Glossary: Cassandra Basics

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
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.
Aggregation	Aggregation is the process of summarizing and computing data values.
Availability	In the context of CAP, availability means that the distributed system remains operational and responsive, even in the presence of failures or network partitions. Availablity is a fundamental aspect of distributed systems.
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.
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.
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 databases using the CQL language.
Decentralized	Decentralized means there is no single point of control or failure. Data is distributed across multiple nodes or servers in a decentralized manner.
Dynamic table	A dynamic table allows flexibility in the columns that the database can hold.
Joins	Combining data from two or more database tables based on a related column between them.
Keyspace	A keyspace in Cassandra is the highest-level organizational unit for data, similar to a database in traditional relational databases.
Partition Key	The partition key is a component of the primary key and determines how data is distributed across nodes in a cluster.
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 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.
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.
Scalability	Scalability is the ability to add more nodes to the cluster to handle increased data and traffic.
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.
Transactions	Transactions are sequences of database operations (such as reading and writing data) that are treated as a single, indivisible unit.