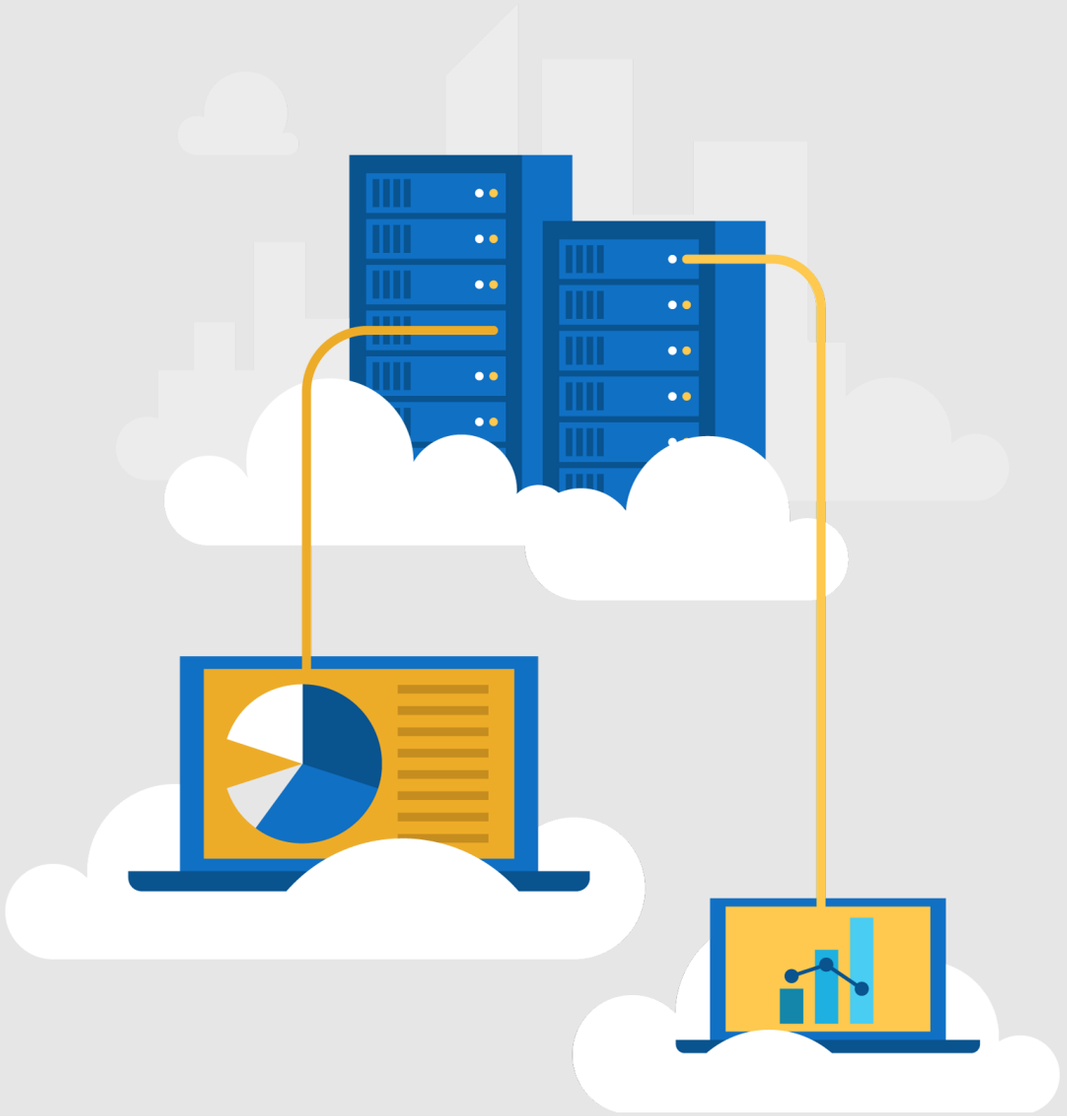
Microsoft

Azure Study Group

AZ-301 - Microsoft Azure Architect Design

Jeff Wagner
Partner Technology Strategist



Microsoft Azure

Design a Data Platform Solution (15-20%)

Agenda

1. Agenda
2. Speaker Introduction
3. Feedback Loop
4. Objective Review
5. Open Mic

Series Agenda

1	Determine Workload Requirements (10-15%)
2	Design for Identity and Security (20-25%)
3	Design a Data Platform Solution (15-20%)
4	Design a Business Continuity Strategy (15-20%)
5	Design for Deployment, Migration, and Integration (10-15%)
6	Design an Infrastructure Strategy (15-20%)

https://aka.ms/azurecsg

Series Agenda

1	Determine Workload Requirements (10-15%)
2	Design for Identity and Security (20-25%)
3	Design a Data Platform Solution (15-20%)
4	Design a Business Continuity Strategy (15-20%)
5	Design for Deployment, Migration, and Integration (10-15%)
6	Design an Infrastructure Strategy (15-20%)

https://aka.ms/azurecsg

Speaker Introduction - Jeff Wagner

- Partner Technology Strategist based in Atlanta
- 21+ years with Microsoft, more in the industry
- Constant learner - *Ancora Imparo*
- Working on the same certifications that you are

Microsoft Azure

Feedback Loop

Objectives

Design a Data Management Strategy

May include but not limited to: Choose between managed and unmanaged data store; choose between relational and non-relational databases; design data auditing and caching strategies; identify data attributes (e.g., relevancy, structure, frequency, size, durability, etc.); recommend Database Transaction Unit (DTU) sizing; design a data retention policy; design for data availability, consistency, and durability; design a data warehouse strategy

Design a Data Protection Strategy

May include but not limited to: Recommend geographic data storage; design an encryption strategy for data at rest, for data in transmission, and for data in use; design a scalability strategy for data; design secure access to data; design a data loss prevention (DLP) policy

Objectives (cont.)

Design and Document Data Flows

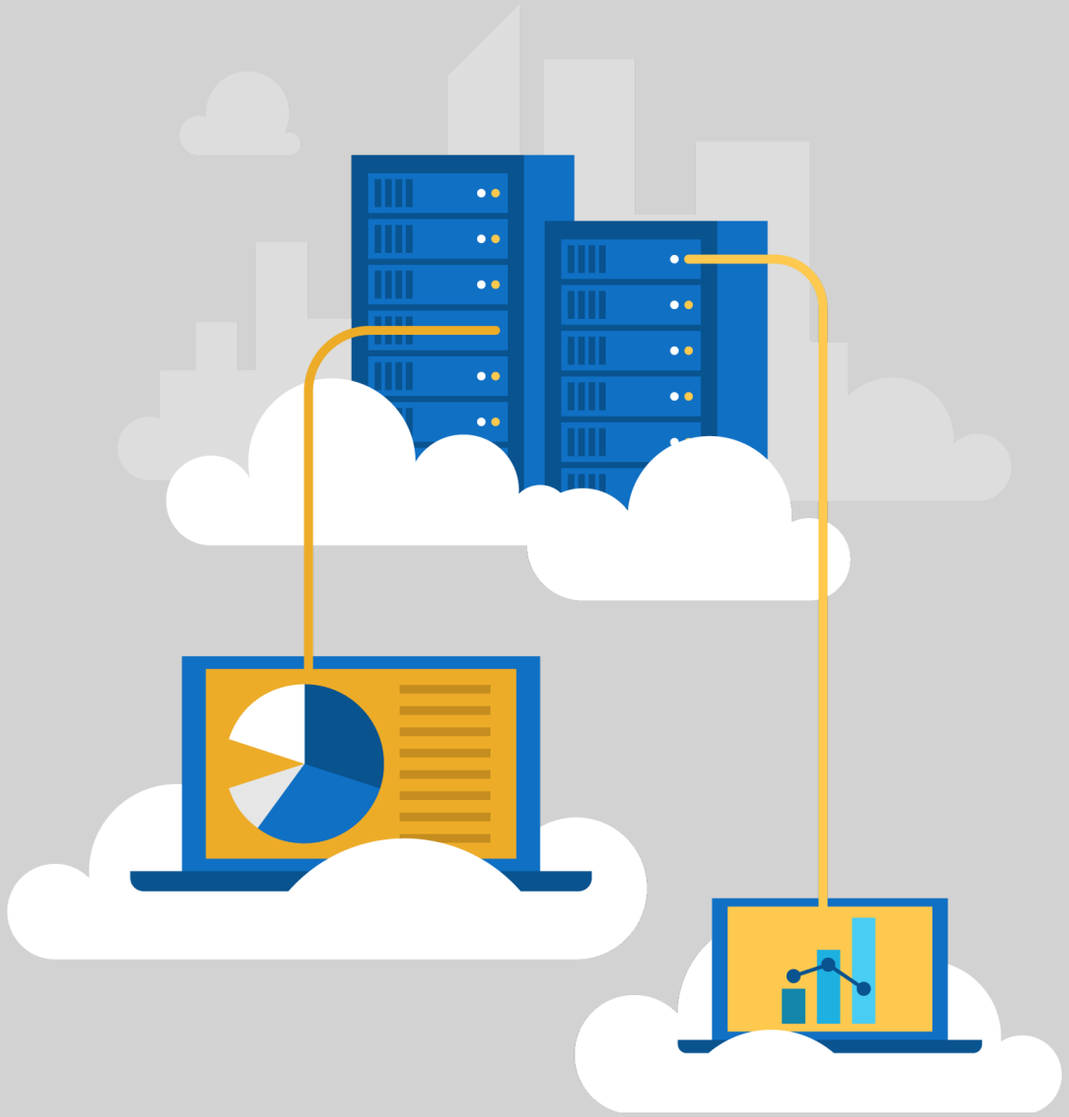
May include but not limited to: Identify data flow requirements; create a data flow diagram; design a data flow to meet business requirements; design a data import and export strategy

