Glossary: Introduction to NoSQL

Welcome! This alphabetized glossary contains many of the terms you'll find within this course. This comprehensive glossary also includes additional industry-recognized terms not used in course videos. These terms are important for you to recognize when working in the industry, participating in user groups, and participating in other certificate programs.

Term	Definition
ACID	This term is an acronym for Atomicity, Consistency, Isolation, and Durability, which is a set of properties that guarantee reliable processing of database transactions in traditional relational databases.
Aggregation	Aggregation is the process of summarizing and computing data values.
Aggregation pipeline	The aggregation pipeline in MongoDB allows for data transformation and processing using a series of stages, including filtering, grouping, sorting, and projecting. The aggregation pipeline is a powerful tool for expressive data manipulation.
Atomic	In the context of database transactions, atomic means that an operation is indivisible and either completed fully or is completely rolled back. It ensures that the database remains in a consistent state.
Availability	Availability in Cassandra, and in the context of CAP, is the system's ability to remain operational and responsive even in the presence of failures. Availability is a fundamental aspect of distributed systems.
B+ Tree	The B+ Tree is a data structure commonly used in database indexing to efficiently store and retrieve data based on ordered keys.
BASE	An alternative to ACID. Stands for basically available, soft state, eventually consistent. BASE allows for greater system availability and scalability, sacrificing strict consistency in favor of performance.
Basically available	A basically available system remains operational even in the presence of failures or faults.
Bigtable	A NoSQL database system developed by Google, designed for handling large amounts of data and providing high performance, scalability, and fault tolerance.
BSON	Binary JSON, or BSON, is a binary-encoded serialization format used for its efficient data storage and retrieval. BSON is similar to JSON but designed for compactness and speed.
Caching	The temporary storage of frequently accessed data in high-speed memory reduces the need to fetch the data from the primary storage, which can significantly improve response times.
CAP	CAP is a theorem that highlights the trade-offs in distributed systems, including NoSQL databases. CAP theorem states that in the event of a network partition (P), a distributed system can choose to prioritize either consistency (C) or availability (A). Achieving both consistency and availability simultaneously during network partitions is challenging.
Cluster	A group of interconnected servers or nodes that work together to store and manage data in a NoSQL database, providing high availability and fault tolerance.
Clustering key	A clustering key is a primary key component that determines the order of data within a partition.
Code-first	Code-first refers to a development approach where developers create the application code first and let the code define the database schema. In MongoDB, this means that the schema is flexible and adapts to evolving application needs.
Collection	In MongoDB, a collection is a group of MongoDB documents. Collections are analogous to tables in a relational database and store related data documents in a schema-free, JSON-like format.
Column database	A NoSQL database model that stores data in column families rather than tables, making it suitable for storing and querying vast amounts of data with high scalability. Examples include Apache Cassandra and HBase.
Consistency	In the context of CAP, consistency refers to the guarantee that all nodes in a distributed system have the same data at the same time.
Consistent	The action of being consistent ensures that a database transaction transforms the database from one consistent state to another.
CQL	Cassandra Query Language, known as CQL, is a SQL-like language used for querying and managing data in Cassandra.
CQL shell	The CQL shell is a command-line interface for interacting with Cassandra

Term	Definition
	databases using the CQL language.
CRUD	CRUD is an acronym for create, read, update, and delete, which are the basic operations for the basic operations for interacting with and manipulating data in a database.
Database	A database in MongoDB is a logical container for one or more collections. It provides an isolation mechanism for collections and their associated data.
DBaaS	Stands for database as a service, a cloud-based service that provides managed database hosting, maintenance, and scalability, allowing users to focus on application development without managing the database infrastructure.
Decentralized	Decentralized means there is no single point of control or failure. Data is distributed across multiple nodes or servers in a decentralized manner.
Denormalized	Denormalization is a database design technique used in NoSQL databases (and sometimes in traditional relational databases) where redundant or duplicate data is intentionally introduced into one or more tables to improve query performance and reduce the need for complex joins.
Document	A NoSQL database model that stores data in semi-structured documents, often in formats like JSON or BSON. These documents can vary in structure and are typically grouped within collections.
Durable	Guarantees that once a transaction is committed, its changes are permanent and will survive any system failures.
Dynamic table	A dynamic table allows flexibility in the columns that the database can hold.
Election	In a MongoDB replica set, an election is the process of selecting a new primary node when the current primary becomes unavailable.
Eventually consistent	An eventually consistent system reaches a consistent state, where all nodes have the same data given that there are no new updates.
Expressive querying	Expressive querying refers to the ability to write complex and flexible queries that address data retrieval and manipulation needs, often facilitated by MongoDB's query language and aggregation framework.
Extract, transform, and load (ETL)	A process of extracting data from its sources, transforming the data into a business-usable format, and then loading the data into a target database, often used with MongoDB for data integration.
Graph database	A NoSQL database model optimized for storing and querying data with complex relationships, represented as nodes and edges. Examples include Neo4j and OrientDB.
High availability (HA)	High availability (HA) in MongoDB refers to the ability of the database system to maintain near-continuous operation and data accessibility, even in the face of hardware failures or other issues. High availability is often achieved through features like replication and failover.
Horizontal scaling	The process of adding more machines or nodes to a NoSQL database to improve its performance and capacity. This is typically achieved through techniques like sharding.
Idempotent changes	Idempotent operations are those that can be safely repeated multiple times without changing the result. MongoDB encourages idempotent operations to ensure data consistency.
Indexing	The creation of data structures that improve query performance by allowing the database to quickly locate specific records based on certain fields or columns.
Isolated	Isolation refers to the property that multiple transactions can run concurrently without affecting each other.
Joins	Combining data from two or more database tables based on a related column between them.
JSON	JSON is an acronym for JavaScript Object Notation, a lightweight data-interchange format used in NoSQL databases and other data systems. JSON is human-readable and easy for machines to parse.
Keyspace	A keyspace in Cassandra is the highest-level organizational unit for data, similar to a database in traditional relational databases.
Key-value	A NoSQL database model that stores data as key-value pairs. It's a simple and efficient way to store and retrieve data where each key is associated with a value.
Lightweight transactions	Lightweight transactions provide stronger consistency guarantees for specific operations, though they are more resource-intensive than regular operations.

