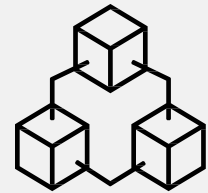


Modern Data Architectures

Cloud Databases

**HO
GENT**

Cloud Models

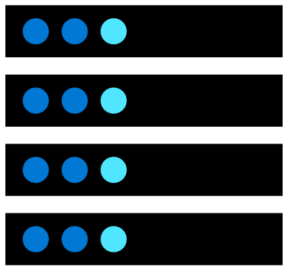


Cloud Models - Objective Domain

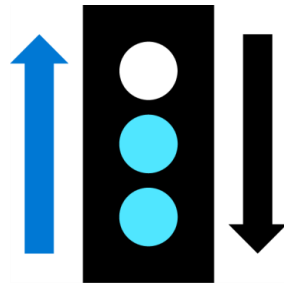
- Define cloud computing
- Describe Public cloud
- Describe Private cloud
- Describe Hybrid cloud
- Compare and contrast the three different cloud models

What is cloud computing?

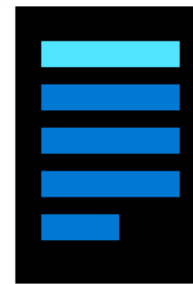
Cloud Computing is the delivery of computing services over the internet, enabling faster innovation, flexible resources, and economies of scale.



Compute



Networking



Storage



Analytics

Public cloud

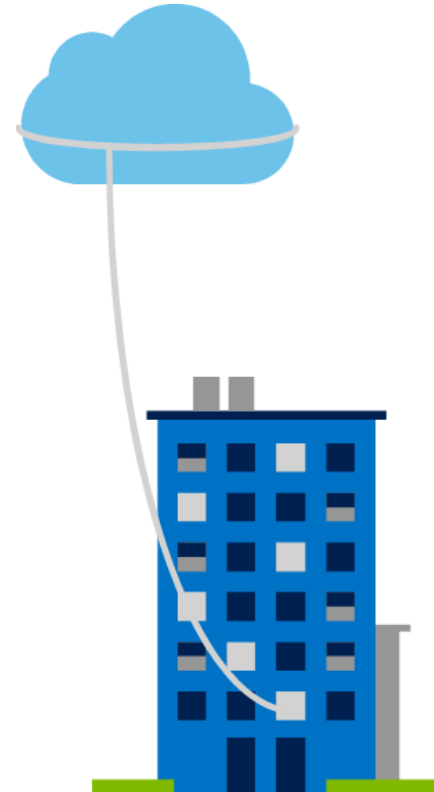
- Owned by cloud services or hosting provider.
- Provides resources and services to multiple organizations and users.
- Accessed via secure network connection (typically over the internet).



Private cloud

- Organizations create a cloud environment in their datacenter.
- Organization is responsible for operating the services they provide.
- Does not provide access to users outside of the organization.

Example: VIC (Virtual IT Company) HOGENT



Hybrid cloud



Combines **Public** and **Private** clouds to allow applications to run in the most appropriate location.

Cloud model comparison

Public Cloud

- No capital expenditures to scale up.
- Applications can be quickly provisioned and deprovisioned.
- Organizations pay only for what they use.

Private Cloud

- Hardware must be purchased for start-up and maintenance.
- Organizations have complete control over resources and security.
- Organizations are responsible for hardware maintenance and updates.

Hybrid Cloud

- Provides the most flexibility.
- Organizations determine where to run their applications.
- Organizations control security, compliance, or legal requirements.

Cloud benefits and considerations



Cloud Benefits - Objective Domain

- Identify the benefits of cloud computing such as High Availability, Scalability, Elasticity, Agility, and Disaster Recovery.
- Identify the differences between Capital Expenditure (CapEx) and Operational Expenditure (OpEx).
- Describe the consumption-based model.

Cloud Benefits

High availability

Fault tolerance

Scalability

Elasticity

Global reach

Customer latency capabilities

Agility

Predictive cost considerations

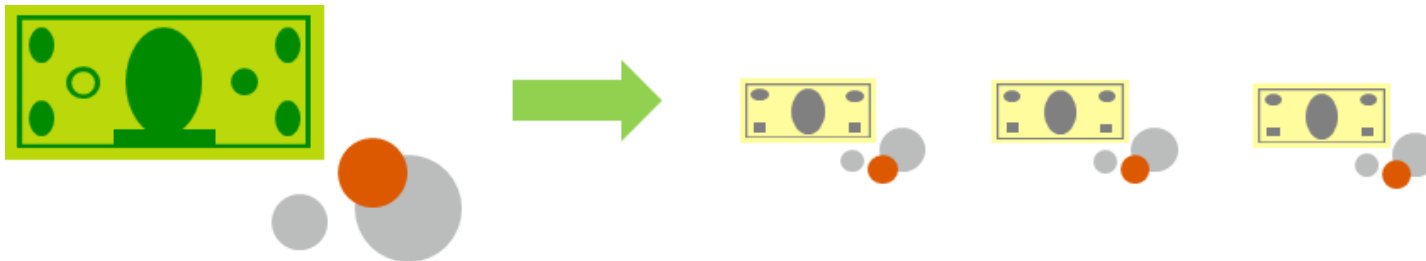
Compare CapEx vs. OpEx

Capital Expenditure (CapEx)

- The up-front spending of money on physical infrastructure.
- Costs from CapEx have a value that reduces over time.

Operational Expenditure (OpEx)

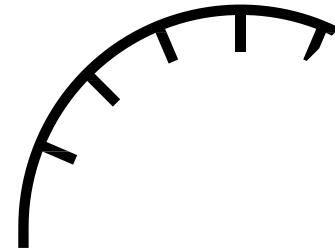
- The spending and billing of services or products as needed.
- Expenses are deducted in the same year.



Consumption-based model

Cloud service providers operate on a consumption-based model, which means that end users only pay for the resources that they use. Whatever they use is what they pay for.