Design a Monitoring Strategy for the Data Platform

May include but not limited to: Design for alert notifications; design an alert and metrics strategy

Microsoft

Design a Data Management Strategy



Microsoft

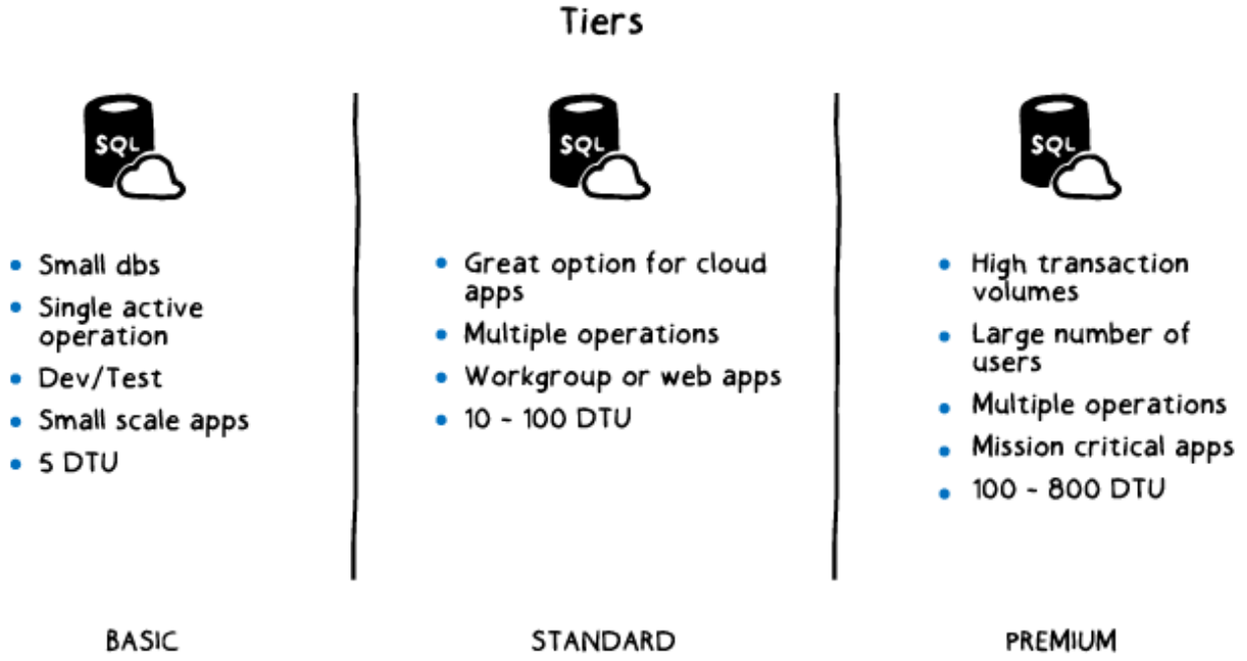
Comparing Database Options in Azure

Azure SQL Database

- A relational Database-as-a-Service offering that provides:
  - **Predictable performance: measured in Database Throughput Units (DTUs)**
  - **High compatibility: supporting existing SQL client applications via a Tabular Data Stream (TDS) endpoint**
  - **Simplified management: including SQL Server-specific and Azure tools**

Azure SQL Database Tiers

- Pricing tiers:
  - **Basic: small databases with single concurrent user**
  - **Standard: Medium size databases that must support multiple concurrent connections**
  - **Premium: Large databases that must support large number of concurrent connections and operations**
- Performance levels:
  - **Categories within service tiers**
  - **Provide more granular scaling**
  - **Determine:**
    - **Maximum DTUs**
    - **Maximum database size**



Azure SQL Database Elastic Scale

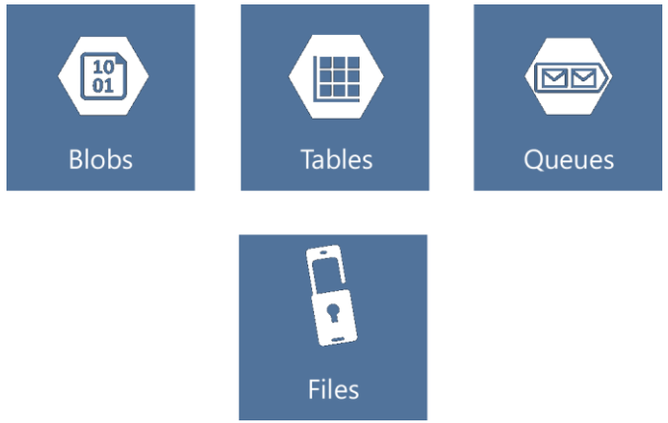
- Facilitates horizontal scaling and includes:
  - **Elastic Scale library for client applications to configure and access shards**
  - **Elastic Scale server-side component that implements sharding**
- Relies on the shard-based scaling strategy:
  - **Uses a shard key to determine partitioning of the database**
  - **Automatically directs transactions to the appropriate shard**
  - **Coordinates data movement between shards to split or merge data ranges**

Third-Party Databases in Azure

- Managed database options:
  - **Offer:**
    - **Built-in high availability with no additional cost.**
    - **Predictable performance, using pay-as-you-go pricing.**
    - **Automatic scaling.**
    - **Encryption at-rest and in-transit.**
    - **Automatic backups and point-in-time-restore for up to 35 days.**
    - **Enterprise-grade security and compliance.**
  - **Include:**
    - **Azure Database for MySQL**
    - **Azure Database for PostgreSQL**
- Non-managed database options:
  - **Windows and Linux Azure VMs hosting MySQL installations**
  - **ClearDB offering managed MySQL instance**

Azure Storage

- Massively scalable
- Uses partitioning system that facilitates load-balanced data access
- Offers different types of storage services:
  - **Blobs**
  - **Tables**
  - **Queues**
  - **Files**



- Accessible via:
  - **REST API**
  - **Client libraries, including .NET, Java/Android, Node.js, PHP, Ruby, and Python**

Replication

- Content of each storage account is always replicated:
  - **Ensures durability and high availability**
- Scope of replication is configurable:
  - **Locally redundant storage (LRS):**
    - **Consists of 3 copies of the same content in the same availability zone in one region**
  - **Zone-redundant storage (ZRS):**
    - **Consists of 3 copies of the same content across multiple zones in one region**
  - **Geo-redundant storage (GRS):**
    - **Consists of 6 copies of the same content in two regions, with 3 copies per region**
    - **Facilitates failover to a remote region in case of the primary region failure**
  - **Read access geo-redundant storage (RA-GRS):**
    - **Consists of 6 copies of the same content in two regions, with 3 copies per region**
    - **Facilitates failover to a remote region in case of the primary region failure**
    - **Provides read access to the remote region**

Azure Storage Tables

- A NoSQL data store:
  - **Provides storage for massive amounts of structured, non-relational data**
  - **Is suitable for datasets that do not require joins, foreign keys, or stored procedures**
  - **Offers access via OData protocol and LINQ queries with WCF Data Service .NET libraries**
  - **Includes a clustered index for data queries**
  - **Includes the following components:**
    - **An Azure Storage account**
    - **A table:**
      - **a collection of entities with no enforced schema**
    - **Entities:**
      - **sets of properties similar to database rows**
    - **Properties:**
      - **Name-value pairs**
      - **Up to 252 custom properties per entity**
      - **3 system properties: PartitionKey, RowKey, and Timestamp**

