1. SQL database support 2 purchasing models:
    - DTUs: Data Transaction Units
    - vCore: let you select CPU & storage separately from each other
2. OLAP: database workload that is the type of deep analytics used by SQL Server Analysis Services (SSAS), Azure Analysis Services,… There is pre-processing operations that turns the data into cubes, with dimensions and measures
3. **Data warehouse**: best for data that is historical and not used in your day-to-day business, going back to 10-15 years → is a massive data storage that includes all the historical data, database optimized for reading by being denormalized. OLTP: day-to-day operations, but leave historical data in the data warehouse
4. 3 charateristics of Azure SQL Database being used as an OLTP system: heavy write & moderate read, schema on write, normalized data. Azure SQL Database: includes a fully managed backup service, allows independent provisioning of throughput & storage, has built-in availability. OLAP: heavy read
5. Elastics pool databases, you pay per pool, not database
6. SQL Database, SQL Managed Instance, SQL Server in VM use SQL Server engine
7. SQL Server runs on T-SQL which stands for Transact SQL
8. SQL Database networking is set with 'no access' or with 'public endpoints' without any other actions (only with configuring its firewall to allow anyone) → no users can connect the database
9. Cheapest option to store data in Azure: Blob storage
10. Azure database for MySQL is chosen because applications that already use MySQL can be migrated without modification
11. Tumbling window can run a job using data for a specific period of time, and not before or after that
12. Power BI Report Builder: build paginated reports
13. Scaling Azure SQL Database does not affect any applications or users using it at the time
14. Azure Data Studio is downloadable software, open source and supported on Windows, Mac OS and Linux. It provides a graphical user interface for managing many different database systems without logging in each and every other system. **Can be restore Database**. Run on apache Spark
15. Benefits of normalization (is the process of breaking a data field down to its composite parts): reduces data duplication, reduce normal typos and errors on inputs, SELECT queries is faster

16. Primary purpose of an index on a relational data table: speed up SELECT queries so that they return faster

17. ARM template is stores in JSON

18. SQL Server in VM: infrastructure as a service. You can choose the operating system, and specific software versions for SQL Server. SQL Server in VM has guaranteed 100% compatibility with SQL Server running in your environment while Azure SQL Database Elastic Pool has almost 100% compatibility with SQL Server running in your environment. It requires the fewest changes when migrating an existing SQL server on-premises solution. You are responsible for all software installation and maintenance and performing back-ups

19. SQL Managed Instance, Azure SQL Database, Cosmos DB: PaaS

20. DDL - Data Definition Language: used to define the database scheme, including CREATE, ALTER, DROP, RENAME

21. DML: Data Manipulation Language includes SELECT, INSERT, UPDATE, DELETE

22. NoSQL database has no fixed schema → you can dynamically add new properties without worrying about that. Previous documents will not be updated

23. Data ingestion: capturing raw data streaming from various sources and storing it

24. Data processing is conversion of raw data to meaningful information through a process. Processing data as its arrive is called streaming, buffering and processing the data in groups is called batch processing

25. Index: a structure that enables you to locate rows in a tbl quickly, using an indexed value

26. View: a virtual table based on the result set of query

27. Table: a structure comprising rows and columns that use for storing data

28. Benefit of using PaaS instead of on-premise system: increase scalability

29. A key-value store: can ingest large volumes of data rapidly

30. Azure SQL Database: scaling up or out will take effect without restarting the SQL Database

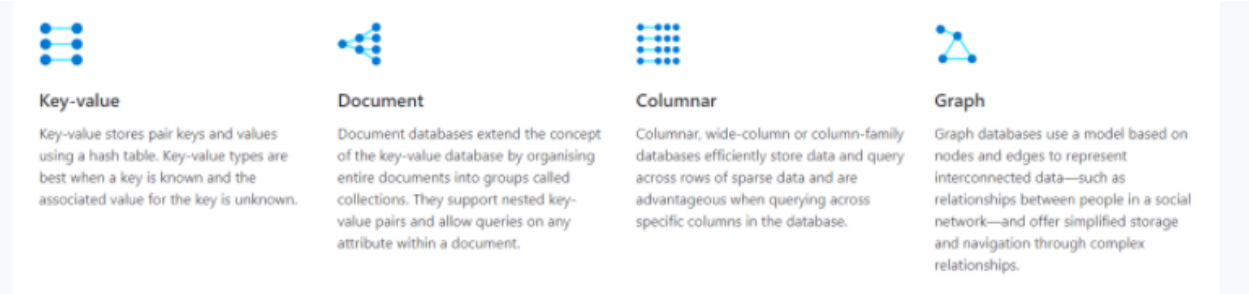
31. Azure Database managed Instance: backups are automatically handled and the ability to restore to a point in time. Use it when you want to lift & shift an on-premises SQL Server Instance and all its database to the cloud, without incurring the management overhead of running SQL Server on a virtual machine