- Better cost prediction
- Prices for individual resources and services are provided
- Billing is based on actual usage



Cloud services

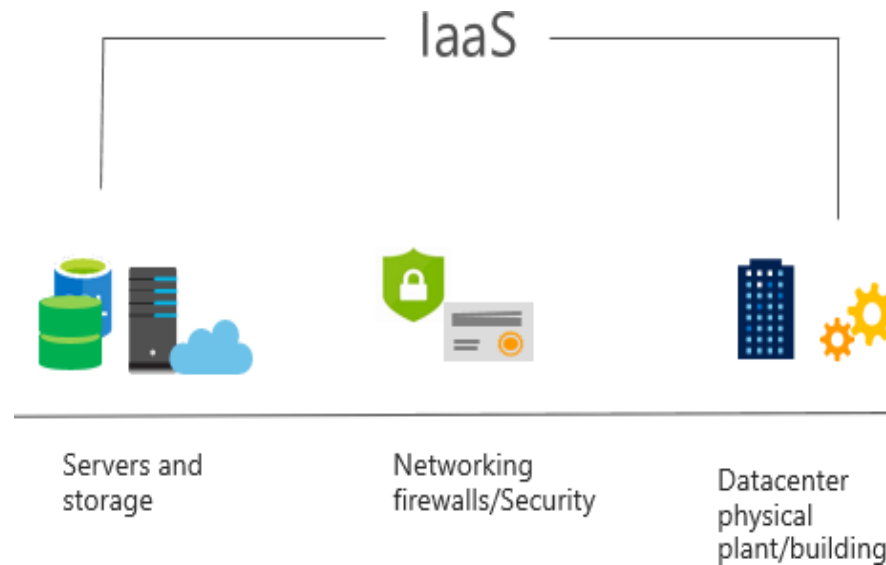


Cloud Services - Objective Domain

- Describe Infrastructure-as-a-Service (IaaS)
- Describe Platform-as-a-Service (PaaS)
- Describe Software-as-a-Service (SaaS)
- Identify a service type based on a use case
- Describe the shared responsibility model
- Describe serverless computing

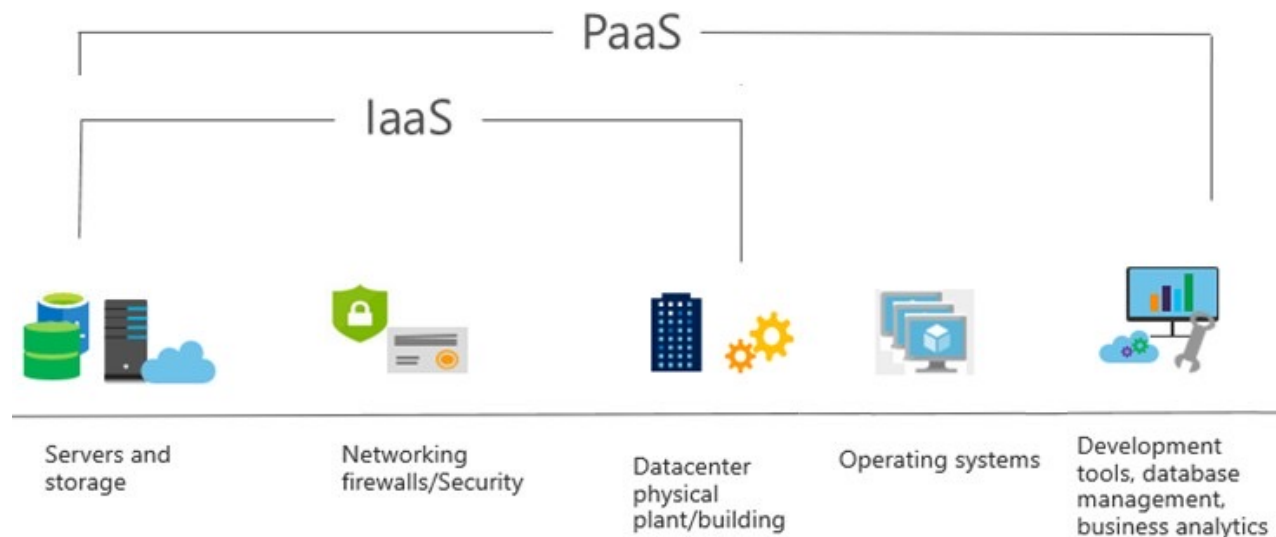
Infrastructure as a Service (IaaS)

Build pay-as-you-go IT infrastructure by renting servers, virtual machines, storage, networks, and operating systems from a cloud provider.



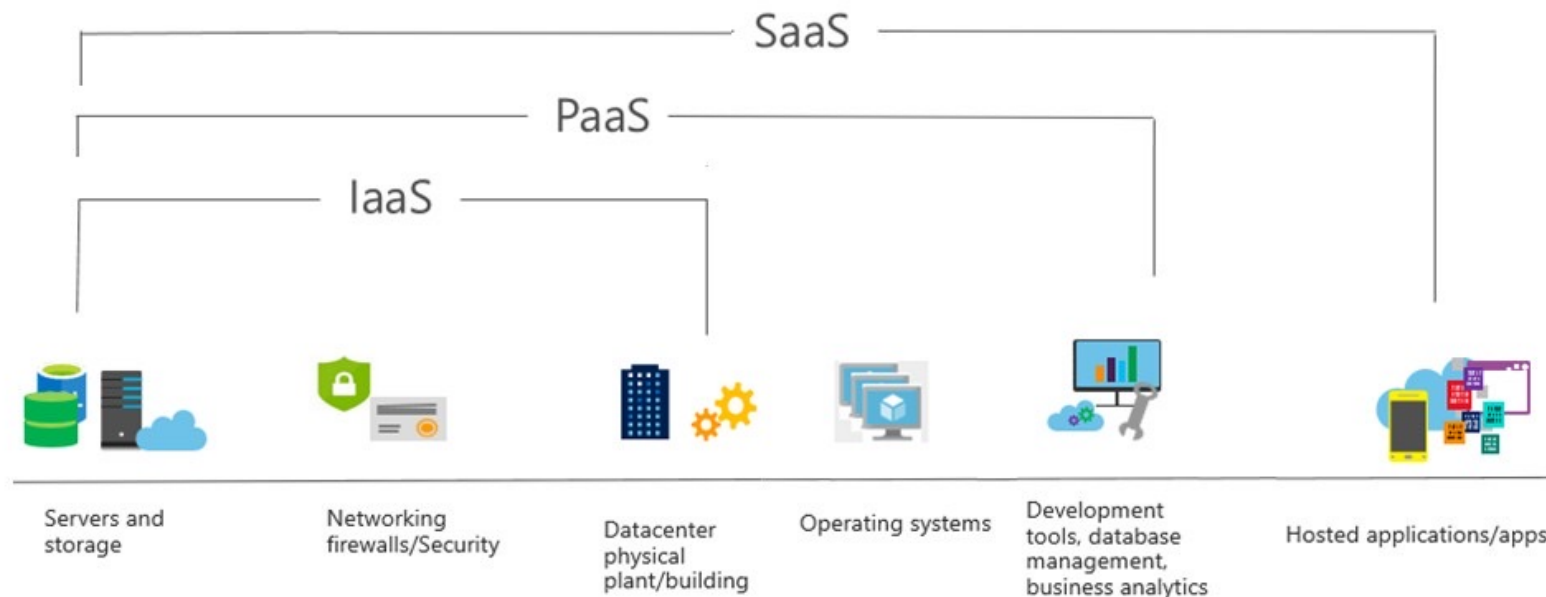
Platform as a Service (PaaS)

Provides environment for building, testing, and deploying software applications; without focusing on managing underlying infrastructure.



