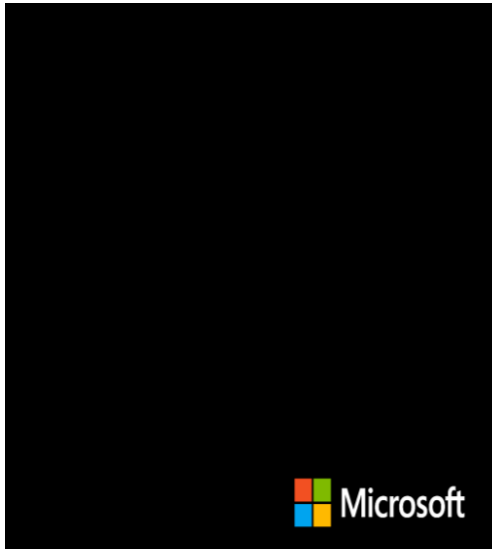
Design Data Implementation

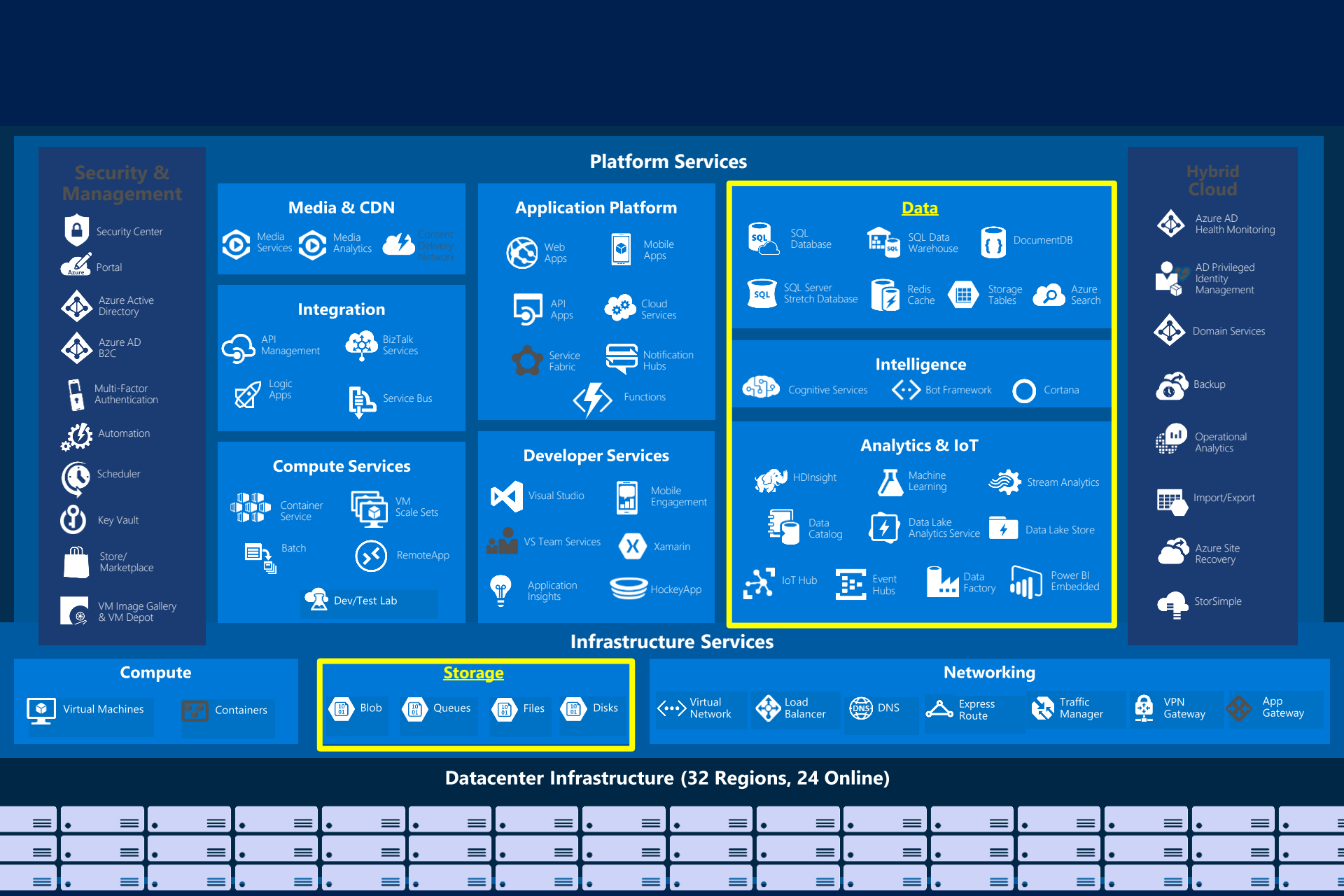


Dan Rey
Cloud Consultant
Technical Trainer | MCT

Design Data Implementation (15 – 20%)

Exam 70-535 : Architecting Microsoft Azure Solutions

- Design for Azure Storage solutions
  - Determine when to use Azure Blob Storage, blob tiers, Azure Files, disks, and StorSimple
- Design for Azure Data Services
  - Determine when to use Data Catalog, Azure Data Factory, SQL Data Warehouse, Azure Data Lake Analytics, Azure Analysis Services, and Azure HDInsight
- Design for relational database storage
  - Determine when to use Azure SQL Database and SQL Server Stretch Database; design for scalability and features; determine when to use Azure Database for MySQL and Azure Database for PostgreSQL; design for HA/DR, geo-replication; design a backup and recovery strategy
- Design for NoSQL storage
  - Determine when to use Azure Redis Cache, Azure Table Storage, Azure Data Lake, Azure Search, Time Series Insights
- Design for CosmosDB storage
  - Determine when to use MongoDB API, DocumentDB API, Graph API, Azure Tables API; design for cost, performance, data consistency, availability, and business continuity


Security & Management
Security Center
Portal
Azure Active Directory
Azure AD B2C
Multi-Factor Authentication
Automation
Scheduler
Key Vault
Store/ Marketplace
VM Image Gallery & VM Depot
Platform Services
Media & CDN
Media Services
Media Analytics
Content Delivery Network
Integration
API Management
BizTalk Services
Logic Apps
Service Bus
Compute Services
Container Service
VM Scale Sets
Batch
RemoteApp
Dev/Test Lab
Application Platform
Web Apps
Mobile Apps
API Apps
Cloud Services
Service Fabric
Notification Hubs
Functions
Developer Services
Visual Studio
Mobile Engagement
VS Team Services
Xamarin
Application Insights
HockeyApp
Data
SQL Database
SQL Data Warehouse
DocumentDB
SQL Server Stretch Database
Redis Cache
Storage Tables
Azure Search
Intelligence
Cognitive Services
Bot Framework
Cortana
Analytics & IoT
HDInsight
Machine Learning
Stream Analytics
Data Catalog
Data Lake Analytics Service
Data Lake Store
IoT Hub
Event Hubs
Data Factory
Power BI Embedded
Hybrid Cloud
Azure AD Health Monitoring
AD Privileged Identity Management
Domain Services
Backup
Operational Analytics
Import/Export
Azure Site Recovery
StorSimple
Infrastructure Services
Compute
Virtual Machines
Containers
Storage
Blob
Queues
Files
Disks
Networking
Virtual Network
Load Balancer
DNS
Express Route
Traffic Manager
VPN Gateway
App Gateway
Datacenter Infrastructure (32 Regions, 24 Online)

Azure Storage solutions

Blob Storage –Standard and Premium (SSD)—Block blob (4.7 TB)- Page blob (8 TB random access files, VHD files for VM)—append blob –append operations like logging to same blob from multiple VMs

File Storage—HA network shared files accessed by SMB protocol (server message block) – multiple VMS can share same file for read/write. Also use REST interface. Diagnostic logs, crash dumps etc. No AD

Queue Storage—store and retrieve messages--millions of messages—list of messages processed asynchrounously

Table Storage—part of CosmosDB– Store high volume of structured NoSQL data, providing key/attrib store with schemaless design. Cheaper than traditional SQL. Data for web apps, address book, device info. Unlimited number of entities and tables in storage account

Blob Tiers—hot (accessed frequ), cool (at least once 30 days), archive (rarely, >180)--
Azure Files—can be used to completely replace or supplement traditional on-prem file serversor NAS devices--
Disks—managed and unmanaged—premium (256 TB per VM, 80K iops, 2k MBps throughput) and standard--
StorSimple—hybrid cloud storage , consolidate storage, DR, compliance,

EXAM TIP!

Know the various storage types and their uses. For example, many times you can use Queues to decouple components of a system.

aka.ms/azure/storage, how they are alike & how they are different

EXAM TIP!

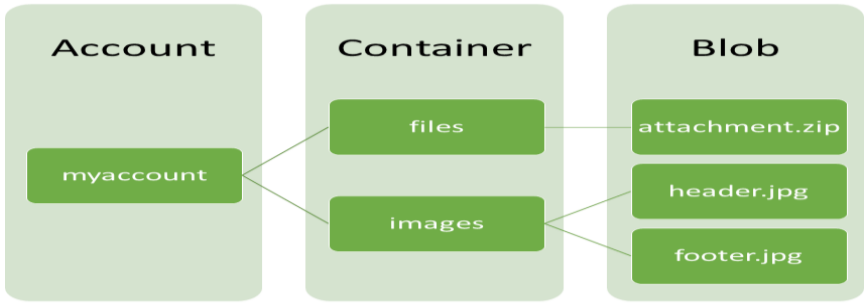
Read

Azure Storage | Share Access Signature - SAS

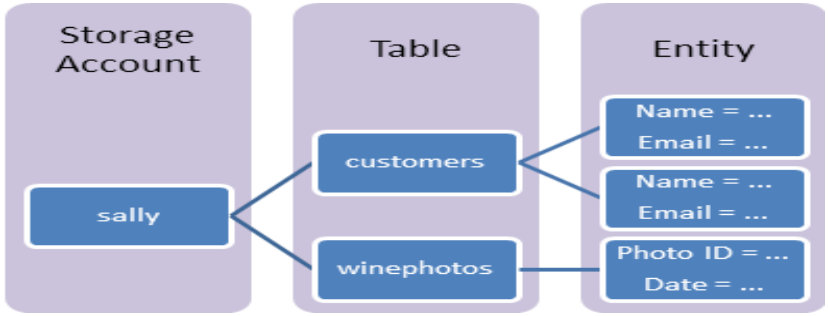
Download and Use:

