Chapter 3
Big Data Computing and the Reference Architecture

M. Baby Nirmala
Holy Cross College, India

Pethuru Raj
IBM India Pvt Ltd, India

ABSTRACT

Earlier, the transactional and operational data were maintained in tables and stored in relational databases. They have formal structures and schemas. However, the recent production and flow of multi-structured data has inspired many to ponder about the new ways and means of capturing, collecting, and stocking. E-mails, PDF files, social blogs, musings, tweets, still photographs, videos, office documents, phone call records, sensor readings, medical electronics, smart grids, avionics data, real-time chats, and other varieties of data play a greater role in presenting highly accurate and actionable, timely insights for executives and decision-makers. The chapter provides an insight into the big data phenomenon, its usability and utility for businesses, the latest developments in this impactful concept, and the reference architecture.

INTRODUCTION

Big data is the "Heart of the talk" in this current era. All Big people think of this Big data and talk about this Big data. Earlier the transactional data were maintained as Tables and stored in relational Databases and Files. All other unstructured data were maintained for few years and then thrown out. There is a lot of potential value in these kinds of non-traditional and less structured data like E-mail, Social media, Weblogs, Photographs, Videos, Power points and Phone calls; Chats play a greater role in Business Intelligence analysis of the Enterprise data. Between now and 2020, the amount of information in the digital universe will grow by an unimaginable 35 trillion gigabytes as all major forms of media- Voice, TV, Radio, Print complete the journey from analog to digital (IDC, Sponsored by EMC[2], 2012).

BIG DATA ANALYTICS

What is Big Data Analytics?

Big Data Analytics is the process of examining large amounts of data of a variety of types (Big data) to uncover hidden patterns, unknown correlations and other useful information. In other words, Big data Analytics is the use of advanced analytical techniques against very large diverse data sets that includes different types such as Structured/Unstructured and Streaming/Batch and different sizes from terabytes to zettabytes.

Figure 1 shows how a Big data processing is done. By facilitating data scientists and other users to analyze huge volumes of transactional data as well as data from other sources which are left untapped by conventional Business Intelligence(BI) Programs, Big data analytics help the organizations to make better business decisions.

These other data sources may include Web server logs and Internet click stream data, Social media activity reports, Mobile-phone call detail records and information captured by the sensors.

Some people exclusively associate Big Data and Big data Analytics with unstructured data. Consulting firms like Gartner Inc. and Forrester Research Inc., consider transactions and structured data to be valid forms of Big data.

Big Data Analytics can be done with the software tools commonly used as part of advanced analytics discipline such as Predictive Analytics and Data Mining.

Three Key Technologies for Extracting Business Value from Big Data

- **Information Management:** Manage data as a strategic, core asset, with ongoing process control for Big data analytics.
- **High-Performance Analytics:** Gain rapid insights from Big Data and the ability to solve increasingly complex problems using more data.

Figure 1. Big data processing
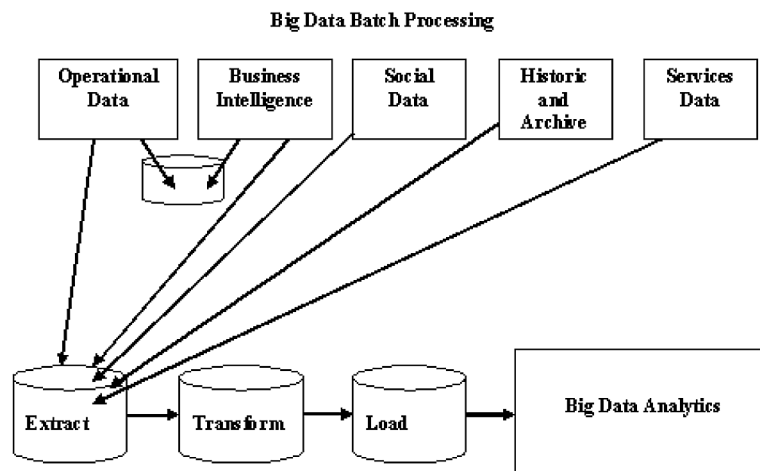
57

Related Content

Efficient and Effective Aggregate Keyword Search on Relational Databases

Luping Li, Stephen Petschulat, Guanting Tang, Jian Peiand Wo-Shun Luk (2012). *International Journal of Data Warehousing and Mining (pp. 41-81).*

www.irma-international.org/article/efficient-effective-aggregate-keyword-search/74755

DMA: Matrix Based Dynamic Itemset Mining Algorithm

Damla Oguz, Baris Yildizand Belgin Ergenc (2013). *International Journal of Data Warehousing and Mining (pp. 62-75).*

www.irma-international.org/article/dma/105120

Optimization of a Hybrid Methodology (CRISP-DM)

José Navaand Paula Hernández (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications (pp. 1998-2020).*

www.irma-international.org/chapter/optimization-hybrid-methodology-crisp/73532

Mining Association Rules in Data Warehouses

Haorianto Cokrowijoyo Tjioeand David Taniar (2005). *International Journal of Data Warehousing and Mining (pp. 28-62).*

www.irma-international.org/article/mining-association-rules-data-warehouses/1755

Data Intensive Cloud Computing: Issues and Challenges

Jayalakshmi D. S., R. Srinivasanand K. G. Srinivasa (2016). *Big Data: Concepts, Methodologies, Tools, and Applications (pp. 639-654).*

www.irma-international.org/chapter/data-intensive-cloud-computing/150186