Software as a Service (SaaS)

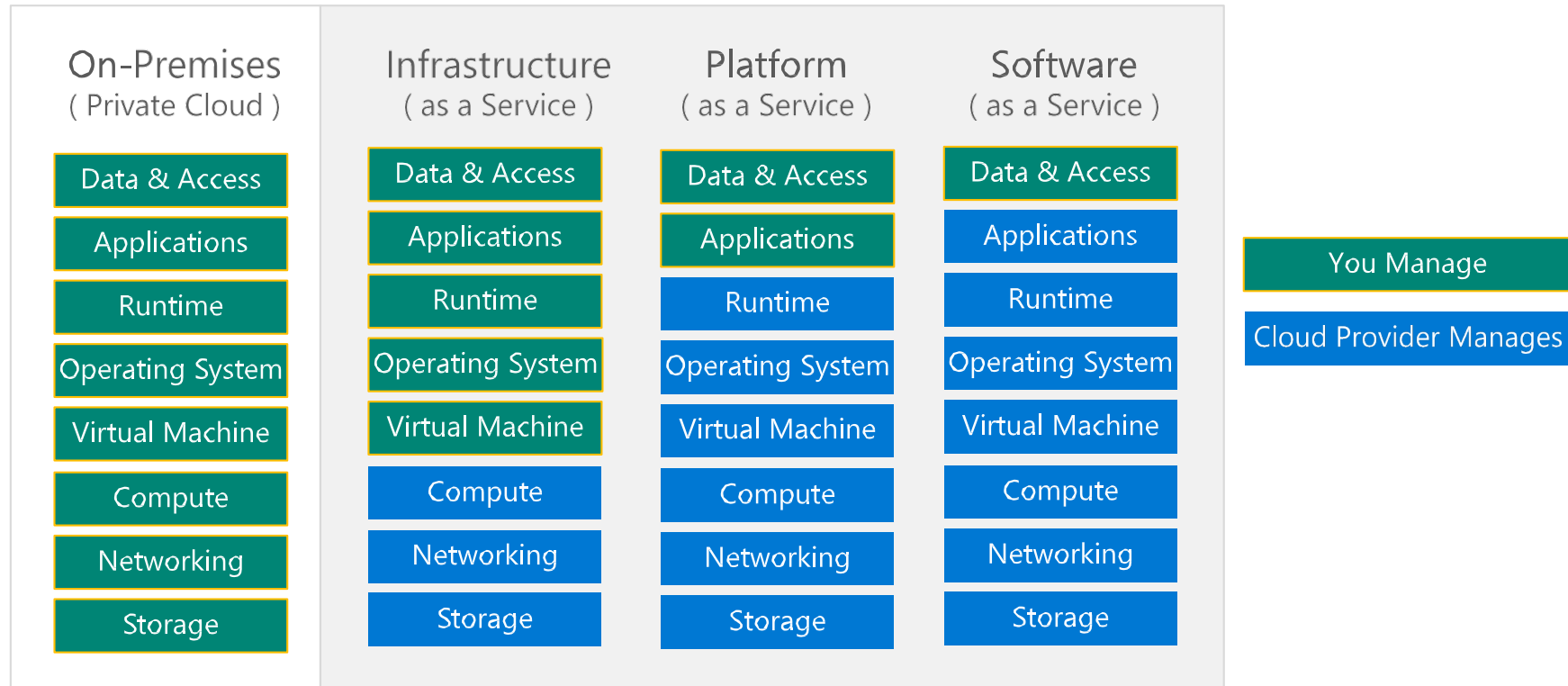
Users connect to and use cloud-based apps over the internet: for example, Microsoft Office 365, email, and calendars.



Cloud service comparison

IaaS	PaaS	SaaS
<p>The most flexible cloud service.</p> <p>You configure and manage the hardware for your application.</p>	<p>Focus on application development.</p> <p>Platform management is handled by the cloud provider.</p>	<p>Pay-as-you-go pricing model.</p> <p>Users pay for the software they use on a subscription model.</p>

Shared responsibility model



Cloud Databases

What is a cloud database?

- Provided by the public or private cloud provider as fully managed software solution
- The data is also managed on cloud storage.
- Data is stored in a cloud storage tier (such as a cloud object store, distributed datastore or other proprietary cloud storage infrastructure).
- These systems are designed to support applications that go from transactional and/or analytical processing to data warehouses and deep learning

Commerical Cloud (Database) Providers

- **Microsoft Azure**
- Amazon AWS
- Alibaba
- IBM
- Google Cloud
- Oracle Cloud Architecture (OCI)
- **Snowflake cloud database**

Data on Microsoft Azure

Roles in Data



Database Administrator

- Database Management
- Implements Data Security
- Backups
- User Access
- Monitors performance



Data Engineer

- Data Pipelines and processes
- Data Ingestion storage
- Prepare data for Analytics
- Prepare data for analytical processing



Data Analyst

- Provides insights into the data
- Visual Reporting
- Modeling Data for Analysis
- Combines data for visualization and analysis

Common Tools – Database Administrator

Azure Data Studio

- Graphical interface for managing on-premises and cloud-based data services
- Runs on Windows, macOS, Linux

SQL Server Management Studio

- Graphical interface for managing on-premises and cloud-based data services
- Runs on Windows
- Comprehensive Database Administration tool

Azure Portal / CLI

- Tools for management and provisioning of Azure Data Services
- Manual and automation of scripts using Azure Resource Manager or Command Line Interface scripting

Common Tools – Data Engineering

Azure Synapse Studio

- Azure Portal integrated to manage Azure Synapse
- Data Ingestion (Azure Data Factory)
- Management of Azure Synapse assets (SQL Pools / Spark Pool)

SQL Server Management Studio

- Graphical interface for managing on-premises and cloud-based data services
- Runs on Windows
- Comprehensive Database Administration tool

Azure Portal / CLI

- Tools for management and provisioning of Azure resources
- Manual and automation of scripts using Azure Resource Manager or Command Line Interface scripting

