

CASE STUDY

Modernizing Flipboard's Hadoop-backed database

Increasing costs and unexpected spikes in usage created a challenge for the Flipboard team to fix their Amazon EC2-backed Cloudera Cluster with Apache HBase. Mactores modernized Flipboard's platform by migrating to Amazon EMR-based Apache HBase backed by S3, which reduced technical debt and eloquently supports massive spikes in user traffic.

Case Study Summary

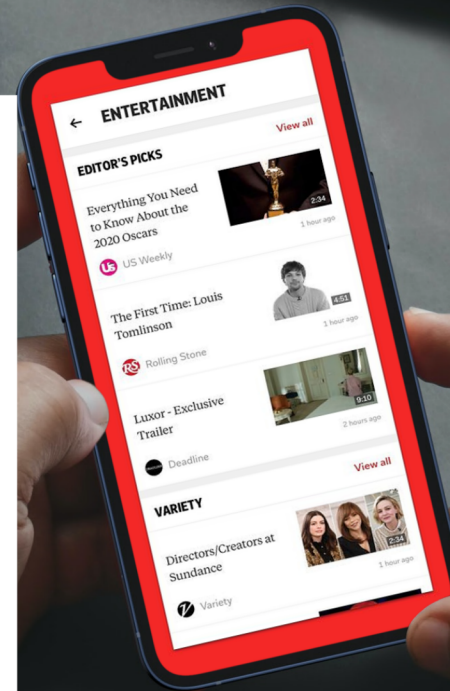
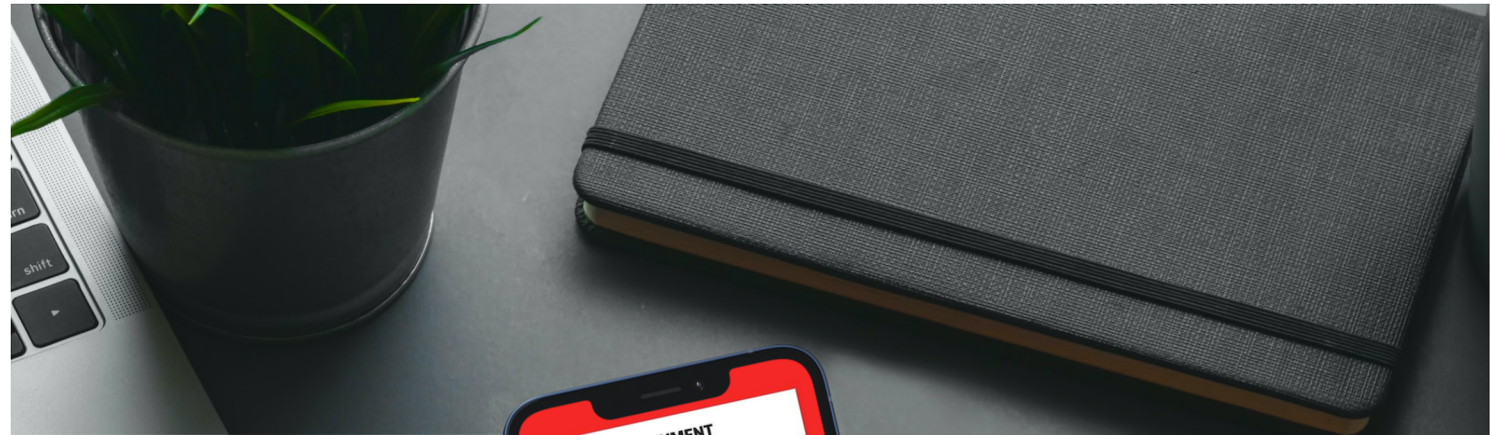
- Flipboard platform has 100 million active users, generating 200k-300k read/sec and 60k-120k writes/sec to Apache HBase.
- As the data volumes grew and user throughput requirements increased, Cloudera-based HBase was not scalable enough to support Flipboard's platform.
- Amazon EMR HBase backed by S3 support read replica
Flipboard now can separate read traffic from write and provide high throughput to users who perform interactive actions on the platform.

