Volume 4, Issue 6, June 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Big Data - Overview and Challenges

Kaiser J. Giri
Department of Computer Science
IUST, Awantipora, Pulwama, J&K, India

Towseef A. Lone
Department of Computer Science
IUST, Awantipora, Pulwama, J&K, India

Abstract: - The processing and management of data, which is growing at a record rate in today's world is the most pressing challenge faced by almost every organization. This gigantic growth of data commonly known as Big Data has received a lot of attention from researches in the last decade and has accordingly gone through a thorough research.This paper gives an overview and also highlights some technological challenges faced in handling the Big Data.

Keywords— Big data, Structured data, Unstructured data, Challenges, Exhabytes.

I. INTRODUCTION

In today's technology and knowledge driven society, data is considered to be the most critical resource of an organization. Due to mass implementation of information technology by most of the organizations for management of their operations, data is growing at an exponential rate these days. According to the 2011, IDC Digital Universe Study, 130 Exabyte of data were created and stored in 2005. The amount grew to 1,227 Exabyte in 2010 and is projected to grow at 45.2% to 7,910 Exabyte in 2015 [1]. Most of this data explosion is originating as a result of Transactional data, Machine data, and Social data and leads us into a new world of data known as Big Data. Big Data is being covered in various columns by prominent news media like New York Times [2], Economist [3, 4], NPR [5, 6]. Government agencies have also shown their concern towards the management of Big Data [7]. The premier scientific journals like Science and Nature also show their respect for the big Data by having columns on the challenge and issues of big data [8, 9].

The term big data is used for a collection of data sets so large and complex that it becomes difficult to process them, as the information comes from multiple, heterogeneous, autonomous sources with complex and evolving relationships, and keeps growing. The datasets are very big, they are measured in Exabyte—one quintillion (1 followed by 18 zeroes) bytes. With this gigantic increase of enormous global data, the term of big data is mainly used to describe enormous datasets. In Comparison to the traditional datasets, big data typically includes masses of unstructured data that need more real-time analysis. Big Data is about growing challenges that organizations face as they deal with large and fast-growing sources of data or information, assessing mixture of data. The data not only have to be just numbers; they can be pictures, videos, maps, words and phrases, and the list goes on. Examples of big data include customer reviews on commercial websites, comments on social networking websites, photos and videos posted online, electronic medical records, and bank records. There are three types of big data: Structured, Semi structured and Unstructured [10]

II. BIG DATA TYPES

Structured data: Data that can be easily categorized and analysed e.g. numbers and words. It is mainly generated because of network sensors embedded in electronic devices, Smartphone, global positioning system (GPS) devices etc. Structured data also include transaction data ,sales figures, account balances etc. This type describes data which is grouped in rows and columns within a standard database i.e. into a relational scheme. The data configuration and consistency allows it to respond to simple queries to arrive at usable information, based on an organization's parameters and operational needs.

Semi structured data: It is a form of structured data that does not conform to an explicit and fixed schema. The data is inherently self-describing and contains tags or other markers to enforce hierarchies of records and fields within the data. Examples include weblogs and social media feeds.

Unstructured data: It comprises of more complex information, such as customer reviews from commercial websites, photos and other multimedia, and comments on social networking sites. These data cannot easily be separated into categories or analysed numerically.
"Unstructured big data is the things that humans are saying," says big data consulting from vice president Tony Jewitt of Plano, Texas. "It uses natural language."[11]
This type of data consists of formats which cannot easily be indexed into relational tables for analysis or querying. Examples include images, audio and video files.