Common Tools – Data Analyst

Power BI Desktop

- Data Visualization tool
- Model and Visualize Data
- Management of Azure Synapse assets (SQL Pools / Spark Pool)

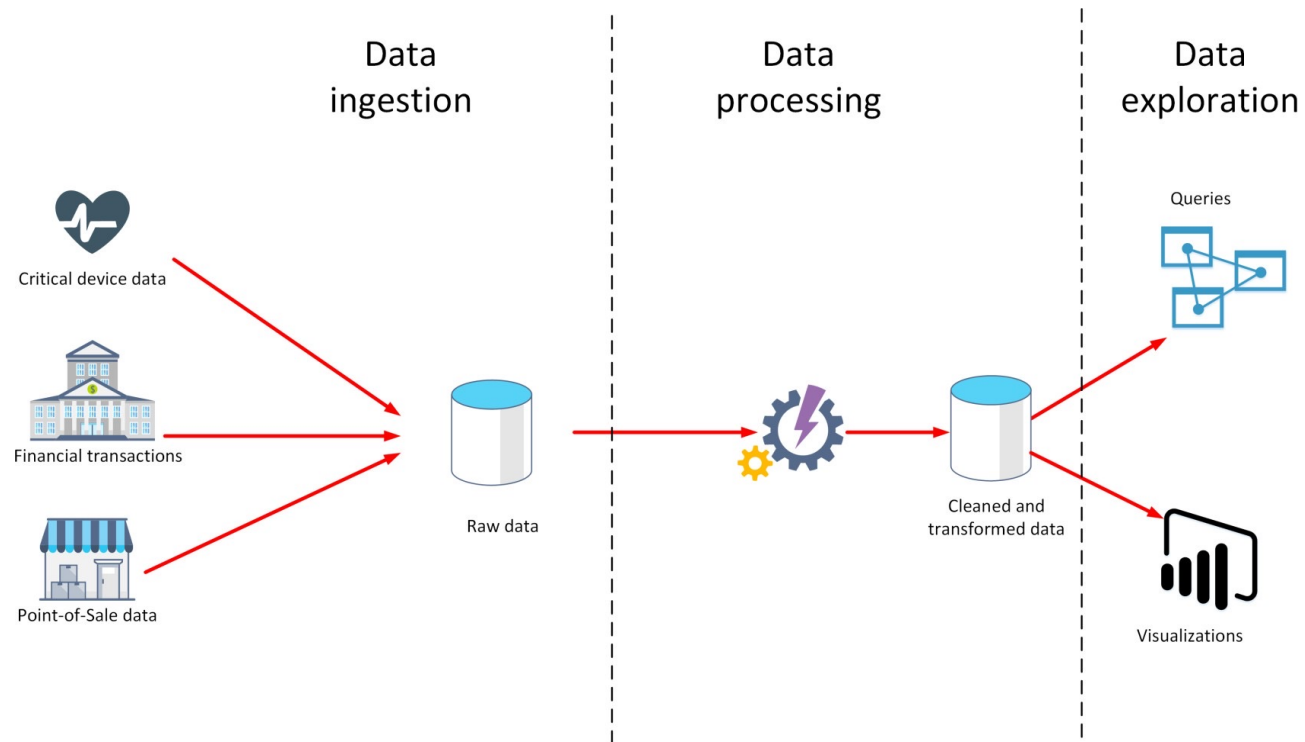
Power BI Portal / Power BI Service

- Authoring and management of Power BI reports
- Authoring of Power BI dashboards
- Share Reports / Datasets

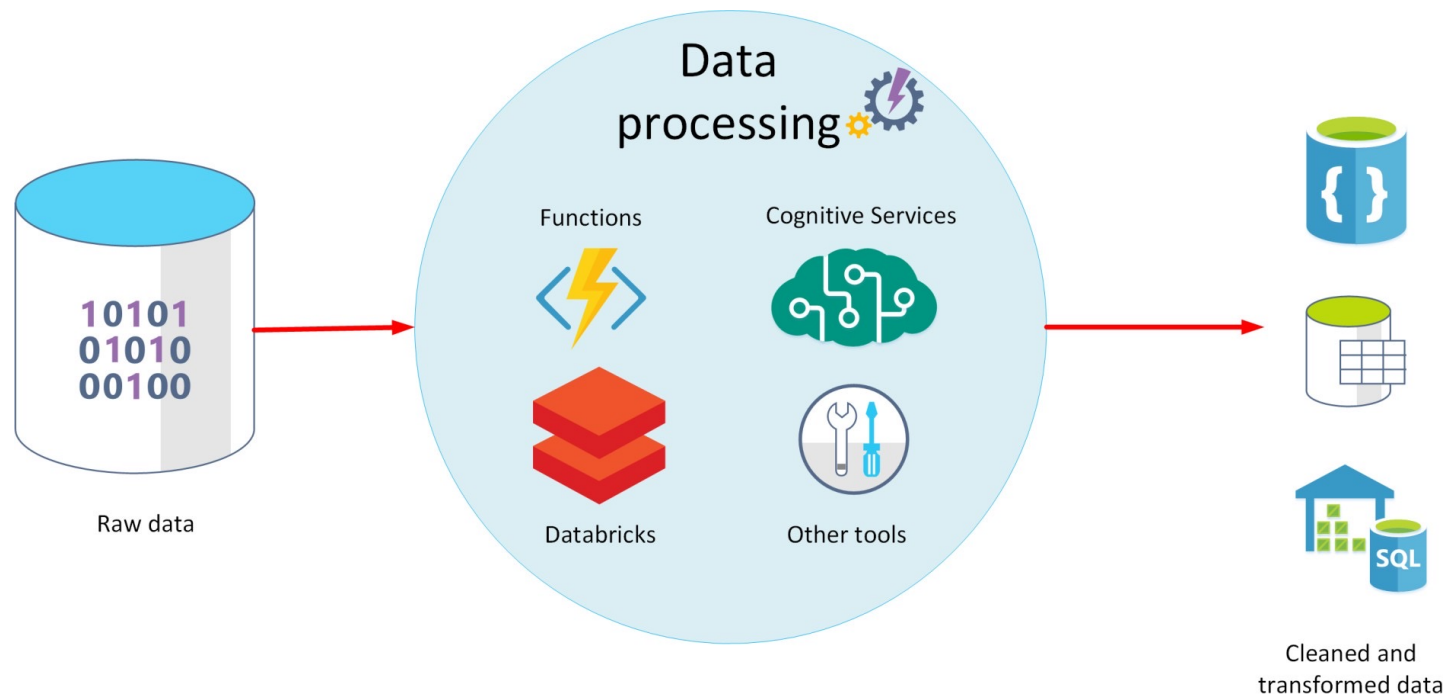
Power BI Report Builder

- Data Visualization tool for paginated reports
- Model and Visualize paginated reports

