



# SQL Server 2019 Roadmap + SQL Server 2019 Big Data Clusters

Andy Roberts  
[andyrob@microsoft.com](mailto:andyrob@microsoft.com)

# Modernize SQL Server

## Customer scenarios

- Mission critical applications with lowest TCO
- Real-time operational analytics
- Run SQL Server anywhere (Windows, Linux, containers)

## Investments

- Friction-free upgrades of SQL Server
- Compliance and security
- Performance and scale improvements

## Customer solutions

Apps

BI

ML

## Partner solutions

ISV apps

Hosted

### Analytics

T-SQL

BI

ML

### Data engine



Optimized query processing  
HTAP  
Security across platforms  
High availability

Intelligent QP, Automatic query tuning, QDS  
In memory OLTP + Column Store  
Always Encrypted, Dynamic Data Masking, RLS, TDE  
Failover Cluster Instance And Availability Groups


### Storage options


 Row


 Column


 In memory


## Platform choice




  
Linux/Windows container





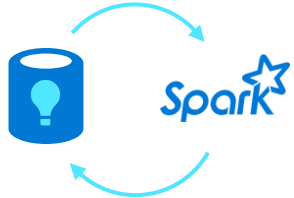


  
Linux

# SQL Server 2019




Industry-leading performance and security, with intelligence over all your data

### Intelligence over any data



AI and Machine Learning over all data with the power of SQL and Apache Spark

### Choice of platform and language




T-SQL  
Java  
C/C++

PHP  
Node.js  
C#/VB.NET

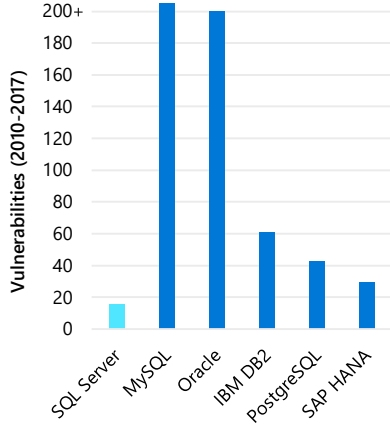
Python  
Ruby

### Industry-leading performance




#1 OLTP performance<sup>1</sup>  
#1 DW performance on 1TB<sup>2</sup>, 10TB<sup>3</sup>, and 30TB<sup>4</sup>  
Intelligent Query Processing

### Most secure over the last 8 years<sup>5</sup>

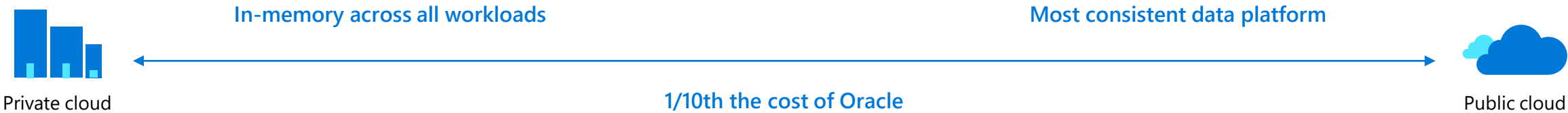


Database	Vulnerabilities (2010-2017)
SQL Server	15
MySQL	205
Oracle	200
IBM DB2	60
PostgreSQL	45
SAP HANA	30

### Insights in minutes and rich reports



The best of Power BI and SQL Server Reporting Services with Power BI Report Server



All TPC Claims as of 1/19/2018.  
<sup>1</sup> <http://www.tpc.org/4081>; <sup>2</sup> <http://www.tpc.org/3331>; <sup>3</sup> <http://www.tpc.org/3326>; <sup>4</sup> <http://www.tpc.org/3321>; <sup>5</sup> National Institute of Standards and Technology Comprehensive Vulnerability Database

# SQL Server 2019 enables intelligence over all your data



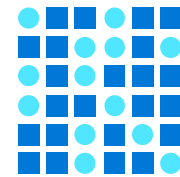
## Integrating all data

Unified access to all your data with unparalleled performance



## Managing all data

Easily and securely manage data big and small



## Analyzing all data

Build intelligent apps and AI with all your data



Simplified management and analysis through a unified deployment, governance, and tooling

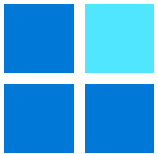
# All on a unified data services platform



Connect to all of your data  
**Including Relational, noSQL, Hadoop**



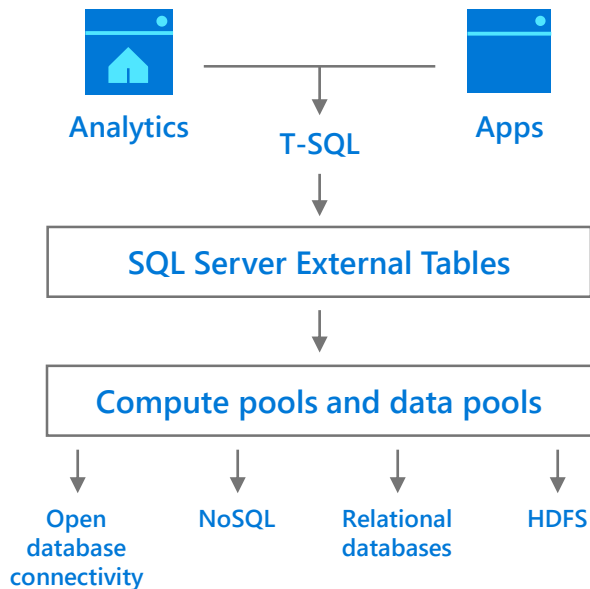
Create intelligence from all your data  
**Using Spark and SQL**



Manage this through a single pane of glass  
**With Azure Data Studio**

# SQL Server 2019 big data, analytics, and AI

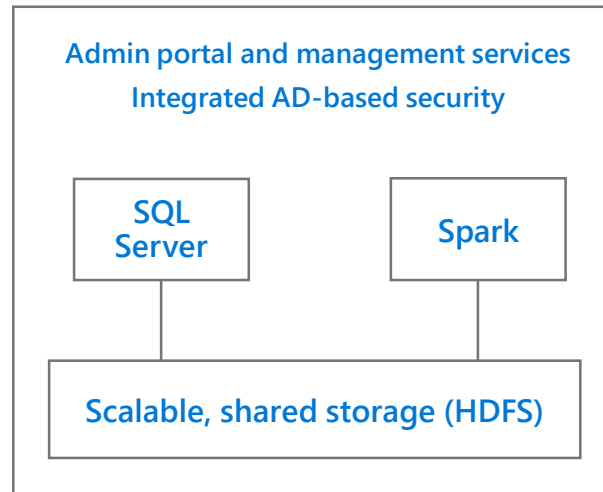
## Data virtualization



Combine data from many sources without moving or replicating it

Scale out compute and caching to boost performance

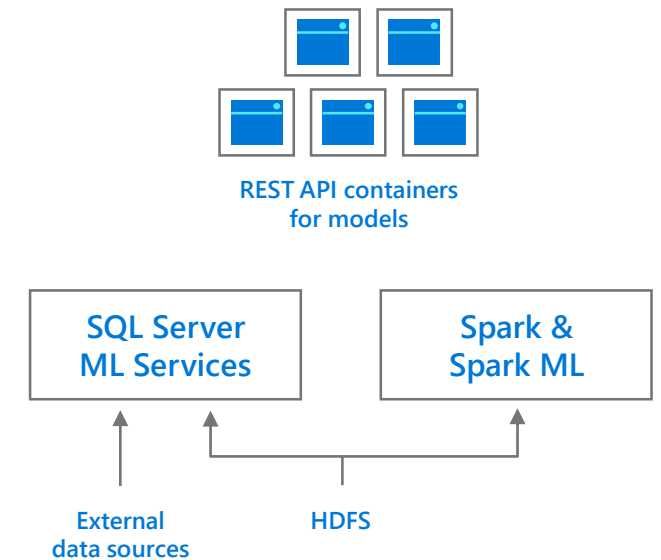
## Managed SQL Server, Spark, and data lake