Microsoft Azure Storage Explorer (Preview) is a standalone app from Microsoft that allows you to easily work with Azure Storage data on Windows, macOS and Linux. http://storageexplorer.com/

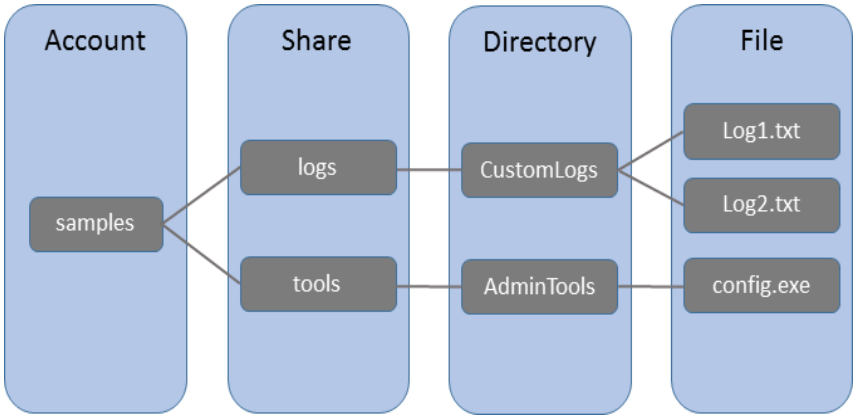
Azure Storage Account Schematic



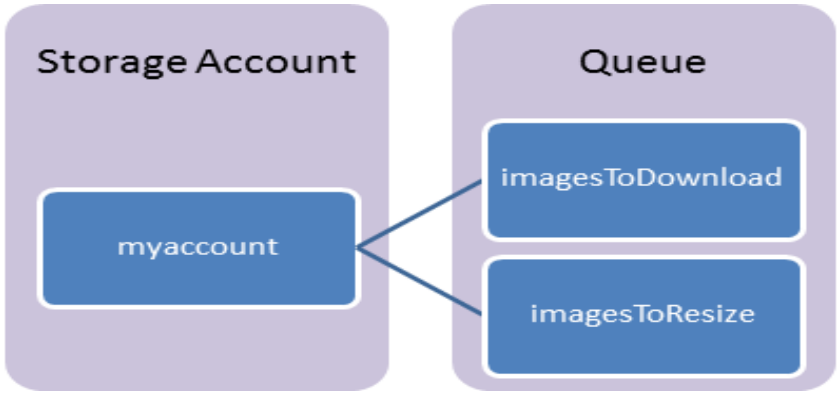
Blob Storage



Azure Table



File Storage



Azure Queue

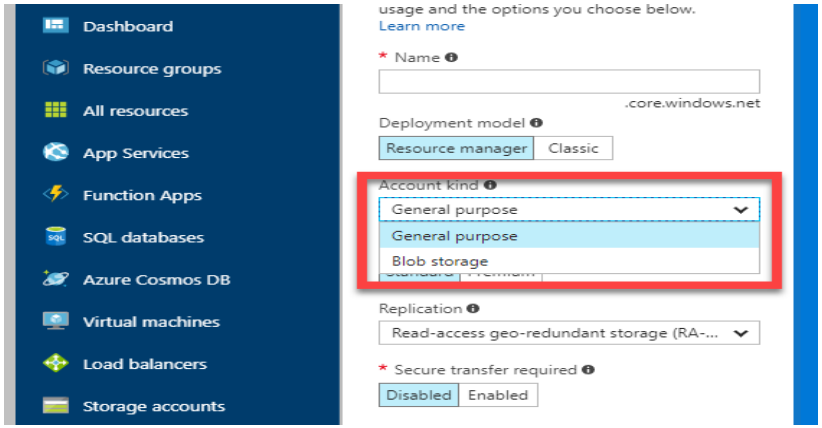
Azure Storage Account

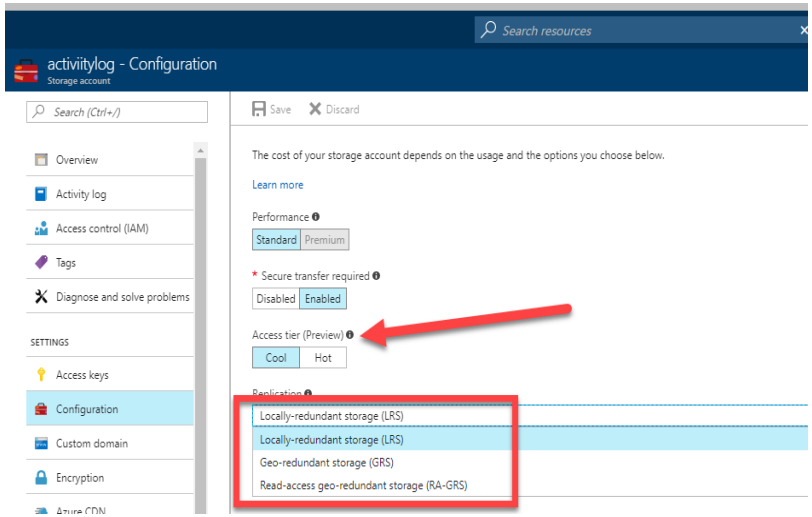
Two Kinds of Storage Account

- General Purpose Storage Account
  - Blobs
  - Files
  - Queues
  - Tables
  - VHD Disks(Page Blob)
- Blob Storage Account
  - Block Blob
  - Append Blob
  - Page Blob

Performance Tier

- General Purpose: **Standard**(HDD based), **Premium**(SSD based) for VM
- Blob Account: **Hot**, **Cool**, **Archival**(Preview)





Azure Table Storage

Stores structure NoSQL

- Key-value pair
- Structured, non relational
- Schemaless

Access data using the Odata & LINQ .

Quickly query data using a clustered index

Data is consistent for other client reads after insert/update

Azure File Storage

- Fully managed file shares in the cloud
- Supports SMB /CIFS protocol
- Can be mounted concurrently
- Azure File Sync(Preview)
- Suitable for Lift & Shift
- LRS, GRS replication

Azure Queue

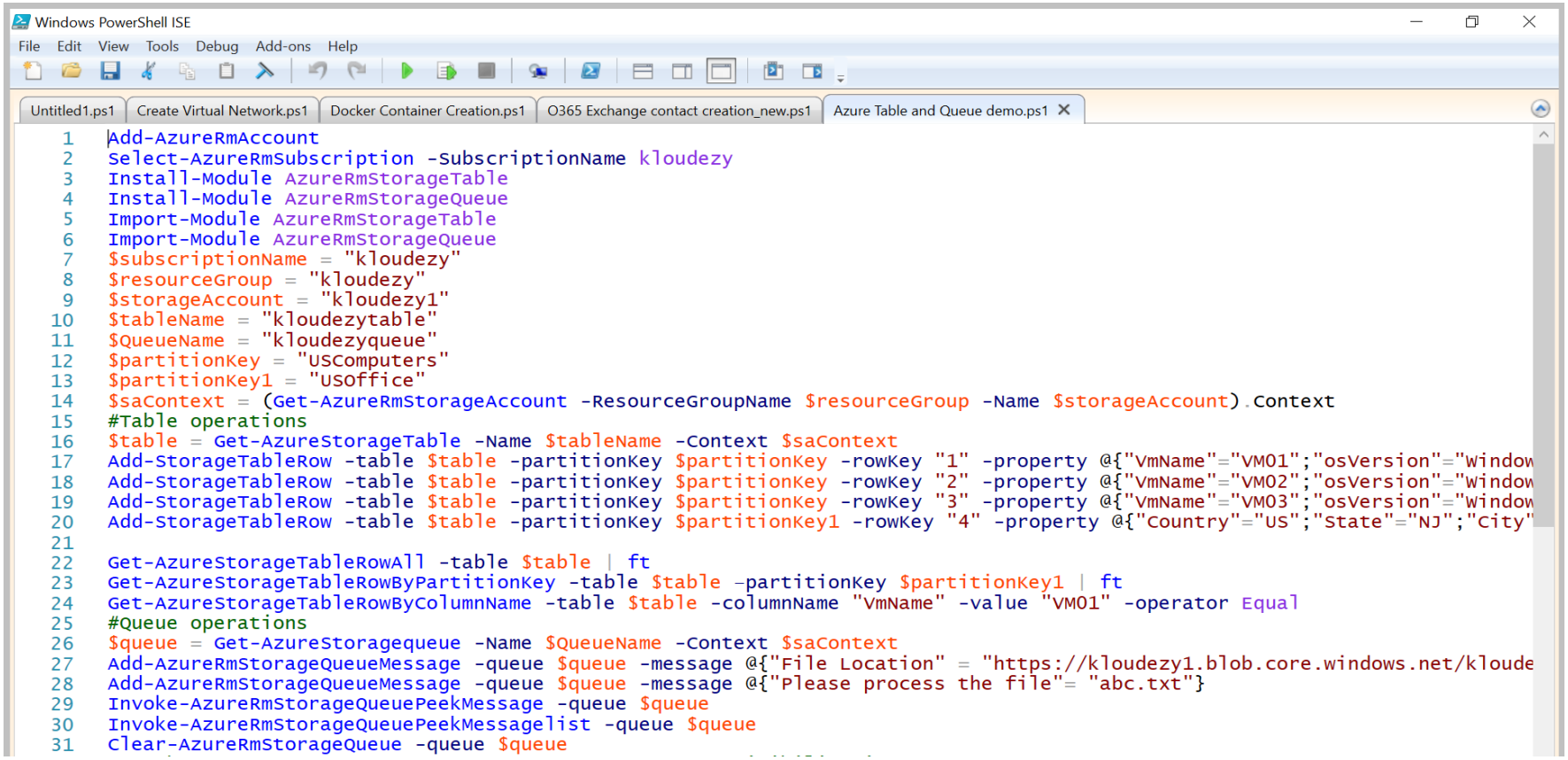
Provides reliable, persistent messaging.

