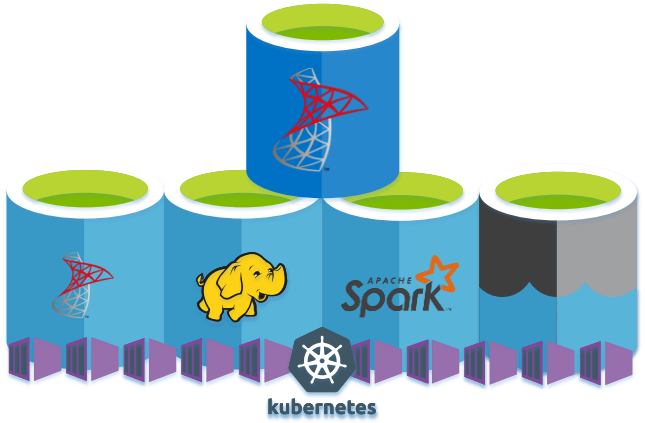
SQL Server 2019

Big Data Clusters



Paul Andrew | Principal Consultant & Solution Architect

Microsoft® Most Valuable Professional

@MrPaulAndrew

Microsoft Partner
Microsoft
Gold Data Analytics
Gold Data Platform
Gold Cloud Platform

adatis

Slide in GitHub

https://github.com/mrpaulandrew

CommunityEvents

Demo code, content and slides from various community events.

C++

{Event/Location}-{Month}-{Year}

adatis

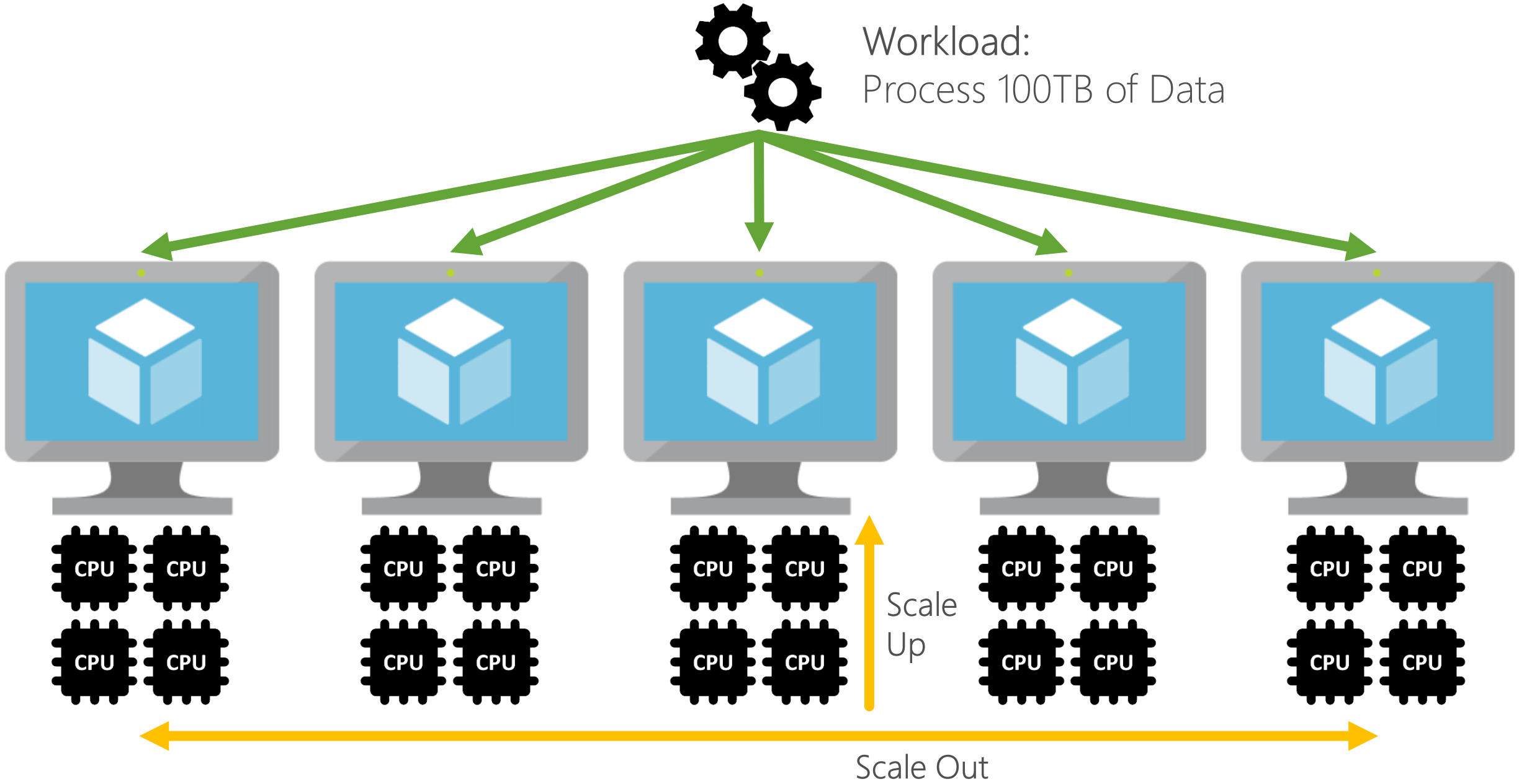
What is big data?

"Any data that you cannot process
in the time that you have/want
using the technology you have."