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

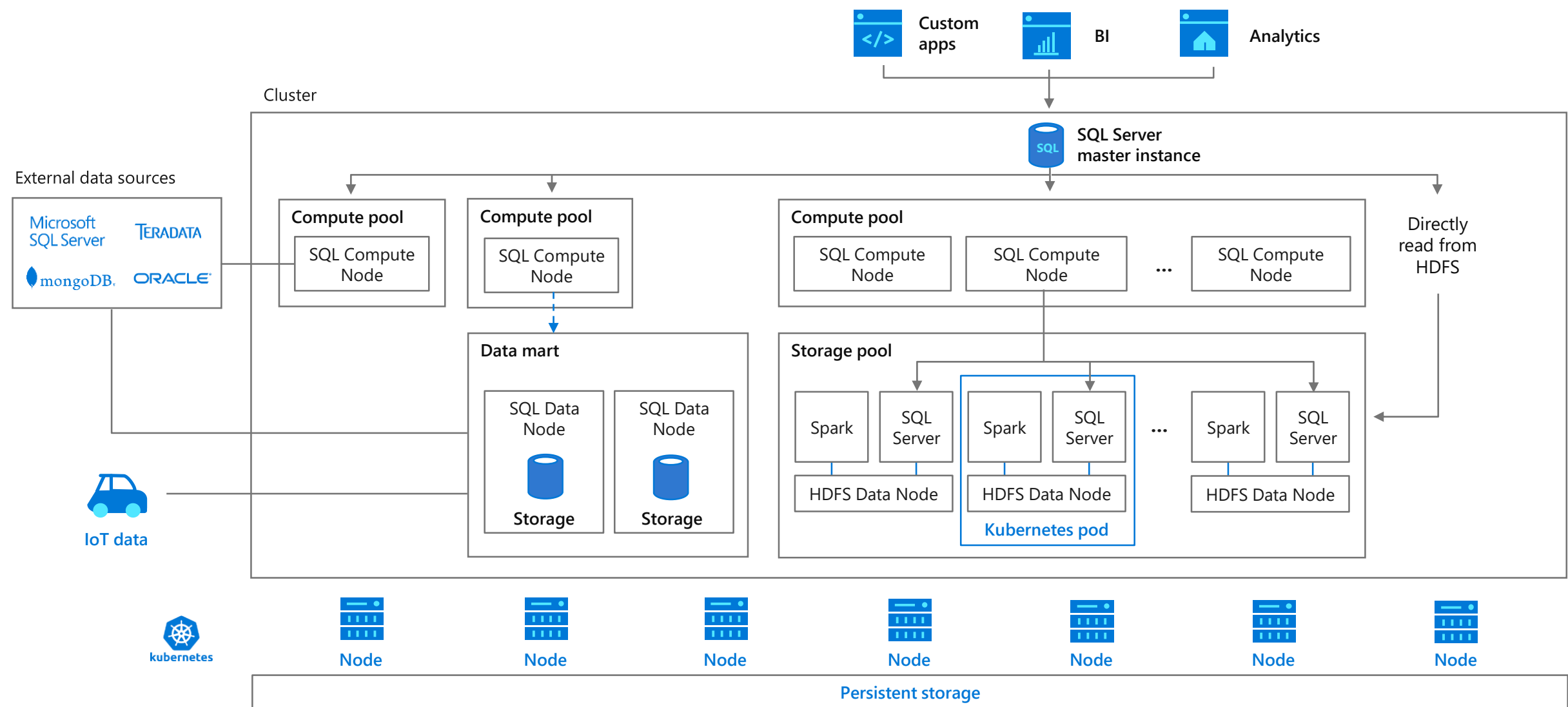
## Complete AI platform



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 big data clusters



# The heart of SQL Server is mission critical performance, security, and availability



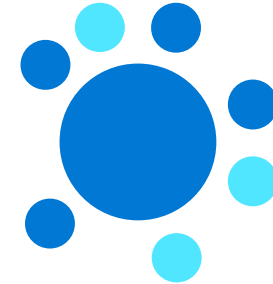
## Performance

Breakthrough performance  
and scalability



## Security

Data protected at rest and  
in motion



## Availability

High availability for business  
critical workloads

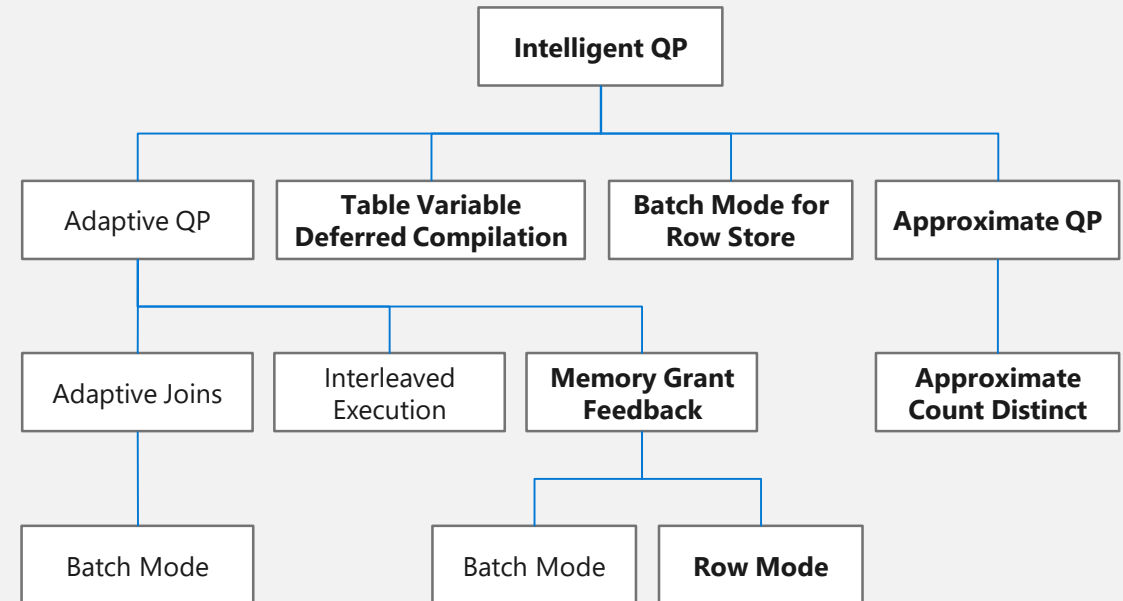


# Mission critical performance

## The intelligent database

- Intelligent Query Processing
- Accelerating I/O performance with Persistent Memory
- Gain performance insights anytime and anywhere with Lightweight Query Profiling

## The Intelligent Query Processing feature family



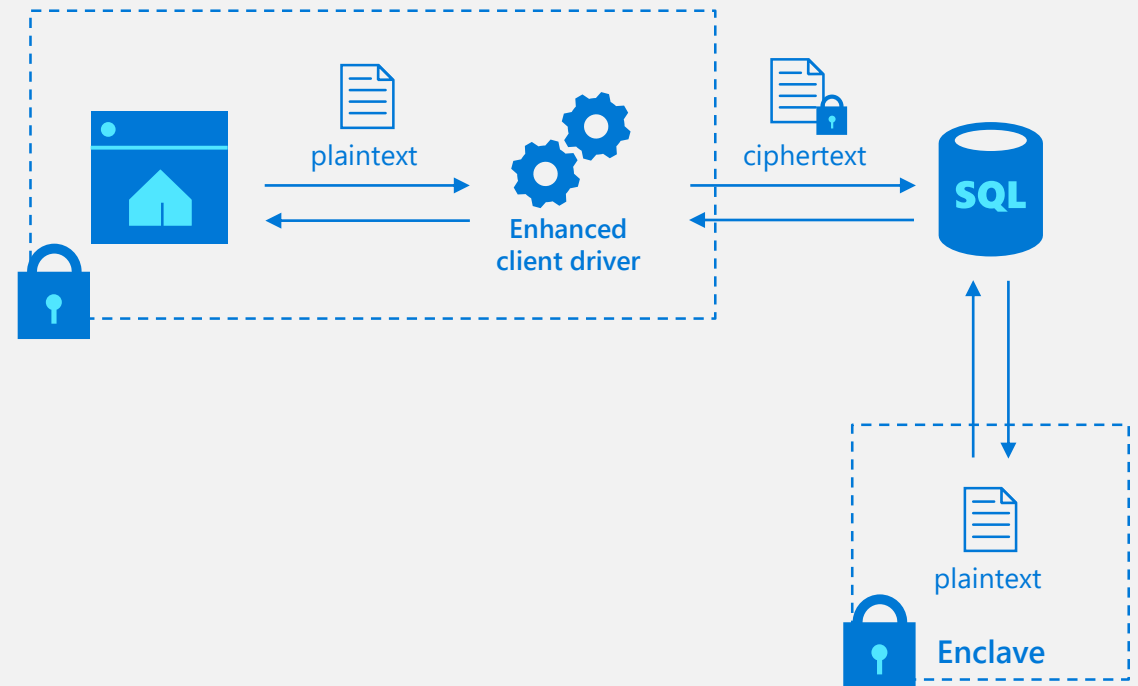
Bold indicates new and improved features in SQL Server 2019

# Mission critical security

## Confidential computing

- Always Encrypted with secure enclaves
- Data Classification and auditing built-in
- Manage certificates easier with SQL Configuration Manager

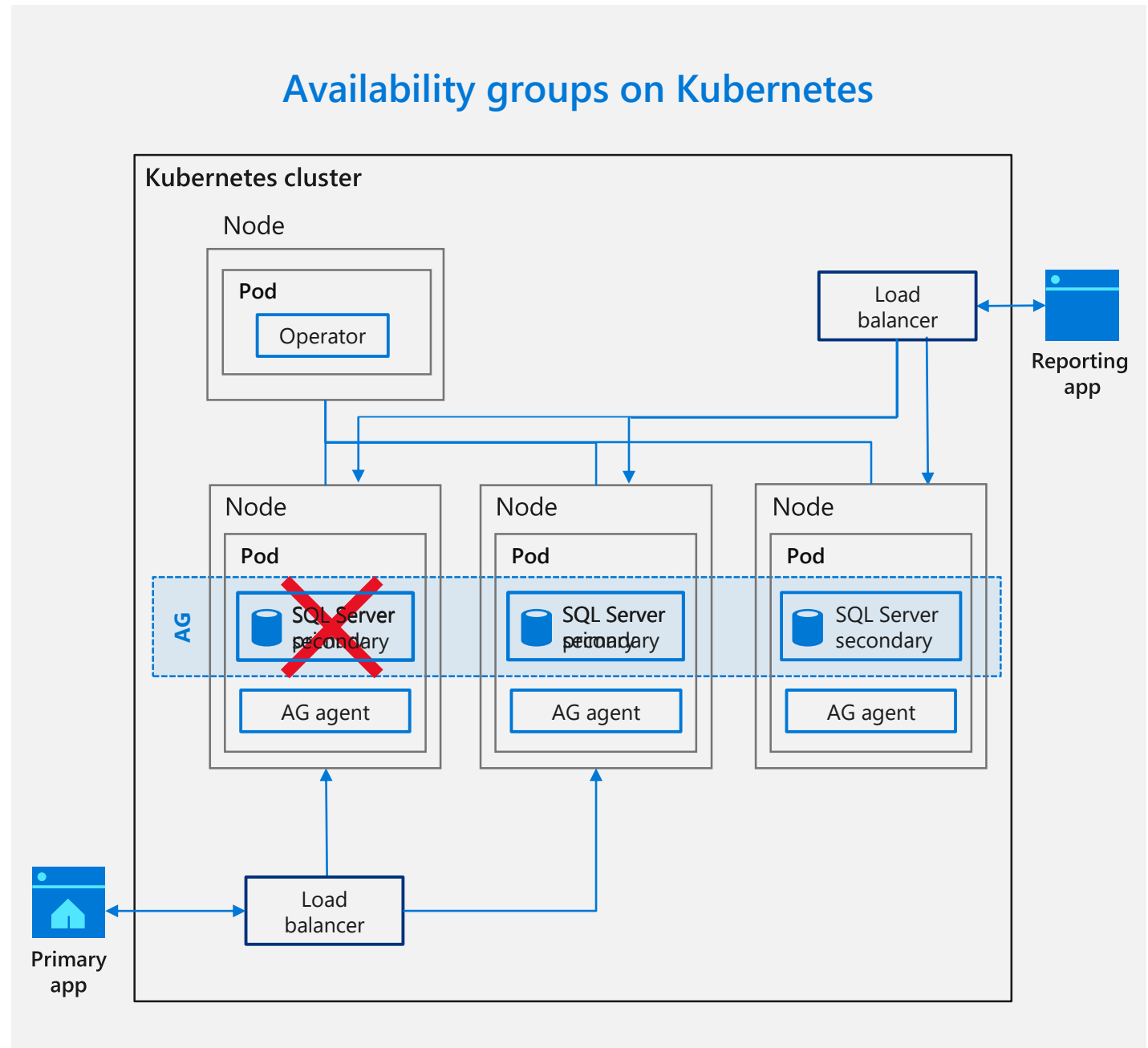
Always Encrypted with secure enclaves



# Mission critical availability

## Keep SQL Server running

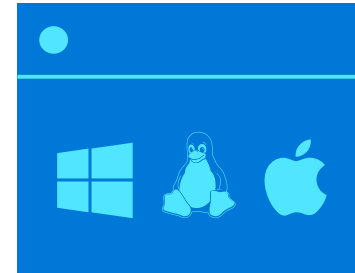
- Always On availability group enhancements
- Resumable online index creation
- Online Clustered Columnstore index creation and rebuild
- Availability groups on Kubernetes



