## Overview

There are no master nodes in casandra

A cassandra cluster is made up of nodes, each running an instance of cassandra

No single point of failure

All nodes can perform all functions

It is a partitioned row store

Data is read / written with a partition key

Partition key is hashed into a token

Token value determines which vnode / physical node is responsible for the data

Token ring

TODO

## Principles

Minimize need to access multiple nodes to satisfy a query.

Favor denormalization over normalization

## Data

Keyspace ⇒ table ⇒ partition ⇒ row ⇒ column

Column

Smallest unit of data, has a name, value and timestamp

Table

Collection of ordered columns fetched by row.

Each table has a single primary key

Partition

One or more columns that uniquely identify a row in a table

Partitions have a fixed limit as to how much data they can store.

TTL functionality is a useful mechanism for controlling storage

Partition key

The first column listed in a primary key definition

Determines on which node data is stored

Primary key

The partition key

First part of the primary key is a column name

May be compound, in which case the other columns in the key define the order of rows in the table

Structure of primary key impacts how many rows are stored in each partition

Clustering

Storage mechanism by which an index is created that determines the order in which data is stored within a partition

Allows multiple rows per partition.

non-primary key columns are stored together within the encompassing partition under clustering key.

Compound primary key

Primary key consisting of a partition key and one or more additional columns that determine clustering

Used when a single key won’t suffice to uniquely identify data

static columns

columns which represent the one side of a materialized one-to-many relationship

Or in other words, a set of columns which are constant within a partition

static column data is associated with a partition instead of any one row within the partition.

Row

Group of columns containing a primary key.

Columns, all of which share the same primary key.

Column family

table

Keyspace

Namespace container that defines how data is replicated across nodes

Top-level building block for organizing your data in Cassandra

Normalization

Process that attempts to minimize duplicate data.

Inefficient in Cassandra as joining tables together would require accessing multiple nodes

Denormalization

Process that attempts to optimize the read performance of a database by removing the need to join taboes together.

Entails duplicating data and grouping data together.

Counter

Special datatype

## Naming

No hyphens or spaces

Surround names that have mixed case or start with number with quotes

Don’t get creative…

Avoid mixed case and don’t start with numbers

## Data storage

### Nodes

All data stored is associated with a token value

If there is only a single node, the node is responsible for all token values

When another node is added, this node becomes responsible with the data associated with a contiguous range of token values

New nodes take over responsibility for tokens that were previously managed by other nodes.

### VNodes

Each node defines a number of vnodes

VNodes are only responsible for a small range of tokens

In a single node cluster, the node assigns the tokens to individual vnodes

With each additional node, the tokens the new node is responsible for is no longer a contiguous range. Only each vnode manages a contiguous range in the total range of tokens

Token space - the set of all available tokens

-2 to the 63 to 2 to the 63

VNodes provide a layer of indirection

Each node doesn’t have to contribute the same number of vnodes.

VNodes allow each token manager (vnode) to take over tokens from a larger number of other vnodes. It also increases the number of vnodes that must be communicated with.

### Partitioner

Dictates where data is stored

## Snitch

Mechanism used by cassandra to understand the environment in which the cluster is being run

Used to make efficient requests for data

Used when data is stored in multiple copies

SimpleSnitch

GossippingPropertyFileSnitch

Used by nodes to keep each other up to date on the state of the cluster

Uses a property file included in each node’s installation to provide info about that node’s location

Nodes gossip this info to each other

Can tell you when nodes are in different racks / datacenter

PropertyFileSnitch

Entire topology described in single file

EC2Snitch

EC2MultiregionSnitch

GoogleCloudSnitch

….

There is

VNode

Token

Indication of range of token values, or an individual token

### Writes

Writes by default are durable

Durability can be turned off, causing Cassandra to skip writing to the commit log before considering the write a success.

Writes are stored both in memory and on disk (commit log)

With durable writes, any crash that occurs before memtables are flushed will result in commit logs to be rerun on subsequent restart.

Local durability is supported by replication to further increase the durability of data