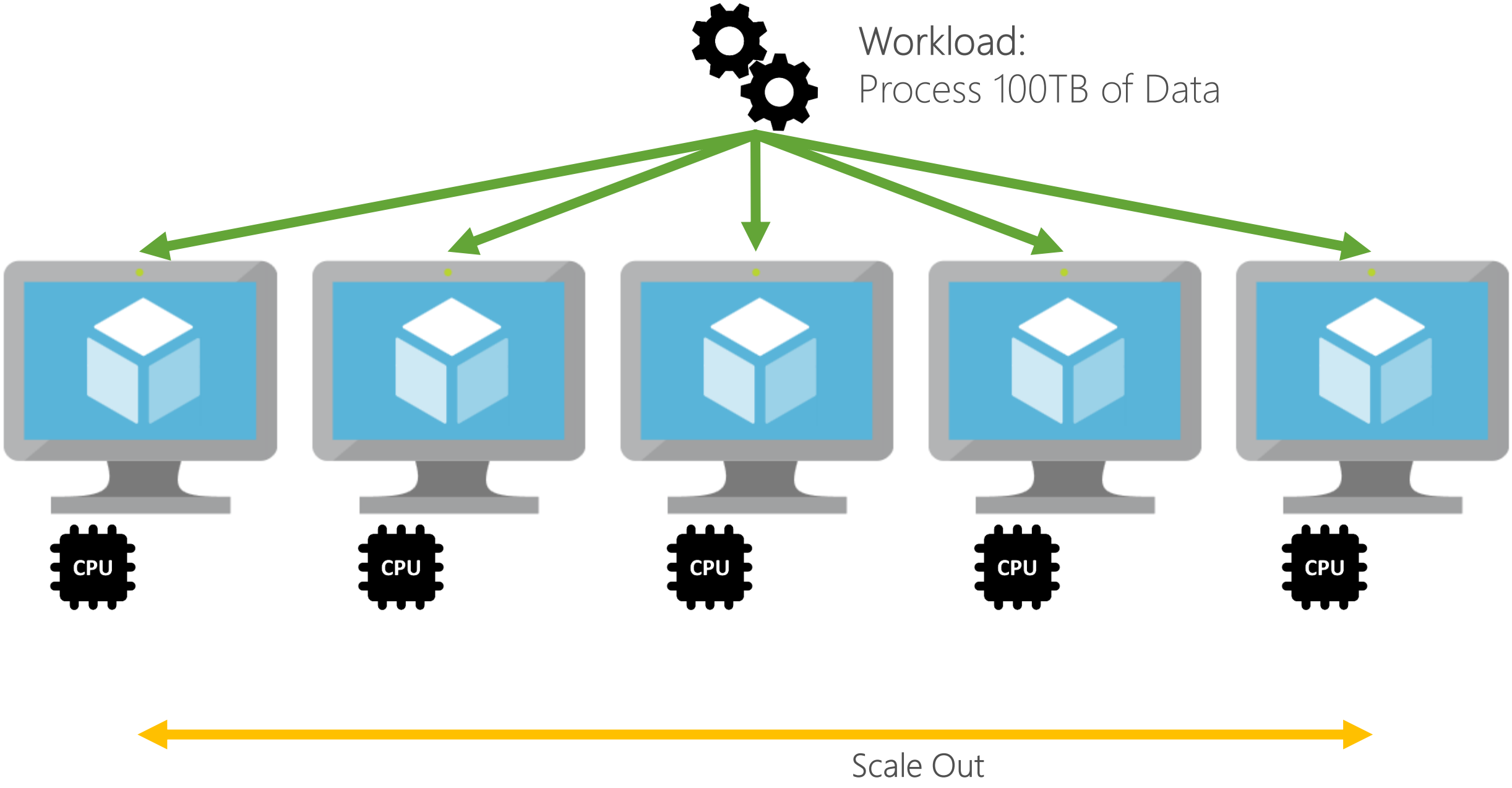
Buck Woody

adatis

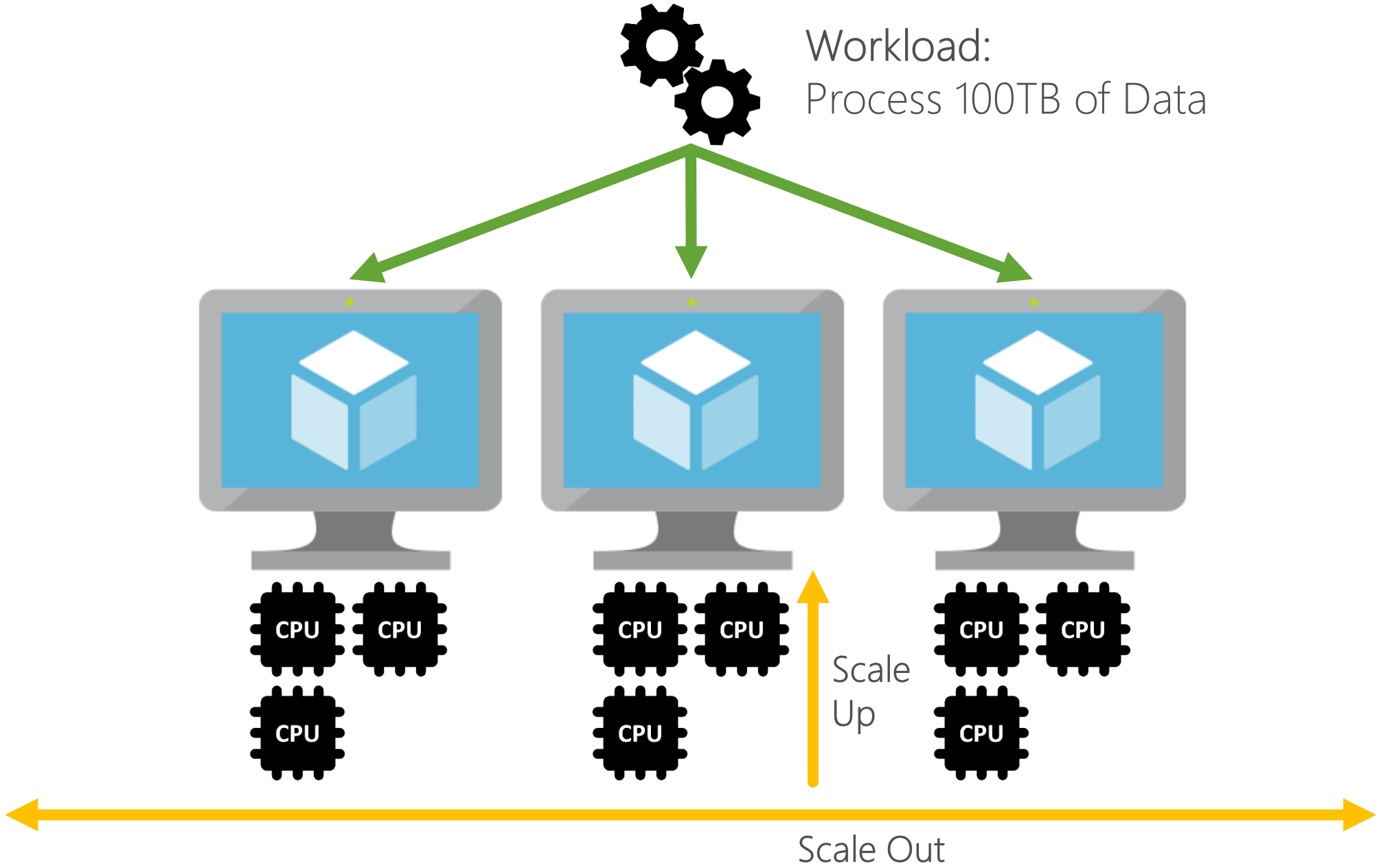
Scale Up and Scale Out



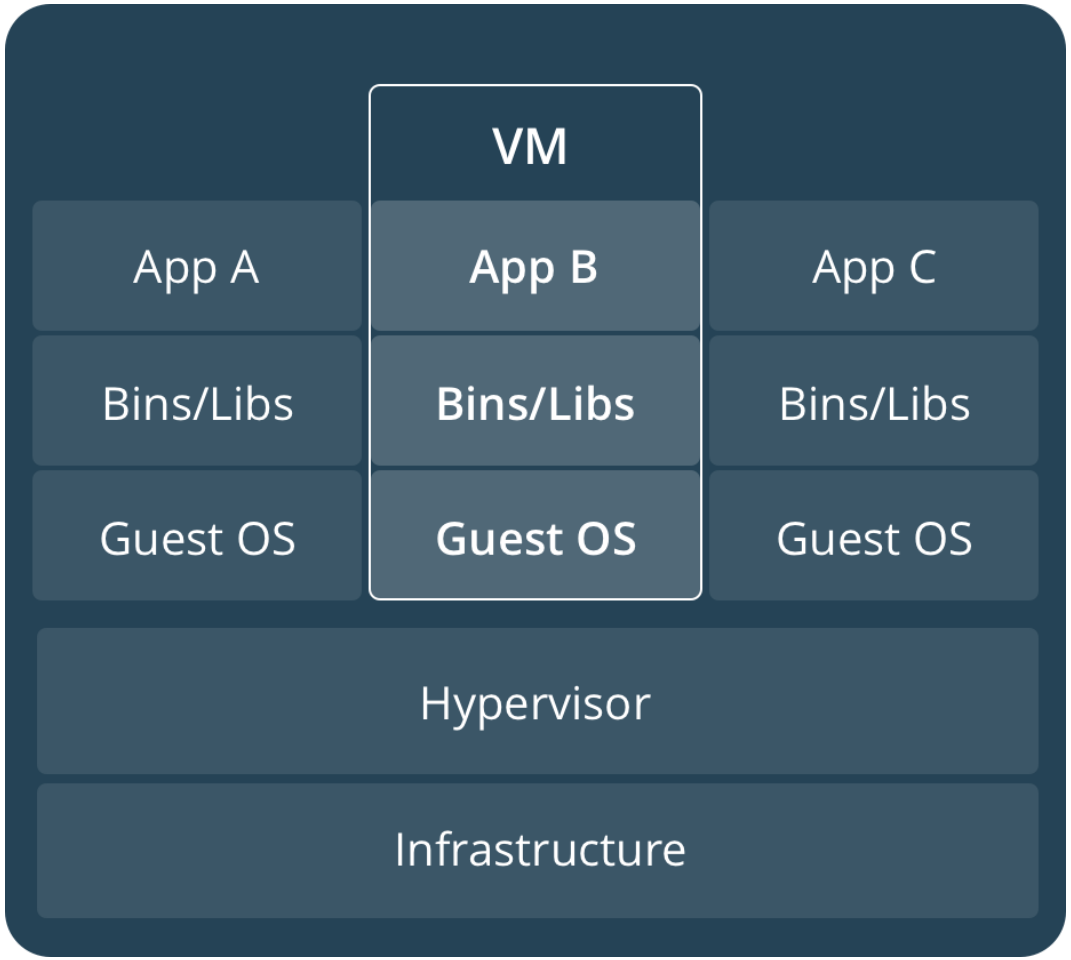
Scale Up and Scale Out

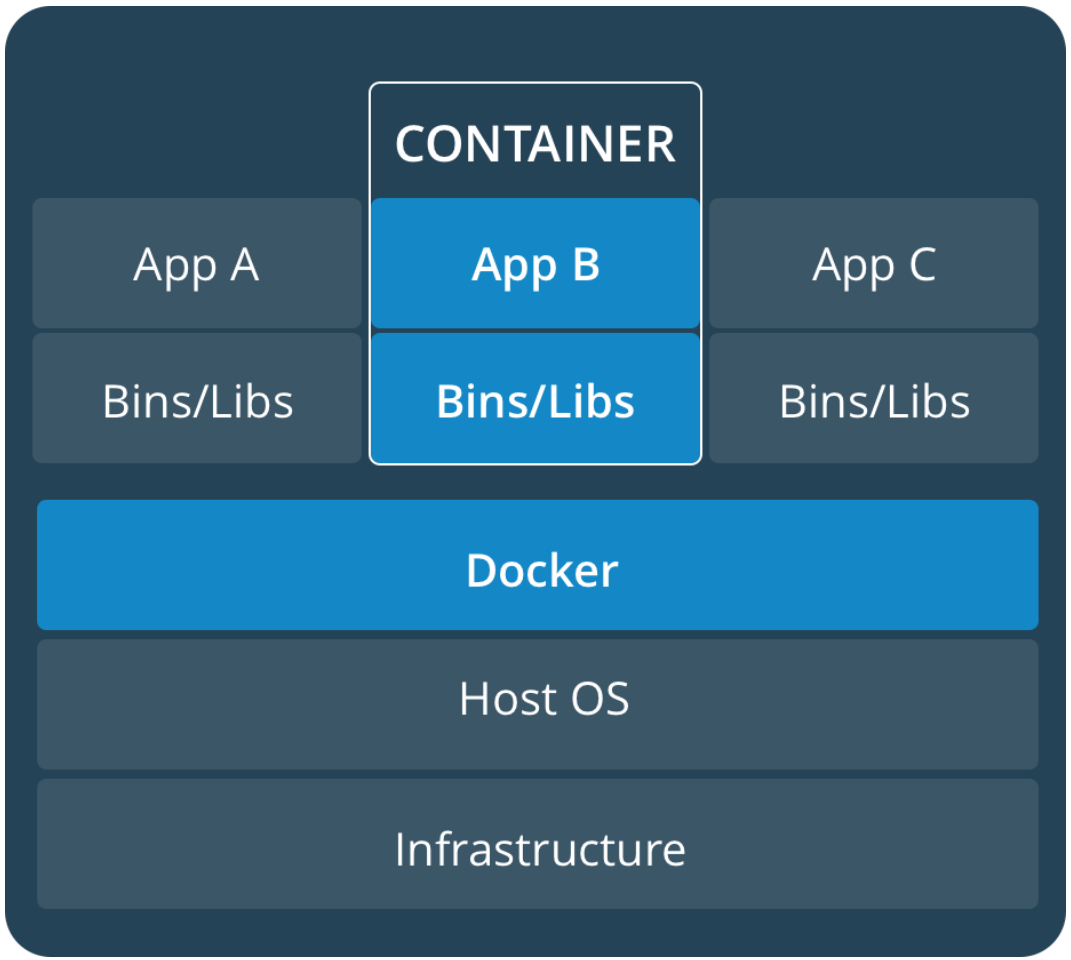


Scale Up and Scale Out



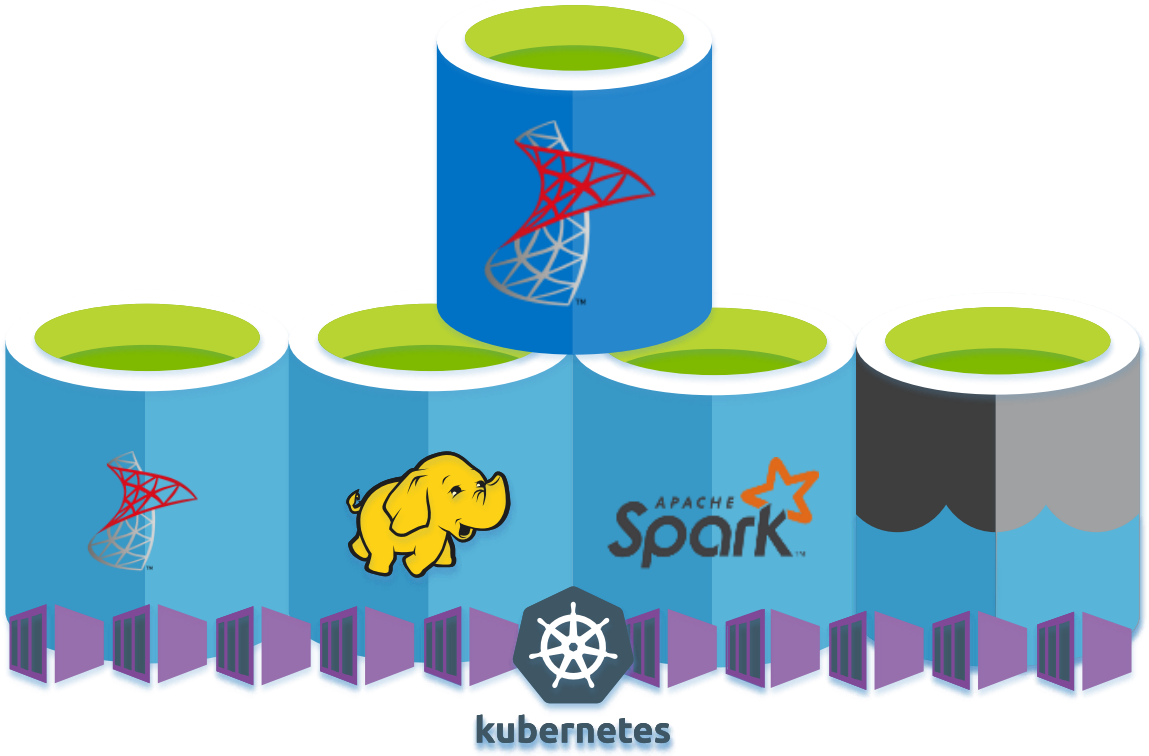
Containers





adatis

What



Why

How

What

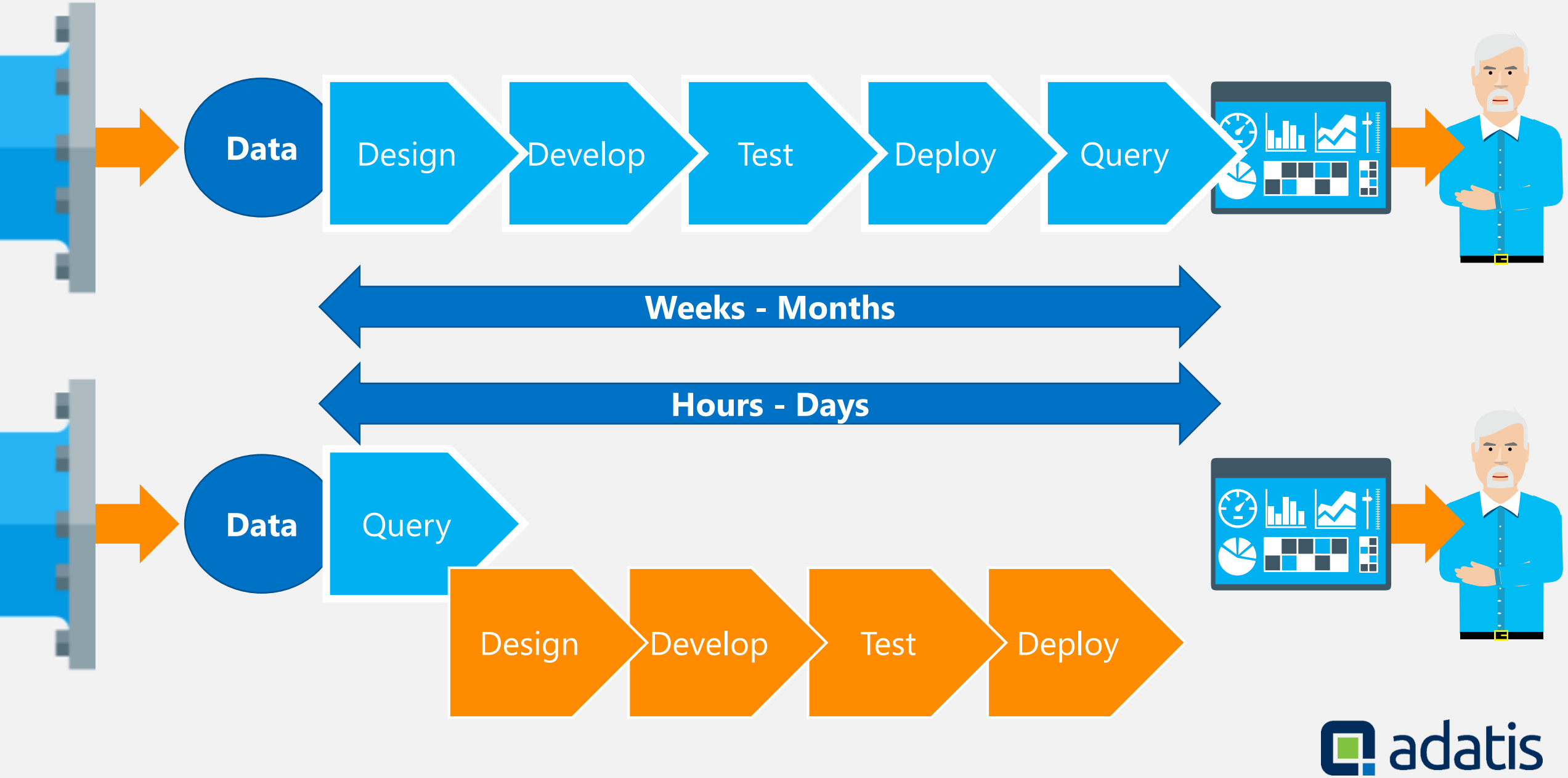
Why

How

Problem 1: All Forms of Data Structure

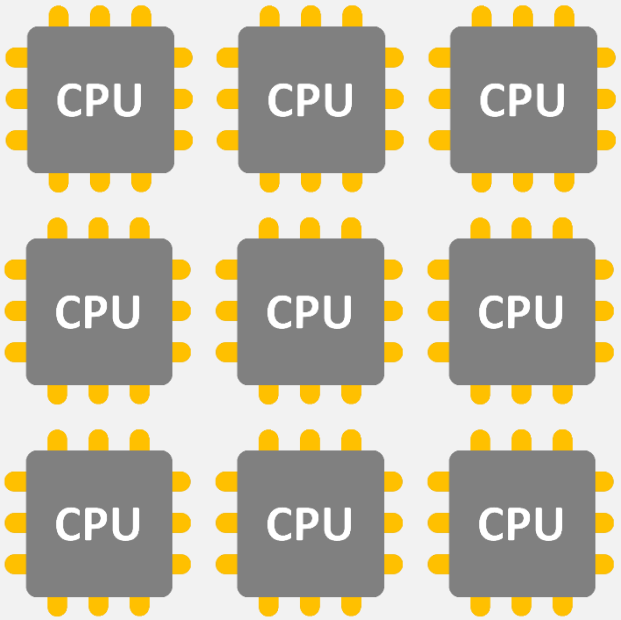
adatis

