



Azure Data

# Who are we ?



Laurent Leturgez

I'm a data and cloud Architect and Spark lover. I worked many years as an Oracle consultant and expert, and now I work with Cloud solutions devoted to solve complex problems with high volumes of data.



Alexandre Bergere

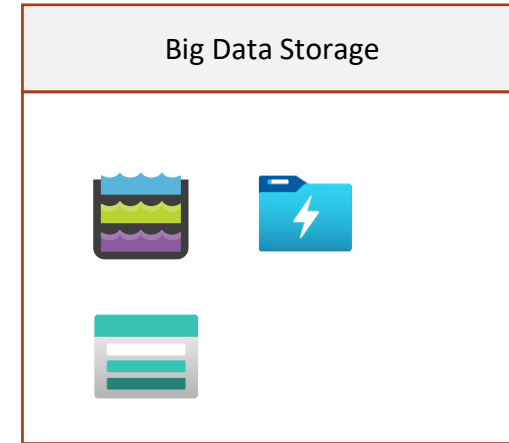
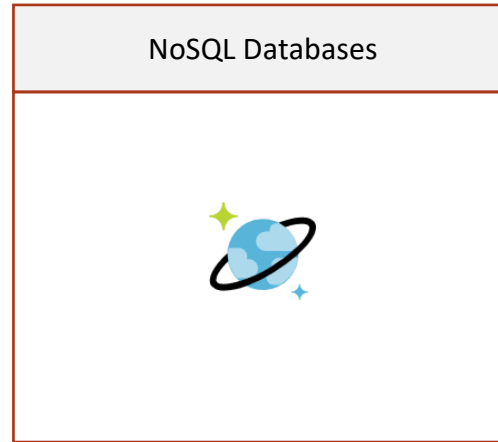
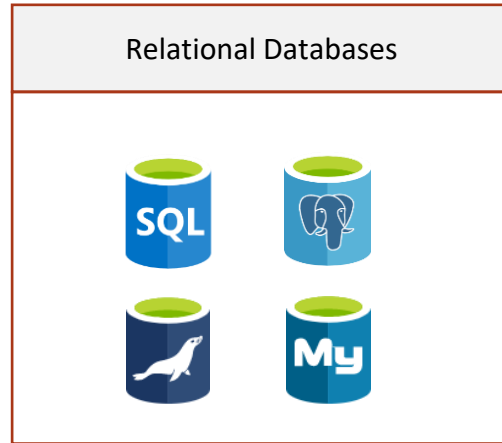
I am a Data Analyst & Solution Architect independent - ☁ MCSE, Cosmos DB & Delta lover. I developed my skills through various clients' projects, teaching at the University and personal proof of concepts. I'm also the Co-Founder of DataRedkite, a product which can quickly give to its user a good management of data in Microsoft Azure DataLake.



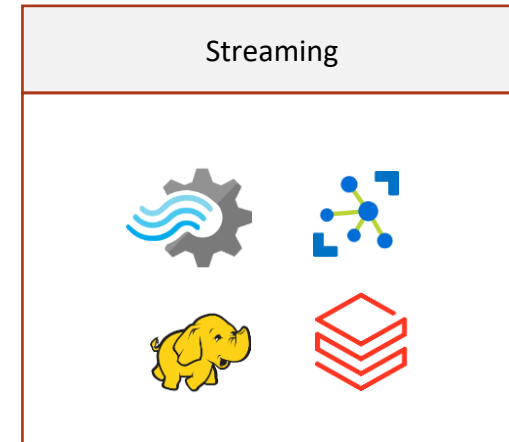
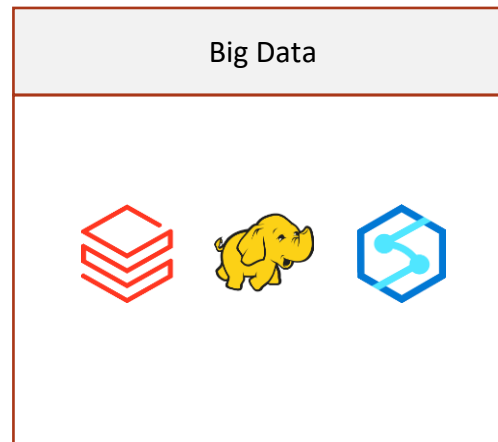
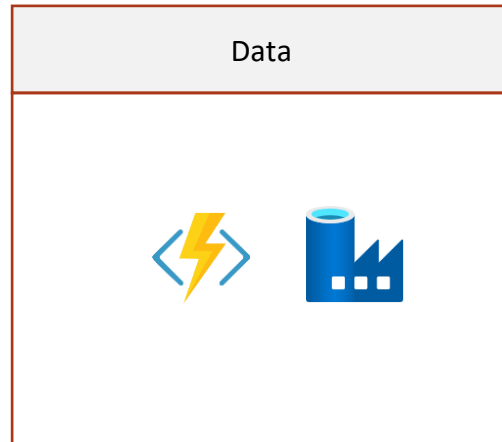
**DataRedKite**

# Summary

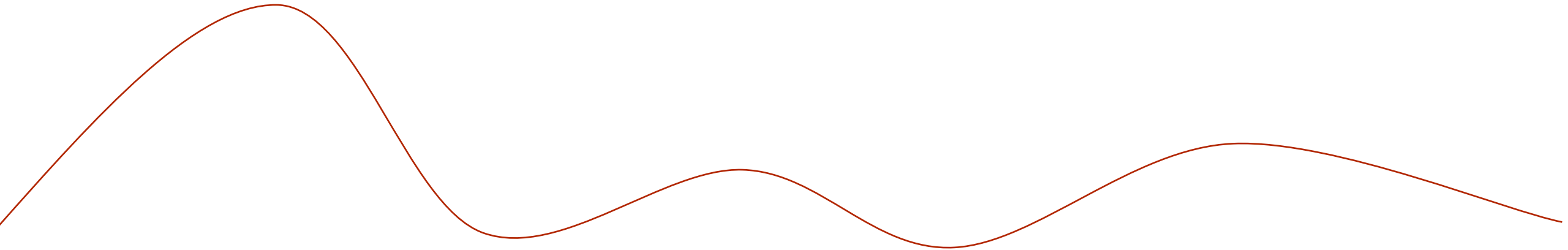
Storage :



Compute :



# Storage



# Relational Databases

## Azure SQL Database



Managed relational SQL Database  
as a service

## Azure Database for PostGres



Managed Postgres database  
service for app developers

## Azure Database for MariaDB



Managed MariaDB database  
service for app developers

## Azure Database for MySQL



Managed MySQL database service  
for app developers

# Relational Databases

## Azure SQL Database



Managed relational SQL Database  
as a service

## Azure Database for PostGres



Managed Postgres database  
service for app developers

## Azure Database for MariaDB



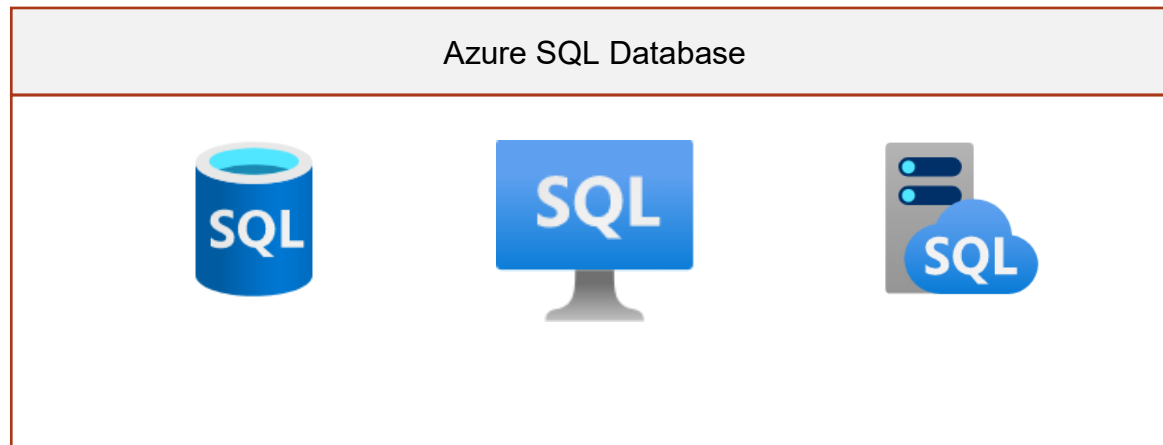
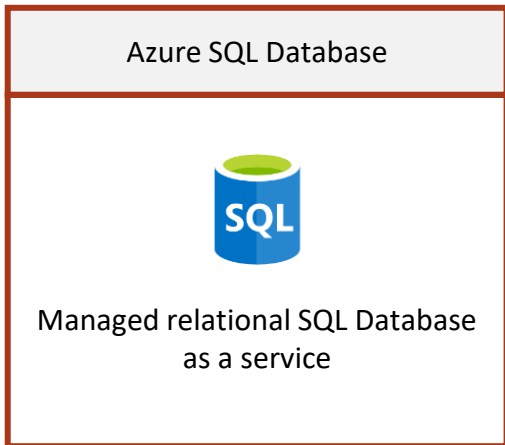
Managed MariaDB database  
service for app developers

## Azure Database for MySQL



Managed MySQL database service  
for app developers

# Relational Databases



# Azure SQL Database

- Azure SQL
  - SQL Server Paas service
    - Managed upgrades, patches, backups and monitoring
    - Latest Stable version of SQL Server
    - 99,99% availability
  - Deployment model
    - Single Database : database runs on non shared resources
    - Elastic Pool : database runs with a collection of databases that share set of resources at a predictable price



