General Information

Sunday, June 5, 2022 11:44 AM

Course Official Website
https://docs.microsoft.com/en-us/learn/certifications/exams/az-305

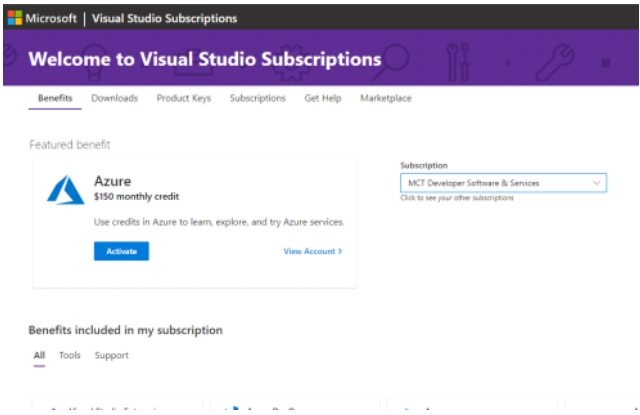
Online Courseware
https://docs.microsoft.com/en-us/users/msftofficialcurriculum-4292/collections/zwm5cy2ownzz08

AZ-900 hands-on lab
https://microsoftlearning.github.io/AZ-900T0x-MicrosoftAzureFundamentals/

Github Version of this notes
https://github.com/Cyrus-Sir/hkic-az-305

Free Azure

https://my.visualstudio.com/



Exam materials

Sunday, June 5, 2022 8:44 PM

Suggested study guide
https://www.thomasmaurer.ch/2021/10/az-305-study-guide-azure-solutions-architect

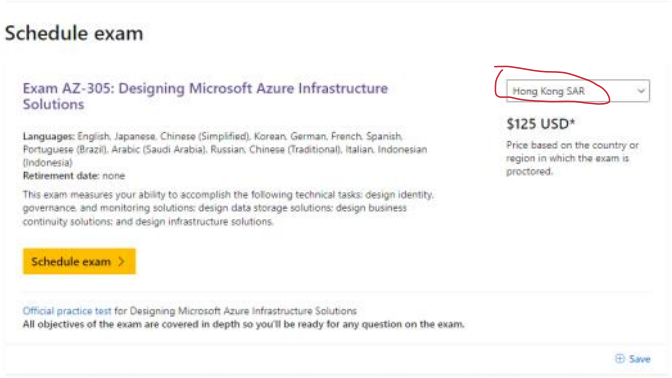
Certification Exam

Certification exams measure your ability to accomplish certain technical tasks for a job role. The study areas are based on the Job Task Analysis that determines what day-to-day tasks are performed in this role.

Each functional area has a percentage indicating the relative weight of the area on the exam. The higher the percentage, the more questions you are likely to see in that area.

Study Area	Percentage
Design identity, governance, and monitoring solutions	25-30%
Design data storage solutions	25-30%
Design business continuity solutions	10-15%
Design infrastructure solutions	25-30%

Exam



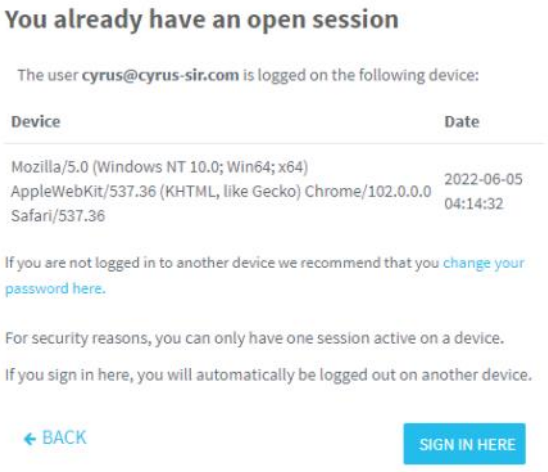
Practical Test (Cannot Share/Concurrent Login)

If you wants to have individual access, you can purchase from me, I have 40% off discount

https://marketplace.measureup.com/login

cyrus@cyrus-sir.com

When you see this, ask is there anyone is using in the WhatsApp Group, before you logout someone

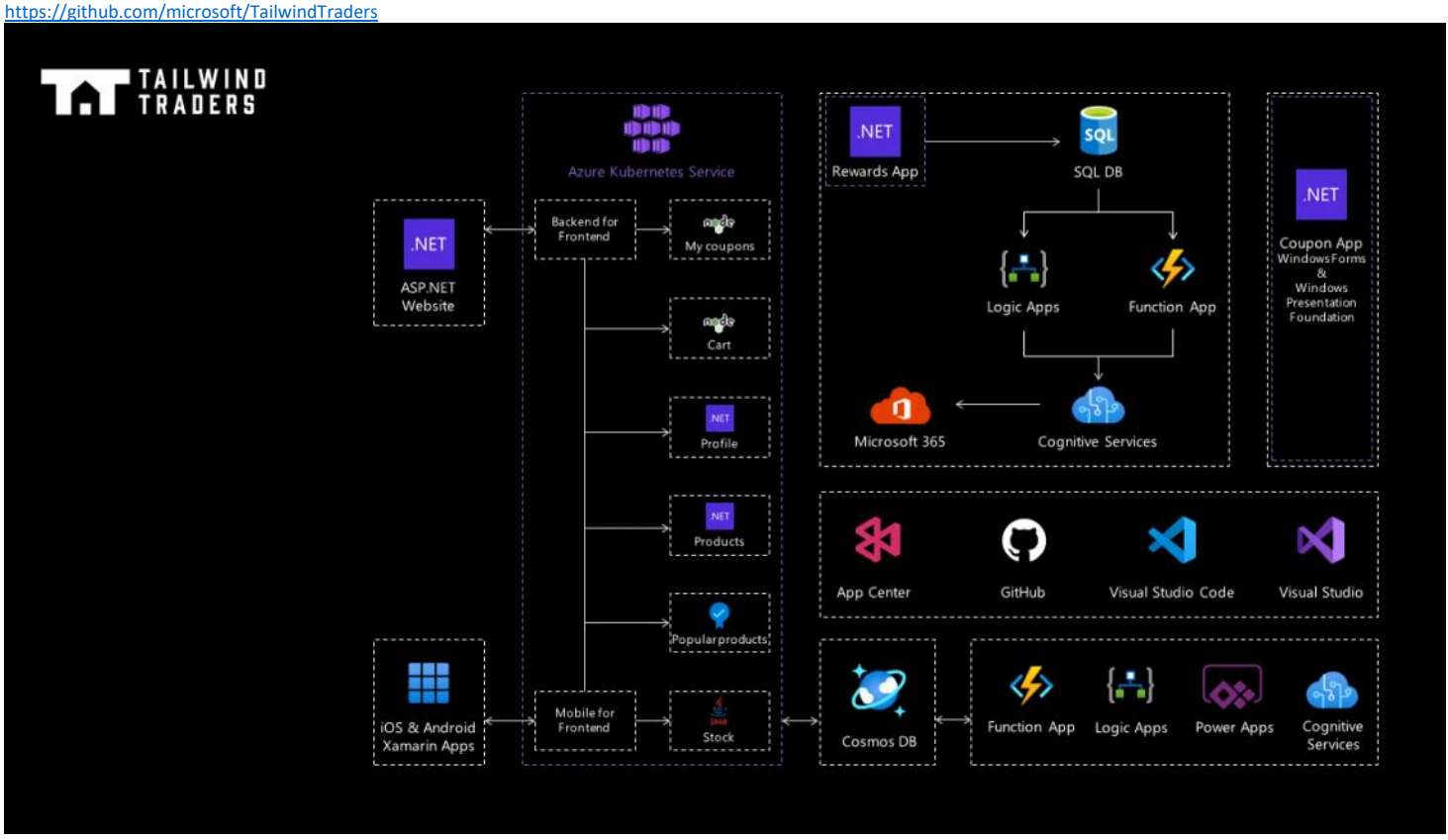


Tailwind Traders

Monday, June 6, 2022 12:16 AM

Sample Application

https://github.com/microsoft/TailwindTraders



Tailwind

Sunday, June 5, 2022 3:29 PM

Case Study GitHub

https://github.com/MicrosoftLearning/AZ-305-DesigningMicrosoftAzureInfrastructureSolutions

Design a governance solution

Sunday, June 5, 2022 11:46 PM

Azure built-in roles
https://docs.microsoft.com/en-us/azure/role-based-access-control/built-in-roles

Exercise

Title	URL
Create a management group	https://docs.microsoft.com/en-us/azure/governance/management-groups/create-management-group-portal
Manage your resources with management groups	https://docs.microsoft.com/en-us/azure/governance/management-groups/manage
Protect a storage account from accidental deletion by using a resource lock	https://docs.microsoft.com/en-us/learn/modules/build-cloud-governance-strategy-azure/4-protect-storage-account-resource-lock
Restrict deployments to a specific location by using Azure Policy	https://docs.microsoft.com/en-us/learn/modules/build-cloud-governance-strategy-azure/7-restrict-location-azure-policy
List access using Azure RBAC and the Azure portal	https://docs.microsoft.com/en-us/learn/modules/secure-azure-resources-with-rbac/4-list-access?source=learn
Grant access using Azure RBAC and the Azure portal	https://docs.microsoft.com/en-us/learn/modules/secure-azure-resources-with-rbac/5-grant-access
View activity logs for Azure RBAC changes	https://docs.microsoft.com/en-us/learn/modules/secure-azure-resources-with-rbac/6-view-activity-logs

Knowledge Check

Title	URL
Using Azure RBAC	https://docs.microsoft.com/en-us/learn/modules/secure-azure-resources-with-rbac/7-knowledge-check-rbac
Build a cloud governance strategy on Azure	https://docs.microsoft.com/en-us/learn/modules/build-cloud-governance-strategy-azure/11-knowledge-check
Intro to Azure blueprints	https://docs.microsoft.com/en-us/learn/modules/intro-to-azure-blueprints/5-knowledge-check
Describe core Azure architectural components	https://docs.microsoft.com/en-us/learn/modules/azure-architecture-fundamentals/knowledge-check

Case Study

Requirements

Tailwind Traders is planning on making some significant changes to their governance solution. They have asked for your help with recommendations and questions. Here are the specific requirements.

- Cost and accounting. Tailwind Traders has two core business units that handle Apparel and Sporting Goods. Each of the business units has three departments: Product Development, Marketing, and Sales. Each business unit and subunit will track their Azure spend. At the same time, the Enterprise IT team will handle providing company-wide Azure cost reporting.
- New development project. The company has a new development project for customer feedback. The CFO wants to ensure all costs associated with the project are captured. For the testing phase, workloads should be hosted on lower cost virtual machines. The virtual machines should be named to indicate they are part of the project. Any instances of non-compliance with resource consistency rules should be automatically identified.

Tasks

1. Cost and accounting.
   - What are the different ways Tailwind Traders could organize their subscriptions and management groups? Which would be the best to meet their requirements?

     https://app.diagrams.net/
2. New development project.
   - What are the different ways Tailwind Traders could track costs for the new development project?
   - How are you ensuring compliance with the requirements for virtual machine sizing and naming?