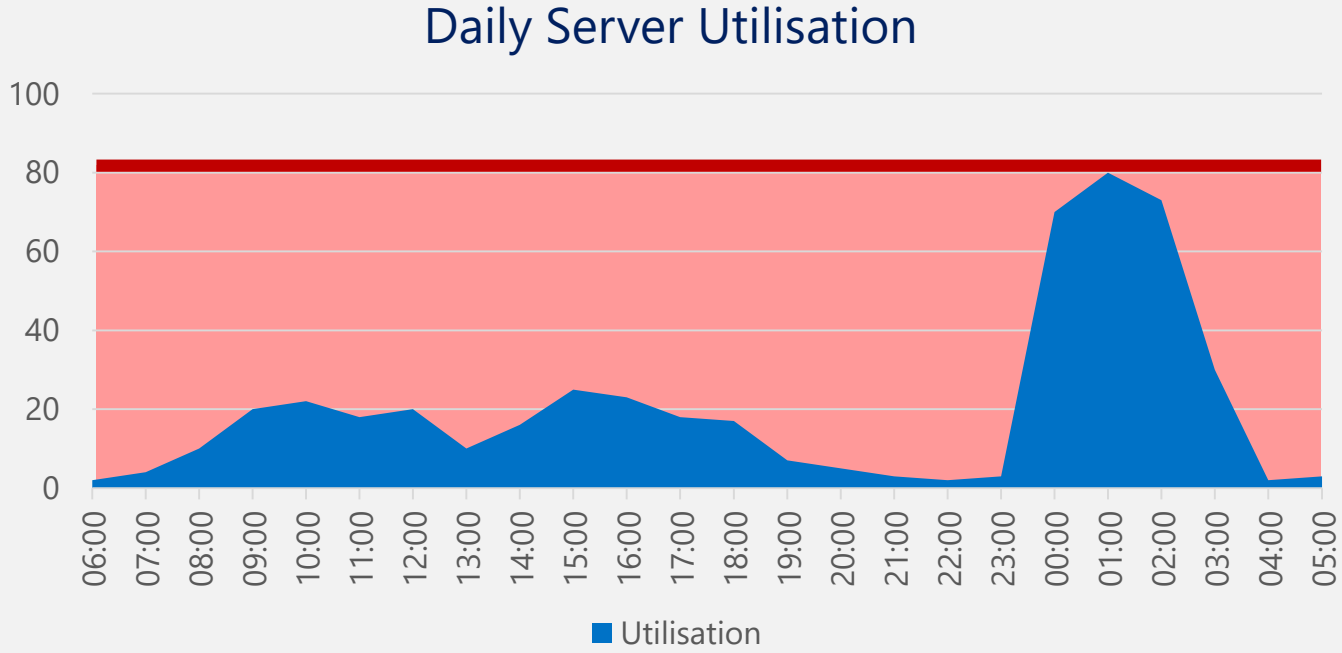
Problem 2: Quick Actionable Insights

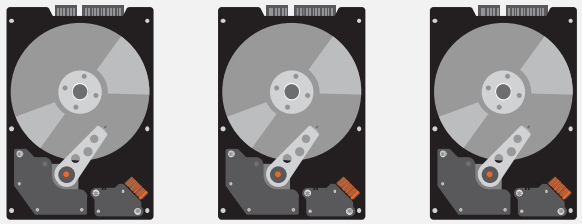


adatis

Problem 3: Infrastructure Costs & Scalability







£££

adatis

The Big Data V's

Volume

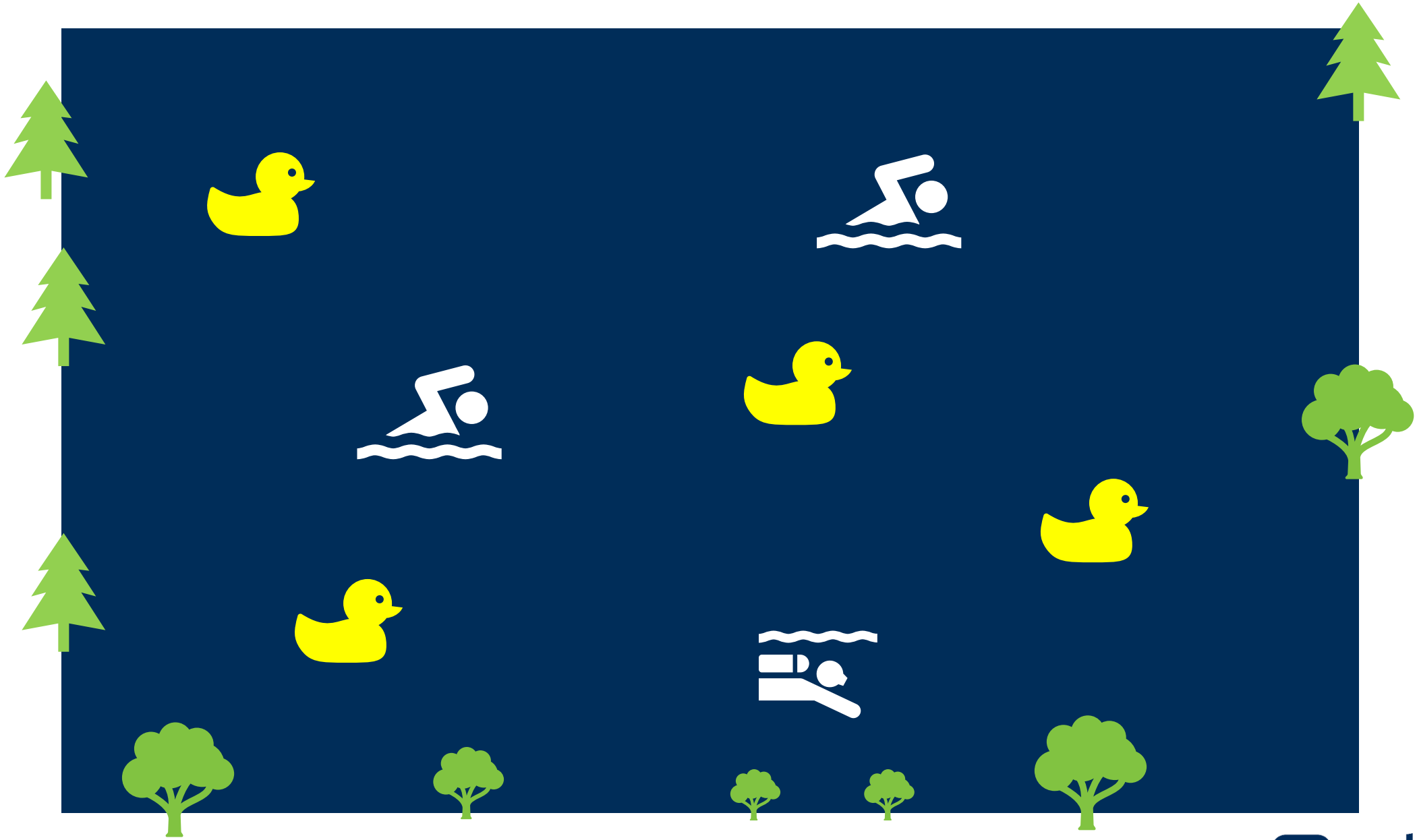
Velocity

Veracity

Variety

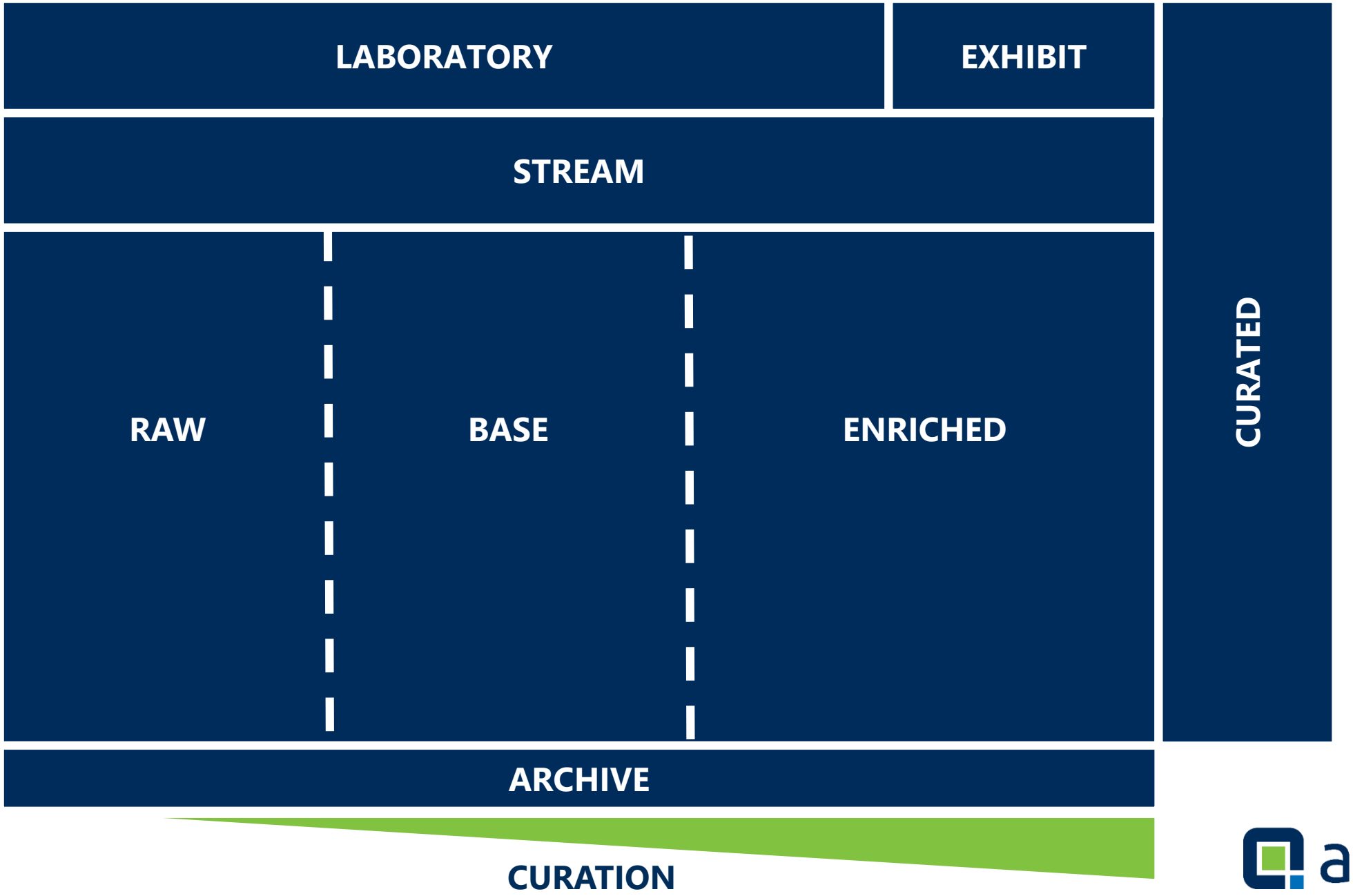
adatis

Data Lakes



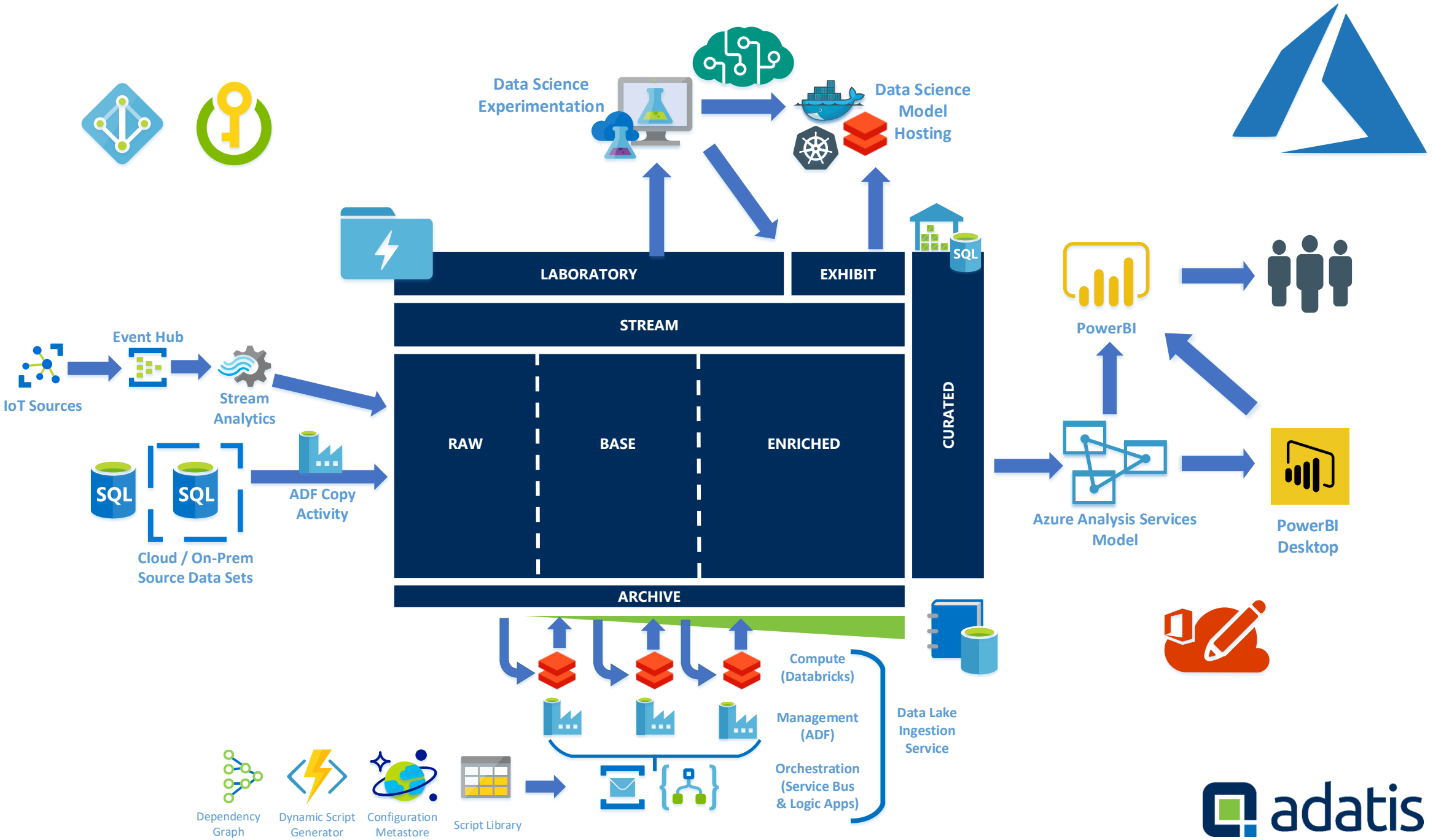
adatis

Data Lake – with Adatis Framework



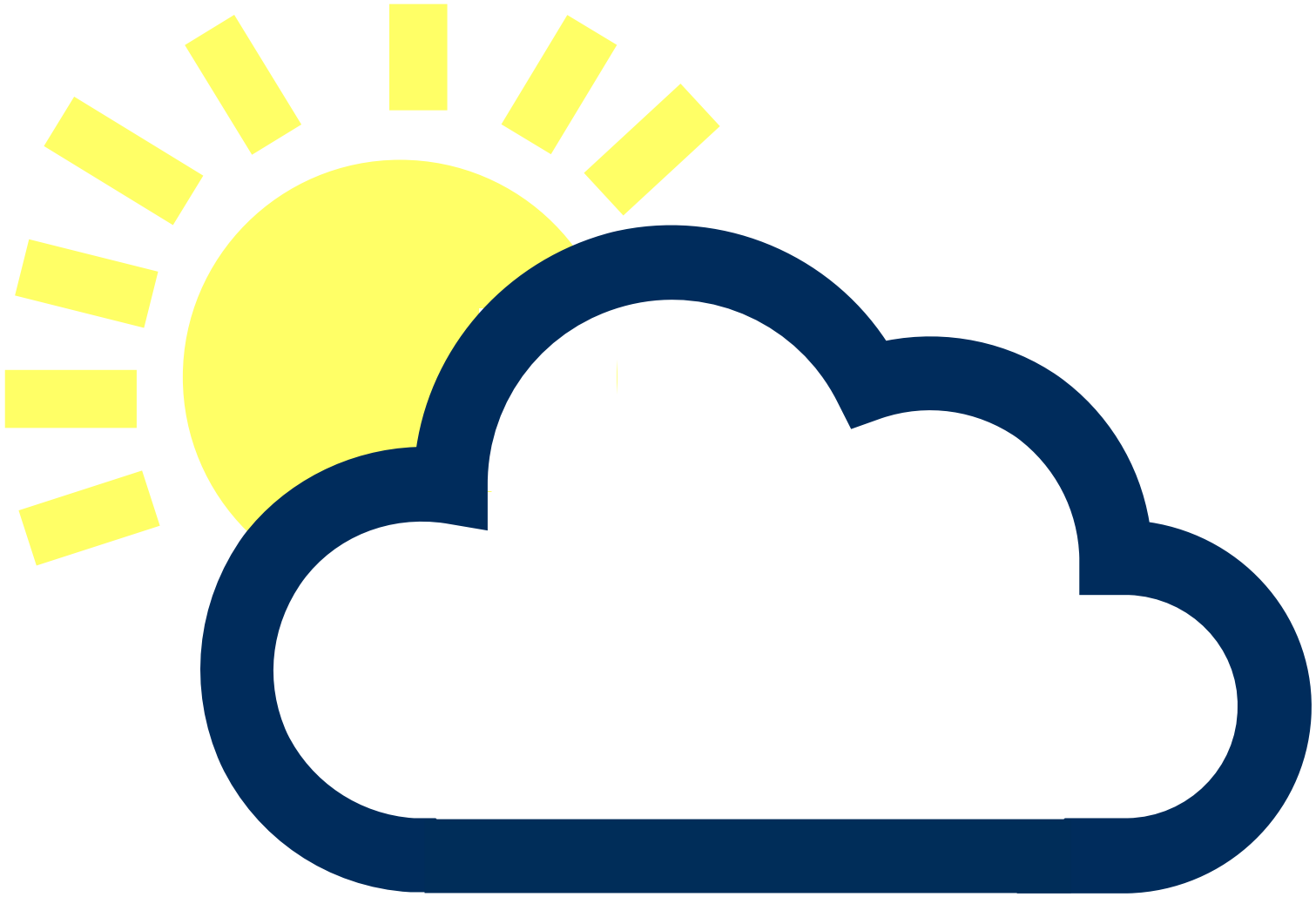
adatis

The Modern Data Warehouse



adatis

