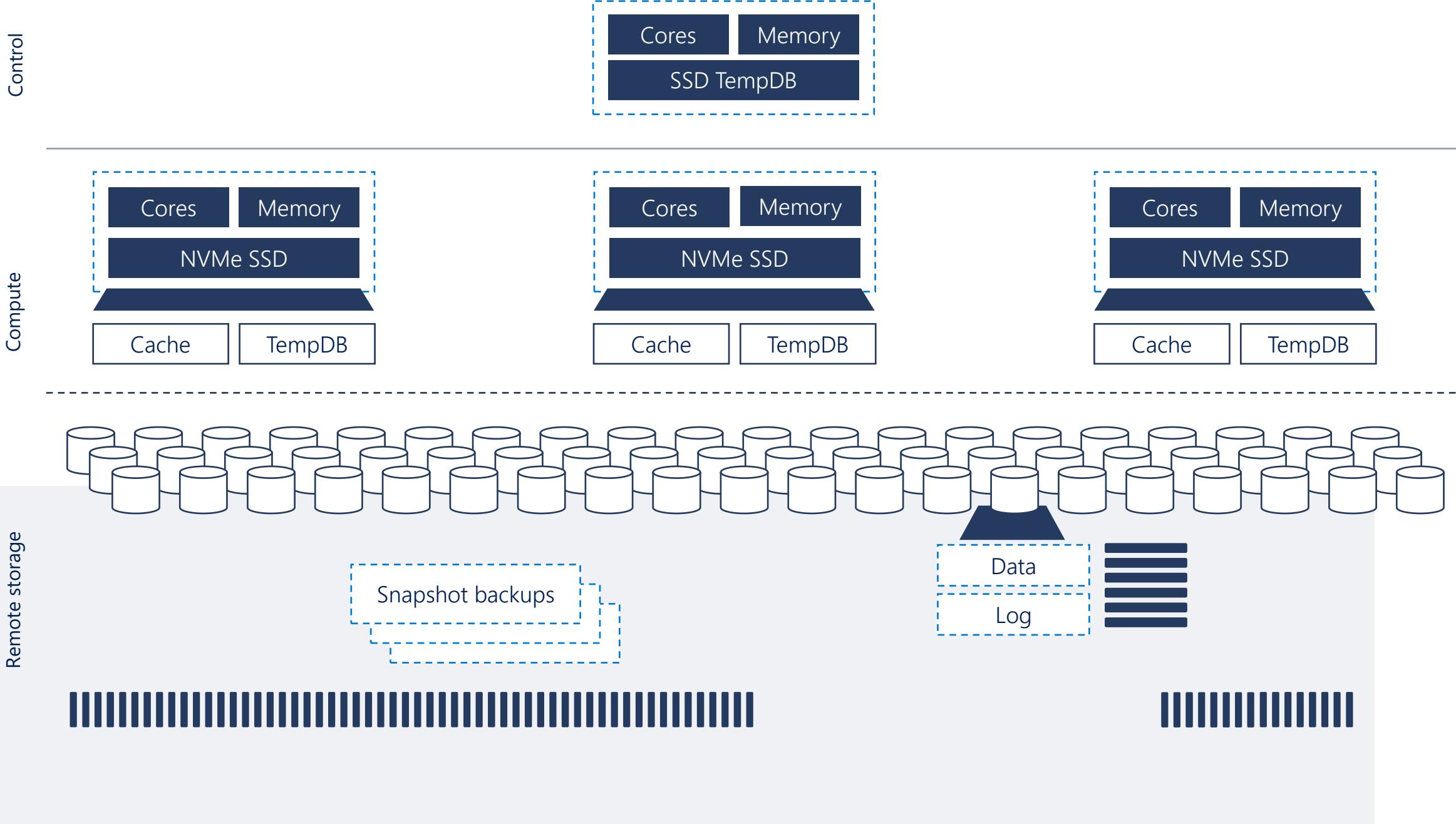
Handles query processing,  
ability to scale up/down

## DMS: Data Movement Services / Instant Data Movement

Coordinates data movement  
between nodes

## Storage

Remote storage. Scales  
independently of compute

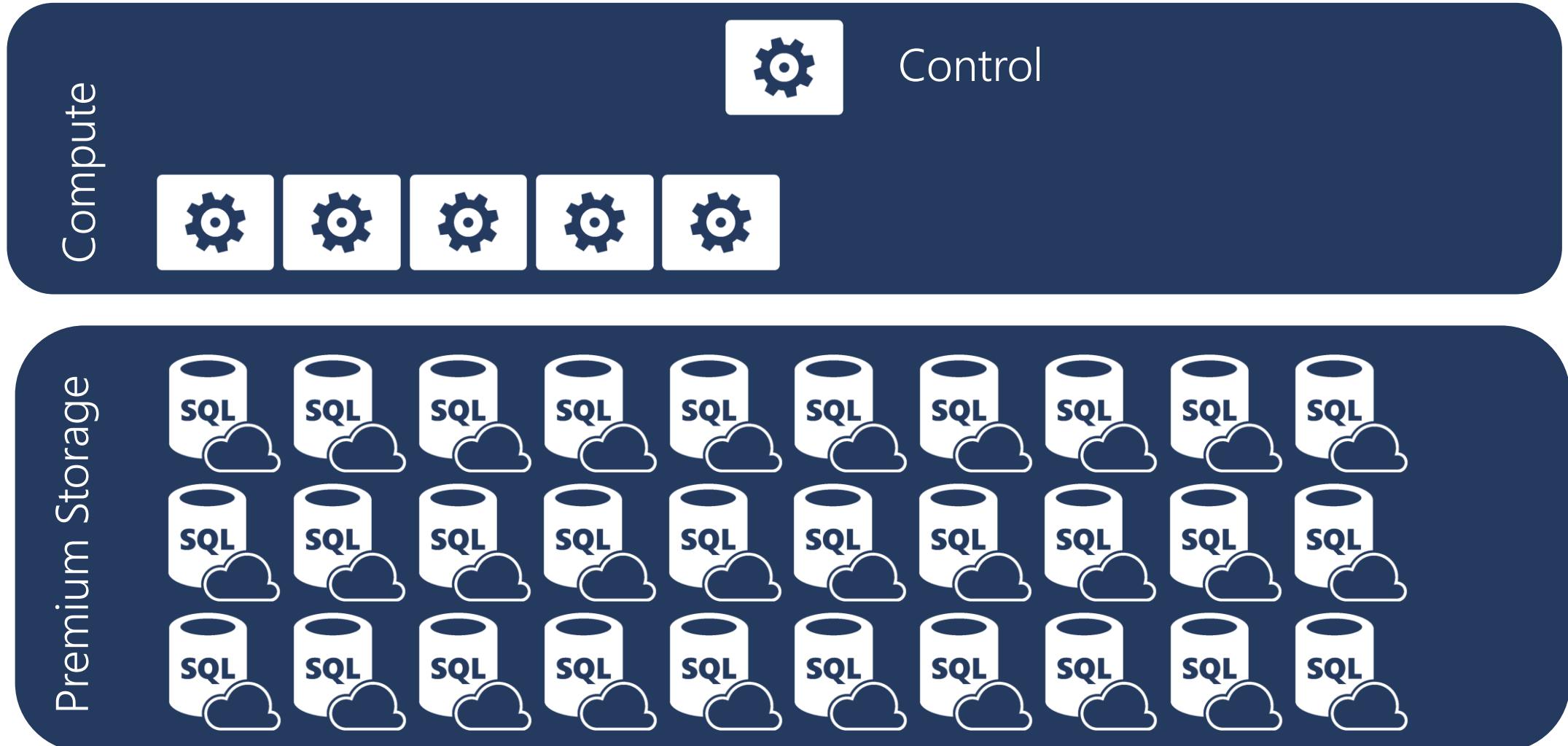


# Compute Data Warehouse Units – Memory and Adaptive Cache

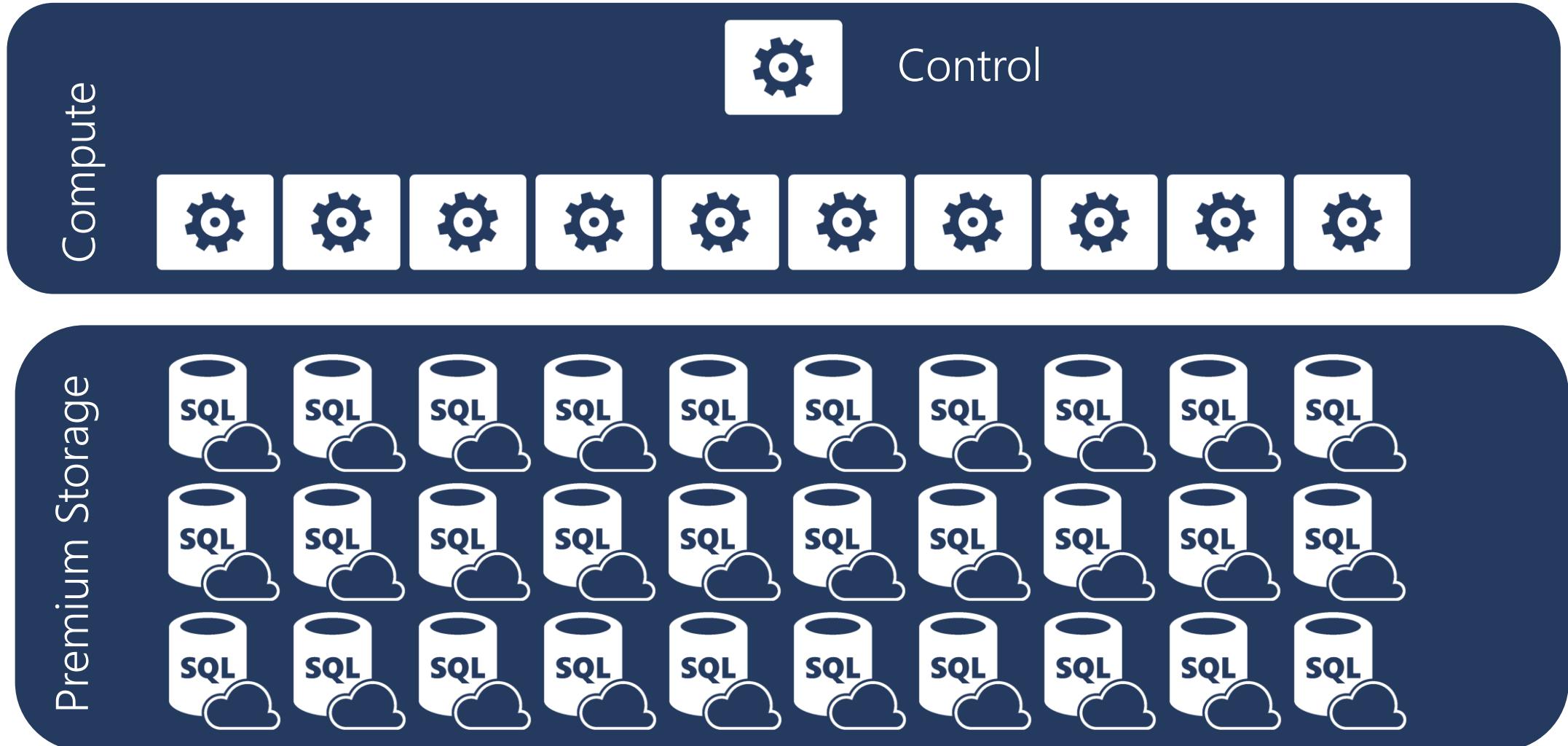
- 300 GB per DWU500
- 1.5TB cache per DWU 500
- EXEC **sp\_spaceused**
- This cache is *per node*