REST-based GET/PUT/PEEK

Maximum Message Size 64 KB. If larger message size needed use Service Bus

Maximum message TTL 7 days

Table and Queue Storage via PowerShell



Storage Options

Standard

- Max total request rate of 20k IOPS
- Billed for actual usage*

Premium

- Designed for Azure VMs
- High performance
- Low latency
- Azure VM disks are implemented as “Page Blobs”
- Billed for Provisioned space and not for actual usage.

Azure Blob Storage

Also called **Object Storage**

Storing large amounts of unstructured object data

Can be accessed from anywhere via HTTP or HTTPS

Three kinds of Blobs:

- Block Blob
- Append Blob
- Page Blob

Common uses of Blob storage include:

- Serving images or documents directly to a browser
- Storing files for distributed access
- Streaming video and audio
- Storing data for backup and restore, disaster recovery, and archiving

Block Blobs

- Blob is broken into pieces
- Block Id for each piece
- Max block size = 100MB
- Up to 50k blocks per blob
- The maximum size of a block blob is ~4.75 TB
- Blocks upload can be multithreaded
- Streaming , Storing documents, media files, backups

Append Blobs

- Similar to Block Blobs
- Optimized for append operations
- Only can add to the end
- The maximum size of an append blob is ~195 GB
- Updating or deleting of existing blocks is not supported
- Unlike a block blob, an append blob does not expose its block IDs.

Page Blob

- Optimized for IaaS disks
- Supports random read/writes
- Collection of 512 byte pages
- Max size = 8TB/4TB(VHD)
- Storage Options
  - Premium Storage(SSD based)
  - Standard Storage(HDD based)

Replication for Table & Blob Storage

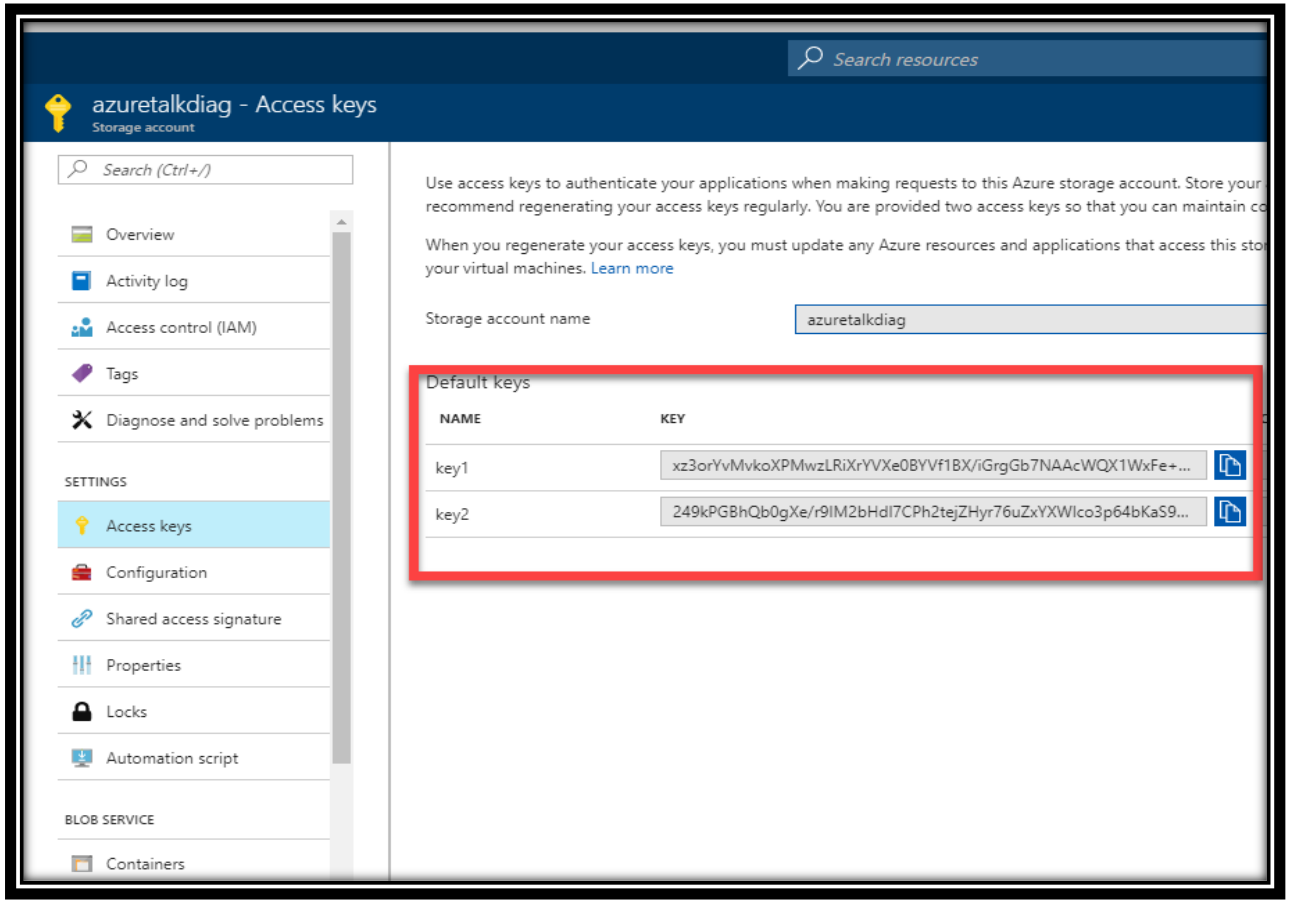
- You can switch from LRS to GRS or RA-GRS but ZRS can't be converted.
- ZRS is only available for General Purpose Account type

Replication Option	Number of copies	Strategy
Locally redundant storage (LRS)	Maintains three copies of your data.	Data is replicated three time within a single facility in a single region.
Zone-redundant storage (ZRS)	Maintains three copies of your data.	Data is replicated three times across two to three facilities, either within a single region or across two regions.
Geo-redundant storage (GRS)	Maintains six copies of your data.	Data is replicated three times within the primary region, and is also replicated three times in a secondary region hundreds of miles away from the primary region.
Read access geo-redundant storage (RA-GRS) (Default)	Maintains six copies of your data.	Data is replicated to a secondary geographic location, and also provides read access to your data in the secondary location.

Blob Access

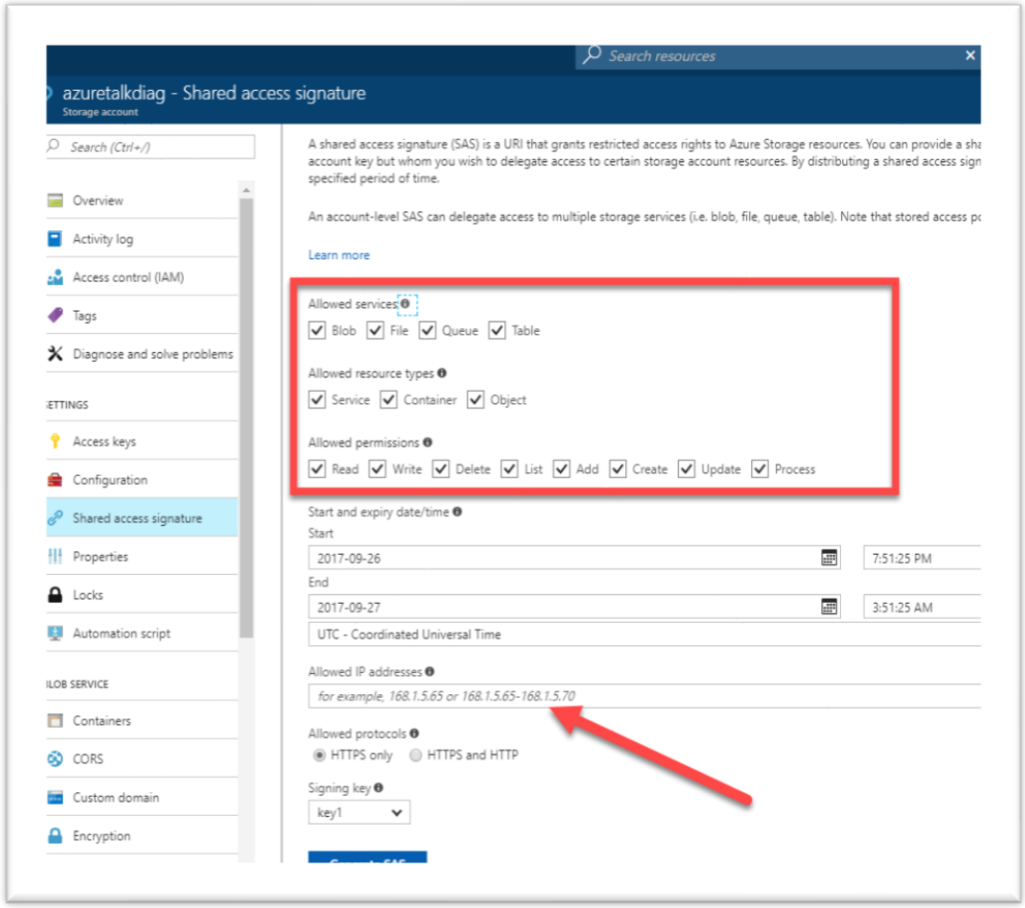
- Anonymous Access: Public access for containers or individual blobs
- Storage Access Key
- Shared Access Signatures (SAS)
  - Delegated access without sharing account key
  - Containers & blobs
  - File shares & files
  - Queues
  - Tables & ranges of table entities

Blob Access


Search resources
azuretalkdiag - Access keys
Storage account
Search (Ctrl+/)
Overview
Activity log
Access control (IAM)
Tags
Diagnose and solve problems
SETTINGS
Access keys
Configuration
Shared access signature
Properties
Locks
Automation script
BLOB SERVICE
Containers
Use access keys to authenticate your applications when making requests to this Azure storage account. Store your
recommend regenerating your access keys regularly. You are provided two access keys so that you can maintain co
When you regenerate your access keys, you must update any Azure resources and applications that access this sto
your virtual machines. Learn more
Storage account name
azuretalkdiag
Default keys
NAME
KEY
key1
xz3orYvMvkoXPMwzLRiXrYVXe0BYVf1BX/iGrgGb7NAAcWQX1WxFe+...
key2
249kPGBhQb0gXe/r9IM2bHdI7CPh2tejZHyr76uZxYXWIco3p64bKaS9...