Design a compute solution

Sunday, June 5, 2022 11:46 PM

VM Size prefix meaning
https://azure.microsoft.com/en-us/pricing/details/virtual-machines/series/

Create a virtual machine in the portal
https://microsoftlearning.github.io/AZ-900T0x-MicrosoftAzureFundamentals/Instructions/Walkthroughs/01-Create%20a%20virtual%20machine.html

Run your first Batch job with the Azure CLI
https://docs.microsoft.com/en-us/azure/batch/quick-create-cli
https://docs.microsoft.com/en-us/azure/batch/quick-create-portal

Run a parallel workload with Azure Batch using the .NET API
https://docs.microsoft.com/en-us/azure/batch/tutorial-parallel-dotnet

Create a Java app on Azure App Service
https://docs.microsoft.com/en-us/azure/app-service/quickstart-java?tabs=javase&pivots=platform-windows

Deploy a container instance in Azure using the Azure CLI
https://docs.microsoft.com/en-us/azure/container-instances/container-instances-quickstart

Deploy an Azure Kubernetes Service cluster using the Azure CLI
https://docs.microsoft.com/en-us/azure/aks/learn/quick-kubernetes-deploy-cli

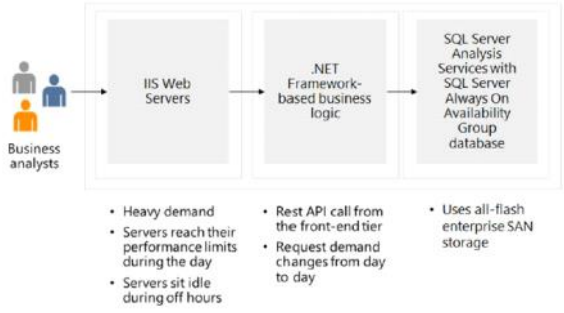
Create a C# function in Azure from the command line
https://docs.microsoft.com/en-us/azure/azure-functions/create-first-function-cli-csharp?tabs=azure-cli%2Cin-process

Module end labs

Create the social media tracker Logic App
https://docs.microsoft.com/en-us/learn/modules/route-and-process-data-logic-apps/4-ex-create-social-media-tracker

Case study

Requirements
Tailwind Traders would like to migrate their product catalog application to the cloud. This application has a traditional 3-tier configuration using SQL Server as the data store. The IT team hopes you can help modernize the application. They have provided this diagram and several areas that could be improved.



- The front-end application is a .NET core-based web app. During peak periods 1750 customers visit the website each hour.
- The application runs on IIS web servers in a front-end tier. This tier handles all customer requests for purchasing products. During the latest holiday sale, the front-end servers reached their performance limits and page loads were lengthy. The IT team has considered adding more servers, but during off hours the servers are often idle.
- The middle tier hosts the business logic that processes customer requests. These requests are often for help desk support. Support requests are queued and lately the wait times have been exceptionally long. Customers are offered email rather than waiting for a representative. But many customers seem frustrated and are disconnecting rather than waiting. Customer requests are 75-125 per hour.
- The back-end tier uses SQL Server database to store customer orders. Currently, the back-end database servers are performing well.
- While high availability is a concern, due to legal requirements the company must keep all the resources in a single region.

Task
- Front-end tier. Which Azure compute service would you recommend for the front-end tier? Explain why you decided on your solution.
- Middle tier. Which Azure compute service would you recommend for the middle tier? Explain why you decided on your solution.

Storage Account

Saturday, June 11, 2022 11:45 AM

Storage Account	Supported Services	Recommended usage
Standard general-purpose v2	Blob (including Data Lake Storage), Queue, and Table storage, Azure Files	Supports all the storage services: Blob, Azure Files, Queue, Disk (Page Blob), and Table.
Premium block blobs	Blob storage (including Data Lake Storage)	Premium block blobs are ideal for applications that require high transaction rates. Also ideal for situations that use smaller objects or require consistently low storage latency. This storage is designed to scale with your applications.
Premium file shares	Azure Files	Recommended for enterprise or high-performance scale applications. Use Premium file shares if you need a storage account that supports both SMB and NFS file shares.
Premium page blobs	Page blobs only	Premium high-performance page blob scenarios. Page blobs are ideal for storing index-based and sparse data structures like OS and data disks for virtual machines and databases.

Considerations

- Cost
- Compliance
- Location
- Replication requirements for different data nature
- Administrative overhead - Prevent deletion / update by date retention? By hold?
- Data sensitivity - Public / Private (Although can use VNet to protect)
- Data isolation - Retention policy differences?

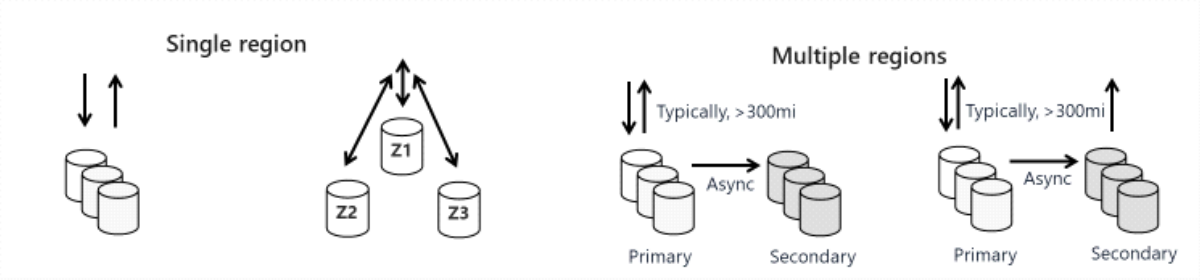
Very important. Remember the endpoint for all storage account services

Storage service	Endpoint
Blob Storage	`https://<storage-account>.blob.core.windows.net`
Static website (Blob Storage)	`https://<storage-account>.web.core.windows.net`
Data Lake Storage Gen2	`https://<storage-account>.dfs.core.windows.net`
Azure Files	`https://<storage-account>.file.core.windows.net`
Queue Storage	`https://<storage-account>.queue.core.windows.net`
Table Storage	`https://<storage-account>.table.core.windows.net`

Redundancy

Saturday, June 11, 2022 11:52 AM

Determine Replication Strategies (1 of 2)



LRS

- Three replicas, one region
- Protects against disk, node, rack failures
- Write is acknowledged when all replicas are committed
- Superior to dual-parity RAID

ZRS

- Three replicas, three zones, one region
- Protects against disk, node, rack, and zone failures
- Synchronous writes to all three zones

GRS

- Six replicas, two regions (three per region)
- Protects against major regional disasters
- Asynchronous copy to secondary

RA-GRS

- GRS + read access to secondary
- Separate secondary endpoint
- Recovery point objective (RPO) delay to secondary can be queried

Continued next slide

Determine Replication Strategies (2 of 2)

GZRS

- Six replicas, 3+1 zones, two regions
- Protects against disk, node, rack, zone, and region failures
- Synchronous writes to all three zones and asynchronous copy to secondary

RA-GZRS

- GZRS + read access to secondary
- Separate secondary endpoint
- RPO delay to secondary can be queried

Blob Storage

Saturday, June 11, 2022 11:58 AM

Access Tier	Immutable Storage Policy
Premium blob storage	Legal hold policies
Hot, cool, and archive access tiers	Time-based retention policies

Feature	Premium	Hot tier	Cool tier	Archive tier
Availability	99.9%	99.9%	99%	Offline
Availability (RA-GRS reads)	N/A	99.99%	99.9%	Offline
Usage charges	Higher storage costs, lower access, and transaction cost	Higher storage costs, lower access, and transaction costs	Lower storage costs, higher access, and transaction costs	Lowest storage costs, highest access, and transaction costs
Minimum storage duration	N/A	N/A	30 days	180 days Subject to early deletion charge
Latency (time to first byte)	Single-digit milliseconds	milliseconds	milliseconds	hours
Use case	Date that small and requires frequent + fast updates Analytical Data	Application Data	Short-term backup Disaster recovery datasets Older media content wouldn't be viewed frequently but must be available immediately	Secondary backups Legally required compliance information

Time-based retention policies

- Before expire : can create and read data / Not update & delete
- After expire : can delete, but not edit

Legal hold policies

- Hold : can create + read; Not update + delete
- Unhold : do everything

Azure Files

Saturday, June 11, 2022 11:26 AM

Performance level	Latency	IOPS	Bandwidth
Standard	Double-digit ms	10,000 IOPS	300-MBps
Premium	Single-digit ms	100,000 IOPS	5-GBps

Premium only support ZRS storage in some region only

Storage tier	Usage
Premium	File shares are backed by solid-state drives (SSDs) and provide consistent high performance and low latency. Used for the most intensive IO workloads. Suitable workloads include databases, web site hosting, and development environments. Can be used with both Server Message Block (SMB) and Network File System (NFS) protocols.
Transaction optimized	Used for transaction heavy workloads that don't need the latency offered by premium file shares. File shares are offered on the standard storage hardware backed by hard disk drives (HDDs).
Hot	Storage optimized for general purpose file sharing scenarios such as team shares. Offered on standard storage hardware backed by HDDs.
Cool	Cost-efficient storage optimized for online archive storage scenarios. Offered on storage hardware backed by HDDs.

File Sync Cloud Tier and Policy
https://docs.microsoft.com/en-us/azure/storage/file-sync/file-sync-cloud-tiering-policy
https://docs.microsoft.com/en-us/azure/storage/file-sync/file-sync-cloud-tiering-overview

Key points

- Volume free space policy
- Date policy
- Windows Server data deduplication
- Cloud tiering heatmap
- Proactive recalling

File Sync Labs
https://docs.microsoft.com/en-us/azure/storage/file-sync/file-sync-extend-servers

Azure Disk

Saturday, June 11, 2022 12:08 PM

Detail	Ultra-disk	Premium SSD	Standard SSD	Standard HDD
Disk type	SSD	SSD	SSD	HDD
Scenario	IO-intensive workloads such as SAP HANA, top tier databases (for example, SQL, Oracle), and other transaction-heavy workloads.	Production and performance sensitive workloads	Web servers, lightly used enterprise applications and dev/test	Backup, non-critical, infrequent access
Max throughput	2,000 MB/s	900 MB/s	750 MB/s	500 MB/s
Max IOPS	160,000	20,000	6,000	2,000

Disks encryption

- Azure Disk Encryption
  Encrypt the VHD
  Only the VM that own the disk can access the disk image
  DM-Crypt for Linux / BitLocker for Windows
  ADE Prerequisites
  a. Create a key vault.
  b. Set the key vault access policy to support disk encryption.
     i. Disk encryption - Required for Azure Disk encryption.
     ii. Deployment – Used by Compute Resource when defined in deployment
     iii. Template deployment – used by template deployment
  c. Use the key vault to store the encryption keys for ADE.