About The Customer



Flipboard is a news aggregator and social network aggregation company with 500 million users and 100 million monthly active users. Flipboard was founded as one place to find the stories for your day, bringing together your favorite news sources with social content to give an in-depth view into everything from political issues to technology trends to travel inspiration.

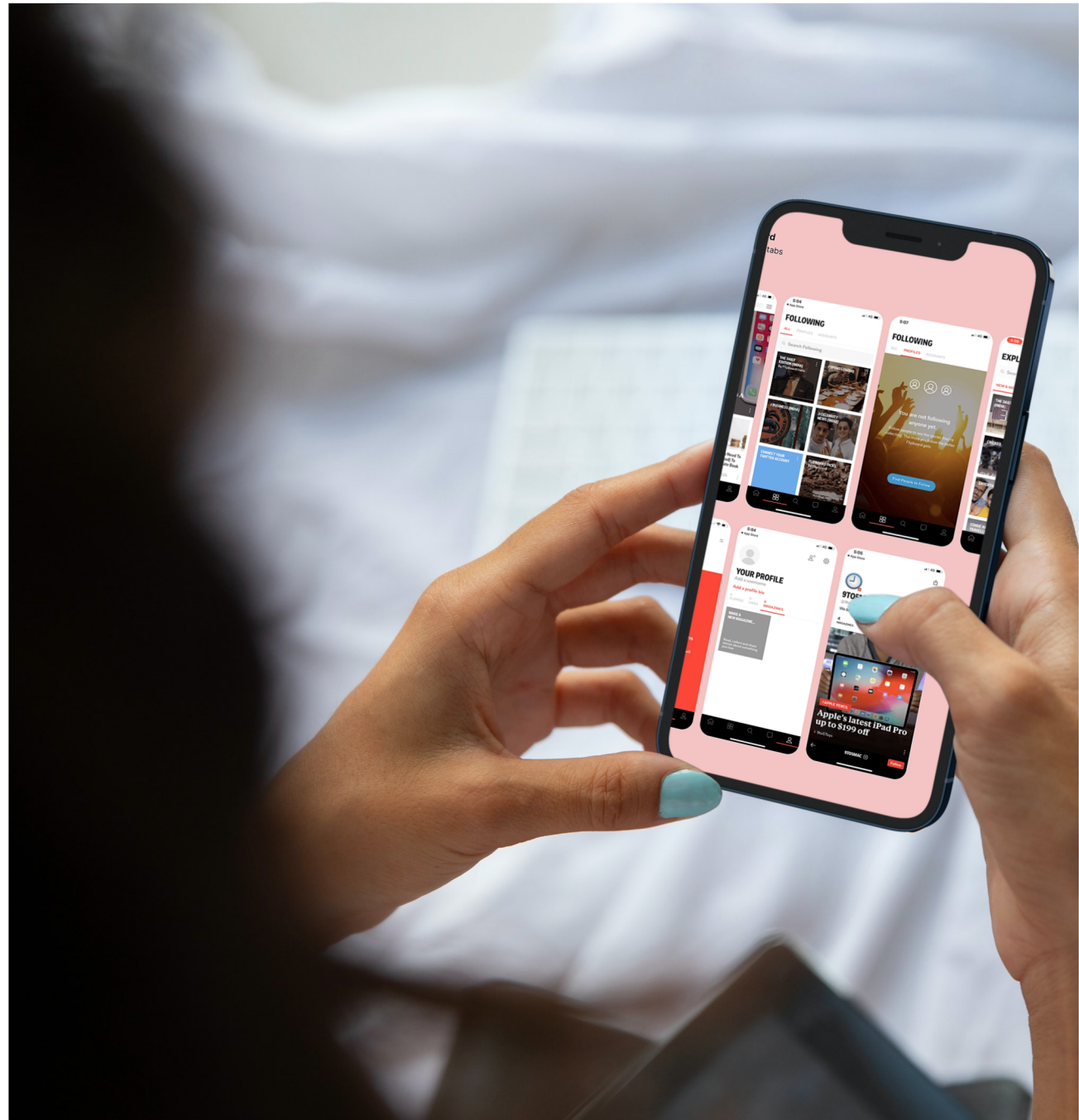
Over the years, in partnership with the world's greatest publishers, we've built a curated experience with a plurality of voices, where people can find quality stories on any interest, investing in their lives and passions.