DWU	Memory (GB)	Tempdb (TB)	Cache (TB)
1000	600	4	3
1500	900	6	4.5
2000	1200	8	6
2500	1500	10	7.5
3000	2000	12	9
5000	3000	19	15
6000	3600	23	18
7500	4500	29	22.5
10000	6000	39	30
15000	9000	58	45
30000	18000	117	90

# Separate Compute From Storage



# Independently Scale Compute



# Pause and Resume Workload



Compute



Control

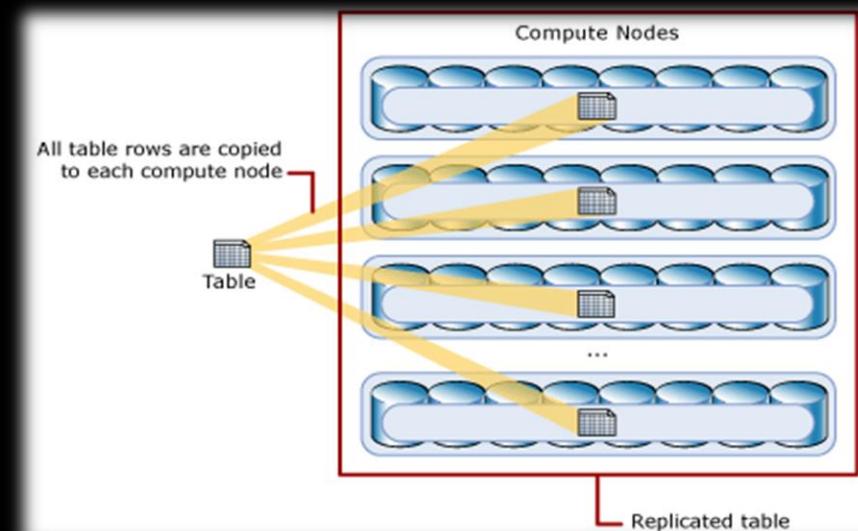
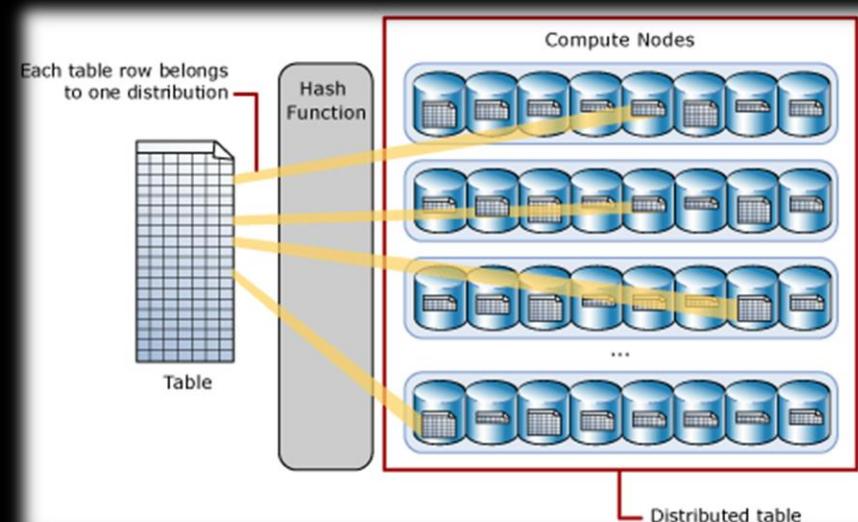


Premium Storage



# Distributions

- The work is divided into 60 distributions
- Various ways to distribute the data:
- Hash-distributed tables
- Round-robin distributed tables
- Replicated tables



# Table Distribution Options

## Hash Distributed

Data divided across nodes based on hashing algorithm

Same value will always hash to same distribution

Single column only  
(Multi column soon)

Check for Data Skew

## Round Robin

(Default)

Data distributed evenly across nodes

Easy place to start, don't need to know anything about the data

Simplicity at a cost

Will incur more data movement at query time

## Replicated

Data repeated on every node

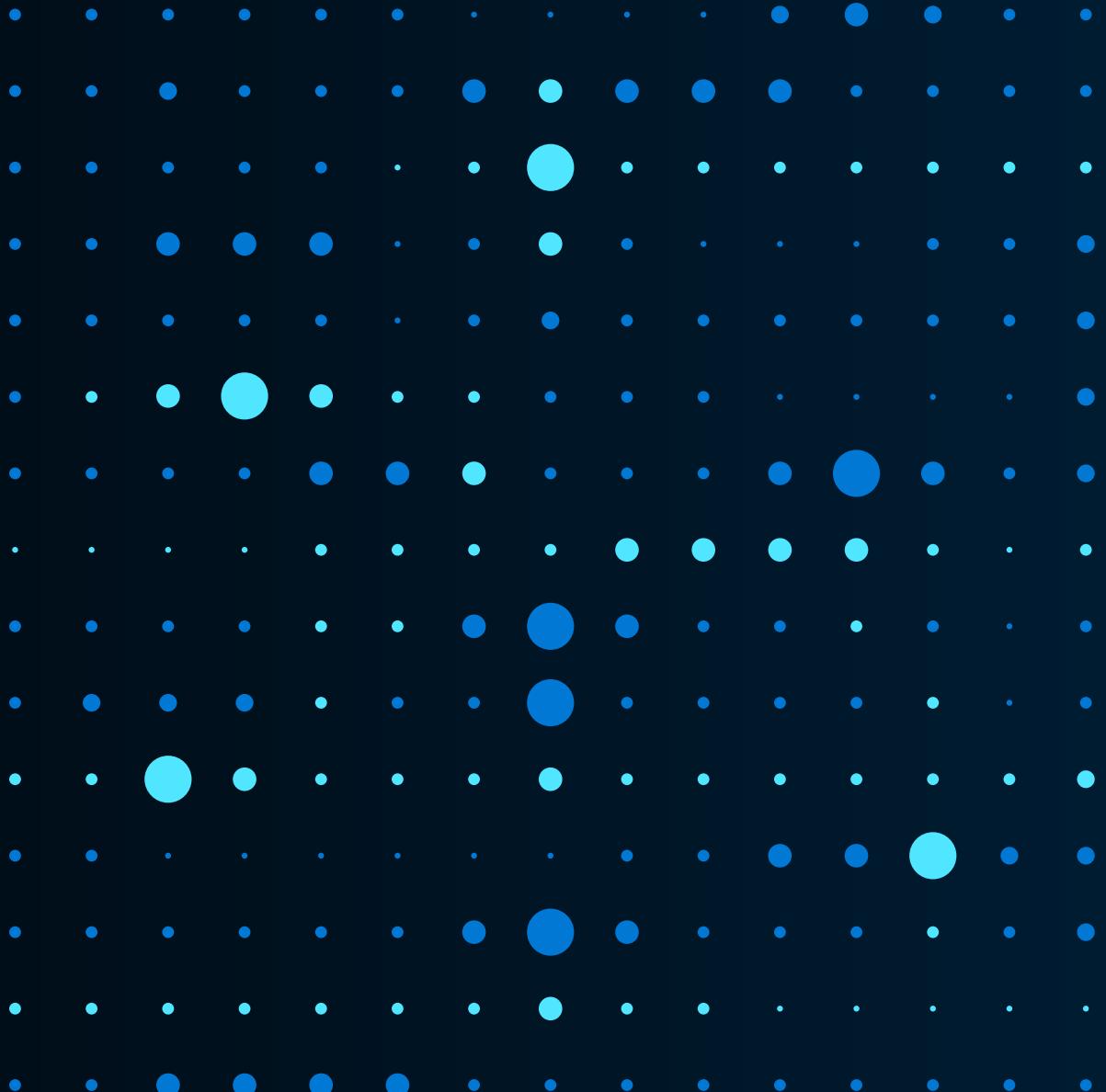
Simplifies many query plans and reduces data movement

Best with joining hash table

Consumes more space  
Joining two Replicated Table runs on one node

# Demonstration One

- Create Synapse Workspace
- Overview of Synapse Studio
- Create SQL Pool
- Create Workspace using PowerShell



# Synapse Studio

Synapse Studio divided into **Activity hubs**.

These organize the tasks needed for building analytics solution.

The screenshot shows the Microsoft Azure Synapse Studio interface. On the left, a vertical sidebar menu is highlighted with a red box, containing the following items: Home, Data, Develop, Integrate, Monitor, and Manage. A red arrow points from the 'Integrate' item in this menu to the 'Integrate' hub icon on the main page. The main content area is titled 'Synapse workspace' and shows the 'wsazuresy' workspace. It features six activity hubs arranged in a grid:

- Home**: Quick-access to common gestures, most-recently used items, and links to tutorials and documentation.
- Data**: Explore structured and unstructured data.
- Develop**: Write code and define business logic of the pipeline via notebooks, SQL scripts, Data flows, etc.
- Integrate**: Design pipelines that move and transform data.
- Monitor**: Centralized view of all resource usage and activities in the workspace.
- Manage**: Configure the workspace, pool, linked service, access to artifacts.

The top navigation bar includes the Microsoft Azure logo, Synapse Analytics, workspace name, user info (someone@microsoft.com), and a search bar.

# Home Hub

It is a starting point for the activities with key links to tasks, artifacts, documentation and sample artifacts for learning purpose

The screenshot shows the Microsoft Azure Synapse Analytics Home Hub. The top navigation bar includes 'Microsoft Azure' and 'Synapse Analytics' with the workspace name 'wsazuresynapseanalytics'. The top right corner shows a user email 'someone@microsoft.com' and the Microsoft logo. On the left, a sidebar menu lists 'Home', 'Data', 'Develop', 'Integrate', 'Monitor', and 'Manage'. The main content area is titled 'Synapse workspace' and 'wsazuresynapseanalytics'. It features a large, stylized 3D bar chart and network graph visualization. Below this, four cards are displayed, each with an icon and a call-to-action:

- Ingest**: Perform a one-time or scheduled data load.
- Explore and analyze**: Learn how to get insights from your data.
- Visualize**: Build interactive reports with Power BI capabilities.
- Learn**: Start with Azure Open Datasets and sample code.

A red box highlights these four cards. Below them is a section titled 'Recent resources' with a table:

Name	Last opened by you
05 Sentiment_Analysis_Cognitive_Services	4 hours ago
Predict NYCTaxi Trip Amount	4 hours ago
001 SQL Pool Security RLS DDM CLE	5 hours ago
005 Predict In-Engine Scoring	a day ago
05 Anomaly_Detection_Cognitive_Services	a day ago

At the bottom left, there is a 'Show more ▾' link.