- Server-Side Encryption (encryption-at-rest)
  Encrypt physical disks in the data center
  When the data is accessed from the disk, it's decrypted and loaded into memory

- Encryption at host
  VM host encrypt the disk and put the encrypted data into Azure Storage

Different type of Disk

https://docs.microsoft.com/en-us/learn/modules/choose-the-right-disk-storage-for-vm-workload/2-managed-unmanaged-local-disk-storage

https://docs.microsoft.com/en-us/learn/modules/choose-the-right-disk-storage-for-vm-workload/3-disk-types-for-virtual-machines

Storage security

Saturday, June 11, 2022 2:21 PM

Use Shared Access Signatures

- Remember to SAS meaning. Try to generate a SAS by yourself

BlobEndpoint=https://cyrus822az305demo.blob.core.windows.net/;

QueueEndpoint=https://cyrus822az305demo.queue.core.windows.net/;

FileEndpoint=https://cyrus822az305demo.file.core.windows.net/;

TableEndpoint=https://cyrus822az305demo.table.core.windows.net/;

SharedAccessSignature=

sv=2021-06-08&

ss=bft&

srt=so&

sp=rwdlacuitfx&

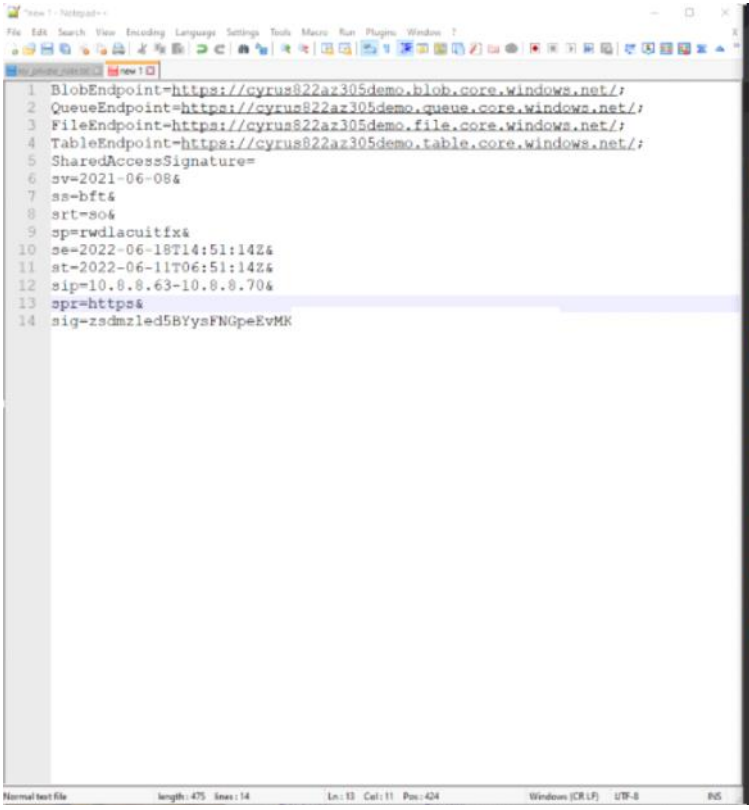
se=2022-06-18T14:51:14Z&

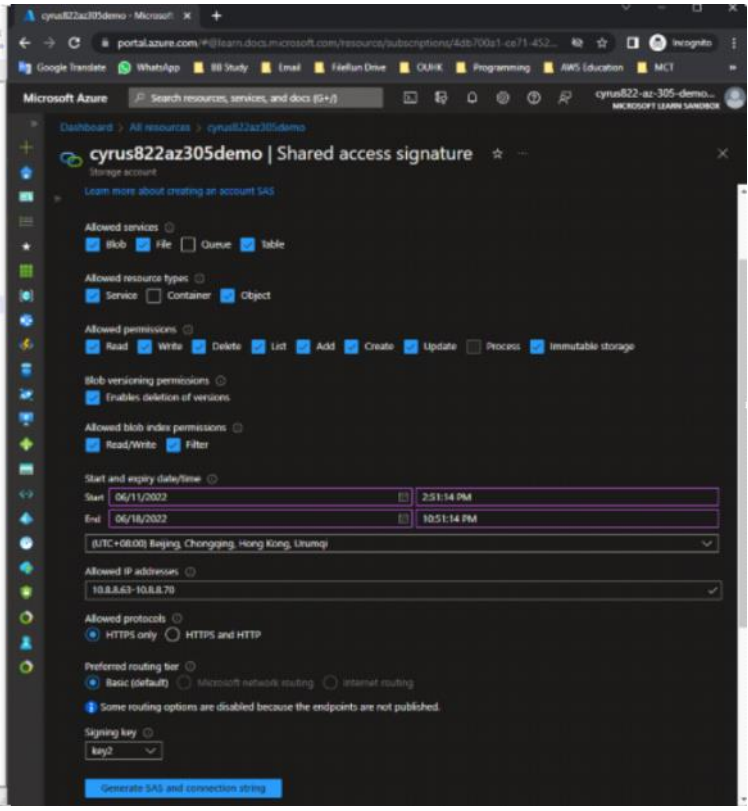
st=2022-06-11T06:51:14Z&

sip=10.8.8.63-10.8.8.70&

spr=https&

sig=zsdmzledXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX





Enable firewall policies and rules

Restrict network access using service endpoints

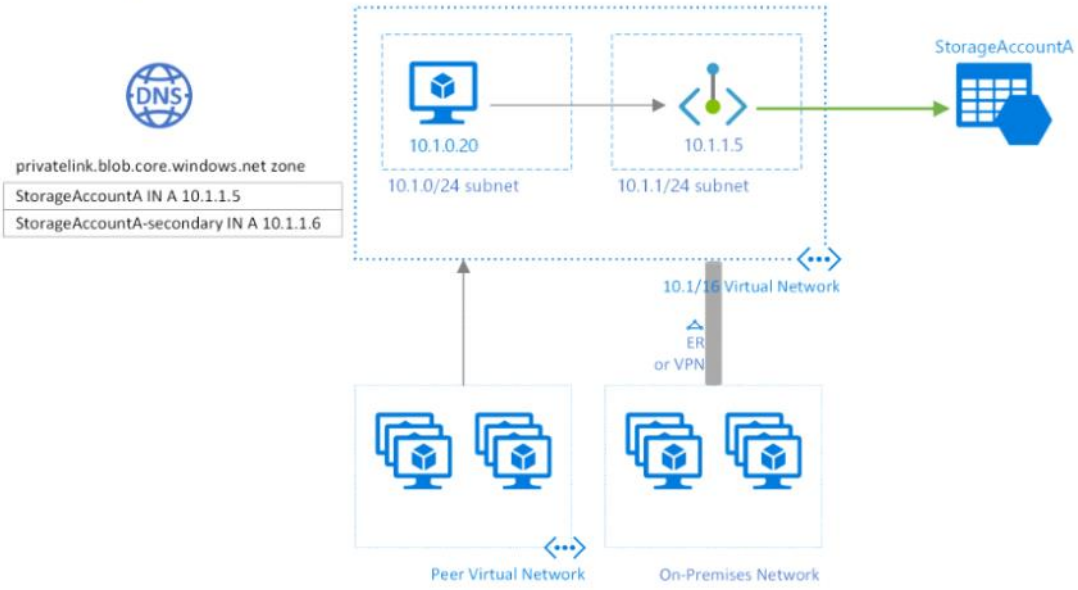
- Enable private IP addresses in the VNet to reach the service endpoint
- Enables on-premises networks to access resources using NAT IP addresses

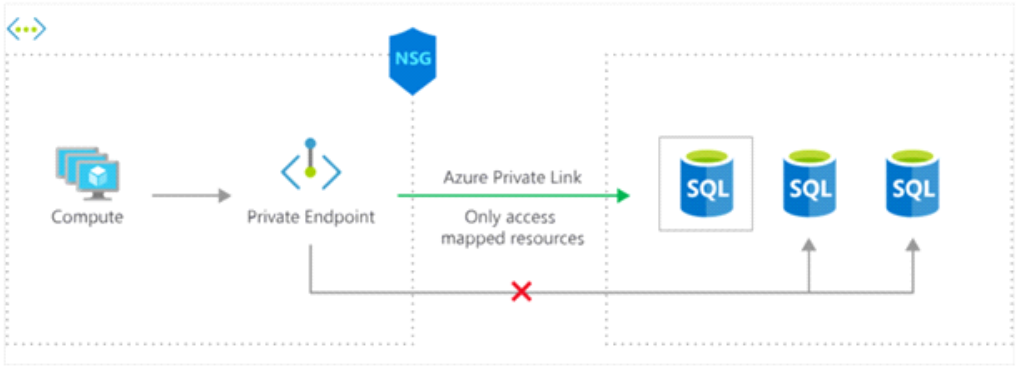
Use private endpoints (Private Link)

Enable secure transfer

Use Customer-managed keys

- Customer-managed keys must be stored in Azure Key Vault


DNS
privatelink.blob.core.windows.net zone
StorageAccountA IN A 10.1.1.5
StorageAccountA-secondary IN A 10.1.1.6
10.1.0.20
10.1.0/24 subnet
10.1.1.5
10.1.1/24 subnet
StorageAccountA
10.1/16 Virtual Network
ER
or VPN
Peer Virtual Network
On-Premises Network


NSG
Compute
Private Endpoint
Azure Private Link
Only access
mapped resources
SQL
SQL
SQL

Exercise

Saturday, June 11, 2022 4:37 PM

Disk Security
https://docs.microsoft.com/en-us/learn/modules/secure-your-azure-virtual-machine-disks/7-knowledge-check

Azure SQL DB

Saturday, June 11, 2022 4:10 PM

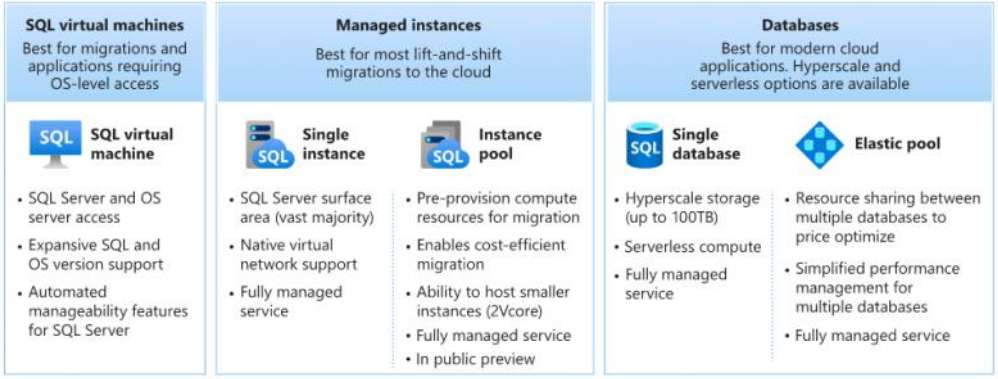
Decision Making Flow
Deployment Model > Purchase Model > Service Tier

Features comparison: Azure SQL Database and Azure SQL Managed Instance
https://docs.microsoft.com/en-us/azure/azure-sql/database/features-comparison?view=azuresql

Deployment Model

- SQL Server on Azure VMs
- Managed instances:
  - Single instances
  - Instance pool
- Databases:
  - Single database
  - Elastic pool