32. Elements of Azure Table storage key: partition key and row key

33. Page block for blobs that require random read and write access. Block blob for discrete objects that change infrequently

34. Azure File storage: enable users at different site to share the files
35. Security principal: an object that represents a user, group, service or managed identity that is requesting access to Azure resource.
36. A name of collection of permissions that can be granted to a service, such as the ability to use the service to read, write, and delete data. Ex: Owner, contributors → Azure Built in Roles
37. Building block in Power BI: visualizations, dataset, reports, dashboard, tiles
38. Provisioning: the act of running series of tasks that a service provider performs to create and configure a service
39. Synapse SQL pool: collection of servers running Transact SQL which is used by Azure SQL Database and Microsoft SQL Server. Synapse Spark pool: cluster of servers running Apache Spark to process data. Spark supports Machine Learning through integration. Azure Synapse Analytics (example of OLAP): to perform very complex queries and aggregations
40. A relational database must be used when strong consistency is required
41. Streaming processing is not preferred for stream processing
42. Polybase is used to query data from external data sources from Azure Synapse Analytics
43. Pipeline which is a component of Azure Data Factory can be triggered to run data ingestion tasks & can pass parameters to a notebook.
44. The Integration Runtime (IR) is the compute infrastructure used by Azure Data Factory to provide the following data integration capabilities across different network environment
45. In Azure Data Factory, you can use a Control Flow to orchestrate the pipeline
46. Azure Data Studio can be used to restore the database / performs backups
47. Azure Databricks is based on apache spark. This is analytics solution & does not act as a storage solution for metric & log
48. RBAC – role based access control is supported by Azure Data Lake Storage at the file and folder level. Besides, it Azure Data Lake Storage also supports: Shared Key authorization, SAS, Access Control List (ACL)
49. Using the existing Microsoft SQL Server licenses to reduce the cost of Azure SQL database
50. Azure Active Directory (Azure AD) authentication allows users to use multi-factor authentication when connect to an Azure SQL Database
51. Event hub, IoT hub, Azure Data Factory → ingestion
52. Steam Analytics Job, Synapse Analytics, Databricks → Processing

53. All Azure Cosmos DB operations must complete within the specified timeout duration
54. Resource Group → Azure Storage Account → Container → Blob
55. Azure resource Manager Template: deploys multiple HDInsight clusters
56. Azure Databricks, Azure Event hub (Ex: twitter) → support the analysis of log files in real time
57.



