

DATABASE SCALE WITHOUT LIMITS ON AWS

The move to cloud computing is changing the face of the computer industry, and at the heart of this change is elastic computing. Modern applications now have diverse and demanding requirements that leverage the cloud to achieve scale. As demand grows, applications add new webservers and storage nodes to add capacity, but scaling your database hasn't been as straightforward. Vertica and Hadoop are good scale-out solutions for offline analytics. Scale-out NoSQL solutions work well for non-critical or unstructured data. But scaling a primary SQL database that can concurrently run transactions and real-time analytics remains a challenge.

Scaling Your Primary Database in the Cloud

AWS is one of the most prominent players in the cloud computing space. By using AWS, companies can avoid the hassle of managing hardware purchases and co-location facilities. This infrastructure outsourcing enables companies to concentrate on their primary business operations with low up-front costs.

AWS excels at providing scale to web applications, but unfortunately there aren't a lot of options on AWS if a company wants a SQL database to:

- Scale and perform well beyond 1TB
- Provide fault tolerance and failover

Developers have tried to work around these limits via NoSQL, sharding, or re-working how their database is used—but to date, there hasn't been a scale-out solution for the primary database.

How Can a Scale-Out SQL Database Help?

Scale-out SQL (also called NewSQL) is a category of RDBMS that leverages the principles of distributed computing to provide scale while maintaining compliance with ACID, SQL, and all the properties of a relational database.

There are a variety of NewSQL players, each with slightly different architectures and benefits. ClustrixDB uses a shared-nothing approach that “brings the query to the data” (rather than forward data to a centralized compute node) to provide near-linear scale. Users can simply add nodes when transactions and users grow—without hitting the wall on database performance.

What is ClustrixDB?

ClustrixDB is the leading scale-out SQL database engineered for the cloud. With ClustrixDB, companies can scale transactions, run real-time analytics, and simplify operations. ClustrixDB uses a combination of intelligent data distribution and distributed query processing so companies can achieve horizontal scale-out by simply adding nodes as the database grows.

Clustrix has been serving large-scale, production workloads worldwide since 2008. Clustrix's largest customers have datasets with billions of rows, multiple terabytes of data, and transactional workloads approaching 100,000 TPS in production.

ClustrixDB on AWS Marketplace

ClustrixDB is easily accessible on the AWS Marketplace. ClustrixDB's scale-out capability goes hand in hand with the elastic scale found on AWS.

ClustrixDB enables the full power of a SQL interface and is a drop-in replacement for MySQL. Because of this compatibility, existing application code and connectors can be used with ClustrixDB with no code changes. For high availability and disaster recovery, ClustrixDB supports full MySQL replication and has fast parallel backup. MySQL replication allows ClustrixDB to replace existing MySQL databases on the fly.

ClustrixDB provides an easy to use interface to get up and running quickly. Additionally, it has an easy-to-understand UI for monitoring workload, CPU utilization, and slow queries while also providing tools to easily manage a database cluster.

ClustrixDB vs. Aurora

ClustrixDB, like AWS Aurora, is a MySQL drop-in replacement that runs in AWS. They both handle common site-visitor traffic very well. But Aurora leaves some customers high and dry, specifically those that depend on the ability to process high volume of transactions immediately and accurately.

However, even as the number of transactions soars, ClustrixDB continues to deliver superior levels of performance, ease of use, and cost effectiveness. As a high-performance relational database, it is specifically designed to meet the needs of customers with large, high-value transactional workloads.

ClustrixDB can deliver 2x, 5x, or even 10x the performance of Aurora, at much lower latencies—without complicated read slaves or sharding. This means that ClustrixDB can process more e-commerce check outs, or ingest more real-time data, or serve more ads, or do more of anything faster than you can do with MySQL or Aurora.

For more information on how ClustrixDB performs better and costs less than Aurora, please visit <http://www.clustrix.com/aurora-vs/>

Scalability

ClustrixDB is the only distributed transactional SQL database on AWS that scales to terabytes of data. ClustrixDB not only scales simple reads and writes, but also complex queries. To demonstrate how ClustrixDB scales on AWS, a performance test was run with an OLAP query that did a join of four tables and returned 100 rows after aggregating data over 32M rows.