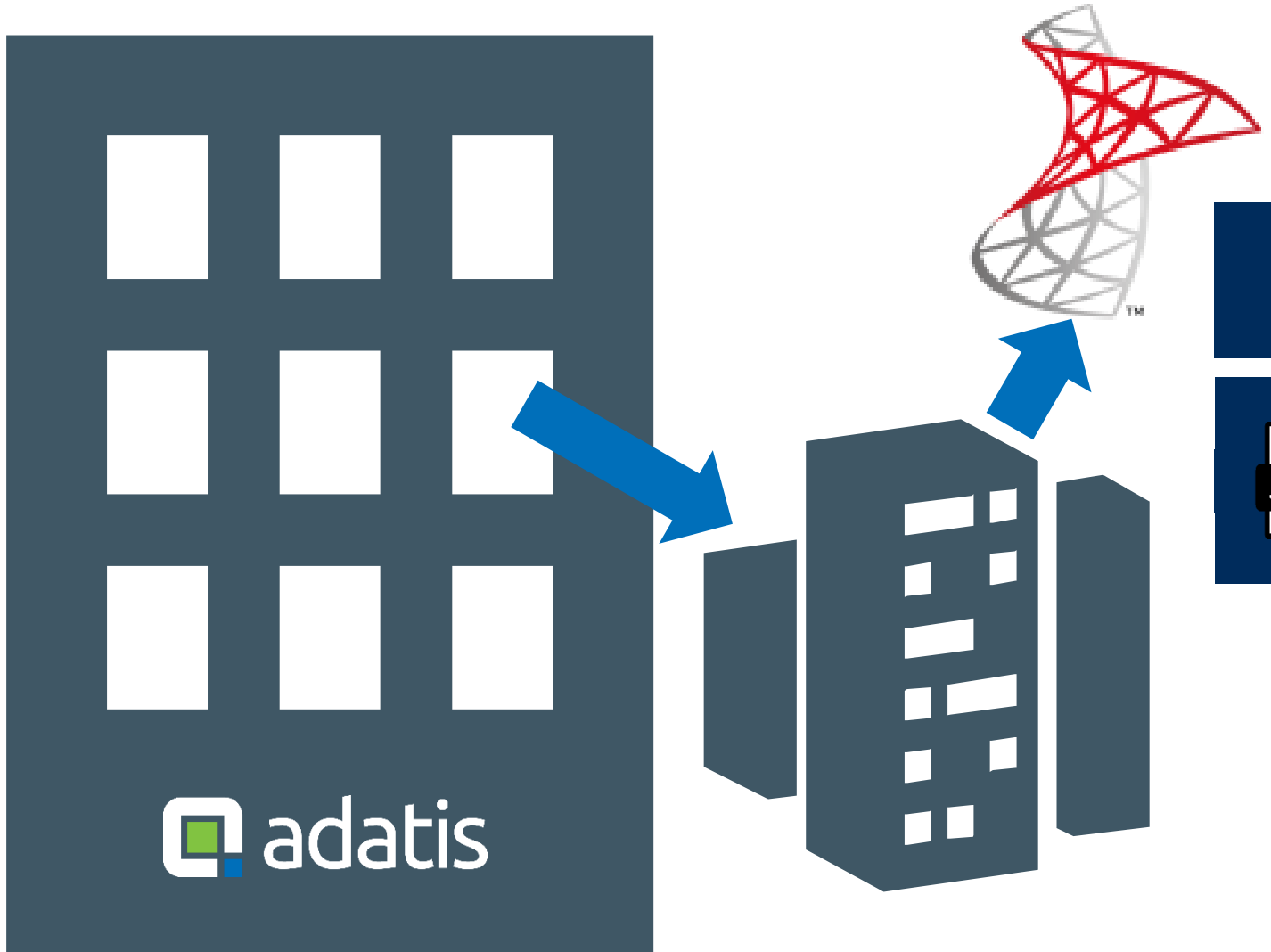
No Cloud?



adatis

adatis

On Premises



Microsoft SQL Server

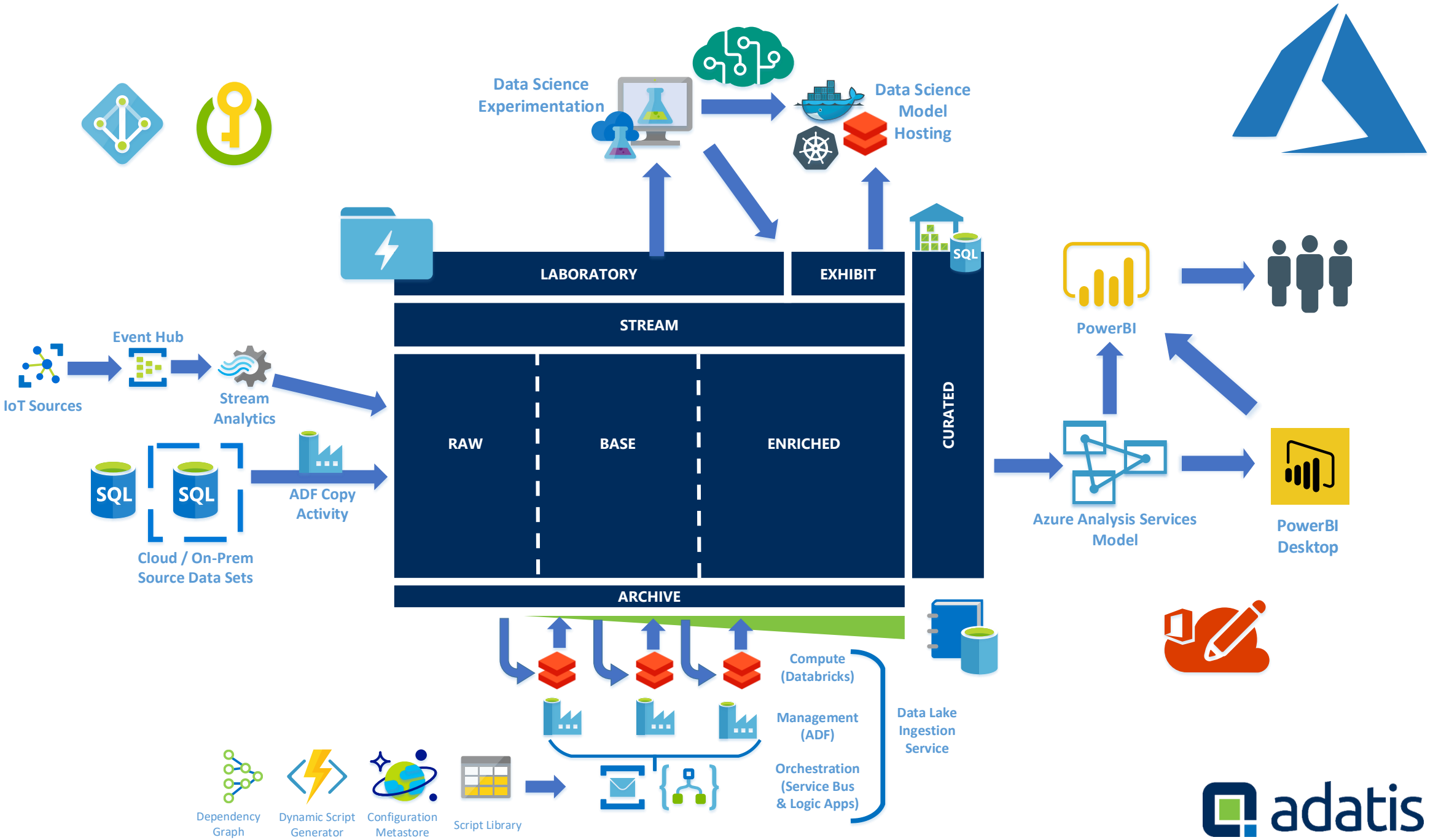
COMPUTE

JSON XML DATABASE STORAGE

DBMS SSIS SSAS

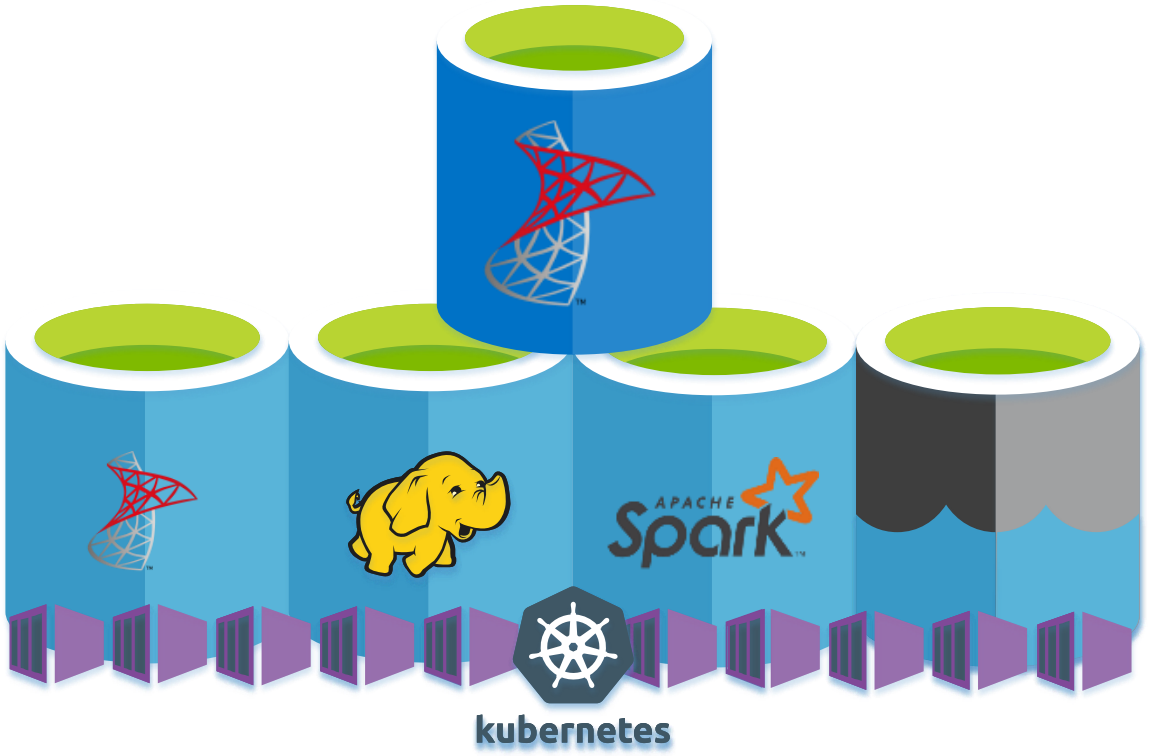
SSRS MDS

The Modern Data Warehouse



adatis

What



Why

How

What

Why

How

adatis

SQL Server - Product Evolution

Hyperscale Tier - 100 TB Databases
Docs Link

Azure SQL Database (SQLDB)

T-SQL

Compute

Compute Node

Data

Database Storage

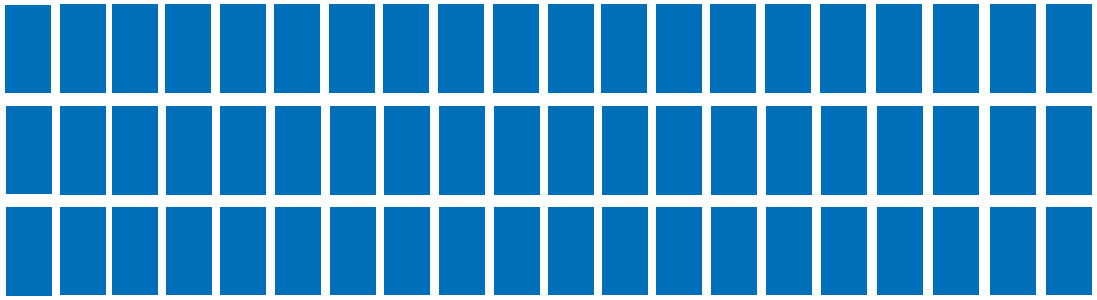
