Use Case: Product Catalogs & Playlists

A product catalog is an organized collection of products, services, inventory or other assets that is constantly modified, searched and analyzed by both internal and external user communities. Playlists refer to user-defined queues of items like songs, movies, games, gift registries and lessons.

Industry

Digital Media

Challenges

- Limited scale across data centers and geographies
- Relational system can't deliver 100% availability of streaming service
- Difficult to analyze massive volumes of data

Solution

- Apache Cassandra for persistent data store
- Multi-data center replication and no single
  point of failure for 100% uptime
- Integration with Apache Spark for real time data processing and analytics

Results

- Over 1.5 billion playlists created and managed in real time
- More than 40,000 requests/second handled without latency
- Over 500 Cassandra nodes across 4000 servers in 4 data centers

---

PRODUCT CATALOGS + PLAYLISTS

SPOTIFY SUPPORTS 40K REQUESTS/ SECOND TO DELIVER THE RIGHT MUSIC TO THEIR CUSTOMERS' FINGERTIPS

Music streaming service Spotify, is a Swedish-born company that delivers a personalized music streaming service of over 24,000,000 songs and growing. Spotify makes it easy for over 40,000,000 active monthly users to translate their mood into music by finding the right playlist for every moment, and on any device.

The Challenge

As one of the fastest growing online and mobile services in the world, Spotify delivers streaming music in real time to over 40 million active users and growing, without interruption. To achieve the level of service demanded by its fast growing userbase, Spotify needed a database technology that could keep up with its growth without performance or availability issues.

Spotify initially started out as a postgreSQL shop, but with skyrocketing popularity, Spotify realized that a relational system couldn't keep up with their performance, scalability and availability requirements. What Spotify needed was a scalable solution that was highly available and could support multiple data centers.

"After we had scaled up to one or two million users we started to experience some scalability problems with certain services," said Axel Liljencrantz Backend Engineer at Spotify. "Once you hit multiple data centers, streaming replication in postgreSQL doesn't really work that well for high write volumes and so on because of its limiting architecture."

The Solution

To meet scale and availability requirements, Spotify selected Apache Cassandra for high-availability thanks to Cassandra's masterless architecture and ability to easily and quickly scale horizontally to meet their growing data store needs.

Spotify uses Cassandra to store data for their entire product catalog and key customer experience capabilities such as playlists, radio stations, notification popups, and the customized lists of artists, etc.

Where Cassandra helps with these services is by providing a persistent datastore that performs at high loads. Spotify user activity generates a large volume of logs and data which must be shipped from all over the world to a Hadoop cluster. This information is then transferred back from Hadoop to Cassandra and back to the users. "For example, the data that delivers your playlists and the data that delivers the discovery page are separated in totally different database clusters. This means that if the playlist databases would get sick for some reason, everything else in Spotify would still work perfectly. Decoupling is the key to scalability," stated Jimmy Mårdell, tech product owner at Spotify.

Today, Spotify has deployed over 500 Cassandra nodes both in production and for various testing clusters across four data centers around the world, so ensuring support for multi-data center replication is critical. "Cassandra gives us a level of trust that we won't lose data," said Liljencrantz.

> "Cassandra gives us a level of trust that we won't lose data. If there are bugs or crashes, we are confident it won't lose our data, that is very important to us."
>
> *- Axel Liljencrantz*
> *Backend Engineer*
> *Spotify*
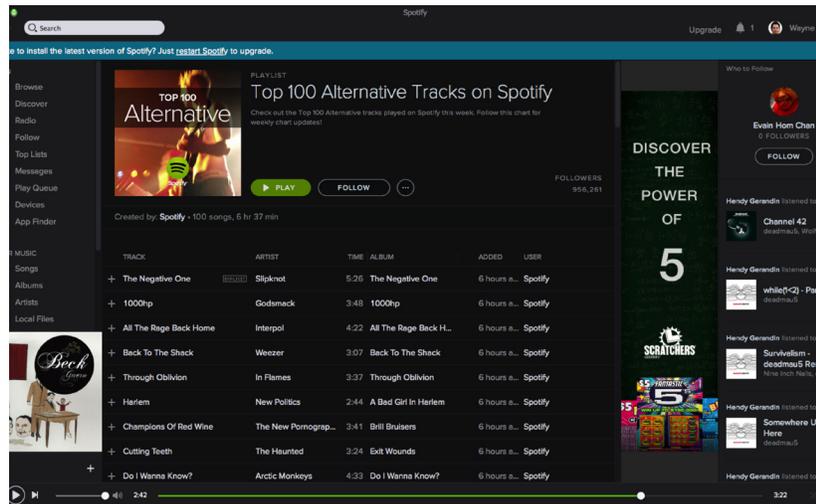
Cassandra provides Spotify with a persistent datastore that scales and never goes down.

The Results

Spotify manages dozens of different services in separate Cassandra clusters creating massive sets of dynamic data required to deliver personalized playlists. "The largest service that we have with the most data has somewhere just north of 50 terabytes of compressed data," said Mårdell.

Mårdell explained that Spotify current has over 40 million monthly active users that have generated over 1.5 billion playlists and counting. The speed at which new data can be generated and returned to users is key. With Cassandra, Spotify is confident that performance won't suffer even as its user base and data sets continue to grow. "At peak hours, we can easily handle over 40,000 requests/second without experiencing performance shortfall or latency," explained Mårdell. "That means when a user is listening to a specific artist or song, we can give back song and playlist recommendations when they want it which enhances their overall listening experience."