BIG DATA PLATFORM AND ADVANCED ANALYTICS ENABLES EXPANSION AND BUSINESS GROWTH

IMPACT
- Over 200 system users including light and heavy users
- Organization currently utilizes Cloudera's solutions on over 50 nodes through server consolidation
- Completing numerous migrations on schedule thanks to Cloudera Professional Services
- Data sizes have grown by almost ten times to over 1 petabyte since Cloudera was initially deployed

Offering various digital content and services online, entertainment company Dwango, Co. Ltd. is a telecommunications and media company based in Japan.

With its company vision of "Born in the net, Connected by the net," Dwango has been driving Japan's Internet industry and develops systems for online games.

The company provides communication platforms such as Niconico Video and Niconico Live Streaming. It creates value by using the power of the Internet to blend opposing concepts - virtual and reality, digital and analog, and innovation and tradition.

Challenge

Dwango was utilizing Apache Hadoop, but with an increasing demand for data analytics, and a desire to reduce the burden on infrastructure management, the organization embarked on a big data journey.

Infrastructure management had become difficult for Dwango's open-source analytics platform, requiring an increase in human power equipped with technical skills. The platform needed to be accessible cross-organization due to increased demand across their business. And last, more resources dedicated toward improving data analytics, including the use and management of a wide range of tools, would be imperative to drive business growth.

Advanced analytics drives new department growth

Dwango analyzes data created on a UGC (user-generated content) platform that includes user videos as well as text. Keiichiro Tsukamoto, General Manager of Data Management Service Department, ICT Service Division of Dwango, says, "The importance of data utilization is growing across the company, with the company forming new departments over the past several years during organizational restructuring as it seeks to create a cross-organizational data analytics platform."

To address growth and the amount of new unstructured data, the company needed to move away from an environment operated and managed on its own. "Instead of spending time and resources on infrastructure operations, we needed to improve data analytics, and allocate more resources to defining and managing data," recalls Tsukamoto.

Solution

After a process of validation to evaluate Dwango's support and specification needs, Cloudera was selected as the Big Data analytics platform to drive their journey forward. Since moving huge amounts of data and consolidating storage nodes were required, the company created a scalable analytics platform with data aggregation and analytics capabilities. The updated platform can be leveraged within and outside the company, allowing as one example, the ability to offer marketing data to owners of the Niconico video sharing site.

1PB

Data being managed

"An environment in which we could procure and migrate necessary resources from the cloud while leaving the on-premises system intact would be ideal. So, for that reason, I have great hopes for CDP as a new solution."

Keiichiro Tsukamoto, General Manager of Data Management Service Department, ICT Service Division, Dwango

Enhanced support across on-premises and the cloud was key to security and governance

During its research into a new environment, the company became interested in the Cloudera enterprise platform. "To reduce the burden of infrastructure management, we needed a solution with a support system. Cloudera's technical support team operates under a clear system, providing quick responses to our questions," says Tsukamoto. Dwango had previously operated UGC related services from its own data center, so Cloudera's extensive experience in delivering on-premises and cloud environments made it a good match. In addition, the authentication technology, including Kerberos used by Dwango, is compatible with Cloudera's platform making it suitable from a security and governance perspective.

The organization quickly applied new technology from Cloudera, including rich query engines capable of ETL style processing through SQL coding. The existence of Cloudera's web-based UI tool and Hue support were also big factors, according to Tsukamoto. "In the past, we offered an internally created web user interface intended to be easy to use for citizen scientists registered as members to each service. However, the task to keep up with open-source updates was also daunting. Cloudera's Hue support is rich in features, and we felt it would easily satisfy user needs."

Measuring effectiveness of service offerings for video and live streaming

The primary application measures the effectiveness of various services such as Niconico Video and Niconico Live Streaming. Separate from web analytics platforms used by the web director to analyze access, the data analytics platform provides detailed analysis of data, including application logs collected from multiple systems, as well as deeper analysis to measure impact. The distributed data storage system Apache HBase integrated with the Cloudera platform is used as part of the system to display comments on Niconico Video.

Dwango now has a conducive environment that supports business growth by providing marketing data to select channel owners. "For example, we provide marketing data to Niconico Channel owners who can set up subscription businesses. Analytics data for users who upload videos within Niconico Video uses data aggregated with Cloudera. The scope of use has expanded, and it has been helpful as a platform that supports business growth," says Tsukamoto.

Dwango has increased the cluster size to over 50 nodes through server consolidation. The data sizes have grown by almost a factor of ten - to over one petabyte since Cloudera was initially deployed.

The company has deployed Cloudera Manager to create an environment capable of operation and management. Now monitoring and distribution can be centrally managed, so the system can continue operating without increasing the number of personnel. In addition to Tsukamoto's department that handles cross-organizational data analytics, there are over 200 accounts using the system.

Results

These solutions have made cross-functional collaboration within the organization much easier. By using various query engines such as Hive for batch processing and Impala for high-speed analysis, it has become possible to obtain much-needed analytics results dynamically during meetings. Use of Impala and the business integration tool Tableau enables on-site visualization and exemplifies the contribution made toward speeding up data analytics operations.

"In addition, because we often undergo frequent organizational restructuring, a rigorous authentication platform is essential, so the ability to implement frameworks such as Kerberos authentication is helpful in creating a secure environment and enabling us to continue data analytics with peace of mind," explains Tsukamoto. In terms of governance, not only do Cloudera tools make data acquisition simple, the ability to manage information trails as an audit log was also part of its appeal.

As for support, Tsukamoto says that generous support from Cloudera engineers has made stable operation possible. "We had a lot of newbies on our team when the system was first deployed, so the leaders took on a lot of responsibility. Because we were able to rely on their support as soon as an issue arose, we were able to focus our human resources on building an environment for the future. I really appreciate their kind support in various situations ranging from migration planning to support for selecting needed equipment for platform expansion. Thanks to them, we were able to complete numerous migrations on schedule."

Moving forward, Tsukamoto says the company would like to use the Cloudera Data Platform (CDP) utilizing the cloud, while also leveraging the environment developed from an on-premises system. "Over many years of operation, we have accumulated a large amount of cold data, so I would like to implement an environment that can manage this data appropriately according to how frequently the data is used. An environment in which we could procure and migrate necessary resources from the cloud while leaving the on-premises system intact would be ideal. So, for that reason, I have great hopes for Cloudera Data Platform (CDP) as a new solution."

The company is looking at ways to further leverage the data analytics platform across other departments. "The fact that we can abstract the required environment for data analytics while centrally managing it with Cloudera Manager means we can count on this contributing to governance, and we'll be able to use all types of environments including on-premises, private clouds, and public clouds. Our data analytics needs will likely further increase as data continues to grow, so I think CDP will be seriously considered," said Tsukamoto in closing.