



Frederic Gisbert Cloud Solution Architect Azure Data & Analytics

# Training session DP900



Ali Bouhaddou Cloud Solution Architect Azure Data & Analytics





## DP-900 Partner Academy

### **Agenda**

### **Describe non-relational data workloads**

- describe the characteristics of non-relational data
- describe the types of non-relational and NoSQL data
- recommend the correct data store
- determine when to use non-relational data
- Describe non-relational data offerings on Azure
- identify Azure data services for non-relational workloads
- describe Azure Cosmos DB APIs
- describe Azure Table storage
- describe Azure Blob storage
- describe Azure File storage

### Identify basic management tasks for non-relational data

- describe provisioning and deployment of non-relational data services
- describe method for deployment including the Azure portal, Azure Resource Manager templates, Azure PowerShell, and the Azure command-line interface (CLI)
- identify data security components (e.g., firewall, authentication, encryption)
- identify basic connectivity issues (e.g., accessing from on-premises, access with Azure VNets, access from Internet, authentication, firewalls)
- identify management tools for non-relational data



## Describe nonrelational data workloads

- describe the characteristics of non-relational data
- describe the types of non-relational and NoSQL data
- recommend the correct data store
- determine when to use non-relational data
- Describe non-relational data offerings on Azure
- identify Azure data services for non-relational workloads
- describe Azure Cosmos DB APIs
- describe Azure Table storage
- describe Azure Blob storage
- describe Azure File storage

## How to choose between relational (SQL) and nonrelational (NoSQL) databases

	NOSQL OR NONRELATIONAL	SQL OR RELATIONAL
BEST FOR:	<ul> <li>Handling large, unrelated, indeterminate, or rapidly changing data.</li> <li>Schema-agnostic data or schema dictated by the app.</li> <li>Apps where performance and availability are more important than strong consistency.</li> <li>Always-on apps that serve users around the world.</li> </ul>	<ul> <li>Handling data that is relational and has logical and discrete requirements that can be identified in advance.</li> <li>Schema that must be maintained and kept in sync between the app and database.</li> <li>Legacy systems built for relational structures.</li> <li>Apps requiring complex querying or multi-row transactions.</li> </ul>
SCENARIOS:	<ul> <li>•Mobile apps.</li> <li>•Real-time analytics.</li> <li>•Content management.</li> <li>•Personalization.</li> <li>•IoT apps.</li> <li>•Database migration.</li> </ul>	<ul> <li>Accounting, finance, and banking systems.</li> <li>Inventory management systems.</li> <li>Transaction management systems.</li> </ul>
SCALE:	•Scales data horizontally by sharding across servers.	•Scales data vertically by increasing server load.
DATA MODEL:	<ul> <li>Database types: key-value, document, columnar, and graph databases.</li> <li>Stores data depending on database type.</li> </ul>	<ul> <li>Database type: tables of rows, grouped into relations.</li> <li>Uses Structured Query Language (SQL).</li> <li>Stores data as rows in tables; related data stored separately and joined for complex queries.</li> </ul>

# Document data stores

### Relevant Azure service:

Azure Cosmos DB (SQL API)







A document data store manages a set of named string fields and object data values in an entity that's referred to as a document.

These data stores
typically store data in the
form of JSON
documents. Each field
value could be a scalar
item, such as a number,
or a compound element,
such as a list or a parentchild collection.

The data in the fields of a document can be encoded in various ways, including XML, YAML, JSON, BSON, or even stored as plain text. The fields within documents are exposed to the storage management system, enabling an application to query and filter data by using the values in these fields.

Key	Document	
1001	<pre>{     "CustomerID": 99,     "OrderItems": [</pre>	
1002	{     "CustomerID": 220,     "OrderItems": [	

# Columnar data stores

### Relevant Azure service:

- •Cosmos DB Cassandra API
- HBase in HDInsight







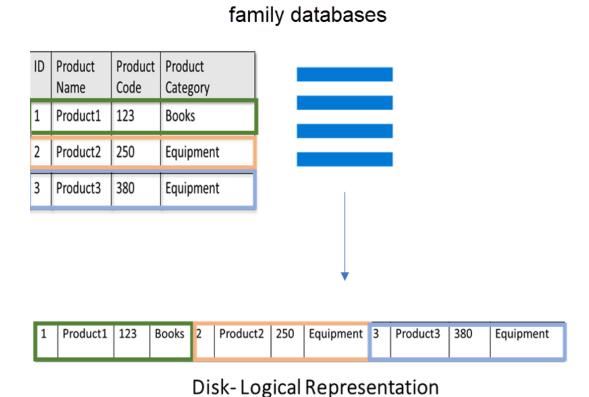
CustomerID	Column Family: Identity
001	First name: Mu Bae Last name: Min
002	First name: Francisco Last name: Vila Nova Suffix Jr.
003	First name: Lena Last name: Adamcyz Title: Dr.

CustomerID	Column Family: Contact Info
001	Phone number: 555-0100 Email: someone@example.com
002	Email: vilanova@contoso.com
003	Phone number: 555-0120

A columnar or columnfamily data store organizes data into columns and rows. Columns are divided into groups known as column families.

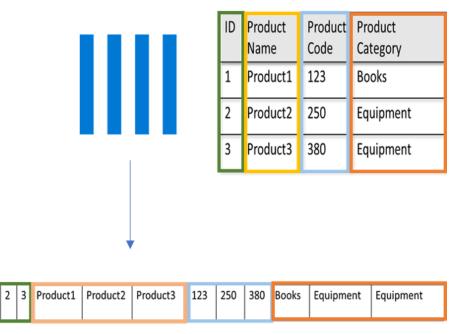
Each column family holds a set of columns that are logically related and are typically retrieved or manipulated as a unit On disk, all of the columns within a column family are stored together in the same file, with a specific number of rows in each file. With large data sets, this approach creates a performance benefit by reducing the amount of data that needs to be read from disk when only a few columns are queried together at a time.

### NB: wide-column stores Vs columnar databases



Wide-column or column

Columnar databases



**Disk-Logical Representation** 

## **Key/value data** stores

### Relevant Azure services

- Azure Cosmos DB Table API
- Azure Cache for Redis
- Azure Table Storage

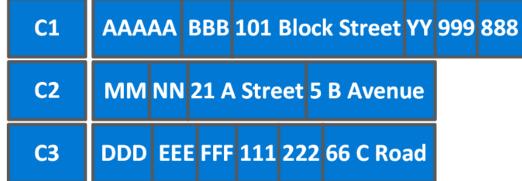


key/value store is essentially a large hash table.

You associate each data value with a unique key, and the key/value store uses this key to store the data by using an appropriate hashing function. The hashing function is selected to provide an even distribution of hashed keys across the data storage.

Kev (Customer ID)

Value (Customer Data)







Most key/value stores only support simple query, insert, and delete operations.

To modify a value (either partially or completely), an application must overwrite the existing data for the entire value. In most implementations, reading or writing a single value is an atomic operation. If the value is large, writing may take some time.

Key/value stores are highly optimized for applications performing simple lookups using the value of the key, or by a range of keys, but are less suitable for systems that need to query data across different tables of keys/values, such

as joining data across multiple tables.