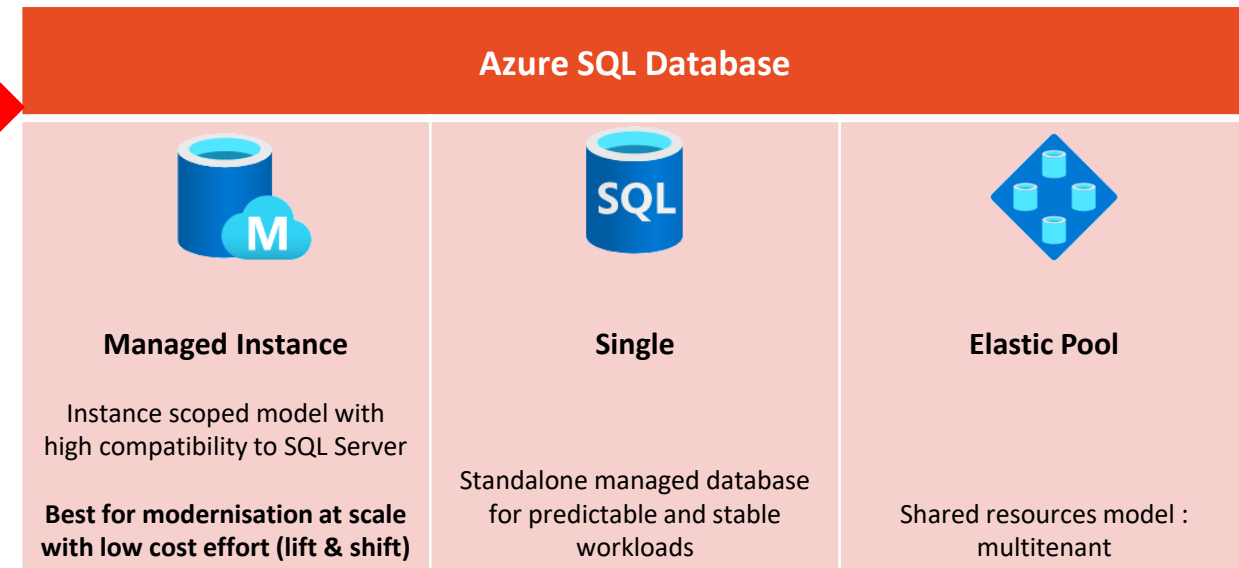
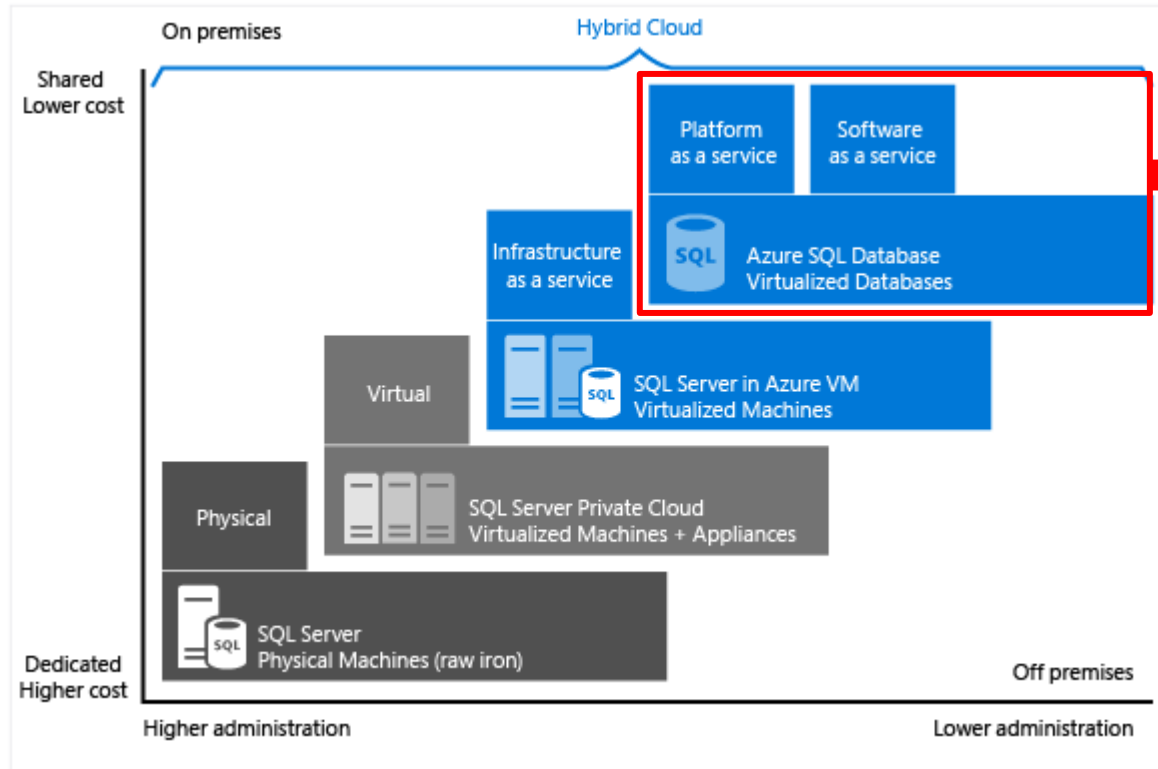
# Azure SQL Database

- Azure SQL
  - Purchasing model
    - DTU (Database Transaction Unit) : <https://docs.microsoft.com/en-us/azure/azure-sql/database/service-tiers-dtu>
      - Basic tier
      - Standard Tier
      - Premium Tier
    - vCore model
    - Serverless
  - Service Tier
    - General Purpose (vCore) / Standard (DTU) : Common workloads
    - Business Critical (vCore) / Premium (DTU) : High transaction and availability / low latency IO
    - HyperScale (vCore) :
      - Up to 100Tb Database
      - Rapid Scale up (compute resources)
      - Rapid Scale out (read only nodes : read workload / hot-standby)

# Azure SQL Database

- Azure SQL Managed Instance
  - Features
    - PaaS platform for lift and shift at scale
    - Broadest SQL Server engine compatibility (network integration, features etc.)
    - With preservation of all PaaS capabilities (patching, updates, backups, HA etc.)
  - vCore purchase model only
  - BYOL available
- SQL Virtual Machine
  - SQL Server deployment on VM (Linux and Windows)
  - Can choose SQL Server version
    - From 2008 R2
    - Up to 2019

# Azure SQL Database



# Relational Databases

## Azure SQL Database



Managed relational SQL Database  
as a service

## Azure Database for PostGre



Managed Postgres database  
service for app developers

## Azure Database for MariaDB



Managed MariaDB database  
service for app developers

## Azure Database for MySQL



Managed MySQL database service  
for app developers

# Azure Database for PostgreSQL

- Paas Service for PostgreSQL
  - Runs on Windows
  - Single Server
    - v9.5 to 11
    - Up to 64 vCores depending on SKU (<https://docs.microsoft.com/en-us/azure/postgresql/concepts-pricing-tiers>)
      - Up to 2 for Basic SKU
      - Up to 64 for General Purpose SKU
      - Up to 32 for Memory Optimized SKU
    - Bunch of PG Extensions available
    - Automated Backup (retention up to 35days)
    - Backup frequency and backup types depend on database size
    - Geo-redundant backup option (General Purpose & Memory Optimized)

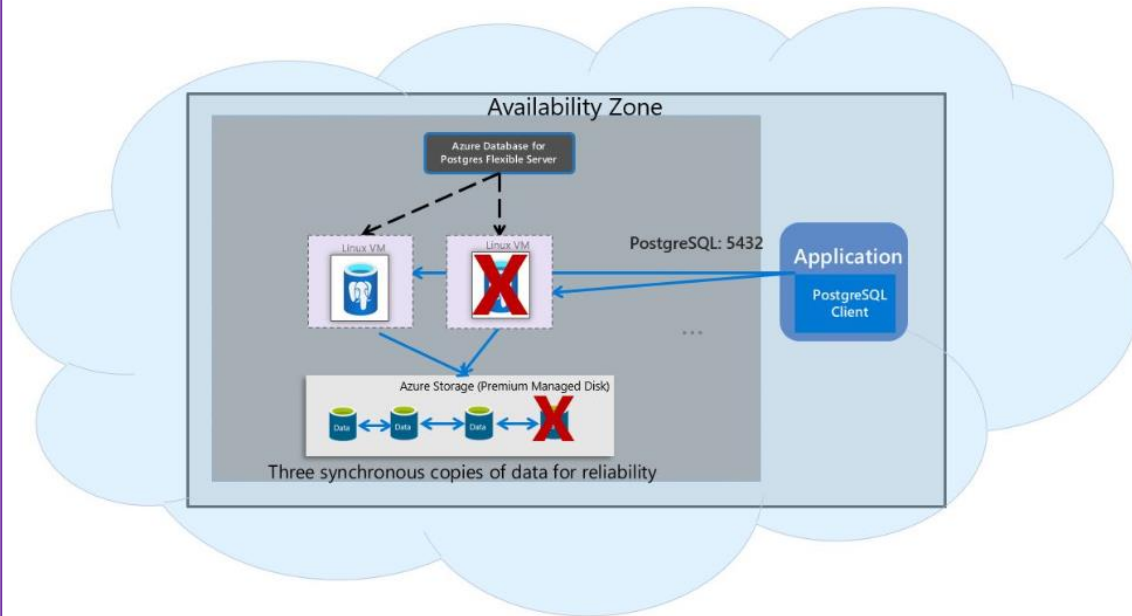
# Azure Database for PostgreSQL

- Paas Service for PostgreSQL
  - HyperScale (Citus)
    - High performance and analytical workloads beyond 100Gb
    - Hyperscale delivers
      - Horizontal scaling across multiple machine (with Sharding)
      - Query parallelization across these servers
      - High performance for analytics
    - Based on server groups
    - Design approach required for table distribution and performance
      - Distributed tables (based on distribution column)
      - Reference tables (content concentrated into a single shard replicated on every worker node)
      - Local tables (ordinary unsharded tables. Perfect for small tables not involded into joins)
  - Automated backup through storage snapshots

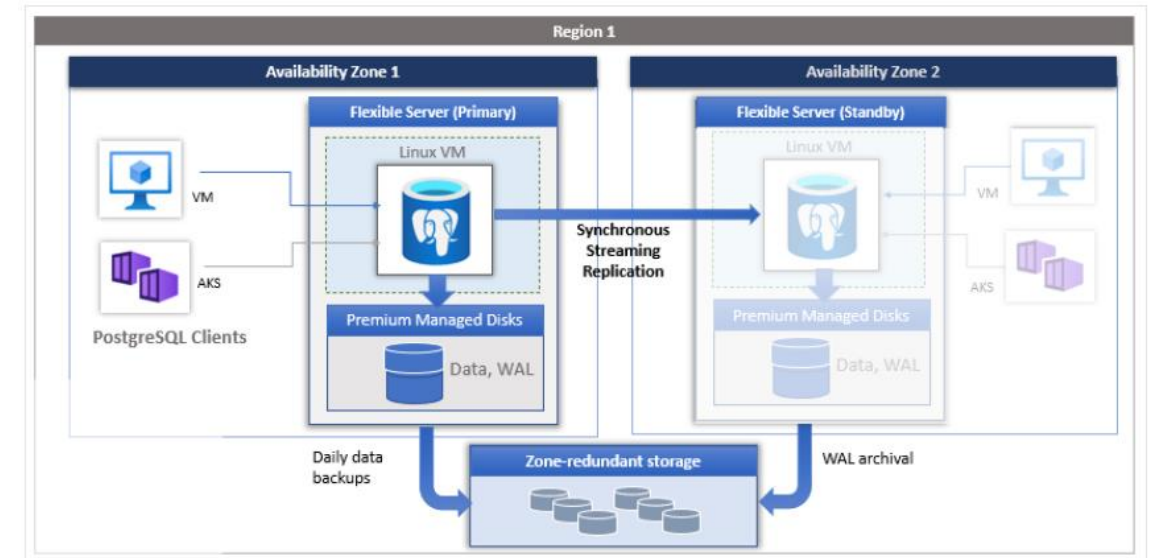
# Azure Database for PostgreSQL

- Paas Service for PostgreSQL
  - Flexible Server (Preview)
    - Automated patching
    - Automatic backups
    - Performance adjustment in three switchable compute tiers : Burstable, GP, Memory Optimized

## High Availability



## Zone Redundant HA (Optional)



# Relational Databases

## Azure SQL Database



Managed relational SQL Database  
as a service

## Azure Database for PostGre



Managed Postgres database  
service for app developers

## Azure Database for MariaDB



Managed MariaDB database  
service for app developers

## Azure Database for MySQL



Managed MySQL database service  
for app developers



# Azure Database for MariaDB

- Paas Service for MariaDB
  - Runs on Windows
  - Single Server
    - V10.2 and 10.3
    - Up to 64 vCores depending on SKU (<https://docs.microsoft.com/en-us/azure/mariadb/concepts-pricing-tiers>)
      - Up to 2 for Basic SKU
      - Up to 64 for General Purpose SKU
      - Up to 32 for Memory Optimized SKU
  - Automated Backup (retention up to 35days)
  - Backup frequency and backup types depend on database size
  - Geo-redundant backup option (General Purpose & Memory Optimized)

# Relational Databases

## Azure SQL Database



Managed relational SQL Database  
as a service

## Azure Database for PostGre



Managed Postgres database  
service for app developers

## Azure Database for MariaDB



Managed MariaDB database  
service for app developers

## Azure Database for MySQL



Managed MySQL database service  
for app developers

# Azure Database for MySQL

- Paas Service for MySQL
  - Runs on Windows
  - Single Server
    - V5.6, 5.7, and 8.0
    - Up to 64 vCores depending on SKU (<https://docs.microsoft.com/en-us/azure/mysql/concepts-pricing-tiers>)
      - Up to 2 for Basic SKU
      - Up to 64 for General Purpose SKU
      - Up to 32 for Memory Optimized SKU
  - Automated Backup (retention up to 35days)
  - Backup frequency and backup types depend on database size
  - Geo-redundant backup option (General Purpose & Memory Optimized)

