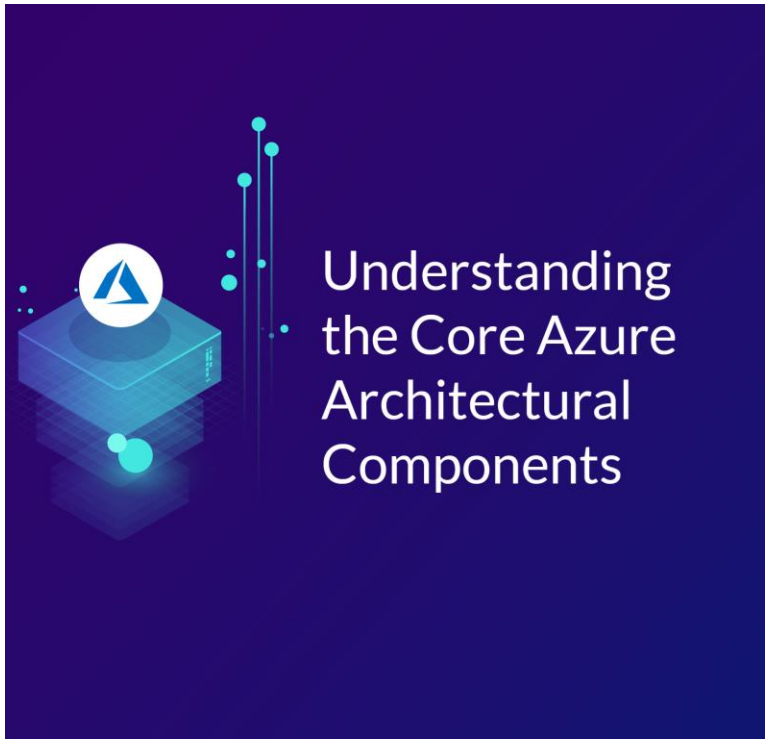
Azure Cloud Services For Big Data

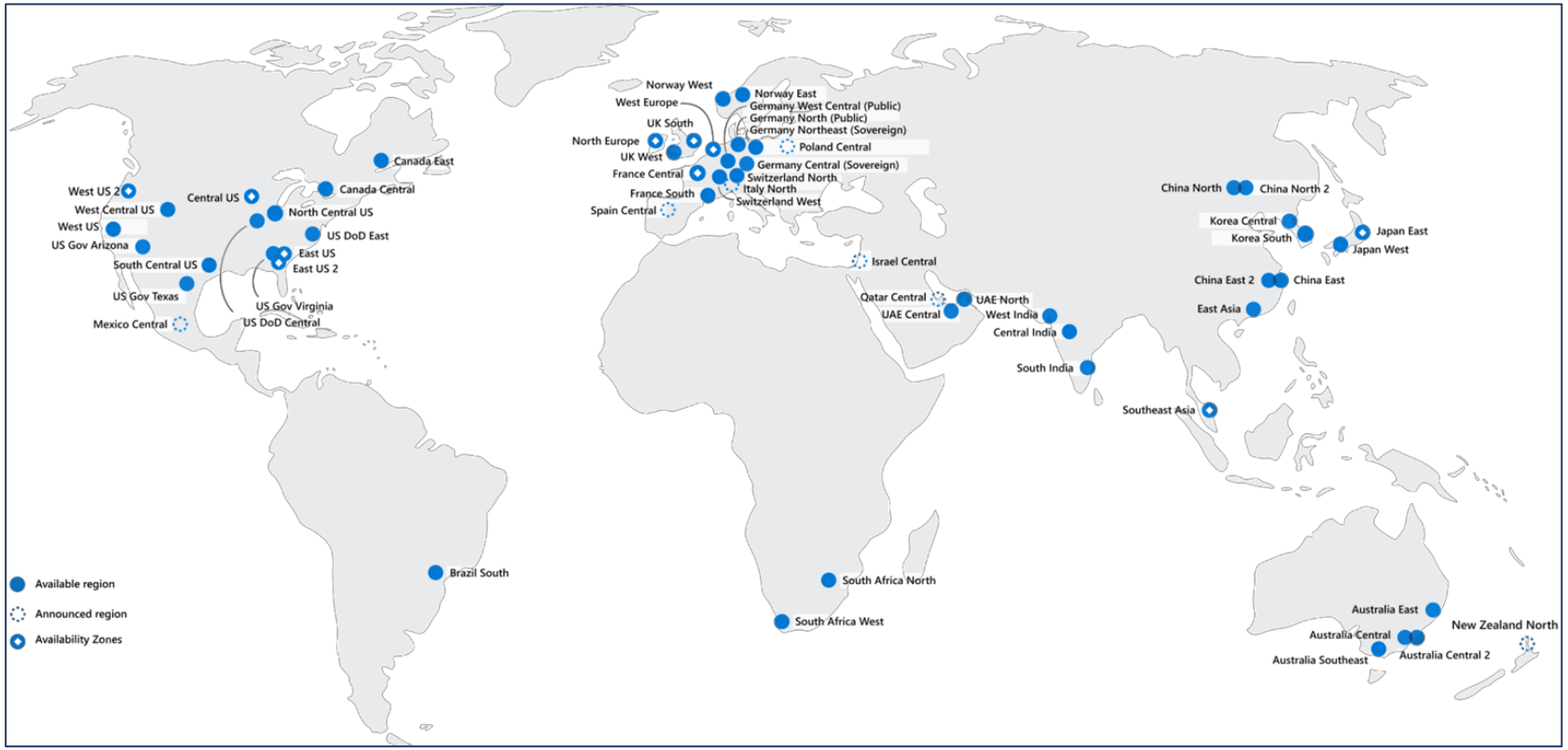
Core Azure architectural components

- Regions and Region Pairs
- Availability Zones
- Azure resources
- Resource Groups
- Azure Resource Manager
- Subscriptions
- Azure Management Groups



Regions

Azure offers more global regions than any other cloud provider with 60+ regions representing over 140 countries



- Regions are made up of one or more data centers in close proximity.
- Provide flexibility and scale to reduce customer latency.
- Preserve data residency with a comprehensive compliance offering.