# The modern development platform



Speed app development and  
admin with new enhancements



Develop on your choice  
of language and tooling

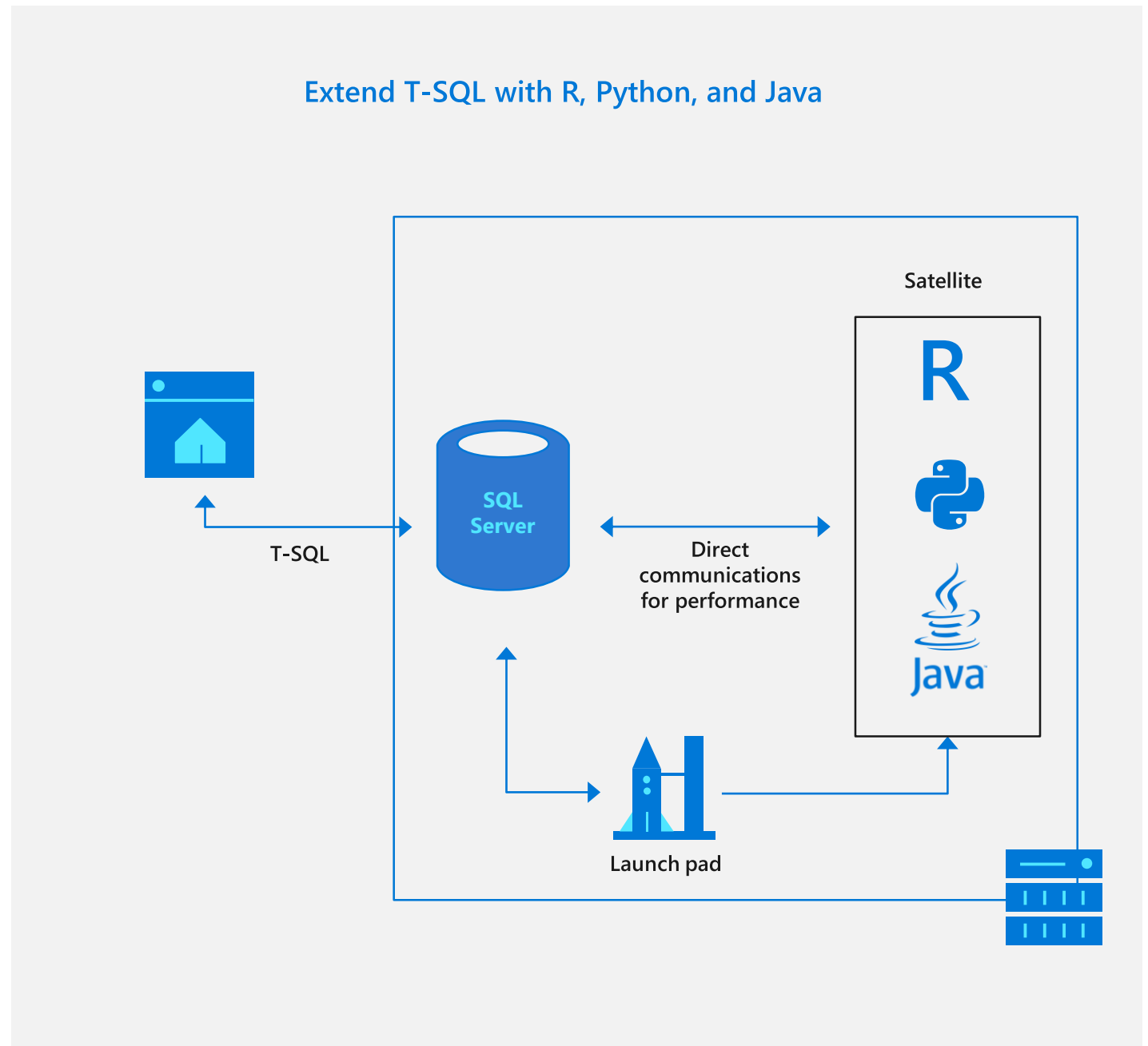
# Enhancing the developer experience

SQL Graph enhancements

UTF-8 support

Machine Learning Services enhancements

SQL Server Java extension



# Enhancing the platform of choice

Closing features gaps for SQL Server on Linux

- Replication
- Distributed transactions
- Machine Learning

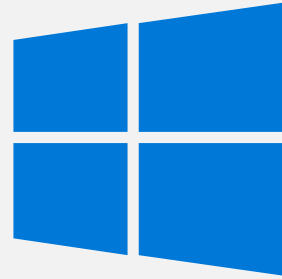
Open LDAP Provider support

The Microsoft Container Registry

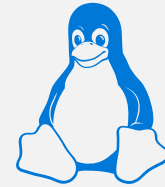
SQL Server RedHat Container Images

Always On Availability Groups on Kubernetes

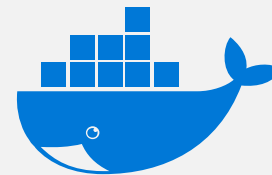
Windows



Linux

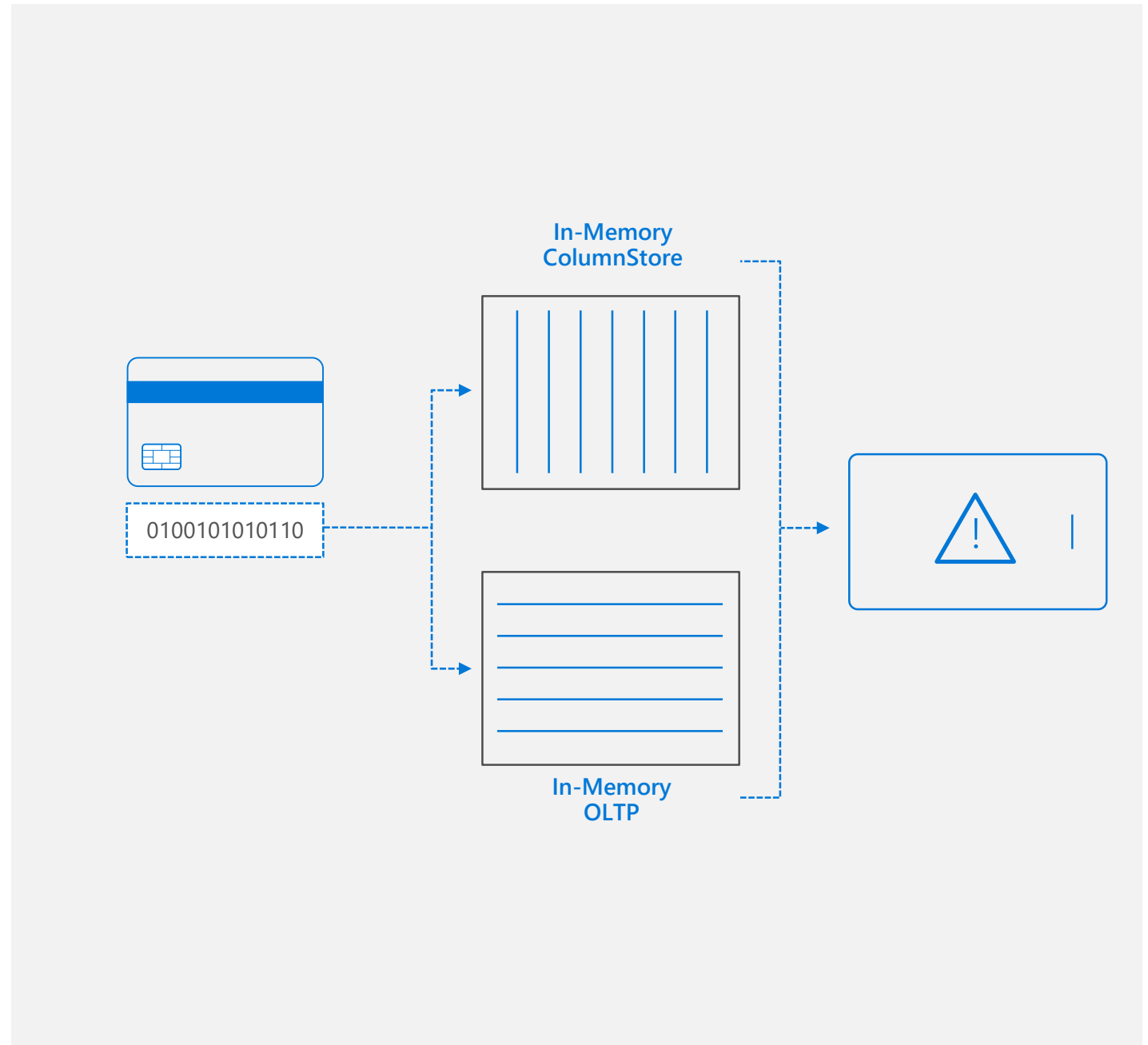


Docker containers and Kubernetes



# Tuning the SQL Server engine

- Columnstore stats in DBCC CLONEDATABASE
- Estimate compression for Columnstore indexes
- Troubleshoot page resource waits with new built-in T-SQL



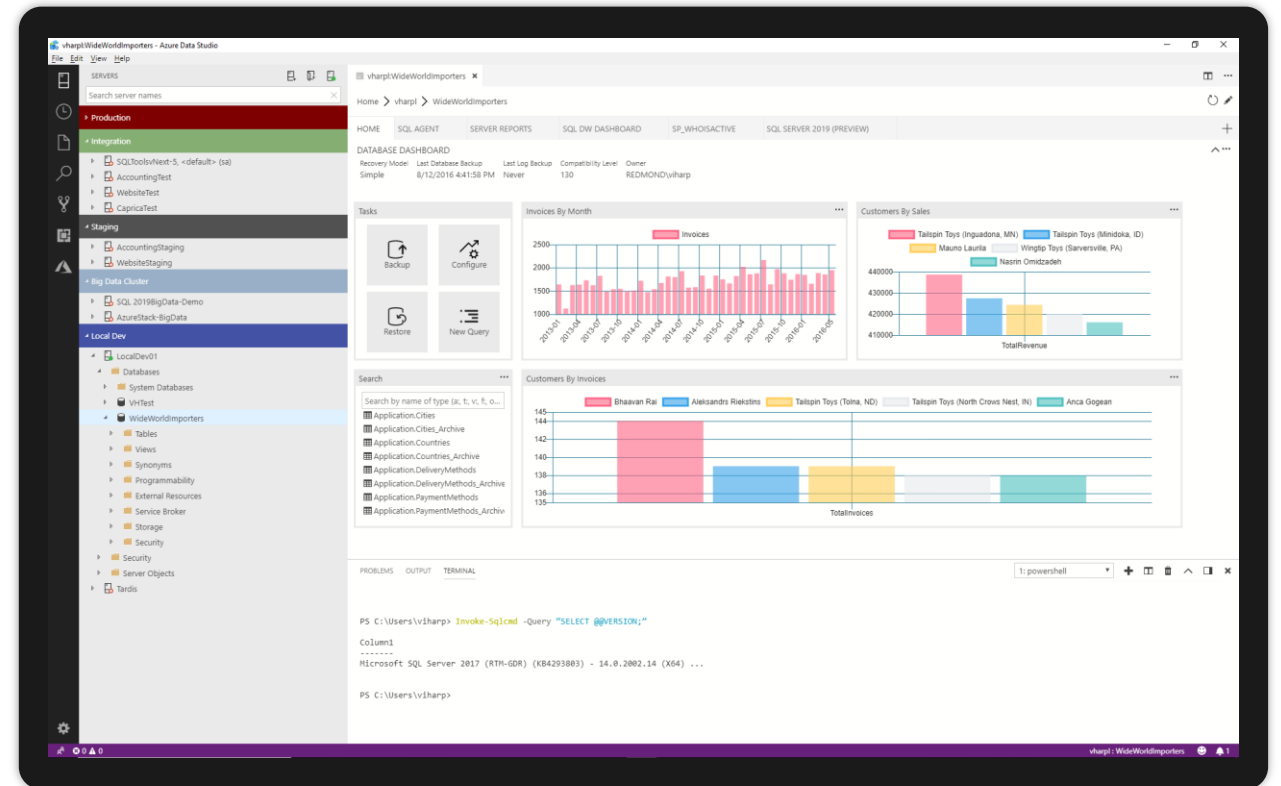
# The Azure Data Studio tools experience

Azure Data Studio is a lightweight, open source, cross-platform graphical management tool and code editor

Enable a modern DevOps experience for database developers and DBAs on their platform of choice

Simplify development, configuration, management, monitoring and troubleshooting for SQL databases on-premises and in the cloud <sup>NEW</sup>

Use SQL Server Management Studio 18.0 Preview to access, configure, manage, and administer all SQL Server components