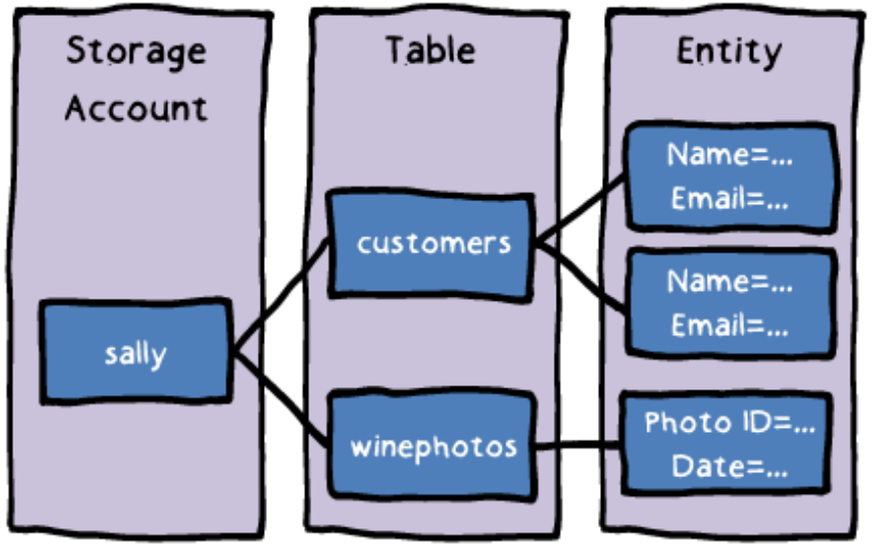
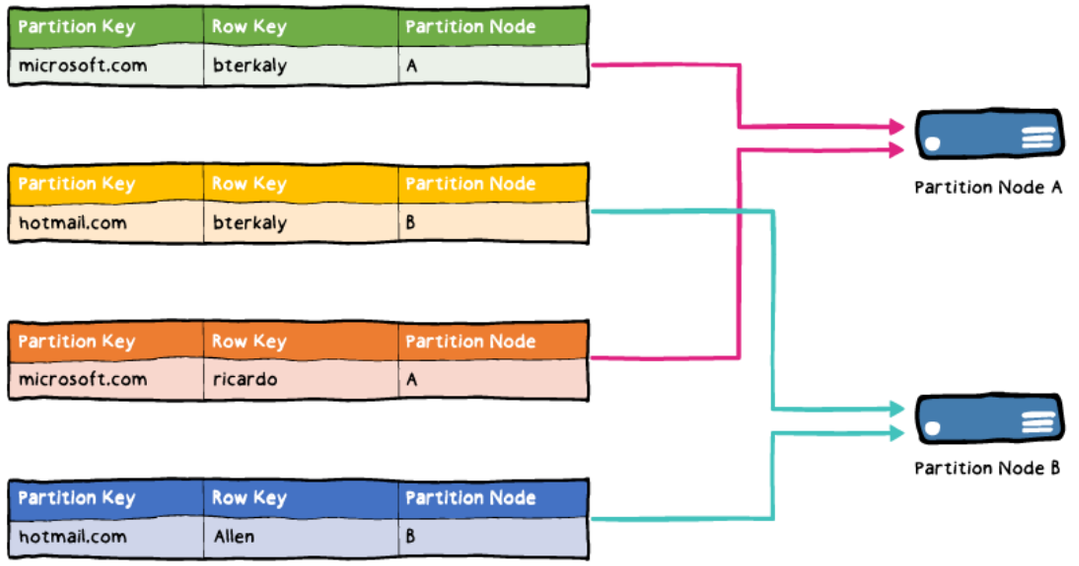


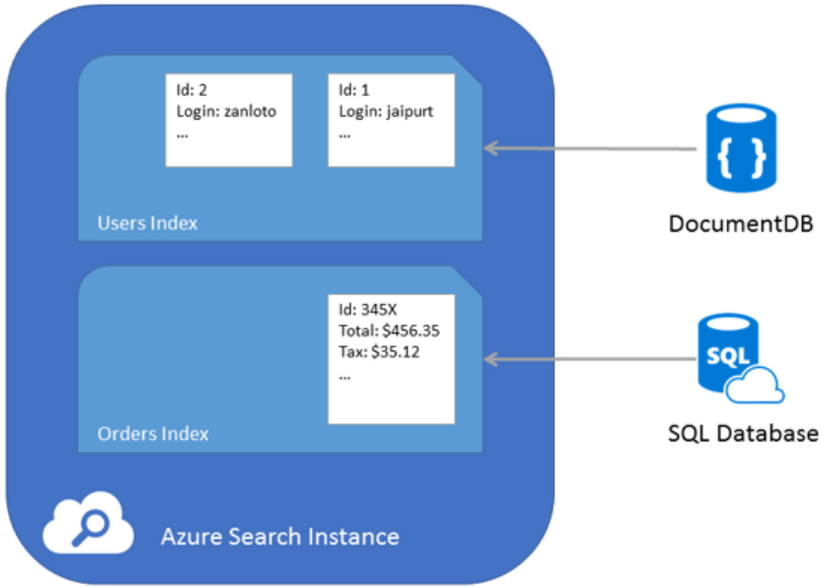
Table Partitioning

- Azure Storage Tables partitions:
  - **Represent collections of entities with the same PartitionKey values**
  - **Determine partitioning of the underlying storage**
  - **A single partition has the scalability target of 500 entities per second**
  - **To scale horizontally, it is necessary to configure multiple partitions**
  - **The choice of PartitionKey is critical:**
    - **It affects horizontal scalability**
    - **It affect indexing behavior:**
      - **The primary key combinesPartitionKey and RowKey**
      - **The clustered index is based on the primary key**



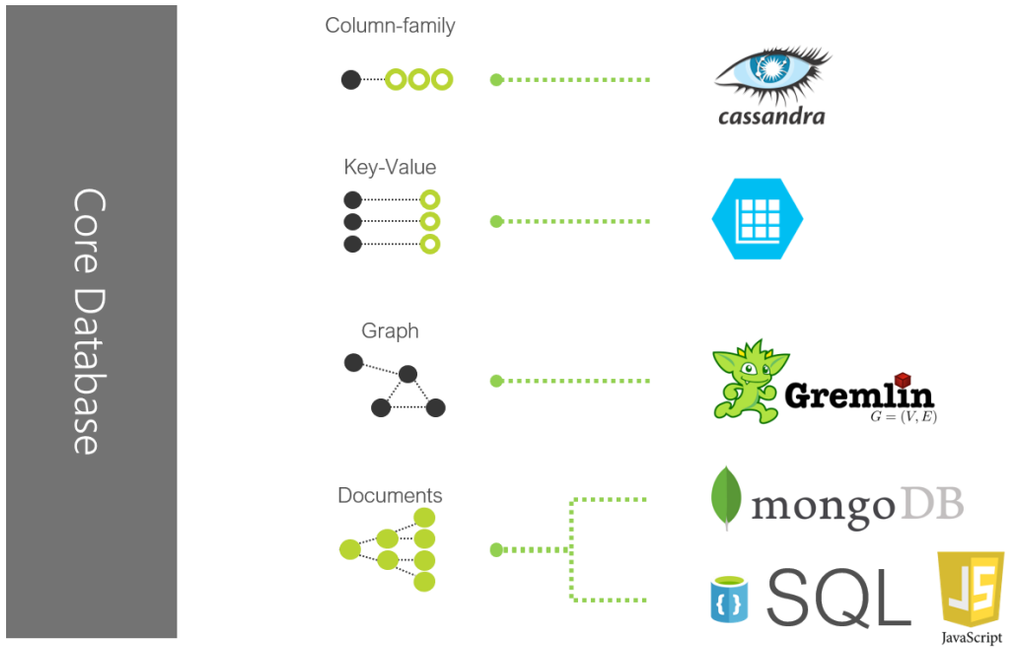
Azure Search

- Managed search service:
  - **Allows developers to build search applications by using .NET SDK or REST API**
  - **Based on automatically generated index with support for custom schemas**
  - **Provides full-text search over custom content**
  - **Offers advanced search capabilities:**
    - **Type-ahead query suggestions**
    - **Hit-highlighting**
    - **Faceted navigation**
    - **Built-in native language support**
  - **Available in two pricing tiers:**
    - **Free: offers limited, shared resources**
    - **Standard: implements autoscaling**
  - **Includes indexing of Azure services:**
    - **Azure SQL Database**
    - **Cosmos DB**



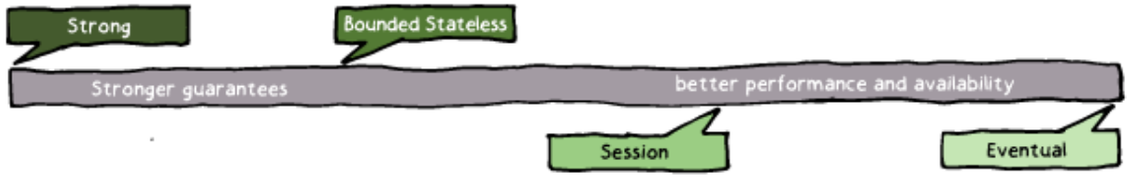
Azure Cosmos DB APIs

- Azure Cosmos DB is a globally distributed, multi-model database:
  - **Accessible via different API (depending on the choice of the model):**
    - **DocumentDB (SQL) API**
    - **MongoDB API**
    - **Graph (Gremlin) API**
    - **Tables (Key/Value) API**
  - **Automatically partitioned:**
    - **based on a custom partition key**
    - **with each partition offering:**
      - **Fixed performance**
      - **Fixed storage capacity**



