

Web Scalability with NoSQL

How to Support Millions of Users with Couchbase

Modern Web Scalability – How Big is Big?

Web applications are driving the need for more easily scalable database technology. Not long ago, 1,000 daily users of an application was a lot and 10,000 was an extreme case. Modern web applications often need to deliver a great interactive experience to millions of users per day, and support wildly-fluctuating numbers of concurrent users. A great user experience means a responsive application, which means fast, predictable access to data.

Supporting large user base is important, but it's just as important to handle rapidly growing (or shrinking) numbers of concurrent users:

- A newly launched app can go viral, growing from zero to a million users overnight – literally.
- Seasonal swings like those around Black Friday, the holiday season or Valentine's Day can create massive spikes for short periods.

With relational technologies, many application developers find it difficult, or even impossible, to get the dynamic scalability and level of scale they need while also maintaining the performance users demand. Many are turning to NoSQL for help.

Massive Scale – Couchbase Style

What does large-scale mean to Couchbase? The world's largest enterprises rely on Couchbase for their mission-critical web and mobile applications. Couchbase Server has proven scale to a billion user profiles in a single application (adding 500k per day!), and we have customers supporting more than 6M concurrent users with systems capable of more than a million operations per second. When a Zynga game called Draw Something exploded from 0-50M users in 50 days, they were able to scale seamlessly as demand increased.

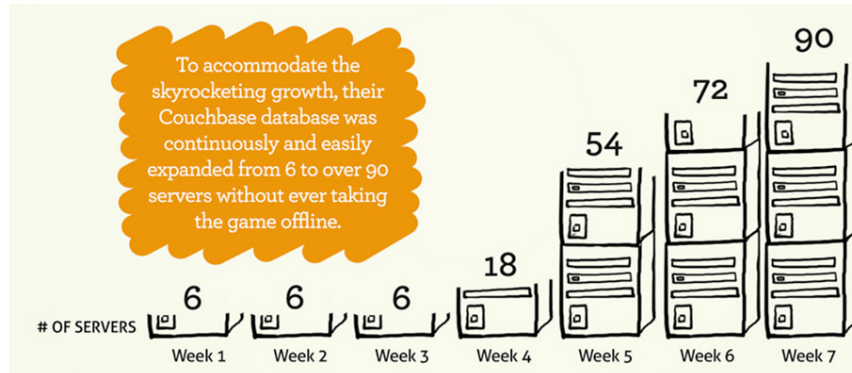


Figure 1 - Draw Something grew from 0 to 50M users in less than two months

Web-facing apps like these have to be fast and reliable at scale, and depend on Couchbase Server's ability to deliver sub-millisecond latency, always-on operation and one-click scalability.

It is important to be able to scale, but that scalability has to come with production manageability.

Why NoSQL Scales Better Than SQL

To deal with increases in concurrent users and data volumes, application databases grow by scaling up or scaling out. Scaling up implies an approach that relies on bigger and bigger servers. Scaling out implies a distributed approach that leverages many commodity physical or virtual servers.

At the web/application tier, a scale out approach has been the default for many years. By distributing the load across many servers (even across geographies) the system is inherently fault tolerant, supporting continuous operations, and makes efficient use of capital and operating resources.

However, prior to NoSQL databases the default scaling approach at the database tier was to scale up. This was dictated by the fundamentally centralized, shared-everything architecture of relational database technology. To support more concurrent users and more data, you need a bigger server with more CPUs, more memory, and more disk storage to keep all the tables. Upgrading a relational database server is an exercise that requires planning, significant hardware investment, and application downtime to complete.

NoSQL databases provide a linear approach to database scaling, using clusters of standard servers to store data and support database operations. Applications can be scaled simply by adding new servers to the cluster.

NoSQL databases are built to tolerate and recover from individual node failures, delivering high availability.