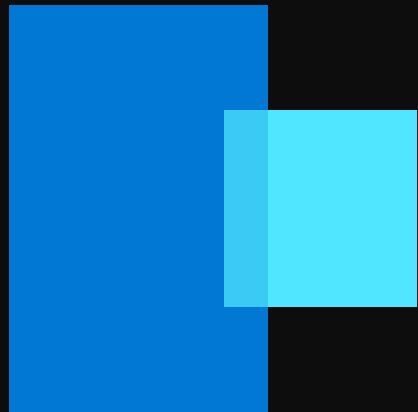




# Investments in the future of SQL Server

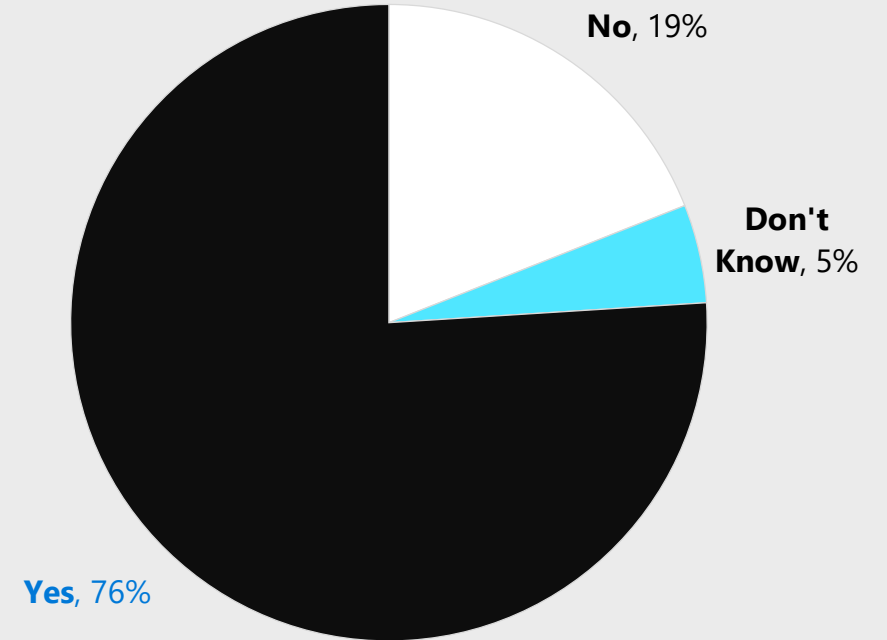
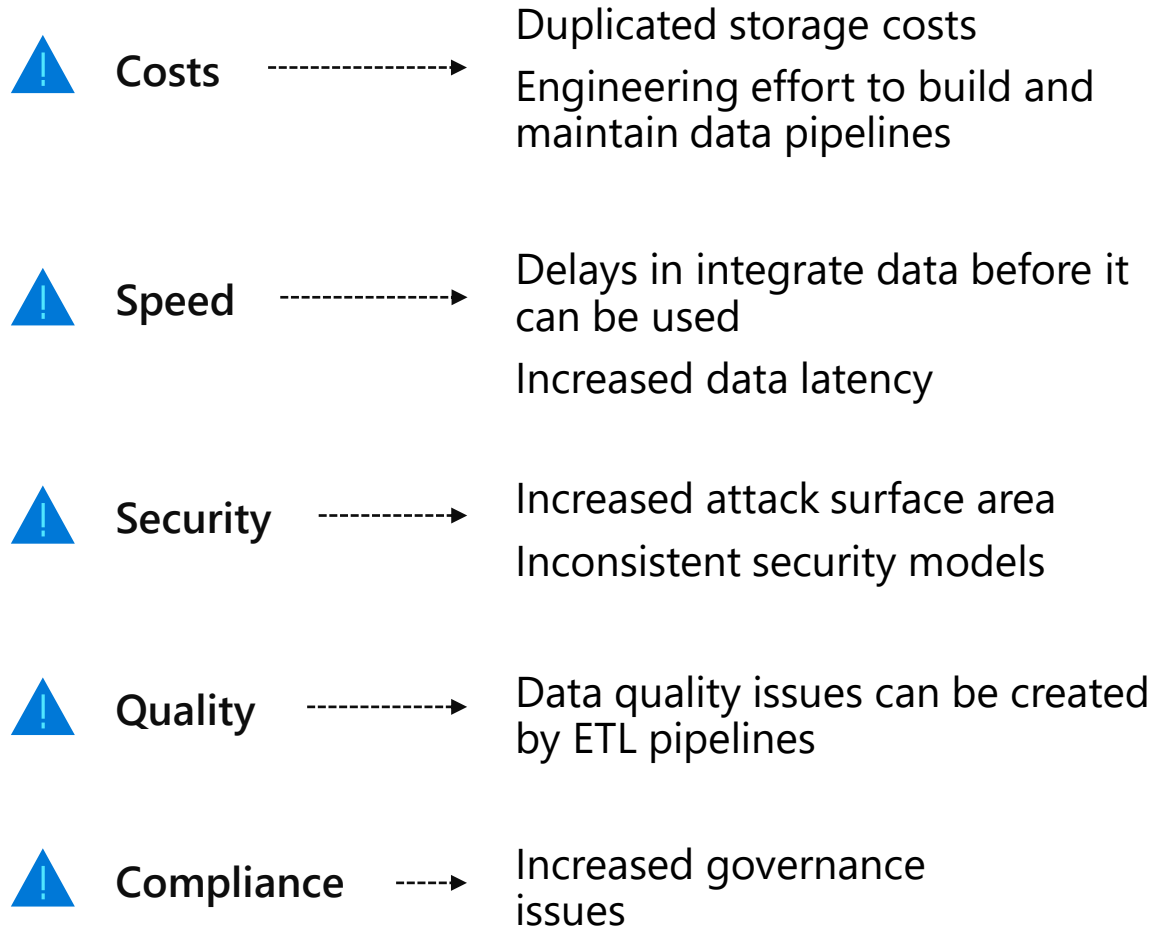
- SQL Server on Edge
- Finish features for big data clusters and data virtualization
- Making SQL Server more available, faster
- Further enhance SQL Server security
- Enhance the engine to align with hardware innovation
- Continue to make the container experience great
- Engine improvements based on customer feedback





Integrating  
all data

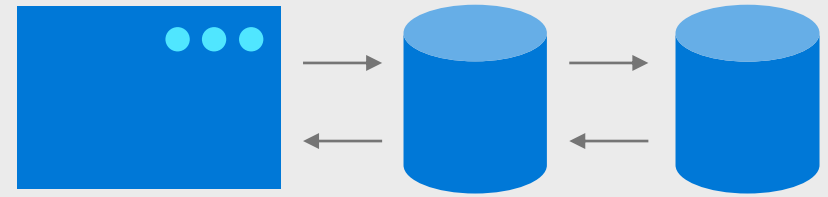
# Data movement is a barrier to faster insights



**3/4** of respondents say that  
**untimely data has inhibited business opportunities**

# Data virtualization creates solutions

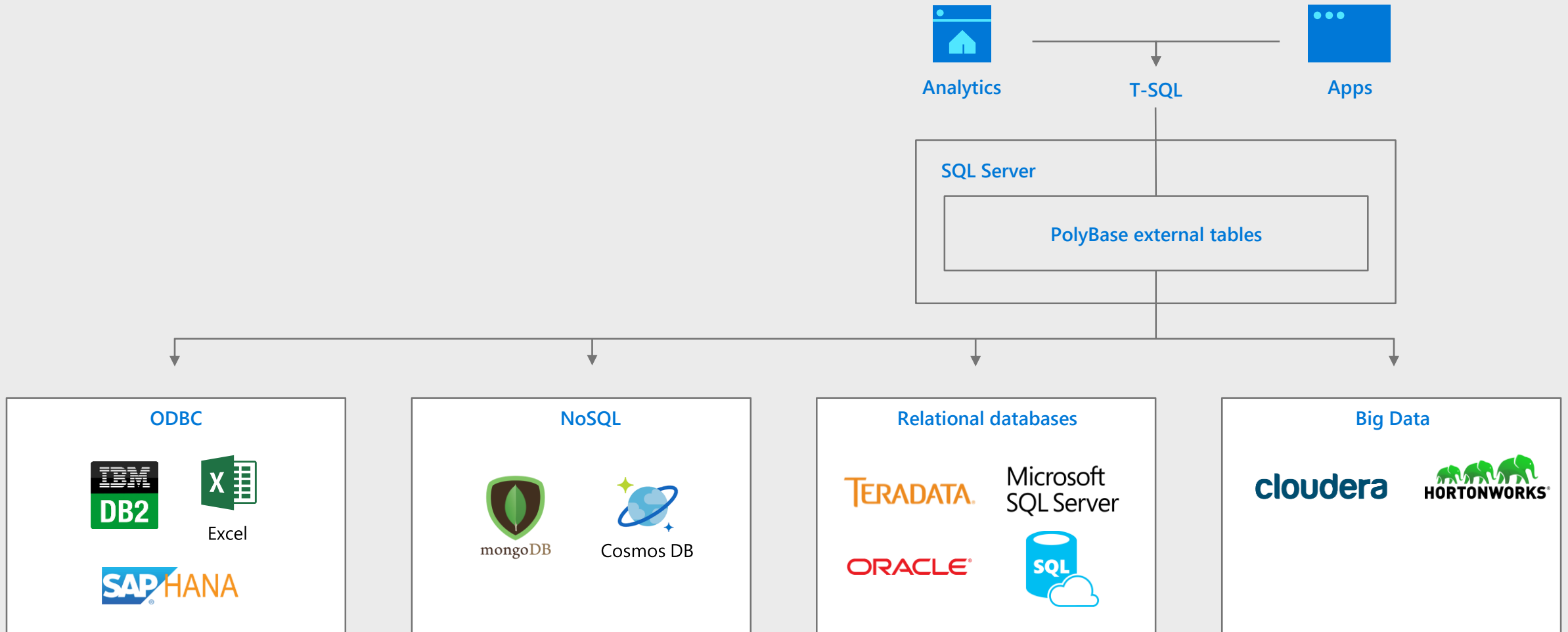
- ✓ **Costs** -----> Lower storage costs  
Less dev time spent on integration
- ✓ **Speed** -----> Rapid iterations and prototypes  
Timely data
- ✓ **Security** -----> Smaller attack surface area  
Consistent security model
- ✓ **Quality** -----> Fresh and accurate data
- ✓ **Compliance** -----> Easier data governance



Data virtualization integrates data from disparate sources, locations and formats, **without replicating or moving the data**, to create a single "virtual" data fabric

# SQL Server is the hub for integrating data

Easily combine across relational and non-relational data stores



# Data virtualization technology comparison

## Linked Servers

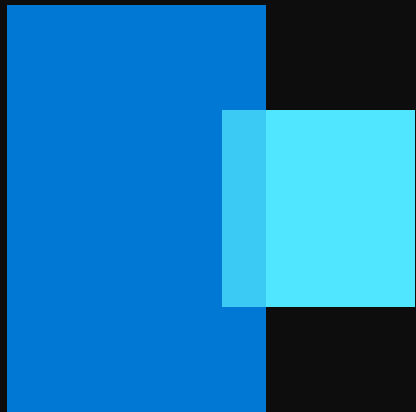
- Instance scoped object
- Uses OLEDB providers
- Supports both read/write & pass-through statements
- Queries are single-threaded & push-down supported
- Separate configuration needed for each instance in Always On Availability Group