Term	Definition
mongo shell	The MongoDB shell, known as mongo shell, is an interactive command-line interface that allows users to interact with a MongoDB server using JavaScript-like commands. The mongo shell is a versatile tool for administration and data manipulation.
MongoClient	MongoClient is the official MongoDB driver that provides a connection to a MongoDB server and allows developers to interact with the database in various programming languages.
MQL	MongoDB Query Language is a query language specific to MongoDB used to retrieve and manipulate data in the database.
Normalized	A database design practice where data is organized to minimize redundancy and maintain data integrity by breaking it into separate tables and forming relationships between them.
NoSQL	NoSQL stands for "not only SQL." A type of database that provides storage and retrieval of data that is modeled in ways other than the traditional relational tabular databases.
Operational data	Operational data in MongoDB refers to the data that the application actively uses and manipulates, as opposed to historical or archived data.
Oplog	The Oplog is a special collection that records all write operations in a primary node. It is used to replicate data to secondary nodes and recover from failures.
Partition key	The partition key is a component of the primary key and determines how data is distributed across nodes in a cluster.
Partition tolerance	In the context of CAP, partition tolerance is the ability of a distributed system to continue functioning even when network partitions or communication failures occur.
Partitions	Partitions in Cassandra are the fundamental unit of data storage. Data is distributed across nodes and organized into partitions based on the partition key.
Peer-to-peer	The term peer-to-peer refers to the overall Cassandra architecture. In Cassandra, each node in the cluster has equal status and communicates directly with other nodes without relying on a central coordinator. If a primary node fails, another node automatically becomes the primary node.
Primary node	In a MongoDB replica set, the primary node is the active, writable node that processes all write operations.
Primary key	The primary key consists of one or more columns that uniquely identify rows in a table. The primary key includes a partition key and, optionally, clustering columns.
Replicas	Replicas in Cassandra refer to the copies of data distributed across nodes.
Replication	Replication involves creating and maintaining copies of data on multiple nodes to ensure data availability, reduce data loss, fault tolerance (improve system resilience), and provide read scalability.
Replication factor	The replication factor specifies the number of copies of data that should be stored for fault tolerance.
Replication lag	Replication lag refers to the delay in data replication from a primary node to its secondary nodes in a replica set. Replication lag can impact the consistency of secondary data.
Replication strategy	The replication strategy determines how data is copied across nodes.
Scalability	Scalability is the ability to add more nodes to the cluster to handle increased data and traffic.
Secondary	Secondary nodes replicate data from the primary and can be used for read-operations.
Secondary indexes	Secondary indexes allow you to query data based on non-primary key columns.
Sharding	Refers to the practice of partitioning a database into smaller, more manageable pieces called shards to distribute data across multiple servers. Sharding helps with horizontal scaling.
Soft state	A soft state acknowledges that the system's state might be transiently inconsistent due to factors like network partitions or concurrent updates. And it's willing to accept a certain level of inconsistency or uncertainty in the data temporarily.
Static table	A static table has a fixed set of columns for each row.
Table	A table is a collection of related data organized into rows and columns.

Term	Definition
Transactions	Transactions are sequences of database operations (such as reading and writing data) that are treated as a single, indivisible unit.
TTL	Stands for "Time to Live," which is a setting in NoSQL databases that determines how long a piece of data should be retained before it's automatically removed from the database.
Unstructured data	Unstructured data in MongoDB is data that does not adhere to a fixed schema. MongoDB allows for flexible and unstructured data storage, making MongoDB suitable for semi-structured or rapidly changing data.
Vertical scaling	Vertical scaling involves upgrading the resources (For example, CPU and RAM) of existing machines to improve performance.
XML	Stands for Extensible Markup Language, another data interchange format used in some NoSQL databases. It's also human-readable and can represent structured data.