# Azure Database for MySQL

- Paas Service for MySQL

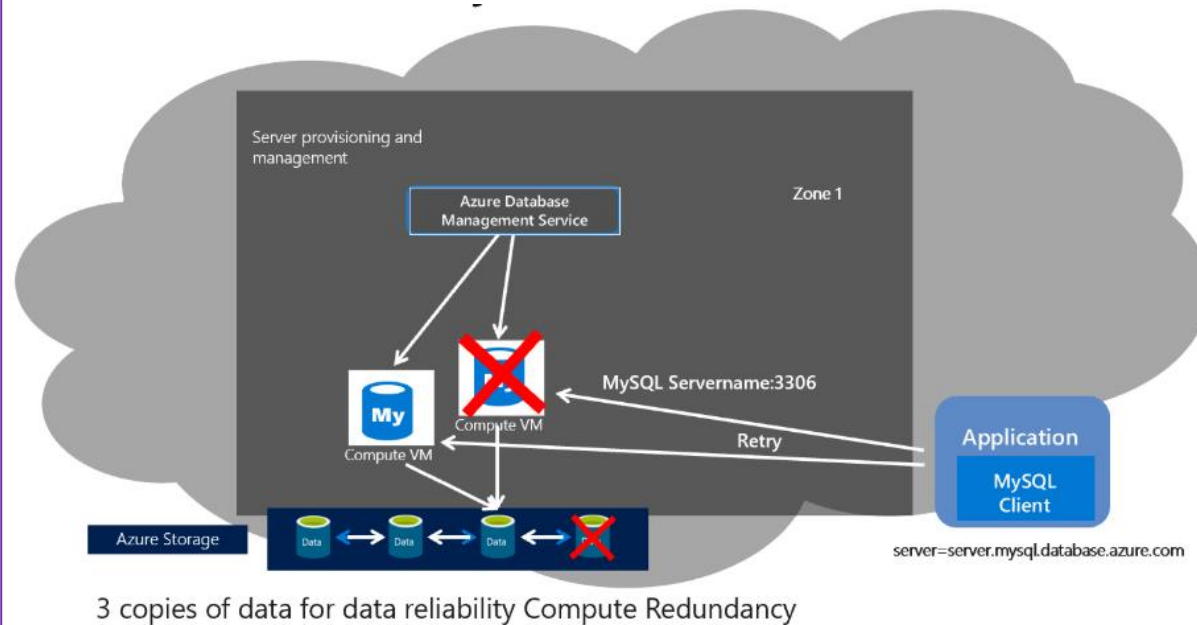
## Flexible Server (Preview)

- V5.7
- Automated patching
- Automatic backups
- Performance adjustment in three switchable compute tiers : Burstable, GP, Memory Optimized
- Network Isolation
  - Private Access through Vnet integration
  - Public Access

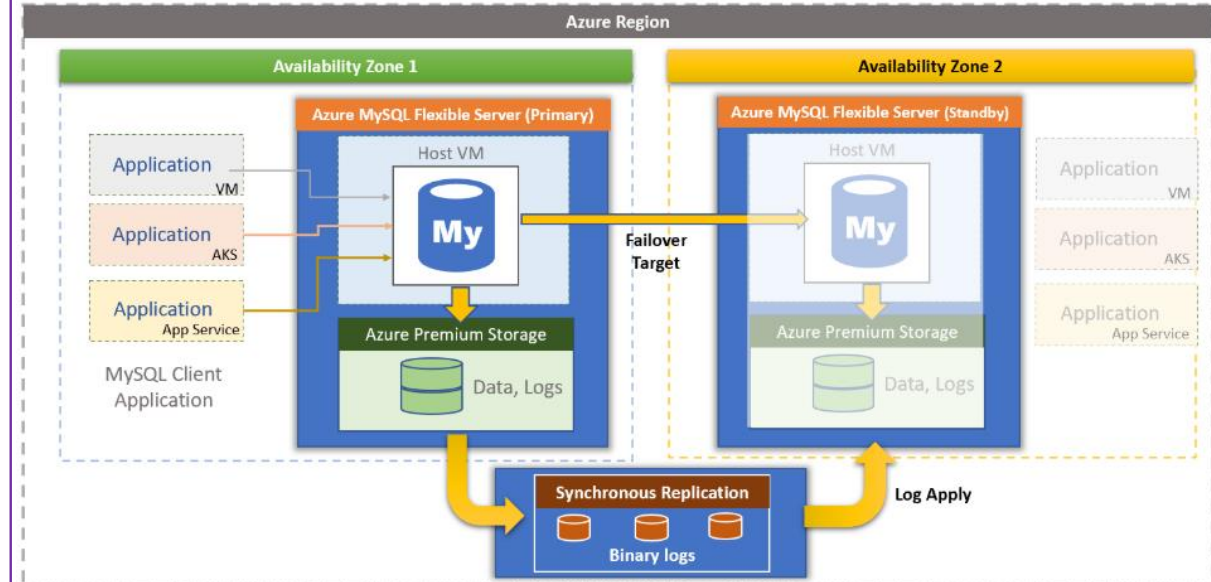
# Azure Database for MySQL

- Paas Service for MySQL  
Flexible Server (Preview)

## High Availability



## Zone Redundant HA (Optional)



# NOSQL Databases

Azure Cosmos DB

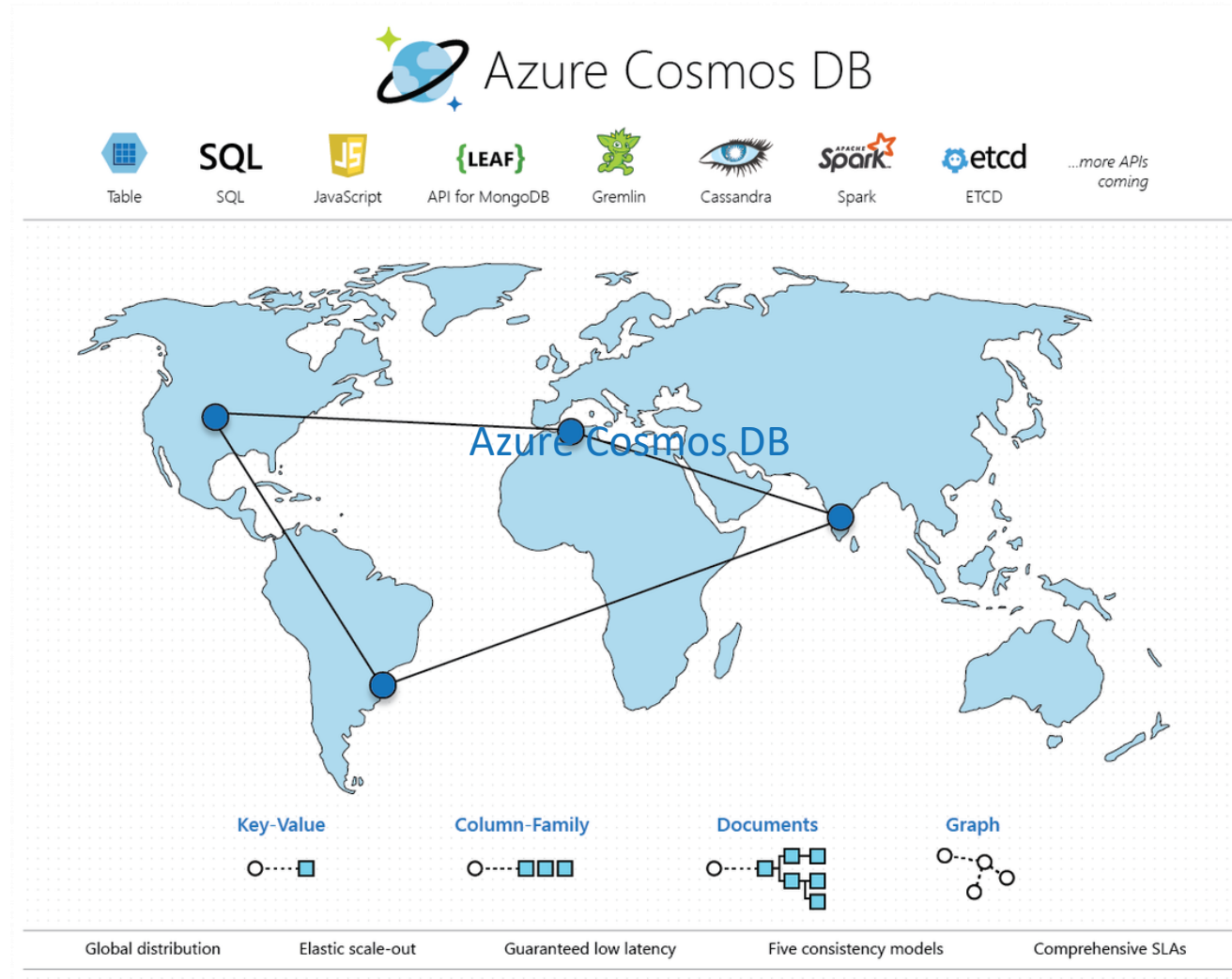


Globally distributed, multi-model  
database for any scale

# Azure Cosmos DB

A globally distributed, massively scalable, multi-model database service

- SQL API
- MongoDB API
- Cassandra API
- Gremlin API
- Table API



# Azure Cosmos DB




## Throughput

What are Request Units (RUs)?

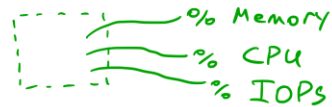
In Cosmos DB, expected performance must be provisioned.

Expressed in Request Units per Second (RU/s)

You can provision *abstracts system resources*

- at  database level
- at  container level
- or both 



*can be changed programmatically*



Learn More!

What are Request Units (RUs)?

Each request consumes # of RUs

- You can find the exact RU charge via  
- 1 read of 1KB item = 1 RU
  - 1 write of 1KB item = 5 RUs
  - Query: Depends on query & data items involved *variable RUs*
  - Indexing also affects RU cost

Learn More!

Cosmic Notes



# Big Data

## Storage

### Storage Account



REST-based object storage for  
unstructured data

### Azure Data Lake Storage



Massively scalable, secure data  
lake functionality built on Azure  
Blob Storage

# Big Data

## Storage

### Storage Account



REST-based object storage for  
unstructured data

### Azure Data Lake Storage



Massively scalable, secure data  
lake functionality built on Azure  
Blob Storage

# Storage Account

Azure Storage accounts are the base storage type within Azure. Azure Storage offers a very scalable object store for data objects and file system services in the cloud. It can also provide a messaging store for reliable messaging, or it can act as a NoSQL store.

Azure selected four of these data services and placed them together under the name Azure Storage. The four services are Azure Blobs, Azure Files, Azure Queues, and Azure Tables. The following illustration shows the elements of Azure Storage



- Azure Blobs : A scalable object store for text and binary data



- Azure Files : Managed file shares for cloud or on-premises deployments



- Azure Queues : A messaging store for reliable messaging between application components



- Azure Tables : A NoSQL store for no-schema storage of structured data

# Storage Account

## Type of Storage Account

Storage account type	Services	Redundancy options
General-purpose V2	Basic storage account type for blobs, files, queues, and tables. Recommended for most scenarios using Azure Storage.	LRS, GRS, RA-GRS, ZRS, GZRS, RA-GZRS
General-purpose V1	<b>Legacy</b> account type for blobs, files, queues, and tables. Use general-purpose v2 accounts instead when possible.	LRS, GRS, RA-GRS
BlockBlobStorage	Storage accounts with premium performance characteristics for block blobs and append blobs. Recommended for scenarios with high transactions rates, or scenarios that use smaller objects or require consistently low storage latency.	LRS, ZRS
FileStorage	Files-only storage accounts with premium performance characteristics. Recommended for enterprise or high performance scale applications.	LRS, ZRS
BlobStorage	Legacy Blob-only storage accounts. Use general-purpose v2 accounts instead when possible.	LRS, GRS, RA-GRS

# Replication Options

## Logically Replicated Storage (LRS)

Replicated three times within a storage scale unit (collection of racks of storage nodes) hosted in a datacenter in the same region as your storage account was created.

## Zone Replicated Storage (ZRS)

Replicated three times across one or two datacenters in addition to storing three replicas similar to LRS. Data stored in ZRS is durable even in the event that the primary datacenter is unavailable or unrecoverable.

## Geographically Replicated Storage (GRS)

Replicates your data to a second region that is hundreds of miles away from the primary region. Your data is curable even in the event of a complete region outage.

## Read Only Geographically Replicated Storage (RA-GRS)

Same replication as per GRS but also provides read access to the data in the other region.

