E-COMMERCE ANALYTICS IN THE CLOUD

BUSINESS OBJECTIVES

The eCommerce platform was open to the social networking of merchants in the cloud. This resulted in a phenomenal increase in the structured and unstructured data to over 10TB per week. Thus an inexpensive solution was to be architected to avoid performing entire business analytics on Oracle and expensive storage like EMC. Unstructured data was also to be mined for business intelligence using SAS.

SOLUTION

- Architected the data warehouse staging on Hadoop/MapReduce using 60 nodes of inexpensive Linux servers to process over 10 TB of unstructured data weekly.
- Used Hive to provide batch reports, thereby reducing the load on the Oracle DW production instance.
- Transformation and Cleaning services were designed to be invoked before the data was stored in Hadoop/MapReduce. The unstructured data source was the merchant discovery and forums and occupied over 90% of Hadoop / MapReduce, but the intelligence from it was occupying less than 20% of Oracle DW storage and less than 5% of computing resources. The hybrid solution utilized Oracle DW and Java / XML framework of BIRT Actuate to provide ad-hoc reporting and drill down in analytical cubes.

SUCCESS CRITERIA & BUSINESS VALUE

TCO was reduced by 30% by providing a hybrid solution using Hadoop / MapReduce. The unstructured data was mined and only the intelligent data was captured in the Oracle DW. The solution was successfully implemented for half million merchants in the eCommerce Cloud.

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. Hive