

HADOOP IMPLEMENTATION FOR A LARGE ONLINE TRAVEL INDUSTRY CUSTOMER



BUSINESS OBJECTIVES

Customer wanted to reduce the total cost ownership for the data warehouse environment by migrating to open-source Hadoop Big Data architecture.

SOLUTION

Architected the Hadoop infrastructure, capacity, security and data strategy. Configured Hive, Scoop and other administration tools. Each existing ETL stored procedure touching the raw LZ Omniture tables was converted to a series of Map-Reduce jobs. Raw Omniture files were stored in Hadoop HDFS in a directory structure organized by TPID, Year, Month and Day – to enable processing / re-processing for different periods of time. The output of Omniture ETL Workflows was exported from Hadoop back into existing DB2 ADS and DataMarts. An export utility was built for R1.0 to bring selective datasets from Hadoop into DB2 scratchpad for ad hoc analysis and for advanced analytics using traditional SQL methods.

SUCCESS CRITERIA & BUSINESS VALUE

The infometry team worked closely with an Online travel industry company in helping them to solve a Big Data Analytics problem, which involved designing and architecting the Hadoop/MapReduce solution to replace their existing DB2-based data warehouse platform, which resulted in the savings of 8.5 million over 2.5 yrs and also enable the customer to analyze extensive volume data leveraging Hive/Hue.

CONTACT US

Infometry INC

+1 (510) 770 6400

+1 (510) 793 5859

39111 Paseo Padre Pkwy, Suite 203E,
Fremont, CA 94538, USA

TECHNOLOGIES USED

Hadoop, ETL, Hive/Hue