

Comparing Fast Data Performance

A comparison of VoltDB and Cassandra Benchmarks

Introduction

NoSQL databases, such as Apache Cassandra, are often used in operational applications. However, a new breed of databases, such as VoltDB, are revolutionizing the fast data landscape. VoltDB is an in-memory NewSQL transactional database for fast data applications, currently in use by major telcos, financial services, and many other markets.

To display the competitiveness in performance between VoltDB and Cassandra, two separate benchmarks were compared. The **first** benchmark was published in 2015, and was sponsored by VoltDB. The **second** was also published in 2015 by End Point, and compared Cassandra against Couchbase, HBase, and MongoDB. This study was commissioned by DataStax, whose main product is powered by Cassandra.

Both of these benchmarks are from 2015. However, there are no reasons to suspect more current versions are less powerful. Both systems have become slightly faster over time, but not significantly so. As such, this comparison is still reasonable.

Benchmark Methodology Comparison

The first benchmark (VoltDB) compared the same system on different hardware, while the second benchmark (Cassandra) compared different systems on the same hardware. Both benchmarks used YCSB workload B, and both looked at costs and operations per second. The full specifications used in this report are as follows.

The Amazon Web Service EC2 Compute Unit (ECU) is referenced a number of times. This is an abstraction of compute power that allows comparisons to be made regardless of the actual hardware involved.

VoltDB ran in 6 AWS c4.8xlarge instances. These are large, compute optimized instances. Amazon's stated use cases for this instance include ad serving and distributed analytics. Each c4.8xlarge has 4,000Mbps bandwidth, 36 CPUs, 132 ECU, 60 GiGB, and EBS-Only Storage at 1.76 \$/hr

Cassandra ran in a system comprising 1 to 32 nodes in i2.xlarge instance(s). These are smaller than c4.8xlarge instances. i2.xlarge is recommended for NoSQL databases, in-memory databases, and analytic workloads. It is worth noting that i2.xlarge has been replaced with i3.xlarge. Each i2.xlarge has 500 Mbps bandwidth, 4 CPUs, 14 ECU, 30.5GiB, and 1x800GB SSD at 0.85 \$/hr

Comparing 6 c4.8xlarge VoltDB nodes with 32 i2.xlarge Cassandra nodes, the summary becomes:

VoltDB: 6x c4.8xlarge has 24 Gbps bandwidth, 216 CPUs, 792 ECU, 360 GiGB at 10.56 \$/hr

Cassandra: 32x i2.xlarge has 16Gbps bandwidth, 128 CPUs, 448 ECU, 976 GiGB at 27.20 \$/hr

While a benchmark using the same exact hardware would be preferred, this configuration is fair for comparing performance relative to hardware. It is worth noting that 6 c4.8xlarge instances are almost 2.5 times cheaper than 32 i2.xlarge instances.

Results

With 6x c4.8xlarge hardware, VoltDB achieved 585,137 operations per second.

With 32x i2.xlarge hardware, Cassandra achieved 227,293 operations per second.

Transactions/second breakdown

	Per Node	Per CPU	Per ECU	Per Mbps	Per GB Memory	Per \$ (AWS costs)
VoltDB	97,523 tps	2709 t/s	739	24.4	1625	55,411
Cassandra	7103	1776	507	14.2	233	8356

There are a number of interesting comparisons to draw regardless of hardware differences.

VoltDB performs 1.72 times more operations per second per Mbps than Cassandra. Fast data is all about processing data the instant it arrives. As such, doing the most with bandwidth is key to creating successful fast data applications. VoltDB provides more performance with the same amount of bandwidth than Cassandra.

VoltDB performs 1.46 times more operations per second per ECU (EC2 Compute Unit) than Cassandra. Profitable fast data applications require getting the most out of the available hardware. By using Cassandra over VoltDB on the same hardware, some ECU is essentially wasted.

VoltDB performs 6.63 times more operations per second per AWS dollar than Cassandra. When looking at fast data technologies, the cost of running an application can be enormous. The staggering difference in operations per cost further pushes the notion that VoltDB can do more with the same.

MaxCDN leverages VoltDB to provide real-time analytics with their content delivery service — “we ... would have had to do an index lookup on this row-level transaction-id 80,000 times in order to guarantee idempotency [with Cassandra]. Now contrast this to 336 lookups in case of VoltDB. So, for every single index-lookup in VoltDB we would have had to do about 238 index lookups in Cassandra or HBase”

— Behzad Pirvali,
Performance Architect for MaxCDN

Conclusions

VoltDB thoroughly beats Cassandra in this comparison, which more than shows that VoltDB is comparable to Cassandra in terms of performance, if not better. In addition, VoltDB provides more performance per Mbps, per ECU, and per dollar than Cassandra. If you are researching translytical databases, VoltDB should be a part of your evaluation.

Stay tuned for future comparative benchmarks at votldb.com.

About VoltDB

VoltDB is the only in-memory transactional database for modern applications that require an unprecedented combination of data scale, volume, and accuracy. Unlike other databases, including OLTP, Big Data, and NoSQL, that force users to compromise, only VoltDB supports all three modern application data requirements: **1. Millions** — VoltDB processes a relentless volume of data points from users and data sources. **2. Milliseconds** — VoltDB ingests, analyzes, and acts on data in less than the blink of an eye. **3. 100%** — Data managed by VoltDB is always accurate, all the time, for all decisions. Telcos, Financial services, Ad Tech, Gaming, and other companies use VoltDB to modernize their applications. VoltDB is preparing energy, industrial, telco and other companies to meet the challenges of the IoT. VoltDB was founded by a team of world-class database experts, including Dr. Michael Stonebraker, winner of the coveted ACM Turing award.

2October2017

