# Azure Data Fundamentals

[Microsoft Certified: Azure Data Fundamentals - Learn | Microsoft Docs](https://docs.microsoft.com/en-us/learn/certifications/azure-data-fundamentals/?azure-portal=true)

## 1 Explorer core data concepts

Graphical user interface, application, website

Description automatically generated

Data structures in which this data is organized often represents entities that are important to an organization

## Structured data

Structured data is data that adheres to a fixed schema, so all of the data has the same fields or properties.

Structured data is often stored in a database in which multiple tables can reference one another by using key values in a relational model.

Table

Description automatically generated

## Semi-structured data

Semi-structured data is information that has some structure, but which allows for some variation between entity instances. For example, while most customers may have an email address, some might have multiple email addresses, and some might have none at all.

One common format for semi-structured data is JavaScript Object Notation (JSON).

Text

Description automatically generated with medium confidence

## Unstructured data

Not all data is structured or even semi-structured. For example, documents, images, audio and video data, and binary files might not have a specific structure. This kind of data is referred to as unstructured data.

A picture containing text, mammal, hippo

Description automatically generated

## Data stores

File stores

Databases

## File storage

Delimited txt, csv,JSON, XML, BLOB Binary Large Object (images, video, audio)

## Databases

Relational databases are commonly used to store and query structured data, SQL.

Table

Description automatically generated

Non relational databases are data management systems that dont apply a relational schema to the data., NoSQL.

Table

Description automatically generated

## Transactional data processing OLTP

A transactional data processing system is what most people consider the primary function of business computing. A transactional system records transactions that encapsulate specific events that the organization wants to track. A transaction could be financial, such as the movement of money between accounts in a banking system, or it might be part of a retail system, tracking payments for goods and services from customers.

OLTP

The work performed by transactional systems is often referred to as Online Transactional Processing (OLTP).

OLTP solutions rely on a database system in which data storage is optimized for both read and write operations in order to support transactional workloads in which data records are created, retrieved, updated, and deleted (often referred to as CRUD operations). These operations are applied transactionally, in a way that ensures the integrity of the data stored in the database. To accomplish this, OLTP systems enforce transactions that support so-called ACID semantics:

Atomicity – each transaction is treated as a single unit, which succeeds completely or fails completely. For example, a transaction that involved debiting funds from one account and crediting the same amount to another account must complete both actions. If either action can't be completed, then the other action must fail.

Consistency – transactions can only take the data in the database from one valid state to another. To continue the debit and credit example above, the completed state of the transaction must reflect the transfer of funds from one account to the other.

Isolation – concurrent transactions cannot interfere with one another, and must result in a consistent database state.

Durability – when a transaction has been committed, it will remain committed. After the account transfer transaction has completed, the revised account balances are persisted so that even if the database system were to be switched off, the committed transaction would be reflected when it is switched on again.

## 2 Explorer data roles and services

### #Identify data services

Some of the most commonly used cloud services for data are described below.

[Identify data services - Learn | Microsoft Docs](https://docs.microsoft.com/en-us/learn/modules/explore-roles-responsibilities-world-of-data/3-data-services)

Text

Description automatically generated

A picture containing text

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Text, application

Description automatically generated

Text

Description automatically generated

A picture containing text

Description automatically generated

Text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated with medium confidence

Graphical user interface, text

Description automatically generated

## 3 Explorer relational in Azure

Graphical user interface, application, website

Description automatically generated

## Understand relational data

In a relational database, you model collections of entities from the real world as tables.

## Understand normalization

Normalization is a term used by database professionals for a schema design process that minimizes data duplication and enforces data integrity.

While there are many complex rules that define the process of refactoring data into various levels (or forms) of normalization, a simple definition for practical purposes is:

* Separate each entity into its own table.
* Separate each discrete attribute into its own column.
* Uniquely identify each entity instance (row) using a primary key.
* Use foreign key columns to link related entities.

## Explorer SQL

SELECT, INSERT, UPDATE, DELETE, CREATE, and DROP to accomplish almost everything that you need to do with a database.

