Daily Data Warehousing helps a Leading American Media Company Localize Content & Ads

DW process provides daily analytics on page views and page impressions on the various websites of the company.

Overview
A leading US-based media company needed daily analytics reports on articles and advertisements published on their online properties to identify those that were driving more traffic. Tavant developed a daily data warehousing process that enabled the customer to get access to important analytics data on page views and page impressions for articles and advertisements across all their news portals.

Customer
The customer is a leading US-based diversified media company that owns a number of online news properties. The company's principal operations include educational services, newspaper print, online publishing, television broadcasting, and cable television systems.

Business Context
The customer required daily analytics reports on articles and ads that appeared on their various news websites and portals to enable personalization and localization of content. The customer was using Omniture and DoubleClick to fetch data on article page views and ad impressions. However, the data thus retrieved was unstructured, non-relational, and in the form of a compressed plain flat file. Our customer wanted to simplify the data analytics process and also reduce the volume of data by denormalizing and storing it in a data warehouse.
Solution

Tavant developed a daily data warehousing process – ETL (Extraction, Transformation, and Loading) - using the Hadoop - Hive framework on Amazon Web Services using Amazon Elastic Map Reduce (EMR) and Amazon Elastic Cloud Compute (EC2).

We used the following steps for daily data processing:

1. Data for all online properties was individually fetched by Omniture and DoubleClick.

2. The data was subjected to the ETL process involving data dimensioning and data crunching.

3. The denormalized data was stored in the data warehouse using Amazon Simple Storage Service (S3). The data warehouse stored multiple terabytes of data in columnar database sourced from Amazon S3.

4. Various tools such as Tableau and Pentaho interfaced with the data warehouse to generate daily combinational and individual analytics report for articles and ads.

The data thus retrieved helped the customer identify which ads and articles were popular in which regions, and accordingly, personalize and/or localize them. The data warehouse offered an accuracy of 99% in data count for page views and page impressions. Tavant also provided maintenance and feature enhancements of the daily data warehousing process.

Tools & Technologies

- Hadoop - Hive framework
- Amazon Web Services: Amazon S3, Amazon EC2, and Amazon EMR
- Perl & Shell scripting
- Tableau

Business Benefits

- Processing of a large amount of data on cloud.
- Scalable framework to accommodate high data volume.
- Reduced costs.

About Tavant

Tavant Technologies is a specialized IT solutions and services provider of choice for customers across the world. We leverage our unrivaled capabilities and domain insights to provide impactful, game changing results for our customers every time, without exception. We are known for our engineering excellence, passionate employees, and long-lasting customer relationships. Founded in 2000, we are headquartered in Santa Clara, California and service customers across North America, Europe, and Asia-Pacific.