

Tencent 腾讯

Tencent News Delivers Exceptional Customer Experience and Predictable Service Levels with Alluxio

Tencent, based in China, is one of the largest technology companies in the world and a leader in sectors such as social networking, gaming, ecommerce, mobile, and web portal. Tencent News provides a rich, tailored news experience to over 100 million active monthly users. In order to meet the strict Service Level Agreements (SLAs) required by the business for optimal customer experience, the company turned to Alluxio for performance, predictability, and scalability.

Customer:
Tencent News

Industry:
Mobile Content Delivery

Use Case:
Analytics

Application Stack:
Spark + Alluxio + HDFS

Benefit Highlights

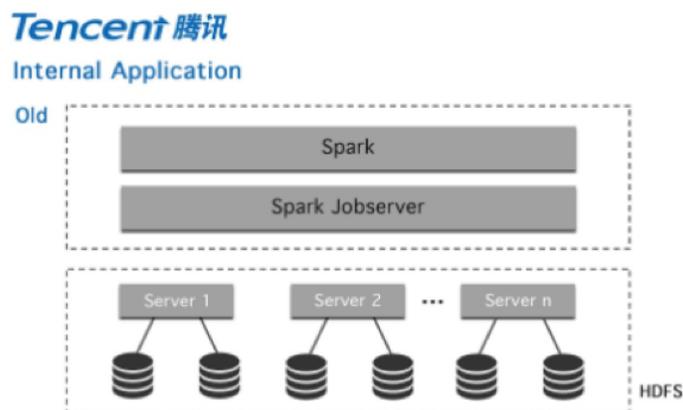
- Separation of compute and storage
- Predictably meet SLAs business requirements
- Independently scale compute and storage resources up and down
- No changes to existing applications

The Challenge

The goal of Tencent News is to deliver the best online customer news experience for over 100 million active monthly users. The application runs Spark jobs in a computation cluster of 150 dedicated servers and pulls data from an HDFS cluster outside the computation cluster in the data center. Jobs are required to complete on the order of seconds.

The architecture ensures Spark jobs can exclusively leverage a dedicated computation resource for performance isolation. Challenges occur when the data to pull is large or the machines/network are under high load. During peak times when the data required increases in size, or there is a heavy workload on the cluster, Spark jobs are not able to guarantee the ability to cache the data as RDD in memory and therefore read data from disk due to resource contention. This results in slow job completion, or worse, failures which take a long time to relaunch the job and reload the data. In these cases the company was not able to meet the business requirements for customer experience.

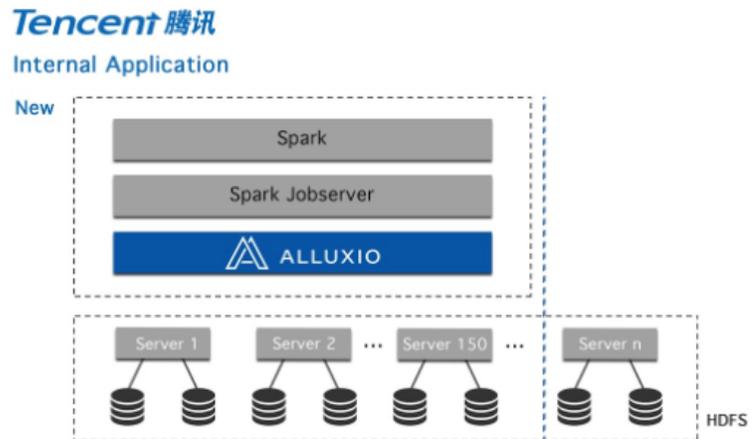
The following figure shows the previous generation Tencent architecture



The Solution

Tencent addressed their challenges by deploying Alluxio as the in-memory data layer for the application. Each machine in the computation cluster running Spark jobs reads or writes to Alluxio instead of caching data inside Spark processes. This effectively separates compute from storage and provides memory-speed data access for the application.

The following figure shows the next generation Tencent architecture expanded to over 600 nodes with Alluxio



The Results

The separation of compute from storage enabled the data to be accessed from memory in Alluxio and served locally to the nodes where Spark compute was running. Initial data requests are pulled from HDFS and subsequent requests are served from memory. As a result, performance is much higher and provides an SLA guarantee that meets the stringent requirements for job completion time.

Additionally, the deployment can be scaled up or down dynamically as requirements change. Changing this layer between compute and storage is transparent to application and independent from the size of the data required by Spark for compute.

Looking Forward

With Alluxio integrated in the data processing stack, Tencent now has an architecture that delivers the scalability, predictability, and performance required by their users. Alluxio is currently deployed on over 600 nodes with plans to expand the footprint further.