Region Pairs

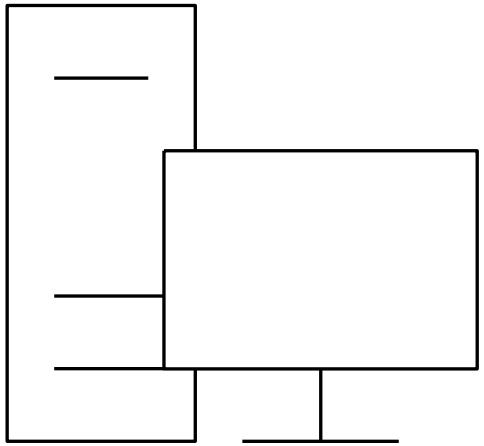
- At least 300 miles of separation between region pairs.
- Automatic replication for some services.
- Prioritized region recovery in the event of outage.
- Updates are rollout sequentially to minimize downtime.

Web Link: https://aka.ms/PairedRegions

Region	Region
North Central US	South Central US
East US	West US
West US 2	West Central US
US East 2	Central US
Canada Central	Canada East
North Europe	West Europe
UK West	UK South
Germany Central	Germany Northeast
South East Asia	East Asia
East China	North China
Japan East	Japan West
Australia Southeast	Australia East
India South	India Central
Brazil South (Primary)	South Central US

