Glossary of Terms

This is a high-level glossary of terms. To get more detail on a particular term, click on one of the related links. To find the information you are looking for, you can also try searching this site. If you would like to see additional terms defined, please email docs@clustrix.com

Search all Clustrix Documentation:

Jump to:

A

ACID

ACID (Atomicity, Consistency, Isolation, Durability) refer to the characteristics of a database that guarantee that transactions are processed reliably.

ALLNODES

Refers to the ability to set REPLICAS = ALLNODES.

Atomic

Refers to the characteristic of a transaction that is all or nothing.

B

B-Tree

Standard computer science data structure used for fast access. See B-Tree at Wikipedia.

Barrier

A barrier is a synchronization method used to control message flow within ClustrixDB. A barrier delineates a group of messages and all nodes must reach that barrier before proceeding.

Base Representation

The representation that contains all the table data and that is indexed by the primary key is the "Base Representation" or baserep. If no primary key is defined, ClustrixDB assigns a unique rowid key.

BigC

BigC is Clustrix's garbage collection process that cleans up undo logs needed to rollback running transactions. ClustrixDB must keep the system's state of a transaction the entire time a transaction is open. Once a transaction is committed, "BigC" removes it from the various undo logs as it is no longer needed.

Long-running transactions can cause BigC to become "pinned". That means that an old transaction must be preserved as it is potentially needed for recovery, yet subsequent activity on the cluster is causing the undo logs to become full.

Broadcast

Broadcasting refers to a method of transferring a message to all recipients simultaneously. ClustrixDB leverages distributed computing to avoid broadcasts. See how ClustrixDB scales joins.

Buffer Manager

ClustrixDB utilizes a Buffer Manager to exchange pages from disk to memory and vice versa.

C

cid

Commit Identifier that marks when transactional changes become visible to other transactions.

Cluster

A group of ClustrixDB nodes connected to provide a redundant, scalable RDBMS.

clxnode

The ClustrixDB database process.

Consistent

A consistent transaction does not violate any referential integrity during its execution.

Container

Containers are the base storage unit used by ClustrixDB. They define how representations are stored and retrieved using an access method such as B-Trees, layered trees, skiplists, etc. Each slice and replica of a representation will have its own container, regardless of whether or not that container is written to disk.

Cost

ClustrixDB uses a cost-based model for the query optimizer (Sierra) that uses a cost factor based on I/O, CPU usage, and latency.

D

Data Distribution

ClustrixDB leverages fine-grained data distribution and a shared-nothing architecture to provide scalability.

Database

A collection of tables, or relations. This is also sometimes referred to as a schema. (The ClustrixDB term for tables is "relations.")

device1

This file represents ClustrixDB's permanent storage and is used for all database data, undo logs, temporary tables, binlogs, and ClustrixDB system usage.

device1-redo

The write-ahead log (WAL) is stored in this file.

device1-temp

ClustrixDB uses this temporary storage for sorting and grouping of large query results.

Distributed Aggregates

Refers to the ability for ClustrixDB to perform aggregate queries (e.g. OLAP) in a distributed manner.

Distribution Key

The distribution key is some prefix of a representation's key columns and is used to distribute data across the cluster.

Durable

E

Evaluation Model

This describes how a query is evaluated in ClustrixDB.

F

Fair Scheduler

ClustrixDB component that prevents long-running queries from monopolizing cpu resources by giving priority to any waiting small queries.

Fanout

Fanout is the ability to use multiple CPUs to execute a query.

Flow Control

The GTM and other subsystems use flow control to prevent message senders from outpacing receivers and to prevent receiving nodes' memory from filling up with unprocessed messages.

Forward

Sending a row or rows to another node for further processing.

Fragment

A pre-compiled part of a query usually sent to another node for processing.

G

Global Transaction Manager (GTM)

A subsystem that manages the 12486558 commitment of transactions across the cluster, and ensures that all nodes involved come to the same decision every time.

Group

The list of all nodes known to the cluster.

Group Change

A group change is the event that occurs when a cluster forms a new group of nodes. This occurs when a node joins or leaves the cluster group.