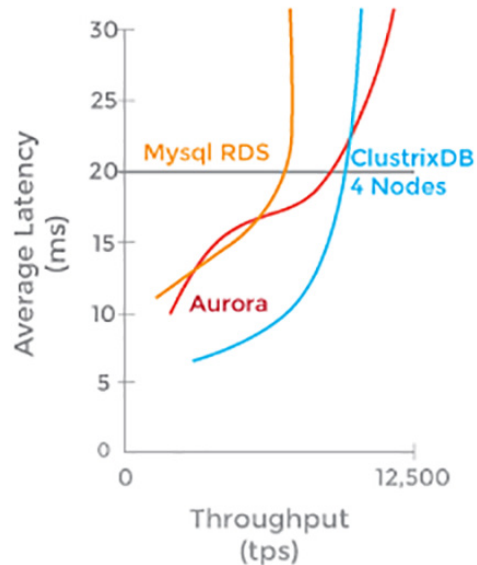
To back up this claim, we ran a Sysbench test with a typical workload that is representative of the kinds of workloads our customers see in the course of their business (e-commerce, gaming, adtech, social, etc).

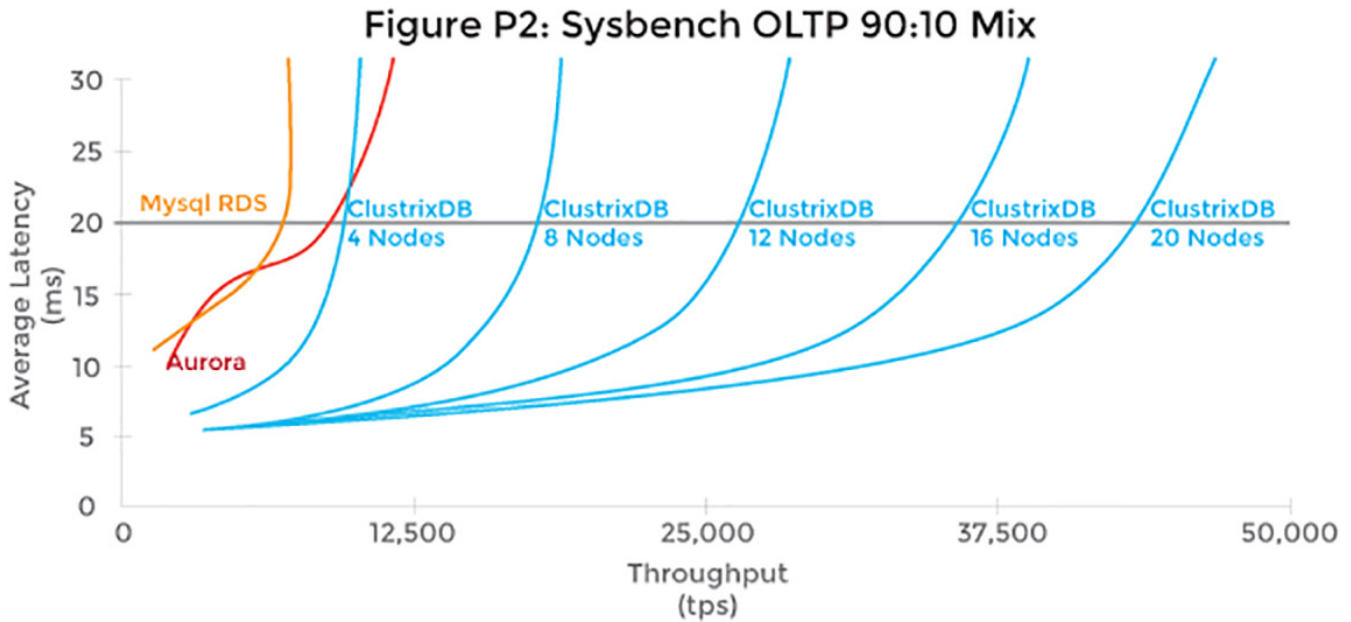
ClustrixDB started faster and stayed faster until the hardware became overwhelmed. Up until that cross-over point, ClustrixDB delivers more throughput at a lower latency rate than Aurora. At the crossover point, only ClustrixDB offers users the ability to instantly add more nodes to the cluster (see Figure P1) in order to increase throughput and keep latency low.