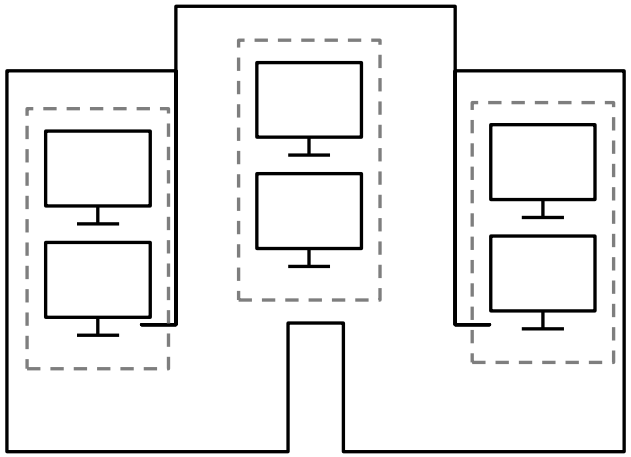
Availability Options

VM SLA
99.9% with Premium Storage



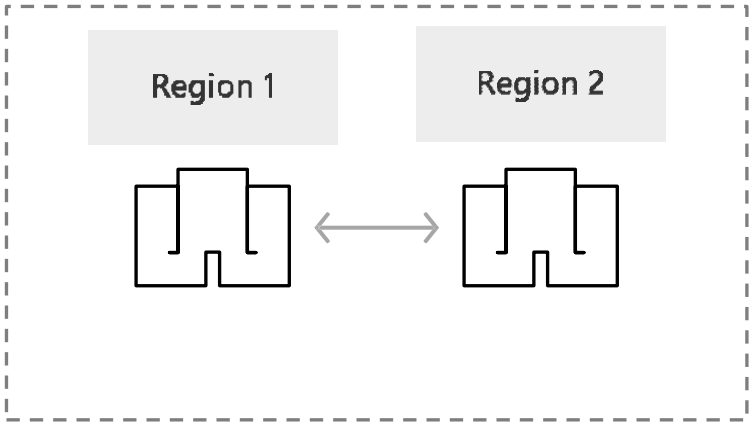
SINGLE VM
Easier lift and shift

VM SLA
99.99%



AVAILABILITY ZONES
Protection from entire datacenter failures

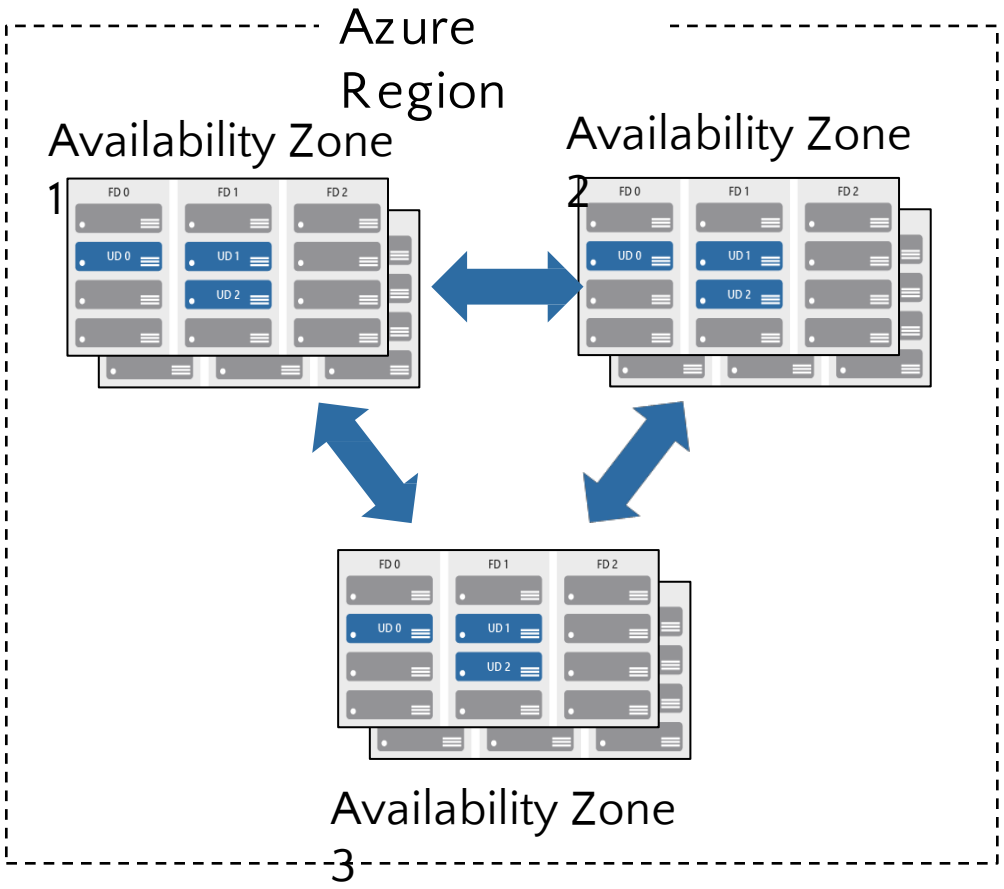
MULTI REGION DISASTER RECOVERY



REGION PAIRS
Regional protection within Data Residency Boundaries

Availability zones

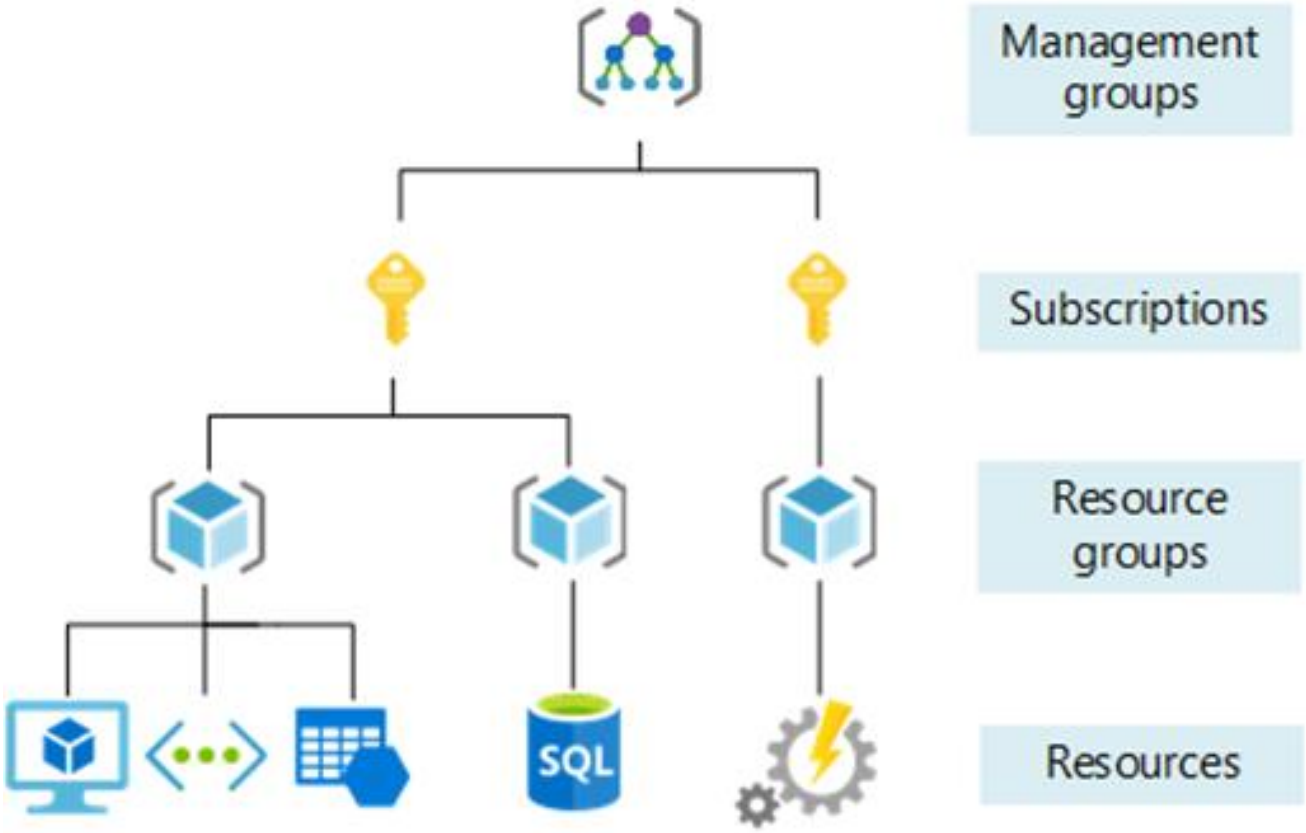
- Provide protection against downtime due to datacenter failure.
- Physically separate data centers within the same region.
- Each datacenter is equipped with independent power, cooling, and networking.
- Connected through private fiber-optic networks.



Management levels and hierarchy

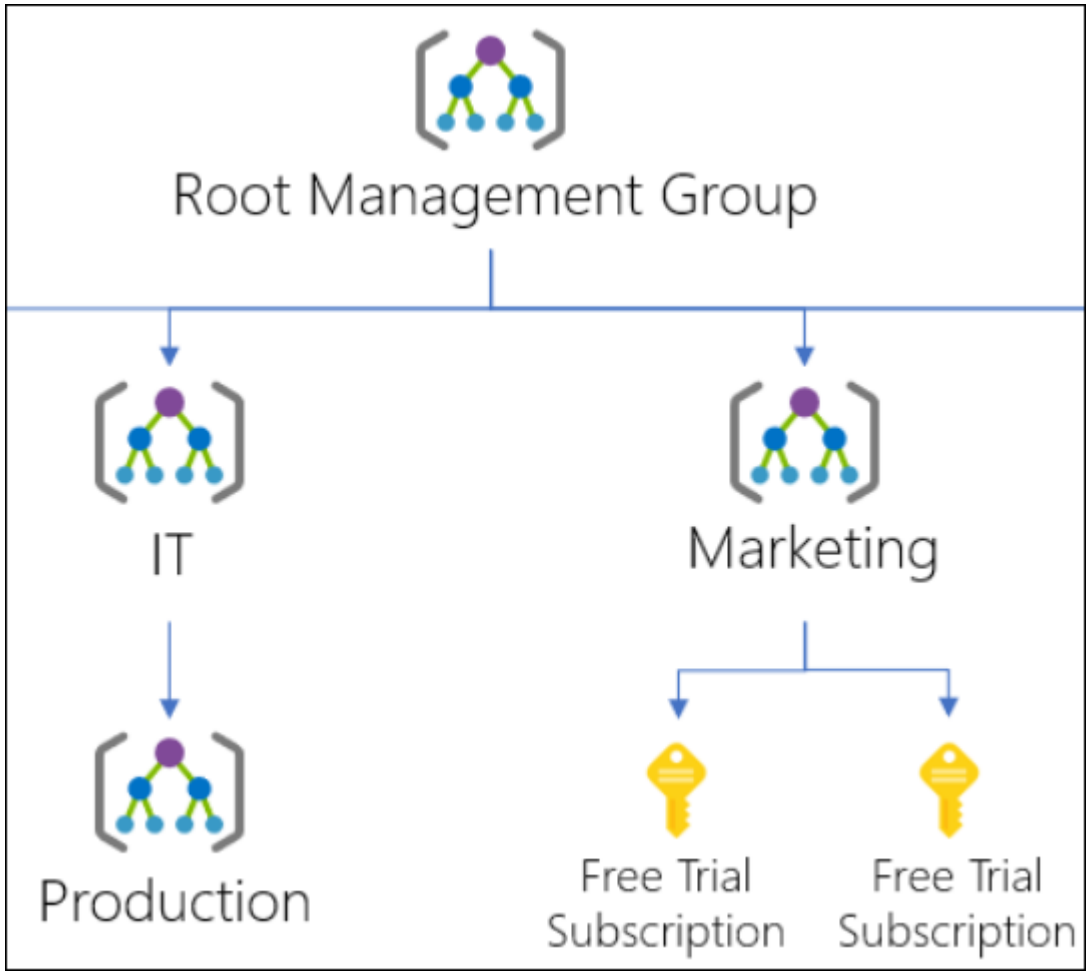
Azure provides four levels of management:

