



MixRank Uses CitusDB for High Performance & Cost Savings

MIXRANK

Highlights

- MixRank stores in excess of 160 TB of data, including cold data.
- Company able to maintain a 50-100 times smaller hardware budget with CitusDB.
- MixRank end users immediately began seeing fresher data.
- IT easily transitioned existing applications to CitusDB.

MixRank's Challenge

MixRank is the Customer Discovery Platform that helps business-to-business (B2B) inside sales teams find new customers. It is the fastest way for sales reps to build prospect lists, prioritize leads, and contact decision makers. MixRank collects and analyzes vast amounts of data across the web and mobile. Customers use the solution daily as a critical step in their workflows to narrow down and qualify their prospects.

Citus Data Solutions at MixRank

CitusDB – Scaling out PostgreSQL to support big data and parallel query processing for real-time end user insights

When Scott Milliken, cofounder and CTO of MixRank, began coding the MixRank platform, he opted to use PostgreSQL. Hadoop was not an option because batch processing would not satisfy the near real-time needs of customers. And the commercial databases Milliken considered supported analytics only on in-memory datasets, which would have been cost prohibitive.

“CitusDB enabled us to dramatically increase performance and cut costs. It has been a competitive and operational game-changer for our business.”

Scott Milliken, Cofounder & CTO of MixRank

With PostgreSQL, MixRank was able to keep only a small amount of data in-memory, while the efficient PostgreSQL storage engine could deliver the required performance with the help of solid state drives (SSDs). PostgreSQL also enabled MixRank to rapidly prototype the solutions its customers needed.

Milliken knew, however, that vanilla PostgreSQL was not sustainable for the long term. The company needed to make tens of terabytes of time series datasets immediately available to customers. They also knew they would eventually need to distribute tables across a cluster, something not possible with plain PostgreSQL.

“We didn’t know when we would need to transition, or what the solution would be, so we had a looming anxiety,” said Milliken.

May 2015

MixRank and CitusDB

Fortunately for MixRank, Milliken met Ozgun Erdogan, CTO at Citus Data, at an industry event. Ozgun explained the value proposition of CitusDB was to scale out PostgreSQL databases across commodity servers and run real-time analytics and ad hoc SQL queries on large, distributed datasets, all without making any major changes to the existing PostgreSQL database. Milliken immediately decided to give it a try.

Easy to test

Milliken quickly launched a pilot of CitusDB: “Because we could use our existing PostgreSQL tables, the pilot was very fast and easy to set up. We just had to make some minor changes to the backend structure. After seeing the stability of the solution and the performance gains for our customers, we quickly decided to go into production.”

Boost in performance

The most obvious benefit of the shift to CitusDB was that customers immediately began seeing fresher data. Because CitusDB did not need to pre-compute as many elements as before, customers began seeing data from minutes or even seconds before their queries. “Most of our competitors still rely on daily batch updates, so this is a huge competitive differentiator for us,” said Milliken. MixRank currently relies on a 20-system cluster and uses SSDs for data storage. “CitusDB and SSDs have enabled us to achieve phenomenal I/O parallelism and I/O throughput.”

Huge cost savings & reduced complexity

CitusDB enabled MixRank to scale out using lower-cost machines. To optimize disk performance, Milliken switched to more expensive SSDs. Still, by avoiding the cost of hardware to support 100 percent in-memory analytics, Milliken believes his hardware budget is 50 to 100 times smaller than it would otherwise be. Further, by having one database for transactional data and analytics, Milliken has avoided costs related to maintaining multiple hardware platforms and multiple software licenses. He was also able to hire Postgres experts instead of generalists who could manage multiple platforms. “CitusDB enabled us to dramatically increase performance and cut costs,” said Milliken. “It has been a competitive and operational game-changer for our business.”

Leverage existing PostgreSQL effort

With CitusDB, MixRank was able to bring all the benefits of a relational database to its big data efforts without any significant changes to its existing database applications.

MixRank by the Numbers

- Over 160 TB of data, including cold data. 10 TB of data that ranges from hot to warm.
- Several petabytes of new data processed every month.
- A 20-node CitusDB cluster. Servers are 32-64GB RAM, 4-6 core (8-12 virtual) Intel Xeons. On each node, arrays of 20 Intel SSDs, totaling 2,400 GB. Effective IO throughput per node up to ~3.8 GB/s, up to ~700,000 IOPS.
- Total cluster capacity is ~48 TB, ~14 million IOPS, ~76 GB/s read.

About Citus Data

Citus Data brings the advantages of relational databases to big data while also decreasing costs and complexity. Citus Data extends PostgreSQL to seamlessly address big data challenges with familiar technology. The result is an extensible distributed SQL database with an open ecosystem that tackles heavy workloads orders of magnitude faster than what has been possible with PostgreSQL before. We make it simple for organizations to ingest, explore and aggregate their large datasets in real time. To learn more about Citus Data solutions, please call us at (415) 688-4279 or email us at sales@citusdata.com now.