Scale Up Compute

Azure SQL Data Warehouse (SQLDW)

D-SQL

Control Node

Compute Node | Compute Node | Compute Node | Compute Node



Scale Out Compute & Distributed Storage

SQL Server - Product Evolution

Hyperscale Tier - 100 TB Databases
Docs Link

Azure SQL Database (SQLDB)

T-SQL

Compute

Compute Node | Read Only Node | Read Only Node

Data

Database Storage

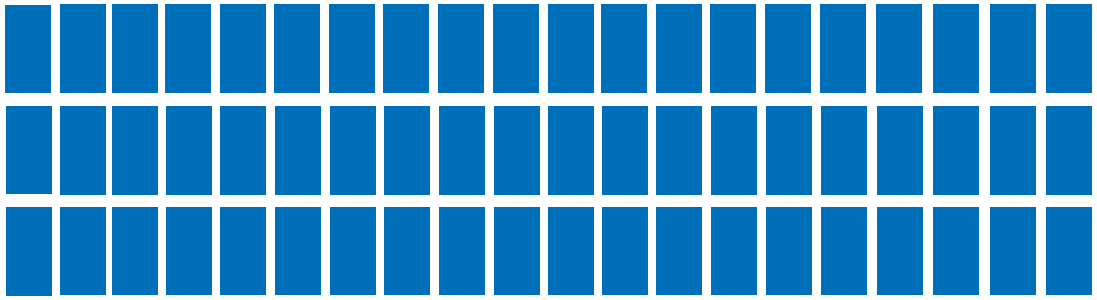
Scale Up Compute & Scale Out Read Only Nodes

Azure SQL Data Warehouse (SQLDW)