- Management groups
- Subscriptions
- Resource groups
- Resources



Management Groups

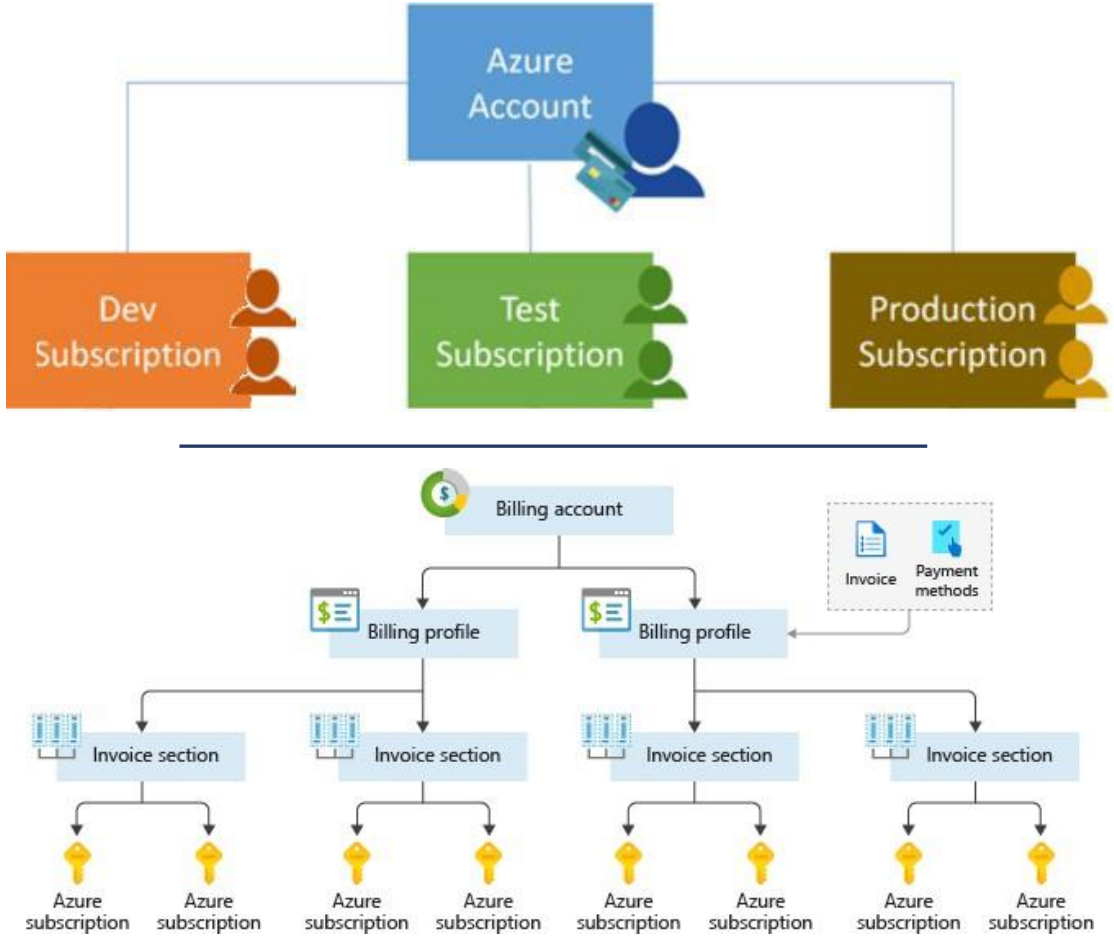
- Management groups can include multiple Azure subscriptions.
- Subscriptions inherit conditions applied to the management group.
- 10,000 management groups can be supported in a single directory.
- A management group tree can support
  up to six levels of depth.



Azure Subscriptions

An Azure subscription provides you with authenticated and authorized access to Azure accounts.