# Replication Strategy

Replication Strategy	LRS	ZRS	GRS	RA-GRS
Data is replicated across multiple datacenters?	No	Yes	Yes	Yes
Data can be read from a secondary location <i>and</i> the primary location?	No	No	No	Yes
Number of copies of data maintained on separate nodes:	3	3	6	6

# Big Data

## Storage

### Storage Account



REST-based object storage for  
unstructured data

### Azure Data Lake Storage



Massively scalable, secure data  
lake functionality built on Azure  
Blob Storage

# Azure Datalake Store

Azure Data Lake Storage is a **Hadoop-compatible data repository** that can store any size or type of data. This storage service is available as Generation 1 (Gen1) or Generation 2 (Gen2).



Azure Datalake Gen 1



Azure Datalake Gen 2

Key features of Data Lake Storage:

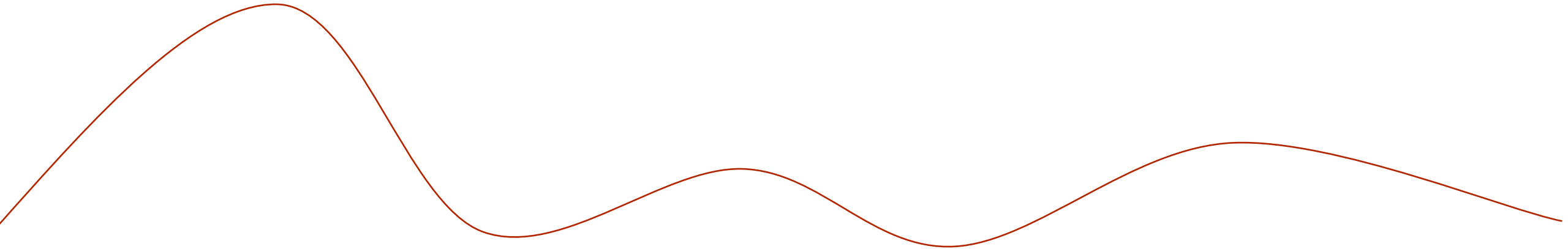
- Unlimited scalability
- Hadoop compatibility
- Security support for both access control lists (ACLs) & RBAC (for Gen 2 only)
- POSIX compliance
- An optimized Azure Blob File System (ABFS) driver that's designed for big-data analytics
- Zone-redundant storage
- Geo-redundant storage



# Choose a storage solution on Azure

	Data classification	Operations	Latency & throughput	Transactional support	Recommended service
Product catalog data	Semi-structured because of the need to extend or modify the schema for new products	<ul style="list-style-type: none"> <li>Customers require a high number of read operations, with the ability to query on many fields within the database.</li> <li>The business requires a high number of write operations to track the constantly changing inventory.</li> </ul>	High throughput and low latency	Required	Azure Cosmos DB
Photos and videos	Unstructured	<ul style="list-style-type: none"> <li>Only need to be retrieved by ID.</li> <li>Customers require a high number of read operations with low latency.</li> <li>Creates and updates will be somewhat infrequent and can have higher latency than read operations.</li> </ul>	Retrievals by ID need to support low latency and high throughput. Creates and updates can have higher latency than read operations.	Not required	Azure Blob storage
Business data	Structured	Read-only, complex analytical queries across multiple databases	Some latency in the results is expected based on the complex nature of the queries	Required	Azure SQL Database Azure Database for MariaDB Azure Database for PostgreSQL Azure Database for MySQL

# Compute



# Data

Compute

Azure Functions



Process events with serverless  
code

Data Factory



Managed data-integration solution

# Data

Compute

Azure Functions



Process events with serverless  
code

Data Factory



Managed data-integration solution

# Azure Function

Azure Functions is the **serverless compute service** from Microsoft. Functions are event-driven: each function defines a trigger — the exact definition of the event source, for instance, the name of a storage queue.



## Uses cases:

If you want to...	then...
Build a web API	Implement an endpoint for your web applications using the <a href="#">HTTP trigger</a>
Process file uploads	Run code when a file is uploaded or changed in <a href="#">blob storage</a>
Build a serverless workflow	Chain a series of functions together using <a href="#">durable functions</a>
Respond to database changes	Run custom logic when a document is created or updated in <a href="#">Cosmos DB</a>
Run scheduled tasks	Execute code at <a href="#">set times</a>
Create reliable message queue systems	Process message queues using <a href="#">Queue Storage</a> , <a href="#">Service Bus</a> , or <a href="#">Event Hubs</a>

# Azure Function

## Azure Functions hosting options : Azure Plan

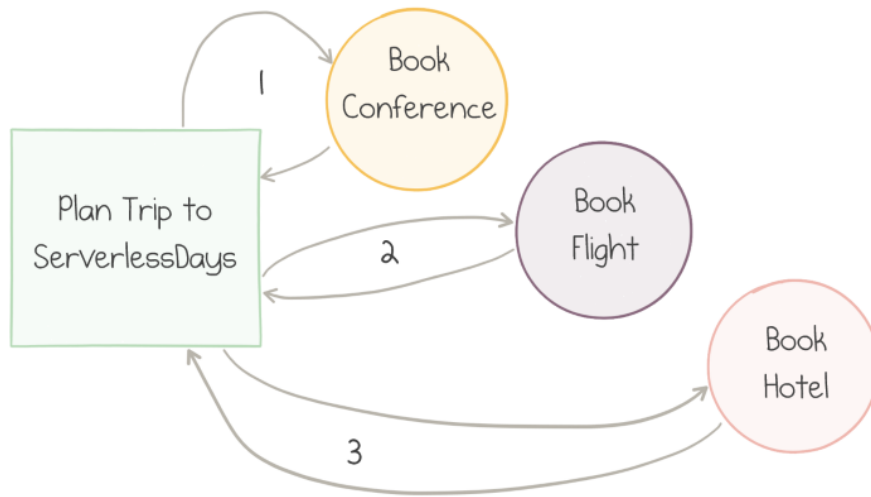
Consumption Plan	Functions
Consumption Plan (B1, B2, B3, S1, S2, S3)	Scale automatically and only pay for compute resources when your functions are running. On the Consumption plan, instances of the Functions host will be dynamically added and removed based on the number of incoming events.
Premium plan (P1v2, P2v2, P3v3)	While automatically scaling based on demand, use prewarmed workers to run applications with no delay after being idle, run on more powerful instances and connect to VNets.
Azure App Service plan	Run Functions within an App Service plan at regular App Service plan rates. Good fit for long-running operations, as well as when more predictive scaling and costs are required.

# Azure Function

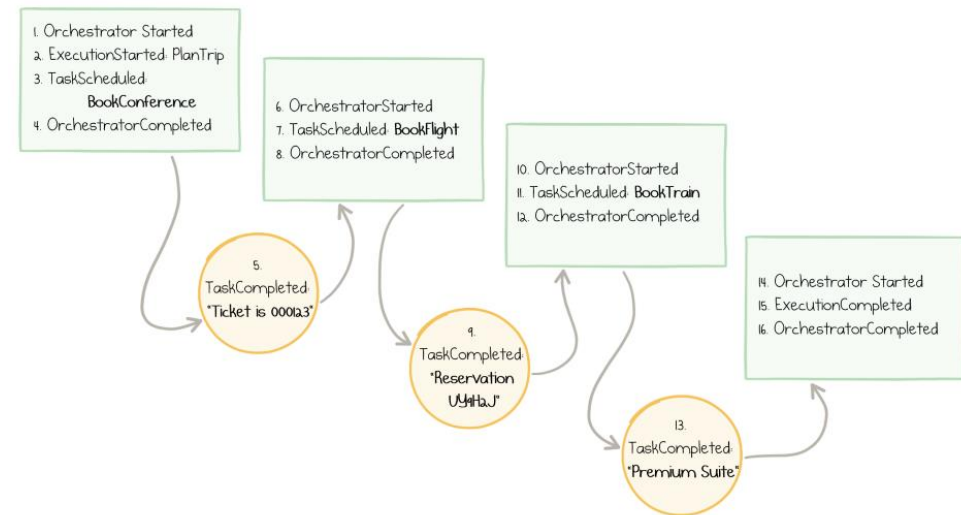
## Azure Durable Functions



**Durable Functions** is a library that brings **workflow orchestration abstractions** to Azure Functions. It introduces a number of idioms and tools to define stateful, potentially long-running operations, and manages a lot of mechanics of reliable communication and state management behind the scenes.



3 steps of a workflow executed in sequence



Log of events in the course of orchestrator progression

# Data

Compute

Azure Functions



Process events with serverless  
code

Data Factory

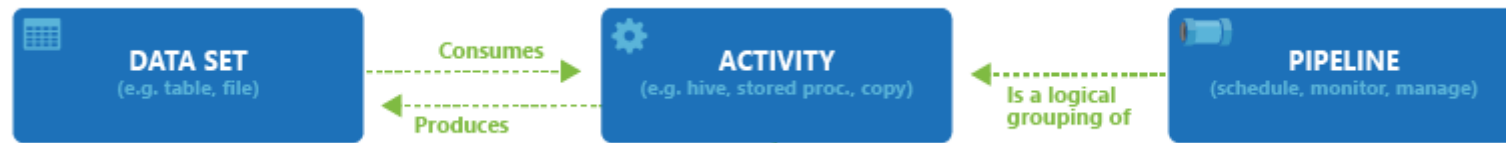


Managed data-integration solution

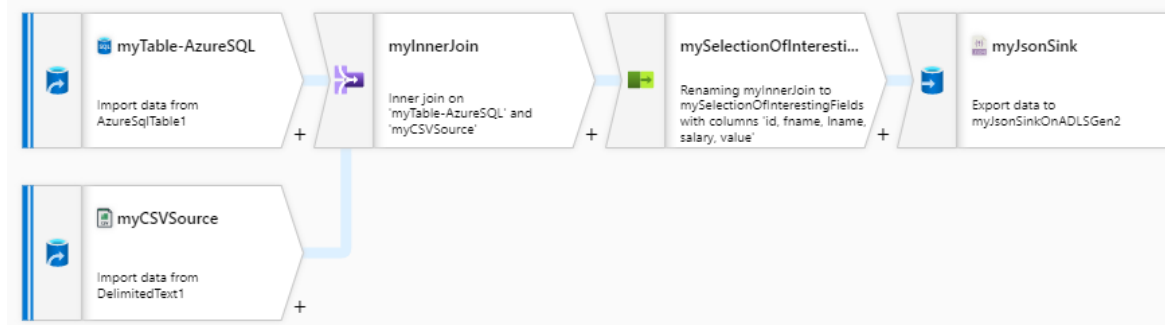


# Azure Data Factory

- Serverless Data Integration service
  - Data Pipeline : logical group of activities



- Data Flow : Data Transformation activity



- Data Copy : Data Transfer activity
- SSIS Integration
- Git integration

# Azure Data Factory

- Serverless Data Integration service
  - Job scheduling
    - Automatically through internal Scheduler
    - Manually
      - SDK : .NET, Python
      - REST API
      - PowerShell

# Azure Data Factory

- Serverless Data Integration service
  - Integration runtime
    - Compute infrastructure used by ADF to provide data integration
      - Azure : Serverless
      - Self Hosted : Onprem or Azure Virtual Machine (Windows)
      - SSIS

	Activity	Features
Azure	Data Flow Data Copy Dispatch Activity (HDI, Databricks, SQL ...)	Cloud to Cloud data transfer/flows
Self-Hosted	Data Flow Data Copy Dispatch Activity (HDI, Databricks, SQL ...)	OnPrem or Virtual Machine deployment (Windows) OnPrem <-> Cloud data transfer/flows When connectors are not available
SSIS	SSIS Package execution	Private or public Network

# Big Data

## Compute

### Azure Databricks



Fast, easy, and collaborative  
Apache Spark-based analytics  
platform

### Azure HDInsight



HDInsight supports the latest open  
source projects from the Apache  
Hadoop and Spark ecosystems.