## **Graph data stores**

### Relevant Azure service:

Azure Cosmos DB Graph API

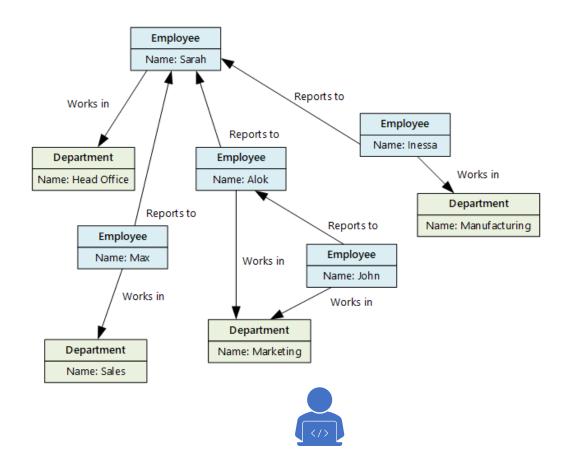


graph data store manages two types of information, nodes and edges.

<<Nodes>> represent entities, and <<edges>> specify the relationships between these entities.

**Both nodes and edges can have properties** that provide information about that node or edge, similar to columns in a table.

**Edges can also have a direction** indicating the nature of the relationship.





The purpose of a graph data store is to allow an application to efficiently perform queries that traverse the network of nodes and edges, and to analyze the relationships between entities.

For large graphs with lots of entities and relationships, you can perform complex analyses quickly.

Many graph databases provide a query language that you can use to traverse a network of relationships efficiently.

## Time series data stores

### Relevant Azure services: Azure DataExplorer

- **Azure DataExplorer**
- **Azure Time Series Insights**
- OpenTSDB with HBase on HDInsight





Time series data stores also handle out-of-order and late-arriving data, automatic indexing of data points, and optimizations for queries described in terms of

windows of time

timestamp	deviceid	value
2017-01-05T08:00:00.123	1	90.0
2017-01-05T08:00:01.225	2	75.0
2017-01-05T08:01:01.525	2	78.0



This last feature enables queries to run across millions of data points and multiple data streams quickly, in order to support time series visualizations, which is a common way that time series data is consumed

Time series data is a set of values organized by time, and a time series data store is optimized for this type of data.

Time series data stores must support a very high number of writes, as they typically collect large amounts of data in real time from a large number of sources.

# **Object data stores**

### Relevant Azure services:

- Azure Blob Storage
- Azure Data Lake Store
- Azure File Storage

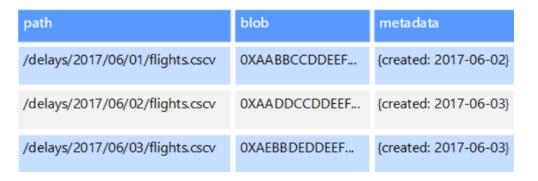
/	

Object data stores are optimized for storing and retrieving large binary objects or blobs such as images, text files, video and audio streams, large application data objects and documents, and virtual machine disk images.



An object consists of the stored data, some metadata, and a unique ID for accessing the object.

Object stores are designed to support files that are individually very large, as well provide large amounts of total storage to manage all files.





Some object data stores replicate a given blob across multiple server nodes, which enables fast parallel reads.

This process, in turn, enables the scale-out querying of data contained in large files, because multiple processes, typically running on different servers, can each query the large data file simultaneously.

# **External index** data stores

### Relevant Azure service:

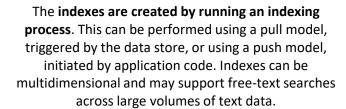
Azure Cognitive Search



External index data stores provide the ability to search for information held in other data stores and services.

An external index acts as a secondary index for any data store, and can be used to index massive volumes of data and provide near real-time access to these indexes.









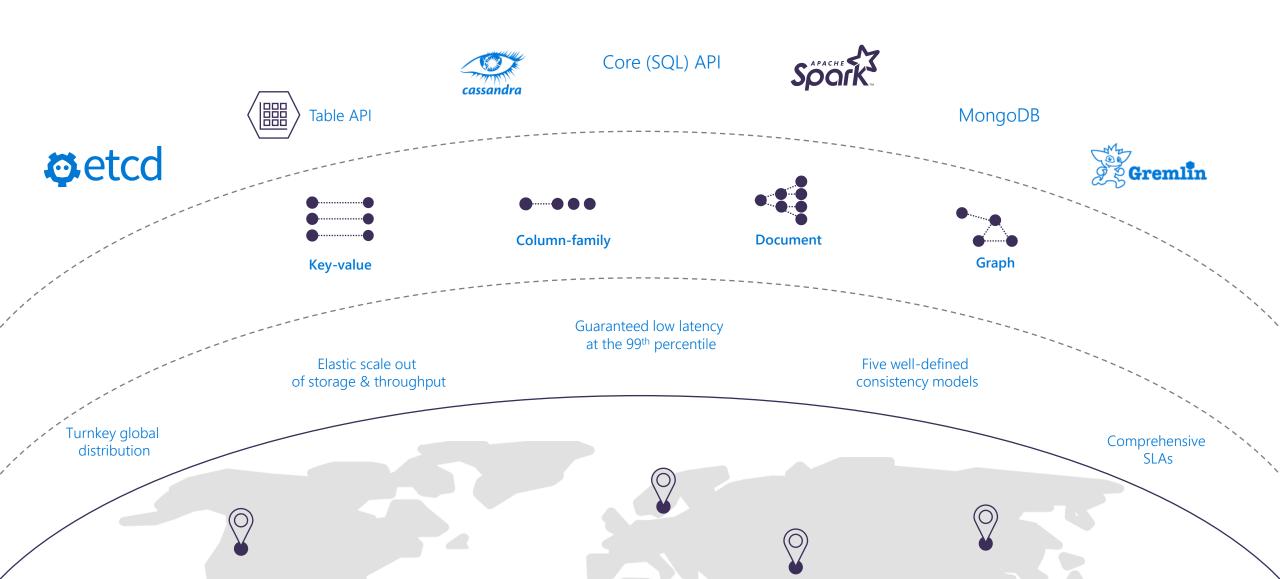
External index data stores are often used to support full text and web-based search. In these cases, searching can be exact or fuzzy. A fuzzy search finds documents that match a set of terms and calculates how closely they match.

Some external indexes also support **linguistic analysis** that can return matches based on synonyms, genre expansions (for example, matching "dogs" to "pets"), and stemming (for example, searching for "run" also matches "ran" and "running").

## Describe nonrelational data workloads

- describe the characteristics of non-relational data
- describe the types of non-relational and NoSQL data
- recommend the correct data store
- determine when to use non-relational data
- Describe non-relational data offerings on Azure
- identify Azure data services for non-relational workloads
- describe Azure Cosmos DB APIs
- describe Azure Table storage
- describe Azure Blob storage
- describe Azure File storage

### What is...AZURE COSMOS DB?



## AZURE COSMOS DB FAST NOSQL DATABASE WITH OPEN APIS FOR ANY SCALE

Azure Cosmos DB is a fully managed NoSQL database for modern app development with SLA-backed speed and availability, automatic and instant scalability, and open source APIs for MongoDB, Cassandra, and other NoSQL engines.



Guaranteed speed at any scale

Gain unparalleled SLAbacked speed and throughput, fast global access, and instant elasticity.



Simplified application development

Build fast with open source APIs, multiple SDKs, schemaless data, and no-ETL analytics over operational data.



Mission-critical ready

Guarantee business continuity, 99.999% availability, and enterprise-level security for every application.



Fully managed and costeffective

End-to-end database management with serverless and automatic scaling matching your application and TCO needs.