Customer Situation

The Flipboard platform has 100 million active users who generate 200k-300k read/sec, and 60k-120k writes/sec to Apache HBase. They used a self-managed Cloudera Cluster on EC2 to support their platform with HBase with 5,600 regions and 40 TB of data on EBS.

As the data volumes grew and user throughput requirements increased, Cloudera-based HBase was not scalable enough to support Flipboard's platform.



Our Approach

Because Flipboard uses a significantly large payload for each key in HBase, the bucket sizes were challenging to identify.

So, the Mactores data engineering team decided it's best to test various Amazon EMR configurations and used Yahoo Cloud Severing Benchmark (YCSB) tool to benchmark all combinations of configurations to arrive at the recommended Amazon EMR deployment.

Due to significantly high traffic and the requirements of the average latency to be less than 100ms for "get" requests, the Mactores team optimized the bucket cache, block cache, and various other HBase parameters to support the S3-based HBase.

Business Outcomes

The Flipboard team has much more time to focus on forwarding motions now that they leverage the managed services.

By modernizing the platform, the Flipboard team now had the availability to add more functionality to improve the user experience.

With the spike concerns now properly managed, costs are much more manageable and predictable.

Technical Outcomes

Post Migration to Amazon EMR Flipboard benefits from the Autoscaling EMR clusters by adding region servers to support user traffic spikes.

With Amazon EMR HBase backed by S3 support read replica, Flipboard now can separate read traffic from writing and provide high throughput to users who perform interactive actions on the platform.

The next step of Modernization for Flipboard is planning to migrate some of the tables to DynamoDB and re-evaluate their use cases to use multi-column family databases.