D-SQL

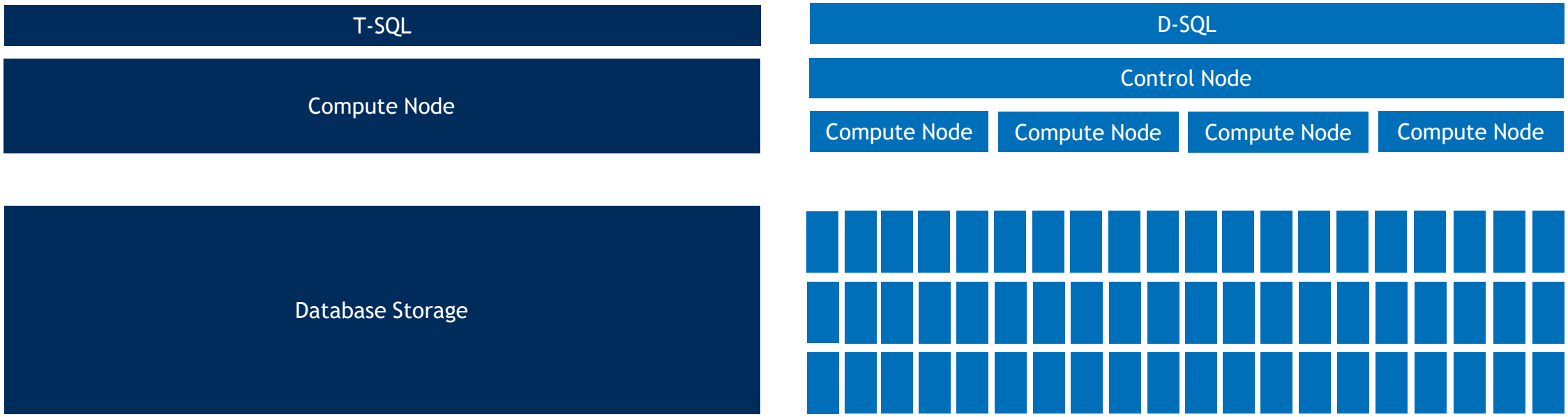
Control Node

Compute Node | Compute Node | Compute Node | Compute Node



Scale Out Compute & Distributed Storage

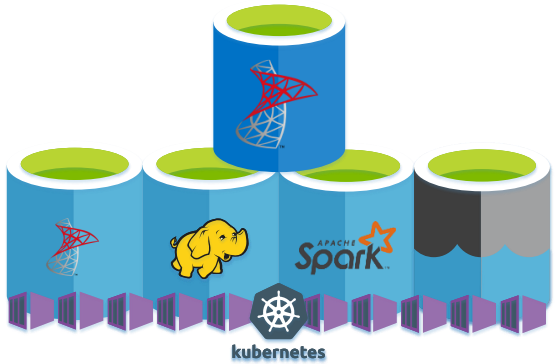
SQL Server - Product Evolution



SQL Server - Product Evolution

SQL Server 2019 Big Data Cluster

Azure Data Studio



SSMS v18

T-SQL

Python/Scala

Control

Master Instance

Compute

Compute Node

Compute Node

Compute Node

Compute Node

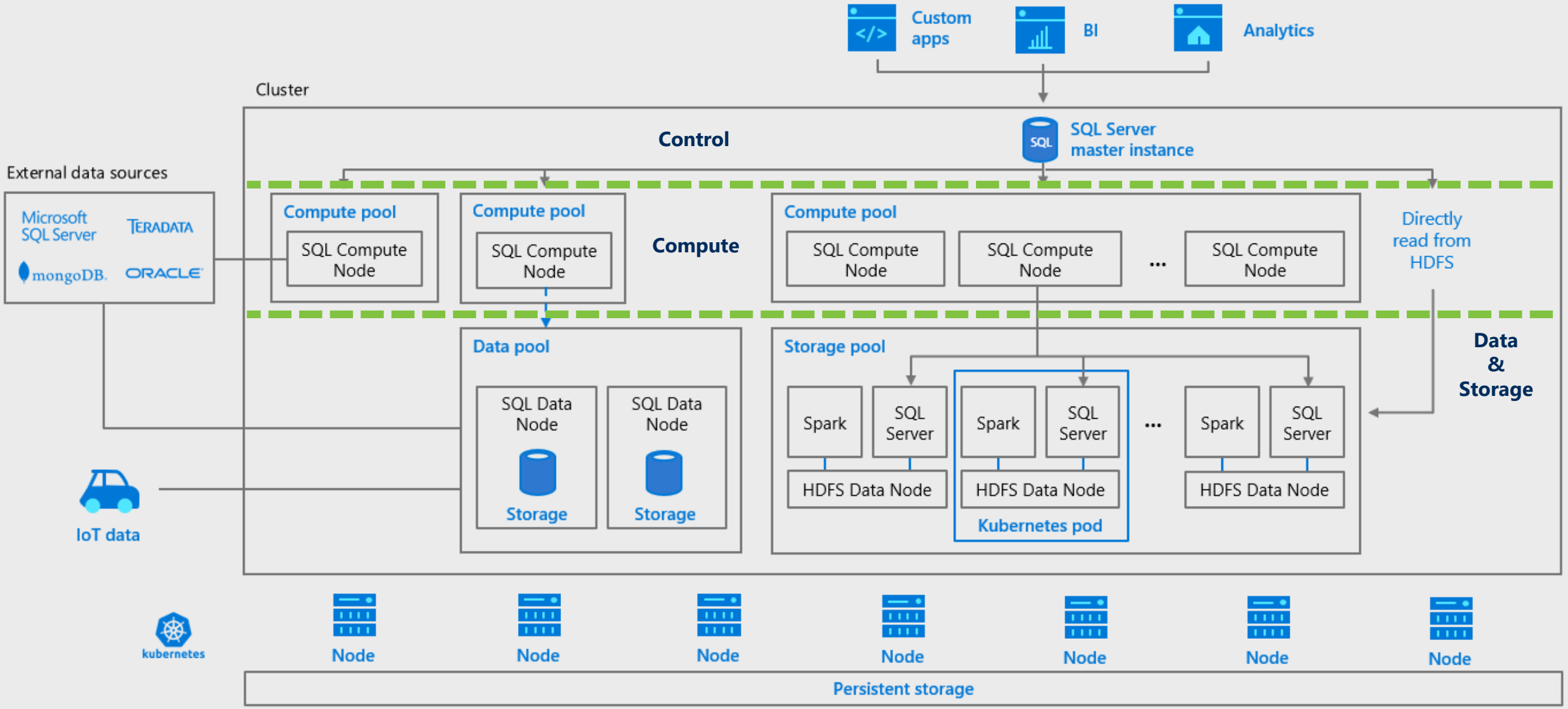
Data & Storage

Data Pool
(Storage Shards)

Storage Pool
(SQL Engine, HDFS, Spark)

Linux Containers

in Kubernetes

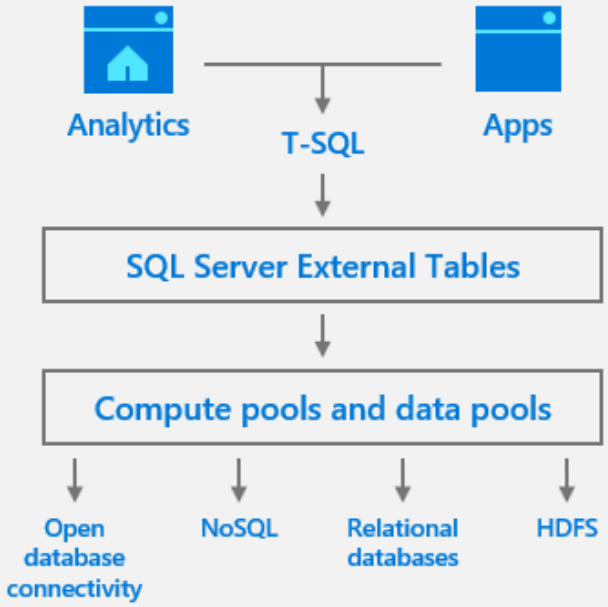


SQL Server 2019 Big Data Cluster

https://cloudblogs.microsoft.com/sqlserver/2018/09/25/introducing-microsoft-sql-server-2019-big-data-clusters/

adatis

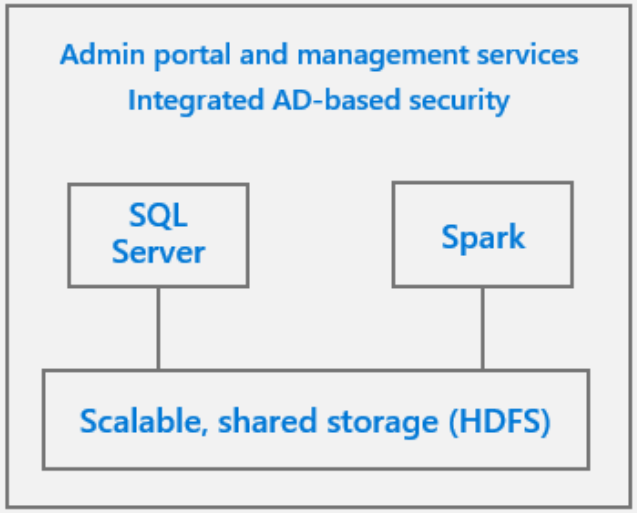
Data virtualization



Combine data from many sources without moving or replicating it

Scale out compute and caching to boost performance

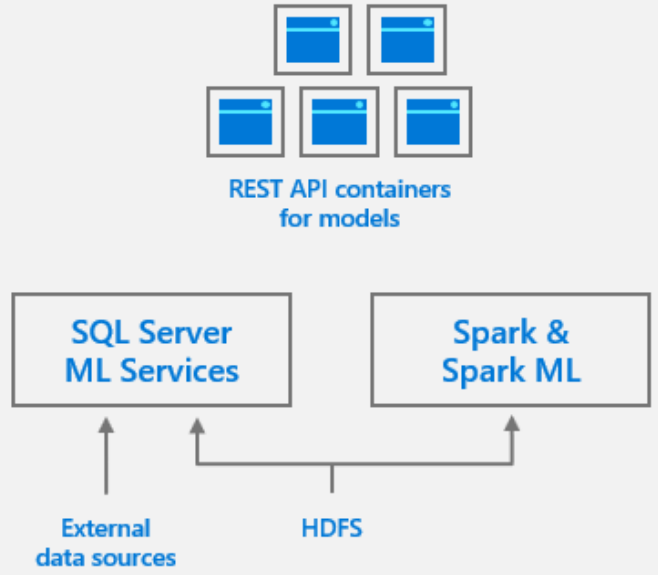
Managed SQL Server, Spark, and data lake



Store high volume data in a data lake and access it easily using either SQL or Spark

Management services, admin portal, and integrated security make it all easy to manage

Complete AI platform



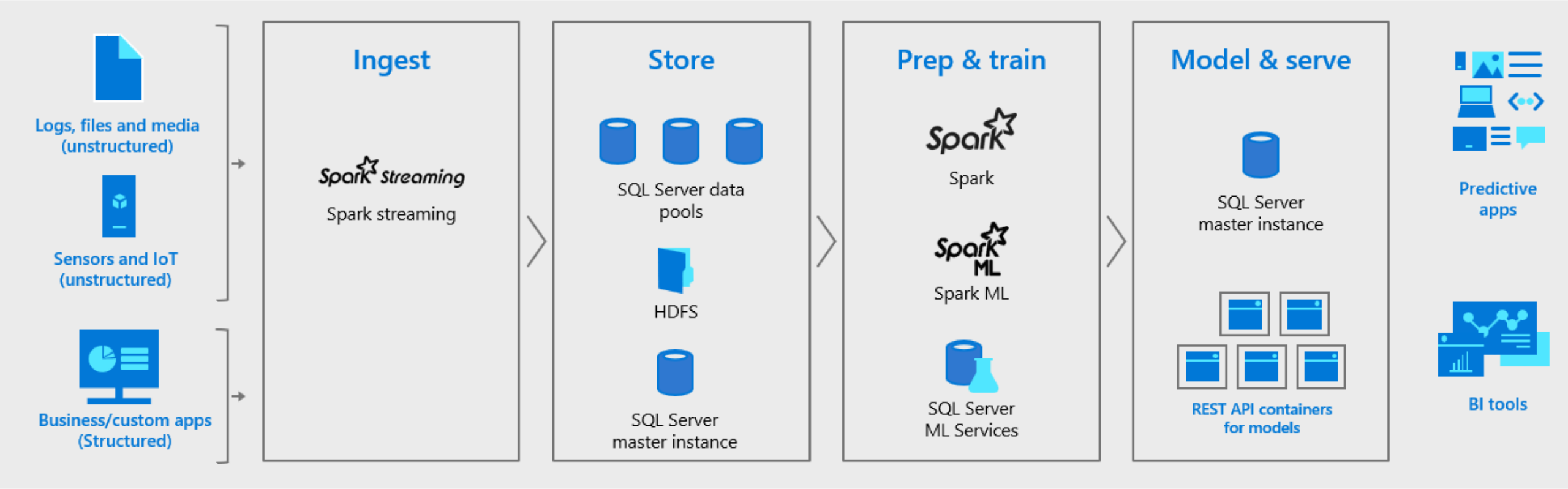
Easily feed integrated data from many sources to your model training