## Top use cases for Azure Cosmos DB

Use Case	Use Case Details	Why Azure Cosmos DB?	Alternatives	Optimization Tips	Customer Examples
Real-time telemetry	<ol> <li>loT</li> <li>Logistics (track assets/inventory)</li> <li>Smart cars (telematics)</li> <li>Smart Cities</li> <li>Smart Buildings</li> <li>User/Customer</li> <li>Track digital actions (clicks; app events)</li> </ol>	<ul> <li>✓ Elastic, unlimited throughput (scale)</li> <li>✓ Can easily ingest large volumes of data from write-heavy scenarios</li> <li>✓ Real-time queries backed by &lt;10 ms low-latency SLA</li> <li>✓ Auto-indexing of semi-structured data</li> </ul>	<ul> <li>Time-consuming and complex connection, integration, and management bundle of OSS tech (e.g., Redis, Cassandra, etc)</li> <li>Maintain performance by actively managing, overallocating, and over-paying for cluster</li> </ul>	<ol> <li>Use Device ID or User ID as partition key</li> <li>Customize index policy (vs autoindex all data)</li> <li>Materialize View design pattern to optimize aggregates</li> <li>Use local region for read/write data in SDK connection policy</li> </ol>	<ul> <li>Teijen</li> <li>Mercedes Benz</li> <li>Johnson Controls</li> <li>Exxon Mobil</li> <li>SitePro</li> </ul>
Real-time recommendatio ns and 360- degree customer view	Rapid response recommendations to enhance user experience (e.g., Netflix recommendations; e- commerce product suggestions)	<ul> <li>✓ Spark integration offers ML and big data processing</li> <li>✓ &lt;10 ms low-latency SLA</li> </ul>	<ul> <li>Build and maintain custom solutions; manage data integration (POS, CRM, etc.)</li> <li>Static, non-responsive, or delayed recommendations based on stale data</li> <li>Limited personalization (e.g., missing location, daypart, etc.)</li> </ul>	<ol> <li>Custom configuration for Spark connector (e.g., batch size)</li> <li>Scale up RU/s for batch processing then immediately scale down for steady-state</li> <li>Use Product ID or User ID as partition key</li> </ol>	<ul><li>ASOS</li><li>Siemens Healthineers</li><li>PureFacts</li></ul>
Mission-critical or geo- distributed applications	Deliver consistent real-time, always-on user experience anywhere in the world  Shopping cart/checkout  IoT device registry	<ul> <li>✓ One-click, fully-managed data replication/distribution around the word</li> <li>✓ &lt;10 ms low-latency SLA</li> <li>✓ 99.999% high-availability SLA</li> <li>✓ Available in all Azure regions</li> </ul>	<ul> <li>Set up and maintain infrastructure, networks, and databases globally</li> <li>Manage data replication manually</li> <li>Time, complexity, and cost to scale capacity and reach</li> </ul>	<ol> <li>Use Session or Bounded Staleness pre-configured data consistency</li> <li>Use local region for read/write data in SDK connection policy</li> <li>Use multi-master writes/replication in production (not dev/test)</li> </ol>	<ul><li>Walmart/<u>Jet.com</u></li><li>Citrix</li><li><u>Chipotle Mexican</u> <u>Grill</u></li></ul>
NoSQL migration	Move from self-managed NoSQL (e.g., MongoDB, Cassandra) to PaaS solution and stop managing clusters.	<ul> <li>✓ Full PaaS offering that handles all database/cluster management</li> <li>✓ APIs for MongoDB, Cassandra, and other NoSQL enable migration with minimal code changes</li> </ul>	<ul> <li>Create, manage, and tune clusters</li> <li>Manage upgrades, patching, etc</li> <li>24-7 internal team for monitoring, site reliability, and performance optimization</li> </ul>	<ol> <li>Use Azure Data Migration Service to migrate to API for MongoDB</li> <li>Use Azure Data Factory to migrate to SQL (core) API</li> <li>Use Striim (or other preferred partner) for any migration</li> </ol>	<ul><li>Bentley Systems</li><li>Symantec (Norton)</li></ul>

## When should the customer use Azure Cosmos DB?

One or more of these requirements...

· Flexible schema

· Sensor data may change over time



- Scalability
  - · Especially for writes

- Our use case needs high #writes/sec
  - · Cosmos guarantees throughput for any workload



 Cosmos guarantees <10ms at P99 for reads, writes of 1kb



· Can use as fast cache for key/val lookups

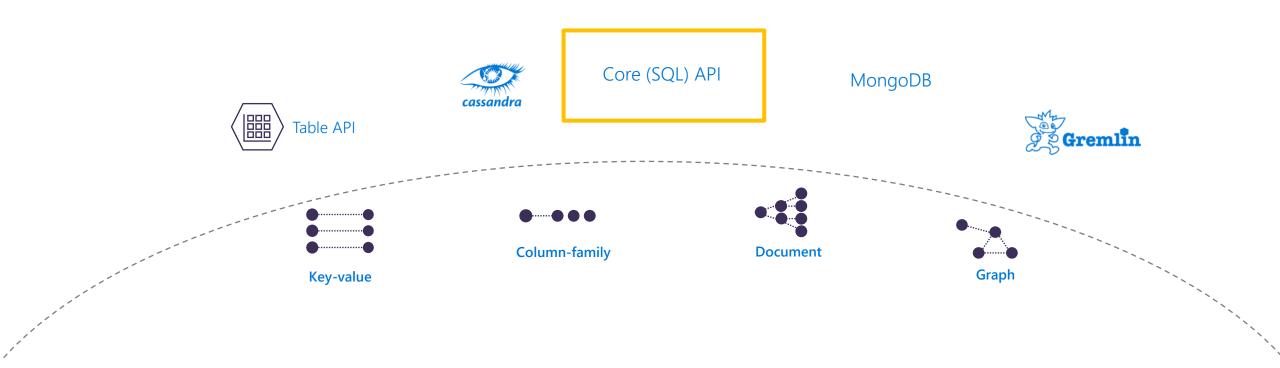
High availability

• 99.999% HA with 2+ regions



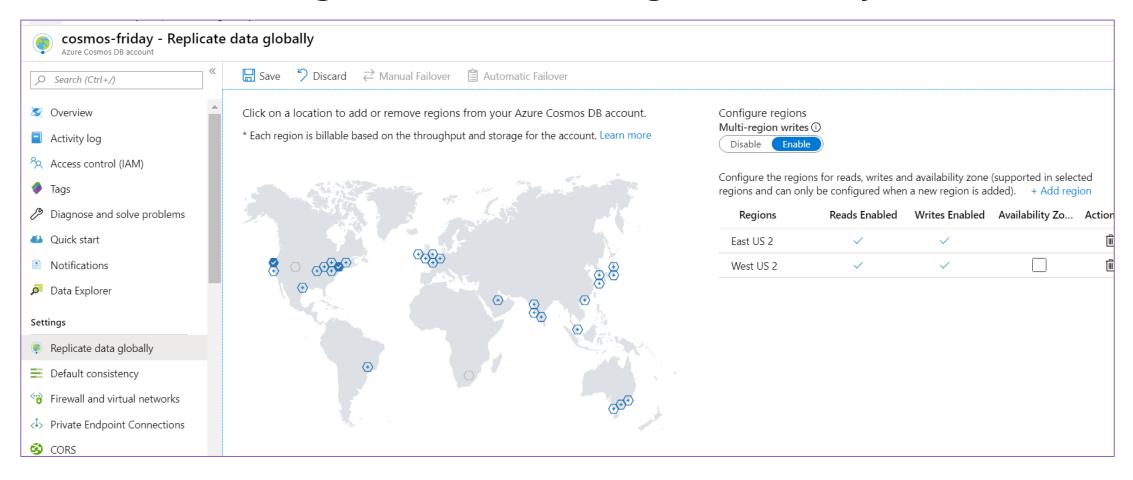
## Which API should the customer use for their app?

