Apache Cassandra

Navya Devineni, Naga Anshitha Velagapudi, Soujanya Janapatla

School of Computer Science & Information Systems

Northwest Missouri State University

Abstract

This article describes how you should write up your research paper.

You must write your article with the same format as this document.

Please read this document carefully.

1. Introduction

Apache Cassandra is a distributed open source database that can be referred to as a "NoSQL database" or a "wide column store". Cassandra has been used by some of the world’s largest companies to leverage the service that expands the globe.

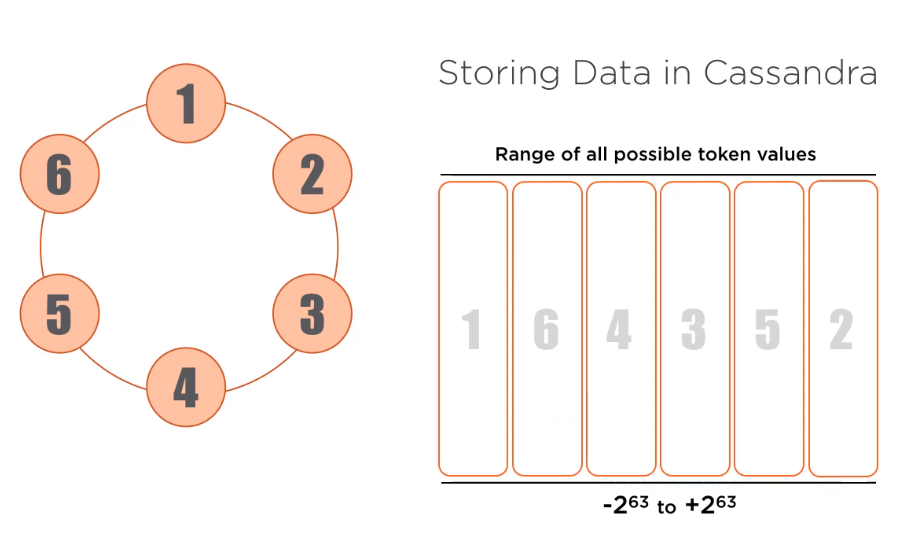
1. What is Cassandra

We shall start with the brief history of Cassandra. Cassandra was originally developed at Facebook, back in 2008. It was designed as the store for a Facebook feature Inbox Search. Such a feature requires a storage system that can handle a tremendous number of rights as well as geographical replication to reduce search latencies for end users.

Like most modern distributed NoSQL systems Cassandra was founded on the principal outline in 2 similar papers on the topic: Googles Cloud Bigtable and amazon Dynamo. Cassandra combines the distributed nature of Dynamo and data model of Googles Bigtable. Facebook open sourced Cassandra in 2008 and in 2009 it became an apache incubator project in 2010 it graduated to top level apache project.

Netflix famously migrated from Oracle to Cassandra running in Amazon public cloud, Apple was also reveled to be a significant user of Cassandra running more than 75000 nodes and storing tons of Terra bytes of data.

1. Cassandra Cluster



This diagram is probably the most common representation of a Cassandra installation. Each dot on the circle is a Node, which represent a running instance of Cassandra. This diagram helps strive on the point that a Cassandra cluster is a true master-less peer to peer system with no single point of failure. All nodes can perform all Cassandra function.

1. How Cassandra operates under the hood

All data stored in Cassandra is associated with a token, there are astronomical number of possible token values, falls in this range -2 63 to + 2 63. As new node are added each node taking over a contiguous range of token values and storing the data associated with it. 2nd node, 3rd node, and so on till 6th node.

1. Sample data

1. Sample
2. We shall start with the brief history of
3. Sample 2

basics of how Cassandra works

Consistency levels and replication strategies

Cassandra Query Language

More recent additions

What is Cassandra,

A brief history of Cassandra

Topology of a Cassandra cluster

Like most mordern distributed noSQL systems Cassandra was founded on the principal outline in 2 similar papers on the topic: Googles Bigtable and amazon Dynamo