Shared Access Signatures (SAS)

- Time interval
  - Start & end time for permissions
- Permissions
  - Read, Write, and/or Delete
- (Optional) IP address or address range
- Allowed protocols
  - Example: only https

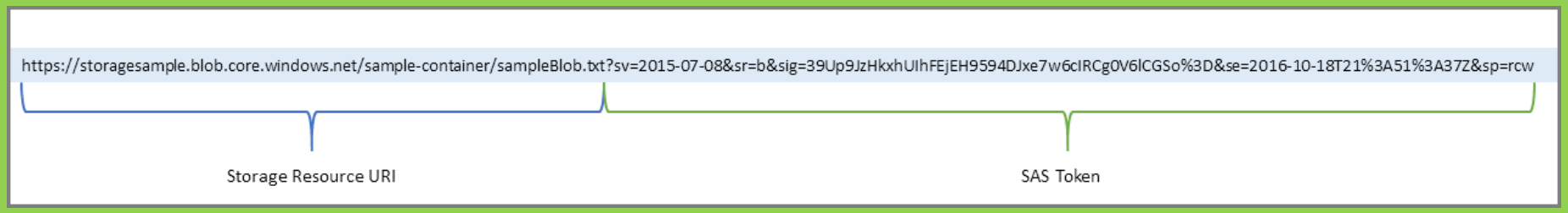


Azure Disk

Two ways to create VM disks

- Unmanaged Disk
  - In an unmanaged disk, you manage the storage accounts
  - VHDs are stored in Page blob
  - 99.99% SLA.
  - Can be converted to Managed.
- Managed Disk
  - Azure manages the storage accounts that you use for your VM disks
  - Available in Standard & premium tier.
  - Standard can be converted in to Premium and vice-versa
  - 99.999% SLA
  - Recommended storage for VMs.
  - Allow you to create up to 10,000 VM **disks** in a subscription
  - Billing for managed disks depends on the provisioned size of the disk.
  - Managed disk Overview

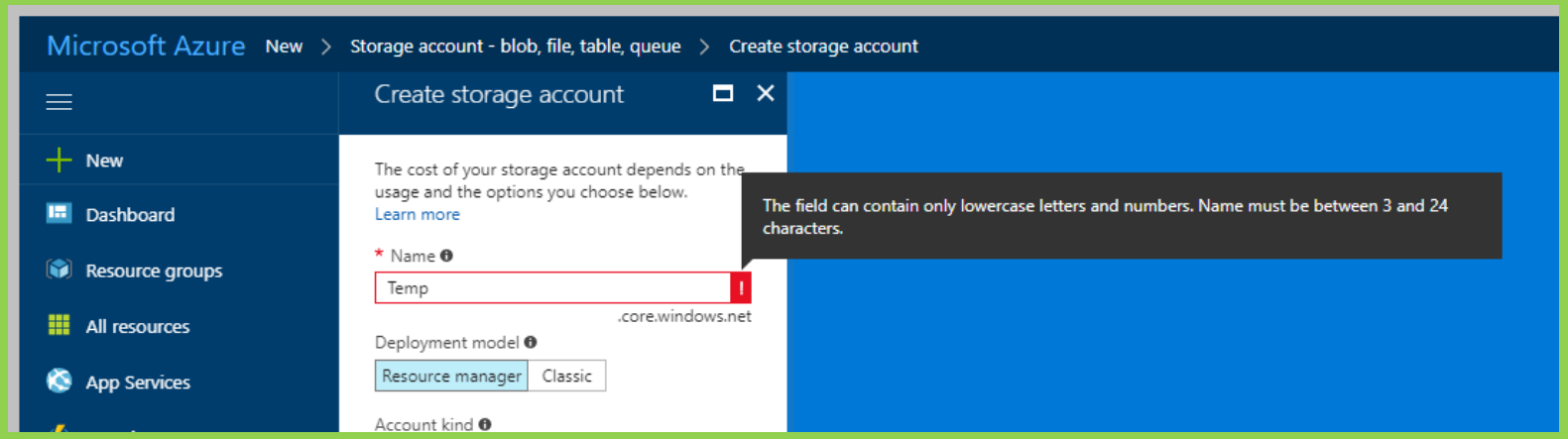
EXAM TIP! *Example of SAS token*



3.1.3 Exam Tip!

EXAM TIP!

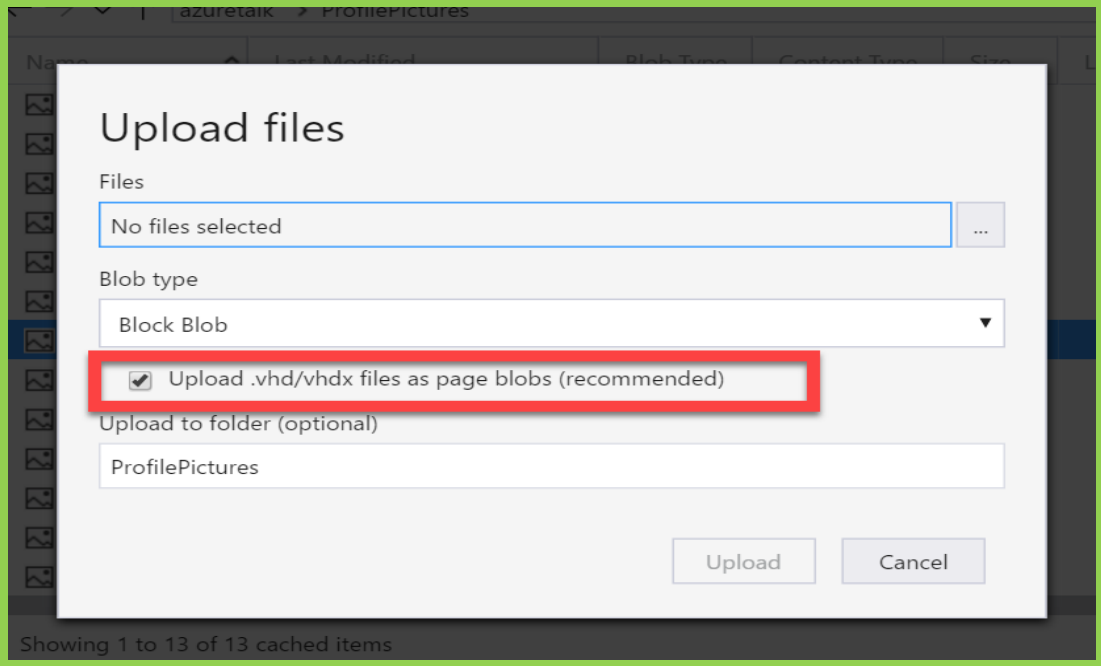
The storage account name should always in lowercase and unique within *.core.windows.net namespace.



3.1.4 Exam Tip!

EXAM TIP!

Choose appropriate blob type for uploading VHD. If VHD files are uploaded in block blob you can’t use those.



3.1.5 Exam Tip!

Blob Encryption/Security

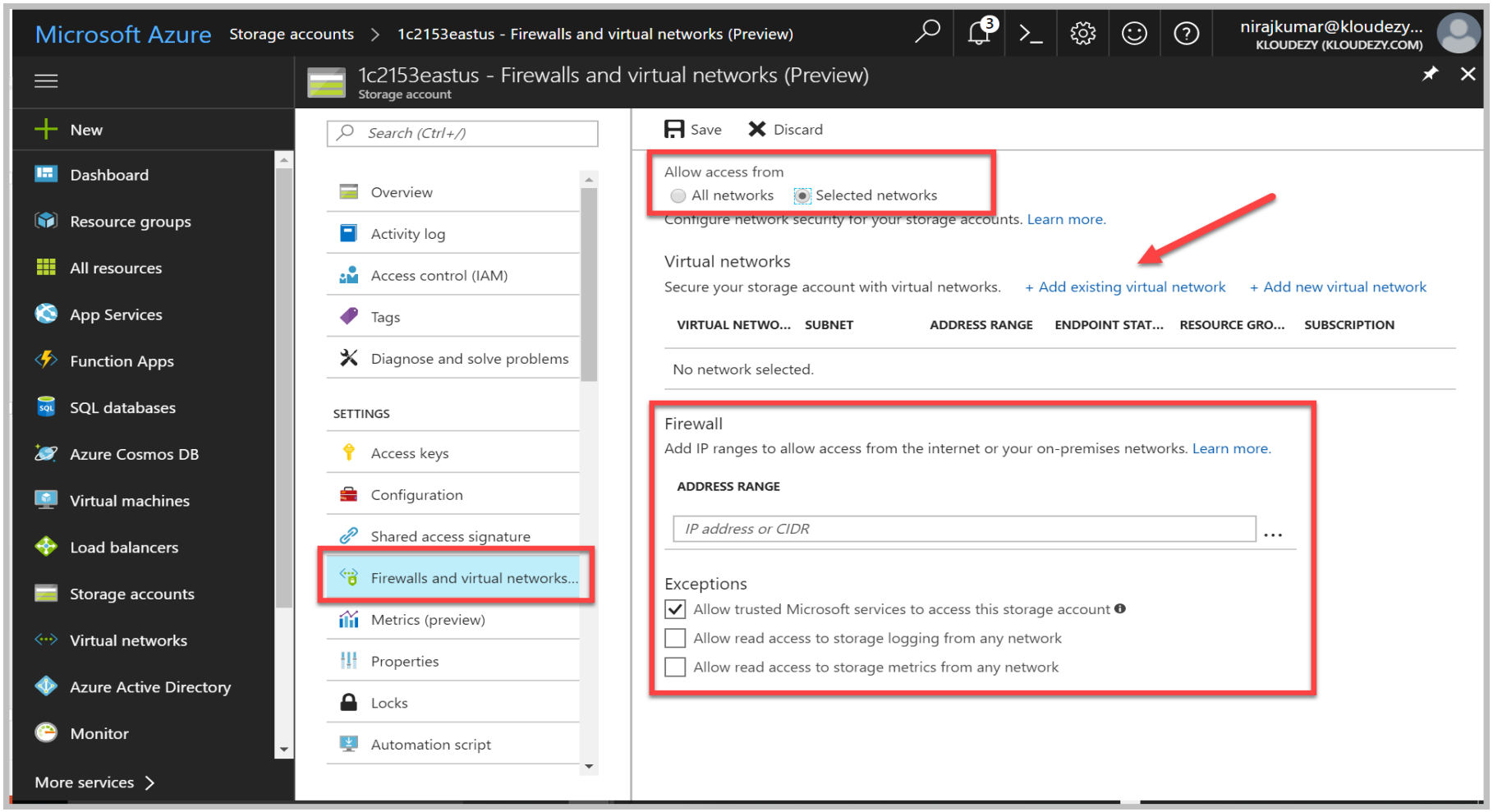


Table & Blob:
Performance, Storage Limits, etc.