Recommendation	Requirement
SQL Virtual machines	When considering migrations and applications requiring OS level access
Managed Instances	When considering Lift and Shift migrations to the cloud
Databases	When considering modern cloud applications solution



Azure SQL feature

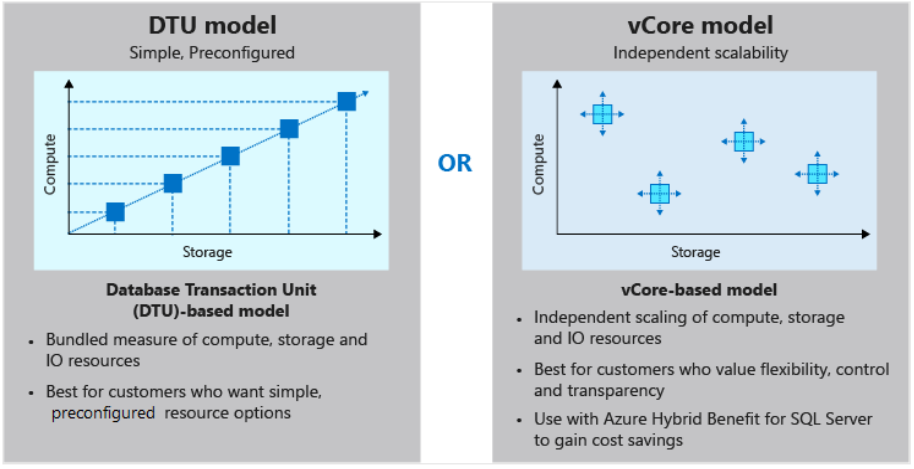
- Very large databases (currently up to 100TB)
- Autoscaling for unpredictable workloads (serverless)

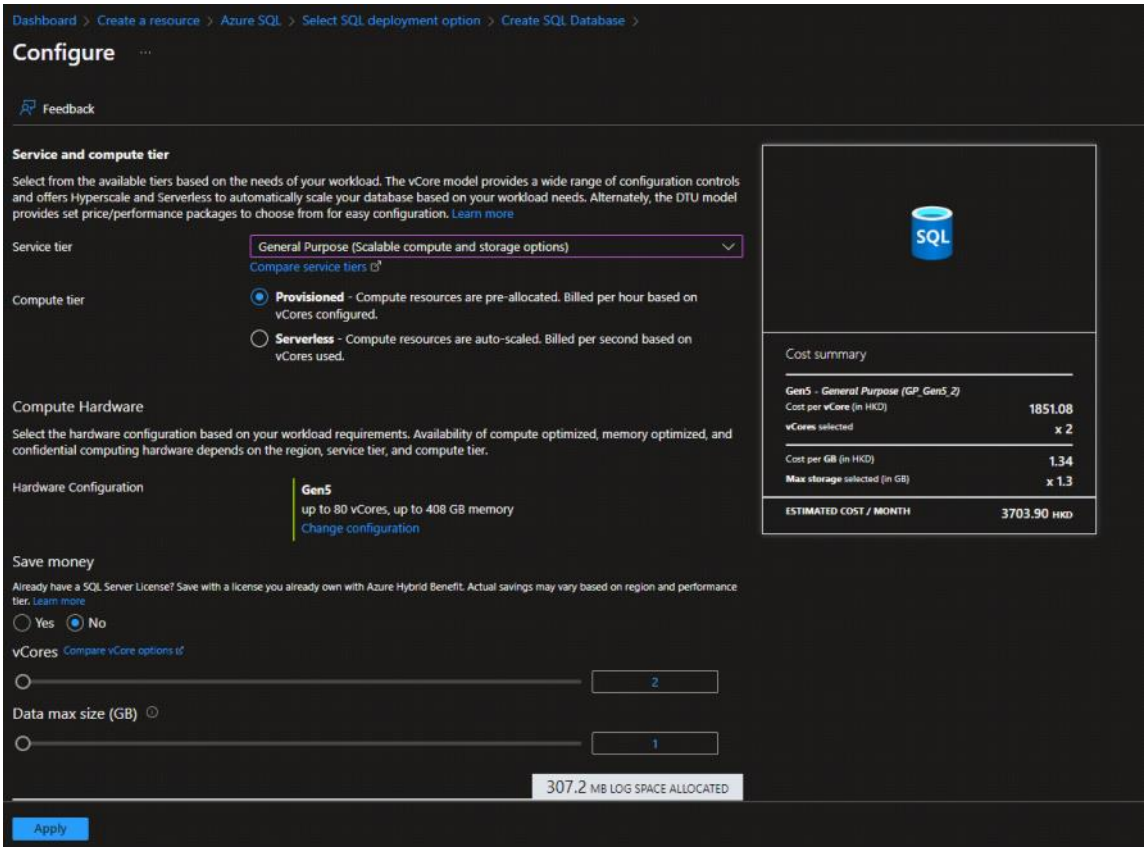
Elastic pool feature

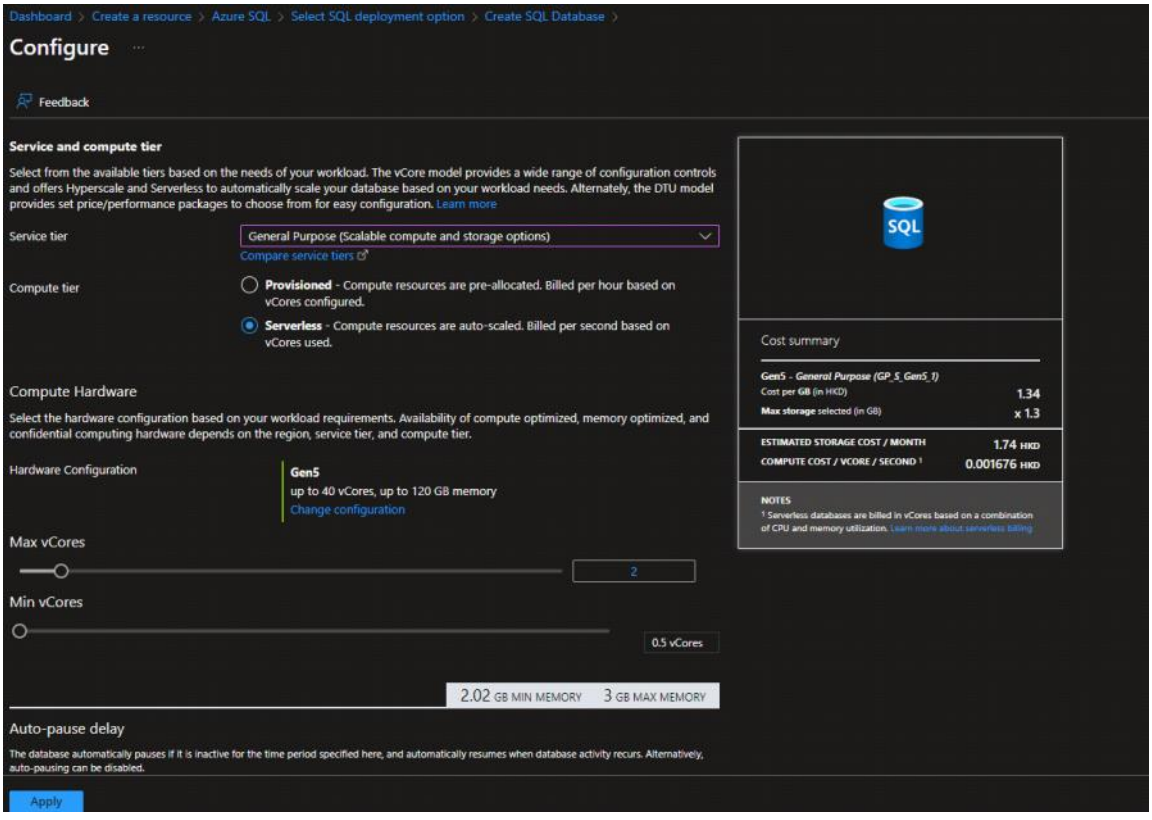
- Buy compute + storage, share across multiple DB
- Each database can use the resources they need, within the limits you set, depending on current load
- Billed for each hour a pool exists at the highest eDTU or vCores

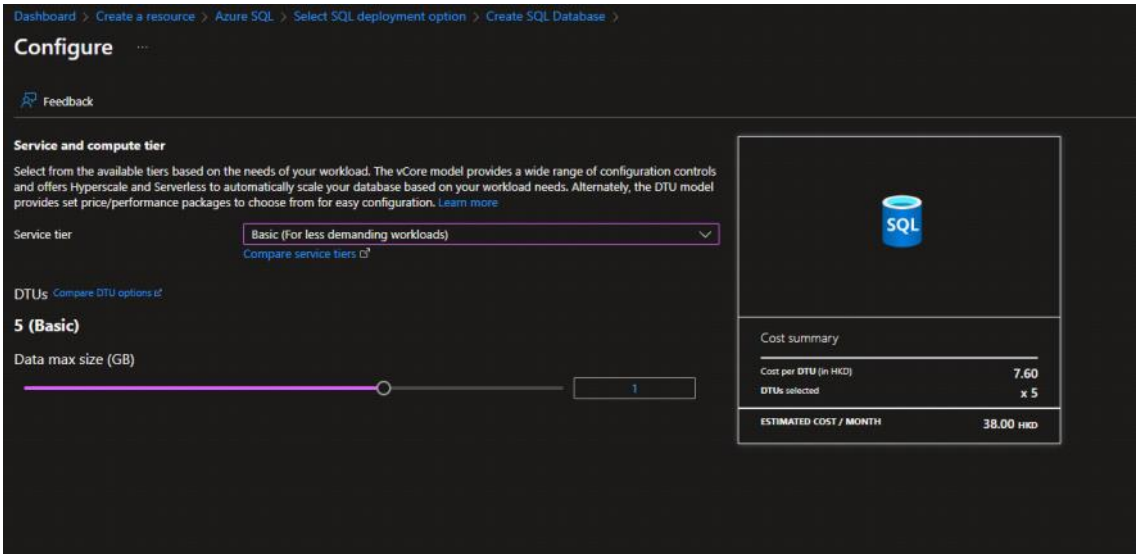
Provisioned + vCore
==> Hybrid Benefit ==> Save money

Pricing model
https://docs.microsoft.com/en-us/azure/azure-sql/database/purchasing-models?view=azuresql