* Data Definition Language (DDL), CREATE, ALTER, DROP, RENAME
* Data Control Language (DCL), GRANT, DENY, REVOKE
* Data Manipulation Language (DML), SELECT, INSERT, UPDATE, DELETE
* VIEW
* Stored procedure
* Index

## 4 Explorer relational database services in Azure

Text

Description automatically generated with medium confidence

### #Describe Azure SQL services and capabilities

[Describe Azure SQL services and capabilities - Learn | Microsoft Docs](https://docs.microsoft.com/en-us/learn/modules/explore-provision-deploy-relational-database-offerings-azure/2-azure-sql)

Azure SQL is a collective term for a family of Microsoft SQL

SQL Server on Azure Virtual Machines (VMs) - A virtual machine running in Azure with an installation of SQL Server. The use of a VM makes this option an infrastructure-as-a-service (IaaS) solution that virtualizes hardware infrastructure for compute, storage, and networking in Azure; making it a great option for "lift and shift" migration of existing on-premises SQL Server installations to the cloud.

Azure SQL Managed Instance - A platform-as-a-service (PaaS) option that provides near-100% compatibility with on-premises SQL Server instances while abstracting the underlying hardware and operating system. The service includes automated software update management, backups, and other maintenance tasks, reducing the administrative burden of supporting a database server instance

Azure SQL Database - A fully managed, highly scalable PaaS database service that is designed for the cloud. This service includes the core database-level capabilities of on-premises SQL Server, and is a good option when you need to create a new application in the cloud.

Azure SQL Edge - A SQL engine that is optimized for Internet-of-things (IoT) scenarios that need to work with streaming time-series data.

Graphical user interface, text, application, email

Description automatically generated

### #Describe Azure services for open-source databases

[Describe Azure services for open-source databases - Learn | Microsoft Docs](https://docs.microsoft.com/en-us/learn/modules/explore-provision-deploy-relational-database-offerings-azure/3-azure-database-open-source)

In addition to Azure SQL services, Azure data services are available for other popular relational database systems, including MySQL, MariaDB, and PostgreSQL.

Text

Description automatically generated

Text

Description automatically generated

Text

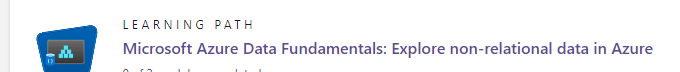
Description automatically generated

### #Azure GUI

Text, application

Description automatically generated

## 5 Explorer non-relational data



## Blob storage

Blob Storage is a service that enables you to store massive amounts of unstructured data as binary large objects, or blobs, in the cloud. Blobs are an efficient way to store data files in a format that is optimized for cloud-based storage, and applications can read and write them by using the Azure blob storage API.

Graphical user interface, text, application

Description automatically generated

Block blobs, Block blobs are best used to store discrete, large, binary objects that change infrequently.

Page blobs, A page blob is optimized to support random read and write operations; you can fetch and store data for a single page if necessary. Azure uses page blobs to implement virtual disk storage for virtual machines.

Append blobs, An append blob is a block blob optimized to support append operations. You can only add blocks to the end of an append blob; updating or deleting existing blocks isn't supported.

## DataLake Storage Gen2

Azure Data Lake Store (Gen1) is a separate service for hierarchical data storage for analytical data lakes, often used by so-called big data analytical solutions that work with structured, semi-structured, and unstructured data stored in files.

Diagram

Description automatically generated

## Files

Diagram

Description automatically generated

## Tables

Azure Table Storage is a NoSQL storage solution that makes use of tables containing key/value data items. Each item is represented by a row that contains columns for the data fields that need to be stored.

Diagram, table

Description automatically generated

An Azure Table enables you to store semi-structured data. All rows in a table must have a unique key (composed of a partition key and a row key), and when you modify data in a table, a timestamp column records the date and time the modification was made; but other than that, the columns in each row can vary.

### Partions

An Azure Table enables you to store semi-structured data. All rows in a table must have a unique key (composed of a partition key and a row key), and when you modify data in a table, a timestamp column records the date and time the modification was made; but other than that, the columns in each row can vary.

## 6 Explorer fundamentals of Cosmos DB

Graphical user interface, text

Description automatically generated

Azure Cosmos DB is a highly scalable cloud database service for NoSQL data.

Diagram

Description automatically generated

Azure Cosmos DB supports multiple application programming interfaces (APIs) that enable developers to use the programming semantics of many common kinds of data store to work with data in a Cosmos DB database. The internal data structure is abstracted, enabling developers to use Cosmos DB to store and query data using APIs with which they're already familiar.