Ingest and prep data and then train, store, and operationalize your models all in one system

SQL Server 2019 Big Data Cluster

https://cloudblogs.microsoft.com/sqlserver/2018/09/25/introducing-microsoft-sql-server-2019-big-data-clusters/

adatis

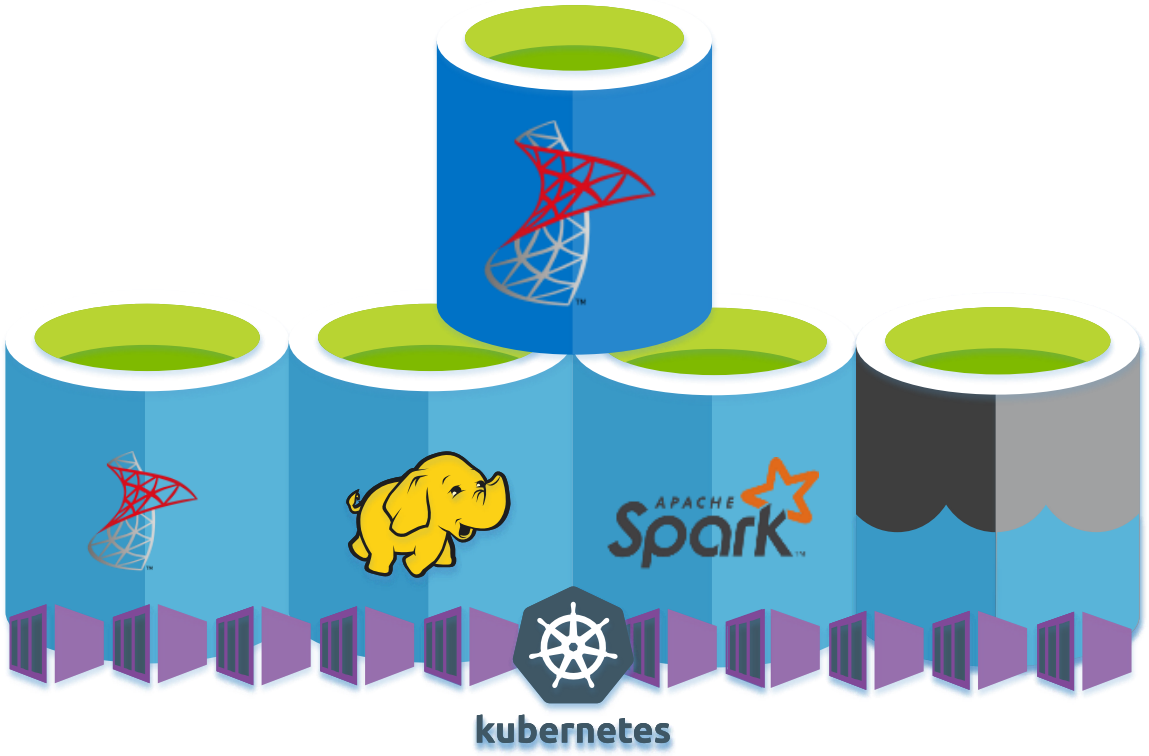


SQL Server 2019 Big Data Cluster

https://cloudblogs.microsoft.com/sqlserver/2018/09/25/introducing-microsoft-sql-server-2019-big-data-clusters/

adatis

What



Why

How

What

Why

How

adatis

What

Why

How... does a SQL Server Big Data Cluster address our big data problems?

do we implement this new product architecture?

adatis

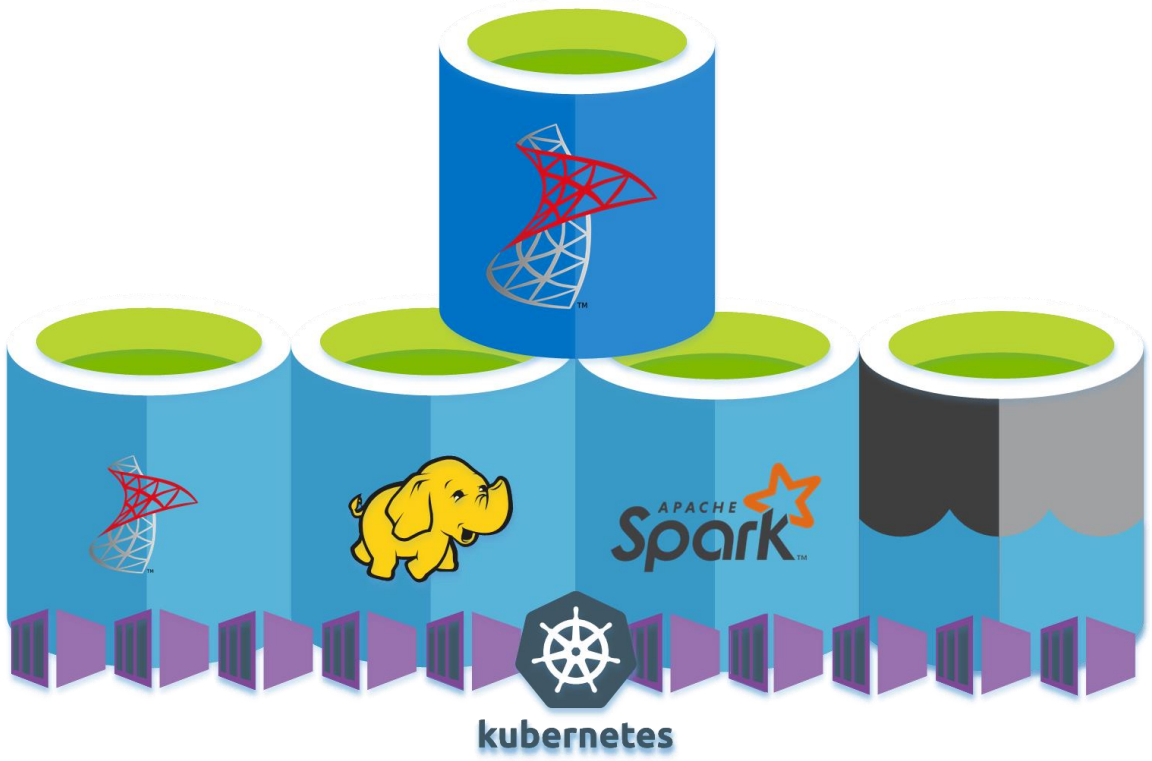
What

Why

How... does a SQL Server Big Data Cluster address our big data problems?

do we implement this new product architecture?

adatis


APACHE
Spark
kubernetes

Volume

Distributed storage pools as standard.

Velocity

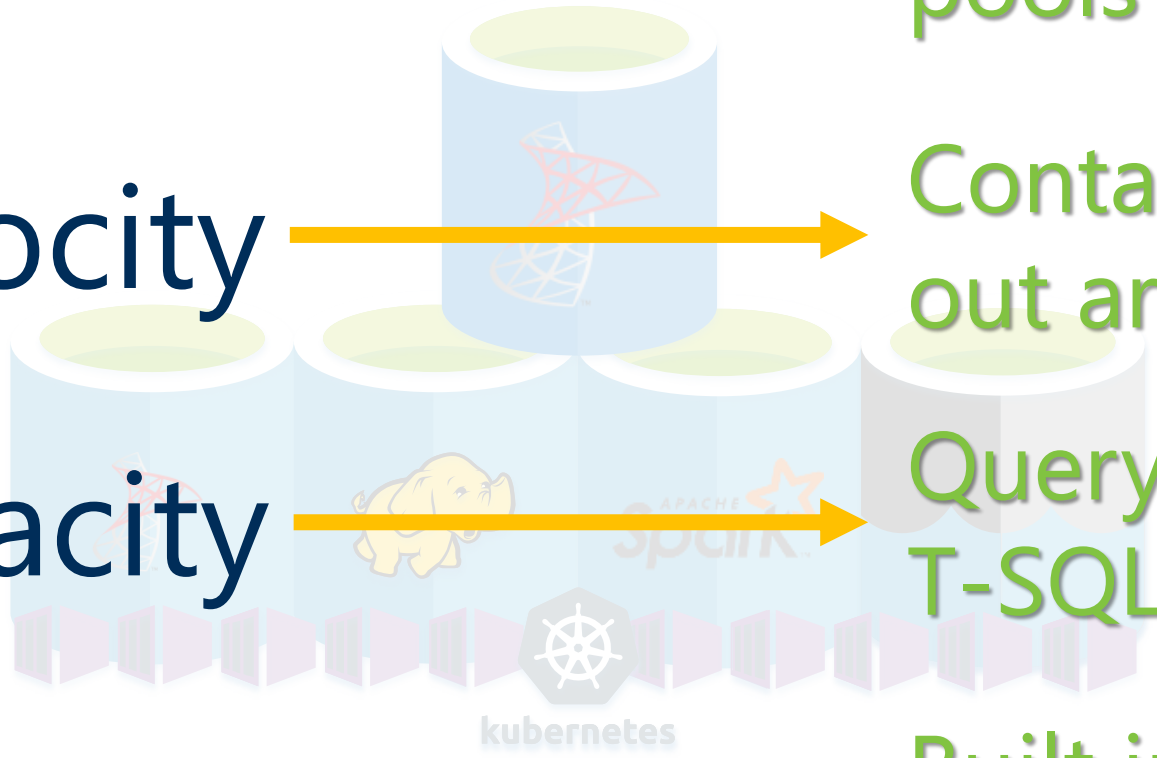
Containerised scale out architecture.

Veracity

Query everything in T-SQL via PolyBase.