https://docs.microsoft.com/en-us/azure/storage/storage-scalability-targets

Resource	Default Limit
Number of storage accounts per subscription	200
TB per storage account	500 TB
Max number of blob containers, blobs, file shares, tables, queues, entities, or messages per storage account	Only limit is the 500 TB storage account capacity
Max size of a single blob container, table, or queue	500 TB
Max number of blocks in a block blob or append blob	50,000
Max size of a block in a block blob	100 MB
Max size of a block blob	50,000 X 100 MB (approx. 4.75 TB)
Max size of a block in an append blob	4 MB
Max size of an append blob	50,000 X 4 MB (approx. 195 GB)
Max size of a page blob	4 TB
Max size of a table entity	1 MB

Resource	Default Limit
Maximum Request Rate per storage account	Blobs: 20,000 requests per second for blobs of any valid size (capped only by the account's ingress/egress limits)
	Files: 1000 IOPS (8 KB in size) per file share
	Queues: 20,000 messages per second (assuming 1 KB message size)
	Tables: 20,000 transactions per second (assuming 1 KB entity size)
Target throughput for single blob	Up to 60 MB per second, or up to 500 requests per second
Target throughput for single queue (1 KB messages)	Up to 2000 messages per second
Target throughput for single table partition (1 KB entities)	Up to 2000 entities per second
Target throughput for single file share	Up to 60 MB per second
Max ingress per storage account (US Regions)	10 Gbps if GRS/ZRS enabled, 20 Gbps for LRS
Max egress per storage account (US Regions)	20 Gbps if RA-GRS/GRS/ZRS enabled, 30 Gbps for LRS
Max ingress per storage account (Non-US regions)	5 Gbps if GRS/ZRS enabled, 10 Gbps for LRS
Max egress per storage account (Non-US regions)	10 Gbps if RA-GRS/GRS/ZRS enabled, 15 Gbps for LRS

EXAM TIP!

Sufficient bandwidth on VM

Make sure sufficient bandwidth is available on your VM to drive disk traffic, as described in Premium Storage-supported VMs. Otherwise, your disk throughput and IOPS is constrained to lower values. Maximum throughput and IOPS are based on the VM limits, not on the disk limits described in the preceding table.

https://docs.microsoft.com/en-us/azure/storage/storage-premium-storage#scalability-and-performance-targets

EXAM TIP! *Premium storage accounts Scalability*

… have the following scalability targets:+

Total account capacity

Disk capacity: 35 TB

Snapshot capacity: 10 TB

Total bandwidth Locally redundant storage account

Up to 50 gigabits per second

for **inbound[1] + outbound[2]**

[1] All data (requests) that are sent to a storage account+

[2] All data (responses) that are received from a storage account

https://docs.microsoft.com/en-us/azure/storage/storage-premium-storage#scalability-and-performance-targets

EXAM TIP!

If Using Storage Spaces Striping

If you stripe premium storage data disks by using Storage Spaces, set up Storage Spaces with 1 column for each disk that you use. Otherwise, overall performance of the striped volume might be lower than expected because of uneven distribution of traffic across the disks. By default, in Server Manager, you can set up columns for up to 8 disks. If you have more than 8 disks, use PowerShell to create the volume. Specify the number of columns manually. Otherwise, the Server Manager UI continues to use 8 columns, even if you have more disks. For example, if you have 32 disks in a single stripe set, specify 32 columns. To specify the number of columns the virtual disk uses, in the New-VirtualDisk PowerShell cmdlet, use the *NumberOfColumns* parameter. For more information, see Storage Spaces Overview and Storage Spaces FAQs.

https://docs.microsoft.com/en-us/azure/storage/storage-premium-storage#scalability-and-performance-targets

EXAM TIP! *Pay Attention to I/O vs IOPs*

If your disk traffic mostly consists of small I/O sizes, your application likely will hit the IOPS limit before the throughput limit. However, if the disk traffic mostly consists of large I/O sizes, your application likely will hit the throughput limit first, instead of the IOPS limit. You can maximize your application's IOPS and throughput capacity by using optimal I/O sizes. Also, you can limit the number of pending I/O requests for a disk.

https://docs.microsoft.com/en-us/azure/storage/storage-premium-storage#scalability-and-performance-targets

Azure Storage Samples

https://azure.microsoft.com/en-us/resources/samples/?service=storage

Table Labs for .NET

https://github.com/Azure-Samples/storage-table-dotnet-getting-started

Blob Labs for .NET

https://azure.microsoft.com/en-us/resources/samples/storage-blob-dotnet-getting-started/

https://github.com/Azure-Samples/storage-blob-dotnet-getting-started

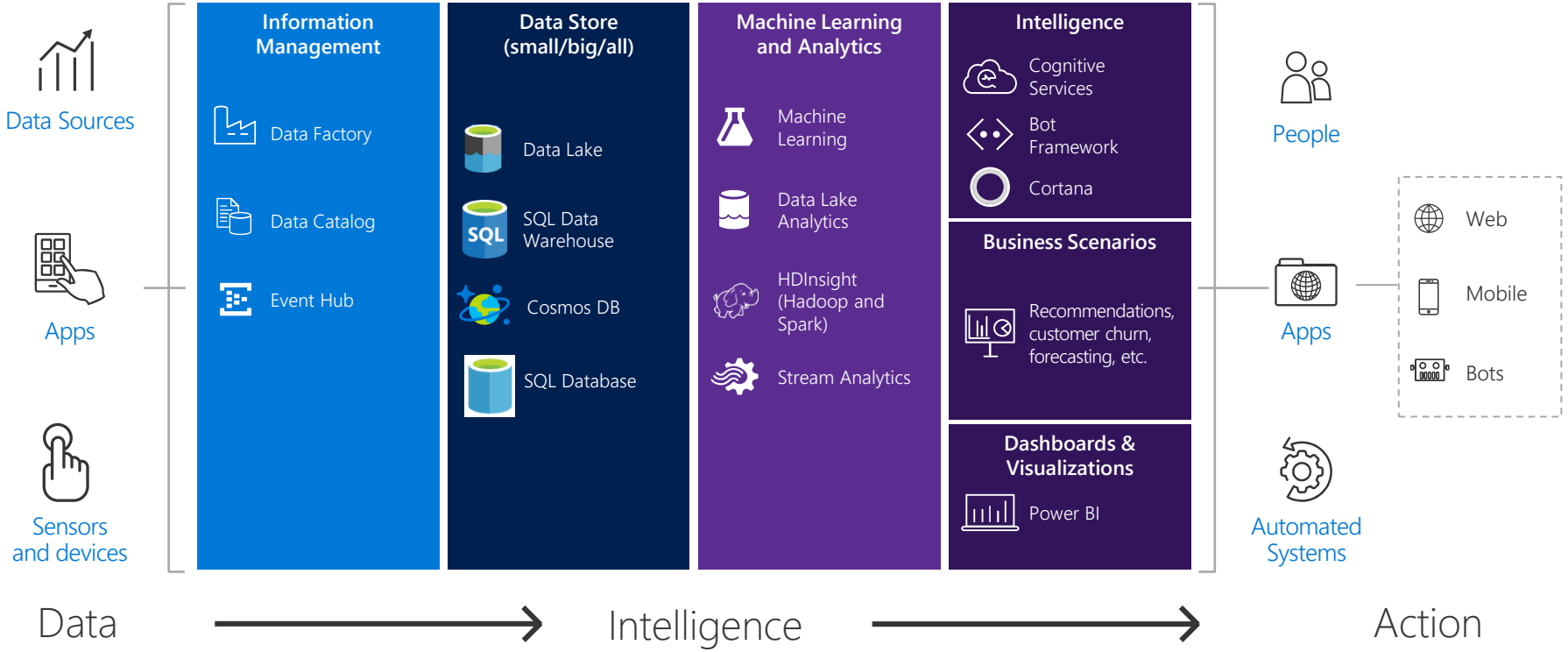
SAS Labs

https://azure.microsoft.com/en-us/resources/samples/storage-dotnet-sas-getting-started/

Azure Storage KeyWords

- General Purpose account
- Blob Account
- Premium
- Standard
- SSE (Server Side Encryption)
- Hot, Cool, Archive
- Block, Append, Page blob
- Azure File, Queue, Table, blob
- SAS(Shared Access Signature)
- Storage Key
- Containers
- Entity
- Message
- Managed Disk
- IOPs
- LRS, ZRS, GRS, RA-GRS
- Storage Explorer
- Azcopy