Consistency Levels

- **Strong:**
  - **Guarantees that a write operation is only committed (and visible) on the primary after it has been committed and confirmed by ALL replicas.**
- **Bounded staleness:**
  - **Allows you to configure how stale documents can be within replicas. Staleness refers to the quantity of time (or version count) a replica document can be behind the primary document.**
- **Session:**
  - **Guarantees that all read and write operations are consistent within a user session.**
- **Consistent prefix:**
  - **Ensures that changes are read in the order that matches the sequence of the corresponding writes.**
- **Eventual:**
  - **Offers the loosest consistency and commits any write operation against the primary immediately. Replica transactions are asynchronously handle and will eventually (over time) be consistent with the primary.**

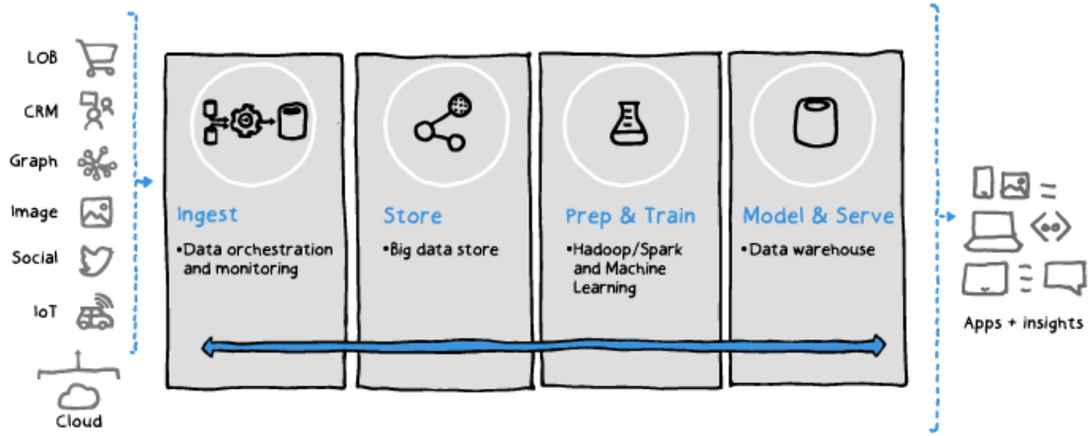


Choosing a Consistency Strategy

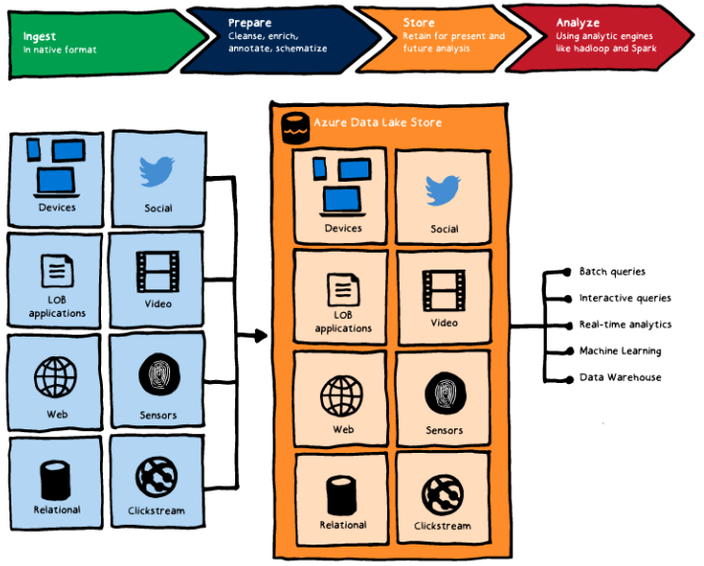
- Two primary considerations:
  - **a stronger consistency level will ensure that versions of documents in replicas do not lag behind the primary:**
    - **Is recommended for applications that require all replicas to exactly match the primary at any point in time**
    - **Affects negatively the speed of write operations, which must wait for every replica to confirm that the operation has been committed**
  - **A weaker consistency level will ensure that your database operates at peak efficiency:**
    - **Is recommended for applications that require the best performance**
    - **It introduces the possibility that read operations against a replica return stale data**

Data Storage & Integration Options

- SQL Data Warehouse:
  - **Operates as managed Data Warehouse**
  - **Delivers Massively Parallel Processing**
  - **Uses PolyBase to query big data stores**
  - **Runs complex queries across PB of data**
  - **Uses columnar storage to optimize cost & performance**



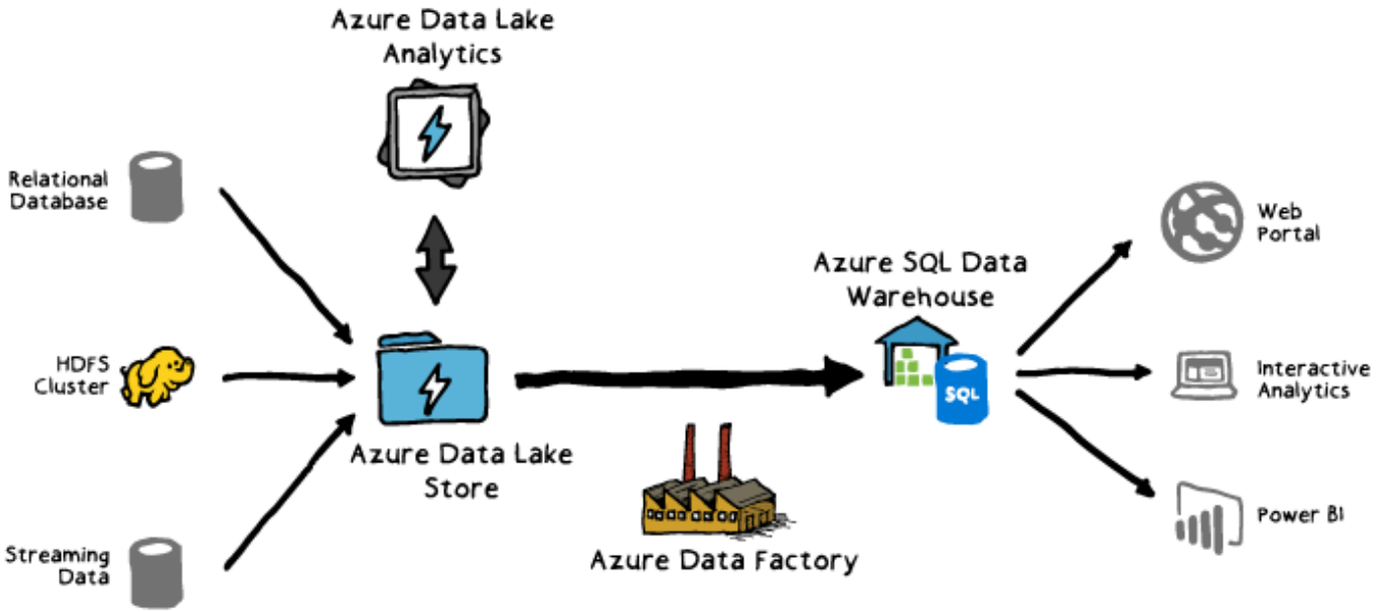
- Azure Data Lake:
  - **Serves as hyper-scale repository for big data workloads**
  - **Stores data of any size, type, and ingestion speed**
  - **Is accessible from HDInsight via WebHDFS REST API**



- Azure Data Lake Analytics:
  - **Offers managed, on-demand analytics job service**
  - **Runs U-SQL distributed queries against Data Lake, SQL Database, and Data Warehouse**

