

INDUSTRY

Manufacturing

BUSINESS NEED

The company needed to replace its Netezza data warehouse in order to avoid a costly on-premises upgrade to continue having it supported. Their data warehouse was critical for the demand forecasting they needed to optimize manufacturing operations. In addition, they wanted a flexible platform to allow new use cases down the road.

SOLUTION

The company engaged Pythian for a phased approach to migrating their data warehouse from Netezza to GCP that involved a proof of concept, productionized proof of concept and a final migration of their analytics workloads to Google Cloud.

TECHNOLOGIES

Netezza, Oracle, Google Cloud Platform, Google BigQuery, Google Cloud Composer, Apache Airflow, SQL

RESULT

The client is now able to run efficient and cost-effective queries in its new BigQuery data warehouse without sacrificing performance.

PYTHIAN MIGRATES MAJOR LOGISTICS COMPANY FROM NETEZZA DATA WAREHOUSE TO GOOGLE CLOUD

A global end-to-end manufacturer of innovative optical, electronic and mechanical products wanted to migrate its complex and robust data warehouse from on-premises Netezza to Google BigQuery, and needed to do it quickly. Pythian's encyclopedic knowledge of Google Cloud Platform (GCP) and years of data migration experience meant they were the logical choice to help shepherd the client through what could have been an extremely time-consuming and sensitive process.

BUSINESS NEED

The company's data warehouse was being used to support complex queries to predict demand based on a range of factors, which were then used to align its manufacturing operations with that demand. They wanted to continue running these queries while migrating to a robust, yet cost-effective and flexible, alternative to Netezza. In addition, they wanted to enable potential new use cases for their data for future enhancements to their demand forecasting capabilities.

In addition, one of the project's major drivers was cost reduction: the client was tired of paying heavy licensing, hardware and maintenance costs to run its on-premises Netezza data warehouse. But that wasn't the only push factor: since IBM had previously announced the sunset of older versions of Netezza by June 2019, the client was staring at a forced migration to Netezza's newest iteration. The company also realized that Netezza's support offering was at its end-of-life, and therefore set a hard deadline to execute a full migration of their Netezza-based solution to a newer, more flexible cloud system.

In their new platform, they needed scalability and flexibility of the cloud while keeping the same or similar high performance and cost structure to which they were accustomed. The client's large, on-prem data

warehouse featured multiple, complex data areas including financial data requiring extreme precision in terms of decimal places.

Because the company uses MicroStrategy as a BI/reporting tool, the new solution also had to work seamlessly with that software. The client is highly regulated and required Pythian to provide compliance-related documentation throughout and at the conclusion of the development process.

SOLUTION

Because of the scale and sensitivity of the project, it was divided into three phases. The first was an initial proof-of-concept to begin exploring whether BigQuery met the client's needs, while the second phase entailed the productionization of the proof-of-concept to further evaluate and fine-tune the platform. The third and final phase was the complete migration of the client's data warehouse from Netezza to GCP. Pythian DevOps consultants were also brought in to develop and configure the environments to provide flexibility through CI/CD capabilities.

To maintain the client's preferred cost structure and performance standards while taking advantage of the flexibility and scalability of the cloud, Pythian expertise was required on several fronts. The first was to advise the client's data users on querying the system in a way that would minimize cost while maintaining top performance—something of a culture shift at the organization. Because running queries through their on-prem Netezza platform had previously entailed no cost other than electricity, the client's data users were accustomed to running virtually any query that came to mind, at any time—something that can quickly become cost-prohibitive on a pay-per-read cloud system. So Pythian provided advice and best practices on preset limits on GCP and query construction and frequency to get the best value, without spending money on needless queries. Secondly, Pythian worked with the client's IT team to both evaluate and expertly optimize GCP to reduce the number of records read per query, thus further improving performance and cost. Pythian then used Cloud Dataflow to redesign the many

data transformations that had previously occurred in the client's on-prem Netezza data warehouse, automating these transformations via a scheduler to ensure data flowed automatically from source to target.

Lastly, because much of the project involved evaluating the suitability of GCP for the client's requirements, Pythian built a state-of-the-art audit and data quality solution to gauge the platform's performance in real time. It provided ongoing and robust auditing of the full data pipeline, using notifications to alert the appropriate stakeholders of data failure at any point, along with providing information about that failure. The solution also included a metadata repository, allowing ongoing performance monitoring.

RESULT

The client is now able to run efficient and cost-effective queries in its new GCP BigQuery data warehouse without sacrificing performance or paying unnecessary costs associated with reads from their data warehouse. They can also keep track of their data pipelines while ensuring top-shelf data quality through Pythian's custom-built auditing solution, along with running accurate and real-time reports through MicroStrategy BI software. And although this client currently doesn't perform advanced analytics requiring data scientists, members of its C-suite now have the option to throw the gamut of GCP's ecosystem—including a data lake and machine learning tools – at any problem they come up against. Pythian was also able to provide documentation to satisfy the client's regulatory obligations throughout the project, along with complete documentation at the project's end.

TECHNOLOGIES

Netezza, Oracle, Google Cloud Platform, Google Cloud Composer, Apache Airflow, SQL

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

© The Pythian Group Inc., 2019