### #When to use Cosmos DB

[Describe Azure Cosmos DB - Learn | Microsoft Docs](https://docs.microsoft.com/en-us/learn/modules/explore-non-relational-data-stores-azure/2-describe-azure-cosmos-db)

* IoT and telematics, These systems typically ingest large amounts of data in frequent bursts of activity.
* Retail and marketing
* Gaming
* Web and mobile applications

### #Cosmos DB API

[Identify Azure Cosmos DB APIs - Learn | Microsoft Docs](https://docs.microsoft.com/en-us/learn/modules/explore-non-relational-data-stores-azure/3-cosmos-db-apis)

Azure Cosmos DB supports multiple APIs, enabling developers to easily migrate data from commonly used NoSQL stores and apply their existing programming skills. When you provision a new Cosmos DB instance, you select the API that you want to use.

* Core SQL API (JSON)
* MongoDB API (MQL)
* Table API (Similar to Azure Table)

Graphical user interface, text, application, email

Description automatically generated

<https://endpoint/Customers(PartitionKey='1',RowKey='124')>

* Cassandra API (SQL like)
* Gremlin API (Graph structure)

## 7 Explorer data analytics

Graphical user interface

Description automatically generated

## Explorer modern data warehousing

* 1 Data ingestion and processing, ETL or ELT. Data ingestion includes both batch processing of static data and real-time processing of streaming data.
* 2 Analytical data store, relational data warehouse, file-system based data lakes, and hybrid.
* 3 Analytical data model, 1 or more data models that pre-aggregate the data to make it easier to produce reports, dashboards and interactive solutions.
* 4 Data visualization.

Application

Description automatically generated with medium confidence

## #Explorer data ingestion pipelines

[Explore data ingestion pipelines - Learn | Microsoft Docs](https://docs.microsoft.com/en-us/learn/modules/examine-components-of-modern-data-warehouse/3-data-ingestion-pipelines)

Large scale:

* Pipelines and ETL
* Pipelines with Azure data factory or use same engine in Azure synapse analytics if you want to manage all of the components of your d w solution in a unified workspace.
* In either case, pipelines consist of one or more activities that operate on data. An input dataset provides the source data, and activities can be defined as a data flow that incrementally manipulates the data until an output dataset is produced. Pipelines use linked services to load and process data – enabling you to use the right technology for each step of the workflow.

Timeline

Description automatically generated with low confidence

## #Explorer analytical data store

[Explore analytical data stores - Learn | Microsoft Docs](https://docs.microsoft.com/en-us/learn/modules/examine-components-of-modern-data-warehouse/4-analytical-data-stores)

There are two common types of analytical data store.

Data warehouse

A data warehouse is a relational database in which the data is stored in a schema that is optimized for data analytics rather than transactional workloads. Commonly, the data from a transactional store is denormalized into a schema in which numeric values are stored in central fact tables,

Diagram

Description automatically generated

Data lake

A data lake is a file store, usually on a distributed file system for high performance data access. Technologies like Spark or Hadoop are often used to process queries on the stored files and return data for reporting and analytics. These systems often apply a schema-on-read approach to define tabular schemas on semi-structured data files at the point where the data is read for analysis, without applying constraints when it's stored.

A picture containing text

Description automatically generated

Hybrid

## Azure services for analytical stores

Three main

Graphical user interface, text, application, Word

Description automatically generated

## #Tutorial Explore Azure Synapse Analytics Pricy

[Exercise: Explore Azure Synapse Analytics - Learn | Microsoft Docs](https://docs.microsoft.com/en-us/learn/modules/examine-components-of-modern-data-warehouse/5-exercise-azure-synapse)

Graphical user interface, application

Description automatically generated

Text

Description automatically generated

### Pricy

Graphical user interface, text

Description automatically generated

## 8 Explorer fundamental of real-time analytics

Graphical user interface

Description automatically generated with medium confidence

## #Understand batch and stream processing

[Understand batch and stream processing - Learn | Microsoft Docs](https://docs.microsoft.com/nb-no/learn/modules/explore-fundamentals-stream-processing/2-batch-stream)

Batch processing, in which multiple data records are collected and stored before being processed together in a single operation.

* New data collected and stored, and whole group is processed togehter as a batch
* Based on schedule, 1 h
* Trigger when data has arrived
* Parking lot, collect cars in lot and then count them in a single operation.

Stream processing, in which a source of data is constantly monitored and processed in real time as new data events occur.

* Processed when it arrives
* No wating, processed as individual units
* Parking lot, counting cars in realtime as they pass.