# Knowledge Center

## Accelerate time to solution

Azure Open Data sets

Pre-built samples to accelerate development

- SQL Scripts
- Notebooks
- Data Pipelines

The screenshot shows the Microsoft Azure Synapse Analytics Sample center interface. The top navigation bar includes 'Microsoft Azure', 'Synapse Analytics', and a workspace name 'wsazuresynapseanalytics'. The top right features icons for notifications, search, and help. The main area is titled 'Sample center' and has tabs for 'Datasets', 'Notebooks', 'SQL scripts', and 'Pipelines'. A 'Datasets' tab is selected. Below the tabs is a 'Filter by keyword' input field and a 'Tags : All' dropdown. The dataset list is organized into four columns:

Dataset	Description	ID	Action
Bing COVID-19 Data	Bing COVID-19 data includes confirmed, fatal, and recovered cases from all regions, updated da...	ID: bing-covid-19-data	Sample
Boston Safety Data	Read data about 311 calls reported to the city of Boston. This dataset is stored in Parquet format and is up...	ID: city_safety_boston	Sample
COVID Tracking Project	The COVID Tracking Project dataset provides the latest numbers on tests, confirmed cases, hospitalizat...	ID: covid-tracking	Sample
Chicago Safety Data	Read data about 311 calls reported to the city of Chicago. This dataset is stored in Parquet format and is ...	ID: city_safety_chicago	Sample
European Centre for Disease Prevention and Control (ECDC) Covid-19 Cases	The latest available public data on...	ID: ecdc-covid-19-cases	Sample
NOAA Integrated Surface Data (ISD)	NOAA Integrated Surface Data (ISD) provides Worldwide hourly weath...	ID: isd	Sample
NYC Taxi & Limousine Commission - For-Hire Vehicle (FHV) trip records	The For-Hire Vehicle trip records i...	ID: nyc_tlc_fhv	Sample
NYC Taxi & Limousine Commission - green taxi trip records	The green taxi trip records include...	ID: nyc_tlc_green	Sample



# Synapse Studio **Data hub**

# Data Hub

Explore data inside the workspace and in linked storage accounts

This screenshot shows the Microsoft Azure Synapse Analytics Data Hub interface. The left sidebar includes Home, Data (selected), Develop, Integrate, Monitor, and Manage. The main area has tabs for Synapse live, Validate all, and Data (Workspace selected). A search bar and a 'Filter resources by name' input field are at the top. Below are sections for Databases (10 items) and Pools (5 items). A red box highlights the 'Workspace' tab.

- Databases (10):
  - newpoll (SQL)
  - NYCTaxi\_Pool (SQL)
  - Predict\_Pool (SQL)
  - Streaming\_Pool (SQL)
  - WWI\_Pool (SQL)
  - NYT2020 (SQL)
  - SQLServerlessDB (SQL)
  - default (Spark)
- Pools (5):
  - NYCTaxi\_Pool
  - Predict\_Pool
  - Streaming\_Pool
  - WWI\_Pool
  - NYT2020

This screenshot shows the Microsoft Azure Synapse Analytics Data Hub interface with the 'Linked' tab selected. The left sidebar is identical to the first screenshot. The main area shows a 'Data' section with tabs for Workspace (selected) and Linked (highlighted with a red box). A search bar and a 'Filter resources by name' input field are at the top. Below are sections for Storage and Databases (24 items total). A red box highlights the 'Linked' tab.

- Storage (3):
  - Azure Blob Storage
  - Azure Cosmos DB
  - Azure Data Explorer
- Databases (24):
  - Azure Data Lake Storage Gen2 (2 items)
    - wsazuresynapseanalytics (Primary...)
    - (Attached Containers)
  - Integration datasets (24 items)

# Data Hub – Linked Storage

Browse Azure Data Lake Storage Gen2 accounts – filesystems, Azure Data Explorer – clusters, Azure Cosmos DB -containers

The screenshot shows the Microsoft Azure Synapse Analytics Studio interface. On the left, the 'Data' sidebar is open, showing a tree view of linked storage resources under the 'Linked' tab. The resources are:

- Linked Cosmos DB Analytical Store
- Linked Azure Data Explorer
- Linked ADLS Gen2 Account
- Container (filesystem)

Red arrows point from each of these labels to their corresponding items in the sidebar. The 'rawdata' container is selected, highlighted with a red border. The 'File path' navigation bar shows 'rawdata > taxidata'. The main pane displays the contents of the 'rawdata' container, listing six cached items:

Name	Last Modified	Content Type	Size
part-00000-0300809f-304e-44bc-81bd-bbd63974c3e4-c000.snappy.parq...	8/27/2020, 12:32:19 AM		121.9 MB
part-00000-6b990121-0341-456c-8723-aec72b03f65f-c000.snappy.parqu...	8/27/2020, 12:32:25 AM		535.4 MB
part-00001-0300809f-304e-44bc-81bd-bbd63974c3e4-c000.snappy.parq...	8/27/2020, 12:32:20 AM		124.5 MB
part-00001-6b990121-0341-456c-8723-aec72b03f65f-c000.snappy.parqu...	8/27/2020, 12:32:23 AM		983.7 MB
part-00002-0300809f-304e-44bc-81bd-bbd63974c3e4-c000.snappy.parq...	8/27/2020, 12:32:19 AM		123.7 MB
part-00002-6b990121-0341-456c-8723-aec72b03f65f-c000.snappy.parqu...	8/27/2020, 12:32:21 AM		966.1 MB

At the bottom, it says 'Showing 1 to 6 of 6 cached items'.



# Synapse Studio Monitor hub

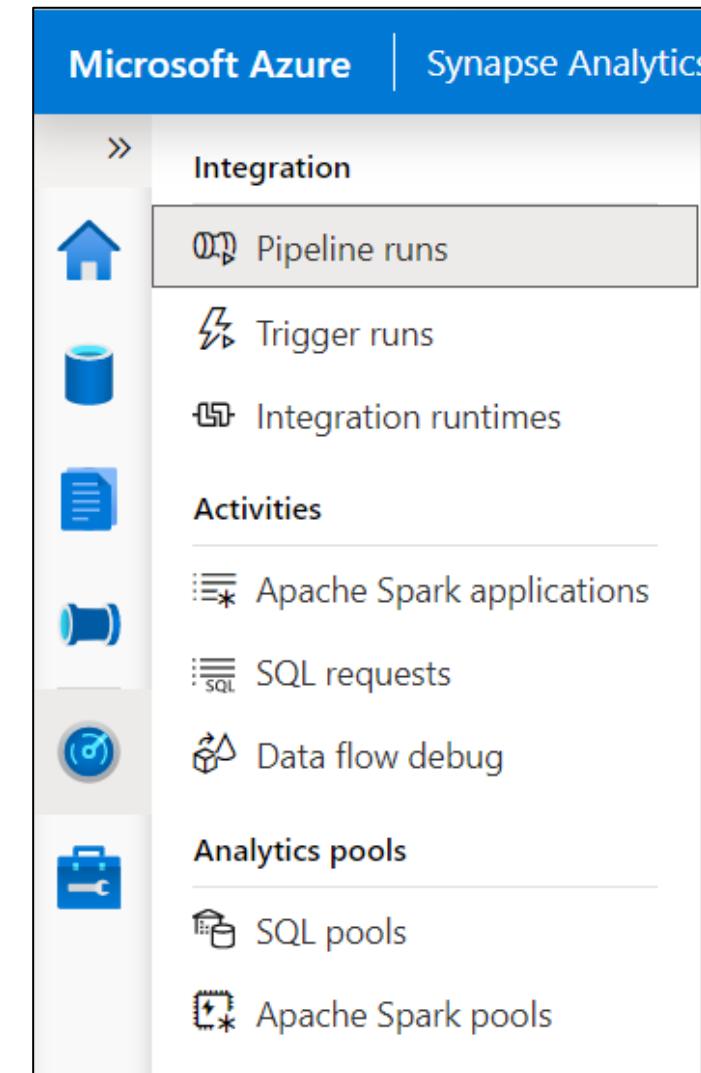
# Monitor Hub

## Overview

This feature provides single pane of glass to monitor orchestration, activities for Apache Spark Application and SQL requests.

## Benefits

Offers additional filters to monitor specific activities or orchestration



# Monitor Hub - Integration

## Overview

Monitor orchestration in the Synapse workspace for the progress and status of pipeline

Pipeline runs				
		Time : Last week (10/24/2019 9:44 AM - 10/31/2019 9:44 AM)	Time zone : Pacific Time (US & Canada) (UT...)	Runs : Latest runs
All status		Rerun	Cancel	Refresh
		PIPELINE NAME	RUN START ↑	DURATION
<input type="checkbox"/>	Load Data to SQLDW	10/25/2019, 3:49:42 PM	00:10:55	Manual trigger
<input type="checkbox"/>	Copy Open Dataset	10/25/2019, 2:17:54 PM	00:14:12	Manual trigger
<input type="checkbox"/>	Pipeline 1	10/24/2019, 1:23:43 PM	00:00:08	Manual trigger

## Benefits

Track all/specific pipelines

Monitor pipeline run and activity run details

Find the root cause of pipeline failure or activity failure

Trigger runs			
		Time : Last 30 days (3/17/20 11:48 PM - 4/16/20 11:48 PM)	Time zone : Pacific Time (US & Canada) (UT...)
All status		Refresh	Edit columns
Showing 1 - 1 items			
TRIGGER NAME	TRIGGER TYPE	TRIGGER TIME ↑	STATUS
TriggerCopy_csvdata100	ScheduleTrigger	4/10/20, 12:14:00 AM	<span style="color: green;">✓ Succeeded</span>



# Synapse Studio Manage hub

# Manage Hub

## Overview

This feature provides ability to manage Analytics pools, Linked Services, Integration, Security and Source Control.

The screenshot shows the Microsoft Azure Synapse Analytics Manage Hub interface. The left sidebar has a 'Manage' section selected. The main area shows the 'SQL pools' section, which lists six pools: 'Built-in' (Serverless, Online, Auto), 'newpool' (Dedicated, Paused, DW200c), 'NYCTaxi\_Pool' (Dedicated, Online, DW100c), 'Predict\_Pool' (Dedicated, Online, DW1000c), 'Streaming\_Pool' (Dedicated, Paused, DW2000c), and 'WWI\_Pool' (Dedicated, Online, DW100c). A 'System assigned managed identity' toggle switch is visible at the top of the pool list.

Name	Type	Status	Size
Built-in	Serverless	Online	Auto
newpool	Dedicated	Paused	DW200c
NYCTaxi_Pool	Dedicated	Online	DW100c
Predict_Pool	Dedicated	Online	DW1000c
Streaming_Pool	Dedicated	Paused	DW2000c
WWI_Pool	Dedicated	Online	DW100c

# Manage – dedicated SQL pools

## Overview

Provides ability to Pause and Resume, change Scale, Assign Tags from Studio.