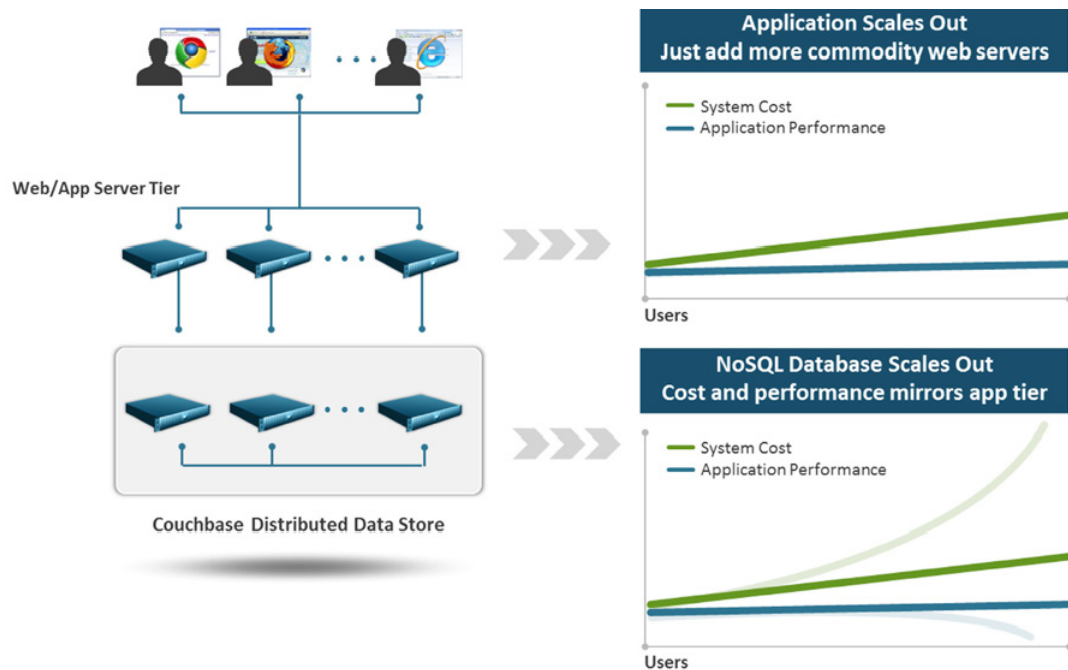


Figure 2 - Horizontal scalability with NoSQL

Why Couchbase Scales Better Than Any Other NoSQL Database

Couchbase Server was developed from the ground up to be a distributed, scale-out database, and has been proven in extremely large applications and those that scaled incredibly rapidly. There's no need to modify the application as you scale since the application always sees a single (distributed) database.

Simple Node Management.

Every server in a Couchbase cluster is the same and easy to set up. Additional servers are added as needed, and the data and database operations are spread across the larger cluster. If 10,000 new users start using your application, simply add another database server to your cluster. Add or remove nodes with one click, no downtime and no changes to your application.

Other NoSQL databases have multiple node types, complicated methods for adding or removing nodes, or must be taken offline to change the number of nodes.

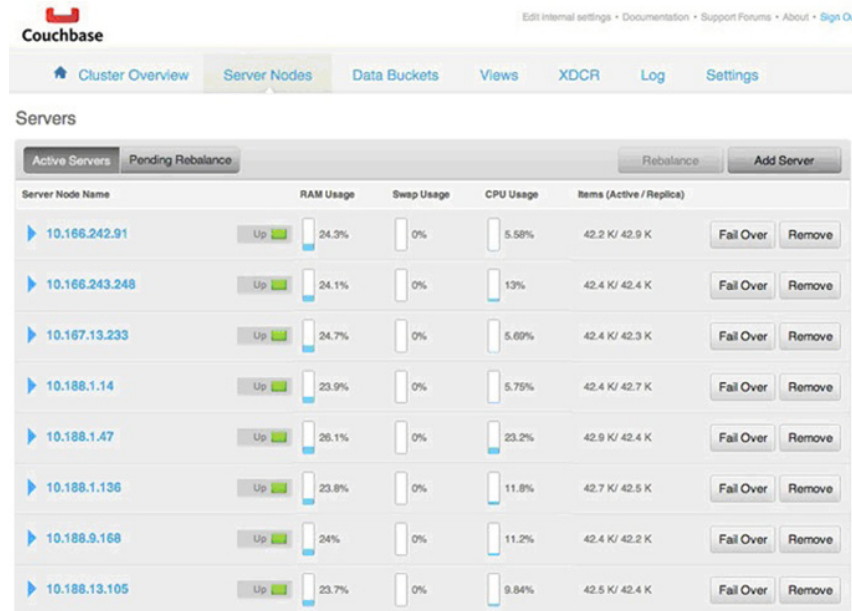


Figure 3 - You can scale your cluster with a click of a button in Couchbase

Effective Auto-Sharding

Because all nodes are identical, Couchbase auto-sharding balances load evenly across a cluster, so there are never any “hot spots” on a few overloaded servers. The limited sharding capabilities of other NoSQL databases require continuous manual tuning, which leads to performance issues and hot-spots.

You can grow your cluster to hundreds of servers with one click of a button, no downtime and no change to your application.

Scale Globally with Multi-Data Center Support

A modern web application is likely to have users around the globe, accessing it around the clock from every time zone. Data must always be available, and delivered with consistent high performance no matter where the user is located.