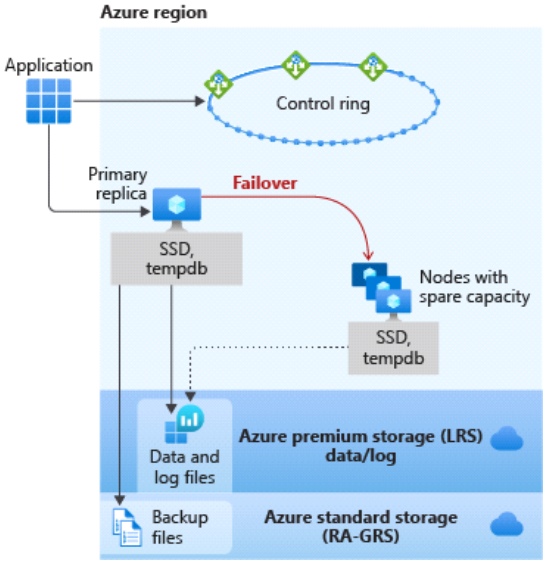

Dashboard > Create a resource > Azure SQL > Select SQL deployment option > Create SQL Database >
Configure
Feedback
Service and compute tier
Select from the available tiers based on the needs of your workload. The vCore model provides a wide range of configuration controls and offers Hyperscale and Serverless to automatically scale your database based on your workload needs. Alternately, the DTU model provides set price/performance packages to choose from for easy configuration. Learn more
Service tier
General Purpose (Scalable compute and storage options)
Compare service tiers
Compute tier
Provisioned - Compute resources are pre-allocated. Billed per hour based on vCores configured.
Serverless - Compute resources are auto-scaled. Billed per second based on vCores used.
Compute Hardware
Select the hardware configuration based on your workload requirements. Availability of compute optimized, memory optimized, and confidential computing hardware depends on the region, service tier, and compute tier.
Hardware Configuration
Gen5
up to 80 vCores, up to 408 GB memory
Change configuration
Save money
Already have a SQL Server License? Save with a license you already own with Azure Hybrid Benefit. Actual savings may vary based on region and performance tier. Learn more
Yes
No
vCores
Compare vCore options
2
Data max size (GB)
1
307.2 MB LOG SPACE ALLOCATED
Apply
SQL
Cost summary
Gen5 - General Purpose (GP_Gen5_2)
Cost per vCore (in HKD) 1851.08
vCores selected x 2
Cost per GB (in HKD) 1.34
Max storage selected (in GB) x 1.3
ESTIMATED COST / MONTH 3703.90 HKD


Dashboard > Create a resource > Azure SQL > Select SQL deployment option > Create SQL Database >
Configure
Feedback
Service and compute tier
Select from the available tiers based on the needs of your workload. The vCore model provides a wide range of configuration controls and offers Hyperscale and Serverless to automatically scale your database based on your workload needs. Alternately, the DTU model provides set price/performance packages to choose from for easy configuration. Learn more
Service tier
General Purpose (Scalable compute and storage options)
Compare service tiers
Compute tier
Provisioned - Compute resources are pre-allocated. Billed per hour based on vCores configured.
Serverless - Compute resources are auto-scaled. Billed per second based on vCores used.
Compute Hardware
Select the hardware configuration based on your workload requirements. Availability of compute optimized, memory optimized, and confidential computing hardware depends on the region, service tier, and compute tier.
Hardware Configuration
Gen5
up to 40 vCores, up to 120 GB memory
Change configuration
Max vCores
2
Min vCores
0.5 vCores
2.02 GB MIN MEMORY
3 GB MAX MEMORY
Auto-pause delay
The database automatically pauses if it is inactive for the time period specified here, and automatically resumes when database activity recurs. Alternatively, auto-pausing can be disabled.
Apply
SQL
Cost summary
Gen5 - General Purpose (GP_S_Gen5_1)
Cost per GB (in HKD) 1.34
Max storage selected (in GB) x 1.3
ESTIMATED STORAGE COST / MONTH 1.74 HKD
COMPUTE COST / VCORE / SECOND 1 0.001676 HKD
NOTES
1 Serverless databases are billed in vCores based on a combination of CPU and memory utilization. Learn more about serverless billing



Recommendation	Requirement
General Purpose	When you need balanced compute and storage options for business workloads
Business Critical	When you need low latency requirements and highest resilience to failures for business applications
Hyperscale	When you need highly scalable storage and have read-scale requirements for business workloads

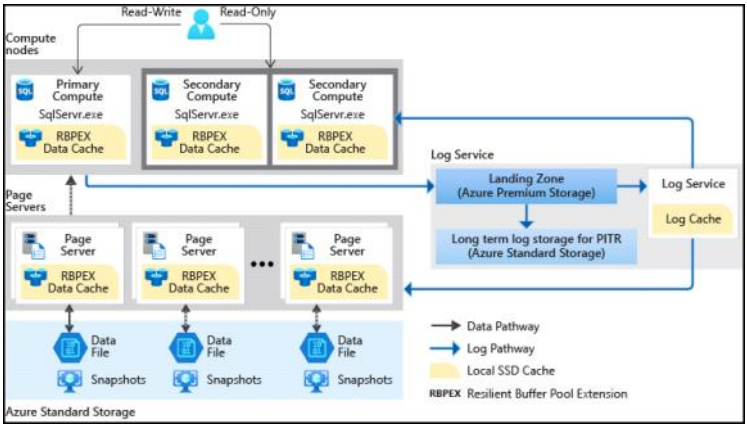
General Purpose



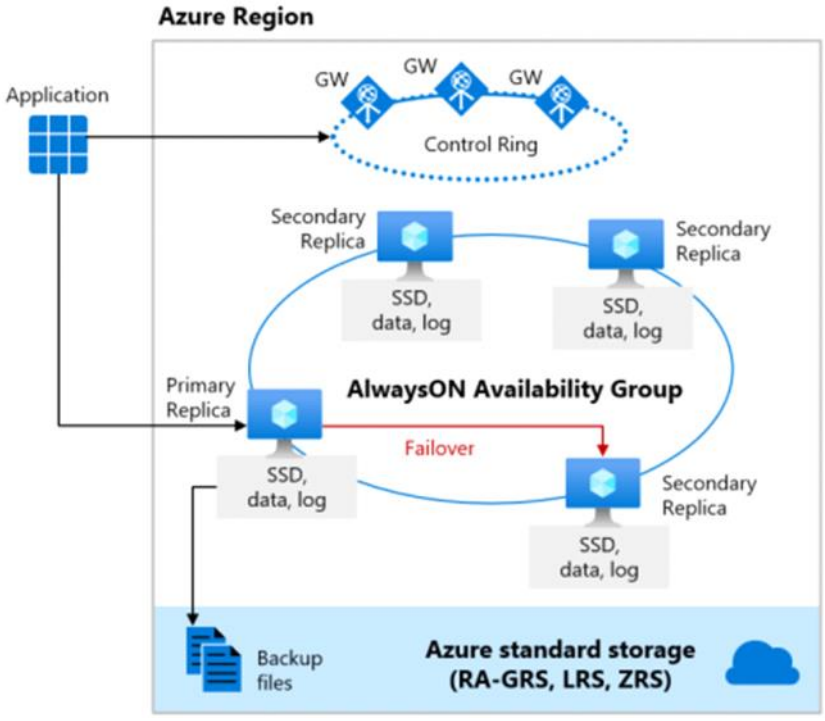
Hyperscale

https://docs.microsoft.com/en-us/azure/azure-sql/database/hyperscale-architecture?view=azuresql

- Azure SQL DB only
- Up to 100TB
- Compute Node + Page Server + Log Services
- 2nd Compute node are hot standby, can read
- Page server serve 128GB or 1 TB page data. Also have replica
- Log from primary node send to Log services
- Log services broadcast to all 2nd Compute and Page server. Data will be sync
- Restores in minutes rather than hours and days



Business Critical


Azure Region
Application
GW
GW
GW
Control Ring
Secondary Replica
SSD, data, log
Secondary Replica
SSD, data, log
Primary Replica
AlwaysON Availability Group
Failover
SSD, data, log
Secondary Replica
SSD, data, log
Backup files
Azure standard storage (RA-GRS, LRS, ZRS)

SQL Managed Instance

Sunday, June 12, 2022 10:27 PM

Features

- Have Instance-scope features, but no need to manage OS
- Azure manage something for you
  - Automatic patching and version updates
  - Automated backups
  - High availability
  - Reduced management overhead

Instance Scope Features

- SQL Server Agent
- Service Broker
- Common language runtime (CLR)
- Database Mail
- Linked servers
- Distributed transactions (preview)
- Machine Learning Services

Features comparison: Azure SQL Database and Azure SQL Managed Instance
https://docs.microsoft.com/en-us/azure/azure-sql/database/features-comparison?view=azuresql

Major differences that may asked in exam

- BACKUP command
- Azure Active Directory (Azure AD) authentication
- Common language runtime - CLR
- Cross-database/three-part name queries
- Linked servers
- Windows authentication

SQL Server on Azure VM

Sunday, June 12, 2022 10:34 PM

- All your SQL Server skills should directly transfer, though Azure can help automate backups and security patches.
- You have access to the full capabilities of SQL Server
- You're responsible for updating and patching the OS and SQL Server

Labs

Sunday, June 12, 2022 11:38 PM

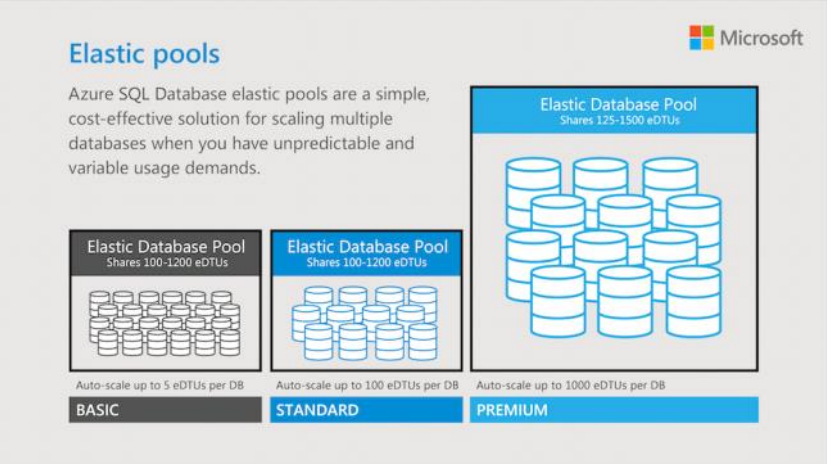
Create a database
https://docs.microsoft.com/en-us/learn/modules/azure-database-fundamentals/exercise-create-sql-database

Scalability

Monday, June 13, 2022 9:50 AM

Vertical – Scale up
Horizontal – Scale out

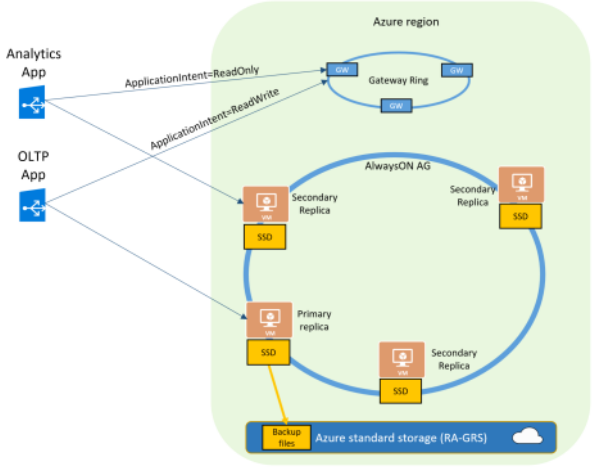
Scale up – Elastic Pool



Scale Out

- Read only scale out (Similar to CQRS)
- Sharding (Partitioning)

