Industry: Media

Iflix Distributes Video Streaming Online With Real-Time Analytics And Recommendations

Authors: September 18, 2018 by Saptarshi Bardhan and Holden Ackerman

In a little over three years, iflix, the Malaysia-based OTT service, has become one of the world's leading entertainment service providers for emerging markets. Today, iflix is available in 28 countries across Asia and Africa and has more than 10 million paid subscribers. Earlier this year, the company made news by bagging the **online live streaming rights to the Malaysian Football league.** Since the early days of the organization, iflix has been leveraging AWS services to power their business with data-driven decision making.

In this article we will recap the recent discussion Bruno Gagliardo, iflix's Global Director of Data Analytics, had with AWS Solution Architect Adrian De Luca on **"going over the top with streaming analytics."**

As is popularly known, the universe of online video streaming and real-time analytics relies heavily on data to deeply — and often in real time — understand the audience and audience engagement patterns, as well as analyze how content, packaging, and quality impact the business. Iflix embraced this data-driven culture throughout the evolution of the organization, and from the beginning, this culture has been a key driver of their exponential growth. As the company continued to grow, the scale of data kept ascending and the opportunities of what could be achieved from data also increased exponentially, demanding iflix look beyond traditional data warehouses and venture forward into big data.

As a new-age enterprise, iflix has been a user of Qubole's big data platform, which empowers the company to operationalize and activate their big data at enterprise scale. In the "chalk talk," Bruno spoke about iflix's use of Qubole within their data architecture and what drove the team to separate compute from storage.

How Iflix Uses Data To Create A Differentiated Competitive Advantage

With user and consumption data literally imploding, iflix takes a multi-dimensional approach toward ultimately boosting audience acquisition and engagement. The ability to seamlessly activate their big data has helped the OTT service fine-tune (and compete on) pricing and video quality, as well as offer a wider, more regional and demographic-focused catalogue than its competitors. The company's tie-ups with telcos in the region geared to enable the purchase of super small content packets also adds to their differentiated competitive advantage, and is yet another testament to their deep understanding of consumer behavior.

The Analytics And Data Science Capabilities Iflix Had In the Past

Previously, iflix would collect all player and internal events from their customers, and plant those into their Kinesis stream. Afterward, they would process and archive the raw data in their Amazon S3 service. On the reporting side, they would leverage RedShift as their data warehouse and as an Extract, Transform, and Load (ETL) tool. Redshift played a heavy role in the whole process, as they would send massive amounts of raw data and conduct around four layers of aggregation. Tableau, their business intelligence tool, sat on top of EC2 to transform the data and provide insights and information to their stakeholders and businesses.

A Renewed Approach – Decoupling Storage And Compute

Iflix was storing all of their raw data on S3, while Redshift was conducting the heavy lifting of ETL and storing raw and aggregated data for reporting. As the amount of data grew, Redshift couldn't service their data volume and ML needs given the effort and time that management would require.

Recently, iflix took a new approach to decouple data storage and compute. Now that iflix has millions of subscribers on board, they had to build new big data operations differently to better leverage AWS infrastructure capabilities — like leveraging Spark and Presto to save a lot of time on ETL. In the prior approach, the company was conducting six to 10 hours of ETL within Redshift, while through this new approach they could decrease the time needed by 40 percent — for just one stream of data. Thus, with Spark, they are enabling much more real-time analytics in a shorter period of time.

Why Does Iflix Now Have A Separate Virtual Private Cloud (VPC)?

iflix has partnered with Qubole as a title management platform that sits on another AWS account. They use a Bastion IAM role to enable the connection between the AWS accounts. Qubole provides an integrated development environment (IDE) and helps iflix seamlessly manage their different instances on the basis of personas and roles. Iflix can allocate resources within the IDE based on roles, allowing for personas such as engineers, data scientists, and business analysts to access specific compute capabilities or data retrieval capabilities.

Hadoop, Spark, And Presto – What Are Their Roles?

Iflix's new data architecture stores all of its raw data in S3, some of which is leveraged exclusively by the product engineering team and not by the data engineering team. Hence, they are isolating data from the production environment and creating a replica in their VPC. Hadoop or Hive, which is their meta store, allows them to access all of the data in S3, and users in different roles — such as data engineers, data scientists, and business analysts — can now access different data sets. Iflix leverages Spark for data engineers to orchestrate the ETL using Airflow, while data scientists and business analysts leverage Presto to query their meta store and get access to data sets. When users are ready to kick off jobs, they are interfacing directly with Qubole through profiles set up for each type of workload and role.

Qubole's Value Add

Qubole's platform enables iflix to assign different resources within the configuration, since a data scientist might not have as many nodes on Spark as a data engineer. Qubole also incorporates AWS Spot instances into autoscaling, which in turn saves significant costs and increases reliability. Given the flexibility of Qubole's cloud big data platform, if any of those workloads change, Qubole allows iflix to quickly resize the clusters to get the jobs done much more quickly.

Test Drive Qubole today to try different examples of Spark and Presto — whether you're a data engineer, an analyst, or a data scientist!