instaclustr

Case Study

Peloton scales data volumes reliably with Instaclustr

PELOTON

Overview

Peloton was founded in 2012 to create a new concept in fitness. The founding team loved cycling but had a hard time finding a workout that consistently fit with their busy schedules, and at-home workouts never felt quite as good as a class. They set out on a mission to create a world-class indoor cycling studio experience that would rival the in-class experience – all from the comfort of home.

Highlights

- Fast growing startup has raised $325 million
- Has tripled revenue and subscriber base in 12 months
- 100,000 paying subscribers
- Data-driven software provides personalized experience
- Data growth is exponential
- Was having issues with Apache Cassandra in production
- Moved to Instaclustr to scale reliably with zero downtime

> ***Working with Instaclustr has been great. They fixed our production issues with Apache Cassandra very quickly, and have since taken on responsibility for running our clusters in production. This means we can focus our efforts on building better products for our growing subscriber base, while feeling safe in the knowledge that our Cassandra clusters are always running at peak performance .***

Kengo Hashimoto
Senior Software Engineer at Peloton Cycle

Challenge

Peloton Cycle has a challenge that any startup would love to have. They are growing so fast from both a revenue and customer perspective, that it can be difficult to keep up with the growing data requirements. In order to provide the very best experience to their customers that they can, data is incredibly important to the company.

Peloton tracks and stores telemetric data from every ride that their customers take on their bikes, and due to the exceptional customer growth data volumes have been exploding. They track everything from how fast someone is pedaling, what resistance setting they are using, what their heart rate is like, and much more. Peloton then uses this data to display information back to their users with charts and graphs, including the top 10 all-time rides, their ratings, and the top 10 routines from all users at any given time.

Apache Cassandra is architected as a massively scalable, distributed database, and will have no issues with the exponential data growth that Peloton has been seeing, and is forecasting over the next three years. Being a small team Peloton does not have a lot of Cassandra skills in house. The engineering team wants to spend their valuable time building an increasingly better product, so doesn't want to wrestle with having to learn how to run open source Cassandra in production at massive scale.

They were starting to have some serious issues with performance in their production systems. The team had opted for fewer dense nodes, instead of a larger number of smaller nodes. One day there was a reconciliation issue between Chef and AWS which led to two nodes being destroyed, which in turn caused a painful and lengthy recovery process. The remaining nodes experienced read pressures, and fell behind.

They needed a partner that could diagnose and fix the issues they were experiencing right then, and then also take over as a managed service to free up important resources.

Solution

Peloton contacted Instaclustr, that provides a host of managed services for best of breed open source technologies, including Apache Cassandra, to see if their consultants could figure out the root cause of the performance issues that they had been experiencing.

Instaclustr diagnosed and fixed the production issues that same day. They found the problems related to running very dense nodes in production, and once the environment was right-sized from a node standpoint, Apache Cassandra resumed writing and reading the packet data in real-time.

Apache Cassandra is an amazing distributed database that is capable of handling incredible volumes and scale with ease. While relatively easy to develop against once the right data model is in place, it takes a considerable amount of experience and expertise to run the database in production.

For ongoing operations Peloton had two choices; they could hire for Cassandra expertise in-house, and have a dedicated team for operations, or look to Instaclustr's managed service. Having a dedicated team was deemed to be too costly, and would detract from core competencies.

Instead, the team decided to engage with Instaclustr to manage their production environment for them with tight SLAs in place. They already were using a managed service for another technology from a different vendor so it was a familiar model.

The Results

Since engaging with Instaclustr Peloton has been able to focus on their core competencies, while Apache Cassandra has experienced tremendous growth with zero downtime. Their data volumes are growing at a monthly rate of 8% and they expect to double their node count by next year, with over 500TB of data planned. Along with this data growth they expect their monthly active users to grow quickly to 350,000, and within two years they forecast that they will have over a million monthly active users.

They are currently running in two data centers in AWS, and as they grow their footprint globally they plan to add more data centers, which Cassandra can handle with ease. Additionally, Peloton is excited that Instaclustr is agnostic and can run in any cloud environment, so it gives them freedom and flexibility to move to or adopt a different public cloud vendor in the future if required.

About Instaclustr

Instaclustr is the open source as-a-service company, delivering reliability at scale with Apache Cassandra, Apache Spark, Elassandra and other related technologies.

Our expertise leverages over 15 million node hours and 1 petabyte under management, allowing us to run the world's most powerful NoSQL distributed database effortlessly.

Our open source data management solutions power mission critical, highly available applications for our customers and help them achieve scalability, reliability and performance in their database.