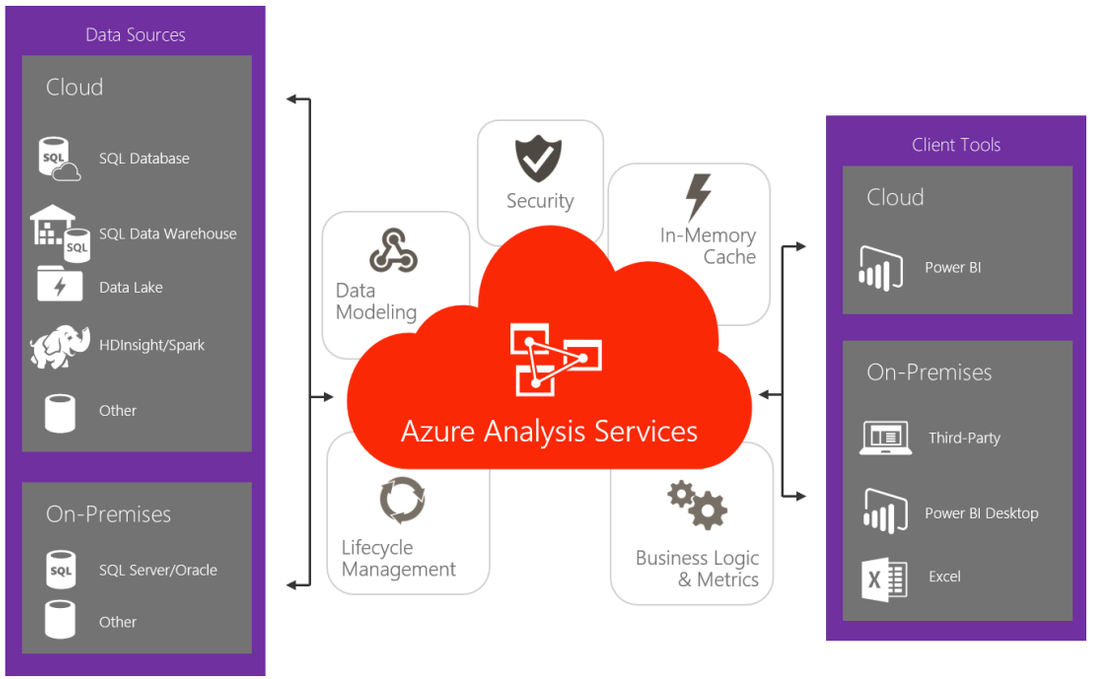
Data Integration

- Azure Data Factory:
  - **Managed Extract-Transform-Load (ETL) and data integration service**
  - **Facilitates data-drive workflows (pipelines) that carry out a range of tasks:**
    - **Connect & Collect:**
      - **Relational databases**
      - **HDFS clusters**
      - **Streaming data**
    - **Transform & Enrich:**
      - **Spark**
      - **Data Lake Analytics**
      - **Machine learning**
    - **Publish:**
      - **Web**
      - **Interactive Analytics**
      - **Power BI**
    - **Monitor**



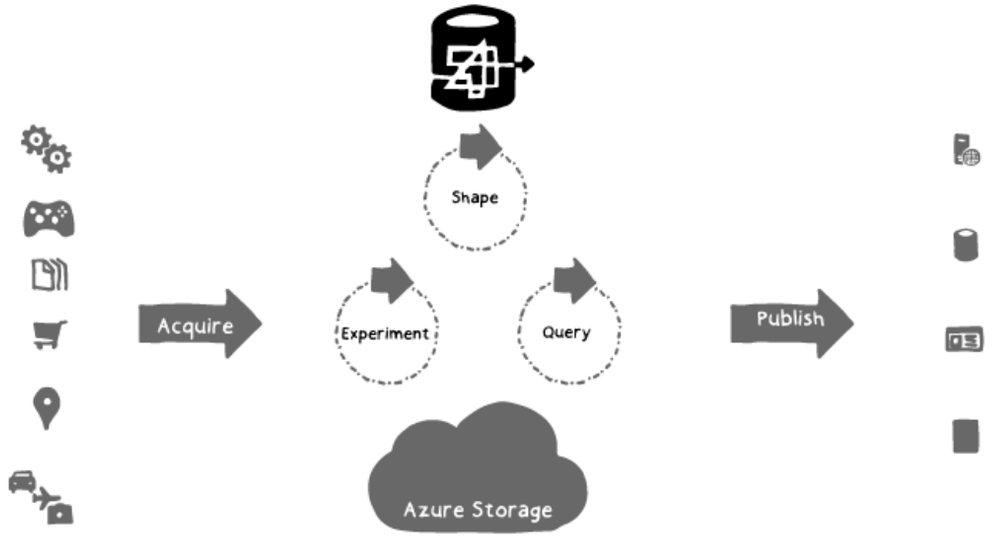
Data Analysis Options

- Azure Analysis Services:
  - **Managed, enterprise-grade data modeling service**
  - **Compatible with a range of features of SQL Server Analysis Services Enterprise Edition:**
    - **Bi-directional relationships**
    - **Tabular data models**
    - **Row-level security**
    - **Translations**
    - **Partitions**
    - **In-memory mode**
    - **DirectQuery mode**
  - **Accessible to client apps:**
    - **Power BI**
    - **Excel**
    - **Reporting services**



HDInsight

- Managed Apache Hadoop service in Azure:
  - **Implements Hadoop components from Hortonworks Data Platform (HDP)**
  - **Supports the most popular open-source frameworks, including:**
    - **Hadoop**
    - **Spark**
    - **Hive**
    - **LLAP**
    - **Kafka**
    - **Storm**
    - **R Server**

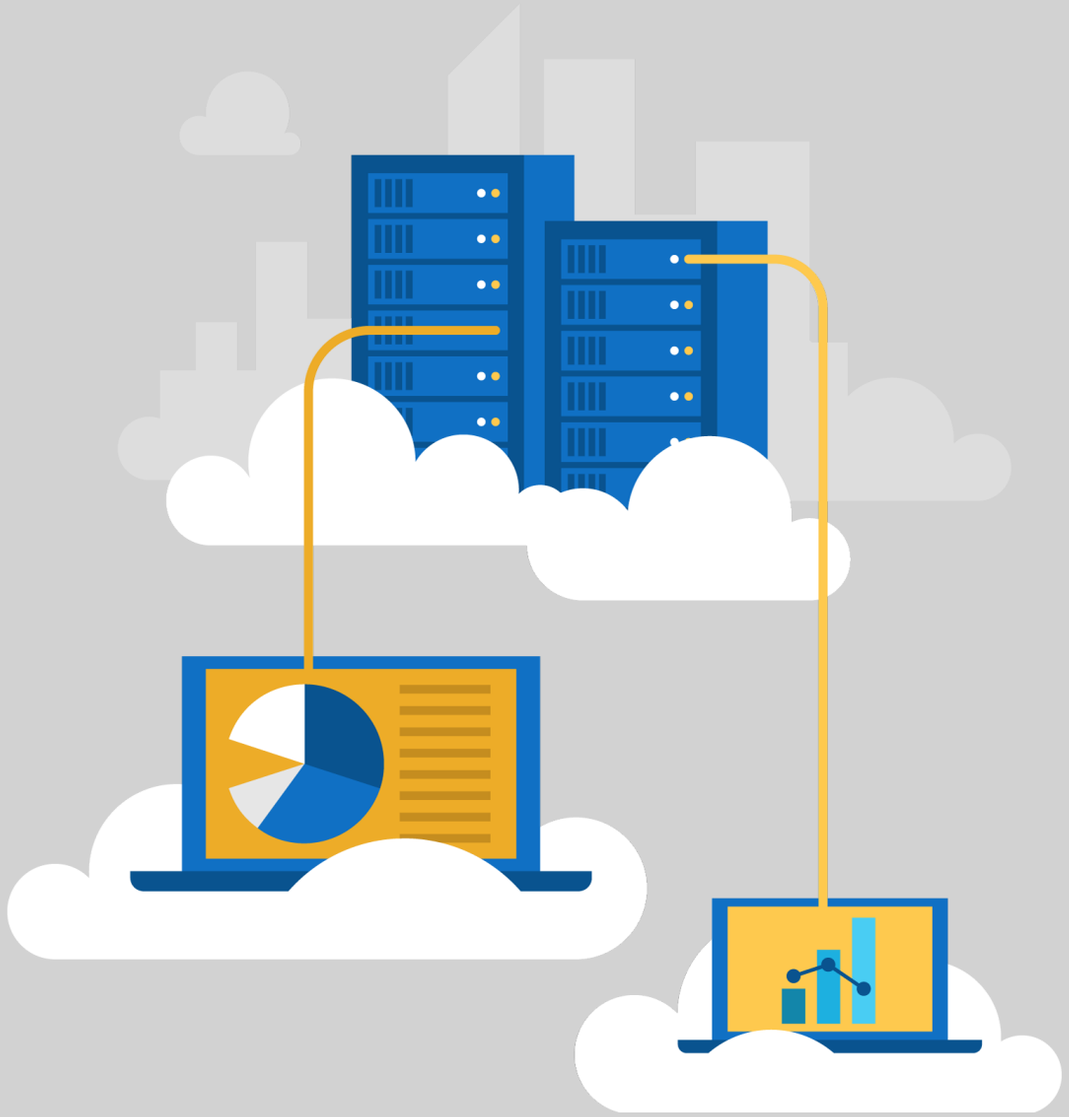


Azure Data Catalog

- A managed data source discovery Azure service:
  - **Helps discover, understand, and consume data sources**
  - **Includes a crowdsourcing model of metadata and annotations**
  - **Constitutes a central place for knowledge sharing and data culture building**
  - **Operates based on registration of data sources:**
    - **Registered data remains in its original location**
    - **Catalog stores metadata and a reference to the data source location**
    - **Metadata is indexed to facilitate discovery through searches**
    - **Users can enrich metadata through annotations (descriptions, tags, documentation)**