## PolyBase External tables

- Database scoped object
- Uses ODBC drivers
- Supports read-only operations only. Will be expanded in future
- Queries can be scaled-out & push-down supported
- No separate configuration needed for Always On Availability Group

# Demo #1

Use SQL Server 2019 PolyBase to query external data  
sources: Oracle



Data virtualization easily combines data from many sources and eliminates data silos





Managing  
all data

# Big Data leads to big problems



**Complex scale-out deployment**



**Time-consuming patching and upgrades**

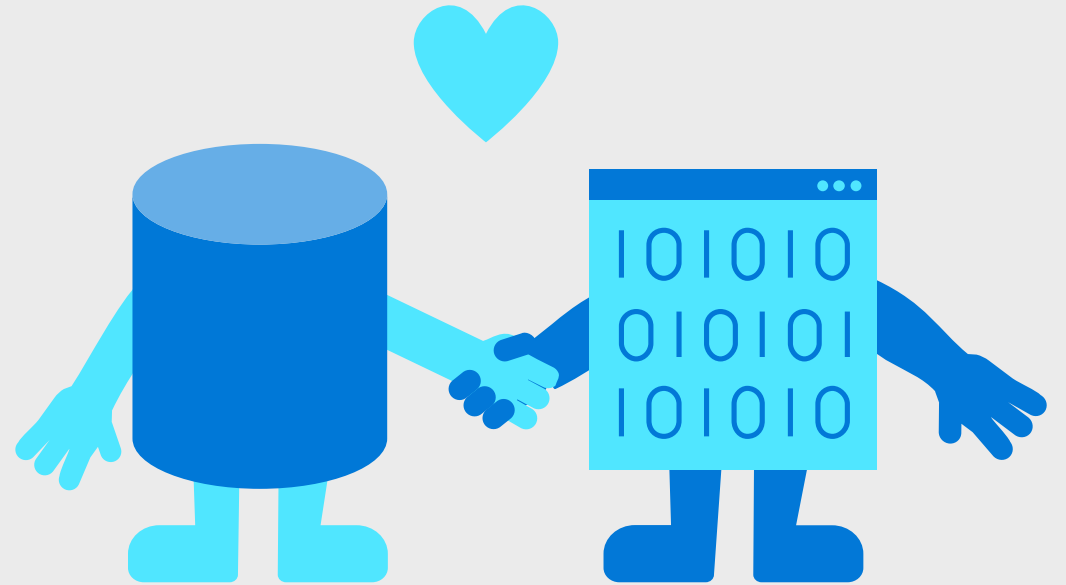


**Cumbersome security management**

# Easily deploy and manage a SQL Server + Big Data cluster

Easily deploy and manage a Big Data cluster using Microsoft's Kubernetes-based Big Data solution **built-in to SQL Server**

Hadoop Distributed File System (HDFS) storage, SQL Server relational engine, and Spark analytics are deployed as containers on Kubernetes in one easy-to-manage package



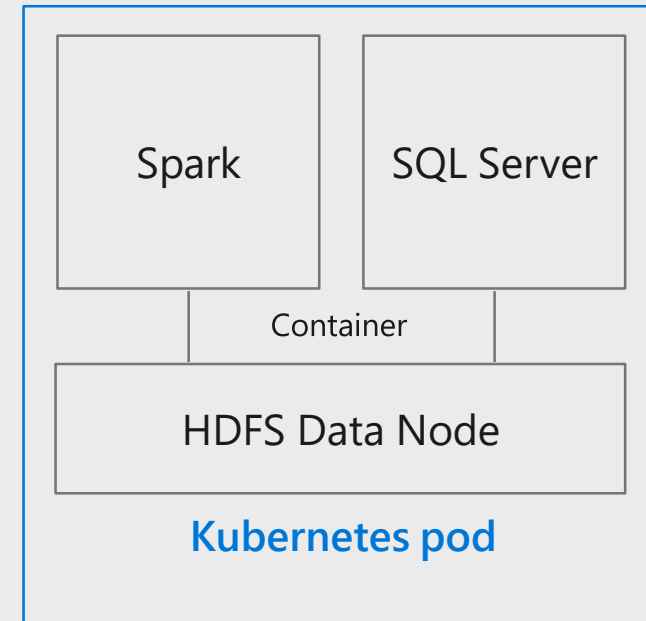
# Simplified deployment with containers & Kubernetes

A container is a standardized unit of software that includes everything needed to run it

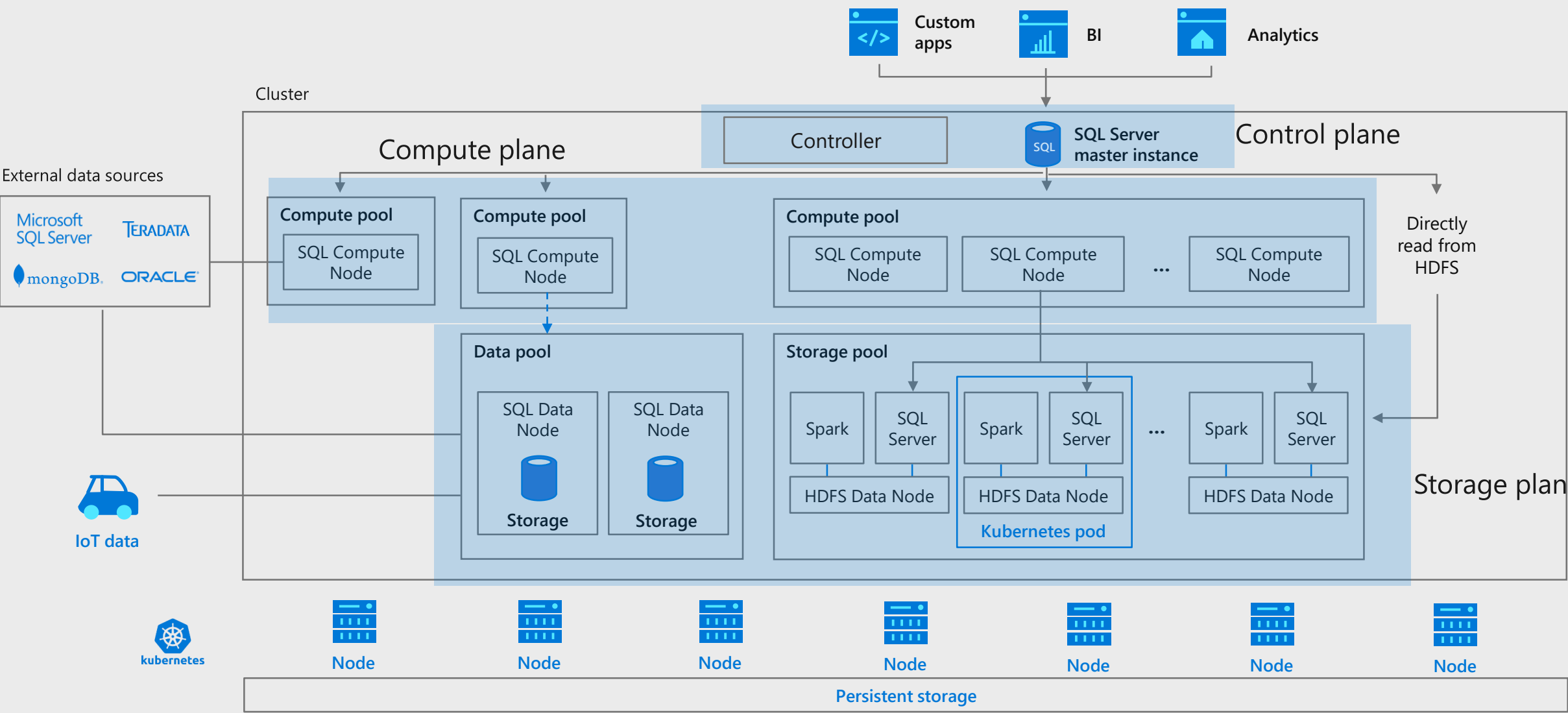
Kubernetes is a container hosting platform

Benefits of containers and Kubernetes:

1. Fast to deploy
2. Self-contained – no installation required
3. Upgrades are easy because - just upload a new image
4. Scalable, multi-tenant, designed for elasticity



# SQL Server Big Data Cluster Layout

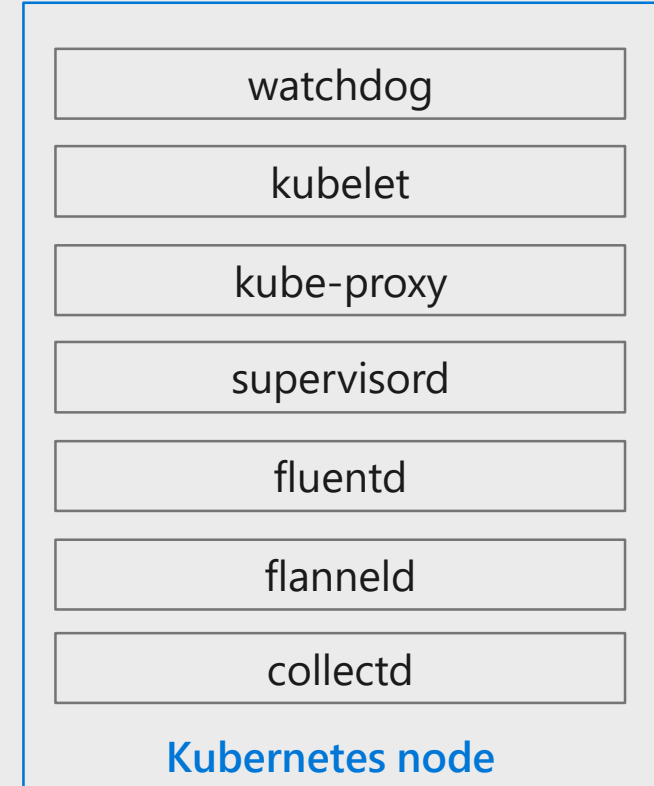


# Base node configuration

Applies to all Kubernetes nodes across all planes

## Services

- kubelet – Kubernetes agent
- kube-proxy – network config and forwarding
- supervisord – process monitor and control
- fluentd – node logging
- flanneld – Software defined network
- collectd – OS and application data collection
- SQL Server big data cluster watchdog– config sync, watchdog, data collector (DMV, etc)



