Abstract

Big Data technologies are becoming a current talk and a new "buzz-word" both in science and in industry. Today data have grown from terabytes to petabytes and now it is in zeta bytes. Increased amount of information increases the challenges in managing and manipulating data. Data integration is a main issue in large data sets which is managed by Extract, Transform and Load (ETL) tools such as Data Warehouses. Data Warehouse is the process of transforming all multiple data formats into a single format and consolidating them in one place. Now days, data generated from social networks, web server logs, sensors used to gather climate information, stock market data, e-mails, transaction records, web click streams, etc. Most of these data are in unstructured or semi structured forms. Today organizations' are trying to find new solutions such as ETLs to manage the situation. The
existing data warehousing tools and techniques were inefficient to handle unstructured and semi structured data. This paper presents the issues and challenges of data integration in Big Data environment and techniques for big data integration. A new ETL framework is proposed open problems for future research of data integration are identified in big data environment.

References

- Steven J. Rysavy, Dennis Bromley, and Valerie Daggett, 2014. "DIVE: A Graph-Based Visual- Analytics Framework for Big Data"; March/April 2014 Published by the IEEE.

Index Terms

Computer Science

Database Management

Keywords

Big Data Integration Data Warehouse Hadoop.