



DataStax: Apache Cassandra™ Backgrounder

AT-A-GLANCE

Built for Big Data

- Peer to peer architecture ensures fast performance at extreme scale.
- Linear performance gains possible via online node additions.

Continuous Availability

- Transparent fault-tolerance built into the design.
- Replication across nodes, multiple data centers, and the cloud ensures no downtime.

Dynamic Data Model

- Flexible schema is built for managing modern types of data.
- Supports structured, semi-structured, and unstructured data.

What is Apache Cassandra?

Apache Cassandra™, an Apache Software Foundation project, is an open-source NoSQL distributed database management system. Cassandra is designed to handle Big Data workloads across multiple data centers with no single point of failure, providing enterprises with extremely high database performance and availability. Cassandra was initially developed at Facebook, and is used by companies such as Twitter, Netflix, and Cisco.

Cassandra's "cluster ring", peer-to-peer architecture, which can include hundreds to thousands of identical nodes, protects from data loss and business disruption because there is no single point of failure. Cassandra is architected to deliver extreme levels of performance for both read and write workloads, and is designed to offer linear performance increases via the addition of new nodes to an existing cluster.

What are the benefits of Apache Cassandra?

- **Massively scalable ring architecture:** Based on the best of Amazon Dynamo and Google BigTable, Cassandra's peer-to-peer architecture overcomes the limitations of master-slave designs and allows for both high availability and massive scalability.
- **Linear scale performance:** Nodes added to a Cassandra cluster (all done online) increase the throughput of your database in a predictable, linear fashion for both read and write operations.
- **No single point of failure:** Data is replicated to multiple nodes to protect from loss during node failure, and new machines can be added incrementally while online to increase the capacity and data protection of your Cassandra cluster.
- **Transparent fault detection and recovery:** Cassandra was designed with the understanding the hardware failures can and do occur, and therefore, failover is handled transparently.
- **Flexible dynamic schema data modeling:** Cassandra's dynamic schema can easily store structured, semi-structured, and unstructured data.
- **Guaranteed data safety:** Cassandra far exceeds other systems on write performance, while ensuring durability, due to its innovative append-only commit log.
- **Location transparency:** Cassandra allows for reading and writing to any node in a cluster.
- **Tunable consistency:** Cassandra gives users the flexibility on a per operation basis to choose between very strong consistency or varying degrees of eventual consistency.
- **Multi-data center replication:** Cassandra offers support for easily spanning multiple data centers and the cloud.
- **Cloud enabled:** Cassandra's architecture maximizes the benefits of running in the Cloud.
- **Data compression:** Cassandra supplies built-in data compression, with some use cases showing up to an 80% reduction in raw data footprint.
- **CQL (Cassandra Query Language):** Cassandra provides a SQL-like language called CQL that mirrors SQL's DDL, DML, and SELECT syntax.
- **No caching layer required:** Cassandra offers caching on each of its nodes, so there's no need for a separate caching layer.
- **Simple install and setup:** Cassandra can be downloaded and installed in minutes, even for multi-cluster installs.

