

AntiEntropy

AntiEntropy means comparing all the replicas of each piece of data that exist (or are supposed to) and updating each replica to the newest version.

Cassandra's implementation is modeled on Dynamo's, with [modifications to support the richer data model](#). Quoting from [Amazon's Dynamo](#) section 4.7,

- To detect the inconsistencies between replicas faster and to minimize the amount of transferred data, Dynamo uses Merkle trees. A Merkle tree is a hash tree where leaves are hashes of the values of individual keys. Parent nodes higher in the tree are hashes of their respective children. The principal advantage of Merkle tree is that each branch of the tree can be checked independently without requiring nodes to download the entire [...] data set.

The key difference in Cassandra's implementation of anti-entropy is that the Merkle trees are built per column family, and they are not maintained for longer than it takes to send them to neighboring nodes. Instead, the trees are generated as snapshots of the dataset during major compactions: this means that excess data might be sent across the network, but it saves local disk IO, and is preferable for very large datasets.

<https://c.statcounter.com/9397521/0/fe557aad/1/> | stats