### Azure Synapse Analytics



Managed Enterprise  
Datawarehouse and BigData  
Analytics service

# Big Data

## Compute

### Azure Databricks



Fast, easy, and collaborative  
Apache Spark-based analytics  
platform

### Azure HDInsight



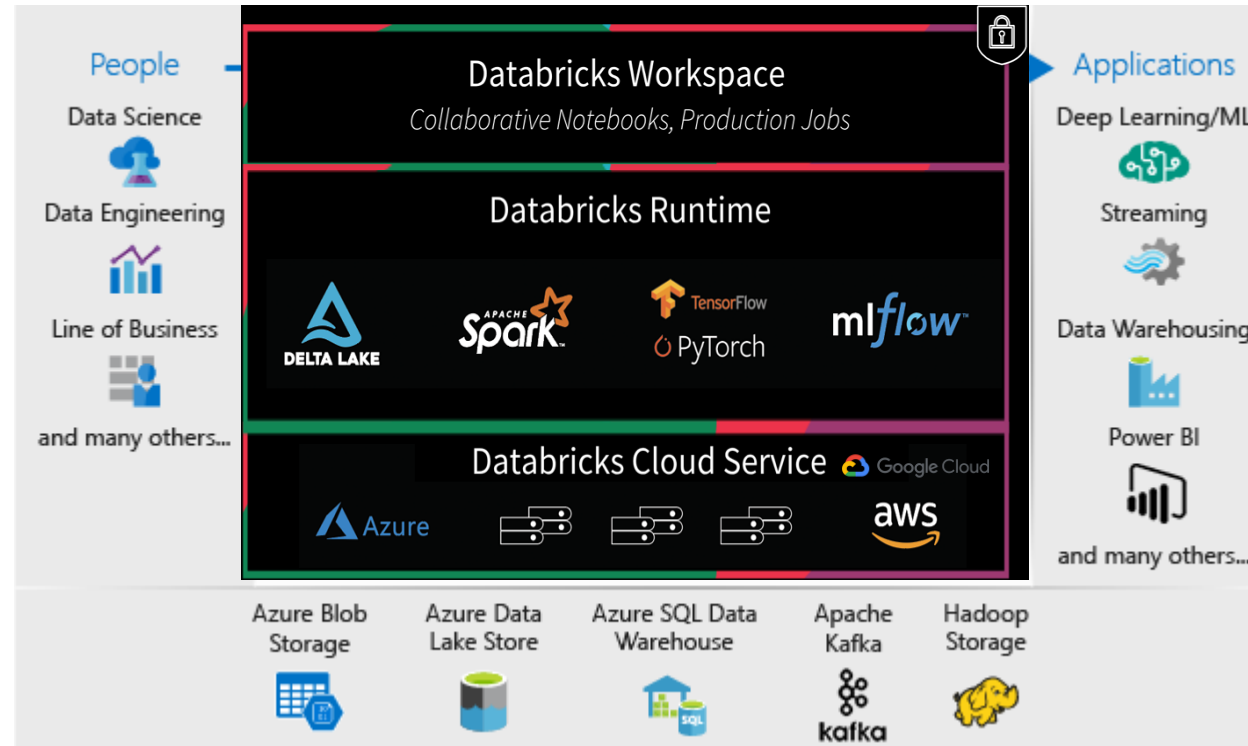
HDInsight supports the latest open  
source projects from the Apache  
Hadoop and Spark ecosystems.

### Azure Synapse Analytics



Managed Enterprise  
Datawarehouse and BigData  
Analytics service

# Azure Databricks



# Azure Databricks

Azure Databricks is a data analytics platform optimized for the Microsoft Azure cloud services platform. Azure Databricks offers two environments for developing data intensive applications:



Workspace  
Data Science and Engineering

Azure Databricks Workspace: provides an interactive workspace that enables collaboration between data engineers, data scientists, and machine learning engineers.



SQL Analytics  
SQL Editor and Dashboarding

Azure Databricks SQL Analytics: provides an easy-to-use platform for analysts who want to run SQL queries on their data lake, create multiple visualization types to explore query results from different perspectives, and build and share dashboards.

# Big Data

## Compute

### Azure Databricks



Fast, easy, and collaborative  
Apache Spark-based analytics  
platform

### Azure HDInsight



HDInsight supports the latest open  
source projects from the Apache  
Hadoop and Spark ecosystems.

### Azure Synapse Analytics



Managed Enterprise  
Datawarehouse and BigData  
Analytics service

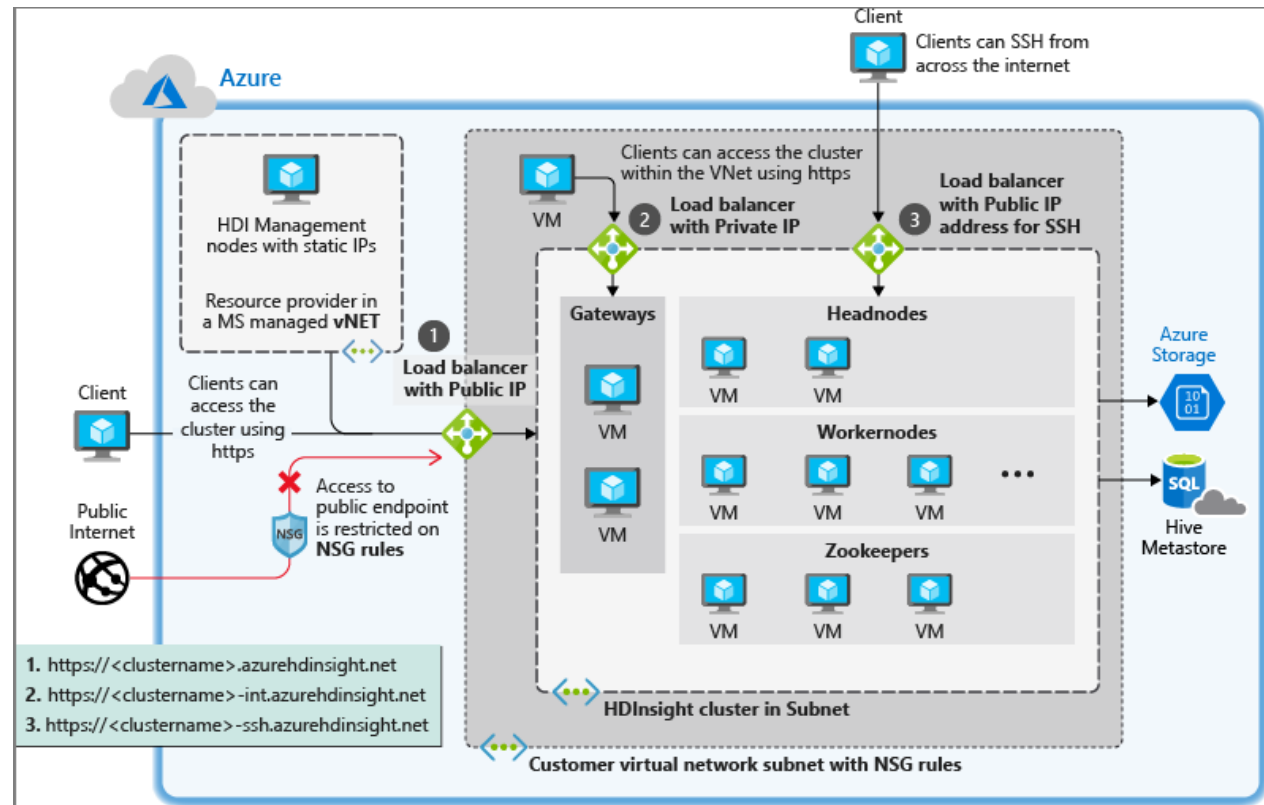


# Azure HDInsights

- Managed Hadoop distribution for Azure
- Based on Cloudera Hortonworks hadoop distribution
- Comes in various flavours / shapes (VM shapes and number)
  - Hadoop : General purpose (HDFS, Yarn, MapReduce, Hive, Pig, Sqoop, Oozie)
  - Spark
  - Kafka
  - HBase
  - Hive / LLAP (Interactive Query)
  - Storm (Stream processing)
  - ML Services with R

# Azure HDInsights

- At least one Storage account mandatory (for libs and binaries)
- External Metastores available for Ambari, Hive and Oozie
- HDInsights architecture



# Big Data

## Compute

### Azure Databricks



Fast, easy, and collaborative  
Apache Spark-based analytics  
platform

### Azure HDInsigh



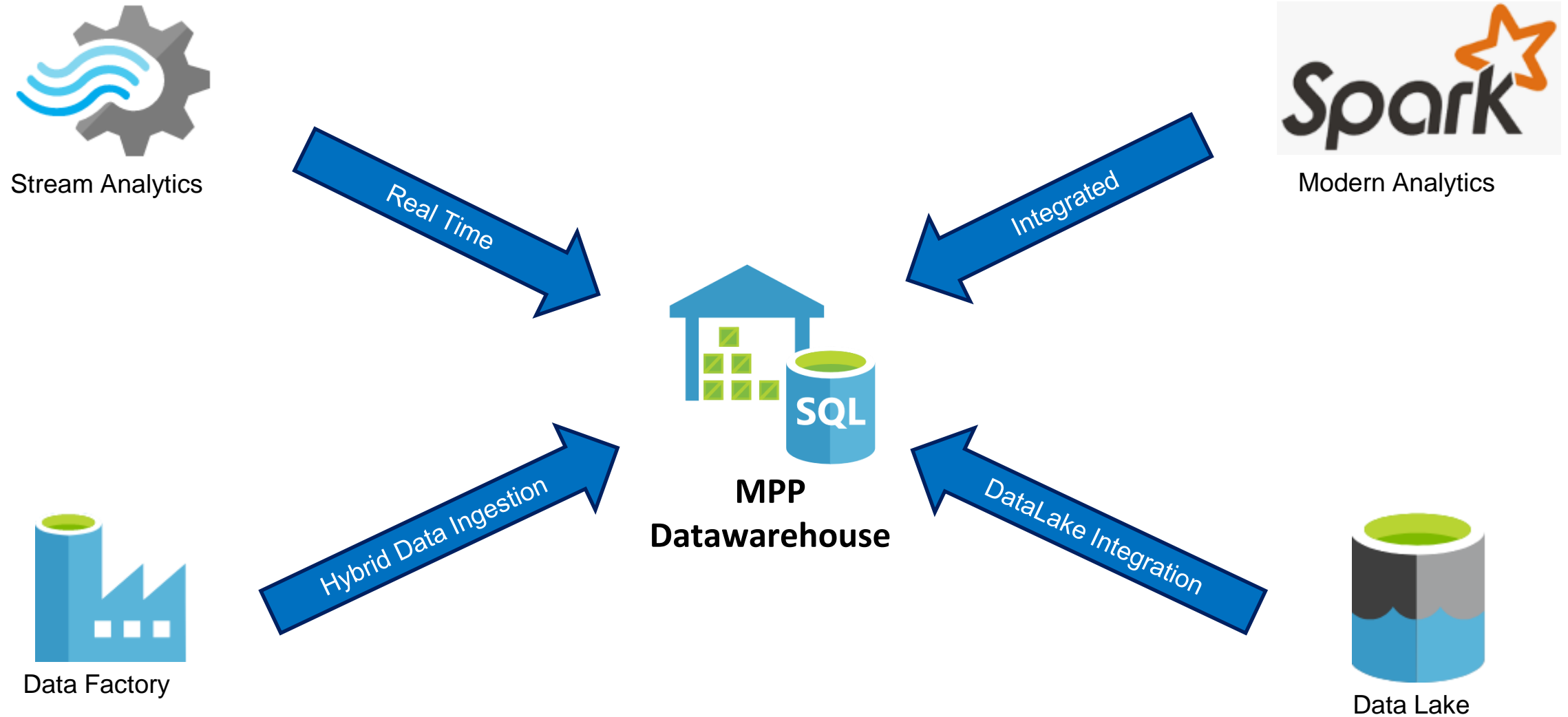
HDInsight supports the latest open  
source projects from the Apache  
Hadoop and Spark ecosystems.

