Abstract

By the improvement and expansion of the internet, social media, internet of things and advanced technology in the fields of healthcare, infrastructure, Agriculture, Education, Scientific fields and in Data Analytics, data generation growth augmented exponentially. In the world of exploding data, storage and speed become the burning issues. To manage, process and to analyze Big Data both academia and industry work together for cost effective solutions. Big Data harmonization is the process of providing a single platform to all heterogeneous data and variety of data. Extraction, transformation and loading is the essential step in the process of data warehouse. Data harmonization is the alternate name for the data warehouse to provide the common level of granularity. Computing of Big Data OLAP requires lot of challenges like scaling of data, speed of processing, storage of data, query performance and lot of others. Author has proposed and implemented a technique name as OOHI (OLAP on Hadoop by Indexing) offer a simplified and efficient multidimensional model. Overall work of OOHI is divided into Data Loading Module, Data Storage Module, Dimension Encoding Module, Dimension traversal Module, Cube Segmentation Module and Block Selection and Process Module. Map reduce based data loading and Block selection worked to achieve parallelism and fault tolerance. Real time oceanography data and super market data is used to compare OOHI model with existing platforms and to show OOHI is not data dependent. Design of test case series helped to compare results with existing techniques like HaOLAP and Olap4cloud. OOHI achieved good performance in data loading and storing with Olap4cloud. This work is useful in big data warehousing domain to speed up the query performance, reduce data loading and processing time.

List of Publications