Hadoop migration: Improving performance and minimizing licensing costs by enhancing trade validation process

**Client**
The largest independent securities regulator in the United States

**Industry**
Capital markets

**Business need addressed**
Developed scalable, cost-effective technology to streamline the trade validation process, address growing volume, and store consolidated audit trail data

**Genpact solution**
Developed a multi-phase approach to port the current application to a Hadoop-based platform, which resolved capacity problems, improved performance, and saved on licensing costs

**Business impact**
With a newer, more scalable solution, the client can handle larger volumes while complying with stringent timelines to maintain business as usual
### Business challenge

This client relies on its audit trail application as a source of timed, sequenced-order events that contain market quotation and trading information submitted by member firms. This key application is used to verify and track member firms’ submission, reporting, and processing statistics for compliance, repair, and resubmission.

The existing technology—a platform built on a Netezza DW/BI appliance and the DataStage ETL tool—was expensive, with limited scalability options, and adversely impacted trade validation with ever-increasing volumes (consistently growing by about 45% year on year). There were frequent misses in SLA due to this growing volume—and application volume was expected to grow five- to six-fold over the next two years as additional asset classes were included. The client also needed to create a central repository to receive and store consolidated audit trail data in order for regulators to view cross-market data.

### Genpact approach

At the beginning of the engagement, Genpact performed a thorough root cause analysis, using Lean and Six Sigma principles. This confirmed the need for a scalable, cost-effective platform to address rising data volumes and licensing costs.

In collaboration with the client’s developers, Genpact engaged in a proof-of-concept and evaluated new designs based on Hive, NoSQL (HBase), and in-memory-based custom Map/Reduce jobs. Based on the findings, Genpact and the developers created a multi-phased solution to migrate the trade validation process to Hadoop.

### Genpact solution

To address the client’s business needs, Genpact developed a multi-phased approach to port the client’s existing Netezza/DataStage application to a Hadoop-based platform:

**Phase 1:** Proof-of-concept to evaluate designs based on Hadoop, HDFS, Hive, and HBase

**Phase 2:** Migrate from current ETL platform (DataStage) to a core Java-based ETL solution

**Phase 3:** Migrate from Netezza to an in-memory-based custom Map/Reduce Hadoop solution, which includes Hive and HBase

This solution enabled the client to move member firms’ trade data from the proprietary NAS system to Hadoop HDFS, along with the required reference data. The data validation engine is fully customizable, and rules are created using an XML-based template that runs on a custom-built Map/Reduce framework, as shown in the below chart.
Business impact

Implementation of Genpact’s trade validation process migration to Hadoop had an immediate positive impact:

• Resolution of capacity problems, with performance improvement over 10x
• Lower license costs for Netezza and DataStage by using more affordable Hadoop platform.
• Affordable scalability potential for expected long-term audit trail data volume growth of 10–15 billion transactions per day
• Long-term, cost-effective Tier 3 storage platform using Hadoop HDFS
• Ability to port the new Hadoop-based solution to cloud-based systems like AWS without design changes

About Genpact

Genpact Limited (NYSE: G) is a global leader in designing, transforming and running business processes and operations, including those that are complex and industry-specific. Our mission is to help clients become more competitive by making their enterprises more intelligent through becoming more adaptive, innovative, globally effective and connected to their own clients. Genpact stands for Generating Impact – visible in tighter cost management as well as better management of risk, regulations and growth for hundreds of long-term clients including more than 100 of the Fortune Global 500. Our approach is distinctive – we offer an unbiased, agile combination of smarter processes, crystallized in our Smart Enterprise Processes (SEP℠) proprietary framework, along with analytics and technology, which limits upfront investments and enhances future adaptability. We have global critical mass – over 65,000 employees in 25 countries with key management and corporate offices in New York City – while remaining flexible and collaborative, and a management team that drives client partnerships personally. Our history is unique – behind our single-minded passion for process and operational excellence is the Lean and Six Sigma heritage of a former General Electric division that has served GE businesses for more than 16 years.


Follow us on Twitter, Facebook, and LinkedIn.

© 2014 Copyright Genpact. All Rights Reserved.