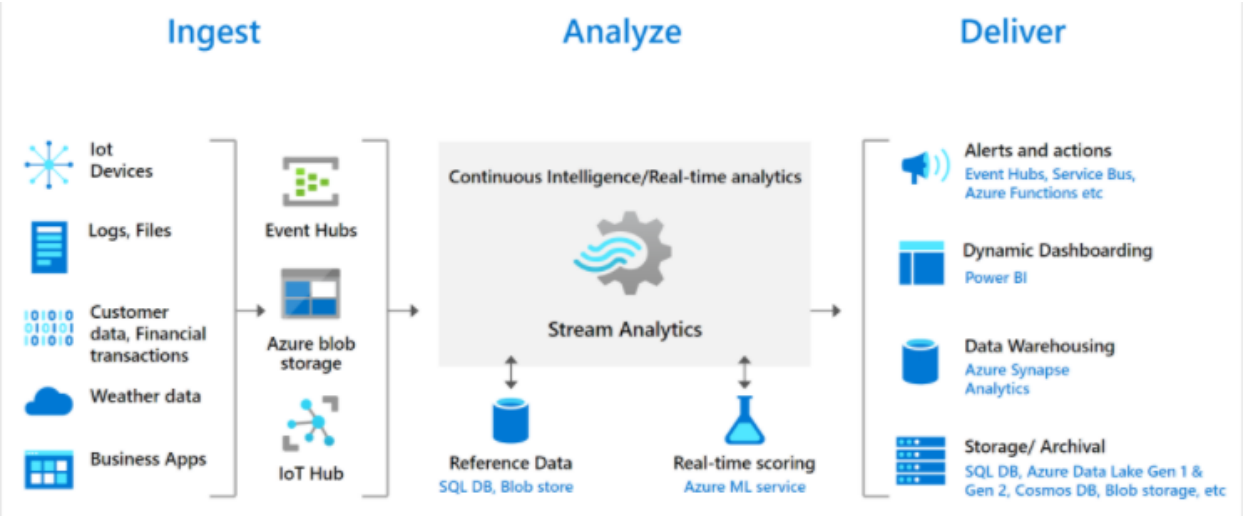
58. A massive parallel processing (MPP) of **Azure synapse analytics distributes processing across compute nodes**
59. Clustering index: an object associated with a table that sorts and stores the data rows in a table based on their keys
60. **At database & container**, we can set the throughput for an Azure Cosmos DB account
61.

Database Administrator tasks and responsibilities

Some of the most common roles and responsibilities of a database administrator include:

- Installing and upgrading the database server and application tools.
- Allocating system storage and planning storage requirements for the database system.
- Modifying the database structure, as necessary, from information given by application developers.
- Enrolling users and maintaining system security.
- Ensuring compliance with database vendor license agreement.
- Controlling and monitoring user access to the database.
- Monitoring and optimizing the performance of the database.
- Planning for backup and recovery of database information.
- Maintaining archived data.
- Backing up and restoring databases.
- Contacting database vendor for technical support.
- Generating various reports by querying from database as per need.
- Managing and monitoring data replication.

62. Hive – Azure HDInsight is a managed, full spectrum, open source analytics service in the cloud for enterprise.
63. Three values is needed to access the server: Server's FQDN, server's Username & Password
64. In order to remain consistency while processing transaction, ACID properties must be adhered: Atomicity, Consistency, Isolation, Durability
65. Azure storage offers different access tiers: Hot – frequently, Cold – at least 30 days, Archive – at least 180 days
66. ETL process requires data that is fully processed before being loaded into the target datasource
67. ETL process have very high load times



68. You are writing a set of SQL queries that administrators will use to troubleshoot an Azure SQL database. You need to embed documents and query results into a SQL notebook → Azure Data Studio

69. Transparent Data Encryption (TDE) encrypts the database to protect the data at rest

70. Cosmos DB Table API supports multiple read replicas, multiple write regions while Azure Table Storage supports only multiple read replicas

71. Azure Cosmos DB uses the ARS – Atom Record Sequence to store the data

72. Azure Storage Explorer is used to access data stored in Azure Storage Account

73. Using Cassandra API to support a column-based database type that use containers to store items: rows (container), table (item)

74. Linked service provides the information needed for Data Factory to connect to a source or destination

75. Azure SQL Database serverless offers support automatic database scaling & pausing of the database during inactive period

76. **Synapse SQL pool**: This is a collection of servers running Transact-SQL. Transact-SQL is the dialect of SQL used by Azure SQL Database, and Microsoft SQL Server. You write your data processing logic using Transact-SQL.