Microsoft

Design a Data Protection Strategy



Azure Security Spectrum

Identity & access	Encryption	Secure networking	Partner solutions	Unified security management
• RBAC • Strong Authentication • Monitoring and Alerting	• Encryption Key Management • Encryption at Rest and In Transit	• Virtual Networks • Traffic Rules • Secure Connectivity	• Antimalware • Network Appliances • Encryption • Monitoring • Application Security • Authentication	• Security Policy • Monitoring • Recommendations • Threat Detection

Encryption

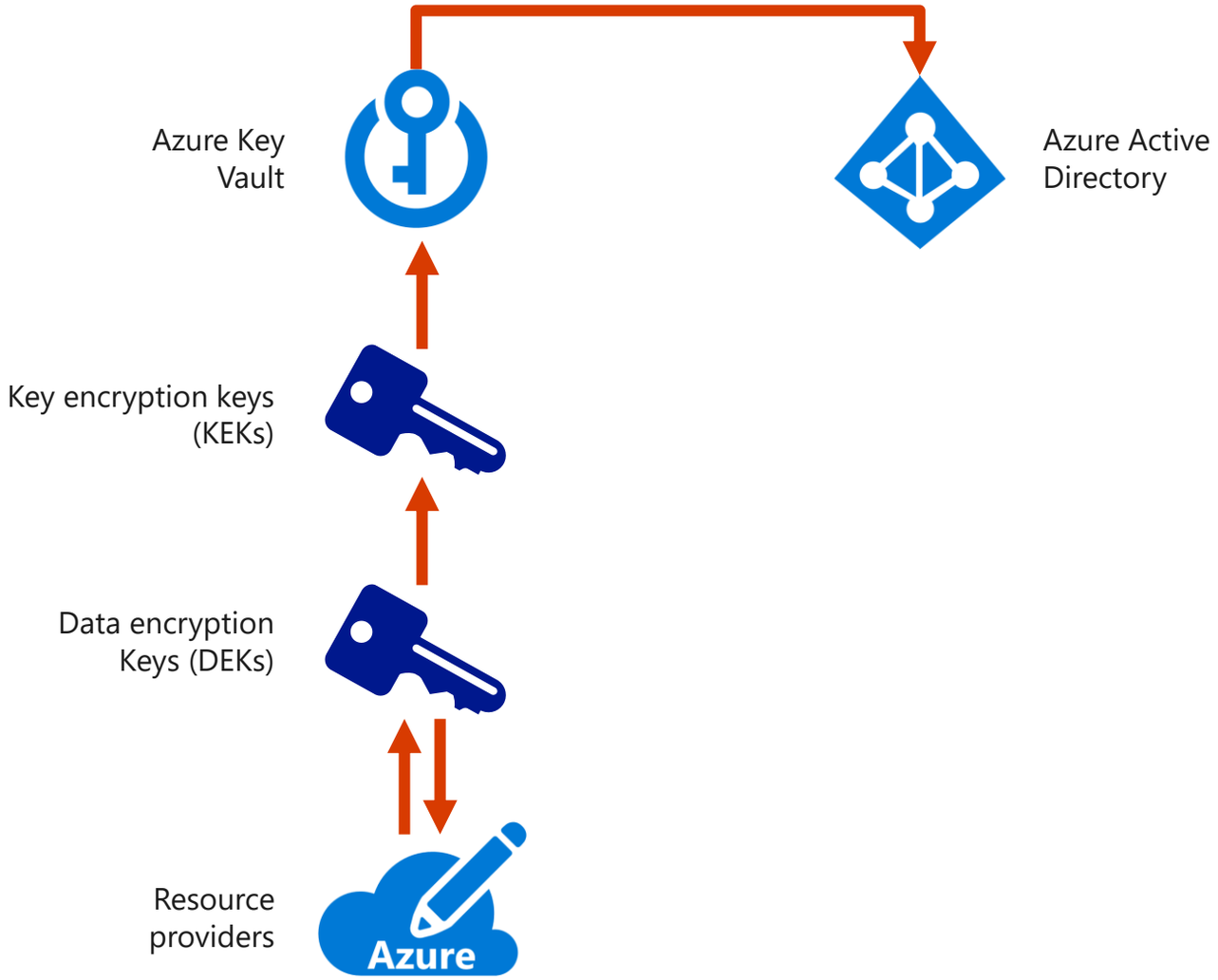
- Encryption
  - Process of translating plain text data (**plaintext**) into something that appears to be random and meaningless (**ciphertext**)
- Decryption
  - Process of converting ciphertext back to plaintext

- Symmetric encryption is used to encrypt more than a small amount of data
  - A symmetric key is used to encrypt the data
  - The same key must be used to decrypt the data

Encryption at rest

- Encryption (or encoding) of data when it is persisted
  - Very common security requirement to encrypt data with a secret encryption key anytime it is persisted to disk
  - Prevents attackers from accessing sensitive data when they have full access to a server's machine, storage or drives
- Encryption at rest design in Azure uses symmetric encryption:
  - A symmetric encryption key is used to encrypt data as it is written to storage
  - The same encryption key is used to decrypt that data as it is readied for use in memory
  - Data may be partitioned, and different keys may be used for each partition
  - Keys are stored in a security-enhanced location with access control policies
  - Data encryption keys are often encrypted with asymmetric encryption to further limit access

Encryption at rest in Azure



Encryption at Rest for Azure Services

- Azure Storage
  - Data is automatically encrypted server-side for all Storage services (Blob, Queue, Table, Files)
  - Keys are managed by the service
- Azure SQL Database
  - Transparent Data Encryption (TDE) is enabled by default on all new databases
  - Supports customer-managed 2048-bit keys in Azure Key Vault
- Azure Cosmos DB
  - Backups and media attachments are stored in Blob storage
  - Databases are automatically encrypted on SSDs

Transparent Data Encryption

- Encrypts database, backups and logs at rest and in flight
- Requires little to no code changes
  - Only requires a modification to connection string in most scenarios
- Can be used with many third-party SQL tools already in the market
- Supported in Azure SQL Database and SQL Server

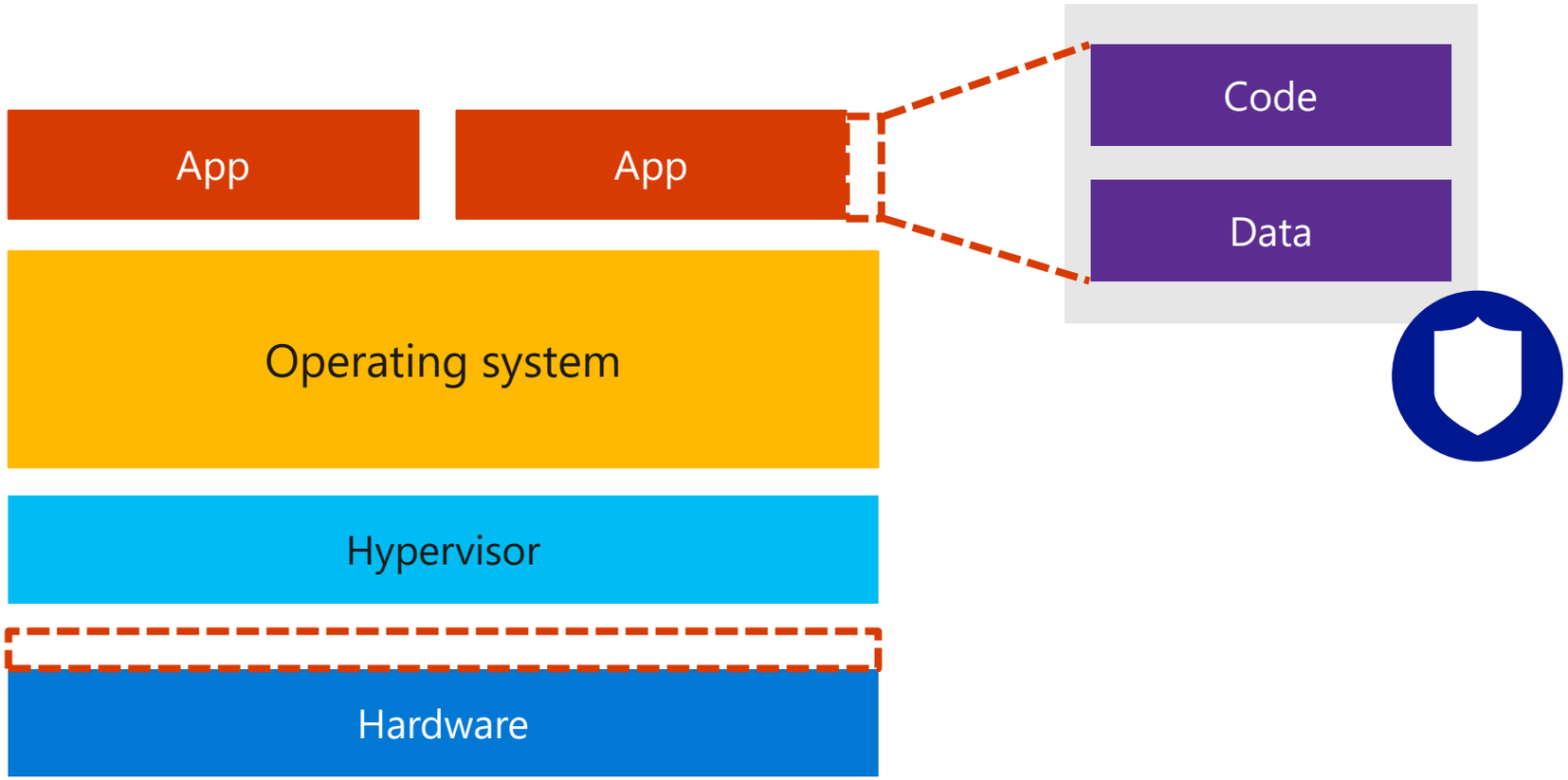
Always Encrypted

- Fully transparent encryption
  - Encrypted data can be queries
- Encrypts data at rest, in flight and in memory
- Requires the use of specific drivers
  - In most applications, requires some rewrite
  - Not always compatible with every third-party tool