Microsoft Azure | Synapse Analytics > wsazuresynapseanalytics

Synapse live ▾ Validate all Publish all 1

- Home
- Data
- Develop
- Integrate
- Monitor
- Manage**

Analytics pools
 

- SQL pools**
- Apache Spark pools

External connections
 

- Linked services

Integration
 

- Triggers
- Integration runtimes

Security
 

- Access control
- Credentials
- Managed private endpoints

Source control
 

- Git configuration

**SQL pools**

Serverless SQL pool is immediately available for your workspace. Dedicated SQL pools can be scaled up or down based on your needs.

+ New Refresh System assigned managed identity

Showing 1-6 of 6 items (1 Serverless, 5 Dedicated)

Name	Type
Built-in	Serverless
newpool	Dedicated
NYCTaxi_Pool	Dedicated
Predict_Pool	Pause
Streaming_Pool	Scale
WWI_Pool	Assign tags

**Scale**

NYCTaxi\_Pool

Scaling can impact workload management settings. Consider using the [workload management scale experience](#) in the Azure portal to configure the settings that best align to your workload needs. [Learn more about performance levels](#)

Performance level DW500c

Estimated price ⓘ  
Est. cost per hour 6.00 USD

Apply Cancel

# Manage – serverless SQL pools

## Overview

It enables to set the budget for amount of data processed within Synapse Studio or T-SQL.

The screenshot shows the Microsoft Azure Synapse Analytics studio interface. On the left, there's a navigation sidebar with Home, Data, Develop, Integrate, Monitor, and Manage sections. Under the Manage section, 'Analytics pools' is selected, which further branches into 'SQL pools' and 'Apache Spark pools'. Below this is an 'External connections' section. The main content area is titled 'SQL pools' and contains a message: 'Serverless SQL pool is immediately available for your workspace. Dedicated SQL pools can constraints. [Learn more](#)'. It has buttons for '+ New', 'Refresh', and a toggle switch for 'System assigned managed identity'. A table lists 1-6 items (1 Serverless, 5 Dedicated), showing rows for 'Built-in' (Serverless) and 'newpoll' (Cost Control). A red box highlights the 'Cost Control' icon for 'newpoll'. At the bottom, a T-SQL query editor window is open, displaying the following code:

```

1 sp_set_data_processed_limit
2     @type = N'daily',
3     @limit_tb = 10
4
5 sp_set_data_processed_limit
6     @type= N'weekly',
7     @limit_tb = 100
8
9 sp_set_data_processed_limit
10    @type= N'monthly',
11    @limit_tb = 1000

```

### Cost Control

Workspace Budget limit for a period. [Learn more](#)

**Daily limit** ⓘ

Enable  Disable

**Data used today**  
2 MB

10	TB
----	----

**Weekly limit** ⓘ

Enable  Disable

**Data used this week**  
23 MB

100	TB
-----	----

**Monthly limit** ⓘ

Enable  Disable

**Data used this month**  
2 MB

0	TB
---	----

**Apply** **Cancel**

# Manage – Apache Spark pools

## Overview

Provides ability to Pause, Scale, Assign Tags, upload Spark configuration, packages from Studio and validate role assignments on respective Spark pool

**Apache Spark pool**

Apache Spark pools can be finely tuned to run different kinds of Apache specific configuration libraries, permissions, etc. [Learn more](#)

Name	Size
analyticspool	Medium (8 vCores)
AnalyticsPool99	
analytics1	
analytics2	

**... More Options**

- Auto-pause settings
- Autoscale settings
- Packages
- Spark configuration
- Assign tags
- View role assignments
- Delete

### Manage packages

Apache Spark pools can be customized with additional libraries by providing a library requirements file. [Learn more](#)

NAME	SIZE	DATE
requirements.txt	73B	9/9/2020, 8:39:40 PM

**... More Options**

Upload Refresh

Apply Cancel

### View role assignment

analyticspool

To manage these pools, users need sufficient Azure RBAC permissions on this workspace, such as the Owner or Contributor role. [Learn more](#)

Synapse role assignments on this Apache Spark pool allow users to submit jobs. All [Workspace users](#) can view this pool. [Learn more](#)

Filter by name Type : All Role : All

Showing 1 - 4 of 4 role assignments (2 user(s), 1 group(s), 1 service principal(s))

Name ↑	Type	Scope
Synapse Administrator		
Priyanka Langade	Individual	Workspace (Inherited)
SynapseWsAdmin	Group	Workspace (Inherited)
SynapseWsAdmin@serv		
Charles Feddersen	Individual	Workspace (Inherited)
wsazuresynapseanalytics	Service Principal	Workspace (Inherited)

**Close** Manage role assignments (Access control)

# Manage – Linked services

## Overview

It defines the connection information needed to connect to external resources.

## Benefits

Offers pre-build 90+ connectors

Easy cross platform data migration

Represents data store or compute resources

**Microsoft Azure | Synapse Analytics > wsazuresynapseanalytics**

Synapse live | Validate all | Publish all

**Linked services**

Linked services are much like connection strings, which define the connection to external resources. [Learn more](#)

**New**

**New linked service**

PayPal (Preview)	Phoenix	PostgreSQL
Power BI	Presto (Preview)	QuickBooks (Preview)
REST	SAP BW Open Hub	SAP BW via MDX
SAP Cloud For Customer	SAP ECC	SAP HANA

Showing 1 - 15 of 15

Name ↑

- AzureDataExplor
- AzureDataExplor
- AzureMLService1
- AzureMLServiceN
- bing-covid-19-d
- Nellies\_Keyvault

Continue Cancel

# Manage – Triggers

## Overview

It defines a unit of processing that determines when a pipeline execution needs to be kicked off.

## Benefits

### Create and manage

- Schedule trigger
- Tumbling window trigger
- Event trigger

### Control pipeline execution

**Triggers**

To execute a pipeline set the trigger to be kicked off.

+ New

Filter by keyword

New trigger

Choose a name for your trigger. This name can be updated at any time until it is published.

Name \*

Description

Type \*

Schedule    Tumbling window    Event

Start Date (UTC) \*

10/29/2019 9:46 PM

Recurrence \*

Every 1 Minute(s)

End \*

No End    On Date

Annotations

Activated \*

Yes    No

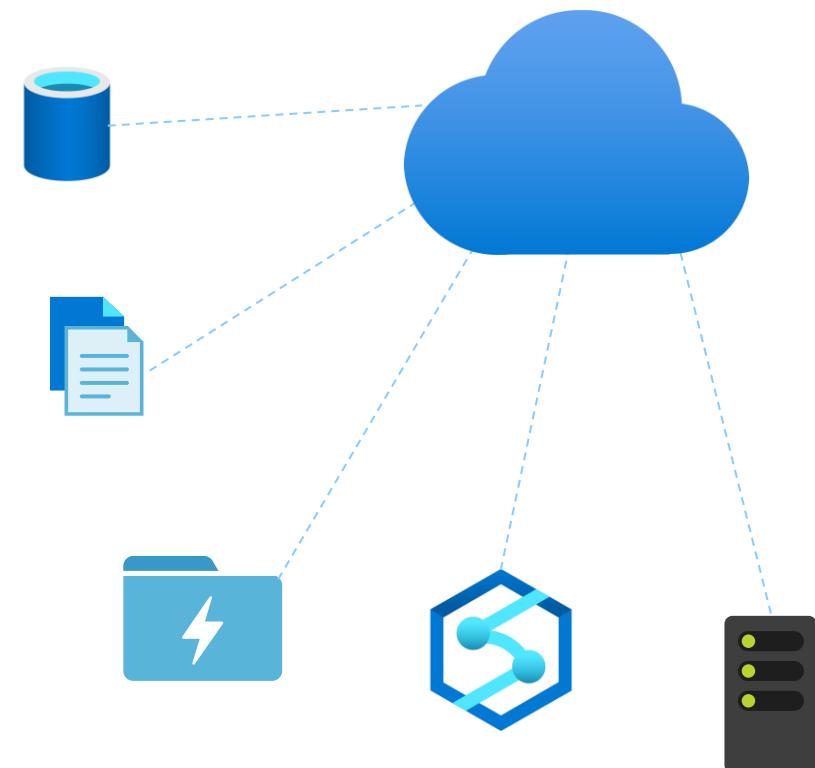
OK Cancel

# Connectivity

We need to ensure we have connectivity between our on-premises server and the cloud.

There are three possible ways to achieve this:

1. Express Route: Private VNET extended between Azure and On premises
2. Side to Side VPN: VNET extension between Azure and On premises
3. **Self-Hosted Integration runtime**: VM Which hosts an Integration runtime, this will allow traffic to pass through between Azure Synapse Workspace Pipelines and Azure resources for data movement



# Manage – Integration runtimes

## Overview

Integration runtimes are the compute infrastructure used by Pipelines to provide the data integration capabilities across different network environments. An integration runtime provides the bridge between the activity and linked services.

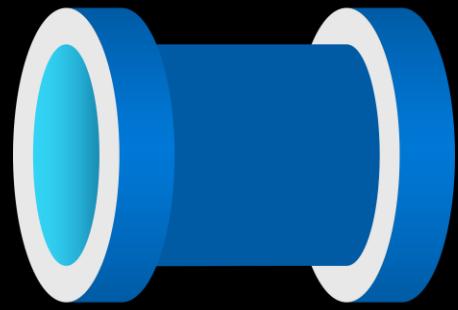
## Benefits

Offers Azure Integration Runtime or Self-Hosted Integration Runtime

Azure Integration Runtime – provides fully managed, serverless compute in Azure

Self-Hosted Integration Runtime – use compute resources in on-premises machine or a VM inside private network

The screenshot shows the 'Integration runtimes' blade in the Azure portal. On the left, a navigation menu lists 'Analytics pools', 'SQL pools', 'Apache Spark pools', 'External connections', 'Linked services', 'Integration', 'Triggers', 'Integration runtimes' (which is selected and highlighted with a red box), 'Security', 'Access control', and 'Credentials'. At the top right, there are buttons for 'Validate all', 'Publish all', '+ New' (also highlighted with a red box), and 'Refresh'. A search bar labeled 'Filter by keyword' is present. Below the search bar, it says 'Showing 1 - 1 of 1 items'. To the right, a table header includes columns for 'Name' (sorted by Type), 'Type' (sorted by Type), 'Sub-type' (sorted by Sub-type), and 'Status' (sorted by Status). A modal window titled 'Integration runtime setup' is open, asking 'Choose the network environment of the data source/destination or external compute to which the integration runtime will connect to for data movement or dispatch activities'. It shows two options: 'Azure' (represented by a cloud icon) and 'Self-Hosted' (represented by a server icon). At the bottom of the modal are 'Continue', 'Back', and 'Cancel' buttons.