### Azure Synapse Analytics

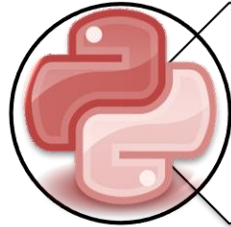
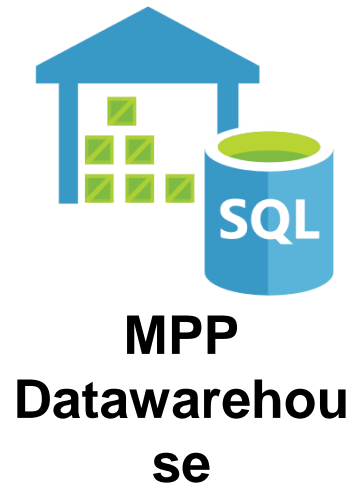


Managed Enterprise  
Datawarehouse and BigData  
Analytics service

# Azure Synapse Analytics



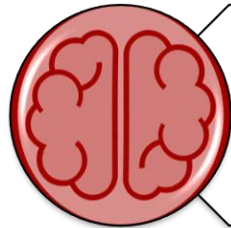
# Azure Synapse Analytics



Choice of language (T-SQL, Spark SQL, Python, Scala, .Net)



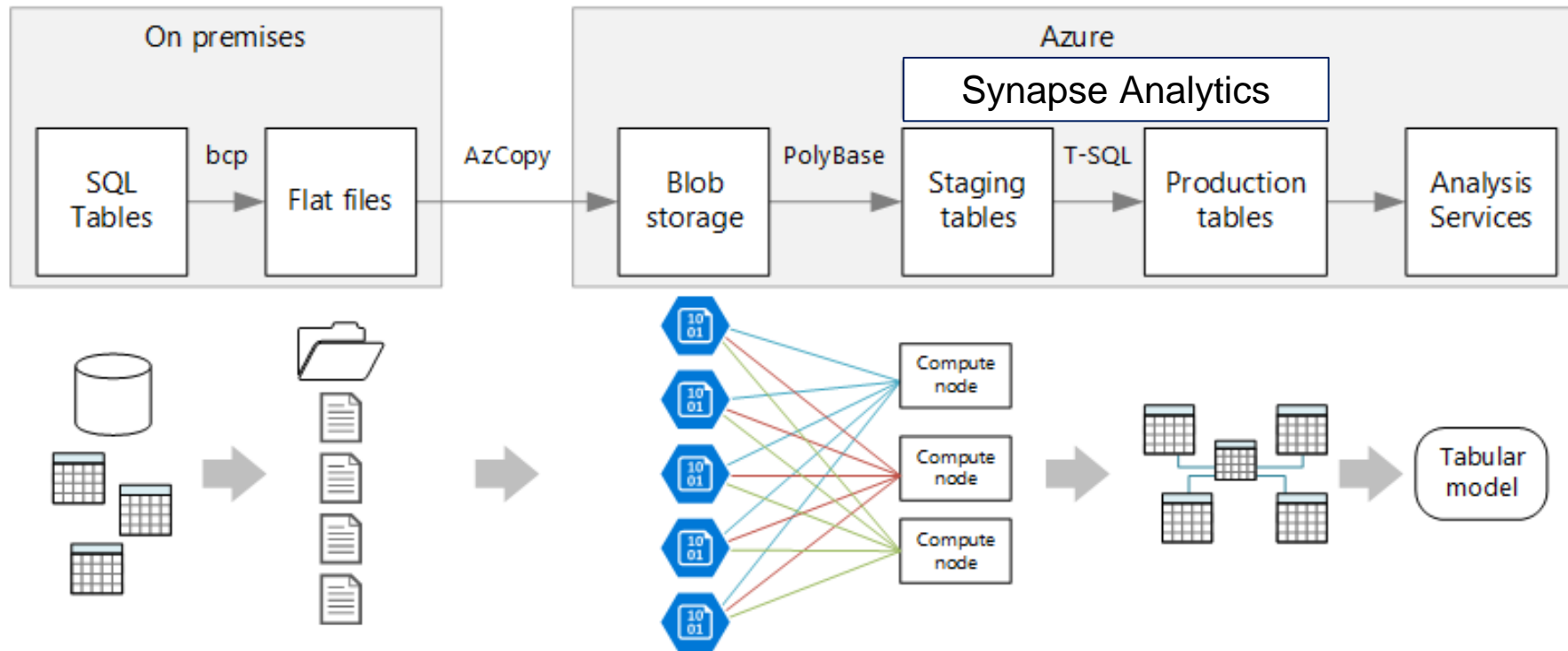
Analytics ready (Analysis Services, Power BI)



Data Science and AI Ready (Azure Machine Learning integration)

# Azure Synapse Analytics

- Sample Use Case : Pure Business Intelligence !

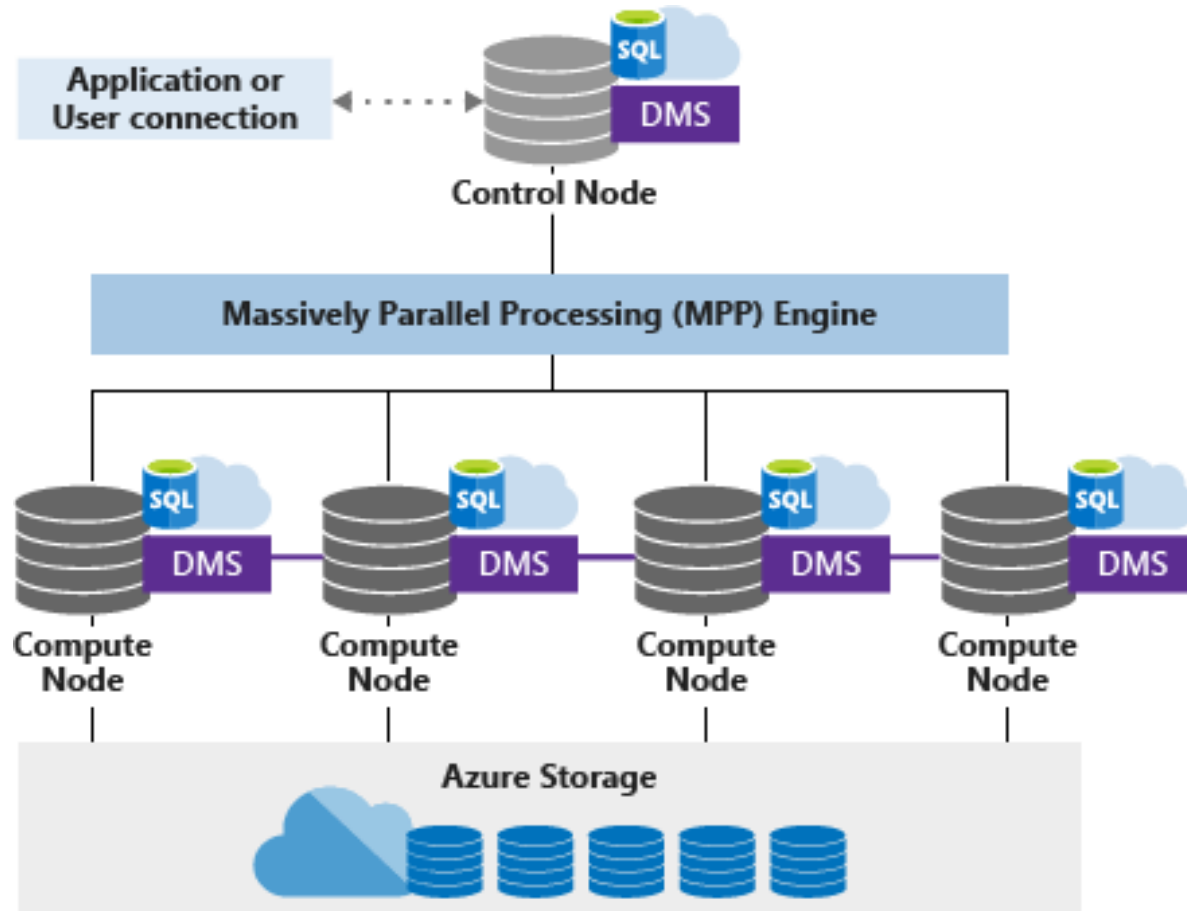


# Azure Synapse Analytics

- Not for small database (Usually > 1Tb)
- Cost Model
  - Synapse Provisioned
    - T-SQL Pool with DWU (Datawarehouse Units)
    - Storage (Geo redundant option)
  - Synapse Serverless
  - Spark Pools
  - Synapse Pipeline

# Azure Synapse Analytics

- Architecture

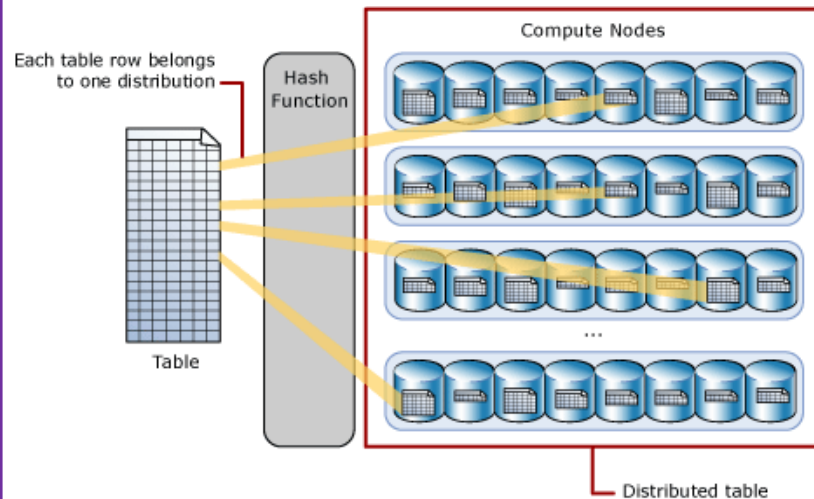


- DMS (Data Movement Service)
  - Used for Data Colocation
- **Key point: Data Partitioning and Data Distribution**