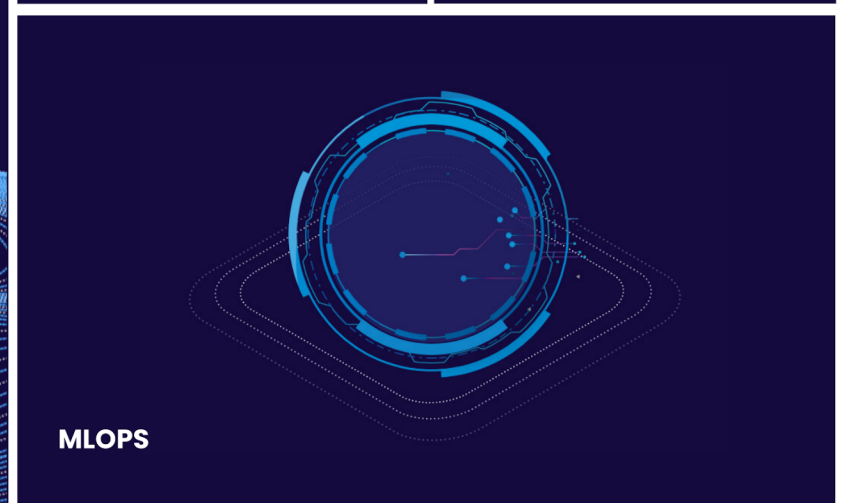
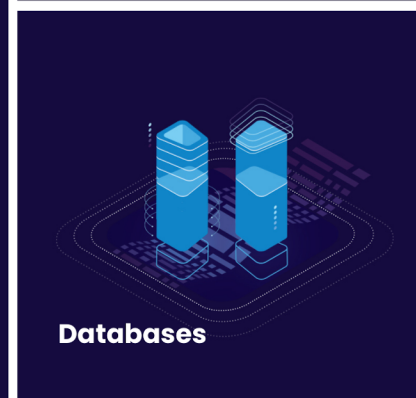
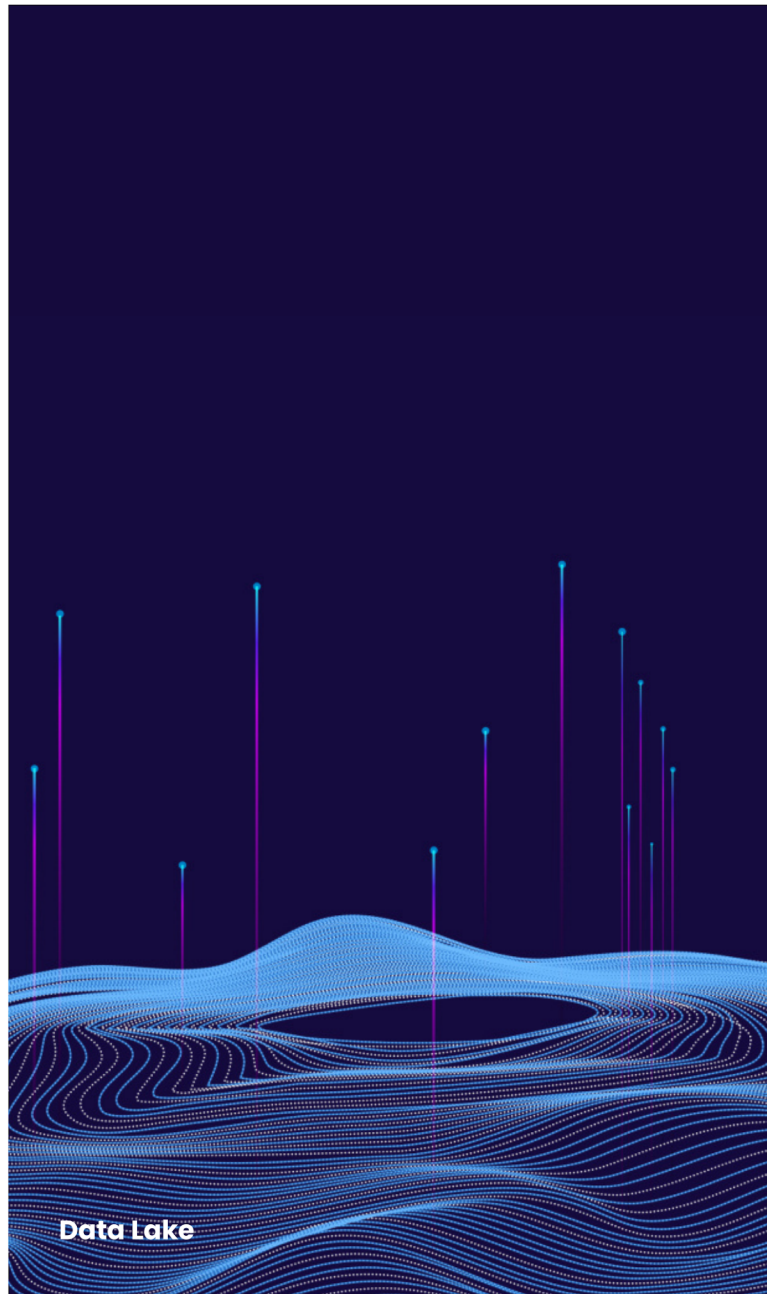
Getting Started

The Flipboard team had already spent too many hours establishing a sensible starting point to resolve this ongoing challenge.

The AWS team recommended using a partner with experience with HBase challenges. Mactores suggested the approach of performing an assessment that aligned with the Flipboard team.

The assessment results clearly defined the previous efforts and needed strategy to move forward with a production solution.

Our Solutions



Our Process

Digital transformation via assessments, migration or modernization

We work alongside your tech team to assess and strategize what you need and how to implement the right data solutions on time, on budget and with c-suite buy in.



Assess

- Discovery Automation
- Future State Assessment
- GAP Analysis
- End State
- Road Map
- TCO



Migrate

- Strategy
- Execution
- Migrate
- Migration Acceleration



Modernize

- StrategyFuture State
- Design
- Build
- Automate

AWS Validated Competencies



AWS Validated Service Deliveries



www.mactores.com/lets-talk

Mactores Cognition Inc

2018 156th Ave NE

Bellevue, WA 98007