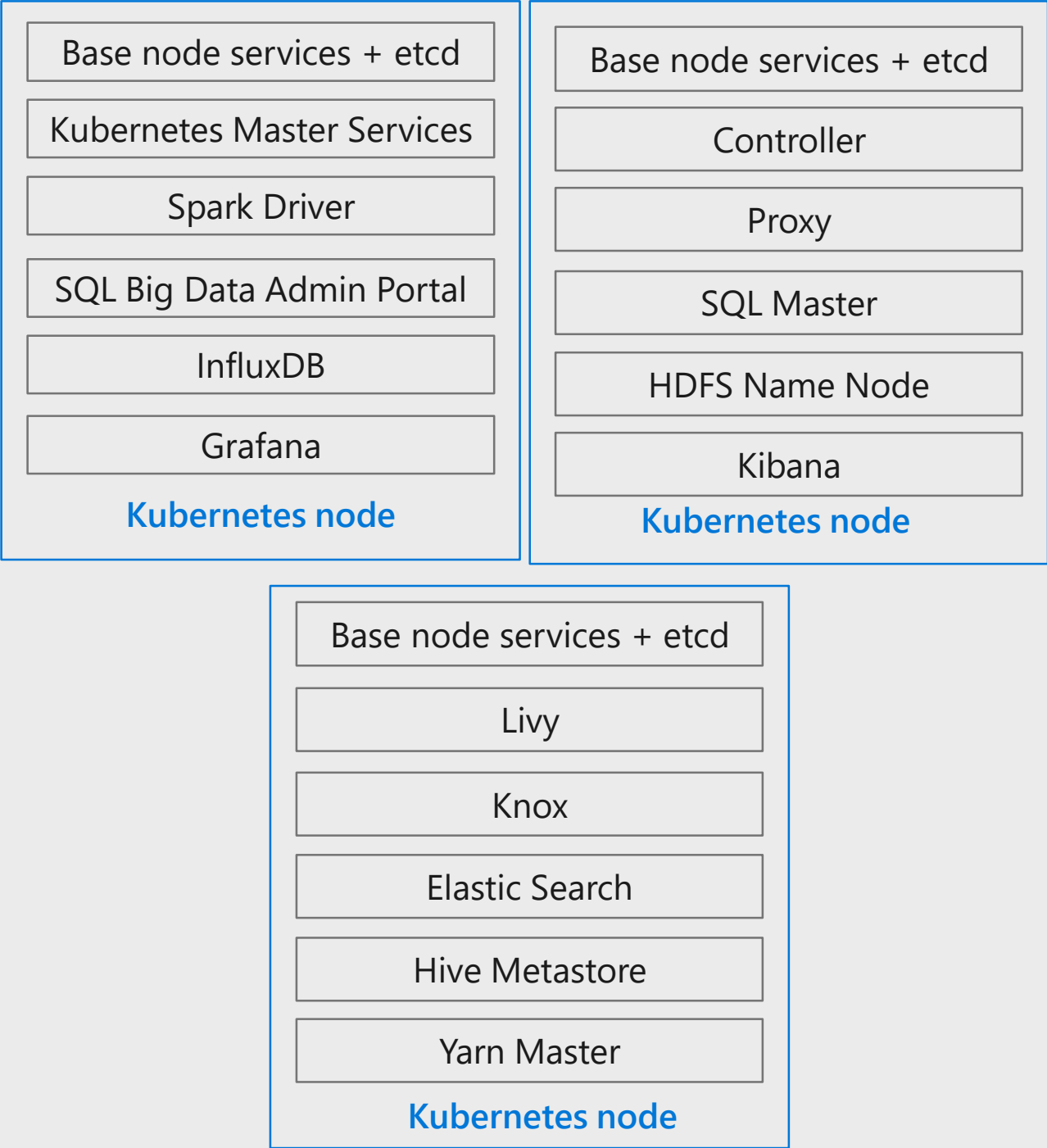
# Control plane

## External Endpoints

- Kubernetes (HTTP)
- Admin portal proxy service (HTTP/REST)
- Controller Service (HTTP/REST)
- Knox Gateway (HTTP/gateway for Hadoop/Spark/Livy)
- SQL Server Master instance (TDS)

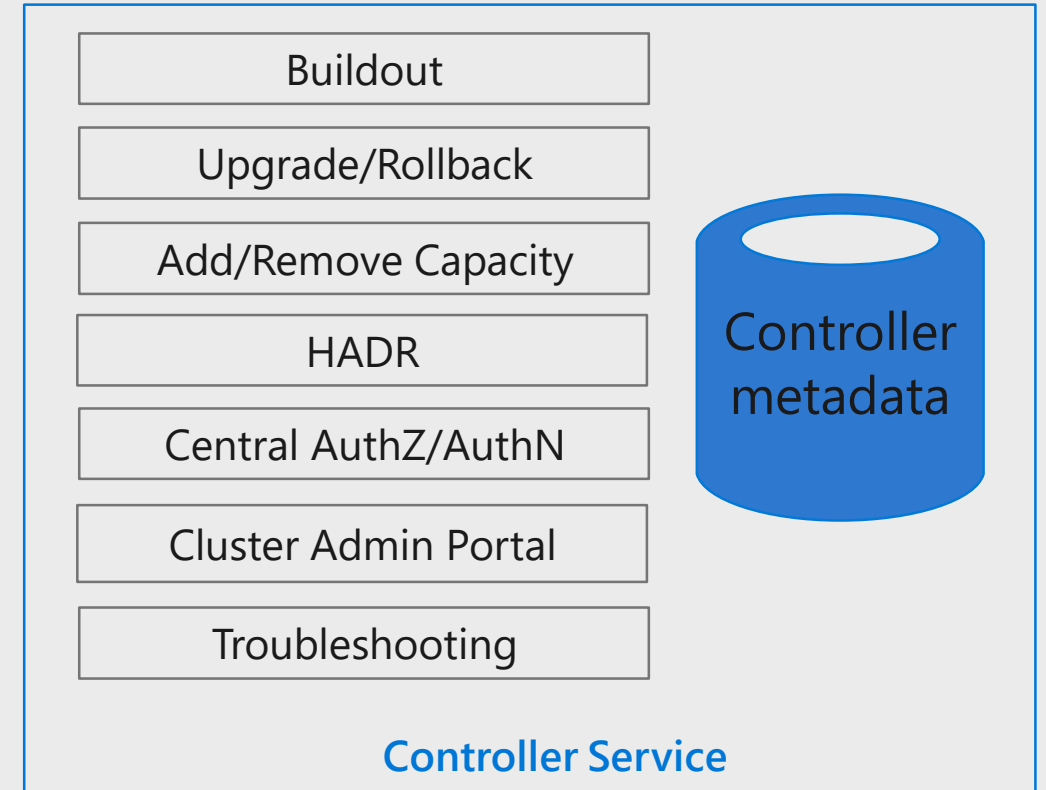
## Services

- etcd
- Kubernetes master services
- Controller service
- SQL Server master instance
- SQL big data cluster admin portal
- Knox gateway
- HDFS name service
- YARN master
- Hive metastore
- InfluxDB (metrics store)
- ElasticSearch (logs store)
- Livy (REST interface for Spark)
- Spark Driver



# Controller

External REST/HTTPS Endpoint  
Bootstrap and Build out  
Manage Capacity  
Configure High Availability and recover from failure (AGs)  
Security (authN, authZ, certificate rotation)  
Lifecycle (upgrade/downgrade/rollback)  
Configuration management  
Monitoring - capacity, health, metrics, logs  
Troubleshooting – performance, failures  
Cluster Admin Portal

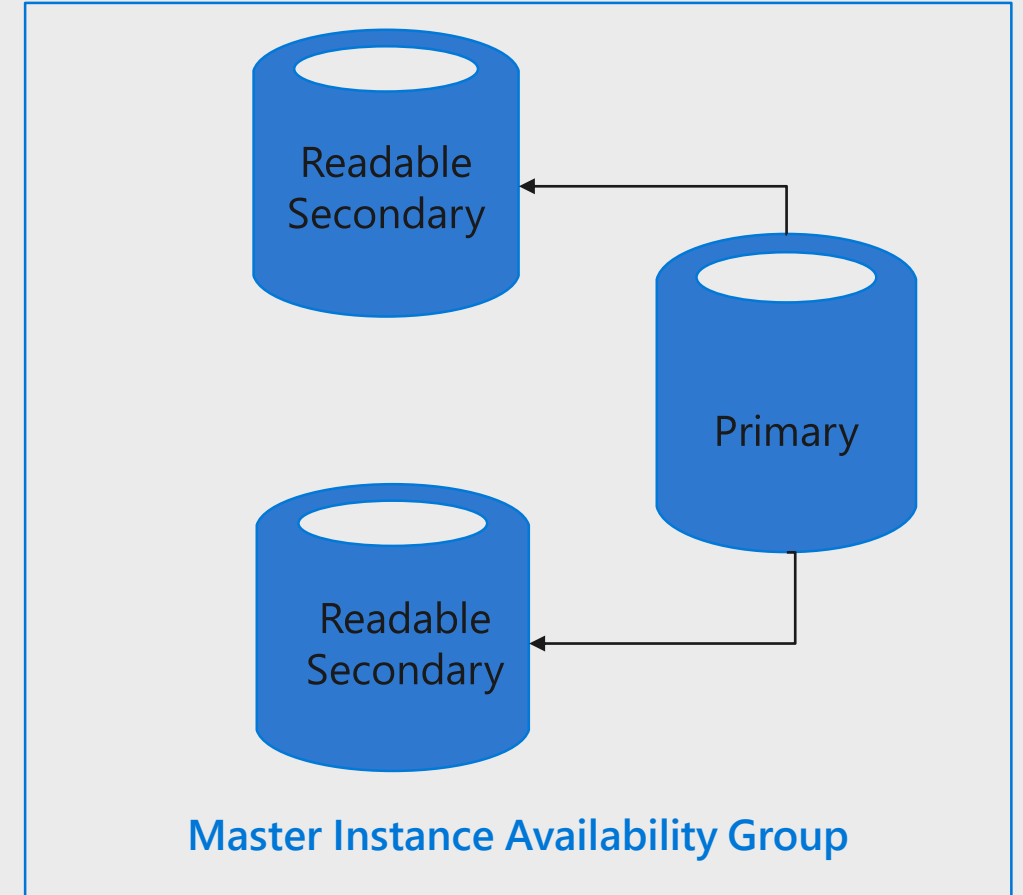




# SQL Master instance

TDS endpoint into the cluster  
High value data  
OLTP server  
Data connectors  
Machine learning & extensibility  
Scalable query engine with readable secondary replicas

Built-in high availability with Always On Availability Groups  
(coming soon)