77. **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). **Spark pools support Azure Machine Learning through integration** with the SparkML and AzureML packages. https://docs.microsoft.com/en-us/learn/modules/explore-data-storage-processing-azure/3-explore-azure-synapse-analytics

78. **Azure Synapse Link** - Synapse Link brings together **Azure Cosmos DB analytical store with Azure Synapse analytics** runtime support. This integration enables you to build cloud native HTAP (Hybrid transactional/analytical processing) solutions that generate insights based on real-time updates to your operational data over large datasets. It unlocks new business scenarios to raise alerts based on live trends, build near real-time dashboards, and business experiences based on user behavior. This is incorrect as our data source is data lake and not cosmos db. https://docs.microsoft.com/en-us/azure/cosmos-db/synapse-link-use-cases

79. SMB 3.0 protocol: Azure File Storage

80. Event hub - **Ingestion.** Stream Analytics Job - **Processing.** Synapse Analytics - **Processing.** Azure blob storage - **Storage.** IoT hub - Ingestion

81. Normalization is the process of organizing data in a database

82. You are designing a data storage solution for a database that is expected to grow to 50 TB. The usage pattern is singleton inserts, singleton updates, and reporting. -> **Azure SQL Database HyperScale**

83. Azure SQL Database is protected by Server Level Firewall

84. At which two levels can you set the throughput for an Azure Cosmos DB account? Each correct answer presents a complete solution -> **Database, Container**

85. You can query graph database in Cosmos DB using -> as node and edge using Gremlin

86. What are two uses of data visualization? Each correct answer present a complete solution. -> represent trends and parttern overtime, communication the signfinication of the data

87. **The Integration Runtime (IR) is the compute infrastructure** used by Azure Data Factory to provide the following data integration capabilities across **different network environments**

88. **FQDN** stand for fully qualified Domain **Name**

89. ACID Properties: A transaction is a very small unit of a program and it may contain several lowlevel tasks. A transaction in a database system must maintain **A**tomicity, **C**onsistency, **I**solation, and **D**urability − commonly known as ACID properties − in order to ensure accuracy, completeness, and data integrity. **Atomicity** − This property states that a transaction must be treated as an atomic unit, that is, either all of its operations are executed or none. There must be no state in a database where a transaction is left partially completed. States should be defined either before the execution of the transaction or after the execution/abortion/failure of the transaction. **Consistency** − The database must remain in a consistent state after any transaction. No transaction should have any adverse effect on the data residing in the database. If the database was in a consistent state before the execution of a transaction, it must remain consistent after the execution of the transaction as well. **Durability** − The database should be durable enough to hold all its latest updates even if the system fails or restarts. If a transaction updates a chunk of data in a database and commits, then the database will hold the modified data. If a transaction commits but the system fails before the data could be written on to the disk, then that data will be updated once the system springs back into action. **Isolation** − In a database system where more than one transaction are being executed simultaneously and in parallel, the property of isolation states that all the transactions will be carried out and executed as if it is the only transaction in the system. No transaction will affect the existence of any other transaction.

90. Application user and their default language -> key/value

91. Which service in Azure can be used to process the data in real-time having three components: input, query, and output? -> Stream Analytic Job

92. **Control flow is an orchestration** of pipeline activities that includes chaining activities in a sequence, branching, defining parameters at the pipeline level, and passing arguments while

invoking the pipeline on-demand or from a trigger. It also includes custom-state passing and looping containers

93. Relational databases use **KEY** to enforce relationship among tables.
94. How much data can be stored in a single Table Storage account -> **500 TB**
95. Which two settings can you configure at the container level? Each correct answer presents a complete solution -> **The Throught, The parition key**
96. Azure SQL Database includes a fully managed backup service / build in high-availability /Azure thread Protection (ATP)
97. Auzre data lake store: When ingesting data from Azure Data Lake Store across Azure regions, you will incur costs for bandwidth / use blob, table, storage in Azure Storage Account / implement Azure data Lake Storage an Azure Storage Account
98. What is benefit of Azure Cosmos DB Table API as compared to Azure Table Storage -> **supports multi-master model**
99. **Azure Files** offers native cloud file sharing service based on SMB protocol.

100. **Data movement** - Copy / **Data Transformation** - Mapping data flow / **Control** - Until
101.

Explanation

A container is specialized into API-specific entities as shown in the following table:

Azure Cosmos entity	SQL API	Cassandra API	Azure Cosmos DB API for MongoDB	Gremlin API	Table API
Azure Cosmos container	Container	Table	Collection	Graph	Table

Cosmos entity	SQL API	Cassandra API	Azure Cosmos DB API for MongoDB	Gremlin API	Table API
Azure Cosmos item	Item	Row	Document	Node or edge	Item

Refer below link for nomenclature for different APIs