Azure SQL Managed Instance	Azure SQL Database
For the basic, standard and general purpose tier, read scale-out feature is unavailable	For the basic, standard and general purpose tier, read scale-out feature is unavailable
For the Business Critical tier, read scale-out is auto-provisioned	For the Premium and Business Critical tier, read scale-out is auto-provisioned
	Read scale-put feature is available in Hyperscale tier if at least one secondary replica is created



Reasons for Sharding include

- If the total amount of data is too large to fit constraints of a single database
- If the transaction throughput of the overall workload exceeds capacities of an individual database
- When different customers or tenents' data needs physical isolation from each other
- Within an organization, there is a geographical separation of data for compliance reasons

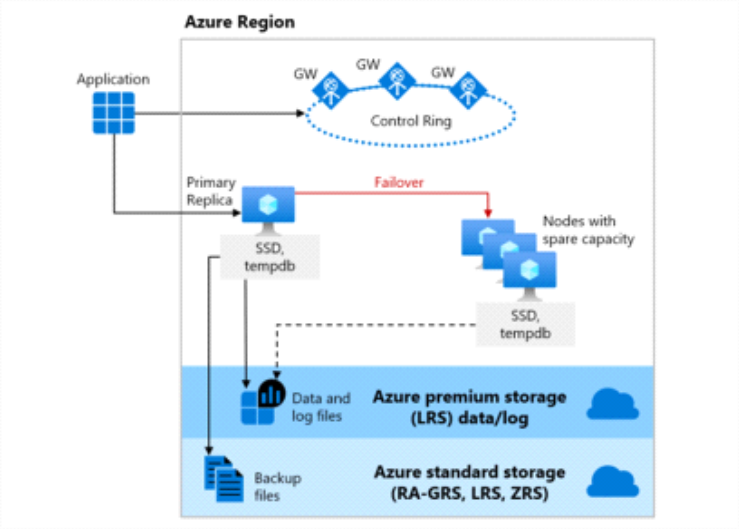
Availability

Monday, June 13, 2022 10:02 AM

High availability with the General Purpose/Standard tier

Azure SQL Database offers three service tiers that are designed for different types of applications:

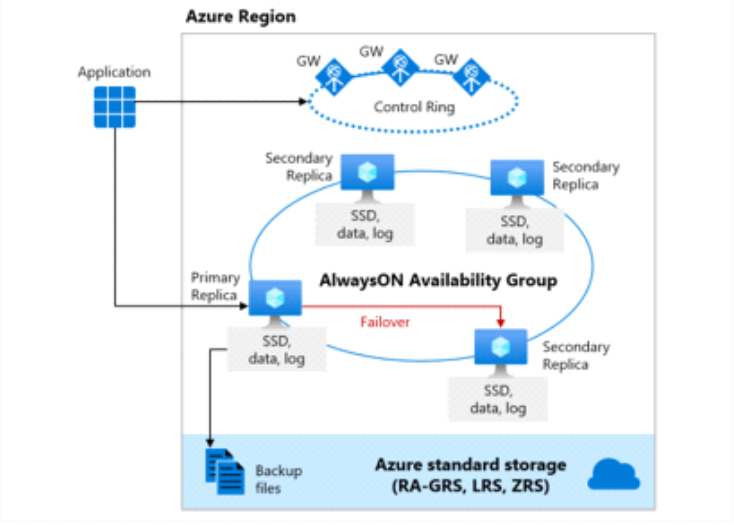
- Designed for common workloads
- Budget oriented balanced compute and storage
- Uses nodes with spare capacity to spin up a new SQL Server instances
- Uses LRS and RA-GRS (backup files)



High availability with the Business Critical/Premium tier

Azure SQL Database offers three service tiers that are designed for different types of applications:

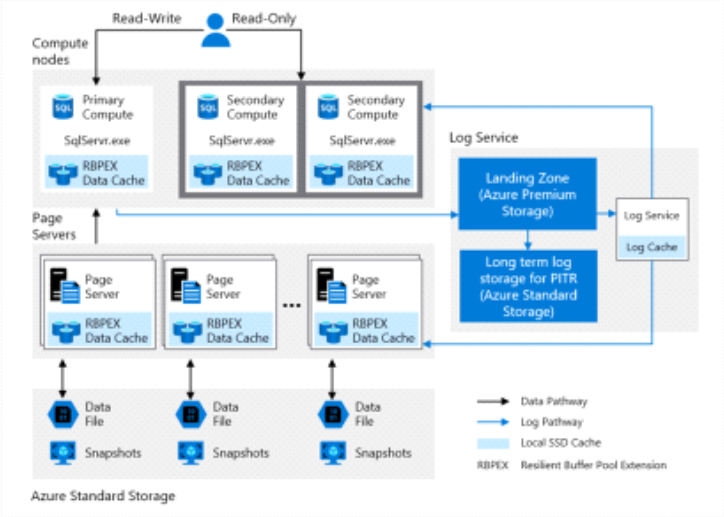
- Designed for OLTP applications
- High transaction rate and low I/O latency
- Offers the highest resilience to failures by using several isolated replicas
- Deploys an Always On availability group using multiple synchronously updated replicas
- Uses local SSD storage and RA-GRS (backup files)



High availability with the Hyperscale tier

Azure SQL Database offers three service tiers that are designed for different types of applications:

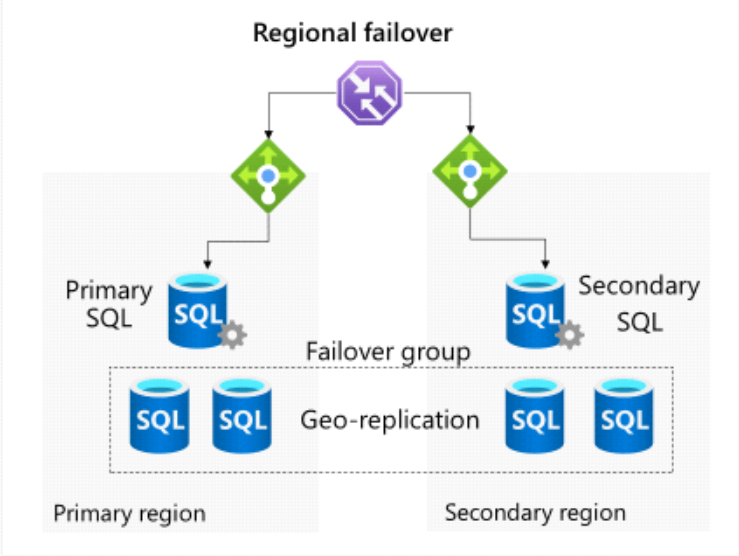
- Designed for very large OLTP databases – as large as 100 TB
- Able to autoscale storage and scale compute
- Captures instantaneous backups (using snapshots)
- Restores in minutes rather than hours and days
- Scale up or down in real time to accommodate workload changes



Select a database failover strategy

Consider datacenter and regional failover.

- In the same region -use AlwaysOn availability zones with failover to secondary replicas
- Across regions – use geo-replication and failover groups



Security

Monday, June 13, 2022 10:25 AM

DATA STATE	ENCRYPTION METHOD
Data-at-rest	Transparent data encryption (TDE), Always Encrypted
Data-in-motion	SSL/TLS, Always Encrypted
Data-in-process	Dynamic data masking

Protect data-at-rest

TDE

- TDE performs encryption and decryption of the data at the page level.
- The data is encrypted as the data is written to the data page on disk and decrypted when the data page is read into memory.
- The end result is that all data pages on disk are encrypted.
- Database backups will also be encrypted because a backup operation just copies the data pages from the database file to the backup device. No decryption is done during the backup operation.
- TDE encrypts the storage of an entire database by using a symmetric key called the Database Encryption Key (DEK).
- Service-managed TDE - where the DEK is protected by a built-in server certificate.
- Customer-managed TDE - the TDE Protector that encrypts the DEK is supplied by customer and stored in a customer-owned and managed in their key management system

Azure's Azure Key Vault ==> RBAC

Protect data-in-transit

SCENARIO	SOLUTION
Secure access from multiple workstations located on-premises to an Azure virtual network	Use site-to-site VPN
Secure access from an individual workstation located on-premises to an Azure virtual network	Use point-to-site VPN
Move large data sets over a dedicated high-speed wide-area network (WAN) link	Use Azure ExpressRoute
Interact with Azure Storage through the Azure portal	All transactions occur via HTTPS. You can also use Storage REST API over HTTPS to interact with Azure Storage and Azure SQL Database.

Protect data-in-use

Dynamic Data Masking

- Data masking policy can be set up in Azure portal only for Azure SQL Database
- Dynamic data masking can be set up using PowerShell cmdlets and REST API
- On Presentation Layer only. Data at storage in fact no masking

Always Encrypted feature for data-at-rest and data-in-transit

- Suggest and encrypt sensitive data in DB
- Real encrypt in storage
- Even DB admin cannot retrieve
- Use key – Bring Your Own Key
- Key can be stored in Windows Certificate Store or in Azure Key Vault

How Always Encrypted works

Step by step process for Always Encrypted is explained below:

- Always Encrypted uses two types of keys: column encryption keys and column master keys.
- A column encryption key is used to encrypt data in an encrypted column. A column master key is a key-protecting key that encrypts one or more column encryption keys.
- The Database Engine only stores encrypted values of column encryption keys and the information about the location of column master keys, which are stored in external trusted key stores, such as Azure Key Vault, Windows Certificate Store
- To access data stored in an encrypted column in plaintext, an application must use an Always Encrypted enabled client driver. Encryption and decryption occurs via the client driver.
- The driver transparently collaborates with the Database Engine to obtain the encrypted value of the column encryption key for the column as well as the location of its corresponding column master key.
- The driver contacts the key store, containing the column master key, in order to decrypt the encrypted column encryption key value, and then it uses the plaintext column encryption key to encrypt the parameter.
- The driver substitutes the plaintext values of the parameters targeting encrypted columns with their encrypted values, and it sends the query to the server for processing.
- The server computes the result set, and for any encrypted columns included in the result set, the driver attaches the encryption metadata for the column, and then the driver decrypts the results and returns plaintext values to the application.