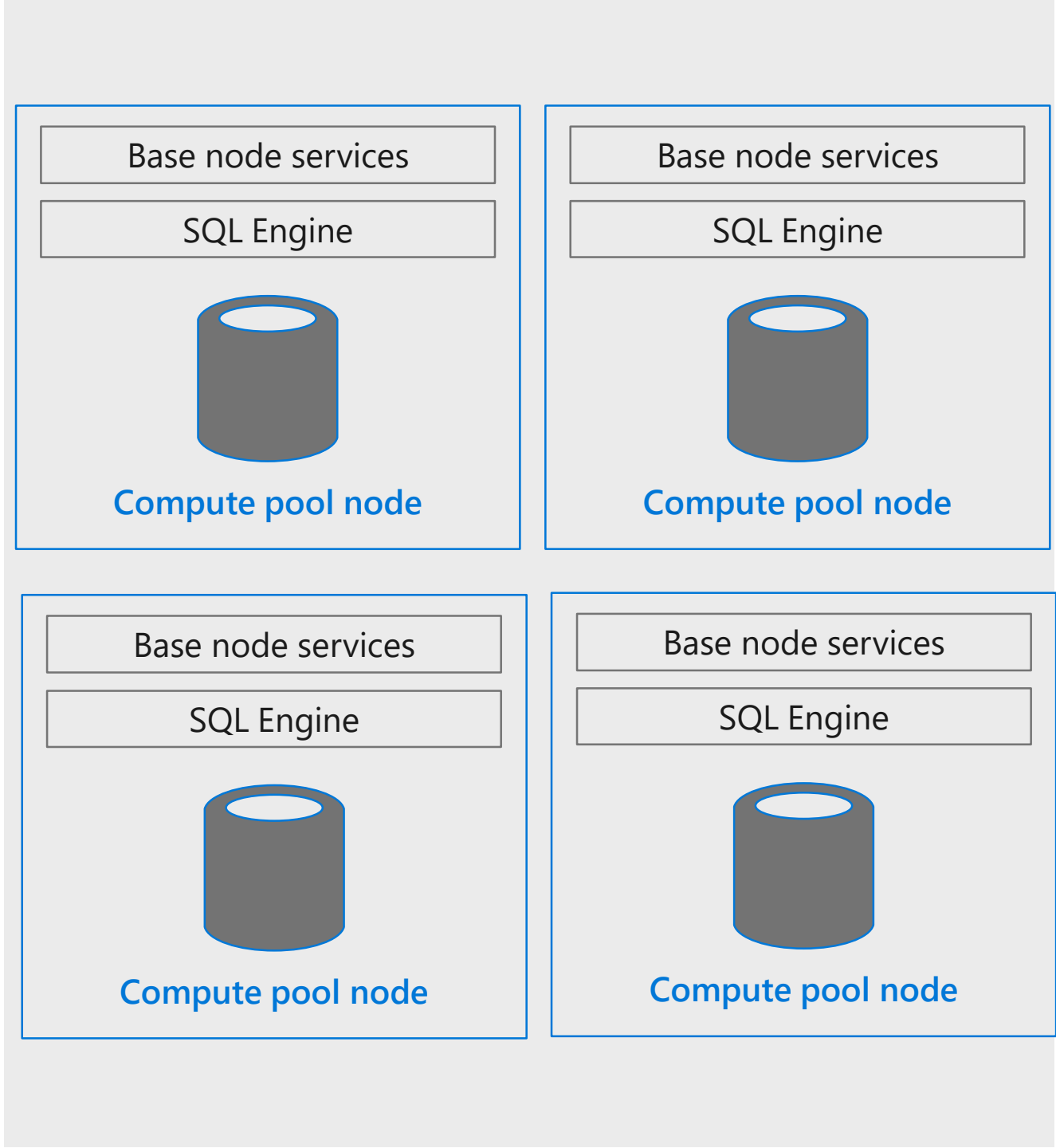
# Compute plane

Hosts one or more SQL Compute Pools

Compute pool is a group of instances that forms a data, security, and resource boundary.

Compute pool processes complex distributed queries against the data plane.

Local storage is used for shuffling data if necessary.



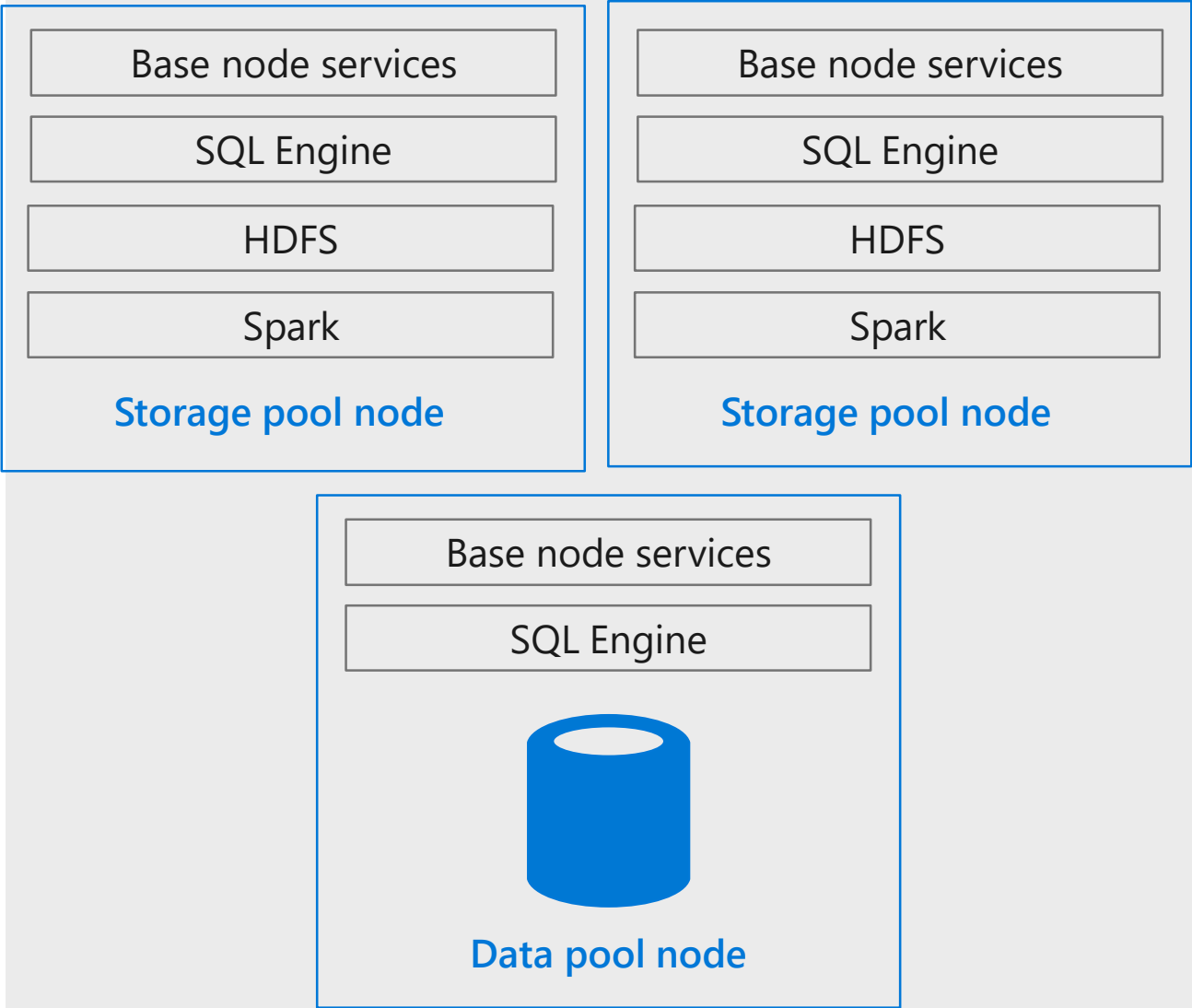
# Data plane

## Storage pool

- Data ingestion through Spark (batch and streaming)
- Data storage in HDFS
- Data access through HDFS and SQL endpoints
- SQL engine reads files in HDFS directly

## Data pool

- Partitioned, in-memory cache for external data or HDFS
- Scale-out data storage for append only data sets
- Data ingestion through Spark



# Unified development and administration

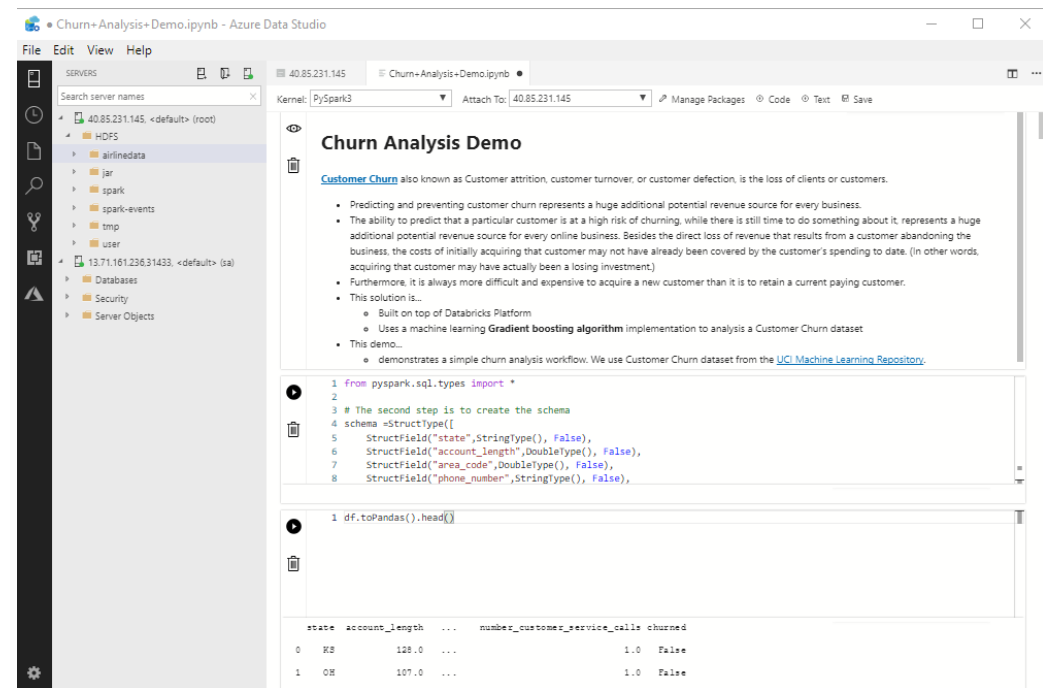
Azure Data Studio provides a unified tool for querying data using a notebook experience for both T-SQL and Spark

Easily access all your data across SQL Server and HDFS

The cluster administration portal provides easy to use cloud-style managed services for HA, monitoring, backup/recovery, security, and provisioning.

The REST API and command line tools simplify automation

The development and management experience is consistent regardless of where you run – on prem or any of the major cloud providers



# Demo #2

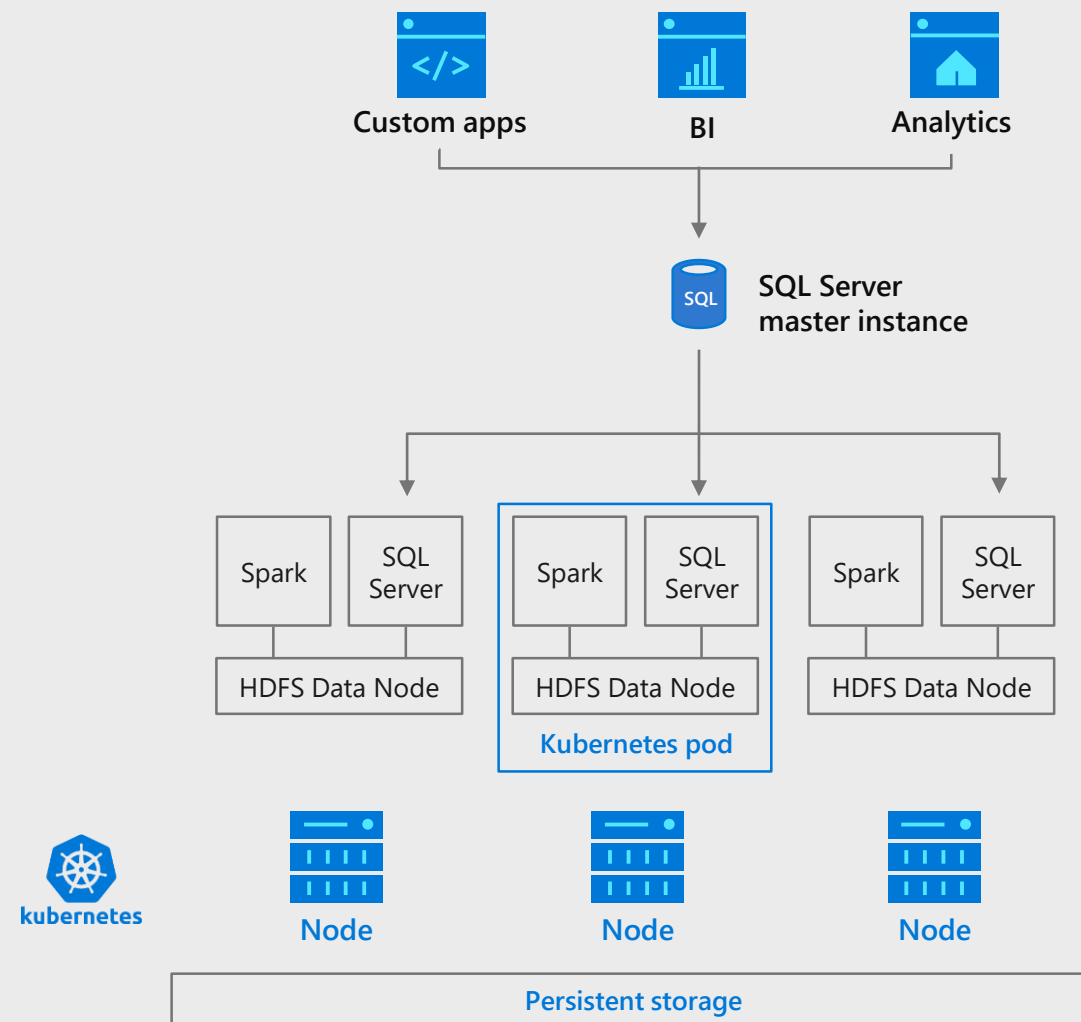
Deploying a SQL Server 2019 big data cluster