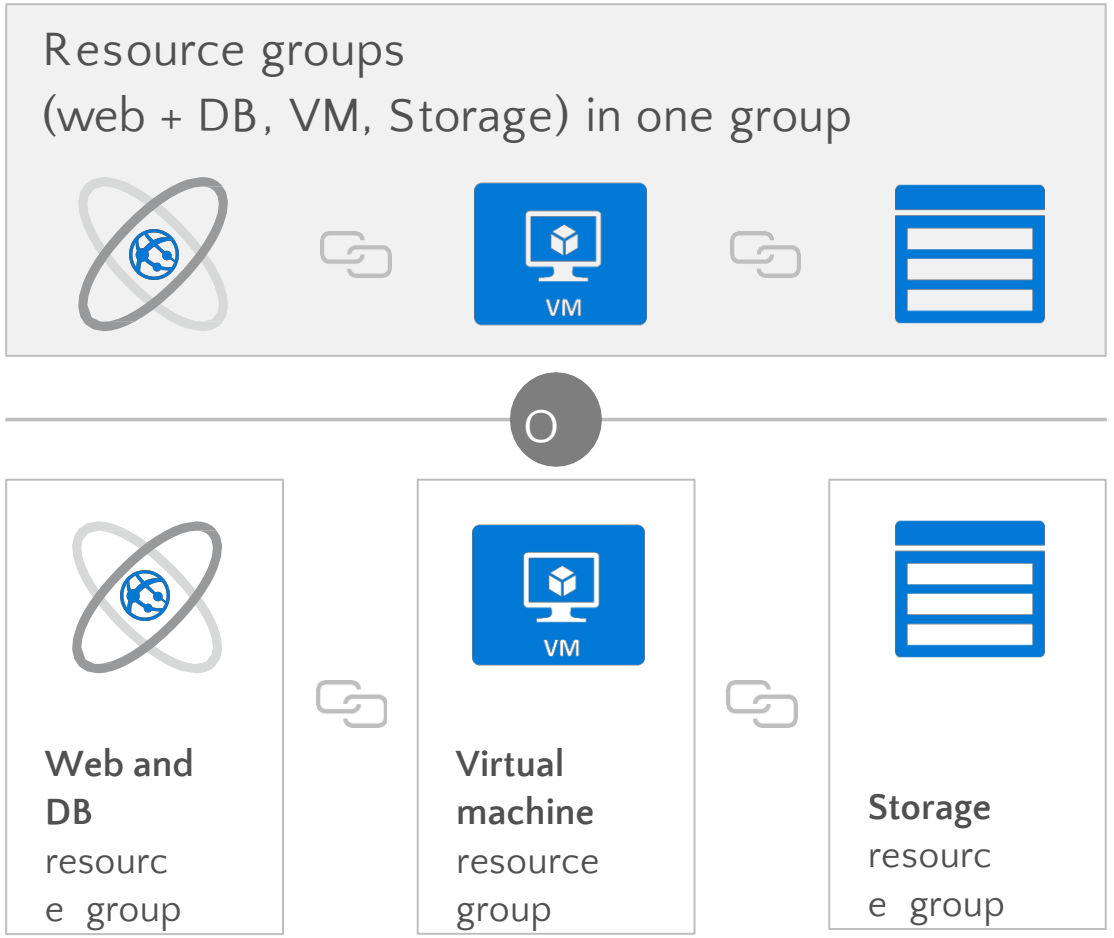
- **Billing boundary:** generate separate billing reports and invoices for each subscription.
- **Access control boundary:** manage and control access to the resources that users can provision with specific subscriptions.



Resource groups

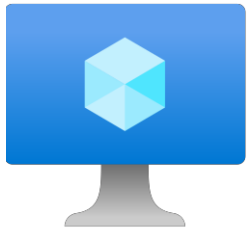
A **resource group** is a container to manage
and aggregate resources in a single unit.

- Resources can exist in only one resource group.
- Resources can exist in different regions.
- Resources can be moved to different resource groups.
- Applications can utilize multiple resource groups.



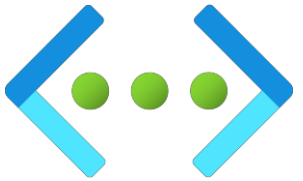
Azure Resources

Azure **resources** are components like storage, virtual machines, and networks that are available to build cloud solutions.