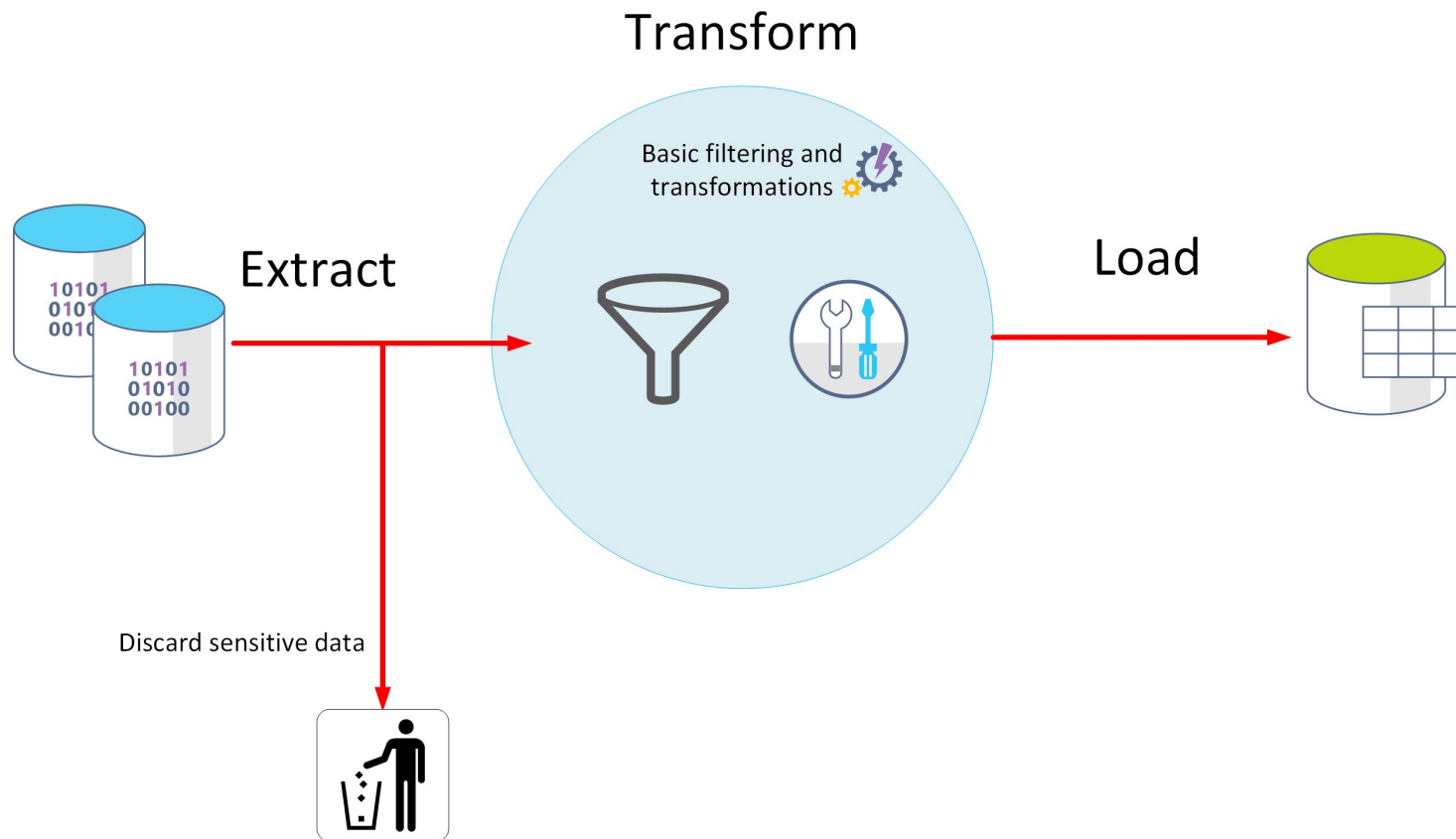
What is data ingestion?



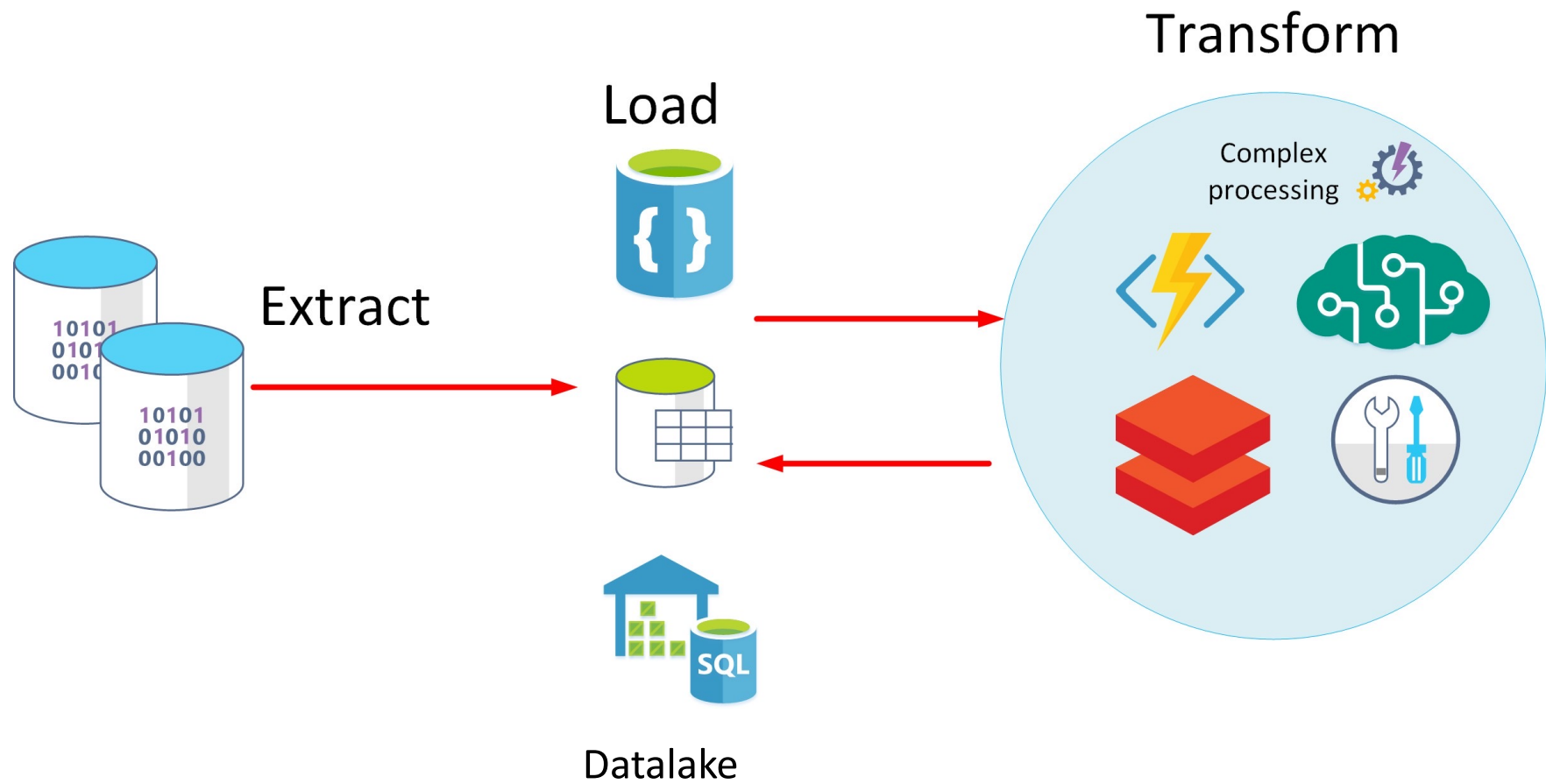
What is data processing?