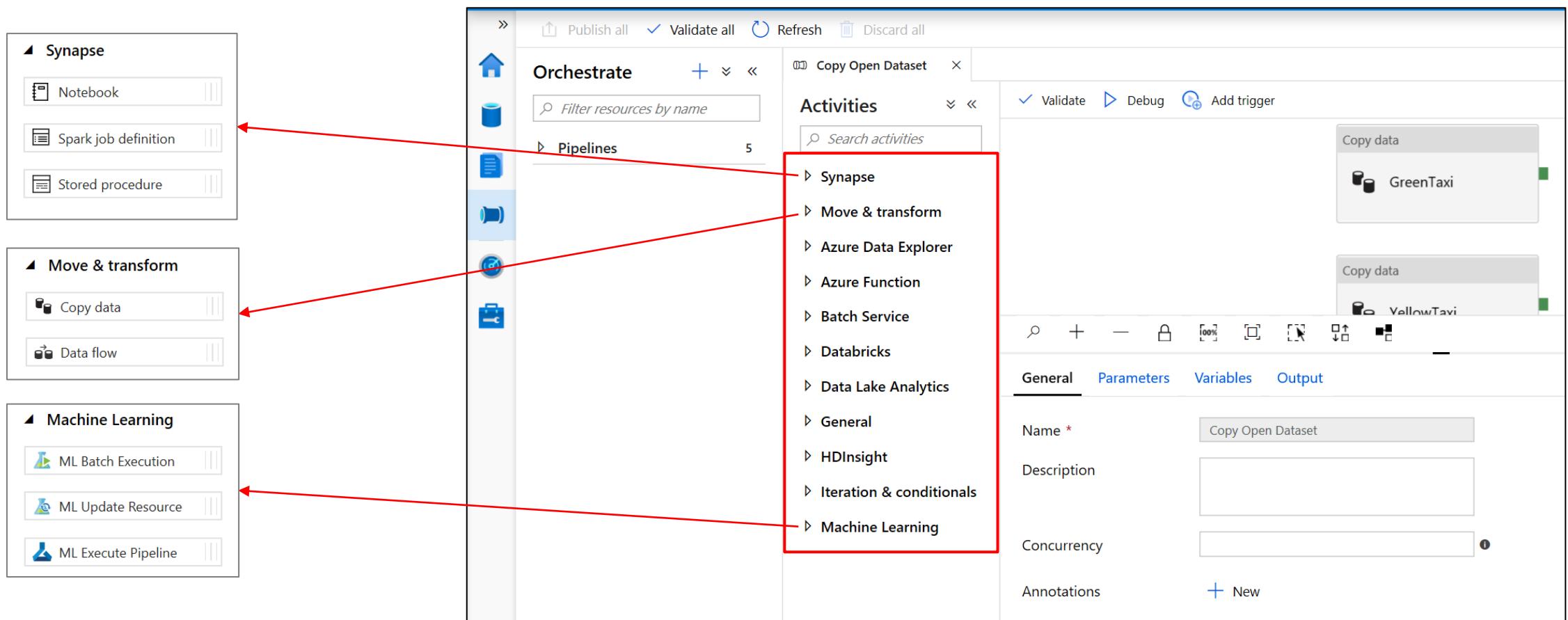


# Synapse Studio **Data Ingestion**

# Integrate Hub

It provides ability to create pipelines to ingest, transform and load data with 90+ inbuilt connectors.

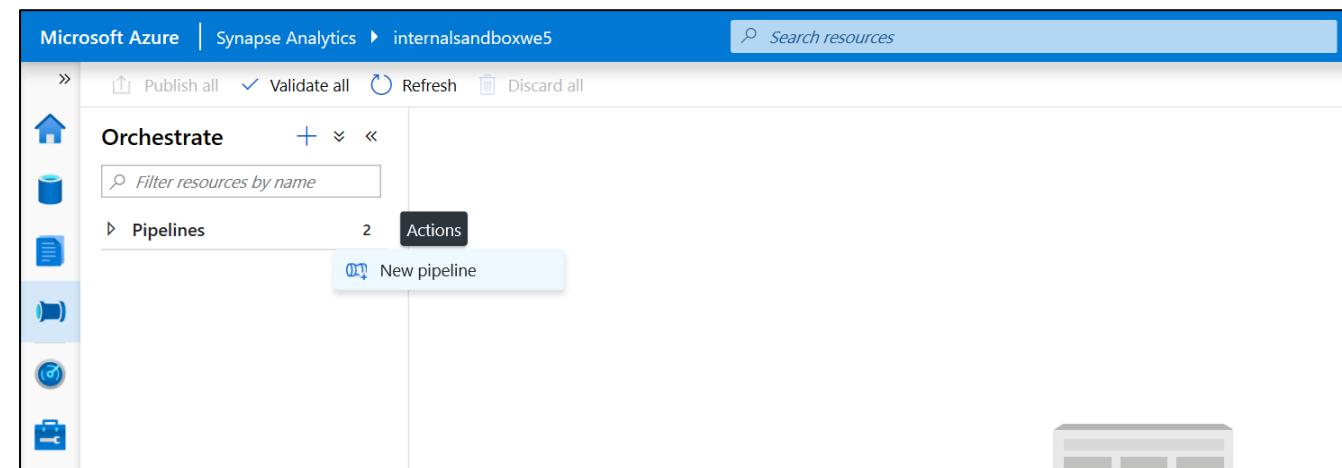
Offers a wide range of activities that a pipeline can perform.



# Pipelines

## Overview

It provides ability to load data from storage account to desired linked service. Load data by manual execution of pipeline or by orchestration

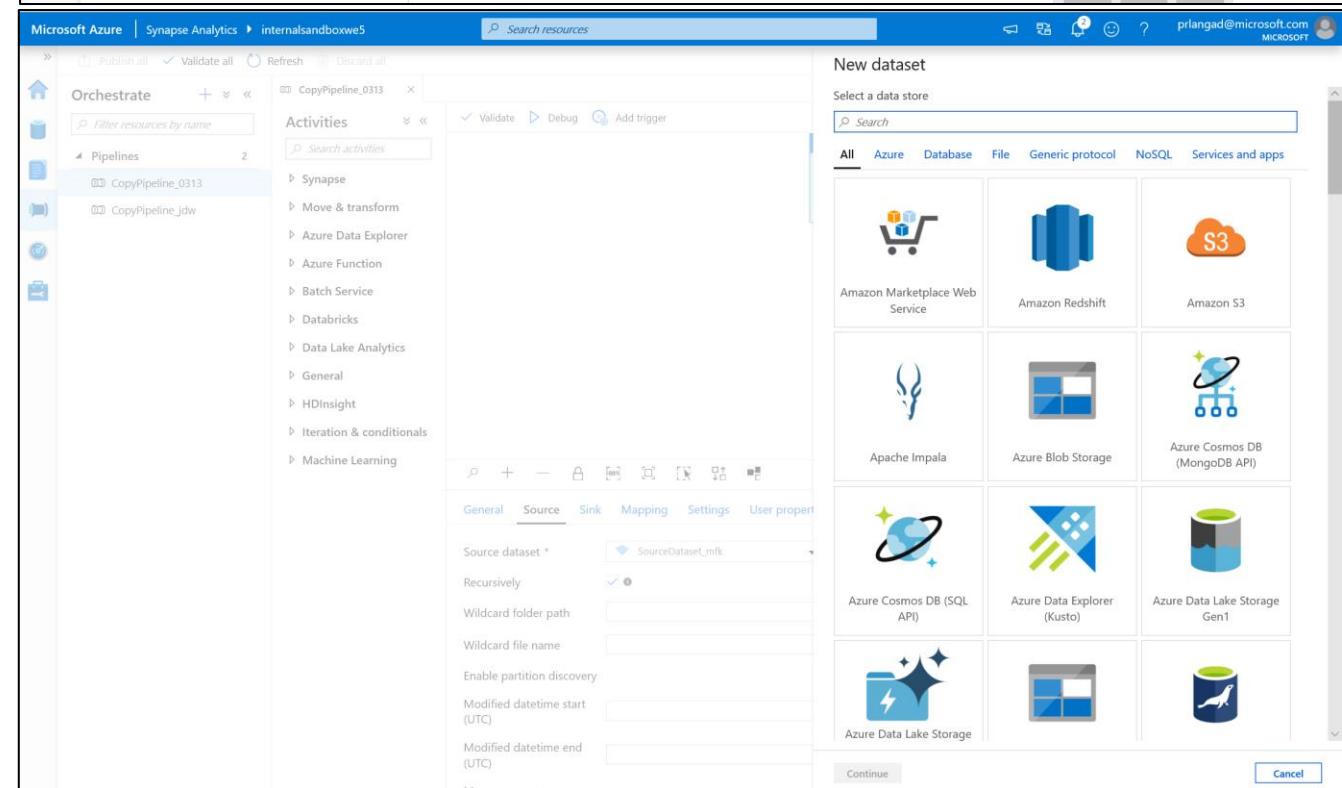


## Benefits

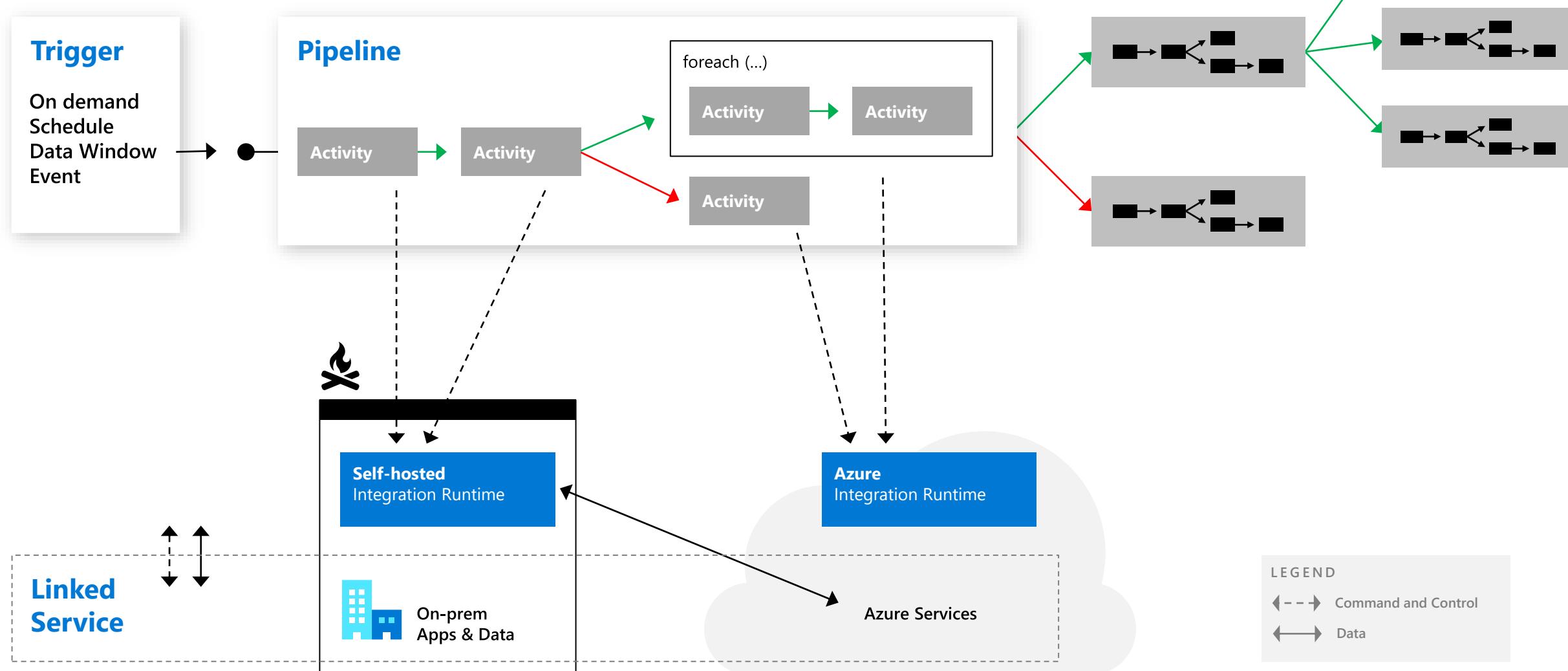
Supports common loading patterns

Fully parallel loading into data lake or SQL tables

Graphical development experience



# Orchestration @ Scale



# Data Movement

## Scalable

per job elasticity

Up to 4 GB/s

## Simple

Visually author or via code (Python, .NET, etc.)