Virtual Machines

Storage Accounts



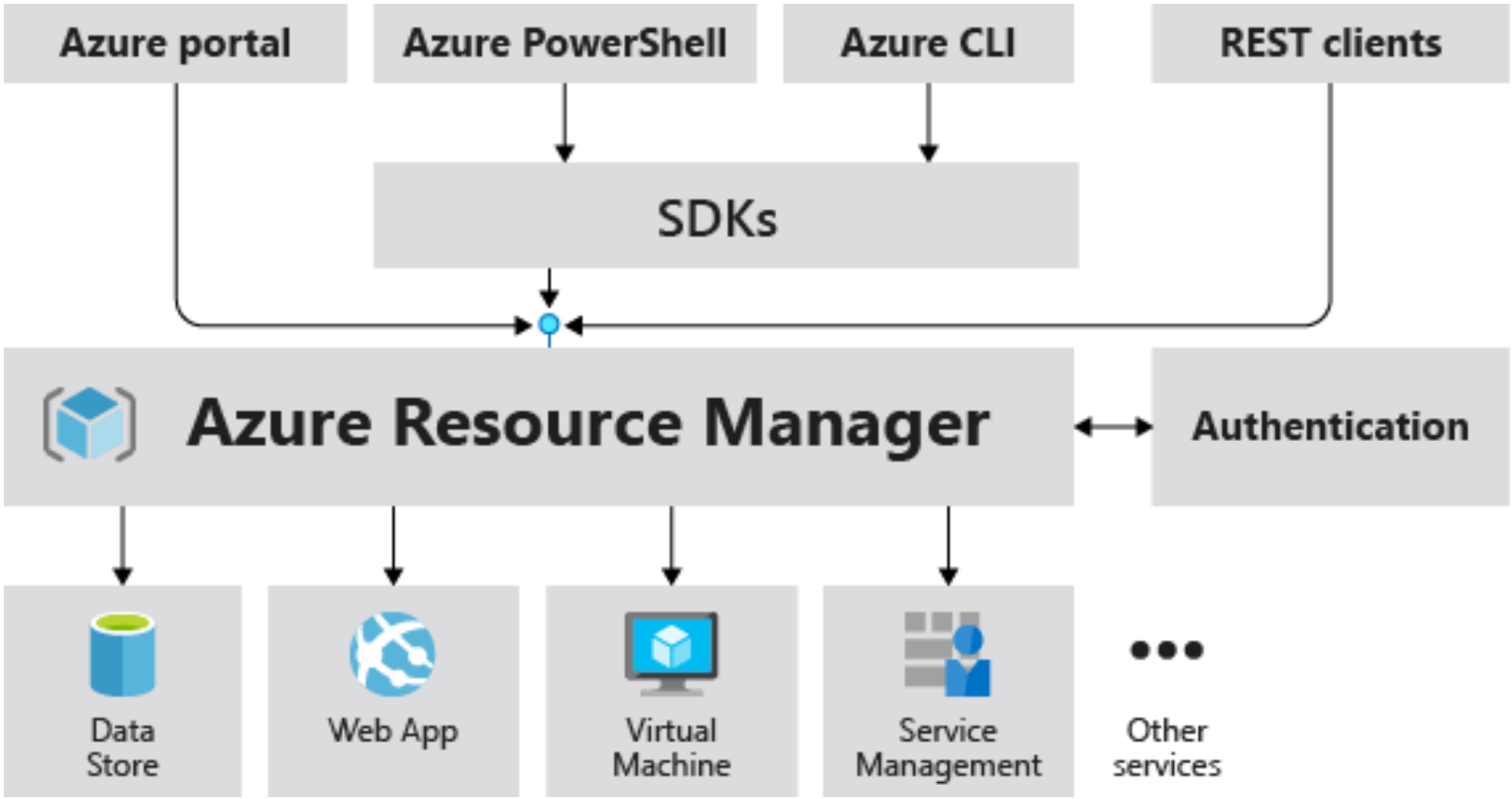
Virtual Networks

App Services

SQL Databases

Functions

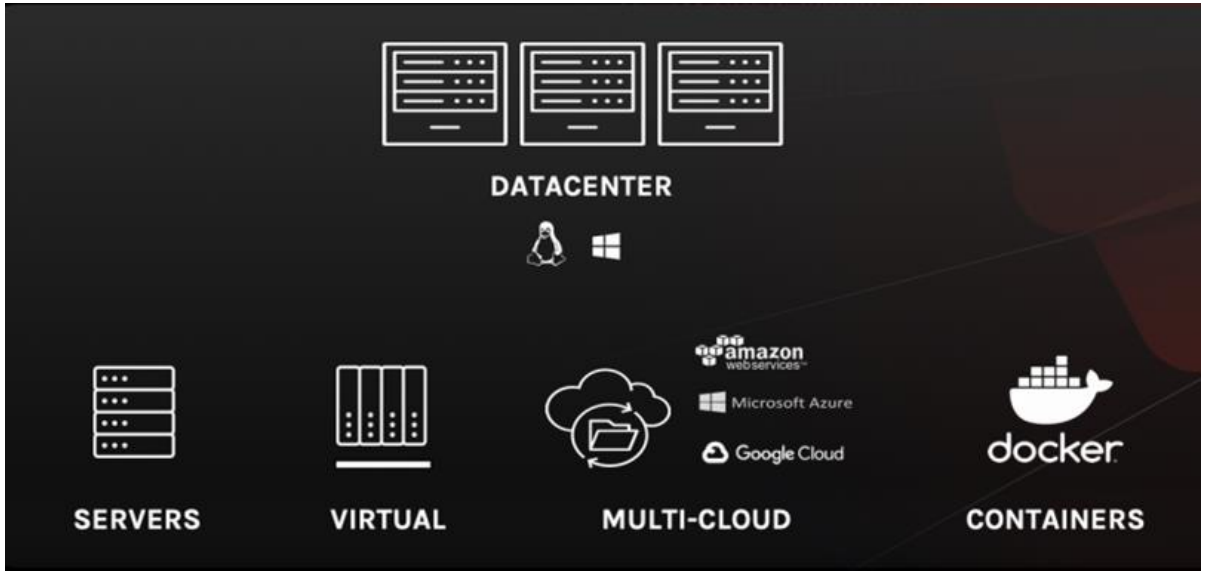
Azure Resource Manager



The **Azure Resource Manager (ARM)** provides a management layer that enables you to create, update, and delete resources in your Azure subscription.

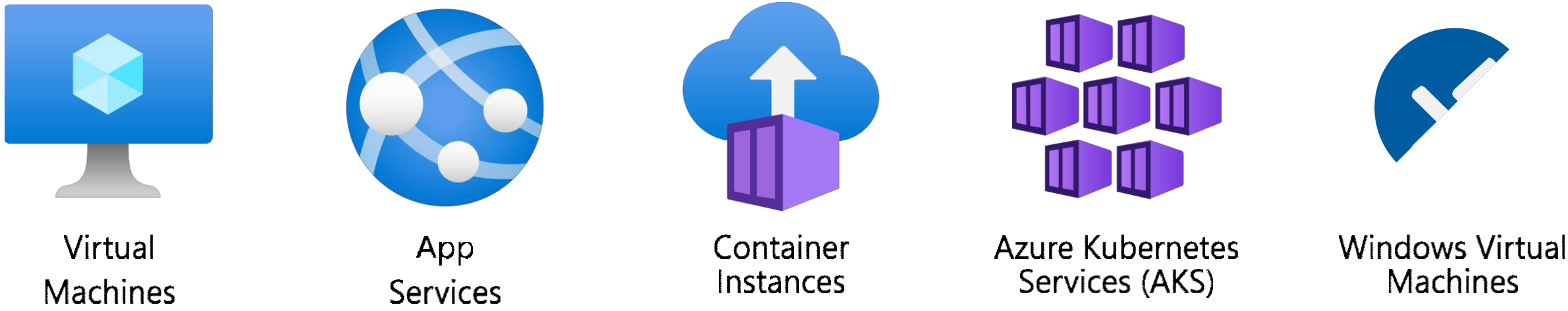
Core Azure Workloads

- Virtual Machines, Azure App Services, Azure Container Instances (ACI), Azure Kubernetes Service (AKS), and Windows Virtual Desktop
- Virtual Networks, VPN Gateway, Virtual Network peering, and ExpressRoute
- Container (Blob) Storage, Disk Storage, File Storage, and storage tiers
- Cosmos DB, Azure SQL Database, Azure Database for MySQL, Azure Database for PostgreSQL, and SQL Managed Instance



Azure compute services

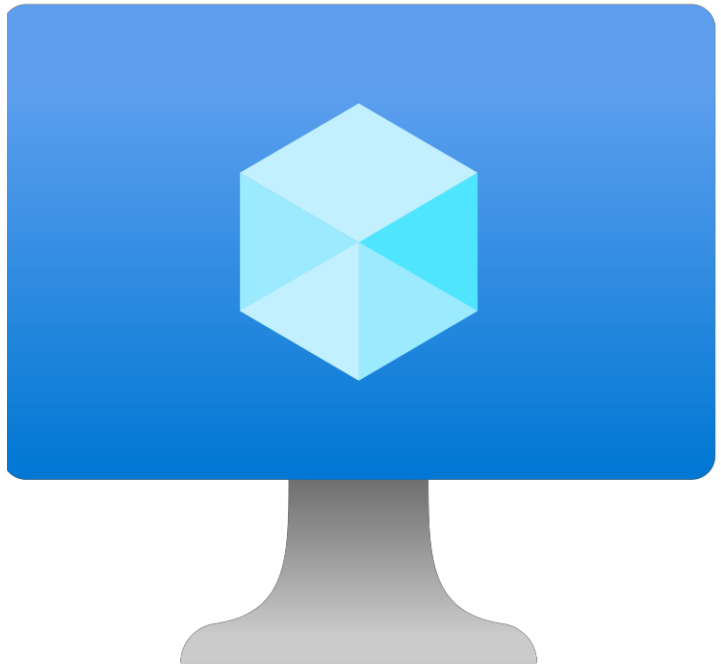
Azure **compute** is an on-demand computing service that provides computing resources such as disks, processors, memory, networking, and operating systems.