What is ETL?



What is ELT?



Datalake

- A repository of raw data.
- Data was simply placed into the data lake for anyone to analyze or use.
- The data lake data comes from a wide variety of sources.
- Currently, an analytical infrastructure is built from the raw data:
 - Identify how data related to each other
 - Identify the timeliness of data
 - Examine the quality of the data
 - Identify the lineage
(NL: herkomst, afstamming)
of data



ELT

- Extract Load Transform, alternative to ETL, mostly used in the cloud.
- Data is only transformed to the required format when analytics is needed.
- Use case:

You have data about money from the US, Australia, and Canada. All three currencies are in dollars.

In a classic DWH setup (with ETL) you would probably convert all data about money to a single currency, so you can easily add them together when making reports.

But you can't just add the dollars together and have a meaningful answer. Instead, to have a meaningful answer, you have to convert (transform) two of the three currencies to a common value. Then, and only then, can you add the values up in a meaningful manner. And even then, the truthfulness of the data is only relevant to the moment that data has been recalculated because exchange rates fluctuate on a daily/hourly basis.

Explore Data Visualization

Power BI: A collection of software, services, apps, and connectors.



Explore Data Analytics

Descriptive



Diagnostic



Predictive



Prescriptive



Cognitive



Relational Data Offerings in Microsoft Azure

What are Azure Data Services?

SQL Server on Azure Virtual Machines



Best for re-hosting and apps requiring
OS-level access and control

Automated manageability features and
OS-level access

Infrastructure as a Service

Azure SQL Managed Instance



Best for modernizing existing apps

Offers high compatibility with SQL Server
and native VNET support

Platform as a Service

Azure SQL Database



Best for building new apps in the cloud

Pre-provisioned or serverless compute
and Hyperscale storage to meet
demanding workload requirements

SQL Server on Azure virtual machines



Customer challenge

I want to migrate to the cloud as fast as possible but maintain operating system control and complete SQL Server functionality



Solution

Get the combined performance, security, and analytics of SQL Server, backed by the flexibility, security, and hybrid connectivity of Azure

Key features

SQL Server and OS server access
Expansive SQL and OS versions
Windows, Linux, Containers
File stream, DTC, and Simple Recovery model
SSAS, SSRS, and SSIS

Azure differentiators

Free Extended Security Updates for SQL Server 2008/R2
Automated Backups and Security Updates
Point in Time Restore with Azure Backup
Accelerated storage performance with Azure Blob Caching
435 percent overall return on an Azure IaaS investment over five years¹

Azure SQL DB



Customer challenge

I want to build modern apps, potentially multi-tenanted, with the highest uptime and predictable performance



Solution

Azure SQL Database is a highly scalable cloud database service with built-in high availability and machine learning

Key features

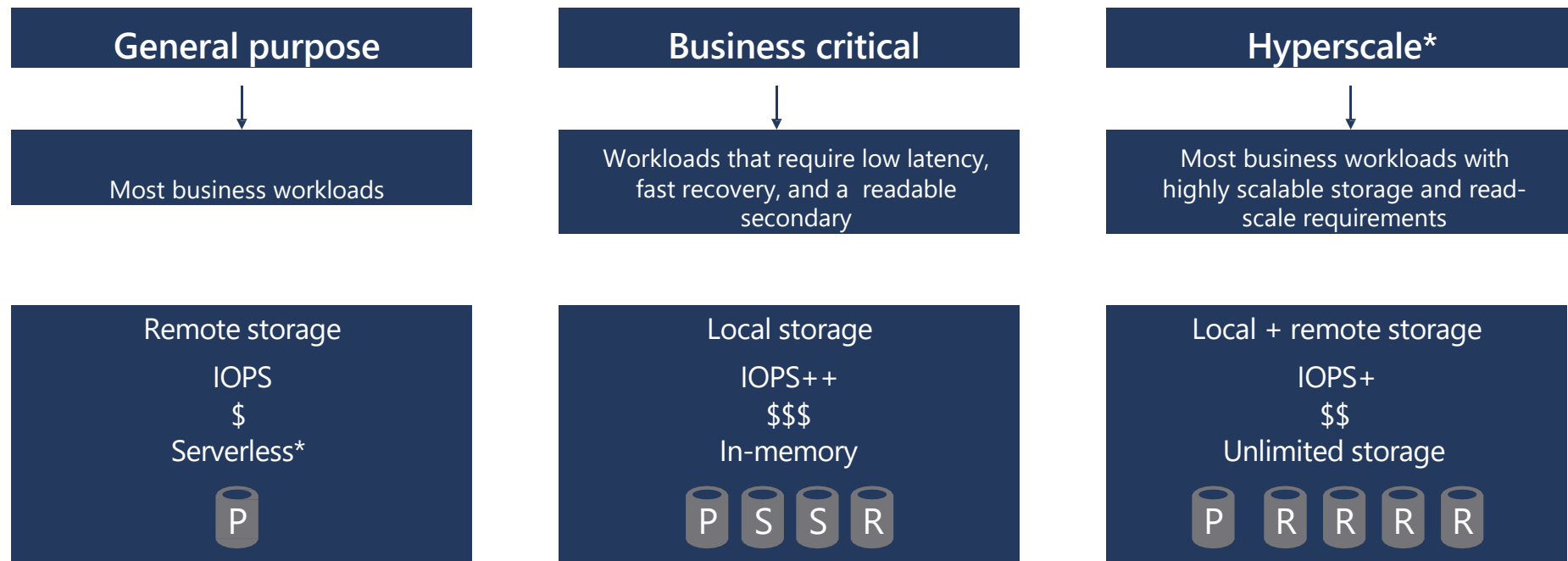
- Single database or elastic pool
- Hyperscale storage (100TB+)
- Serverless compute
- Fully managed service
- Private link support
- High availability with AZ isolation