- For all new workloads Core (SQL) API
  - Best developer experience Cosmos builds interface, service, and SDKs
  - Gremlin API for graph data
- Other APIs for migration or DB platform preference/requirement

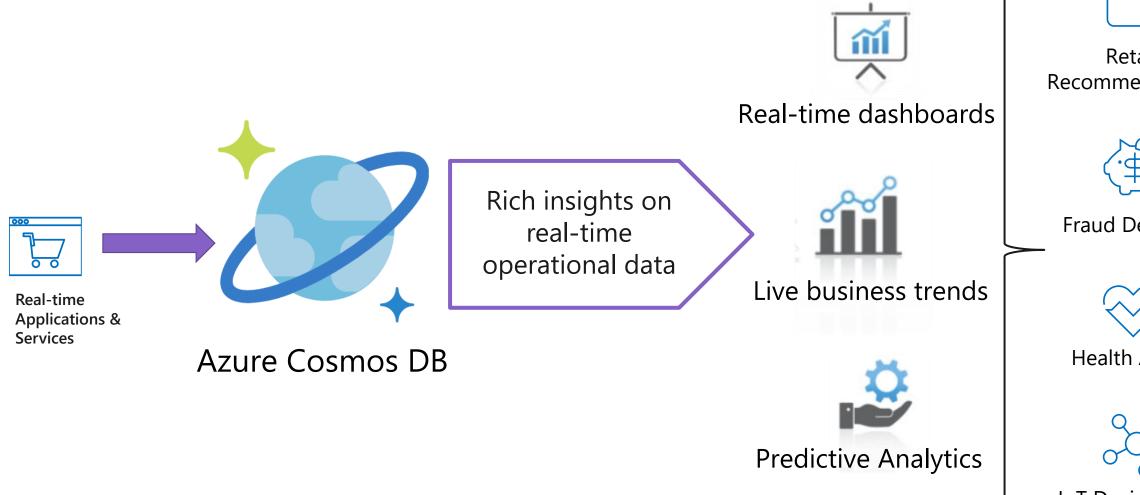


## Production readiness checklist

· Add a second region for 99.999% High Availability



## **Azure Cosmos DB + Real-time analytics**





Retail Recommendations



Fraud Detection



**Health Alerts** 

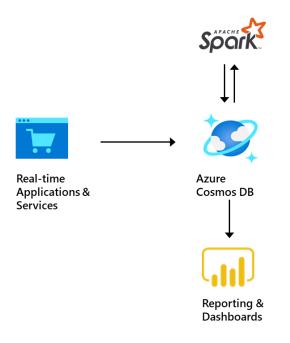


**IoT Device Faults** 

Transform your business with actionable insights

## **Azure Cosmos DB analytics : Options**

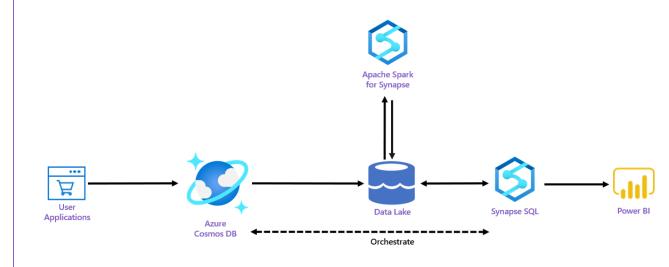
## Run OLTP & OLAP workloads on the same database



HUGE performance impact on the OLTP workloads at scale

As volume of operational data increases, latency of analytical queries increases and more resource intensive