# Scale Big Data on demand

SQL Server can now read directly from HDFS files

Elastically scale compute and storage using HDFS-based storage pools with SQL Server and Spark built in

Apps, BI, and analytics access Big Data through the SQL Server master instance



# Increase performance for data virtualization

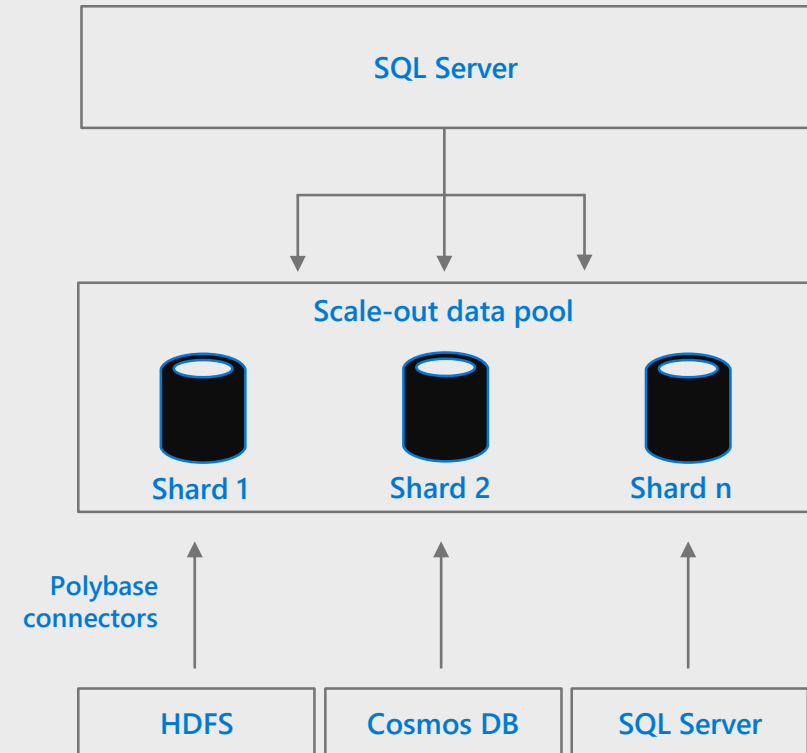
Scale-out data pools combine and cache data from many sources for fast querying

## Scenario

- A global car manufacturing company wants to join data from across multiple sources including HDFS, SQL Server, and Cosmos DB

## Solution

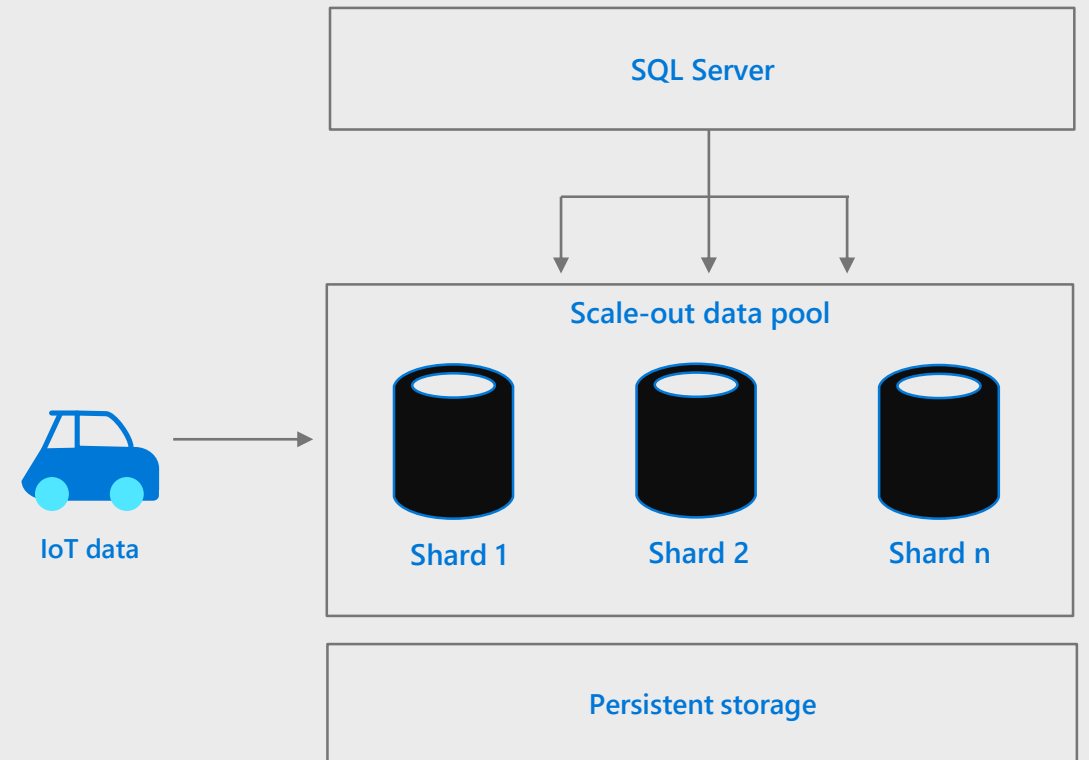
- Query data in relational and non-relational data stores with new PolyBase connectors
- Create a scale-out data pool cache of combined data
- Expose the datasets as a shared data source, **without writing code to move and integrate data**



# Scale out storage and query compute for better performance

Extend SQL Server with a scale-out storage tier by partitioning the data across multiple instances

Speed up query performance by scaling out the filtering and local aggregation across multiple instances





# Query execution over HDFS

## PolyBase (SQL Server 2016/2017)

- Uses Java libraries
- Push-down via Map-Reduce jobs
- Works against 3<sup>rd</sup> party HDFS data sources
- No integration with Hive metastore

## PolyBase (SQL Server 2019)

- Uses SQL instance natively
- Push-down using SQL queries to storage pool instance(s)
- Works against SQL big data cluster & others via HDFS tiering
- Built-in integration with Hive metastore

# Scale out storage

SQL Server can now read directly from HDFS files

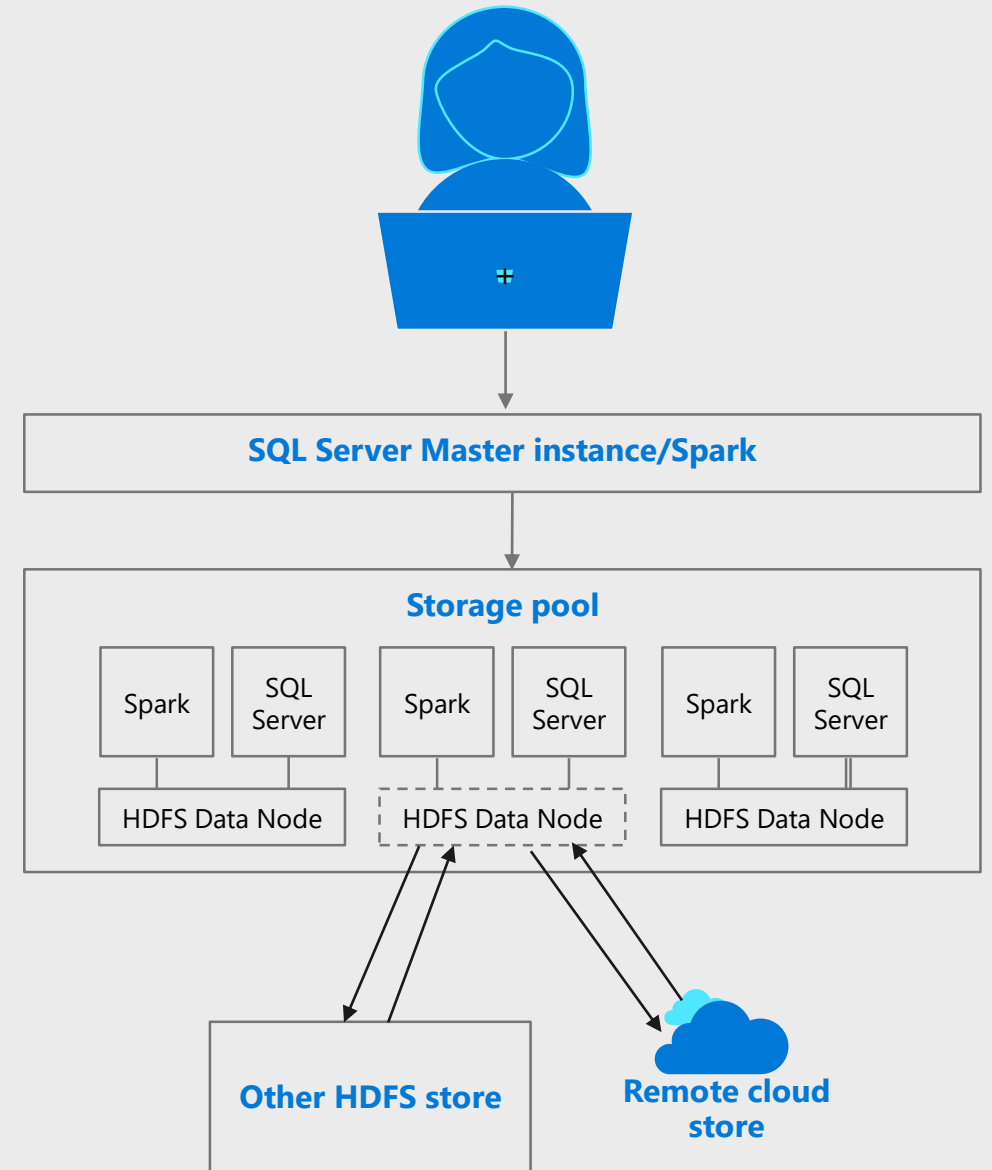
Elastically scale compute and storage using HDFS-based storage pools with SQL Server and Spark built in

Mount and manage remote stores through HDFS

Mount various on-prem and cloud data stores

Accelerate computation by caching data locally

Disaster recovery/Data backup



# Learn more

Preview SQL Server 2019 now: <https://aka.ms/ss19>

Join the early adoption program: <https://aka.ms/eapsignup>

See what's new in CTP 2.0: <http://docs.microsoft.com/sql/sql-server/what-s-new-in-sql-server-ver15>