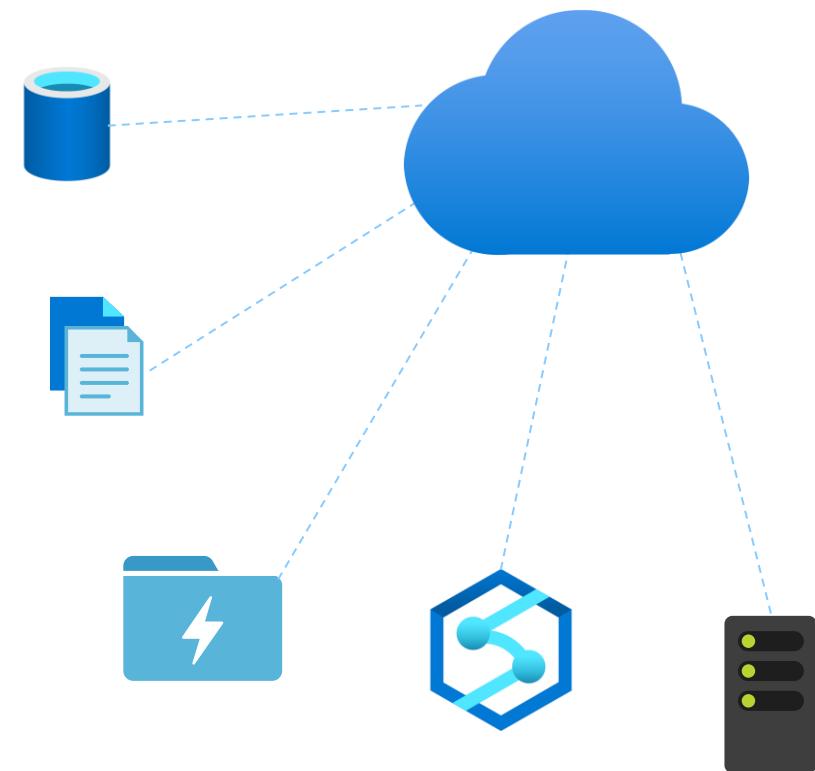
Serverless, no infrastructure to manage

## Access all your data

90+ connectors provided and growing (cloud, on premises, SaaS)

Data Movement as a Service: 25 points of presence worldwide

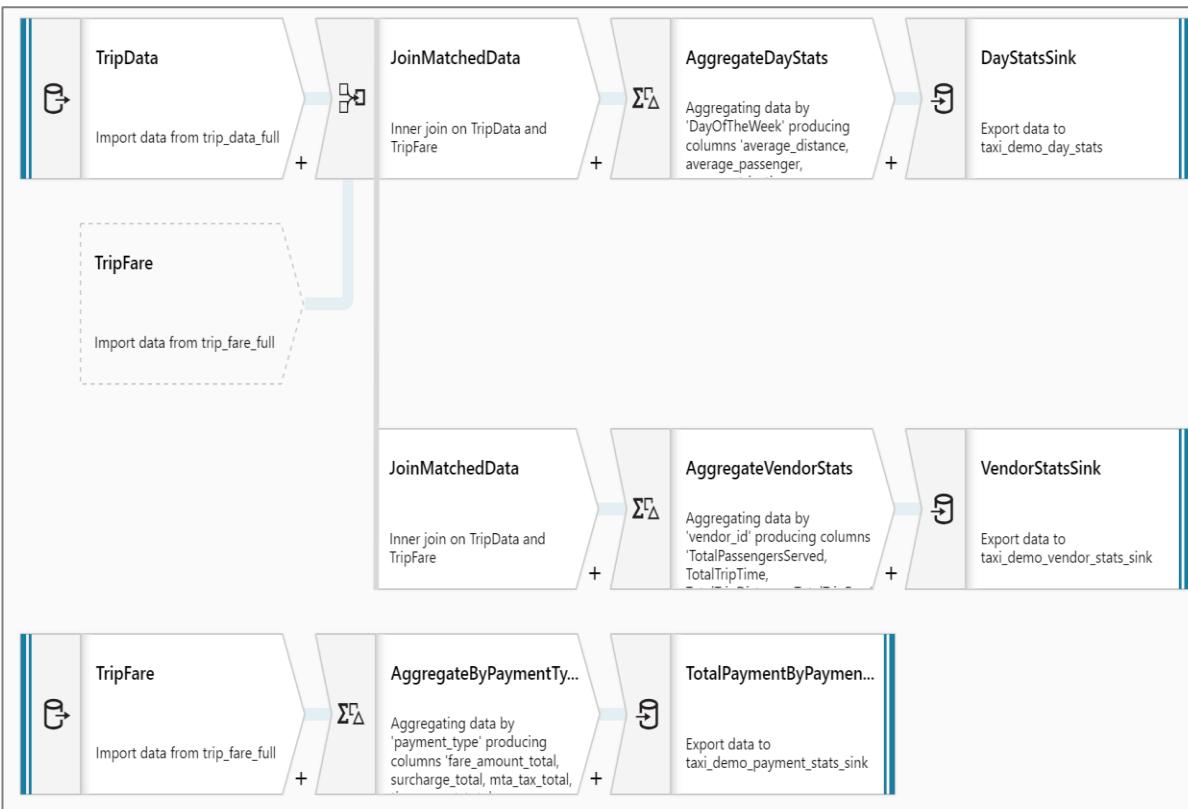
Self-hostable Integration Runtime for hybrid movement



# Prep & Transform Data

## Mapping Dataflow

Code free data transformation @scale



## Power Query

Code free data preparation @scale

The screenshot shows the Microsoft Power Query Editor interface within the Azure Synapse Analytics workspace. The main area displays a table of data with 8 columns and 99+ rows. The columns are labeled: ab\_storeid, ab\_productCode, 12\_quantity, 1.2\_logQuantity, ab\_advertising, ab\_price, ab\_weekStarting, and ab\_id. The data consists of various surface.go entries with corresponding values for each column.

The Power Query ribbon at the top includes tabs for Home, Transform, Add column, and View. The Transform tab is selected, showing various tools like Enter data, Options, Manage parameters, Refresh, Advanced editor, Properties, and a large set of data wrangling functions. The Home tab shows the current query name as PQSalesPrep and indicates it's a dev branch.

On the right side, there are panes for Query settings (with applied steps like 'Changed col...'), a resources pane, and a step history pane indicating the step was completed in 2.95 seconds.

# Triggers

## Overview

Triggers represent a unit of processing that determines when a pipeline execution needs to be kicked off.

Data Integration offers 3 trigger types as –

1. Schedule – gets fired at a schedule with information of start date, recurrence, end date
2. Event – gets fired on specified event
3. Tumbling window – gets fired at a periodic time interval from a specified start date, while retaining state

It also provides ability to monitor pipeline runs and control trigger execution.

The screenshot shows the Azure Synapse Analytics Data Integration interface. On the left, there's a navigation sidebar with options like 'Analytics pools', 'SQL pools', 'Apache Spark pools', 'External connections', 'Linked services', 'Orchestration', 'Triggers' (which is selected and highlighted in blue), 'Integration runtimes', 'Security', 'Access control', and 'Managed Virtual Networks'. The main area is titled 'Triggers' with the sub-instruction: 'To execute a pipeline set the trigger. Triggers represent a unit of processing that determines when a pipeline execution needs to be kicked off.' Below this, there's a search bar with the placeholder 'Search to filter items...' and a table showing one item: 'HolidayUpdateTrigger' (Type: Schedule, Status: Started). At the top, there's a 'New trigger' dialog box with fields for Name (set to 'Trigger 1'), Description, Type (selected 'Schedule'), Start Date (set to '10/30/2019 11:20 PM'), Recurrence (set to 'Every 1 Minute(s)'), End (set to 'No End'), Annotations (with a '+ New' button), and Activated (set to 'Yes').

## Demonstration Two

- **Setup Integration Runtime**
- **Create Synapse Pipeline**
- **Data Ingestion from On-Premises Server**



## Demonstration Three

- **Data Ingestion from On-Premises Server (Lookup)**
- **Structured Data to ADLS**



## Demonstration Four

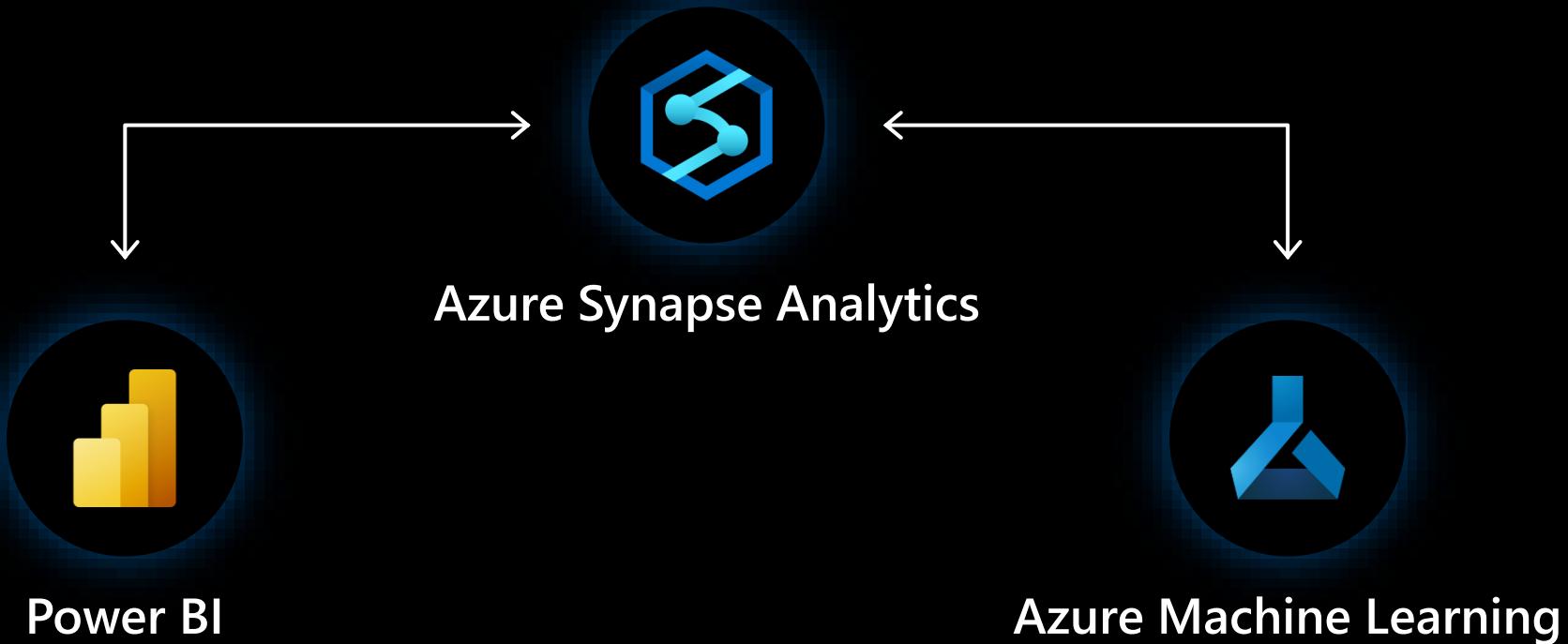
- Ingest Data using Polybase
- Ingest Data using COPY INTO