### **Separating OLTP & OLAP**



Ingest data periodically from Azure Cosmos DB to Data Lake

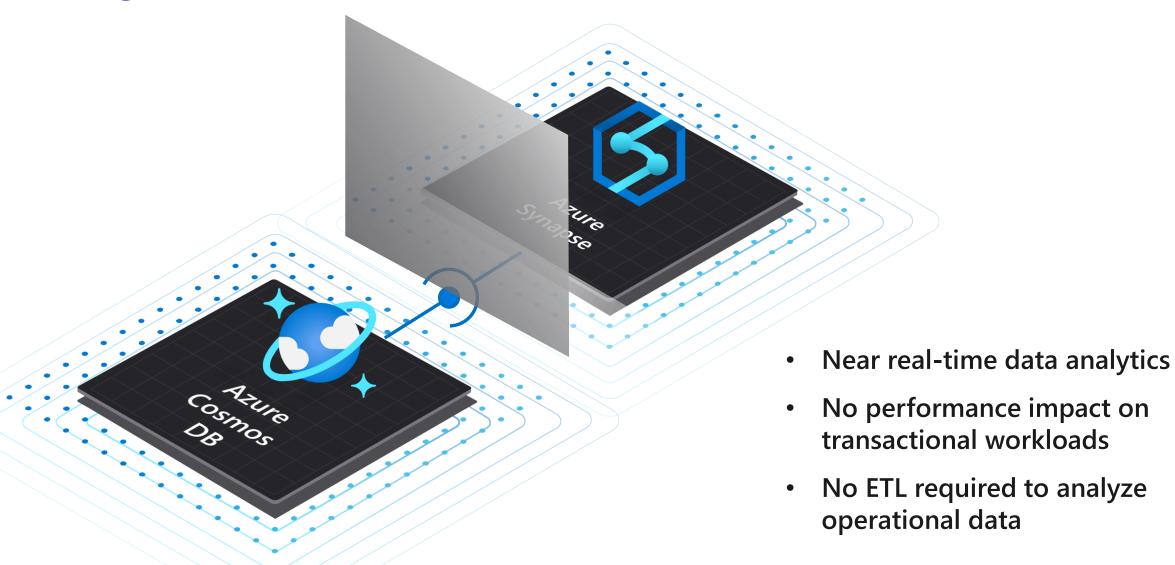
Delayed insights on data

Manage data formats and storage layer for analytics

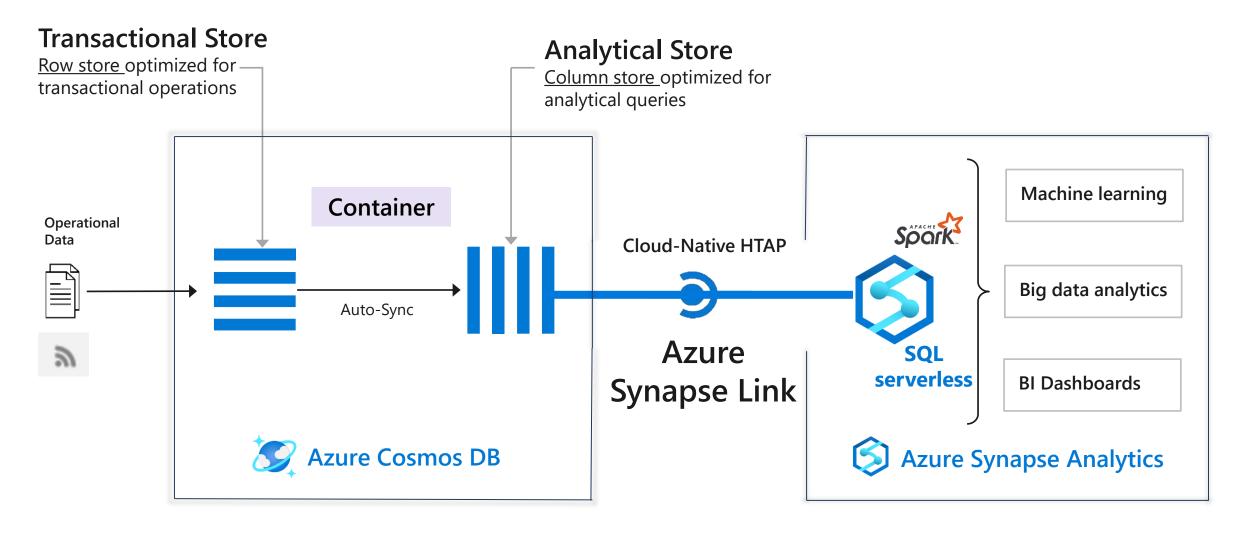


## **Azure Synapse Link for Azure Cosmos DB**

Breaking down the barrier between OLTP & OLAP



## **Azure Synapse Link: How it works?**

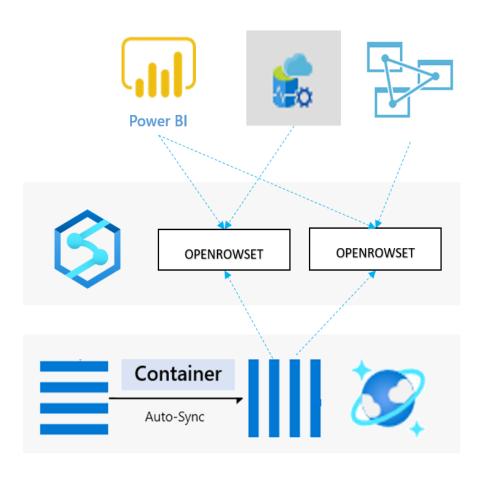


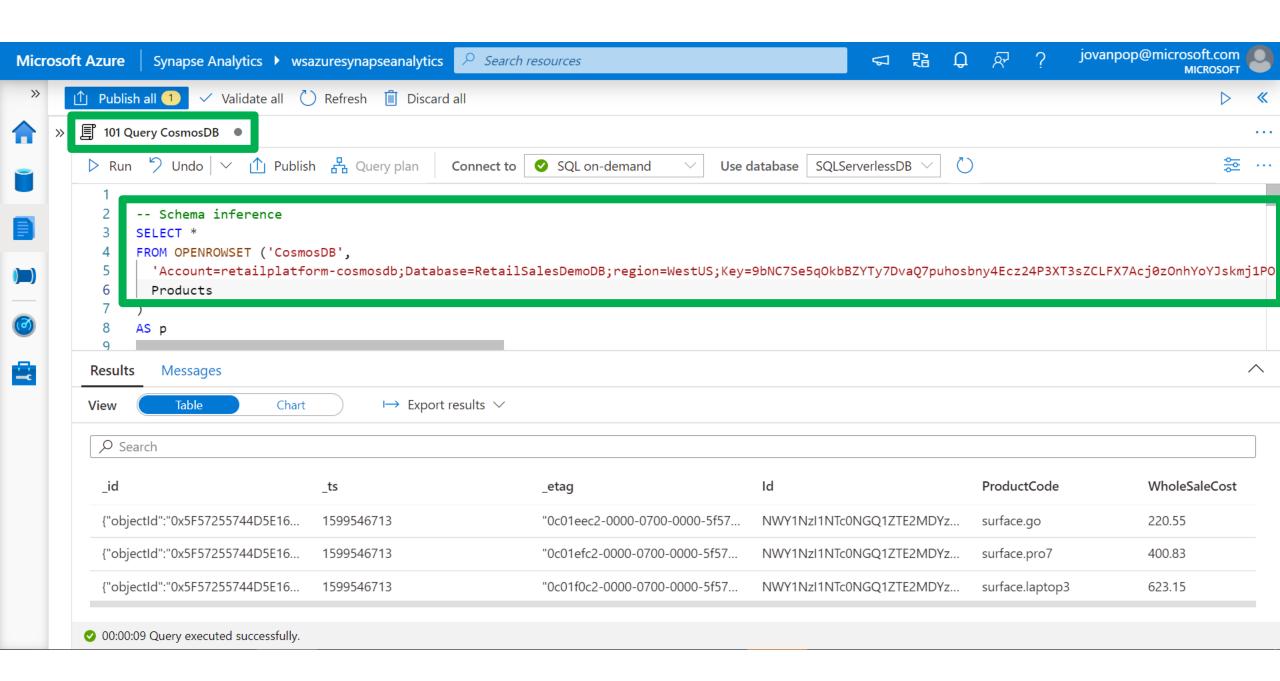
Generate near real-time insights on your operational data

## Azure Synapse Link: Query Azure Cosmos DB Data in Synapse SQL serverless



- Run analytical T-SQL queries:
   Query Azure Cosmos DB data in place, in seconds
- Build rich near real-time dashboards:
   Using Power BI integrated in Azure Synapse
- ➤ Use wider range of *BI and ad-hoc querying* tools: Integrated using T-SQL interface
- Build a logical data warehouse:
   Analyze unified view of data across Azure Cosmos DB,
   Azure Data Lake Storage & Azure Blob Storage



