

Chapter 46

A Survey of Big Data Analytics Systems: Appliances, Platforms, and Frameworks

M. Baby Nirmala
Holy Cross College, India

ABSTRACT

In this emerging era of analytics 3.0, where big data is the heart of talk in all sectors, achieving and extracting the full potential from this vast data is accomplished by many vendors through their new generation analytical processing systems. This chapter deals with a brief introduction of the categories of analytical processing system, followed by some prominent analytical platforms, appliances, frameworks, engines, fabrics, solutions, tools, and products of the big data vendors. Finally, it deals with big data analytics in the network, its security, WAN optimization tools, and techniques for cloud-based big data analytics.

INTRODUCTION

In this technological era of big data, the important issue that arises is how such huge amount of data which is semi structured, unstructured, machine generated /or sensor data, mobile data and large scale data can be stored and processed. It is fair to say that we are now entering an era of analytics 3.0 in which analytics will be considered to be a “table stake” capability for most organizations to find out the great insights of that enormous data. There is a high-level categorization of big data platforms to store and process them in a scalable, fault tolerant and efficient manner. Data growth, particularly of unstructured data poses a special challenge as the volume and diversity of data types surpass the capabilities of older technologies such as relational databases. Organizations are investigating next generation technologies for data analytics. In this increasingly digital world, achieving the full transformative potential of big data requires not only new data analysis algorithms, but also a new generation of systems and distributed computing environments to handle the spectacular growth in the volume of data, the lack of structure and the increasing computational needs of massive-scale analytics.

DOI: 10.4018/978-1-4666-9840-6.ch046

In this increasingly digital world, achieving the full transformative potential from the use of data requires not only new *data analysis algorithms* but also *a new generation of systems* and distributed computing environments to handle the spectacular growth in the volume of data, the lack of structure for much of it and the increasing computational needs of massive-scale analytics.

This chapter covers technical details on the “categories of analytical processing system, how to effectively analyze data from the different analytical processing systems primarily with the help of white papers of many companies and organisations that were identified by IDC, Forrester and Gartner surveys during their analysis. The big data technology allows storing of bulk of data, searching meaningful data for visualization, enabling predictive analysis, thereby internalizing business process for application just by giving valuable insights of big data. By taking all these into consideration, this chapter strives to provide a glimpse of various platforms, frameworks, appliances, products and solutions offered by the leading BIG DATA ANALYTICS PROVIDERS though many of them are not dealt here because of space constrain.

CATEGORIES OF ANALYTICAL PROCESSING SYSTEMS

In his blog, Eckerson, (2013) explained, at a high-level, there are four categories of Analytical Processing Systems available in this era of Big data:

- Transactional RDBM Systems.
- Hadoop Distributions.
- NoSQL Databases.
- Analytic Platforms.

Other than these categories, there are analytical engines, frameworks, fabrics, etc., which also play a prominent role in big data analytics.

TRANSACTIONAL RELATIONAL DATABASE MANAGEMENT SYSTEMS

To make Transactional RDBM systems, more pleasant to analytical processing, most of them have been retrofitted with various types of indexes; join paths, and custom SQL bolt-ons, although they were originally designed to support transaction processing applications. There are two types of transactional RDBM systems- Enterprise and Departmental.

HADOOP DISTRIBUTIONS

Hadoop is an open source software project run within the Apache Foundation for processing data-intensive applications in a distributed environment with built-in parallelism and failover. The most important parts of Hadoop are the Hadoop Distributed File System, which stores data in files on a cluster of servers, and MapReduce, a programming framework for building parallel applications that run on HDFS (Hadoop Distributed File System)

27 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/a-survey-of-big-data-analytics-systems/150204

Related Content

Adaptation of Error Adjusted Bagging Method for Prediction

Selen Yilmaz Isikhan, Erdem Karabulut, Afshin Samadian and Saadettin Kılıçkap (2019). *International Journal of Data Warehousing and Mining* (pp. 28-45).

www.irma-international.org/article/adaptation-of-error-adjusted-bagging-method-for-prediction/228936

Automatic Item Weight Generation for Pattern Mining and its Application

Yun Sing Koh, Russel Pears and Gillian Dobbie (2013). *Developments in Data Extraction, Management, and Analysis* (pp. 187-207).

www.irma-international.org/chapter/automatic-item-weight-generation-pattern/70798

XML Similarity Detection and Measures

Sanjay Kumar Madria and Waraporn Viyanon (2012). *XML Data Mining: Models, Methods, and Applications* (pp. 53-77).

www.irma-international.org/chapter/xml-similarity-detection-measures/60904

Schema Evolution in Multiversion Data Warehouses

Waqas Ahmed, Esteban Zimányi, Alejandro A. Vaisman and Robert Wrembel (2021). *International Journal of Data Warehousing and Mining* (pp. 1-28).

www.irma-international.org/article/schema-evolution-in-multiversion-data-warehouses/290268

Weak Ties and Value of a Network in the New Internet Economy

Davide Di Fatta, Roberto Musotto, Vittorio D'Aleo, Walter Vesperi, Giacomo Morabito and Salvatore Lo Bue (2017). *Social Media Data Extraction and Content Analysis* (pp. 66-84).

www.irma-international.org/chapter/weak-ties-and-value-of-a-network-in-the-new-internet-economy/161959