Azure differentiators

- Industry highest availability SLA of 99.995%
- Industry only business continuity SLA with 5 second RPO and 30 second RTO
- Price-performance leader for mission-critical workloads while costing up to 86 percent less than AWS RDS (GigaOm)

Multi-tenant = multi-customers (≠users) each working in their own environment.

Azure SQL DB Service Tiers



*Not in managed instance

Azure SQL DB Managed Instance



Customer challenge

I want to migrate to the cloud, remove management overhead, but I need instance-scoped features (Service Broker, SQL Server Agent, CLR...)



Solution

Managed instance combines leading security features with SQL Server compatibility and business model designed for on-premises customers

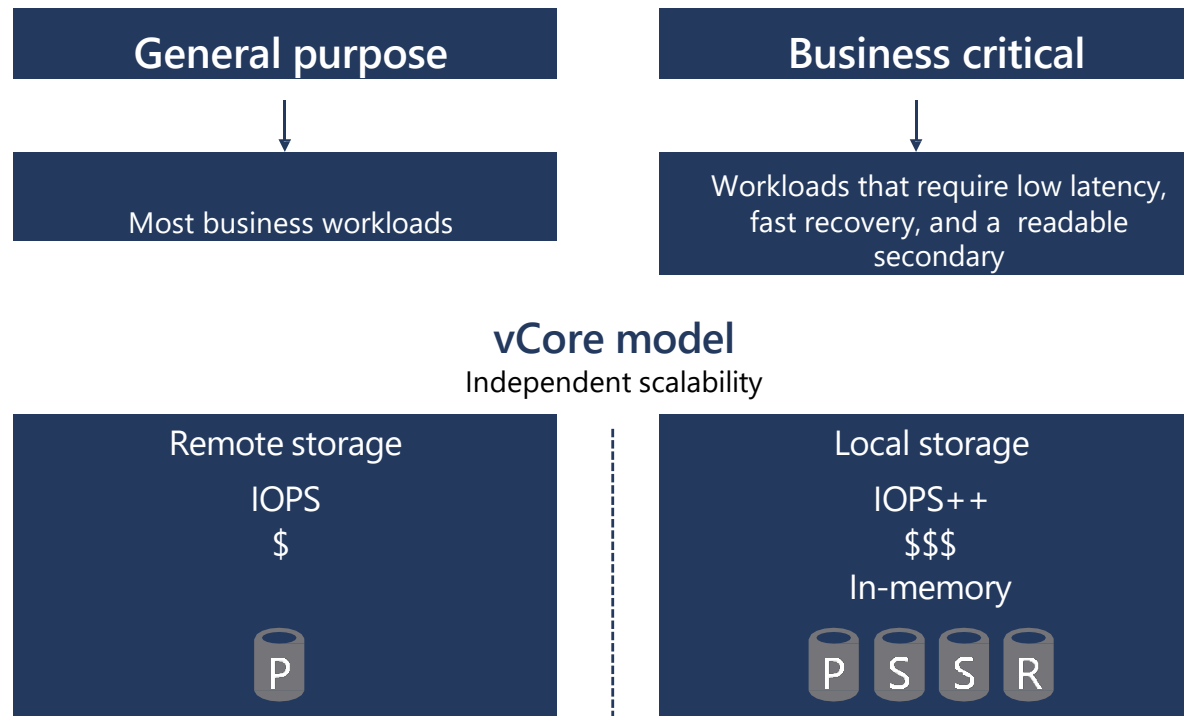
Key features

- Single instance or instance pool
- SQL Server surface area (vast majority)
- Native virtual network support
- Fully managed service
- On-premise identities enabled with Azure AD and AD Connect

Azure differentiators

- Near zero downtime migration using log shipping
- Fully managed business continuity with failover groups
- Projected return on investment of 212 percent over three years¹
- The best of SQL Server with the benefits of a managed service

Managed Instance Service Tiers



Azure SQL Managed instance or DB



Azure SQL managed instance

Single instance

SQL Server surface area (vast majority)

Native virtual network support

Fully managed service

Instance pool

Pre-provision compute resources for migration

Enables cost-efficient migration.

Ability to host smaller instances (2Vcore)

Currently in public preview



Azure SQL Database

Single database

Hyperscale storage (up to 100TB)

Serverless compute

Fully managed service

Elastic pool

Resource sharing between multiple databases to price optimize

Simplified performance management for multiple databases

Fully managed service

PostgreSQL, MariaDB, MySQL



PostgreSQL is the most popular and wanted database for modern apps



MySQL is a leading open source relational database for LAMP stack apps



MariaDB is a community-developed fork of MySQL with strong focus on the user community

Benefits of Azure Database for MySQL, PostgreSQL, MariaDB



Fully managed community database

Take advantage of a fully managed service while still using the tools and languages you're familiar with



Built-in high availability for lowest TCO

Ensure your data is always available without the need for additional costs



Intelligent performance and scale

Improve performance with built-in intelligence and up to 16TB storage and 20K IOPs



Industry-leading security and compliance

Protect your data with enhanced security features including Advanced Threat Protection



