Success Story

Big Data Processing for Real-time Consumer Engagement

Business:
Big Data

Domain:
Telecommunications

Tools:
Hadoop Ecosystem, Oozie, Solr, Sqoop, Hive, HDFS, Hbase, Phoenix

Key Highlights

Key Success:
» ETL nightly process duration reduced from 11 hours to 2 hours.
» The Big Data Hadoop architecture implemented could handle 500 million messages per day.

Engagement:
Offshore engagement with:
» 1 Senior Big Data Architect
» 2 Big Data Engineers

Client
The client is one of those pioneers who has developed a Mobile Engagement Platform that enables enterprises to drive their marketing outreach through mobile messaging technology.

Overview
Due to the client's rapid growth in recent years, the volume (and characteristics) of data it collects and the analytics it must perform have increased exponentially. The original big data infrastructure was being pushed beyond what it was intended to support and consequently errors were being thrown up, time taken for basic analytics work shot up, maintenance costs were spiraling out of control etc. Indium Software was charged with upgrading the client's infrastructure to guarantee: the lowest latency possible, superior responsiveness, seemless integration and cost effectiveness.

1 Status Quo

The client is one of those pioneers who has developed a Mobile Engagement Platform that enables enterprises to drive their marketing outreach through mobile messaging technology. They are an aggregator in the US with direct connectivity to all major wireless carriers with a best-in-class campaign management platform.

The existing architecture of the client used:

MySQL (RDBMS) Server for storing messages. This limited insertion rate leading to an IO bottle-neck. The system needed a Horizontal scale up i.e. adding more hardware and also configure time sensitive features to meet marketing SLAs. In spite of handling key functions: Disaster Recovery, Back up and Reporting, the data container was stagnant beyond 100 million capacity leading to process inefficiencies.

Pentaho Server for ETL Process requiring heavy IO and computation. ETL process took an average of 11 hours and observed spikes up to 15 hours during promotional days/ special campaigns. The system significantly slowed down the output productivity and efficiency.

PostgreSQL Server for Reporting and walling Aggregated Logs. Business users were unable to access the reporting data in real time. Aggregator logs contained only recent 2 month data for reporting; any data requests beyond that would need special data access requests that would typically take 24-48 hours to retrieve.

2 Business Requirements

Indium Software analyzed the platform from functional and operational perspectives and defined the following goals to be met to make the current system a robust one:

- Lowest Latency
- Superior responsiveness
- Powerful Integration
- Cost effective architecture

3 Indium Software's Approach And Implementation

To achieve the above, Indium Software proposed a phased approach to implement a potential Big Data Processing Solution.

Enhance the ETL Process Solution: Enforcing a centralized data system

The ETL function streamlines aggregated data to MySQL and logs to PostgreSQL. These batch jobs were run on an expansive Pentaho architecture. We migrated the set up to a powerful Hadoop ecosystem. Data is centralized in the Big Data containers offering high flexibility, highly scalability, fault tolerance benefits and cost advantage.

Merge log database into Big Data ecosystem

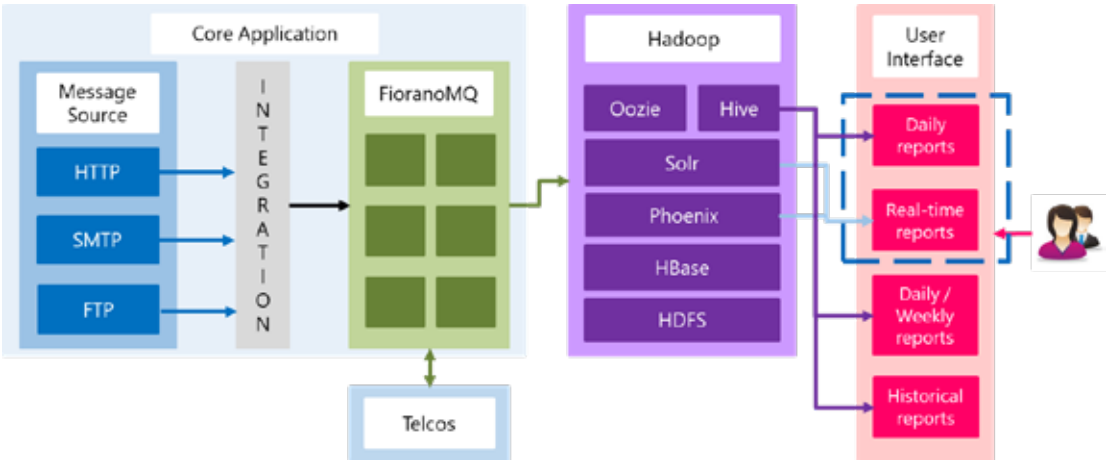
Log insertions into PostgreSQL were tedious, taking up about 4-5 man hours. The logs were also stringent on access to end users for viewing the daily/weekly and historic reports. Indium Software's hindsight on this pain point resulted in migrating the log database into the Hadoop clusters. This enabled the system with distributed processing and thus minimized effort. The data required for real time reporting was easily generated using Hive tables.

Remove the Data Insertion bottleneck by introducing direct inserts into Hbase

There were significant delays in data insertions into the SQL database. The direct consequence of this bottleneck was that the downstream processes were affected i.e. the business users were unable to view real time message delivery status. Indium Software's architects introduced HBase on top of HDFS and Phoenix (SQL on Hadoop). This loaded real-time data into HBase and made accessing reports in real-time effortless.

4 Business Impact

» Post implementation, ETL nightly process duration reduced from 11 hours to 2 hours.
» The Users were able to access daily reports each day, early in the morning.
» The updated technology and approach to migrate the legacy architecture reduced the infrastructure costs.
» Time taken to generate Historical Reports was reduced to just a few minutes.
» The Big Data Hadoop architecture implemented can handle 500 million messages per day.
» The solution enabled all the data to be brought into one single platform while maintaining high scalability.



Architecture Diagram

INDIUM
Digital | QA | Gaming

INDIA
Chennai
+91 44 6606 9100
Bengaluru
+91 80 4645 7777
Mumbai
+91 022 6215 4028

USA
Cupertino | Princeton
Toll-free: 1 888 207 5969

SINGAPORE
+65 9630 7959

UK
London
+44 773 653 9098

General Inquiries : info@indiumsoftware.com
Sales Inquiries : sales@indiumsoftware.com