For high-performance and high-value workloads, ClustrixDB offers superior performance over Aurora in transactions per second and response time. And once Aurora is running on Amazon’s largest node, their ability to scale writes ends without resorting to database gymnastics like master/slave or sharding.

But ClustrixDB offers scale-out performance of both reads and writes. So we re-ran the benchmark with additional configuration (Figure P2). We ran an 8-node, 12-node, 16-node and 20-node cluster and showed how we can deliver 2x, 5x and 10x the performance of Aurora—all without modifying your enterprise applications or sharding your database.

Figure P1: Sysbench OLTP 90:10 Mix





Clustrix Linear Scalability

Clustrix built a distributed planner, optimizer, and compiler from the ground up for distributed query processing. Queries run in parallel across multiple nodes and cores for fast execution. The compiler transforms SQL into machine code for the fastest possible execution. Complex queries with joins and aggregates see significant increases in speed. Add more nodes and more cores, and the database uses multiple cores for a single node.

The distributed query processing enables massively parallel processing (MPP) that allows you to run fast real-time analytics on your primary database for operational intelligence. ClustrixDB offers Multi-Version Concurrency Control (MVCC) for lockless reads. The distributed multi-version concurrency control system ensures readers and writers never interfere with each other, making reads and writes fast even under highly concurrent loads.

ClustrixDB enables online schema changes that are completely lockless as well as online DDL operations.

As data grows and changes, ClustrixDB automatically splits and redistributes data to achieve a uniform distribution—each slice having copies on other nodes. As an application grows and the database reaches its limits, simply add nodes to increase capacity—no costly downtime or migrations to expand the database. The system maintains uniform data distribution as nodes are added, removed, or if data is inserted unevenly. There is no need to shard or worry about data distribution.

Fault Tolerance & Availability

ClustrixDB simplifies operations with built-in fault tolerance for high availability and self-managing operations. It retains all the power of SQL and ACID, yet delivers a fully distributed, shared-nothing architecture. ClustrixDB provides automatic fault-tolerance, linear scalability, online expansion, and online schema changes.

Within a cluster, ClustrixDB maintains multiple copies of all your data. In case of failure, extra copies are automatically generated to replenish lost ones. This self-managing database ensures you have high availability with no interaction required. ClustrixDB is transactional, immediately consistent, and durable.

You get all the guarantees you’ve come to expect from your database so you know your business-critical data is safe and secure.

ClustrixDB clusters can be set up to replicate asynchronously across geography for high availability in case of a geographically regional event (such as a regional electricity outage). ClustrixDB parallel backup runs fast regardless of the cluster size, and adds to disaster recovery options.

Configuration

ClustrixDB requires a minimum 3-node configuration and instance types with at least 7GB of memory (m1.large or larger). All three instances must be of the same type.

ClustrixDB also provides a non-clustered developer edition that can be used for compatibility testing.

Conclusion

Getting started with ClustrixDB on AWS is easy. You simply reserve the instances you want and connect to the Clustrix UI. The configuration wizard then guides you through setup.

When is the right time to start using ClustrixDB on AWS? The best approach is to use ClustrixDB from the beginning, before you begin to feel database scale pains. By implementing ClustrixDB early, companies can take advantage of the operational simplicity and high availability that makes ClustrixDB the “MySQL database on steroids.” When your business hits a growth curve, applications just keep working—no moving your database when it’s a critical time for your business’s success.

Clustrix

Clustrix provides the leading scale-out SQL database engineered for the cloud. With ClustrixDB you can build innovative business critical applications that deliver real-time analytics on live operational data with massive transactional volume. Our exceptional customer service team supports more than one trillion transactions per month across a range of industry segments including Ad Tech, e-commerce, and social analytics. Clustrix customers include AOL, engage:BDR, MedExpert, Photobox, Rakuten, Symantec, and Twoo.com. Headquartered in San Francisco, Clustrix is funded by HighBAR Partners, Sequoia Capital, U.S. Venture Partners, Don Listwin, and ATA Ventures. ClustrixDB is available as a free community edition for developers, a software download that runs on any cloud, and on the AWS marketplace.

To learn more about Clustrix, visit us at www.clustrix.com