# Power BI + Azure Synapse

---

An unmatched combination



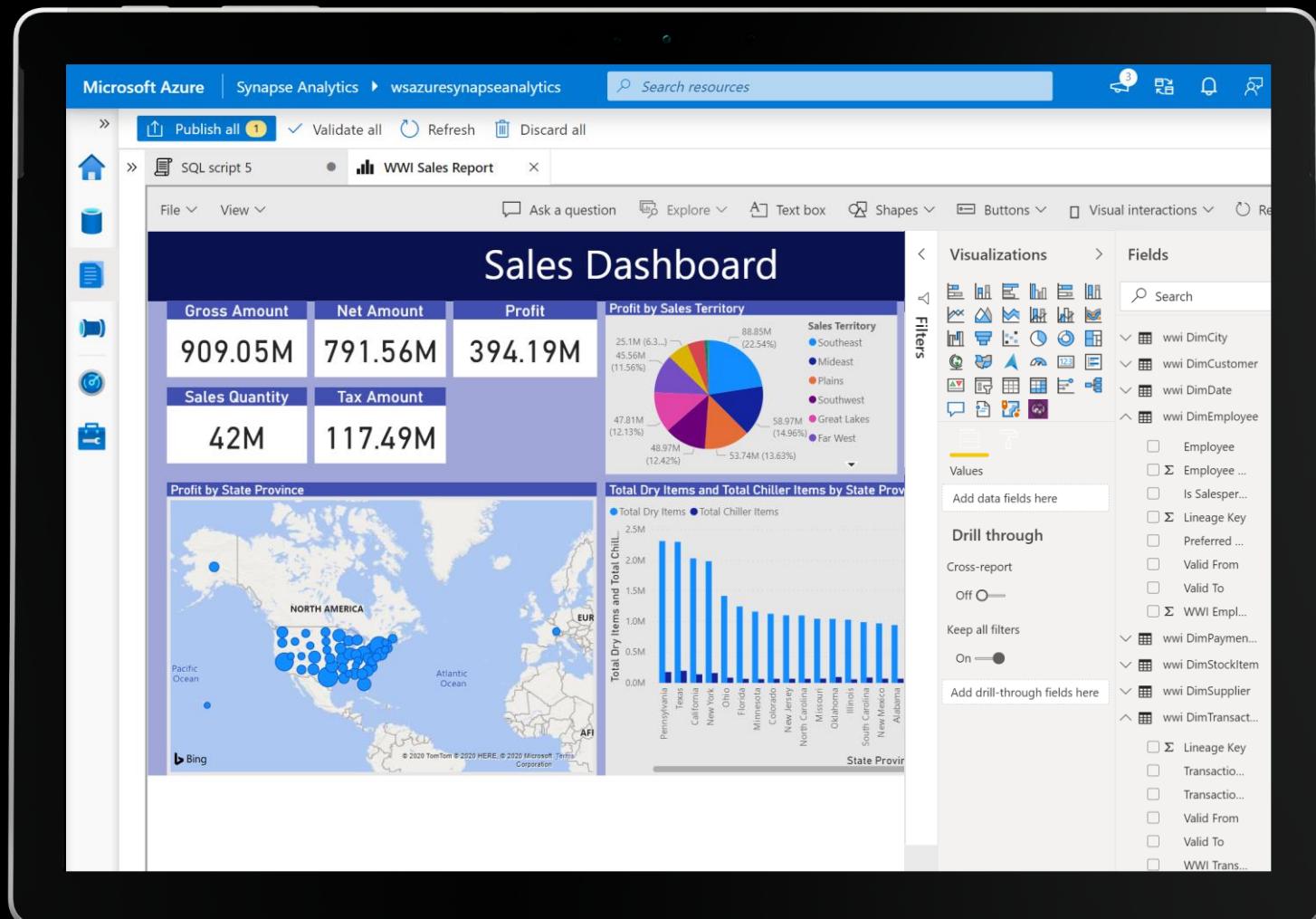
Unified experience to enrich data  
Automated ML for rapid development  
Seamless collaboration