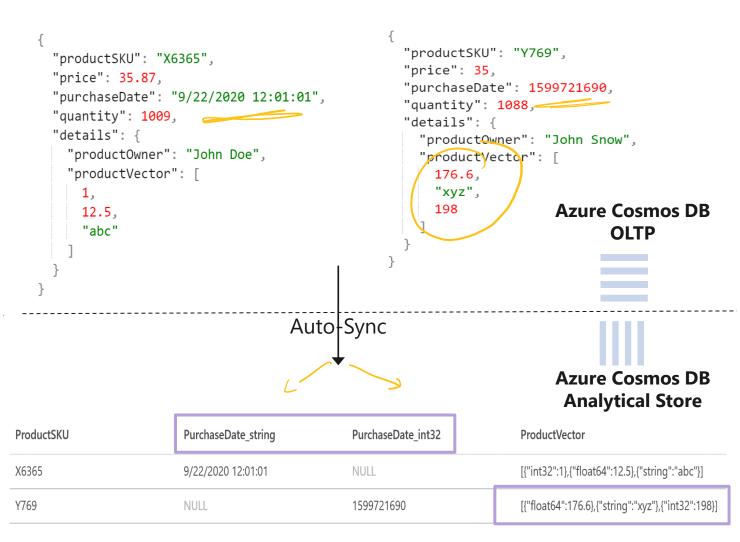


## Azure Synapse Link: Support for Azure Cosmos DB API for MongoDB

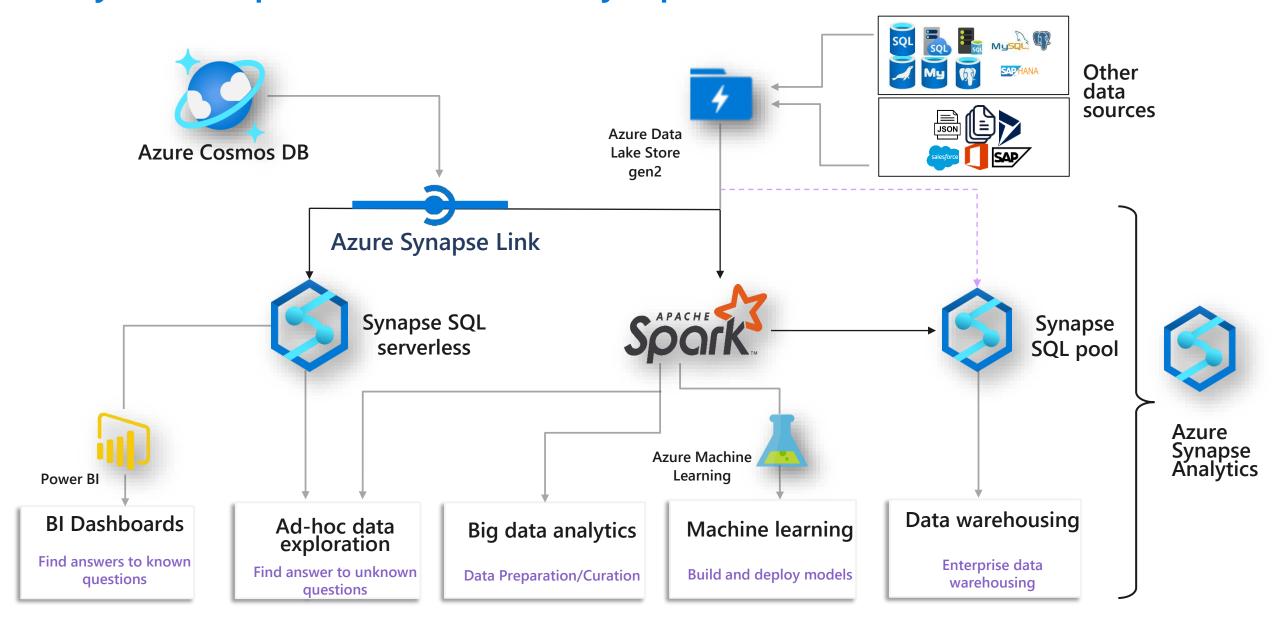


- Cloud native HTAP for MongoDB data
- Run Synapse T-SQL & Spark queries against MongoDB:
   No impact to OLTP workloads
   No need to ETL to OLAP store
- Analytics against highly polymorphic MongoDB BSON schemas:
  Full-fidelity schema representation in

Full-fidelity schema representation in Azure Cosmos DB analytical store (fully managed in auto-sync)



### **Analytics + BI patterns with Azure Synapse Link**



## Describe nonrelational data workloads

- describe the characteristics of non-relational data
- describe the types of non-relational and NoSQL data
- recommend the correct data store
- determine when to use non-relational data
- Describe non-relational data offerings on Azure
- identify Azure data services for non-relational workloads
- describe Azure Cosmos DB APIs
- describe Azure Table storage
- describe Azure Blob storage
- describe Azure File storage

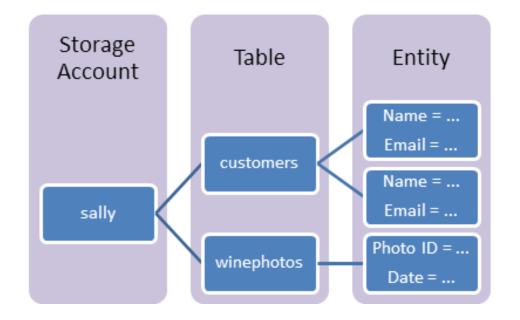
## Table storage concepts

•URL format: Azure Table Storage accounts use this format: http://<storage account>.table.core.windows.net/

Azure Cosmos DB Table API accounts use this format: http://<storageaccount>.table.cosmosdb.azure.com/

You can address Azure tables directly using this address with the OData protocol. For more information, see <a href="OData.org">OData.org</a>.

- •Accounts: All access to Azure Storage is done through a storage account. For more information about storage accounts, see <a href="Storage account overview">Storage account overview</a>. All access to Azure Cosmos DB is done through a Table API account. See <a href="Create a Table API account">Create a Table API account</a> for details creating a Table API account.
- •**Table**: A table is a collection of entities. Tables don't enforce a schema on entities, which means a single table can contain entities that have different sets of properties.
- •Entity: An entity is a set of properties, similar to a database row. An entity in Azure Storage can be up to 1MB in size. An entity in Azure Cosmos DB can be up to 2MB in size.
- •Properties: A property is a name-value pair. Each entity can include up to 252 properties to store data. Each entity also has three system properties that specify a partition key, a row key, and a timestamp. Entities with the same partition key can be queried more quickly, and inserted/updated in atomic operations. An entity's row key is its unique identifier within a partition.



## Describe nonrelational data workloads

- describe the characteristics of non-relational data
- describe the types of non-relational and NoSQL data
- recommend the correct data store
- determine when to use non-relational data
- Describe non-relational data offerings on Azure
- identify Azure data services for non-relational workloads
- describe Azure Cosmos DB APIs
- describe Azure Table storage
- describe Azure Blob storage
- describe Azure File storage

## blob storage concepts



## Évolutivité, durabilité et disponibilité

Excellente durabilité, avec géoréplication et flexibilité pour une mise à l'échelle en fonction des besoins



#### Sécurité

Authentification auprès d'Azure Active Directory et contrôle d'accès en fonction du rôle (RBAC), plus chiffrement au repos et protection avancée contre les menaces



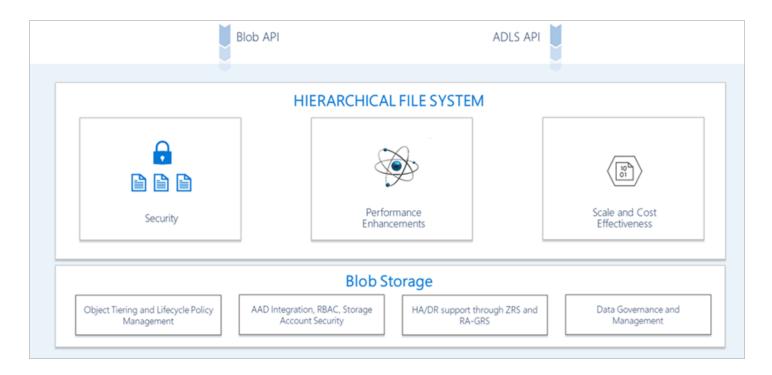
#### Optimisé pour les lacs de données

Espace de noms de fichier et prise en charge de l'accès multiprotocole permettant des charges de travail analytiques pour obtenir des insights sur les données



### Gestion complète des données

Gestion du cycle de vie de bout en bout, contrôle d'accès basé sur des stratégies et stockage immuable (WORM)



## Describe nonrelational data workloads

- describe the characteristics of non-relational data
- describe the types of non-relational and NoSQL data
- recommend the correct data store
- determine when to use non-relational data
- Describe non-relational data offerings on Azure
- identify Azure data services for non-relational workloads
- describe Azure Cosmos DB APIs
- describe Azure Table storage
- describe Azure Blob storage
- describe Azure File storage

## **Azure File Storage**



#### Partages de fichiers serverless

Système de fichiers distribués entièrement conçu pour être hautement disponible et durable (pas de machines virtuelles laaS ni de fichier NAS masqués)



#### Coût total de possession optimisé

Différents niveaux sont utilisés pour adapter vos partages à vos besoins en matière de performances et à des prix compétitifs