Azure confidential computing

- A collection of features across a broad spectrum of Azure services designed to encrypt data in use
- Ideal for scenarios where data needs to be processed in the cloud
  - The services maintain encryption that prevents the data from being exposed as plaintext

Azure confidential computing



Trusted Execution Environments

- Data is processed within Trusted Execution Environments (TEEs)
  - Ensures that there is no way to view data or operations inside the TEE from the outside
  - If code is tampered with or altered, all operations are halted, and the environment is disabled
  - TEEs can be hardware-based, software-based
  - There are frameworks available to take advantage of TEEs
    - Example: Confidential Consortium Blockchain Framework
- TEEs are being developed by multiple groups
  - Intel (SGX)
  - Microsoft Research

Transport Layer Security and Secure Sockets Layer

- Cryptographic protocols that help provide communications security over a computer network
  - Secure Sockets Layer (SSL) encryption is the most commonly used method of helping secure data sent across the internet
  - Transport Layer Security (TLS) is the successor to SSL and was developed to address vulnerabilities in SSL.

- Azure services (not an exhaustive list) supporting SSL encryption:
  - Azure SQL Database
  - Azure Database for MySQL
  - Azure Storage
  - Azure Application Gateway
  - Azure App Service

TLS in Azure Storage

- SSL 1.0, 2.0 and 3.0 have been found to be vulnerable
  - These protocols have been prohibited by an Internet Engineering Task Force (IETF) RFC
- Many services have already moved from SSL to TLS 1.0
- TLS 1.0 also has known vulnerabilities
  - TLS 1.0 is insecure for using insecure block ciphers (Data Encryption Standard [DES] CBC and RC2 CBC) and an insecure stream cipher (RC4)

- Azure Storage recommends using TLS 1.2
  - TLS 1.0 and 1.1 are still supported by Azure Storage but not recommended

TLS 1.2 in .NET and PowerShell

.NET

```
System.Net.ServicePointManager.SecurityProtocol = System.Net.SecurityProtocolType.Tls12;
```

PowerShell

```
[System.Net.ServicePointManager]::SecurityProtocol = [System.Net.SecurityProtocolType]::Tls12;
```

Azure Key Vault

- Safeguard cryptographic keys and other secrets used by cloud apps and services
  - Increase security and control over keys and passwords
  - Applications have no direct access to keys
  - Use FIPS 140-2 Level 2 validated Hardware Security Modules (HSMs)
  - Create and import
    - Encryption keys
    - API Keys
    - Secrets
    - Passwords
    - SSL/TLS certificates

Azure Key Vault in Azure CLI

```
az group create --name SecurityGroup --location westus
```

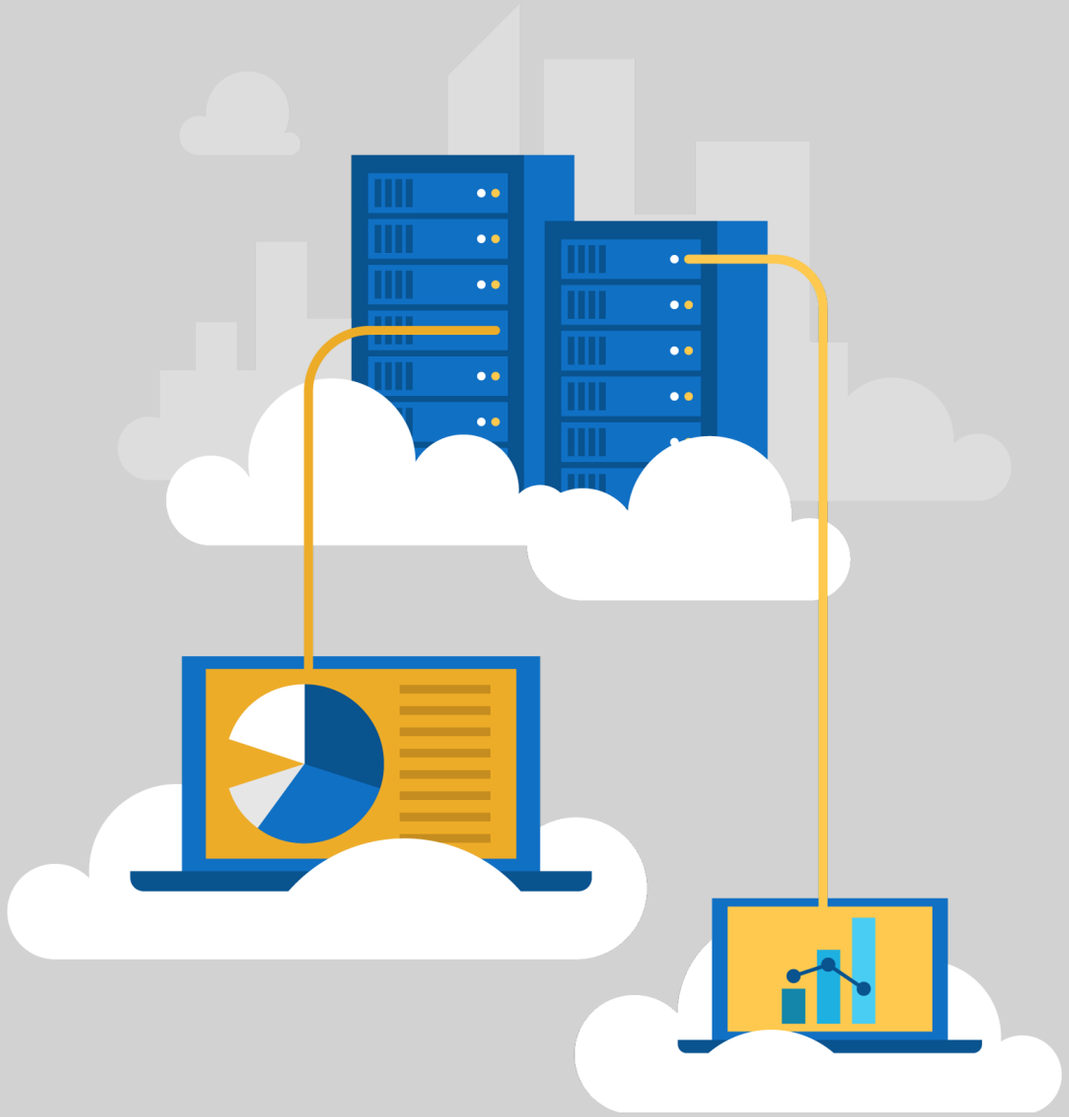
```
az keyvault create --name contosovault --resource-group SecurityGroup --location westus
```

```
az keyvault secret set --vault-name contosovault --name DatabasePassword --value 'Pa5w.rd’
```

```
az keyvault secret show --vault-name contosovault --name DatabasePassword
```