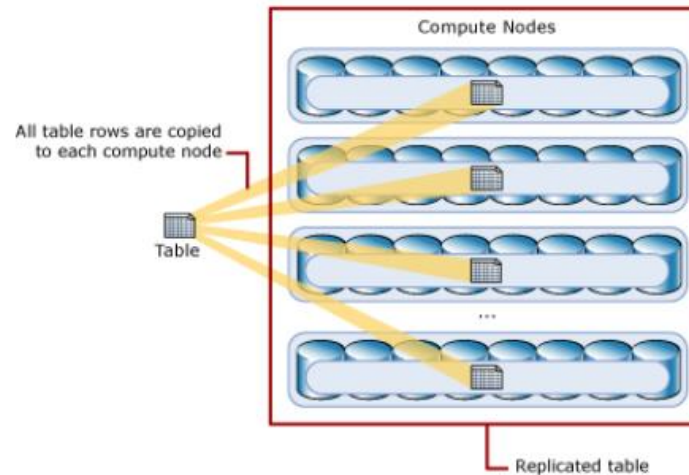


# Azure Synapse Analytics

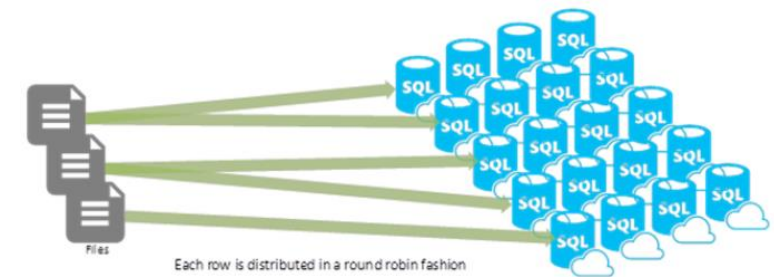
- Hash distributed table



- Replicated Table



- Round Robin distributed Table



- Example

- Dimension to Fact table join

# Big Data

## Streaming

### Azure Stream Analytics



Real-time data stream processing  
from millions of IoT devices

### Azure IoT Hub



Connect, monitor and manage  
billions of IoT assets

### Azure HDInsight & Kafka



Real-time data stream with Kafka

### Spark Streaming with Databricks



Use Spark Streaming with  
Databricks

# Big Data

## Streaming

### Azure Stream Analytics



Real-time data stream processing  
from millions of IoT devices

### Azure IoT Hub



Connect, monitor and manage  
billions of IoT assets

### Azure HDInsight & Kafka



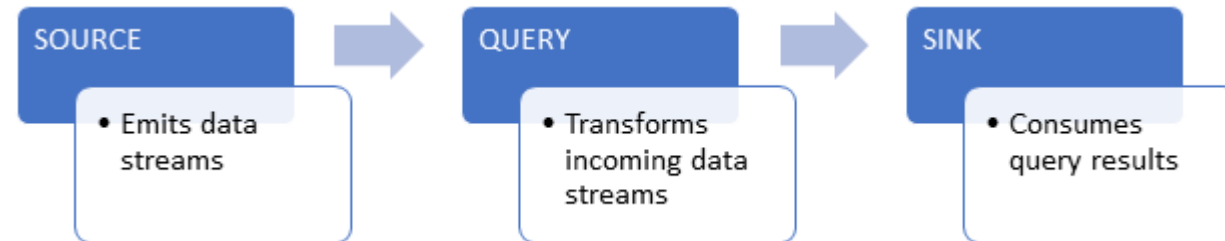
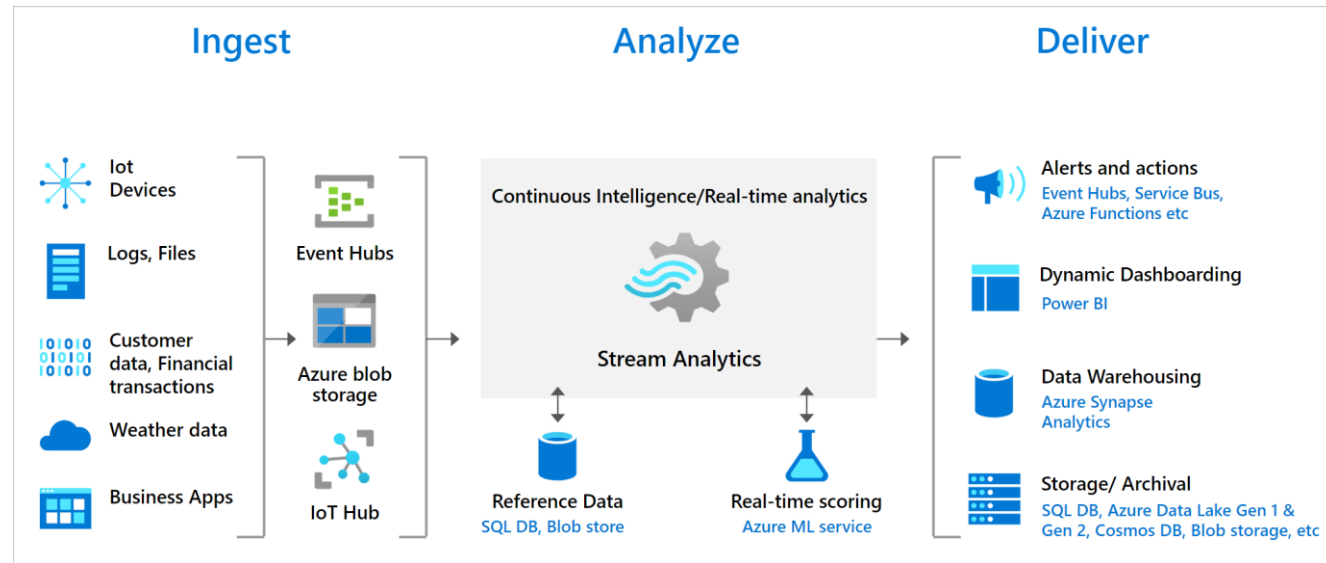
Real-time data stream with Kafka

### Spark Streaming with Databricks



Use Spark Streaming with  
Databricks

# Azure Streaming Analytics



# Azure Streaming Analytics

## UDFs, UDAs, and custom deserializers:

- Azure Stream Analytics supports user-defined functions (UDF) or user-defined aggregates (UDA) in JavaScript for cloud jobs and C# for IoT Edge jobs



## Examples scenarios:

- Analyze real-time telemetry streams from IoT devices
- Web logs/clickstream analytics
- Geospatial analytics for fleet management and driverless vehicles
- Remote monitoring and predictive maintenance of high value assets
- Real-time analytics on Point of Sale data for inventory control and anomaly detection

# Big Data

## Streaming

### Azure Stream Analytics



Real-time data stream processing  
from millions of IoT devices

### Azure IoT Hub



Connect, monitor and manage  
billions of IoT assets

### Azure HDInsight & Kafka



Real-time data stream with Kafka

### Spark Streaming with Databricks

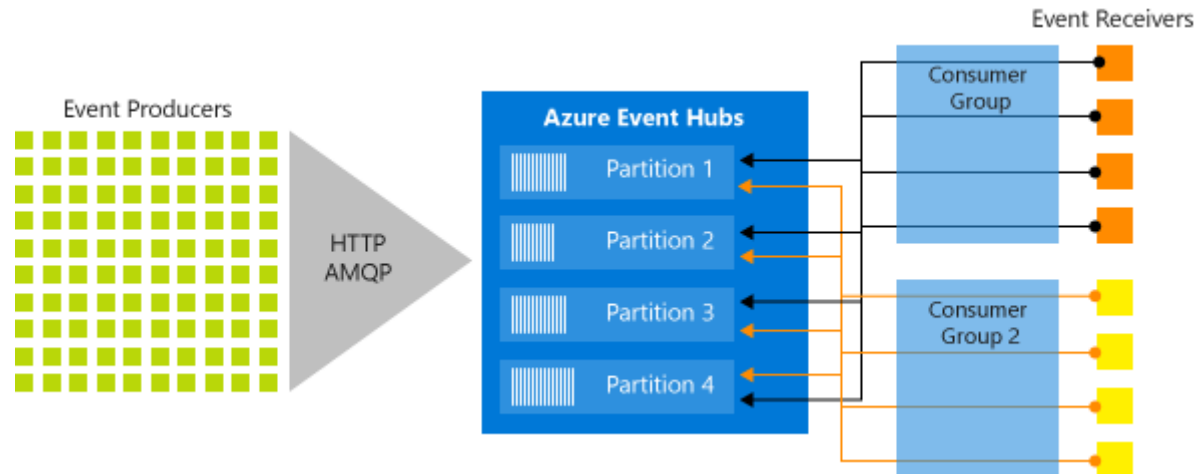


Use Spark Streaming with  
Databricks

# Azure IoT Hub

Azure IoT Hub :

- The cloud gateway that connects IoT devices to gather data and drive business insights and automation.
- The big data streaming service of Azure. It is designed for high throughput data streaming scenarios where customers may send billions of requests per day.
- Bi-directional communication capabilities



# IoT Hub or Event Hubs

IoT Hub was developed to address the unique requirements of connecting IoT devices to the Azure cloud while Event Hubs was designed for big data streaming. Microsoft recommends using Azure IoT Hub to connect IoT devices to Azure.

IoT Capability	IoT Hub standard tier	IoT Hub basic tier	Event Hubs
Device-to-cloud messaging	✓	✓	✓
Protocols: HTTPS, AMQP, AMQP over webSockets	✓	✓	✓
Protocols: MQTT, MQTT over webSockets	✓	✓	
Per-device identity	✓	✓	
File upload from devices	✓	✓	
Device Provisioning Service	✓	✓	
Cloud-to-device messaging	✓		
Device twin and device management	✓		
Device streams (preview)	✓		
IoT Edge	✓		



# Big Data

## Streaming

### Azure Stream Analytics



Real-time data stream processing  
from millions of IoT devices

### Azure IoT Hub



Connect, monitor and manage  
billions of IoT assets

### Azure HDInsight & Kafka



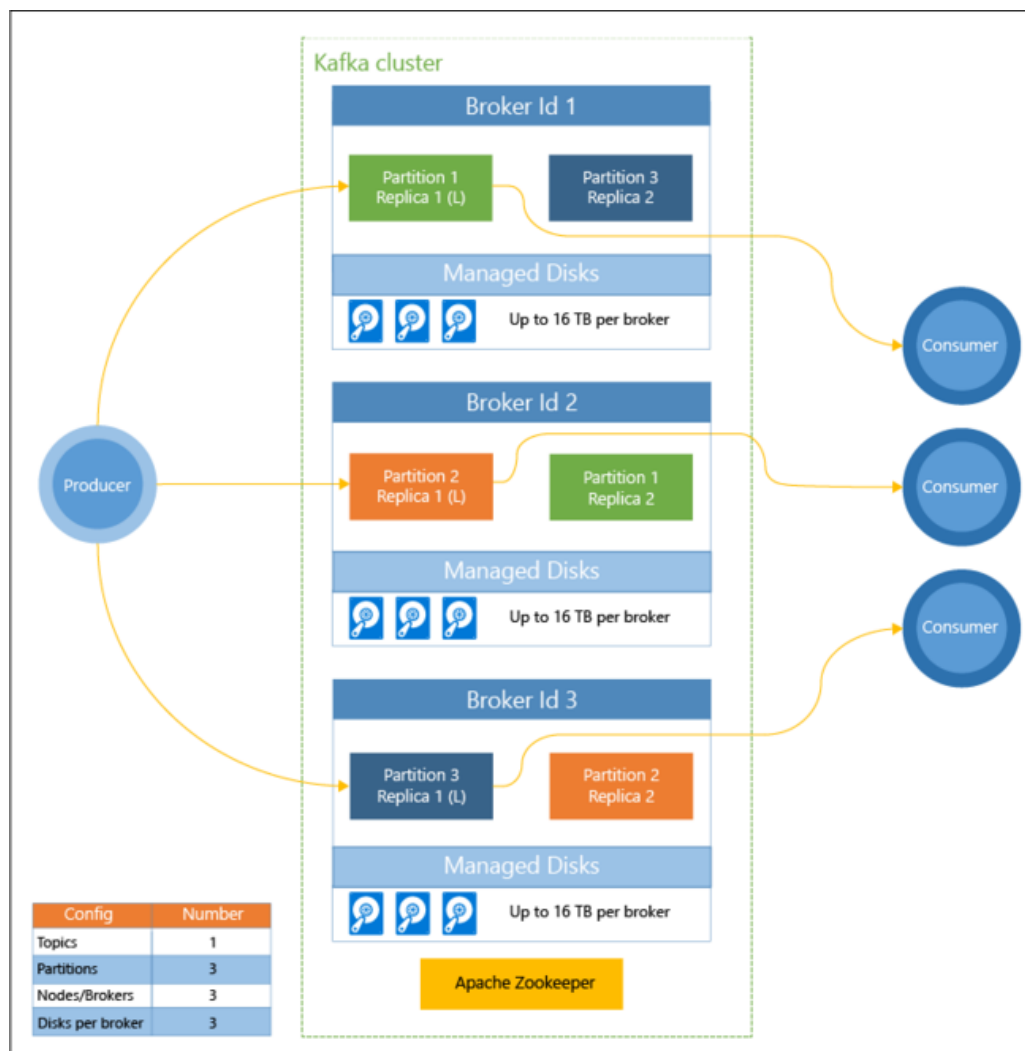
Real-time data stream with Kafka

### Spark Streaming with Databricks



Connect, monitor and manage  
billions of IoT assets

# Apache Kafka on HDInsight architecture



## Basics

Subscription  
Resource group  
Location  
Cluster name  
Cluster type  
Cluster login username  
Secure Shell (SSH) username  
Use cluster login password for SSH

Azure  
(new) myResourceGroup  
East US  
(new) MyKafka2019  
Kafka 1.1.0 (HDI 3.6)  
admin  
sshuser  
Enabled

## Cluster configuration

Head  
Zookeeper  
Disks  
Worker

2 nodes, D3 v2 (4 Cores, 14 GB RAM)  
3 nodes, A4 v2 (4 Cores, 8 GB RAM)  
8 nodes, S30  
4 nodes, D3 v2 (4 Cores, 14 GB RAM)