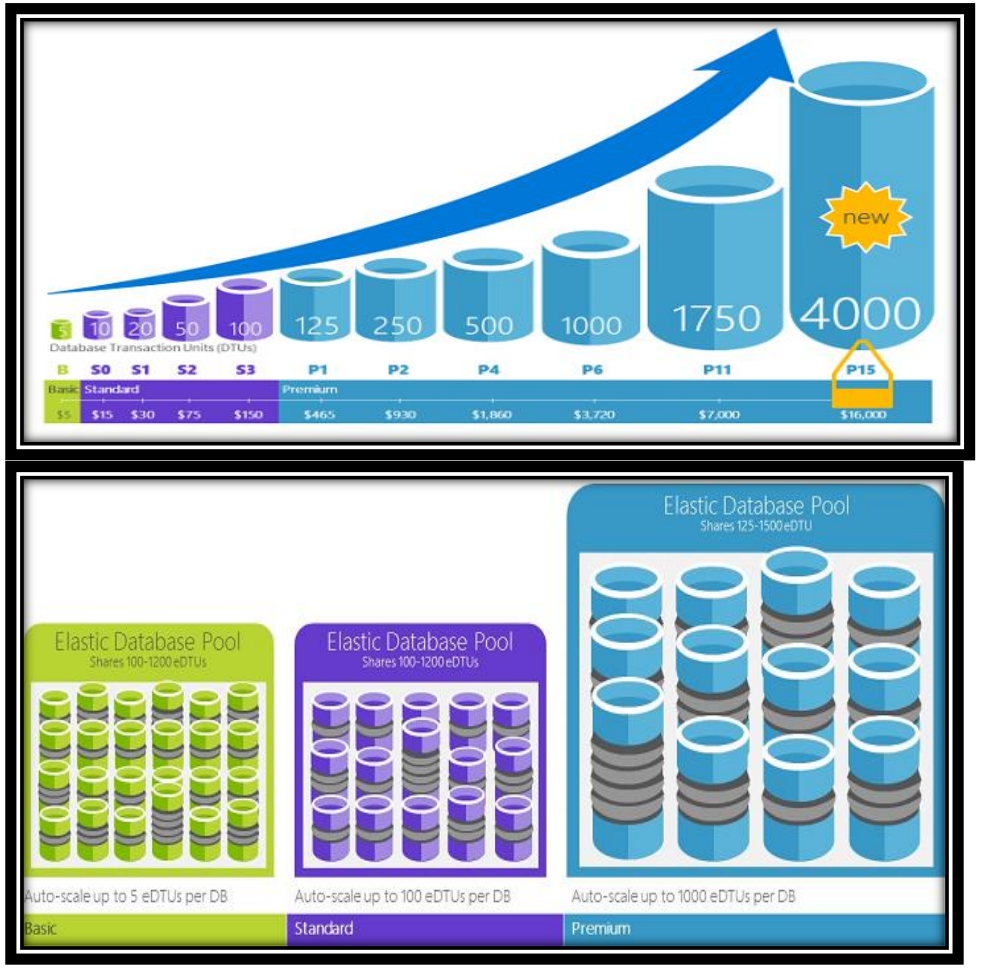
Azure Data Services



SQL Database

Azure SQL Database

- PaaS Offering. Database-as-a-Service
- Relational database.
- Service tiers:
  - Basic
  - Standard
  - Premium
  - Premium RS
- Elastic Pools. Pooled databases delivers performance



Source: Microsoft

Service Tiers

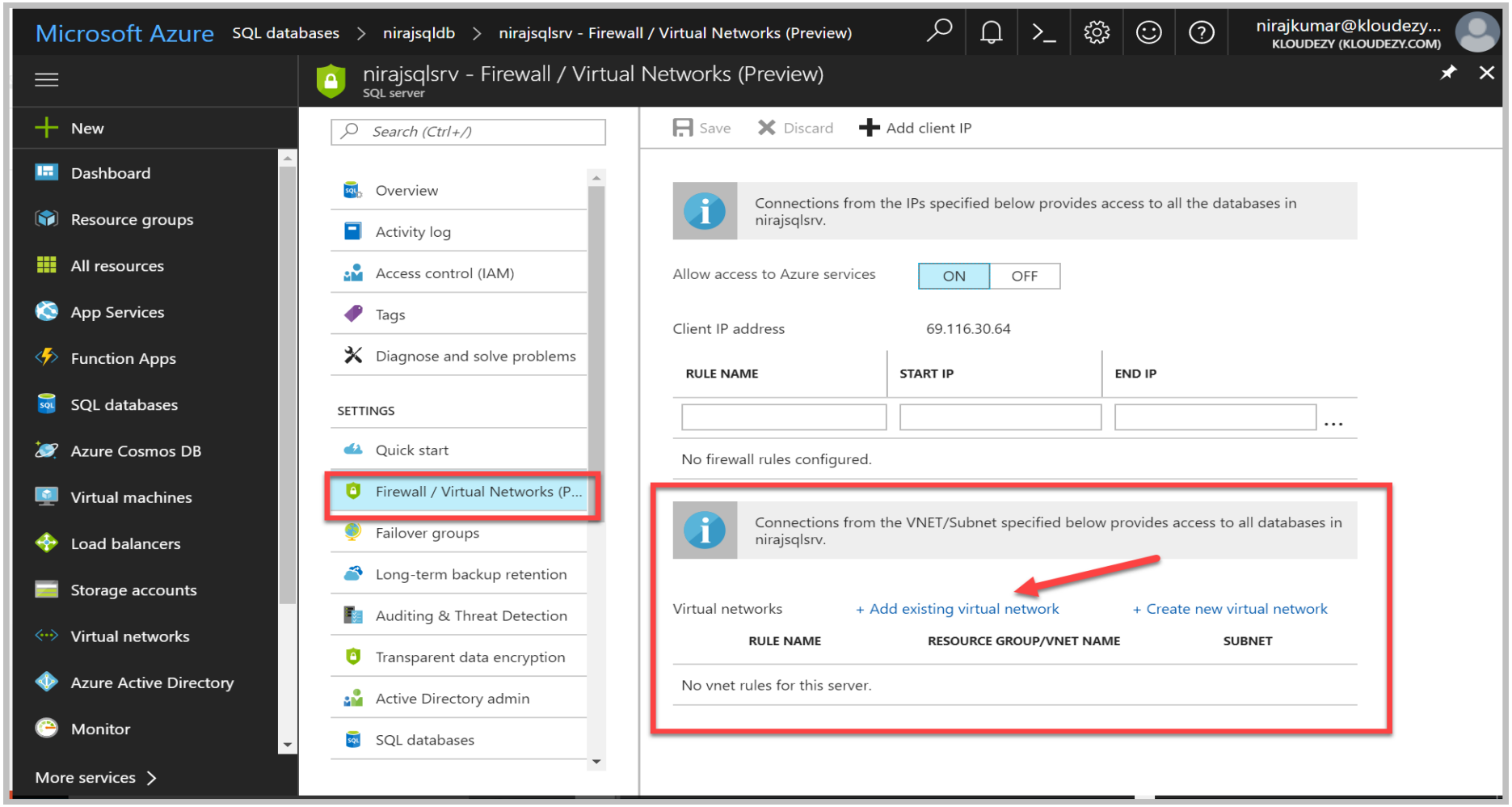
- Basic
  - Small size
  - Low traffic
  - Single active operation
- Standard
  - Low to medium IO performance requirements
  - Multiple current queries
- Premium
  - High transaction volume
  - High IO performance requirements
  - Many concurrent users
  - High availability
  - For mission critical apps
- Premium RS
  - IO-intensive apps
  - Does not require high availability
  - Analytical workloads

Service Tier Attributes

Service tier features	Basic	Standard	Premium	Premium RS
Maximum single database size	2 GB	250 GB	4 TB*	1 TB
Maximum elastic pool size	156 GB	4 TB	4 TB*	1 TB
Maximum database size in an elastic pool	2 GB	250 GB	500 GB	500 GB
Maximum number of databases per pool	500	500	100	100
Maximum single database DTUs	5	100	4000	1000
Maximum DTUs per database in an elastic pool	5	3000	4000	1000
Database backup retention period	7 days	35 days	35 days	35 days
SKUs	Basic	S0-S12	P1-P6, P11, P15	PRS1, PRS2, PRS4, PRS6