# Power BI integration

## Build dashboard in Synapse Studio

Code-free experience for development rich visualizations

One-click publishing to for secure consumption across the enterprise

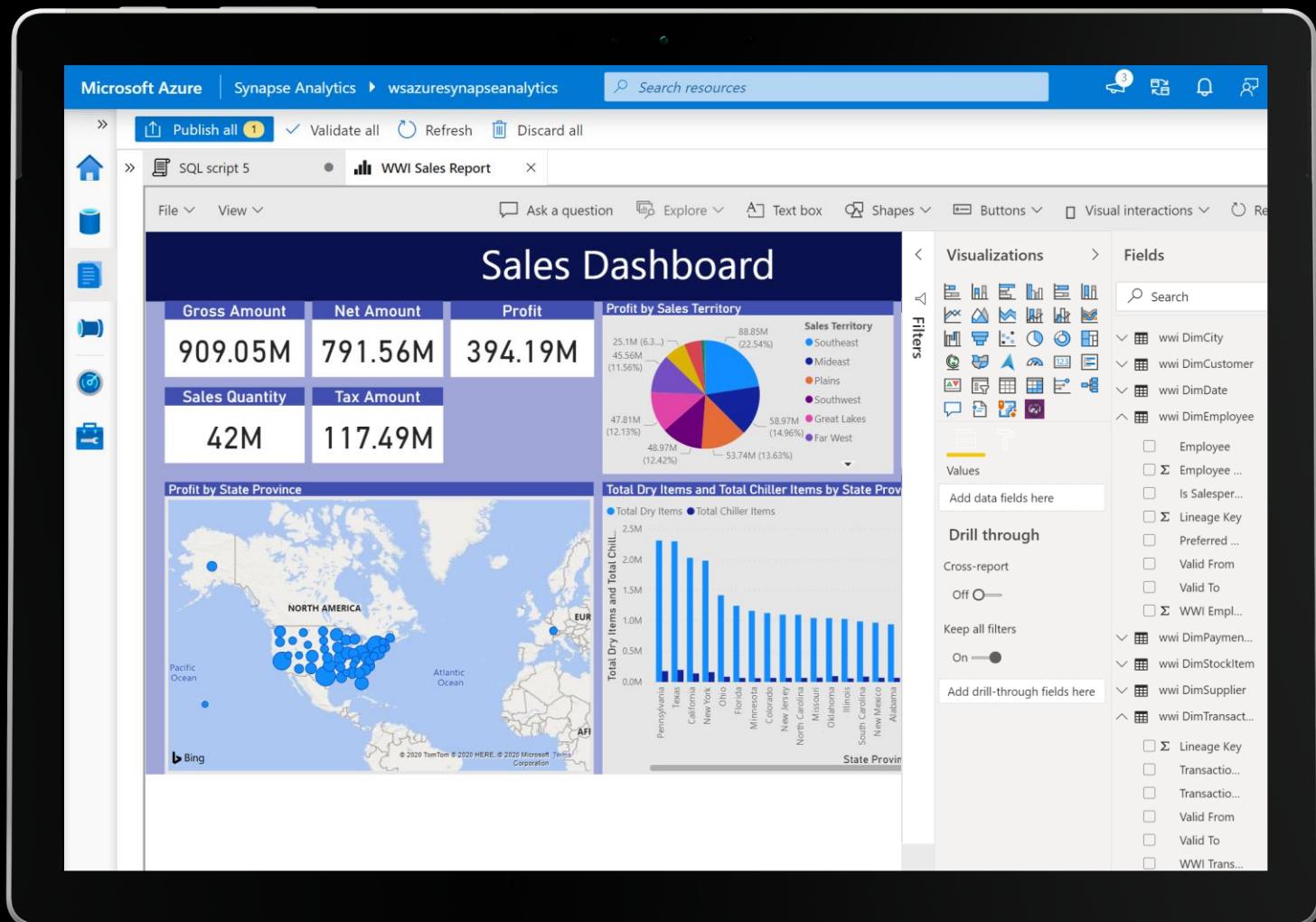


# Power BI integration

## Build dashboard in Synapse Studio

Code-free experience for development rich visualizations

One-click publishing to for secure consumption across the enterprise

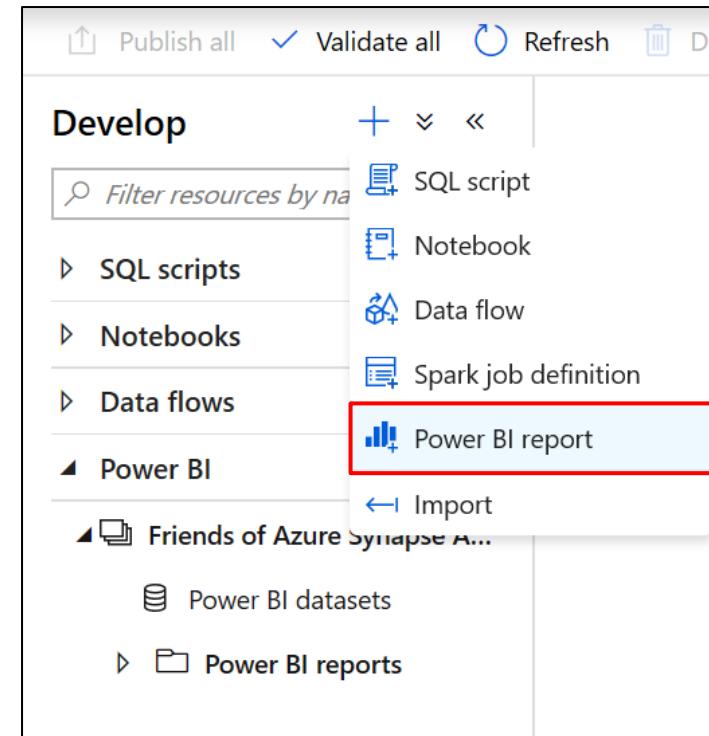


# Develop Hub – Power BI

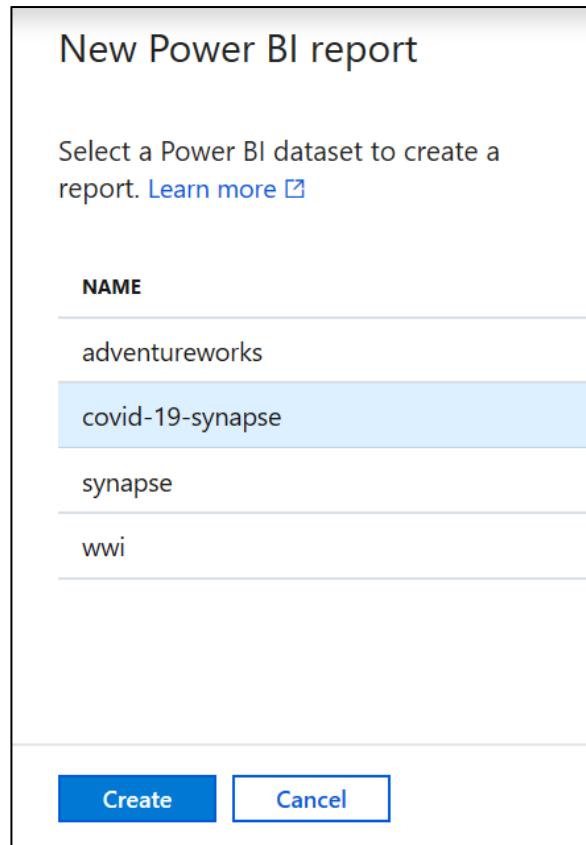
Create new reports from existing published Power BI datasets

Create new Power BI datasets

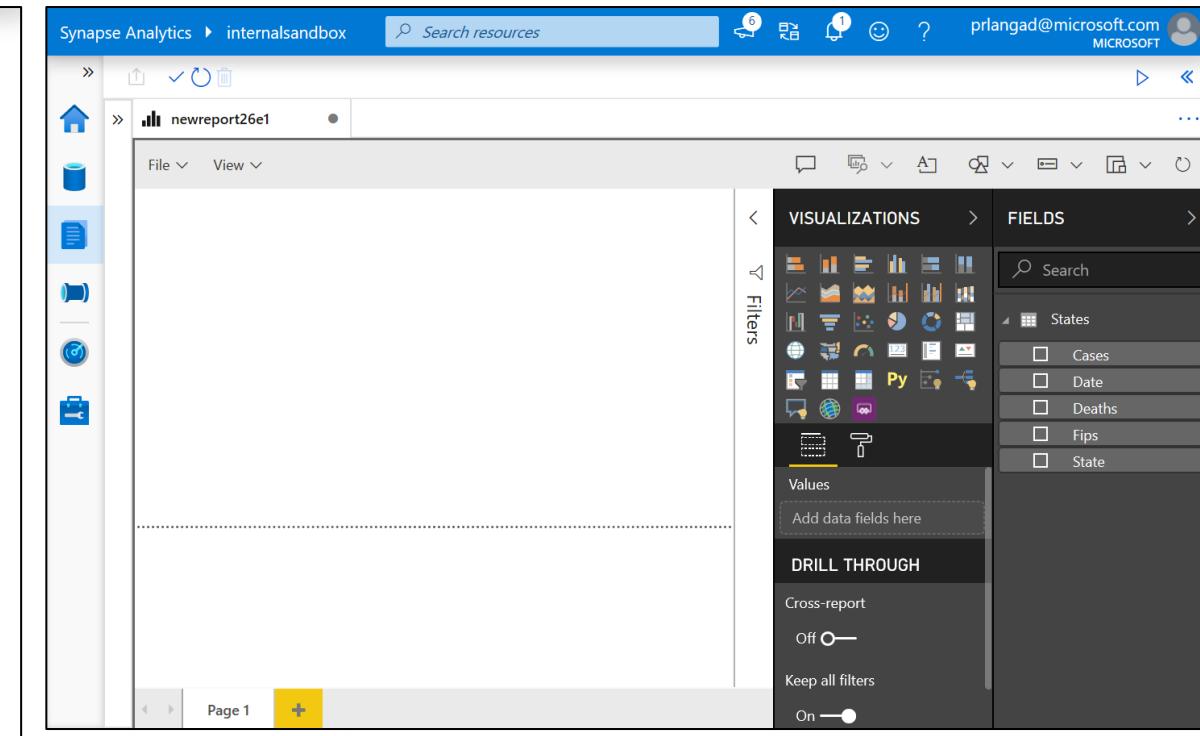
1



2



3



# Develop Hub – Power BI

View published reports in Power BI workspace

The screenshot shows the Microsoft Azure Power BI workspace interface. The left sidebar displays the 'Develop' section with a list of resources: SQL scripts (9), Notebooks (6), Data flows (1), Power BI (1), and a folder named 'gaming-telemetry' containing Power BI datasets and reports. The 'Report' item under 'Power BI reports' is selected.

The main content area displays a report titled 'GAME STUDIO'. The report features a header with a game console image and a 'Total Users' summary of 24.5M. It includes a 'What If...' card asking about increasing free game addons, showing a forecast of 7,361,707 users and an increase of 252.8K. Below this are two cards: 'Users (Forecast)' showing 7,361,707 users (7,346,291 last month) and 'Extra Users' showing a 3.4% increase. The central part of the report contains a table titled 'Total Users vs "What If" Analysis' comparing actual users, forecasts, and extra users across regions and age groups. To the right is a line chart titled '"What If" Analysis Forecast' showing user growth from August 2019 to November 2019.

The right sidebar contains sections for 'VISUALIZATIONS' and 'FIELDS'. The 'VISUALIZATIONS' section lists various chart types, and the 'FIELDS' section lists categories like agegroup, forecast, historical, platform, predictions, realtime, regions, scenario, and weekdays. There are also options for 'DRILL THROUGH', 'Cross-report', 'Keep all filters', and 'Add drill-through fields here'.

# Get started today



Create a free Azure account and get started with Azure Synapse Analytics:

<https://azure.microsoft.com/en-us/free/synapse-analytics/>



Submit info for free proof of value package for modernizing SAP workloads:

<https://aka.ms/synapse-qlik>



Submit info for free proof of value package for migrating on-prem data warehouse:

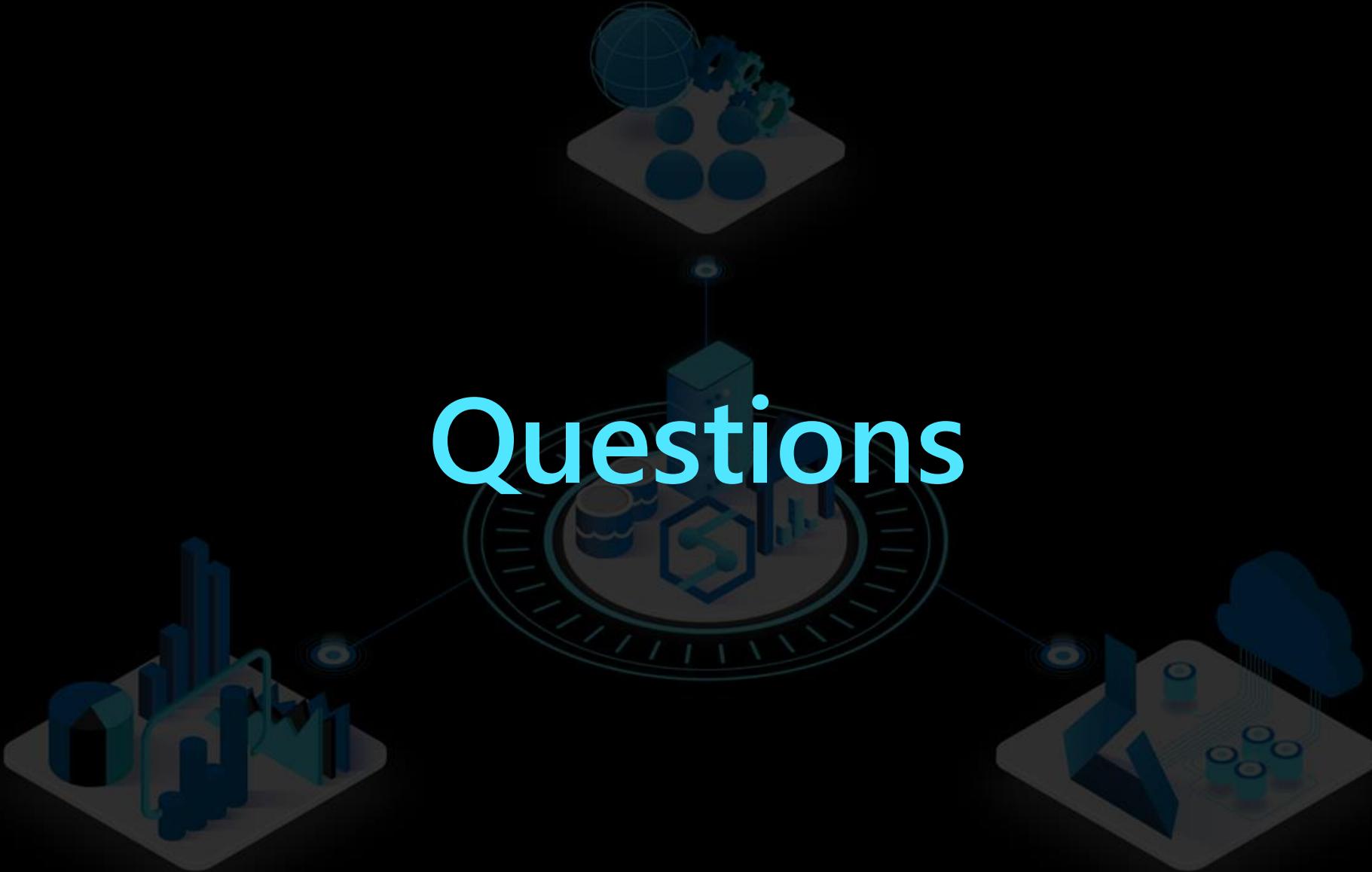
<http://aka.ms/synapse-informatica>



Learn more:

<https://aka.ms/synapse>

# Questions



Dankie Faleminderit **Shukran** Chnorakaloutioun Hvala Blagodaria  
Děkuji **Tak** Dank u Tänan Kiitos **Merci** Danke Ευχαριστώ A dank  
Mahalo මතිල. **Dhanyavād** Köszönöm Takk Terima kasih **Grazie** Grazzi

# Thank you!

감사합니다 Paldies Choukrane Ačiū **Благодарам** ありがとうございました  
謝謝 Баярлалаа **Dziękuję** Obrigado Multumesc **Спасибо** Ngiyabonga  
Ďakujem Tack Nandri Kop khun **Teşekkür ederim** Дякую Хвала Diolch