Variety

Built in data lake capabilities via HDFS.



adatis

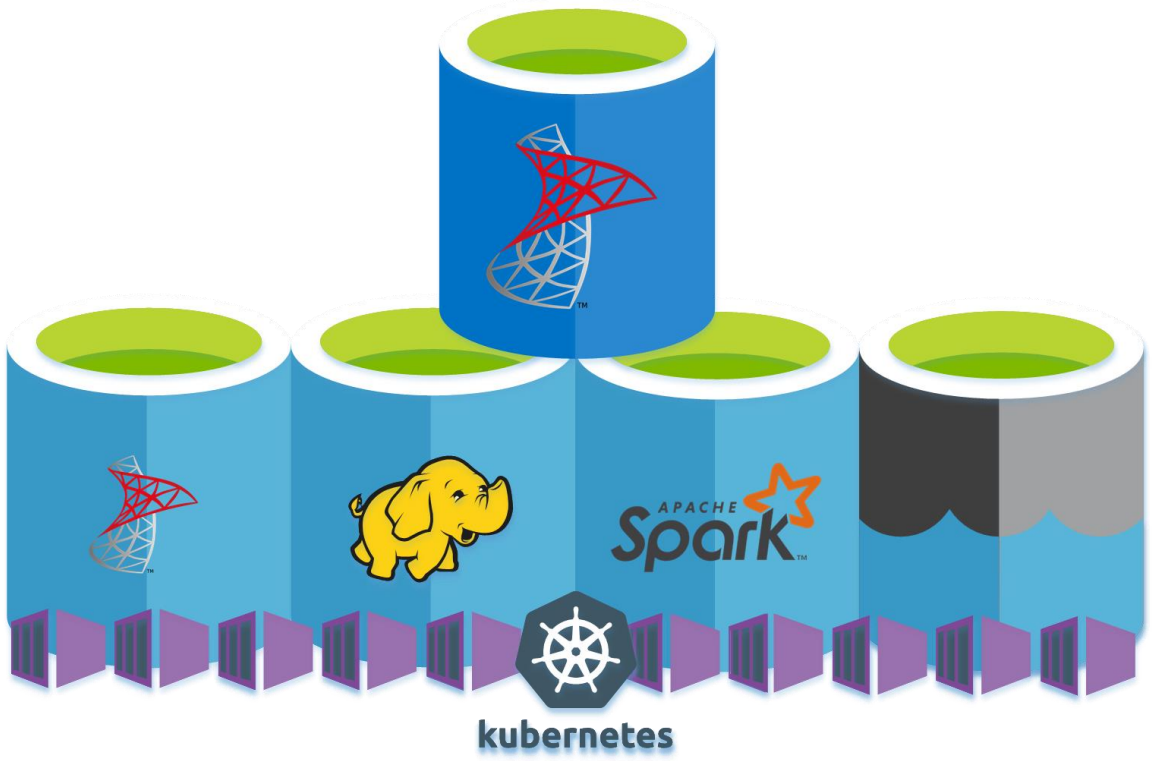
What

Why

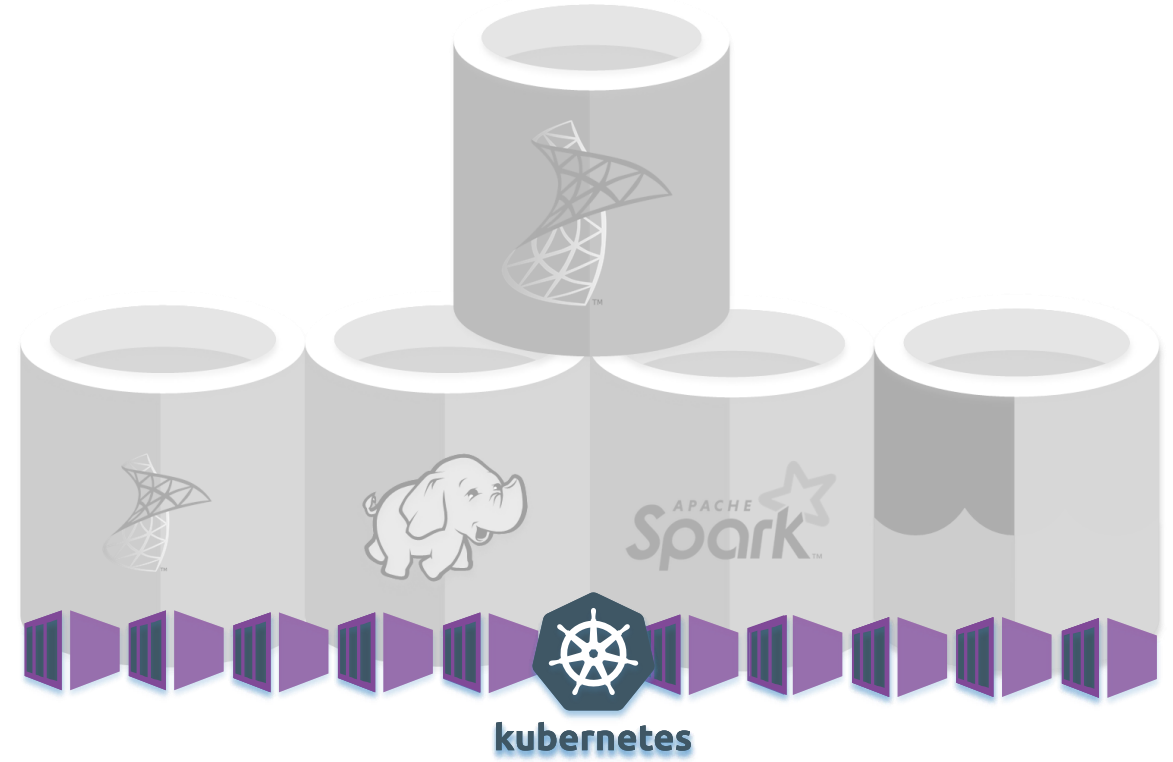
How... does a SQL Server Big Data Cluster address our big data problems?

do we implement this new product architecture?

adatis


APACHE
Spark
kubernetes

kubernetes minikube

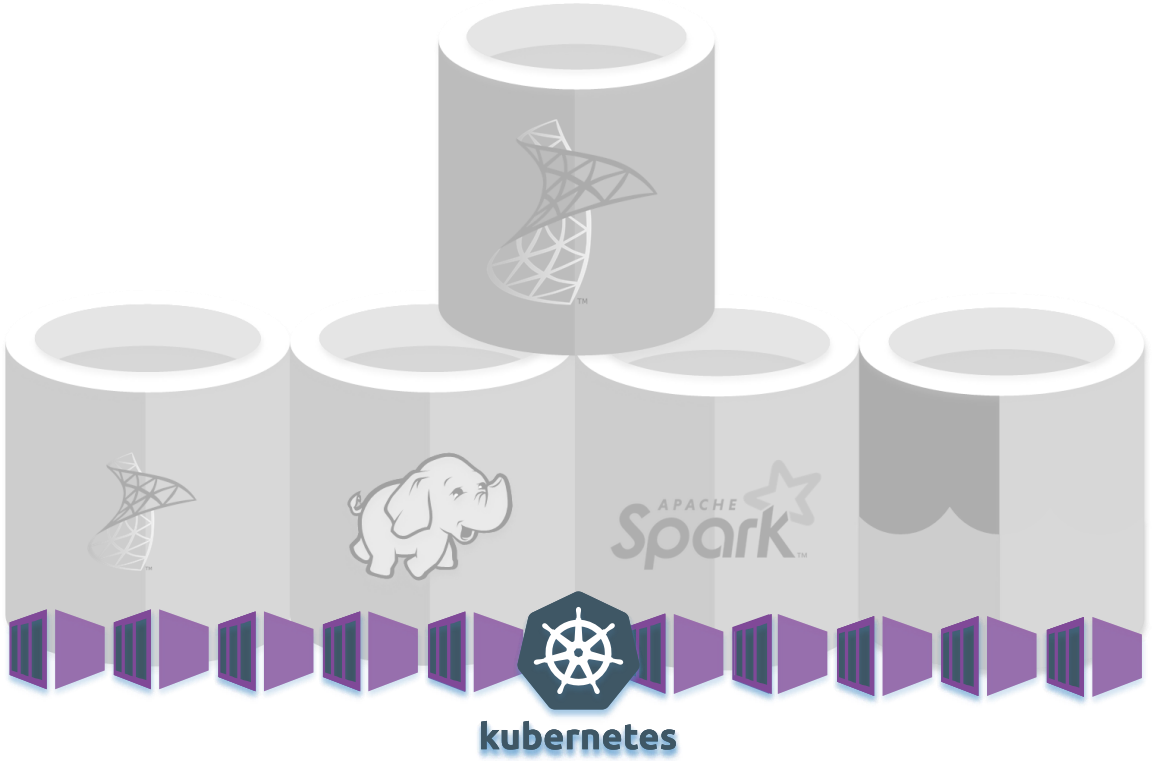


Windows ubuntu Mac OS

SQL Server 2019 Big Data Clusters

kubernetes

minikube



Azure

amazon web services™

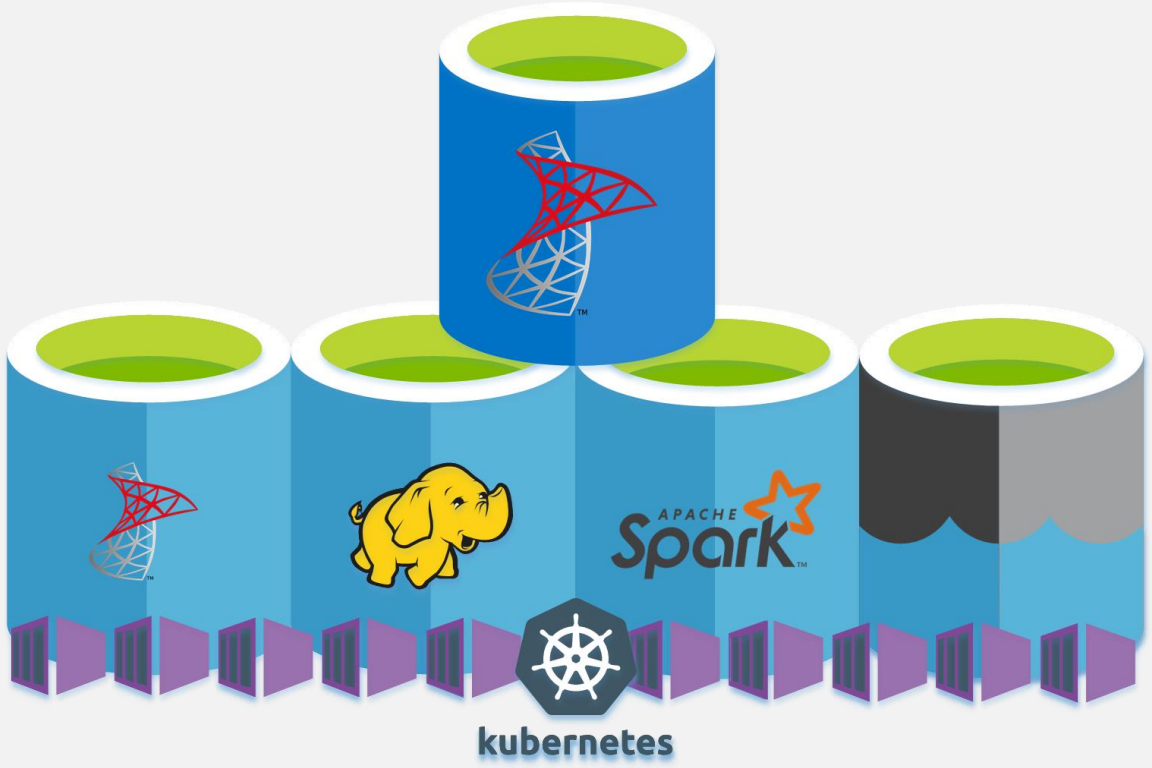
Google Cloud Platform

adatis

SQL Server 2019 Big Data Clusters

Cloud tech, using cloud concepts, now available on premises.

Not only cloud.



adatis

Reference Material

What's new in SQL Server 2019?
https://docs.microsoft.com/en-us/sql/sql-server/what-s-new-in-sql-server-ver15?view=sql-server-ver15

SQL Server 2019 White Paper
https://info.microsoft.com/ww-landing-SQLDB-Microsoft-SQL-Server-WhitePaper.html

Introducing SQL Server 2019 Big Data Clusters – Travis Wright
https://cloudblogs.microsoft.com/sqlserver/2018/09/25/introducing-microsoft-sql-server-2019-big-data-clusters/

MS Ignite Session – The future of querying big data with PolyBase and SQL Server
https://myignite.techcommunity.microsoft.com/sessions/66201

MS Ignite Session – Deep dive on SQL Server and big data
https://myignite.techcommunity.microsoft.com/sessions/65967

Buck Woody SQLBits Precon
https://aka.ms/sqlworkshops

How to deploy SQL Server big data cluster on Kubernetes
https://docs.microsoft.com/en-us/sql/big-data-cluster/deployment-guidance?view=sql-server-ver15

Deploy SQL Server big data cluster on Azure Kubernetes Service
https://docs.microsoft.com/en-us/sql/big-data-cluster/quickstart-big-data-cluster-deploy?view=sql-server-ver15

adatis

Thanks for Listening

Paul Andrew

@MrPaulAndrew

Microsoft® Most Valuable Professional

adatis

Email: paul@mrpaulandrew.com
Blog: mrpaulandrew.com
GitHub: github.com/mrpaulandrew