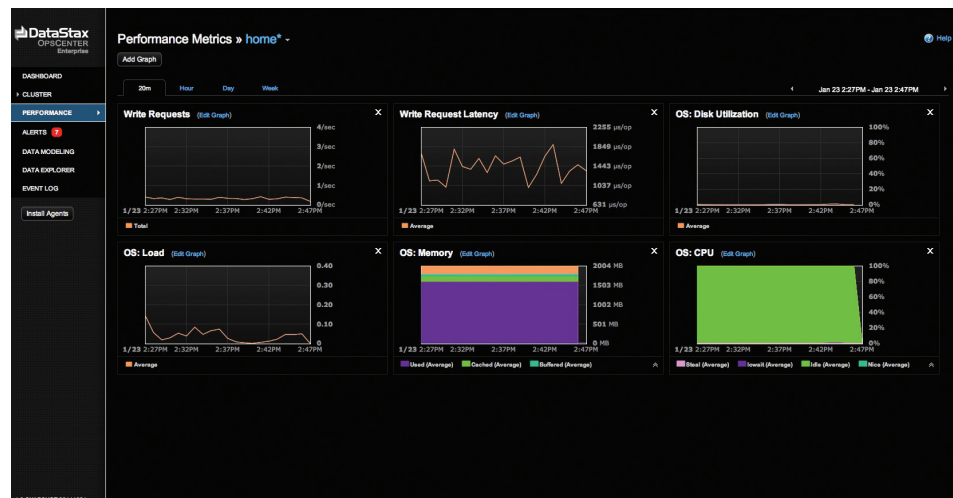
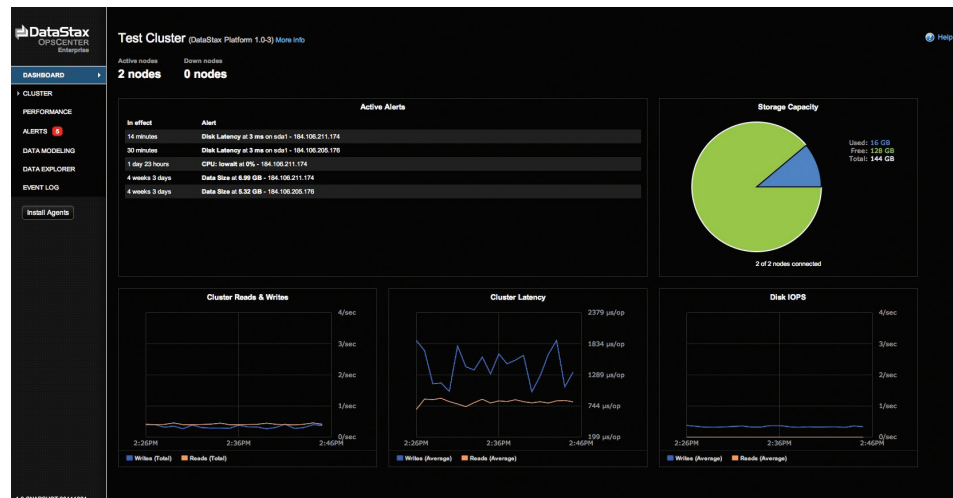
DataStax: Apache Cassandra Backgrounder

How does DataStax work with Apache Cassandra?

More businesses are turning to Cassandra to take control of large, active data sets. DataStax makes it easy for enterprises to build and operate high-volume, real-time applications that are powered by Apache Cassandra, insuring business continuity and preventing data loss. Unlike disruptive approaches that require developers to learn a whole new development stack, DataStax's technology allows development organizations to achieve massive scalability with the same Java and web development platforms already in use in their organizations.

DataStax OpsCenter

DataStax OpsCenter is the first platform for managing, monitoring and operating web and enterprise Cassandra applications, and offers sophisticated visualizations of an organization's Cassandra cluster. OpsCenter is available in two editions: an enterprise version that is bundled with a subscription to DataStax Enterprise, and a free community version that is available as part of DataStax's free DataStax Community bundle.



DATASTAX
270 East Lane #1
Burlingame, CA 94010

For more information visit www.datastax.com

DataStax, offers products and services based on the popular open-source database, Apache Cassandra™ that solve today's most challenging big data problems. DataStax Enterprise (DSE) combines the performance of Cassandra with analytics powered by Apache Hadoop™, creating a smartly integrated, data-centric platform. With DSE, real-time and analytic workloads never conflict, giving you maximum performance with the added benefit of only managing a single database. The company has over 100 customers, including leaders such as Netflix, Cisco, Rackspace and Constant Contact, and spanning verticals including web, financial services, telecommunications, logistics and government. DataStax is backed by industry leading investors, including Lightspeed Venture Partners and Crosslink Capital and is based in San Mateo, CA.