# Big Data

## Streaming

### Azure Stream Analytics



Real-time data stream processing  
from millions of IoT devices

### Azure IoT Hub



Connect, monitor and manage  
billions of IoT assets

### Azure HDInsight & Kafka



Real-time data stream with Kafka

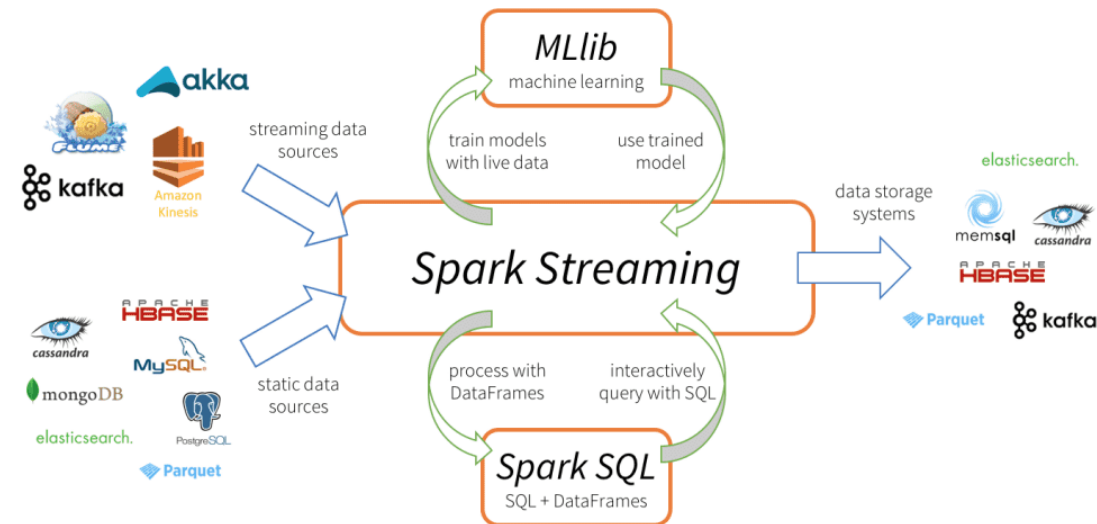
### Spark Streaming with Databricks



Use Spark Streaming with  
Databricks

# Azure Databricks

- Apache Spark Streaming is a scalable fault-tolerant streaming processing system that natively supports both batch and streaming workloads.
- Spark Streaming is an extension of the core Spark API



# Data Tools



# Azure Data Studio

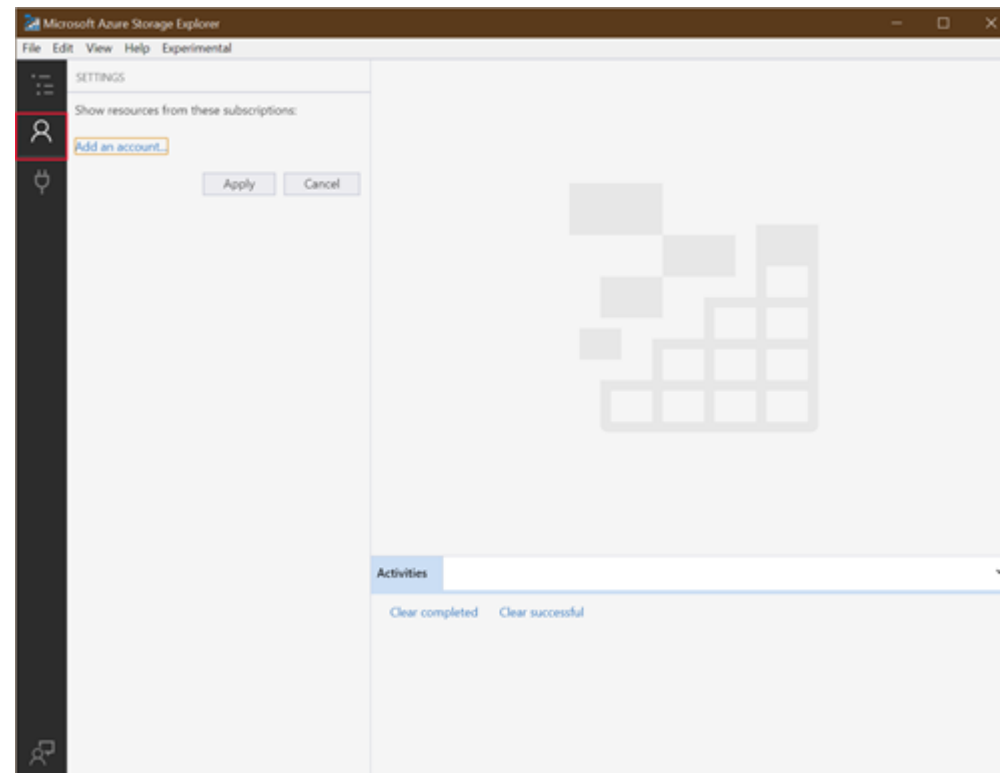
Azure Data Studio is a cross-platform database tool that you can run on Windows, macOS, and Linux. You'll use it to connect to SQL Data Warehouse and Azure SQL Database.

Previously released under the preview name SQL Operations Studio, Azure Data Studio offers a modern editor experience with IntelliSense, code snippets, source control integration, and an integrated terminal. It is engineered with the data platform user in mind, with built in charting of query result sets and customizable dashboards.

# Storage Explorer

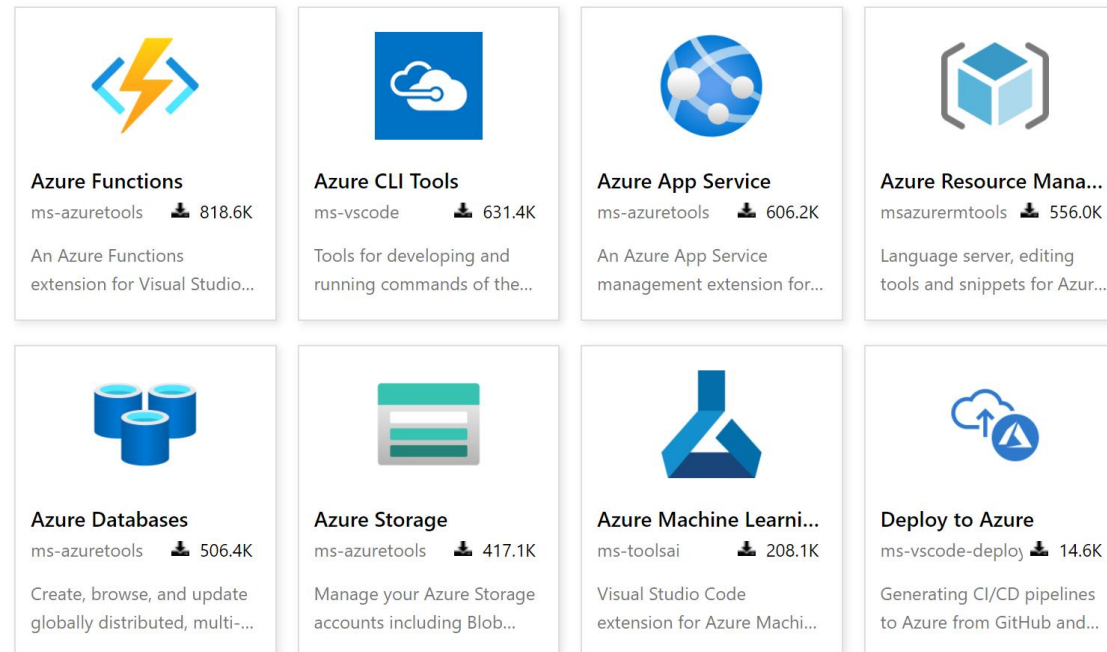
Begin by downloading and installing Storage Explorer. You can use Storage Explorer to do several operations against data in your Azure Storage account and data lake:

- Upload files or folders from your local computer into Azure Storage.
- Download cloud-based data to your local computer.
- Copy or move files and folders around in the storage account.
- Delete data from the storage account.



# Visual Studio Code

Visual Studio Code is a lightweight source code editor which runs on your desktop and is available for Windows, macOS and Linux. It comes with built-in support for JavaScript, TypeScript and Node.js and has a rich ecosystem of extensions for other languages (such as C++, C#, Java, Python, PHP, Go) and runtimes (such as .NET and Unity).





# Data Migration Tools



# Summary

Scenario	Some recommended solutions
Disaster Recovery	<a href="#">Azure geo-redundant backups</a>
Read Scale	<a href="#">Use read-only replicas to load balance read-only query workloads (preview)</a>
ETL (OLTP to OLAP)	<a href="#">Azure Data Factory</a> or <a href="#">SQL Server Integration Services</a> or <a href="#">Databricks</a>
Migration from on-premises SQL Server to Azure SQL Database	<a href="#">Azure Database Migration Service</a>
Kept up-to-date across several Azure SQL databases or SQL Server database	<a href="#">Azure SQL Data Sync</a>
Detecting compatibility issues that can impact database functionality in your new version of SQL Server or Azure SQL Database	<a href="#">Data Migration Assistant</a> (DMA)

# Resources



# Azure charts

AI + Machine Learning	Analytics	Compute	Databases	Development	Identity + Security	IoT	Integration	Management + Governance	Media	Migration	Networking	Storage
Bot Service	Analysis Services	App Service	Blockchain Service	App Configuration	Azure Active Directory	Azure Maps	API Management	Automation	Azure CDN	Azure Migrate	Application Gateway	Azure vFXT
Cognitive Search	Data Catalog	App Service (Linux)	Cosmos DB	Azure DevOps	Azure AD B2C	Azure Sphere	Azure API for FHIR	Azure Advisor	Media Services	Data Box	Azure Bastion	Azure NetApp Files
Cognitive Services	Data Explorer	Azure Batch	Database for MariaDB	DevTest Labs	Azure AD DS	Digital Twins	Event Grid	Azure Arc		DB Migration Service	Azure DNS	Azure Storage
Machine Learning	Data Factory	Azure Functions	Database for MySQL	Lab Services	Azure Key Vault	IoT Central	Logic Apps	Azure Backup		Site Recovery	Azure Firewall	Data Lake Storage
Microsoft Genomics	Data Lake Analytics	Azure VMware Solutions	Database for PostgreSQL	SignalR Service	Azure Sentinel	IoT Edge	Notification Hubs	Azure Blueprints			Azure Front Door	Data Share
Open Datasets	Databricks	Cloud Services	Redis Cache	Visual Studio App Center	DDoS Protection	IoT Hub	Service Bus	Azure Lighthouse			ExpressRoute	Managed Disks
	Event Hubs	Container Instances	SQL Database		Dedicated HSM	Spatial Anchors		Azure Monitor			Load Balancer	StorSimple
	HDInsight	Container Registry	SQL Server Stretch DB		Information Protection	Time Series Insights		Azure Policy			Network Watcher	
	Power BI Embedded	CycleCloud			Security Center			Azure Portal			Private Link	
	Stream Analytics	Dedicated Host						Cloud Shell			Traffic Manager	

<https://azurecharts.com/>

# Sources

Just few sources in Microsoft Learn:

- [Azure for the Data Engineer](#)
- [Store data in Azure](#)
- [Work with relational data in Azure](#)
- [Large Scale Data Processing with Azure Data Lake Storage Gen2](#)
- [Implement a Data Streaming Solution with Azure Streaming Analytics](#)
- [Implement a Data Warehouse with Azure SQL Data Warehouse](#)

# Fill the form



Your turn !

[https://forms.office.com/Pages/ResponsePage.aspx?id=M3s0akU8nUyLePs4Zpn6Tp\\_2uFsS8cJJsHCSwweCY5JUNVIMMIIQNU4yRUVVWjFEOU5GVVc2SVU3Si4u](https://forms.office.com/Pages/ResponsePage.aspx?id=M3s0akU8nUyLePs4Zpn6Tp_2uFsS8cJJsHCSwweCY5JUNVIMMIIQNU4yRUVVWjFEOU5GVVc2SVU3Si4u)



## Next Session: Azure Databricks

Azure Databricks



Fast, easy, and collaborative  
Apache Spark-based analytics  
platform

# Thank you