Azure virtual machines

Azure **Virtual Machines (VM)** are software emulations of physical computers.

- Includes virtual processor, memory, storage, and networking.
- IaaS offering that provides total control and customization.



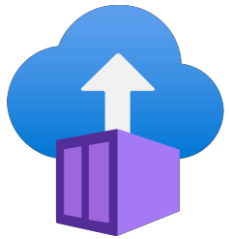
Azure App Services

Azure **App Services** is a fully managed platform to build, deploy, and scale web apps and APIs quickly.

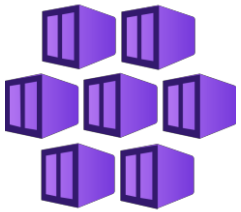
- Works with .Net, .NetC Core, Node.js, Java, Phython, or php.
- PaaS offering with enterprise-grade performance, security, and compliance requirements.

Azure Container Services

Azure **Containers** are a light-weight, virtualized environment that does not require operating system management, and can respond to changes on demand.

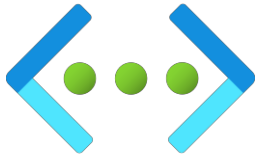


Azure Container Instances: a PaaS offering that runs a container in Azure without the need to manage a virtual machine or additional services.



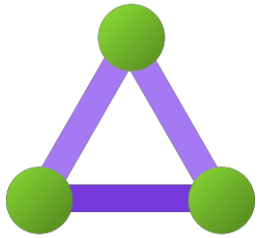
Azure Kubernetes Service: an orchestration service for containers with distributed architectures and large volumes of containers.

Azure networking services



Azure Virtual Network (VNet) enables Azure resources to communicate with each other, the internet, and on premises networks.

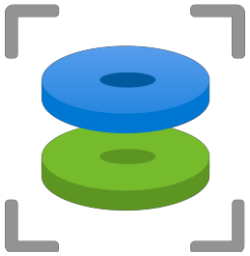
Virtual Private Network Gateway (VPN) is used to send encrypted traffic between an Azure virtual network and an on premises location over the public internet.



Azure Express Route extends on premises networks into Azure over a private connection that is facilitated by a connectivity provider.

Azure storage services

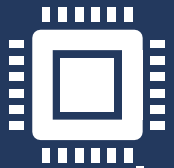
Container storage (blob) is optimized for storing massive amounts of unstructured data, such as text or binary data.



Disk storage provides disks for virtual machines, applications, and other services to access and use.

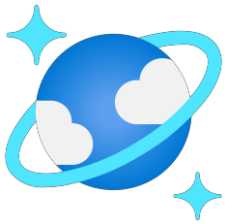
Azure Files sets up a highly available network file shares that can be accessed by using the standard Server Message Block (SMB) protocol.

Azure storage access tiers

 Hot	 Cool	 Archive
Optimized for storing data that is accessed frequently.	Optimized for storing data that is infrequently accessed and stored for at least 30 days.	Optimized for storing data that is rarely accessed and stored for at least 180 days with flexible latency requirements.

You can switch between these access tiers at any time.

Azure database services



Azure Cosmos Database is a globally distributed database service that elastically and independently scales throughput and storage.

Azure SQL Database is a relational database as a service (DaaS) based on the latest stable version of the Microsoft SQL Server database engine.

Azure Database for MySQL is a fully managed MySQL database service for app developers.

Azure Database for PostgreSQL is a relational database service based on the open source Postgres database engine.

Azure SQL Managed Instance

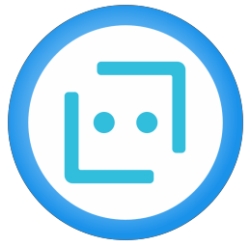
Azure SQL Managed Instance allows existing SQL Server customers to lift and shift their on-premises applications to the cloud with minimal application and database changes.

- Fully managed and evergreen platform as a service.
- Preserves all PaaS capabilities (automatic patching and version updates, automated backups, and high availability)
- Exchange existing licenses for discounted rates on SQL
  Managed Instance using the Azure Hybrid Benefit

Artificial Intelligence & Machine Learning

Azure Machine Learning: cloud based to develop, train, and deploy machine learning models.

Cognitive Services: quickly enable apps to see, hear, speak, understand, and interpret a user's needs.



Azure Bot Service: develop intelligent, enterprise grade bots.

Big data and analytics

Azure Synapse Analytics

A cloud based Enterprise Data Warehouse.

Azure HDInsight

A fully managed, open source analytics service for enterprises.

Azure Databricks

Apache Spark based analytics service.

Azure Data Factory

- ETL service for scale-out serverless data integration and data transformation
- Code free UI
- Monitoring and Management

Code-Free ETL as a service

Ingest

- Multi-cloud and on-premise hybrid copy data
- 100+ native connectors
- Serverless and auto-scale
- Use wizard for quick copy jobs

Control Flow

- Design code-free data pipelines
- Generate pipelines via SDK
- Utilize workflow constructs: loops, branches, conditional execution, variables, parameters, …

Data Flow

- Code-free data transformations that execute in Spark
- Scale-out with Azure Integration Runtimes
- Generate data flows via SDK
- Designers for data engineers and data analysts