Microsoft

Design a Monitoring Strategy for the Data Platform



Azure Network Watcher

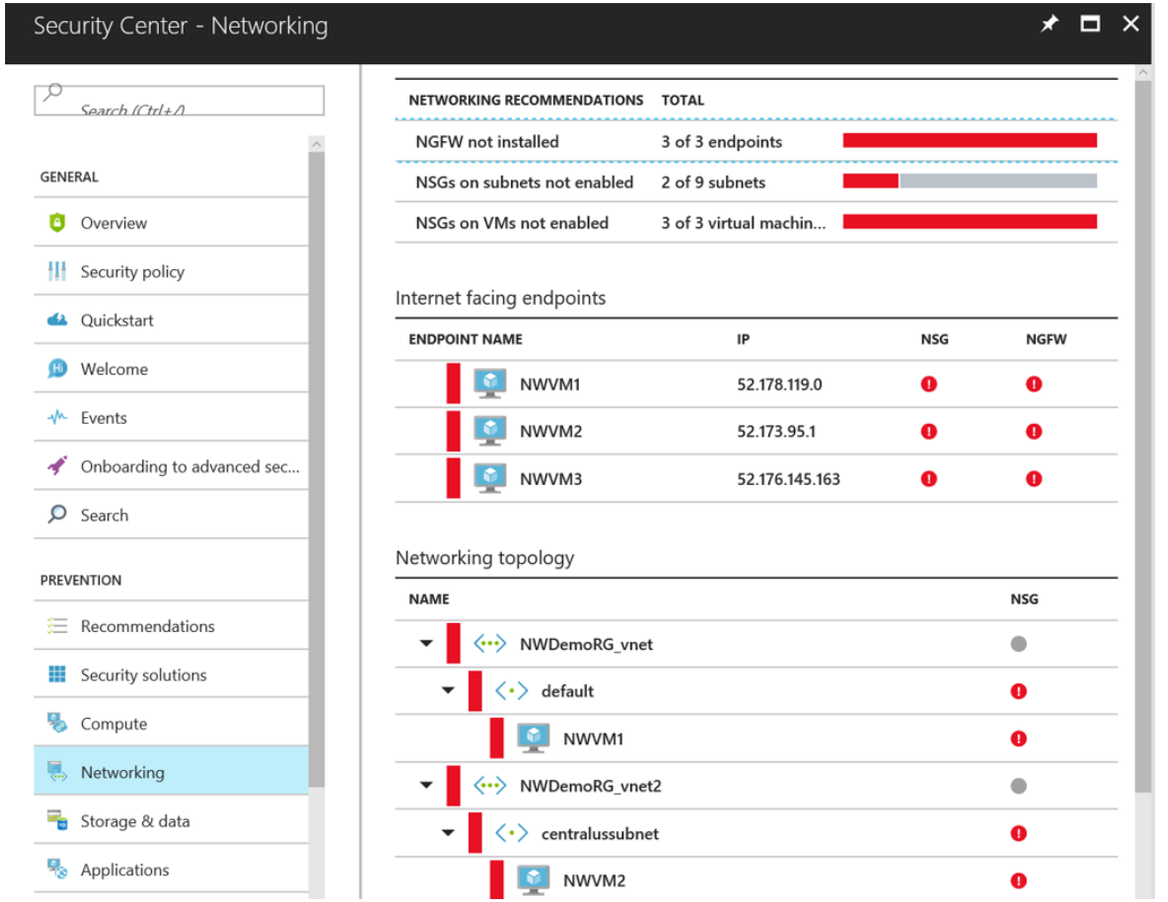
- **Topology**
- **Variable Packet Capture**
- **IP Flow Verify**
- **Next Hop**
- **Diagnostics Logging**
- **Security Group View**
- **NSG Flow Logging**
- **VPN Gateway Troubleshooting**
- **Network Subscription Limits**
- **Role Based Access Control**
- **Connectivity**

Network Monitor

- Azure facilitates network monitoring by using:
  - **Metrics (network-related performance counters)**
  - **Logging of operations performed as part of network configuration:**
    - **Available in the Azure portal and via Power BI**
  - **Logging of network events generated by network resources**
    - **Events can be stored in Azure Storage**
    - **Events can be sent to Event Hub or Log Analytics**
  - **Troubleshooting blades in the Azure portal, available for:**
    - **ExpressRoute**
    - **VPN Gateway**
    - **Application Gateway**
    - **Network Security Groups**
    - **User-Defined Routes**
    - **Azure DNS**
    - **Azure Load Balancer**
    - **Traffic Manager**

Azure Security Center

- Provides unified security management and threat protection:
  - **Supports workloads:**
    - **In Azure**
    - **On-premises**
    - **Hosted in other public clouds**
  - **Increases visibility and control**
  - **Offers active defenses**
  - **Implements intelligent detection**



Azure Monitor & Diagnostics

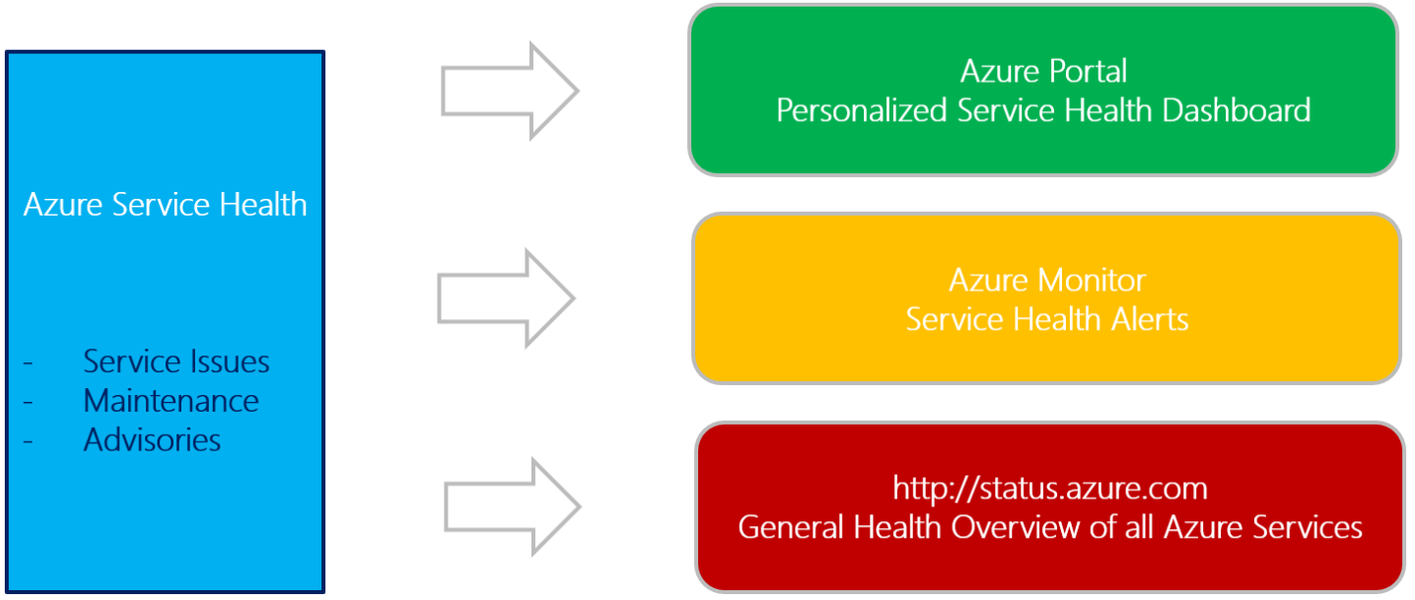
- Azure Monitor is a core component of the Azure monitoring strategy:
  - **Relies on metrics, which:**
    - **Represent telemetry data emitted by Azure resources**
    - **Are available by default**
    - **Are generated with 1 minute frequency**
    - **Are, by default, retained for 30 days**
    - **Include support for name-value pair attributes known as dimensions (limited to some metrics)**
  - **Consumes metrics in order to:**
    - **Track performance of resources**
    - **Generate alerts**
    - **Trigger automated actions, such as autoscaling or launching a runbook**
    - **Perform advanced analytics or reporting**
    - **Archive health history of a resource for compliance or auditing purposes**

Azure Advisor

- Personalized cloud service for optimizing Azure deployments:
  - **Analyzes resource configuration and usage telemetry**
  - **Offers recommendations grouped in four categories:**
    - **High availability: To ensure and improve the continuity of your business-critical applications.**
    - **Security: To detect threats and vulnerabilities that might lead to security breaches.**
    - **Performance: To improve the speed of your applications.**
    - **Cost: To optimize and reduce your overall Azure spending.**

Azure Service Health

- Offers assistance with events affecting availability of Azure services:
  - **Service issues - problems with the Azure services that affect you right now.**
  - **Planned maintenance – events that affect the availability of your services in the future.**
  - **Health advisories - changes affecting Azure services that require your attention – e.g.:**
    - **Azure features are deprecated**
    - **you exceed a usage quota**



Operations Management Suite – Log Analytics

- Log Analytics is a core component of the Azure monitoring strategy:
  - **Supports workloads:**
    - **In Azure**
    - **On-premises**
    - **Hosted in other public clouds**
  - **Collects data generated by resources and by other monitoring tools**
  - **Performs comprehensive analysis with correlation across data sources**
  - **Provides Azure portal-based interface, which supports:**
    - **Log searches**
    - **Dashboards**
  - **Offers extensibility through a range of solutions, geared towards specific functionality**

Application Insights

Extensible Application Performance Monitoring (APM) service:

Intended primarily for development and DevOps purposes

Supports apps:

In Azure

On-premises

Hosted in other public clouds

Provides:

Monitoring of live web apps

Detection of performance anomalies

App diagnostics

Usage analysis

Supports a wide range of development frameworks, including .NET, Node.js, and J2EE

Supports mobile apps by integrating with Visual Studio App Center and HockeyApp

Questions?

Microsoft

Homework Assignment

https://aka.ms/az301asg

Microsoft Azure

Open Mic