

INDUSTRY

Technology

TECHNOLOGIES

Cloudera; Apache Hadoop; AWS RedShift, DynamoDB and Simple Storage Service (S3); Google Cloud BigTable, BigQuery and Google Cloud Storage; Commvault; Swift; Ceph; MariaDB/AX; ClickHouse.

BUSINESS NEED

The client's very large, on-prem HBase database and Hadoop cluster was unable to perform even the most basic backup and disaster recovery actions. Should the database fail, they were facing the prospect of catastrophic data loss and a recovery process of several weeks if not months.

SOLUTION

Several on-premise and cloud options were presented that the client could employ to meet its goal of 12-hour recovery time to recover the past 12 months of data, including a new Hadoop cluster in another data center. Ultimately, a cloud-native object storage/column store option was recommended.

RESULT

The client's IT leadership was able to verify their well-researched assumptions through Pythian's vendor-agnostic expertise, including details on all their on-premises and cloud options. This gave the client much greater confidence that they were making the correct technology choice when building their BDR solution.

PYTHIAN PROVIDES INVALUABLE BACKUP & DISASTER RECOVERY (BDR) CONSULTING FOR A COMPANY WITH NO BACKUP OR RECOVERY CAPABILITIES

BUSINESS NEED

A large software-as-a-service (SaaS) company's on-premise Hadoop cluster was used primarily to collect and store tens of thousands of online audio files generated by their proprietary software, along with the associated metadata. However, the company's massive, 215-plus terabyte Apache HBase database was so big it wasn't able to support the necessary snapshots long enough to perform Hadoop cluster backups. Data loss was a certainty if their Hadoop Distributed File System (HDFS) failed or became corrupted. There was also no disaster recovery cluster, so if the existing cluster was lost it would need to be built from scratch – a process that would likely take several weeks if not months. And although the company had previously attempted to fix the problem using Commvault, there were simply too many cluster performance issues for that initiative to succeed.

SOLUTION

Thanks to our experience providing expert backup and disaster recovery (BDR) advice, the company turned to Pythian for help navigating its on-premise or public cloud options. They wanted a robust and cost-effective BDR solution able to meet a performance benchmark of 12 hours or less to recover the last 12 months.

To achieve this while continuing to use their current on-prem cluster, Pythian proposed a new Hadoop cluster be created in another data center and fed from the backups performed on the existing cluster, or through a cluster-to-cluster-sync process using CDH enterprise or MapR. Pythian also advised the client, however, that this route would require an expensive capital investment likely to exceed their desired budget. A few other on-prem options were considered -- Cloudera BDR (which

requires a Cloudera Enterprise license) and WANdisco Fusion being the first two, but both were similarly deemed too expensive – along with the possibility of using Swift, Ceph, MariaDB/AX or ClickHouse open-source object storage along with the current Hadoop cluster.

Pythian also outlined several cloud-native options for solving the client's BDR dilemma, with the two most compelling and cost-effective being:

- **Google Cloud BigTable combined with BigQuery**, since the former supports magnetic storage and includes automatic multi-region backups to cloud storage.
- **A cloud object storage/column store** approach using either AWS Simple Storage Service (S3) or Google Cloud Storage, which would automatically support multi-region redundancy for backup and disaster recovery, while also being the most cost-effective option.

RESULT

Ultimately Pythian's recommendation was to go with the cloud object storage/column store option, which would provide the required BDR capabilities for the most affordable cost – a finding the client's IT leadership strongly agreed with but needed to verify via a third-party expert. Ultimately, the client came away from its Pythian engagement with the confidence that comes with knowing you've looked at all possible options through an expert lens.

ABOUT PYTHIAN

Pythian excels at helping businesses around the world use data and the cloud to transform how they compete and win in the data economy. From cloud automation to machine learning, Pythian leads the industry with proven innovative technologies and deep data expertise. For more than 20 years Pythian has built its reputation by delivering solutions to the toughest data challenges faster and better than anyone else.

WORLDWIDE OFFICES

Ottawa, Canada

New York City, USA

London, England

Hyderabad, India

© Pythian Services Inc. 2020