



Case Study: Retail Chain

Introduction

The multiplication of retail channels and the increasing use of social media are empowering consumers. With a wealth of information readily available online, consumers are now better able to compare products, services and prices—even as they shop in physical stores. When consumers interact with companies publically through social media, they have greater power to influence other customers or damage a brand.

In order for retailers to capitalize on these and other changes in the industry, they need ways to collect, manage and analyze a tremendous volume, velocity and variety of data.

Hence one of the Big multi-store retail chains wanted to generate valuable insights for personalizing marketing and improving the effectiveness of marketing campaigns, optimizing assortment and merchandising decisions, and removing inefficiencies in distribution and operations. This is where Corpus got in to address these challenges with Big Data Analytics.

Requirement

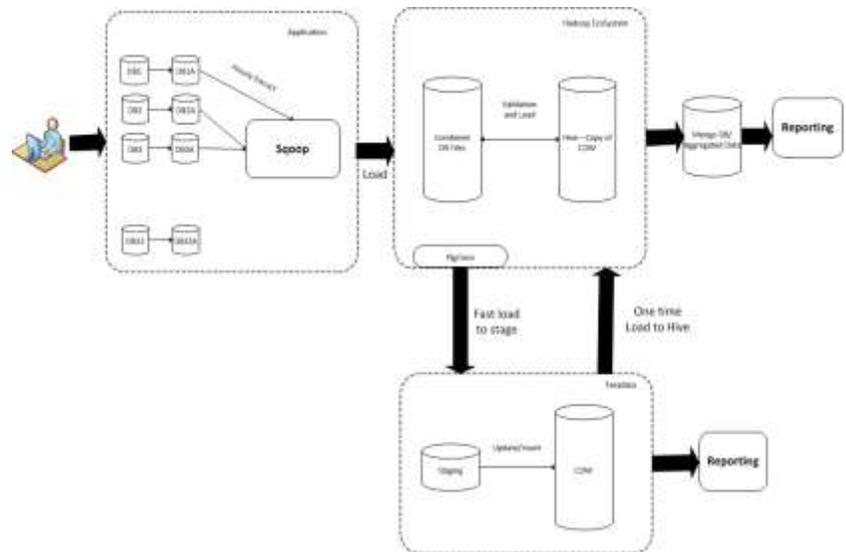
- Data Volume: Struggling to manage and extract value from the growing volume and variety of data and need to unify information across federated sources. Each disparate MySQL server records around 2 million updates a total of 50-60 Million updates daily
- Unable to relate “raw” data collected from system logs, sensors, or click streams with customer and line-of-business data managed in enterprise systems.
- Speed: Each end-to-end ETL process consuming 14 hours. Wanting this process to get complete within an hour
- Real-time visibility into business operations including customer experience and behavior.
- Bottle neck: Most of the time spent on the transformations with in data warehouse staging

Solution

- Replaced Java program ETL with Hadoop for data transformation from MySQL to Data Warehouse using Apache Sqoop
- Replaced Multi-staged data warehousing system with a common centralized warehouse
- Apache Pig was used to transform data that is in HDFS (Hadoop Distributed File System). Apache Pig is a platform, for analyzing large data sets, that consists of a high-

level language for expressing data analysis programs, coupled with infrastructure for evaluating these programs and with substantial parallelization.

- Integrated 'Hive' with the existing data warehousing system for faster batch process; which facilitates easy data summarization, ad-hoc queries, and the analysis of large datasets stored in Hadoop compatible file systems. Hive provides a mechanism to project structure onto this data and query the data using a SQL-like language called HiveQL. At the same time this language also allows traditional map/reduce programmers to plug in their custom mappers and reducers when it is inconvenient or inefficient to express this logic in HiveQL.



- Achieved under 1 hour ELT process to meet the Retail chain's Data analytics needs.
- Planned to implement MongoDB in next phase to achieve near real-time analytics.

Benefits:

- Predict optimal pricing and maintain a price leadership position by analyzing price and demand elasticity
- Select the right merchandise for each channel and fine-tune local assortment planning by drawing on insights from social media, market reports, internal sales data and customer buying patterns
- Optimize inventory across multiple channels by using leading indicators such as customer sentiment and promotional buzz to anticipate future demand
- Reduced costs and improved operational performance for the business
- Improved decision making and greater efficiencies in business processes
- Gained new insights by combining and analyzing data types in different ways
- Improved customer engagement

About Corpus: Corpus Software is one of the faster growing IT solution and services company focused on Digital Media Entertainment, Embedded technology and Business Analytics with offices and partners across Americas, Europe, APAC, Middle east & Africa. We work with clients in most emerging technology, that's where we make their business strong and bring in real difference in the way peer operates. A diverse workplace with continues focused towards developing unique ideas and contributions to make our clients business grow, and to keep the momentum going.

Offices: Dallas, London, Singapore, Johannesburg, Hyderabad and Bangalore