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-service-tiers

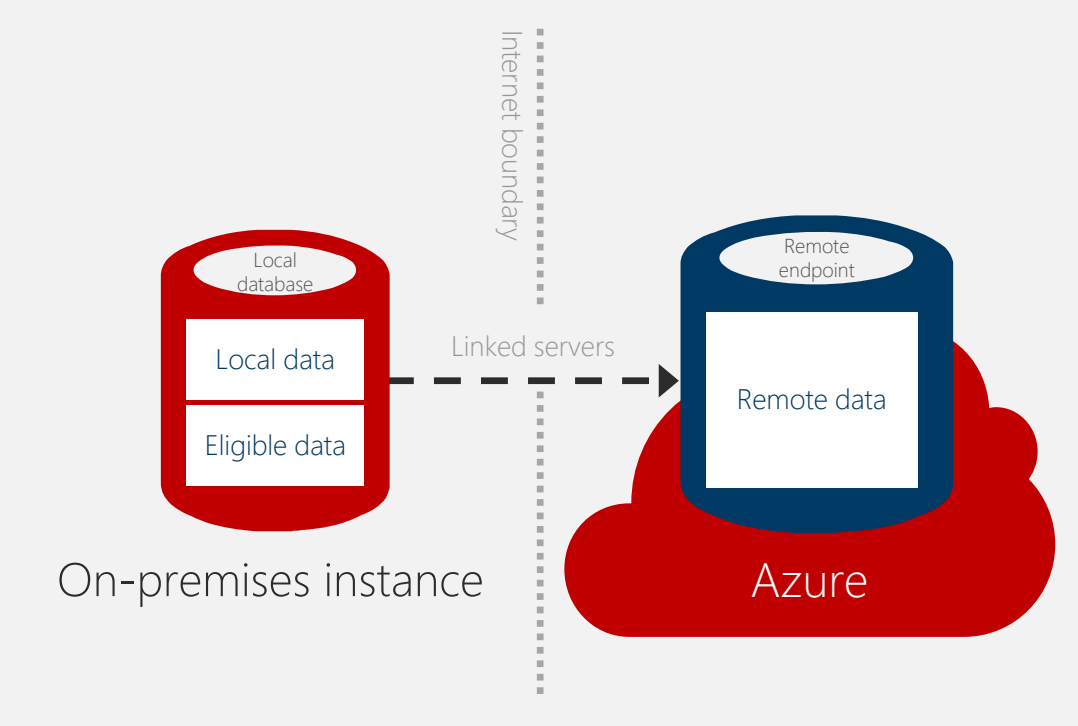
SQL Database Security



Azure SQL KeyWords

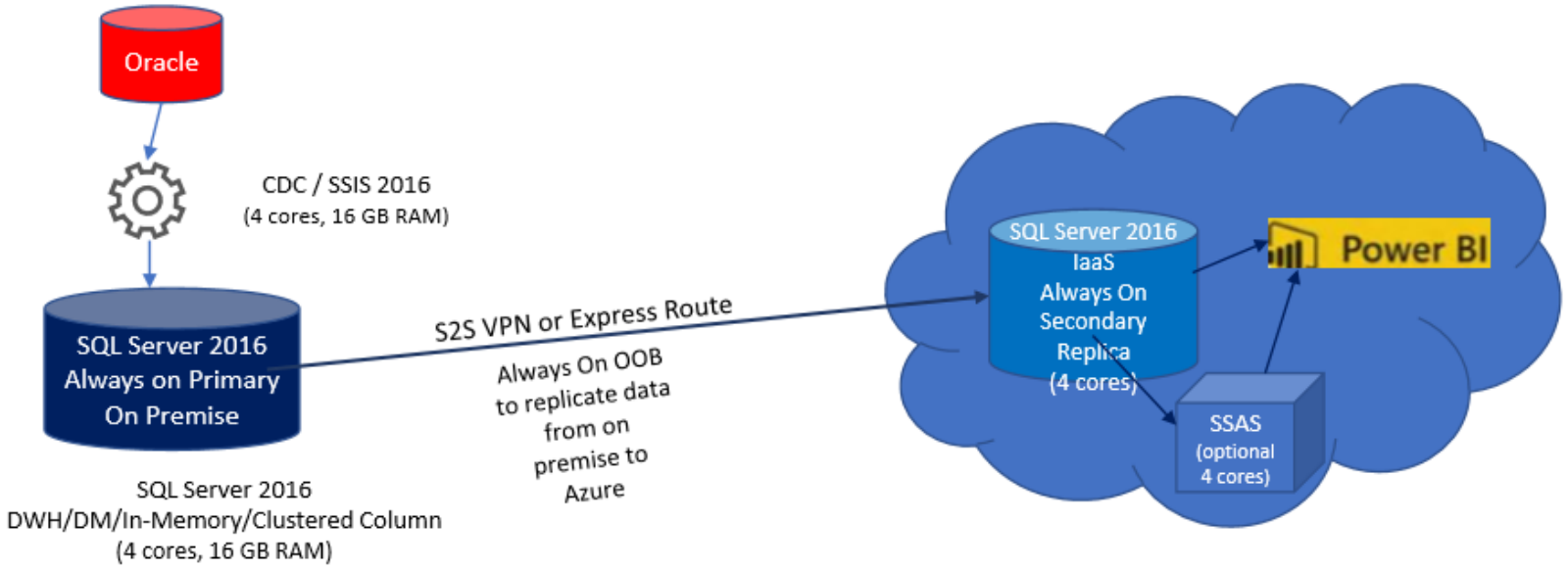
- TDE
- SQL Firewall
- DTU(Database Transaction Unit)
- eDTU(Elastic DTU)
- Always Encrypt
- Azure AD integrated Authentication
- Dynamic data masking
- Long term backup retention
- Row Level Security
- Failover Groups
- Audit and threat detection
- Active directory admin
- Elastic pool
- Geo Replication
- Vulnerability Assessment
- Data Sync

Stretch Database architecture



How it works

- Creates a secure linked server definition in the on-premises SQL Server
- Targets remote endpoint with linked server definition
- Provisions remote resources and begins to migrate eligible data, if migration is enabled
- Queries against tables run against both local database and remote endpoint



- Start developing solution with above configuration.
- D12 instances can be used in Azure.

SQL Database Videos

https://azure.microsoft.com/en-us/resources/videos/index/?services=sql-database

EXAM TIP!

Azure might update to change feature availability. The exam is updated over time, as well, to reflect these changes. However, because of the way Azure is steadily being updated, the newest features might not be on the exams.

EXAM TIP!

Performance levels of the database are important to a company, and the ability to change this at times is equally important. For example, the limits of each level of SQL Database can help the architect to determine the minimum level needed to satisfy those needs.

Azure Database for MySQL

A fully managed MySQL database service for app developers
Integrate the MySQL community edition with Azure for scalability, high availability, and your choice of languages and frameworks

1. Native MySQL that’s fully managed
2. Languages and frameworks of your choice
3. Built-in high availability
4. Scale within seconds
5. Simple and Flexible pricing
6. Unparalleled security and reach

Azure Database for PostgreSQL

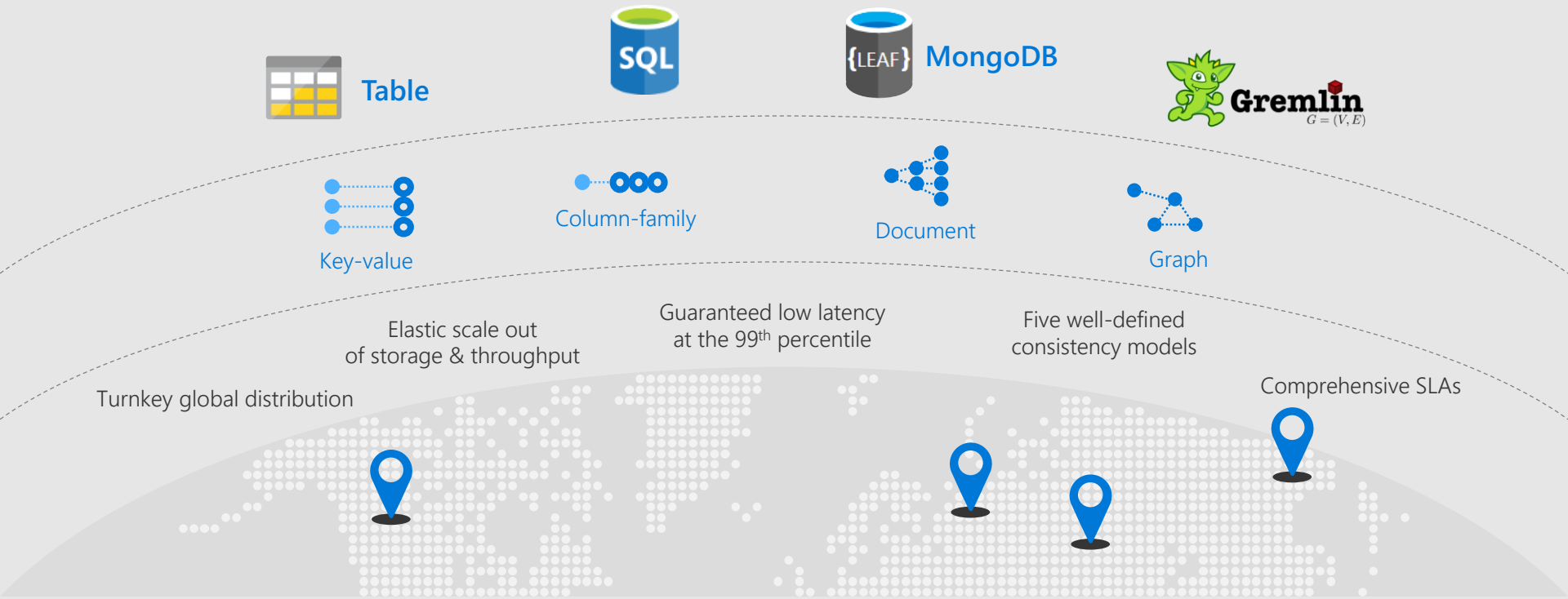
Managed PostgreSQL database service for app developers

Integrate the PostgreSQL community edition with Azure for scalability, high-availability, and your choice of languages and frameworks.

1. Use native PostgreSQL that’s fully managed
2. Choose your languages and frameworks
3. Built-in high availability
4. Scale within seconds
5. Simple and Flexible pricing
6. Unparalleled security and reach

Azure Cosmos DB

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



- SQL API: A schema-less JSON database engine with rich SQL querying capabilities.
- MongoDB API: A massively scalable *MongoDB-as-a-Service* powered by Azure Cosmos DB platform. Compatible with existing MongoDB libraries, drivers, tools, and applications.
- Cassandra API: A globally distributed Cassandra-as-a-Service powered by Azure Cosmos DB platform. Compatible with existing Apache Cassandra libraries, drivers, tools, and applications.
- Graph (Gremlin) API: A fully managed, horizontally scalable graph database service that makes it easy to build and run applications that work with highly connected datasets supporting Open Graph APIs (based on the Apache TinkerPop specification, Apache Gremlin).
- Table API: A key-value database service built to provide premium capabilities (for example, automatic indexing, guaranteed low latency, global distribution) to existing Azure Table storage applications without making any app changes.

					
SQL API					
MongoDB API					
Graph API					
Table API					
Cassandra API					

Cosmos DB

- Azure Cosmos DB is a globally distributed database service
- “one of the supported APIs and data models”
- Document != *.docx
- Document == JSON
- NoSQL, Schema free database
- 99.99% availability within a single region.
- It is a good choice for new web, mobile, gaming, and IoT applications
- **HIPAA-compliant**
- PaaS