Schedule

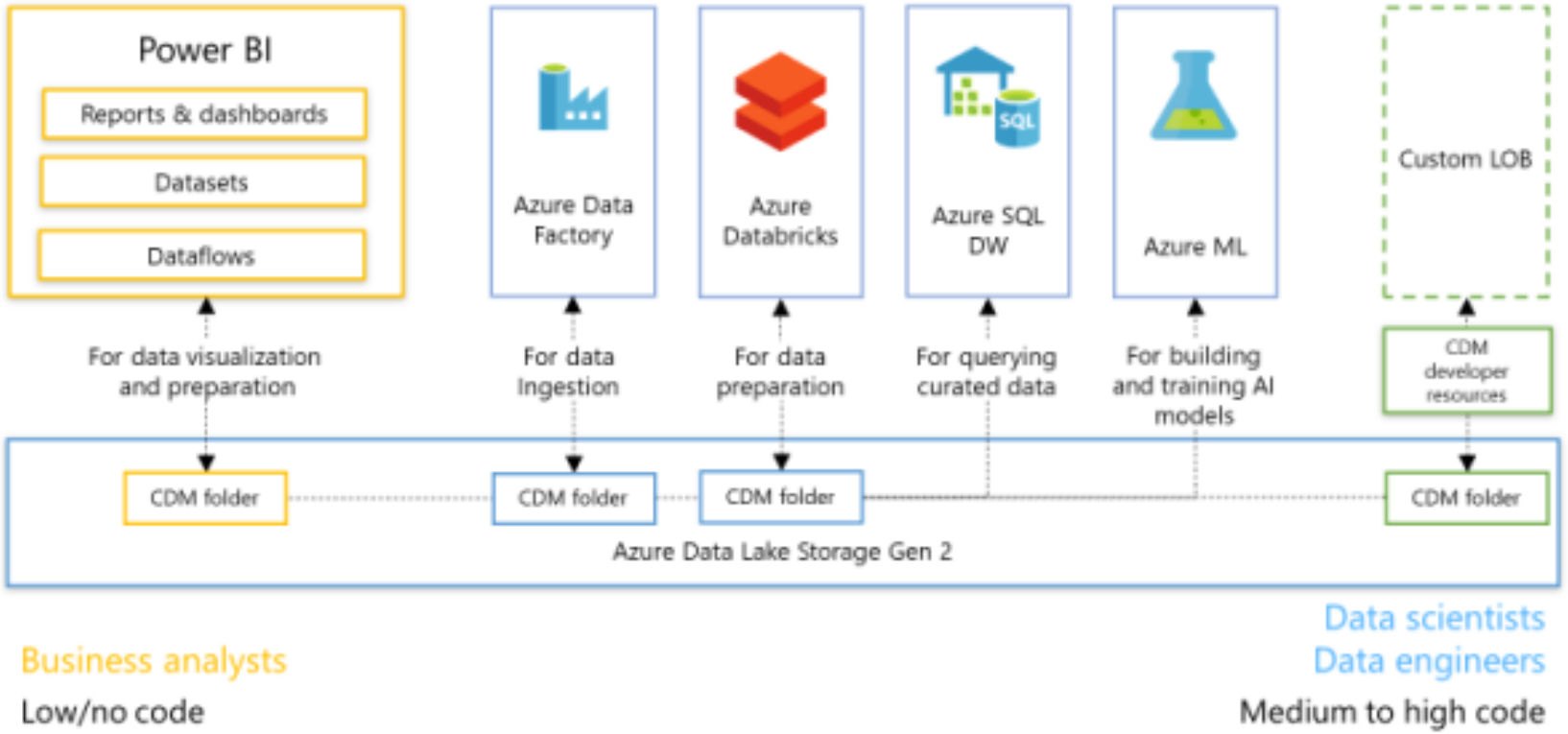
- Build and maintain operational schedules for your data pipelines
- Wall clock, event-based, tumbling windows, chained

Monitor

- View active executions and pipeline history
- Detail activity and data flow executions
- Establish alerts and notifications

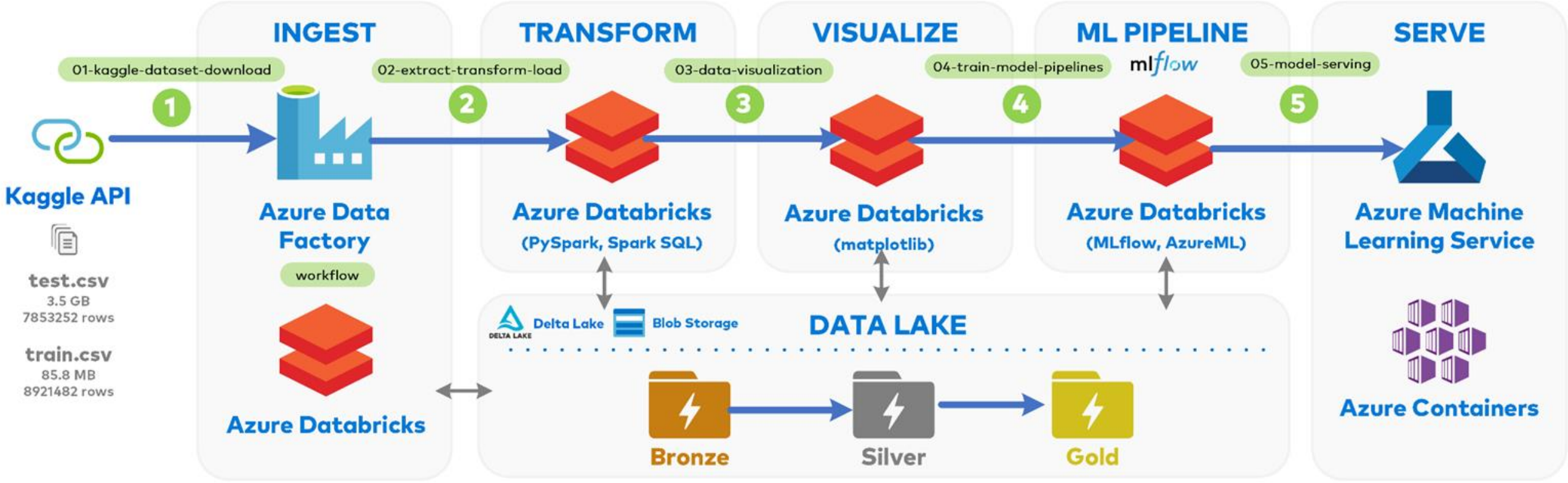
Azure Data Lake

- Provides different types of storages
- Scalable
- Built on YARN



Azure Databricks

- Powerful ETL tool
- Based on Apache Spark
- Provides mlflow and machine learning capabilities



Azure Synapse

- Powerfull data warehousing tool
- Limitless data analytics
- Data integraion