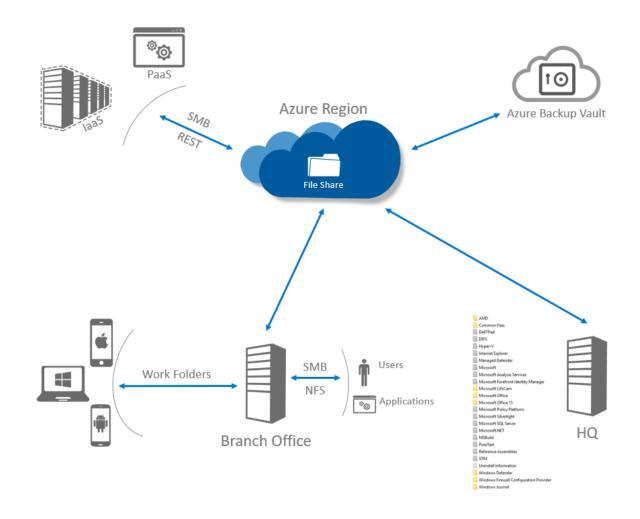
### Conçu pour les scénarios hybrides avec File Sync

Transformez vos serveurs Windows en caches de vos partages cloud sans compromettre les performances locales



### Prise en charge de plusieurs protocoles

Prend en charge plusieurs protocoles permettant de stocker les données à partir de différentes sources





Identify basic management tasks for non-relational data



describe provisioning and deployment of non-relational data services



describe method for deployment including the Azure portal, Azure Resource Manager templates, Azure PowerShell, and the Azure command-line interface (CLI)



identify data security components (e.g., firewall, authentication, encryption)



identify basic connectivity issues (e.g., accessing from on-premises, access with Azure VNets, access from Internet, authentication, firewalls)



identify management tools for non-relational data

## Services réseau Azure

Virtual Network

Virtual WAN, ExpressRoute, VPN

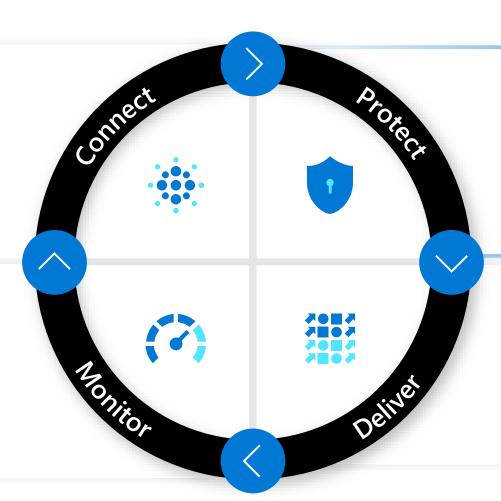
DNS, Azure Bastion

Virtual network NAT Gateway

Azure Peering Service

Azure Edge Zones, Azure Orbital

Network Watcher
ExpressRoute Monitor
Azure Monitor
Virtual Network TAP



Azure Firewall
Azure Firewall Manager
DDoS Protection
Azure Web Application Firewall
Network Security Groups
Service Endpoints, Private Links

CDN
Azure Front Door
Traffic Manager
Application Gateway
Load Balancer

## Services de protection réseau













## DDoS protection

Protection DDOS adaptée aux modèles de trafic de vos applications



Protection centralisée des applications Web entrantes contre les attaques et les vulnérabilités courantes

### Azure Firewall

Filtrage réseau et application centralisé (L3-L7) sortant et entrant (non HTTP / S)

### Network Security Groups

Filtrage distribué du trafic sur les réseaux entrants et sortants (L3-L4) sur une VM, un conteneur ou un sous-réseau

### **Private link**

Limitez l'accès aux ressources de service Azure (PaaS) à votre réseau virtuel uniquement

### Security Appliances

Tirez parti de vos compétences, processus et licences existants en ajoutant des technologies à partir de la Place de marché Azure

**Protection des applications** 

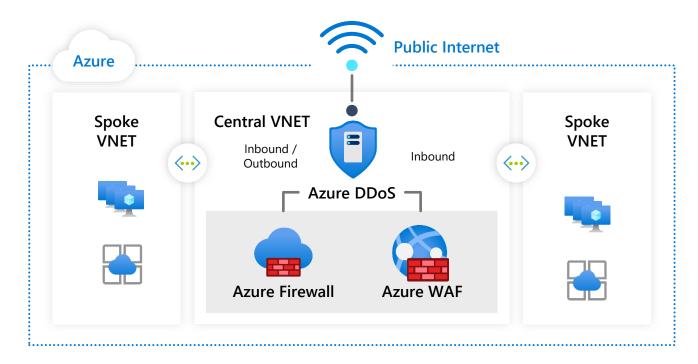
**Segmentation** 

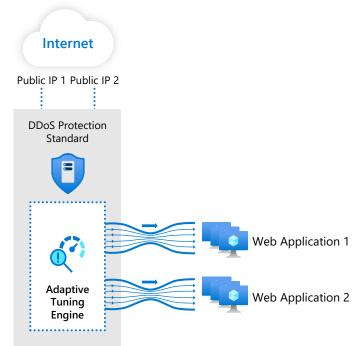
Et plus...

## **Azure DDoS Protection Standard**

**Cloud scale DDoS protection for Azure** 

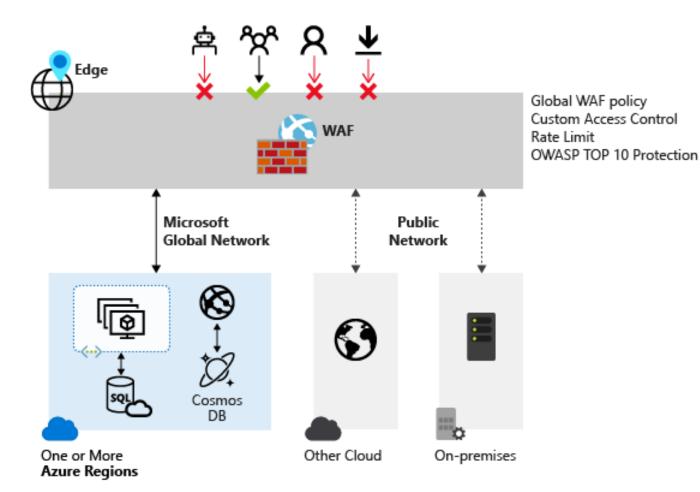






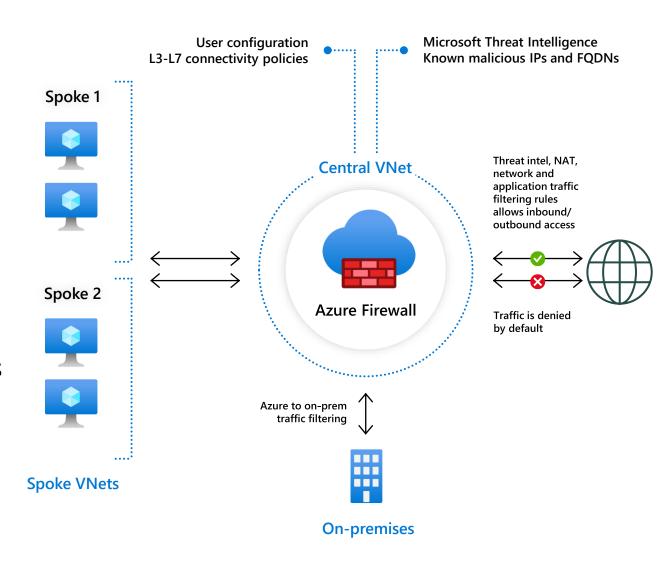
## **Web Application Firewall**

- Une protection centralisée de vos applications
   Web contre les vulnérabilités et les exploits
   courants (SQL injection, cross-site scripting...)
- Une solution WAF peut réagir plus rapidement à une menace de sécurité en corrigeant de manière centralisée une vulnérabilité connue
- Service supporté : Azure Application Gateway,
   Azure Front Door et Azure CDN de Microsoft
   (Public preview)
- WAF possède des fonctionnalités personnalisées pour chaque service spécifique