H

I

iid

Invocation Id that marks the beginning of a statement within a transaction.

Index Distribution

ClustrixDB distributes indexes based on hashing the first column of an index unless advised otherwise. Also see Distribute.

Invocation

An invocation represents a single use of the query engine. Typically, queries use a single invocation, but DDL queries and those that call a stored procedure or function can use multiple invocations.

Isolation

is a property that defines how and when changes made by one transaction are visible to other concurrent transactions.

J

K

L

Layer Trees

Layer trees are a set of B-Trees that appear as a single container. This is the default container type used by ClustrixDB.

Lumpy

Refers to a poor data distribution. See how the Rebalancer and Managing Data Distribution for additional information.

M

Massively Parallel Processing

Refers to the ability to leverage a large number of processors to perform a set of coordinated computations in parallel.

Max_Failures

Defines the number of simultaneous failures that the cluster can survive. Also known as nResiliency.

Multi-Version Concurrency Control (MVCC)

A method used to implement concurrency and consistency in a distributed database environment. One of the original papers on this topic is Concurrency Control in Distributed Database Systems. ClustrixDB implements a modified version of this algorithm that provides optimizations for modern database workloads.

N

Node

A single server running the ClustrixDB software. Multiple nodes connect to form a cluster.

nResiliency

Another term for 12486558.

O

OID

Internal Object Identifier is a data type used by ClustrixDB internal structures.

OLAP

Online Analytical Processing.

OLTP

Online Transaction Processing.

P

PD (Probability Distribution)

Probability Distributions are tracked for values in each relation to aid in query planning.

Protected

Refers to the status of the cluster when at least two replicas of every slice are available. See also 12486558.

Q

Query Optimizer

The job of the optimizer is to determine which execution plan uses the least amount of resources. Typically this is done by assigning costs to a query's plan and then choosing the plan with the lowest cost.

Queue (Recovery)

Queues are used to track changes to data that may have occurred for a given node while it was unavailable to the cluster.

Quorum

ClustrixDB requires that a minimum number of nodes planned for a cluster are operational at any one time for it to be able to operate as configured. That minimum for ClustrixDB is called a quorum and it is calculated as one more than half of all the nodes configured for a cluster or (Total Nodes/2 +1). ClustrixDB cannot form a cluster without a quorum.

R

Ranked (or Ranking) Replica

Same as Read Replica.

Read Replica

ClustrixDB designates one replica of each slice of a table as the Read Replica. All reads are directed exclusively to that replica.

Rebalancer

The ClustrixDB Rebalancer automatically moves, copies, redistributes, and re-ranks data across the cluster.

Relation

A table in ClustrixDB.

Replica

ClustrixDB maintains multiple copies of data for fault tolerance and availability.

Representation

Every index, including the primary key, is called a "Representation" in ClustrixDB. Each representation is made up of a series of slices. The table's data is stored in the Base Representation.

Reprotect

When a slice has fewer replicas than desired, the Rebalancer will create a new copy of the slice on a different node.

Reslicing

As the dataset grows, ClustrixDB will automatically and incrementally redistribute the dataset one or more slices at a time.

S

Sierra

Sierra is the name given to the ClustrixDB Query Optimizer.

Sigma Containers

Sigma containers are temporary containers used to store intermediate results of some queries.

Skip Lists

ClustrixDB uses skip list data structures for In-Memory tables and some internal processes.

Slice

ClustrixDB breaks up each representation into a collection of logical slices. Rows are assigned to slices according to the results of a hashing function. See also 12486558.

Soft Fail

An operation that removes a node from a cluster.

T

U

Under-Protected

Refers to the state of the cluster when it does not have at least two copies (replicas) of each slice.

V

vrel

Virtual relation, often used to represent system information.

W

WAL

The Write Ahead Log is used to log every command that the user executes.

X

xid (Transaction ID)

An identifier used by ClustrixDB internals to denote the logical start of a Transaction.

Y

Z

Zones

ClustrixDB can be deployed across mutilple fault tolerance zones (AWS Availability Zones within the same Region, different server racks, different network switches, different power sources, or even separate servers in different data centers).