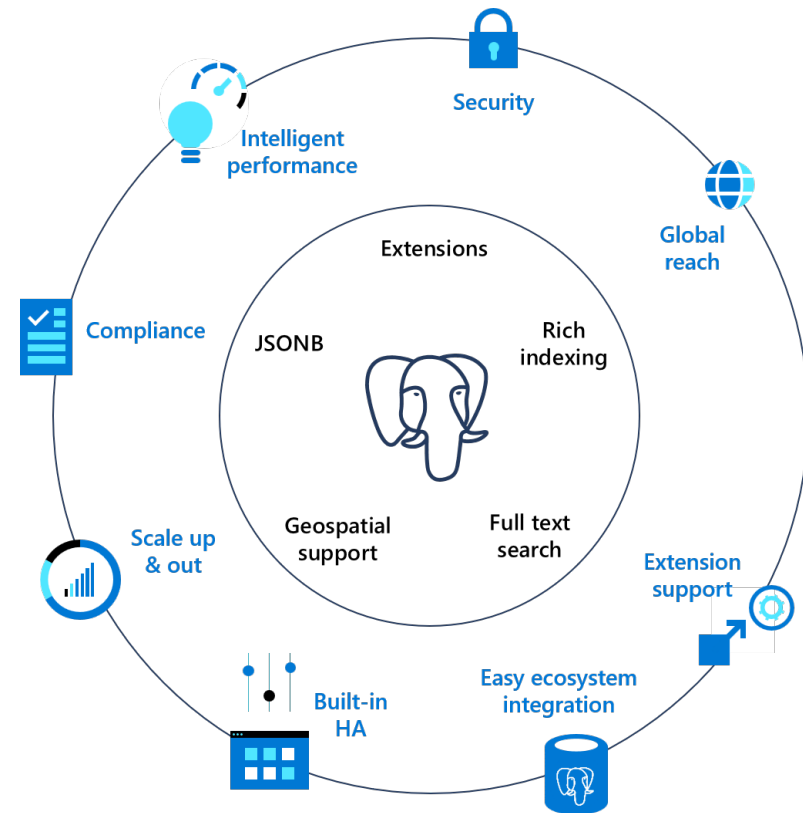
Integration with the Azure ecosystem

Build apps faster with Azure services and safeguard your innovation with Azure IP Advantage

Azure Database for PostgreSQL

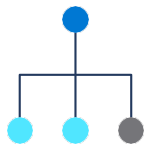
Azure builds upon
the core benefits of
PostgreSQL and
Open Source

Azure Database for PostgreSQL is
fully-managed, community PostgreSQL



The benefits of Azure Database for PostgreSQL

Build or migrate your workloads with confidence and optimized for value



Fully managed and secure

Focus on your apps while Azure manages resource intensive tasks, supports a large variety of Postgres versions and provides best in industry indemnification coverage



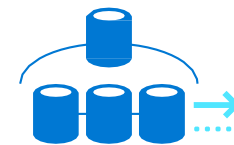
Intelligent performance optimization

Improve performance and reduce cost with customized recommendations



Flexible and open

Stay productive with your favorite Postgres extensions and leverage Microsoft's contributions to the Postgres community



High performance scale-out with Hyperscale

Break free from the limits of single node Postgres and scale out across 100s of nodes

Single Server

Hyperscale

Non-relational Data Offerings in Microsoft Azure

Explore Azure Table storage

Key (Customer ID)	Value (Customer Data)
C1	AAAAA BBB 101 Block Street YY 999 888
C2	MM NN 21 A Street 5 B Avenue
C3	DDD EEE FFF 111 222 66 C Road

Explore Azure Blob Storage

Block blobs	Page blobs	Append blobs
<ul style="list-style-type: none">• Has a maximum size of 4.7TB• Best for storing large, discrete, binary objects that changes infrequently• Each individual block can store up to 100MB of data• A block blob can contain up to 50000 blocks	<ul style="list-style-type: none">• Can hold up to 8TB of data• Is organized as a collection of fixed sized-512 byte pages• Used to implement virtual disk storage for virtual machines	<ul style="list-style-type: none">• The maximum size is just over 195GB• Is a block blob that is used to optimize append operations• Each individual block can store up to 4MB of data

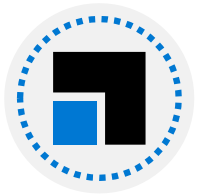
Explore Azure File Storage

The screenshot displays the Azure portal interface for a storage account named 'jpwsstorageaccount'. The breadcrumb navigation at the top shows the path: Home > Microsoft.StorageAccount-20200520142053 | Overview > jpwsstorageaccount. The left-hand navigation pane includes sections for 'Overview' (with links to Activity log, Access control (IAM), Tags, Diagnose and solve problems, Data transfer, Events, and Storage Explorer (preview)) and 'Settings' (with links to Access keys, Geo-replication, CORS, Configuration, Encryption, Shared access signature, Firewalls and virtual networks, Private endpoint connections, Advanced security, Static website, and Properties). The main content area features a search bar, action buttons (Open in Explorer, Move, Refresh, Delete, Feedback), and an information banner about classic alerts in Azure Monitor. Below this, a table of account properties is shown:

Resource group (change)	: learnrg	Performance/Access tier	: Standard/Hot
Status	: Primary: Available, Secondary: Available	Replication	: Read-access geo-redundant storage (RA-GRS)
Location	: UK West, UK South	Account kind	: StorageV2 (general purpose v2)
Subscription (change)	:		
Subscription ID	:		
Tags (change)	: Click here to add tags		

Below the properties table, there are four service tiles: 'Containers' (Massively scalable data lake storage), 'File shares' (Serverless SMB file shares, highlighted with a red border), 'Tables' (Tabular data storage), and 'Queues' (Effectively scale apps according to traffic). Each tile includes a 'Learn more' link. At the bottom, the 'Tools and SDKs' section lists various development tools and languages: Storage Explorer (preview), PowerShell, Azure CLI, .NET, Java, Python, and Node.js.

Explore Azure Cosmos DB



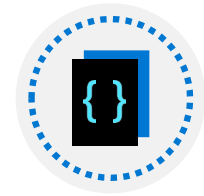
Scalability



Performance



Availability



Programming
model

Use cases for Azure Cosmos DB

Web and retail

Using Azure Cosmos DB's multi-master replication model along with Microsoft's performance commitments, Data Engineers can implement a data architecture to support web and mobile applications that achieve less than a 10-ms response time anywhere in the world

Gaming

The database tier is a crucial component of gaming applications. Modern games perform graphical processing on mobile/console clients but rely on the cloud to deliver customized and personalized content like in-game stats, social media integration, and high-score leader boards.

IoT scenarios

Hundreds of thousands of devices have been designed and sold to generate sensor data known as Internet of Things (IoT) devices. Using technologies like Azure IoT Hub, Data Engineers can easily design a data solution architecture that captures real-time data. Cosmos DB can accept and store this information very quickly

Snowflake

- De facto standard in world of cloud scale datawarehouses.
- Dataplatform and datawarehouse that supports SQL.
- Can work cross-cloud.