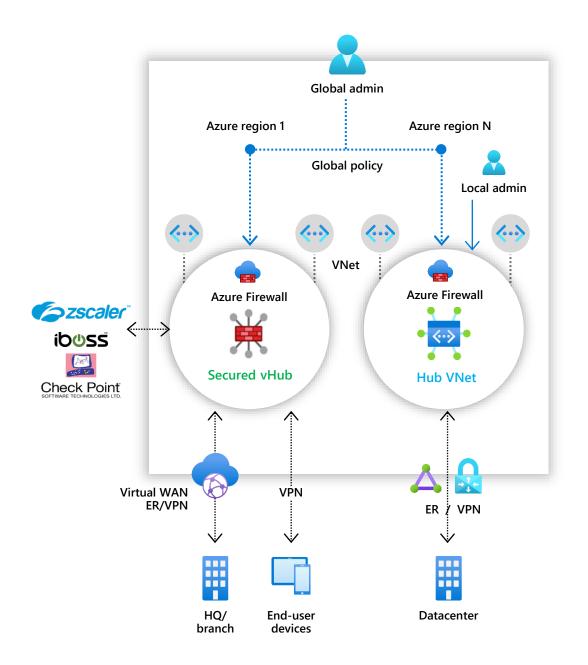
## **Azure Firewall**

- Firewall as a Service
- · Statefull (L3,4,7) hautement disponible et scalable
- Complète les NSG et autres mécanismes de sécurité
- Filtrage par FQDN des flux pour contrôler le trafic <u>sortant</u> HTTP(S)
- Gestion centralisée (multi abonnements) des règles de filtrage
- Passerelle SNAT pour les flux sortants. NAT entre le VNet et l'IP publique
- Journalisation dans Azure Monitor



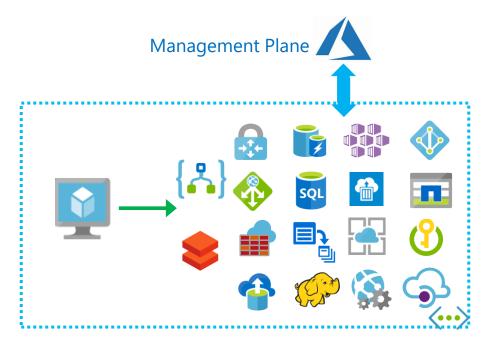
## **Azure Firewall Manager**

- Un nouveau service qui fournit une console centralisée pour déployer et configurer plusieurs instances d'Azure Firewall
- Déployer les stratégies de pare-feu Azure via différents Secure Virtual Hubs et de Azure Hub Vnet
- Intégration avec des solutions tierces de SECurity as a Service
- Couvre différentes régions et différents abonnements Azure pour la gouvernance et la protection du trafic
- Est nécessaire pour gérer les hubs sécurisés (vWAN)

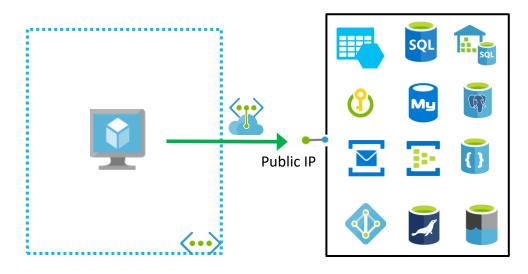


Modèles d'intégration réseau des services Azure

Instances de service exécutées dans le VNet du client



Accès privé aux ressources PaaS Gestion exposée à l'aide d'adresses IP publiques Configuration réseau complexe Service s'exécutant en dehors du VNet du client



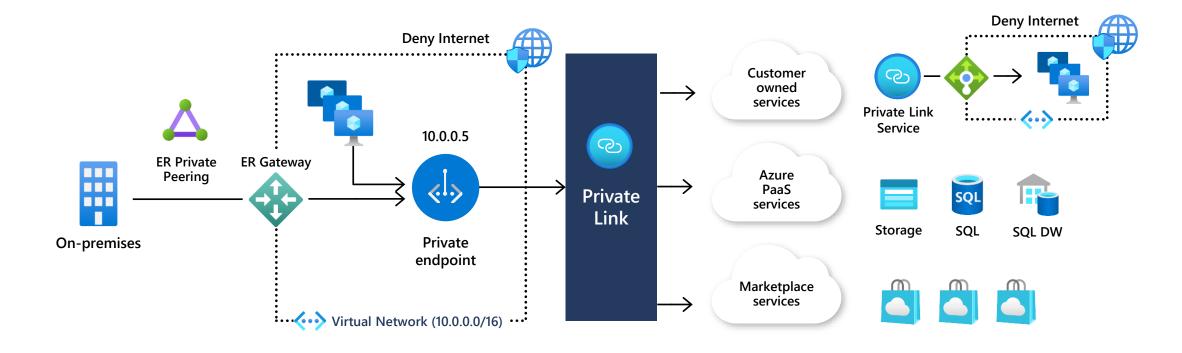
Accès public aux ressources PaaS Plan de gestion isolé Ressource PaaS sécurisée pour le VNet du client

Services dédiés

Services partagés

## **Azure Private Link**

Highly secure and private connectivity solution for Azure Platform



Private access from Virtual Network resources, peered networks and on-premise networks

In-built Data Exfiltration Protection

Predictable private IP addresses for PaaS resources

Unified experience across PaaS, Customer Owned and marketplace Services

## **Azure Bastion**

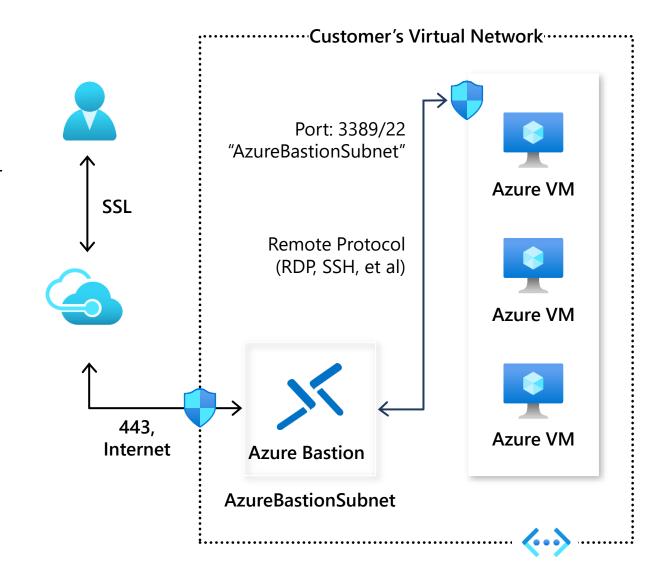
Secure and seamless RDP and SSH access to your virtual machines using zero trust

RDP/SSH to your workload using HTML5 standards-based web-browser, directly in Azure Portal

Resources can be accessed without public IP addresses

Supported Azure resources include VMs, VM Scale Sets, Dev-Test Labs

No agent required



## How it all works together

