Apache Cassandra 1.2 QuickDoc

What is Apache Cassandra?

C* is a NoSQL database with a flexible schema. Built for highly-available applications, it is fully distributed with support for multiple data centers and no single point of failure.

Open sourced in 2008, C* was built with scalability in mind. Used by thousands of companies for OLTP workloads, it handles extremely large data sets with ease.

Getting Started

Download Server
http://www.planetcassandra.org/Home/GettingStarted

Download Client
http://www.planetcassandra.org/Download/ViewDownloads/ClientDriver

Getting Started Documentation
http://www.planetcassandra.org/Home/GettingStarted

Operating Systems

32/64 Bit Architectures
Macintosh OS X
Windows 7
Windows Server 2008
Centos 5.x, 6.x
Debian 5.x, 6.x
Red Hat Linux Enterprise 5.x, 6.x
Ubuntu 10.x, 11.x, 12.x

Additional Resources

Find a Meetup Near You
Check out the Planet Cassandra Meetup Dashboard, to find a local meetup near you.

Cassandra Community Podcasts
Looking to improve your knowledge of Apache Cassandra? Listen to our weekly podcast series "Cassandra Community Podcasts".

C*ollege Credit Webinars
More of a visual learner? Don't forget to register for the C*ollege Credit webinar series.

White Papers
Read up on Apache Cassandra 1.2, NoSQL, Hadoop vs. Cassandra File Systems, Multi-Data Center Ops and much more.

What's New in 1.2

Apache Cassandra 1.2 White Paper
Check out the entire Apache Cassandra 1.2 White Paper

VNodes
Prior to this release, Cassandra assigned one token per node, and each node owned exactly one contiguous range within the cluster. Virtual nodes change this paradigm from one token and range per node to many tokens per node.

Collections
Collections provide easier methods for inserting and manipulating data that consists of multiple items that you want to store in a single column; for example, multiple email addresses for a single employee. There are three different types of collections: set, list, and map.

CQL
CQL 3, which was previewed in Beta form in Cassandra 1.1, has been released in Cassandra 1.2. CQL 3 is now the mode for cqlsh. CQL 3 supports schema that map Cassandra storage engine cells to a more powerful and natural row-column representation than earlier CQL versions and the Thrift API.

Use Cases

Netflix
As Netflix moved into the cloud, they needed to find the appropriate mechanisms to persist and query data within their highly distributed infrastructure. Their goal is to build fast, fault tolerant systems at Internet scale; Netflix realized that in order to achieve this goal, they needed to move beyond the constraints of the traditional relational model.

Spotify
Spotify uses Apache Cassandra for a variety of different data storage needs to enable them to stream millions of songs to millions of people. The open source Cassandra database provides a high availability, high performance database that scales to multi-terabyte datasets.

Barracuda Networks
The Barracuda Central Research Database is using Cassandra to battle the Zombies. Before adopting Cassandra, we could not monitor every malicious site and IP forever - the data volumes were just too great. We would monitor a site or IP for a while, and once we saw that the IP address was no longer alive we would stop monitoring it or need to truncate our history. The big problem however, was once we stop monitoring a site or domain they frequently come back to life - hence the Zombie moniker.

Comcast
Comcast's next generation cloud TV platform, their open sourcing of Amazon's Simple Notification Service (SNS) and Simple Queue Service (SQS) are built on top of Cassandra for its multi-data-center availability, extreme scalability, and very low latency.

Read More Use Cases At
www.PlanetCassandra.org/companies