Azure SQL Edge

Monday, June 13, 2022 10:36 AM

Suitable for IoT and IoT Hub

- **Streaming**
- **Time series storage engine to process time-indexed data**

Azure SQL Edge is ideal for

Requirement	SQL Edge capability
Connectivity limitations	Azure SQL Edge supports solutions that work with, or without, network connectivity.
Slow or intermittent broadband connection	Azure SQL Edge provides a powerful, local database. It negates needing to forward all data to a cloud-based database, which eliminates latency.
Data security and privacy concerns	Azure SQL Edge implements RBAC and ABAC, encryption, and data classification. This helps you secure and control access to your IoT apps' data.
Synchronization and connectivity to back-end systems	Azure SQL Edge provides ease of exchanging data with other systems like Azure SQL Database, SQL Server, and Azure Cosmos DB.
Familiarity	Azure SQL Edge shares the same codebase as SQL Server. Developers with skills in SQL Server or SQL Database can reuse their code and skills

2 deployment mode

1) In Azure
2) Containerization : Docker image in Docker Hub

CosmosDB

Monday, June 13, 2022 10:41 AM

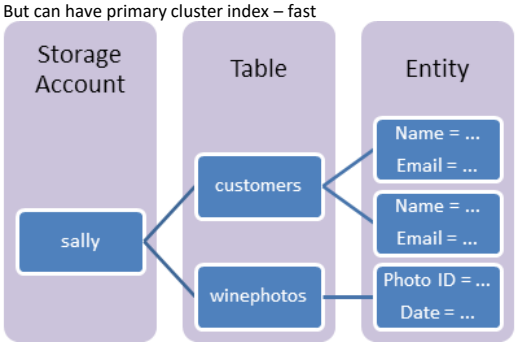
What is CosmosDB?

- NoSQL database
- Aggregate API for
  - Tables
  - Core(SQL for JSON)
  - MongoDB
  - Cassandra
  - Gremlin

Image from Google serach https://www.property.hk/article_content.php?author=PHK_TML&id=57760

Azure Storage Table

- Key value pair
- Cannot have complex joining
- But can have primary cluster index – fast

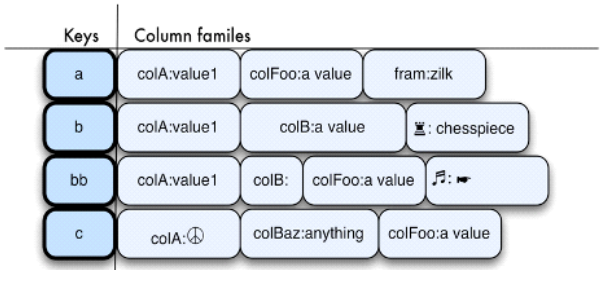


CosmosDB Table API limitation

- Not sorted in order of partition key and row key
- Row key limited to 255 bytes
- Support Cross-Origin Resource Sharing (CORS)
- Table name case-sensitive, while Storage Acct Table is case-insensitive
- Charge on provision created, while Tabla charge when capacity are start using.
- CosmosDB faster, 10ms. While Storage Acct Table may up to 10 seconds

MongoDB, Core (SQL) ==> JSON
Cassandra ==> Wide Columnar No SQL database
Table ==> Key value pair tabular
Gremlin ==> Graph

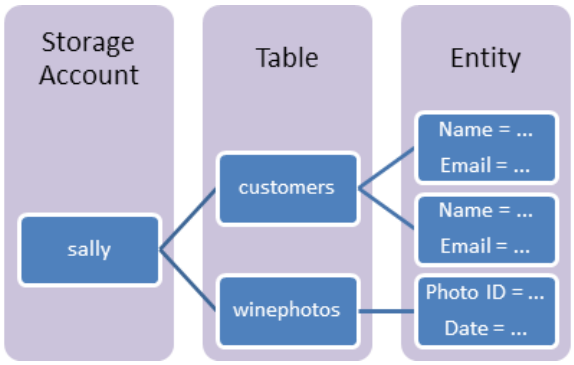
Cassandra



Germlin


bb
colA:value1
colB:
colFoo:a value
c
colA:☮
colBaz:anything
colFoo:a value

Table


Storage Account
Table
Entity
sally
customers
winephotos
Name = ...
Email = ...
Name = ...
Email = ...
Photo ID = ...
Date = ...

Overall

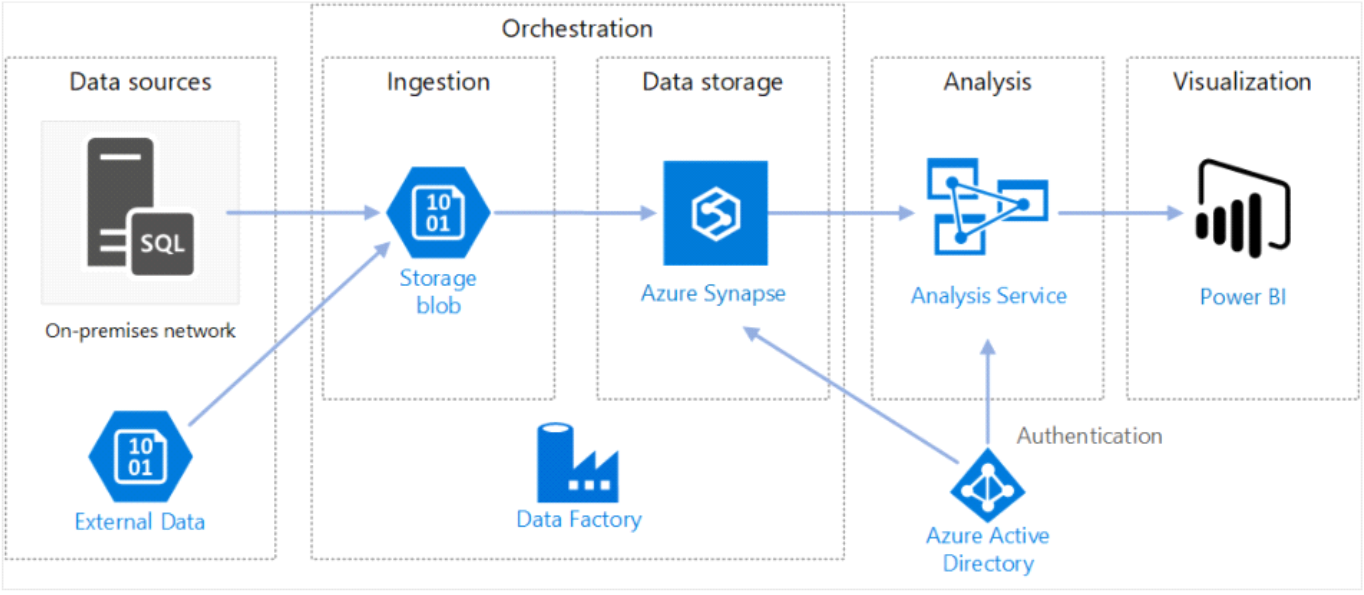
Monday, June 13, 2022 5:10 PM

Data Factory ==> ETL and data integration service (like workflow engine)
Data Lake ==> Repository of data
Databrick ==> BigData & Machine Learning engine. Can query process and analysis, and then feed Azure ML
Synapse Analytics ==> no code ELT to feed BI and ML

Azure Data Factory

Monday, June 13, 2022 5:37 PM

- ELT Tools
- Create and schedule data-driven workflows
- Main functions
  - Orchestrate data movement
  - Transform data at scale



Components of Azure Data Factory

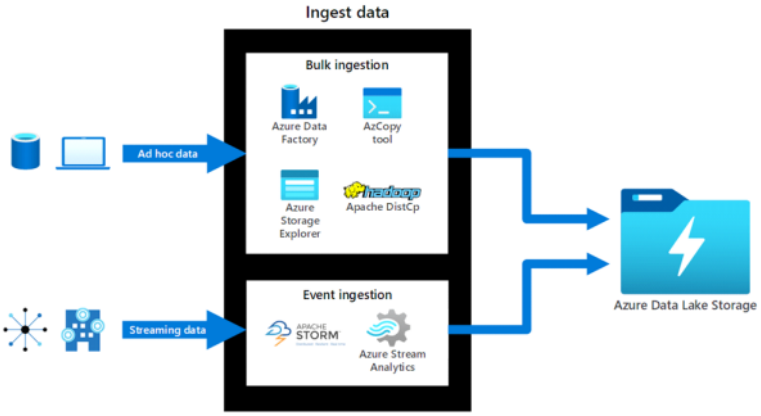
- Linked services
  - Ingest of different data source
- Activities
  - data movement
  - data transformation
  - control activities
- Pipelines
  - Group of activities
- Datasets
  - Source data
- Data Flows
  - develop data transformation logic without writing code
- Integration Runtimes
  - Bridge between the activity and linked Services objects
  - Azure, Self-hosted, and Azure-SSIS

Azure Data Lake Storage

Monday, June 13, 2022 7:27 PM

Use Azure Data Lake when you need

- a data repository on the cloud for managing large volumes of data
- Data types: JSON files, CSV, log files, and other formats in real time
- Real-time data ingestion and storage (e.g. Azure Data Factory)



Ingesting data

- Ad hoc data
  - AzCopy, CLI, PowerShell, Storage Explorer
- Relational data
  - Azure Data Factory sources => Cosmos DB, SQL Database, Managed instances
- For streaming data
  - Apache Storm on Azure HDInsight, Azure Stream Analytics.

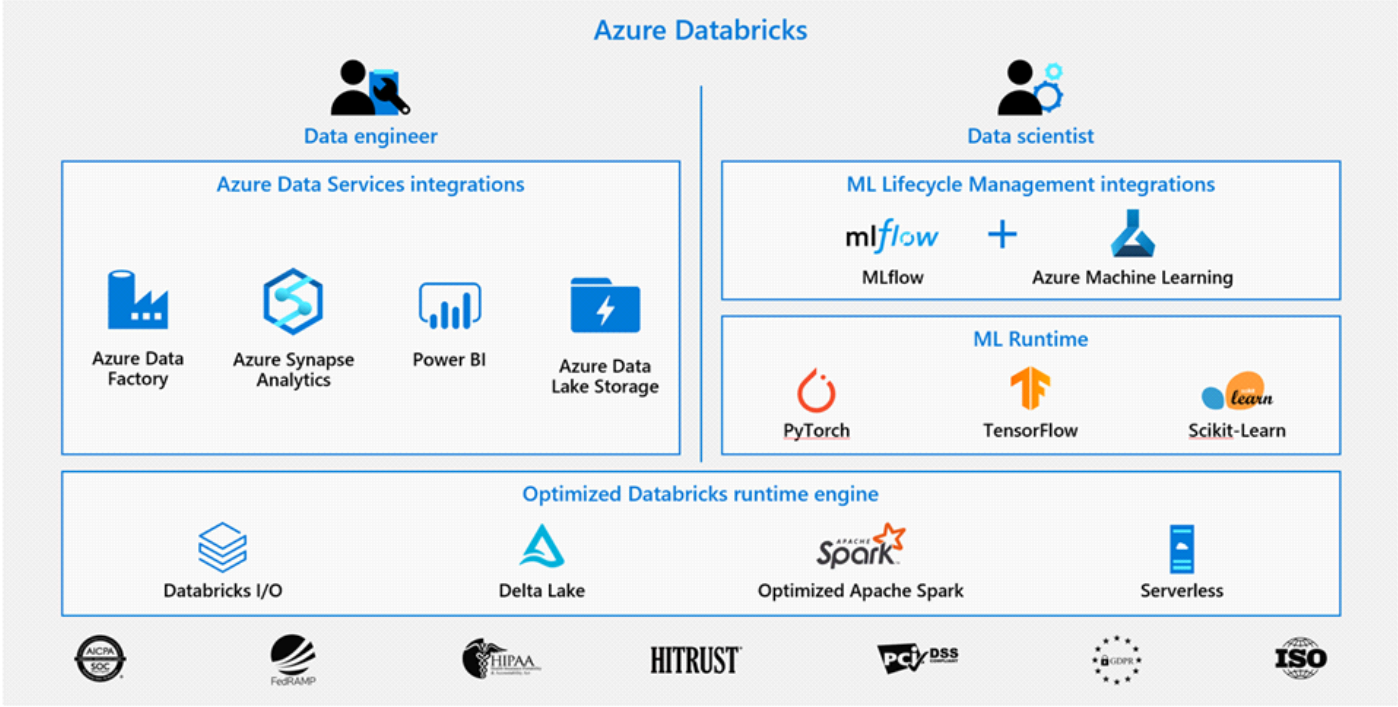
Criteria	Azure Data Lake	Azure Blob Storage
Data type	Good for storing large volumes of text data	Good for storing unstructured non-text based data such as photos, videos, backup etc.
Geographic redundancy	Need to set up replication of data	By default, provides geo redundant storage
Namespaces support	Supports hierarchical namespaces	Supports flat namespaces
Hadoop compatibility	Hadoop services can use data stored in Data Lake	Is not Hadoop compatible
Security	Allows for more granular access	Granular access not supported