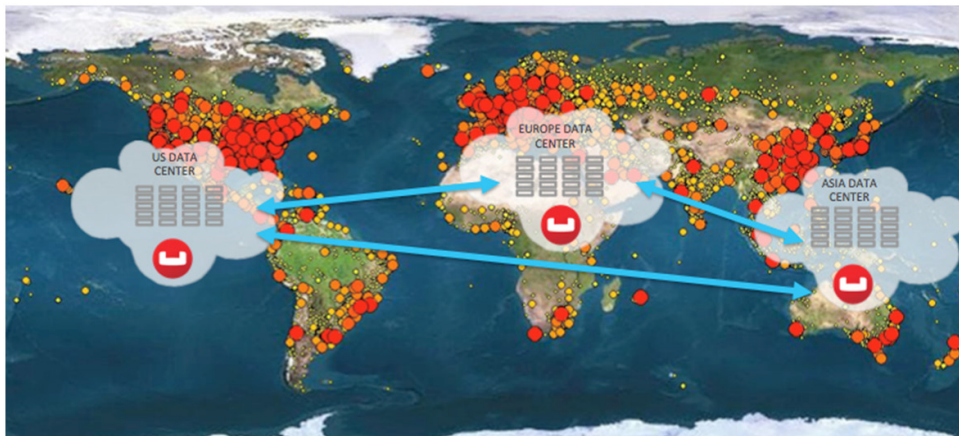


Figure 4 - Cross Datacenter Replication with Couchbase

Beyond simple intra-cluster replication, Couchbase Cross Datacenter Replication (XDCR) replicates active data to multiple, geographically diverse datacenters to bring data closer to its users for faster data access. With XDCR, an application based on Couchbase is scalable across geographies and time zones, delivering the best possible experience to every user all the time.

While replication within a database cluster helps with single server failures giving you high availability, your system still remains at risk from catastrophic failures – loss of power, natural disaster, etc. Couchbase XDCR provides complete disaster recovery and continuous operation, even in the event of a complete data center outage.

Conclusions

Application needs have been changing dramatically, due in part to three trends: 1) growing numbers of users; 2) growth in the volume and variety of data; and 3) the rise of cloud computing.

For interactive web and mobile applications, NoSQL has become a high-performance alternative to relational databases because it offers data management capabilities that meet the needs of modern applications:

- Greater ability to scale dynamically to support more users and data;
- Improved performance with sub-millisecond response time;
- Better development productivity through a more flexible data model.

Couchbase Server has proven to be the most scalable NoSQL database available, currently used by some of the world's largest enterprises for their latest mission critical applications.

The Five Couchbase Server Use Cases

Our customers have found Couchbase Server NoSQL database particularly well suited to five specific application problems:

Global User Profile Management

A great user experience means fast access to individual profiles and flexible support for rich profile data. Couchbase customer applications support billions of users every day. Couchbase offers:

- Proven scale to 1BN users of a single application
- JSON-based data model means flexibility to add and update user attributes
- Sub-millisecond latency means maximum responsiveness
- Easy scalability; grow from 1 to 100 servers with no application changes

Session Store Management

Modern web applications must be able to support rapidly changing numbers of concurrent users, and may need to support millions of users per day. Couchbase offers:

- Sub-millisecond latency with high read/write throughput for session, enabling easier management of session values or Cookies
- Easy scalability: simply add servers to the cluster and the data and database operations are spread across the larger cluster. Developed from the ground up to be a distributed, scale out database, Couchbase uses clusters of servers to store data.

High-Performance Caching

Relational database technology alone is inadequate for solving the scalability and performance problems related to short-term object storage and database retrievals. Couchbase offers:

- Consistent performance without shifting load to the RDBMS layer.
- High-availability caching of application objects, popular search query results, session information, and heavily accessed web landing pages speeds up RDBMS and allows for consistently low response times for document/key lookups.
- Data that is automatically replicated across the Couchbase cluster, providing high availability of data with easy non-disruptive expansion.

Content and Metadata Management

Couchbase's flexible JSON document model, easy scalability and high performance make it a great platform for storing frequently-accessed content and metadata. Couchbase provides:

- Consistently high performance that ensures a great user experience.
- A JSON document model that accommodates varied data and metadata requirements without any changes or downtime to your app.
- Easy scalability that helps you manage spiky loads. With cross data center replication (XDCR), scale across geographies and bring data closer to your users.
- Real-time distributed full-text search for JSON documents via integration with Elasticsearch.

Third-Party Data Aggregation

Data storage applications need the flexibility to store any kind of content and handle schema changes, plus high speed data ingestion and scalability. Couchbase provides:

- A JSON-based data model for schema flexibility
- Fast access to documents at any scale via high-performance built-in cache.
- Easy scalability and indexing and querying capabilities.

About Couchbase, Inc.

We're the company behind the [Couchbase open source project](#), a vibrant community of developers and users of Couchbase document-oriented database technology. Our flagship product, [Couchbase Server](#), is a packaged version of Couchbase technology that's available in [Community and Enterprise Editions](#). We're known for our [easy scalability, consistent high performance, 24x365 availability](#), and a [flexible data model](#). Companies like AOL, Cisco, Concur, LinkedIn, Orbitz, Salesforce.com, Shuffle Master, Zynga and [hundreds of others around the world](#) use Couchbase Server for their interactive web and mobile applications. www.couchbase.com