Understand differences between batch and streaming data

## #Explore common elements of stream processing architecture

[Explore common elements of stream processing architecture - Learn | Microsoft Docs](https://docs.microsoft.com/nb-no/learn/modules/explore-fundamentals-stream-processing/3-explore-common-elements)

General architecture

1. An event generates some data
2. The generated data is captured in a streaming source for processing (folder or database or queue)
3. The event data is processed (query time-based or windows)
4. The result of streaming process operation are written to an output or sink (file, database, dashboard or another queue)

## #Realtime analytics in Azure

[Explore common elements of stream processing architecture - Learn | Microsoft Docs](https://docs.microsoft.com/nb-no/learn/modules/explore-fundamentals-stream-processing/3-explore-common-elements)

Text

Description automatically generated

## Sources for stream processing

Text

Description automatically generated

## Sinks for stream processing

Graphical user interface, text

Description automatically generated

## #Explore Azure Stream Analytics SQL syntax

[Explore Azure Stream Analytics - Learn | Microsoft Docs](https://docs.microsoft.com/nb-no/learn/modules/explore-fundamentals-stream-processing/4-stream-analytics)

Azure Stream Analytics is a service for complex event processing and analysis of streaming data. Stream Analytics is used to:

* Ingest data from an input, such as an Azure event hub, Azure IoT Hub, or Azure Storage blob container.
* Process the data by using a query to select, project, and aggregate data values.
* Write the results to an output, such as Azure Data Lake Gen 2, Azure SQL Database, Azure Synapse Analytics, Azure Functions, Azure event hub, Microsoft Power BI, or others.

Once started, a Stream Analytics query will run perpetually, processing new data as it arrives in the input and storing results in the output.

Azure Stream Analytics is a great technology choice when you need to continually capture data from a streaming source, filter or aggregate it, and send the results to a data store or downstream process for analysis and reporting.

## #Explore Apache Spark on Microsoft Azure

[Explore Apache Spark on Microsoft Azure - Learn | Microsoft Docs](https://docs.microsoft.com/nb-no/learn/modules/explore-fundamentals-stream-processing/6-spark-streaming)

Apache Spark is a distributed processing framework for large scale data analytics.

## #Explore Azure Data Explorer KQL, Kusto syntax

[Explore Azure Data Explorer - Learn | Microsoft Docs](https://docs.microsoft.com/nb-no/learn/modules/explore-fundamentals-stream-processing/8-data-explorer)

Azure Data Explorer is a standalone Azure service for efficiently analyzing data. You can use the service as the output for analyzing large volumes of diverse data from data sources such as websites, applications, IoT devices, and more. For example, by outputting Azure Stream Analytics logs to Azure Data Explorer, you can complement Stream Analytics low latency alerts handling with Data Explorer's deep investigation capabilities.

## 9 Explore fundamentals of data visualization

[Explore fundamentals of data visualization - Learn | Microsoft Docs](https://docs.microsoft.com/nb-no/learn/modules/explore-fundamentals-data-visualization/)

Text

Description automatically generated

## #Describe Power BI tools and workflow

[Describe Power BI tools and workflow - Learn | Microsoft Docs](https://docs.microsoft.com/nb-no/learn/modules/explore-fundamentals-data-visualization/2-power-bi)

A typical workflow for creating a data visualization solution starts with Power BI Desktop, application in which you can import data from a wide range of data sources, combine and organize the data from these sources in an analytics data model, and create reports that contain interactive visualizations of the data.

publish them to the Power BI service; a cloud service in which reports can be published and interacted with by business users.

Users can consume reports, dashboards, and apps in the Power BI service through a web browser, or on mobile devices by using the Power BI phone app.

## #Analytical modeling in Microsoft Power BI

[Describe core concepts of data modeling - Learn | Microsoft Docs](https://docs.microsoft.com/nb-no/learn/modules/explore-fundamentals-data-visualization/3-data-modeling)

Power BI to define an analytical model from tables of data, which can be imported from one or more data source.

## Describe considerations for data visualization

* Tables, text
* Charts, graphs
* Etc

## #Tutorial Visualize data with Power BI

[Exercise – Visualize data with Power BI - Learn | Microsoft Docs](https://docs.microsoft.com/nb-no/learn/modules/explore-fundamentals-data-visualization/5-exercise-power-bi)

* Install Power BI
* Import data
  + web
* Explorer data model
* Create report