Azure Databricks

Monday, June 13, 2022 11:31 PM

Provides data science and engineering teams with a single platform for Big Data processing and Machine Learning. Offers three environments for developing data intensive applications

Environment	Description
Databricks SQL	Provides an easy-to-use platform for analysts who want to run SQL queries on their data lake, create multiple visualization types to explore query results from different perspectives, and build and share dashboards.
Databricks Data Science & Engineering	Provides an interactive workspace that enables collaboration between data engineers, data scientists, and machine learning engineers. For a big data pipeline, the data (raw or structured) is ingested into Azure through Azure Data Factory in batches, or streamed near real-time using Apache Kafka, Event Hub, or IoT Hub. This data lands in a data lake for long term persisted storage, in Azure Blob Storage or Azure Data Lake Storage. As part of your analytics workflow, use Azure Databricks to read data from multiple data sources and turn it into breakthrough insights using ***Spark***.
Databricks Machine Learning	An integrated end-to-end machine learning environment incorporating managed services for experiment tracking, model training, feature development and management, and feature and model serving.



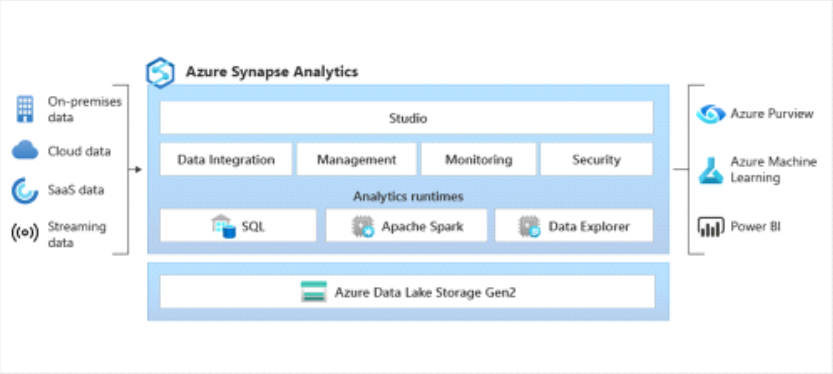
Azure Synapse Analytics

Monday, June 13, 2022 11:41 PM

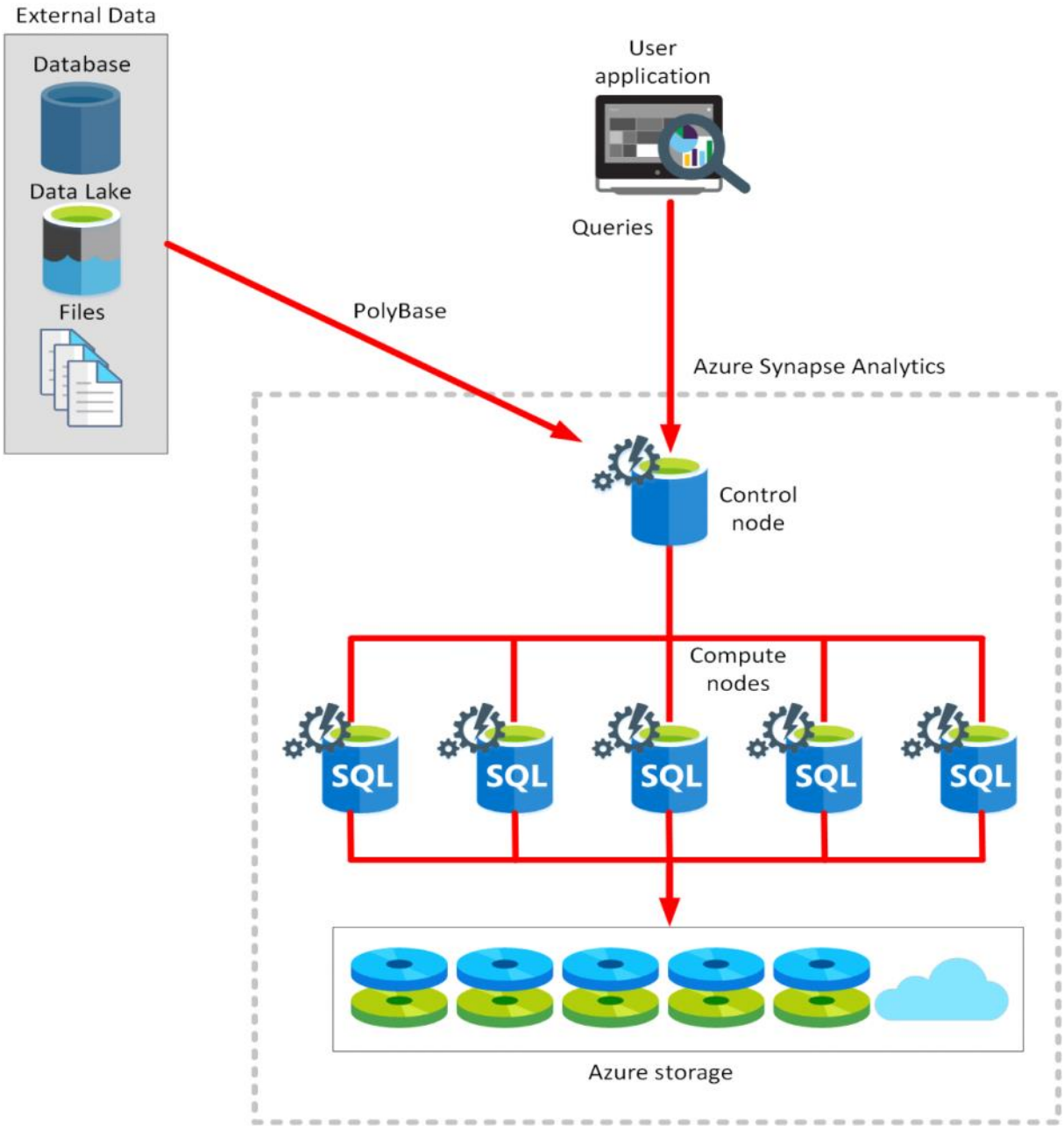
Azure Synapse Analytics

Azure Synapse Analytics is an integrated analytics platform that brings together data integration, enterprise data warehousing, big data analytics and visualization into a single service. Azure Synapse Analytics is an evolution of Azure SQL Data Warehouse.

- Modern data warehousing
- Advanced analytics
- Data exploration and discovery
- Real time analytics
- Data integration
- Integrated analytics
- Machine Learning



- Ingest from different external source
- Enable Parallel Processing
- User submit T-SQL like query statement
- Azure Synapse Analytics process it
  - Distribute by Control node
  - Compute in Compute node
- Use PolyBase to retrieve data from both relational and non-relation storage



Components

- **Synapse SQL pool**: Synapse SQL offers both serverless and dedicated resource models to work with using node-based architecture. For predictable performance and cost, you can create dedicated SQL pools, for unplanned or ad hoc workloads, you can use the always-available, serverless SQL endpoint.

- **Synapse Spark pool**: This is a cluster of servers running Apache Spark to process data. You write your data processing logic using one of the four supported languages: Python, Scala, SQL, and C# (via .NET for Apache Spark). Apache Spark for Azure Synapse integrates Apache Spark-the open source big data engine used for data preparation, data engineering, ETL, and machine learning.

- **Synapse Pipelines**: Azure Synapse Pipelines leverages the capabilities of Azure Data Factory and is the cloud-based ETL and data integration service that allows you to create data-driven workflows for orchestrating data movement and transforming data at scale. You could include activities that transform the data as it is transferred, or you might combine data from multiple sources together.

- **Synapse Link**: This component allows you to connect to Cosmos DB. You can use it to perform near real-time analytics over the operational data stored in a Cosmos DB

database.

- **Synapse Studio**: This is a web-based IDE that can be used centrally to work with all capabilities of Azure Synapse Analytics. You can use Synapse Studio to create SQL and Spark pools, define and run pipelines, and configure links to external data sources.

Compare Azure Data Factory to Azure Synapse Analytics

Criteria	Azure Data Factory	Azure Synapse Analytics
Integration runtime sharing	Can be shared across different data factories	No sharing
Solution templates	Provided with Azure Data Factory template gallery	Provided with Synapse Workspace Knowledge center
Integration Runtime cross region support	Support Cross region data flows	Does not support cross region data flows
Monitoring of Spark Jobs for Data Flow	Not supported	Supported by the Synapse Spark pools

Hot, warm, cold data path

Monday, June 13, 2022 11:57 PM

When to use Hot/Warm/Cold data path

Path	Requirement
Hot data path	• When data requirements are known to change frequently • When processing or displaying data in real time
Warm data path	• When you need to store or display a recent subset of data • Used for data that is consumed for small analytical and batch processing
Cold data path	• When data is rarely used. The data might be stored for compliance or legal reasons • Used for data that is consumed for long term analytics and batch processing

Path	Suitable storage on Azure	Processing
Warm	Azure SQL CosmosDB	Stream Analytics
Cold	Azure Blobs (objects) Azure Data Lake Storage Gen2 Azure Files Azure Queues Azure Tables	Azure Data Factory generate and put to Azure Data Lake or direct ingest by Databrick

Azure Stream Analytics

Tuesday, June 14, 2022 12:15 AM

- fully managed (PaaS offering)
- real-time analytics
- complex event-processing engine
- real-time analytics on multiple streams of data
  - IoT
  - Sensor
  - Clickstreams
  - Social media feeds
- Ingest source
  - Azure Event Hubs
  - Azure IoT Hub
  - Azure Blob Storage
- Analyze by
  - SQL like query to filter/sort/aggregate
  - Extends by JS & C#
- Deliver to
  - Downstream by Azure Event Hubs/ Service Bus/Functions
  - Visualize in Power BI in real-time
  - Train ML by placing output to Azure Synapse Analytics
  - Store ==> SQL/ Cosmos/Blob.....