DocumentDB Programming Options

Server-side programming options

Stored procs, triggers, & user-defined functions

Written in JavaScript

https://docs.microsoft.com/en-us/azure/cosmos-db/programming

DocumentDB Resources

Stored Proc Programming Video

https://docs.microsoft.com/en-us/azure/cosmos-db/programming

DocumentDB vs MongoDB

https://medium.com/@th0maswe1ss/azure-documentdb-vs-mongodb-6d5806c16239

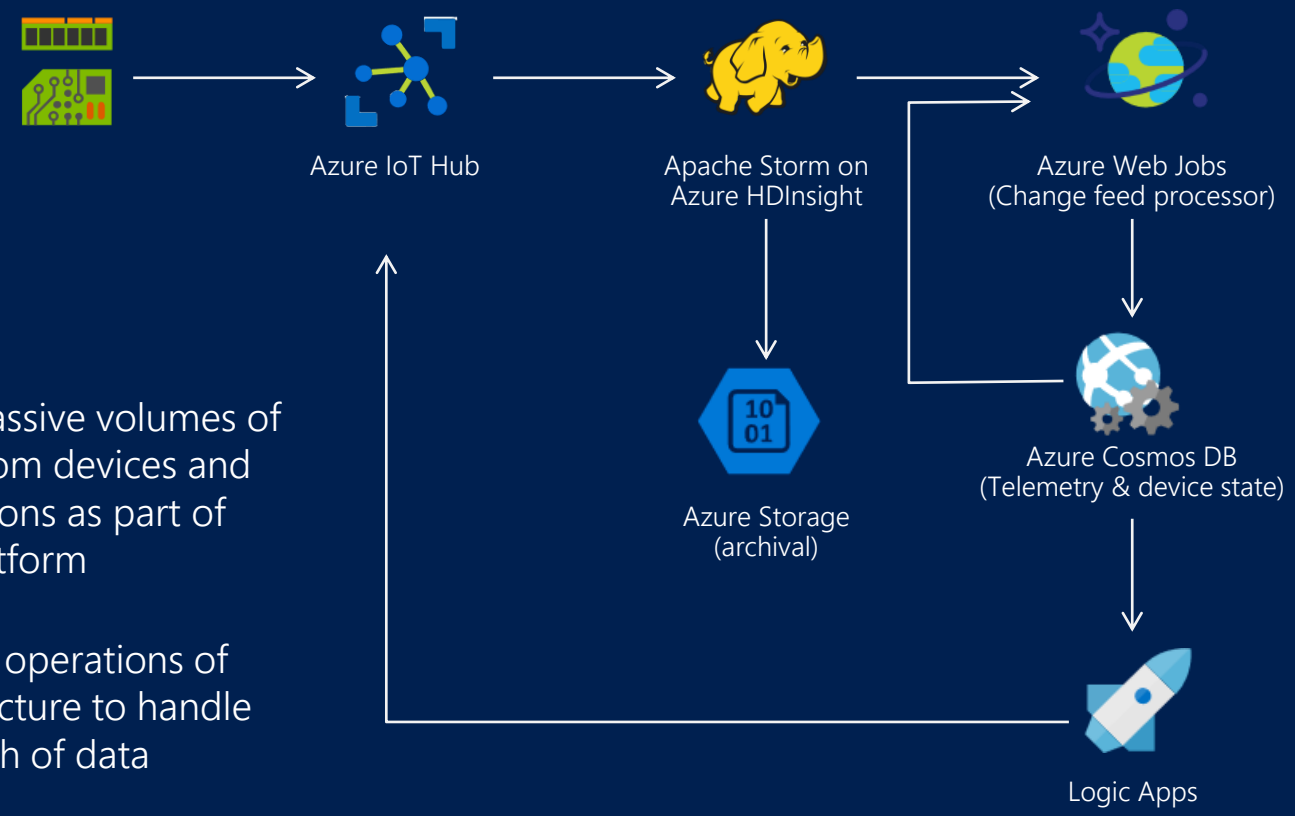
Getting Started

https://docs.microsoft.com/en-us/azure/cosmos-db/documentdb-get-started

Azure CosmosDb KeyWords

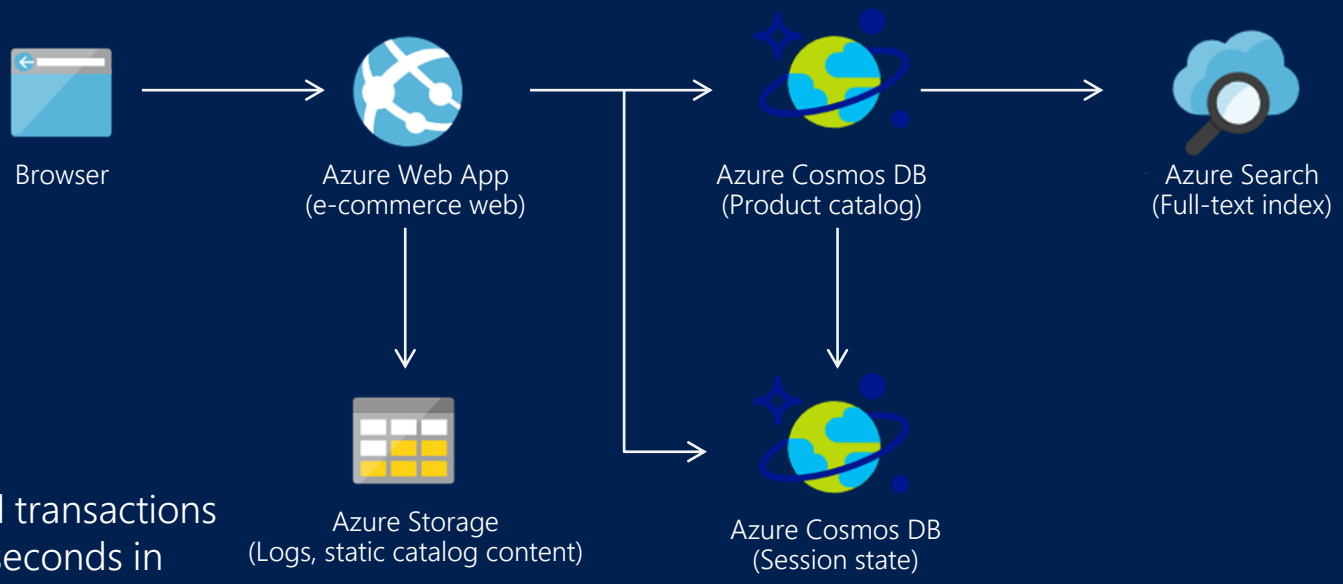
- Request Unit(RU)
- Key-Value Pair
- No SQL
- Schema less
- ARS (atoms, records and sequences)
- Master Key
- PreferredLocations
- Collection
- Connection String
- Indexing Policy
- Partition Key
- Row Key

IoT and Telematics Solution Architecture



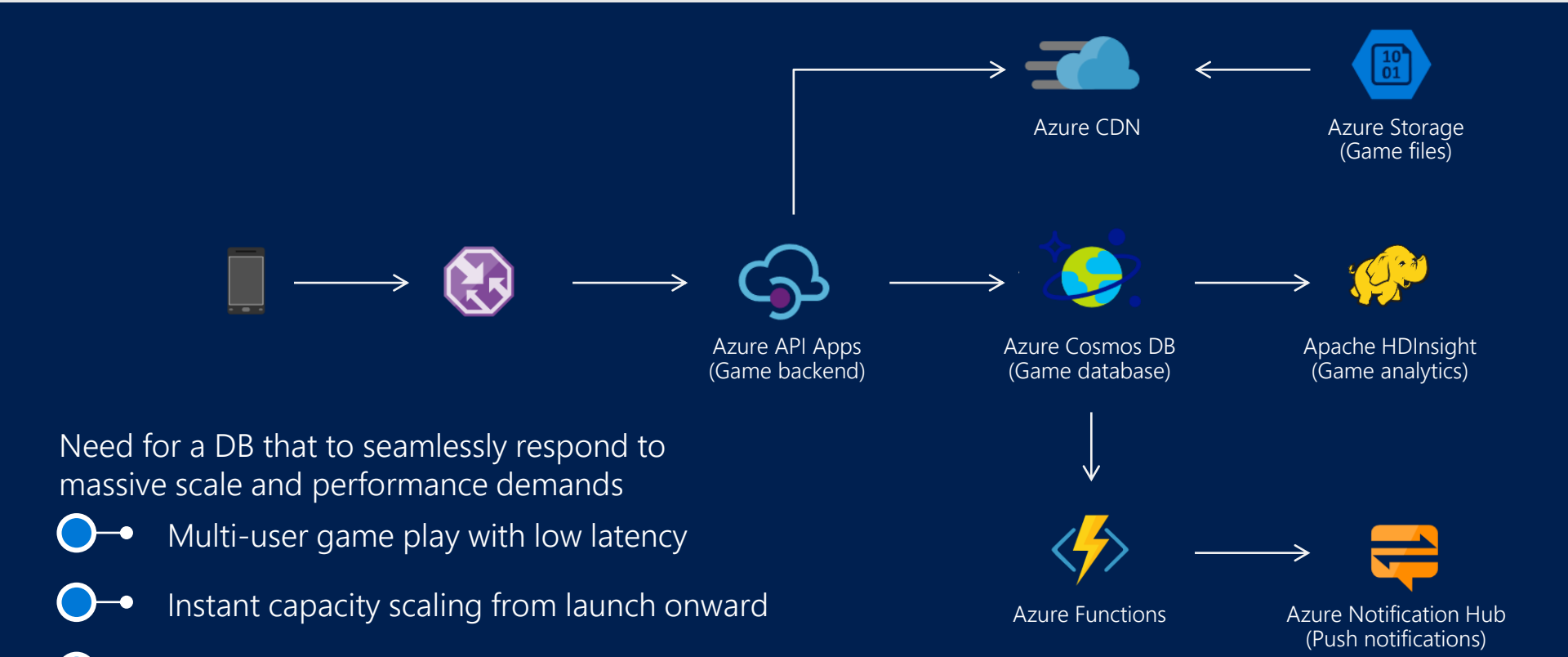
- Need to ingest massive volumes of diagnostic data from devices and take real-time actions as part of connected car platform
- Management and operations of database infrastructure to handle exponential growth of data

Retail and Marketing Solution Architecture



- Process ms of retail transactions per second in milliseconds in inventory pipeline during peak
- Fast development cycles and loosely coupled micro-services

Gaming



Need for a DB that to seamlessly respond to massive scale and performance demands

- Multi-user game play with low latency
- Instant capacity scaling from launch onward
- Uninterrupted global user experience

Scenarios

- Horizontal
  - Serverless, Real time
  - Time Series, Audit log
  - Social Signals
  - Mobile notification/logging
  - Fraud Detection – WITH SPARK or other Analytics
  - Personalization/Recommendation store
  - Content Metadata
  - C360
  - User profile store
    - Customer Registration
- Security
  - Audit log
  - SOA request logging
  - NW request logging

- Ecommerce/Retail
  - Order payment management logging
  - C360
  - Shipment logging
  - Supply chain
  - Catalogs
  - Inventory
  - User profile store
    - Customer Registration
- Messaging/Chat Channel
  - Inbox
- Energy/Manufacturing/Iot
  - Energy meters measurements
  - Readings/gauges measurements
  - C360
- Telecom –
  - User profile management
  - C360

Concepts

- A technical overview of Cosmos DB
- Global distribution with Azure Cosmos DB.
- Automatic regional failover for business continuity in Azure Cosmos DB
- How to partition and scale in Azure Cosmos DB
- Tunable data consistency levels in Azure Cosmos DB
- Performance and scale testing with Azure Cosmos DB
- Schema-agnostic indexing in Cosmos DB (VLDB paper on how we do indexing in Cosmos DB)
- How does Azure Cosmos DB index data
- Request Units in Azure Cosmos DB
- Expire data in Azure Cosmos DB collections automatically with time to live
- Automatic online backup and restore with Azure Cosmos DB
- Integration of Cosmos DB with Spark
- Azure Cosmos DB SLA