

Implementing a Centralized, Worldwide Customer Account Master

Case Study | **Data Management**

LARGE U.S. RETAILER

THE PROBLEM

The client wanted to identify customers who interact on their online portal but shop in-store and vice-versa.

The client also wanted to identify changes in customer preferences and buying patterns, and their correlation with various dependent factors such as Demography, Weather, Rewards, etc.

The business wanted to build a Data Lake to store unstructured and structured data from various channels to commence advance analytics on the Data Lake.

THE SOLUTION

Designed and engineered a managed Data Lake to perform advanced analytics and generate business insights.

The Data Lake was designed to ensure scalability so that any future data sources to be added, regardless of the type of data source, can be easily achieved.

Mastech InfoTrellis devised a strategic roadmap for the client that will serve as a guideline for effective future implementation of Big Data programs.

Recommendations for data governance and associated tools usage were developed.

Hands-on training on Datameer was provided to the client's IT team for enabling them to perform ad-hoc analytics on their Data Lake post go-live.

THE OUTCOME

Successfully built a Managed Data Lake that can hold multi-channel data of various formats and serve as the foundation for all analytic and reporting activities.

Ingested more than 100 TB (terabytes) of data from over 40 different sources. Technology stack included Hortonworks, Oracle DB, Spark, Sqoop, Hbase, Groovy, and Datameer.

Implemented 50+ Datameer jobs for data preparation and analytics to lay the foundation for self-serve analytics by the business.