Big Data Analytics for Business

B

Raymond Kosala
Bina Nusantara University, Indonesia

Richard Kumaradjaja
Renaissance Consulting, Indonesia

INTRODUCTION

Since the inception of the Web, Maes (1994) has been pointed out that we were faced with data or information overload. Since then the volume of data and information that are available digitally has grown to a tremendous size and today digital data is everywhere. According to International Data Corporation, the volume of digital data that existed in 2012 were estimated to be around 2.7 Zettabytes (Gens, 2011) and were expected to grow 40 percent yearly from 2012 to 2020 (Gantz & Reinsel, 2012). This steep growth in the digital data volume is caused by several factors, which are the following.

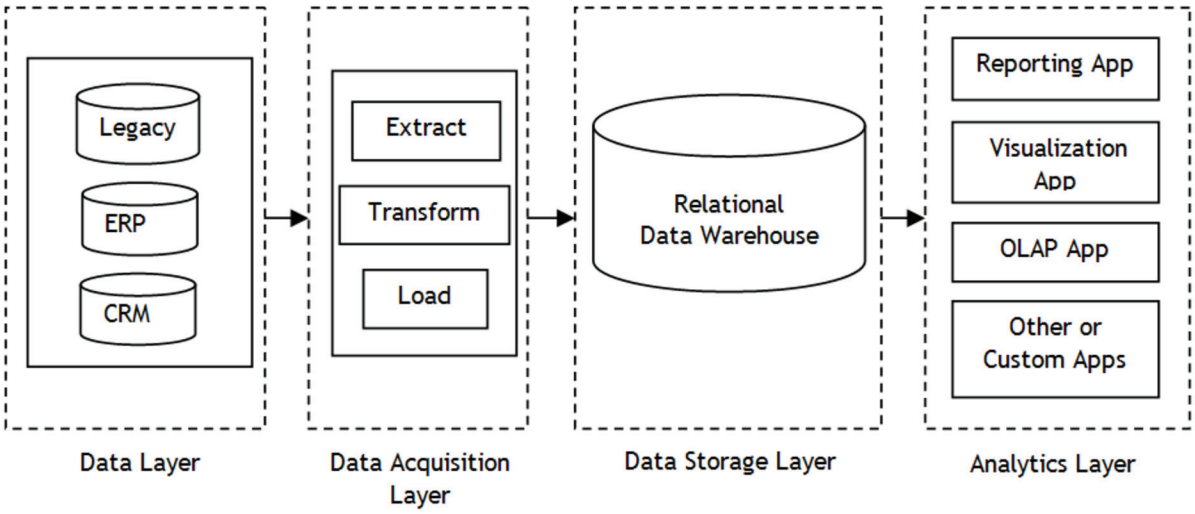
The first factor is our ability to keep increasing the data storage capacity. According to (Grochowski & Halem, 2003), the areal density capacity of hard drive had grown for around 100 percent or doubling every year since the first hard drive introduced in 1956. The second factor is the steady decrease in the data storage costs, which was decreased from 700 U.S. dollar per megabyte in 1981 to 0.002 U.S. cents per megabyte in 2010 (Smith & Williams, 2008). The third factor is the popularity of social media that increases the volume of user-generated contents and their logs in the Internet, which is also known as Web 2.0 (O'Reilly, 2005). The fourth factor is the ever increase volume of machine-generated data streams, such as scientific experiments data, sensor networks data, video surveillance data, medical imaging data, RFID data in the supply chain processes, and data generated from the Internet of Things (Gershenfeld, Krikorian, & Cohen, 2004). In the Internet of Things, everyday devices such as lamp, alarm clock, coffee maker, and others can communicate to each others as they are connected to the Internet. This pervasive connectivity among everyday devices allows, for instance, the following scenario to happen: an alarm clock that can turn the lights on when people are awake and might turn on the coffee maker if it knows the behavior of the person in the bedroom where it is located in.

One of the goals of business organizations is to increase the value of their businesses, and the huge volume of data that can be stored digitally today presents the next frontier for business organizations to increase their values. So far business intelligence (BI), which is a broad category of applications, technologies, and processes for gathering, storing, accessing, and analyzing data to help business users make better decisions, has been instrumental in increasing the value of business organizations (Watson, 2009). However, as shown in Figure 1, many traditional business intelligence systems are relying on the traditional ETL (Extraction Transformation and Load) process that integrates data mostly from different transactional business systems such as Enterprise Resources Planning (ERP), Customer Relationship Management (CRM), etc. The traditional ETL process and the relational data warehouse that are used in many traditional BI systems have failed to capture the Big Data because the volume, velocity, and variety of Big Data exceed the traditional database storage and capacity to compute for accurate and timely decision making.

Although the huge volume of digital data that is available today is very promising to improve the value of business, according to Gantz and Reinsel (2012) very few business organizations, which is only 0.5 percent, have tapped this source for analysis. The purpose of this article is to clarify the concept of data analysis from Big Data, the issues involved with the Big Data analysis effort, to show cases where Big Data have been successful and not successful in improving values, and to propose a Big Data Analytics framework. Eventu-

DOI: 10.4018/978-1-4666-5888-2.ch034

Figure 1. Traditional business intelligence and business analytics framework



ally the goal of this article is to improve the number of successful implementation of Big Data utilization efforts by business organizations.

BACKGROUND

In this section, we would like to point out some confusion, and to clarify the concept of Big Data and Big Data Analytics and its relation to Business Intelligence and Business Analytics. For some time, the terms Big Data and Big Data Analytics has been interchangeably used in many organizations/institutions. This confusion needs to be clarified so that awareness and widespread adoption of Big Data Analytics can be achieved especially in small and medium sized companies where typically only have limited IT resources. Related to this, we will show in the next section what are the issues involved and a framework for business organizations, from small to large, to start Big Data Analytics.

First of all, Big Data refers to the size and varieties of data an organization can access, integrate, store, and analyze. Furthermore, Big Data also represents a situation where the volume, velocity, and variety of data exceed an organization's traditional database storage and capacity to compute for accurate and timely decision making (adapted from Troester, 2012). An organization's traditional database capability to access, integrate, store, and analyze the data will grow overtime. However, the Big Data that is potentially useful to be analyzed grows much faster and in fact growing exponentially (Manyika, et al., 2011). Currently, the term Big Data generally refers to a mixture of structured and unstructured data with size exceeding one Terabytes that needs to be processed daily.

Big Data Analytics, on the other hand, refers to the process, techniques, and tools used to gain insights from Big Data so that the decision making process could be optimized. By using Big Data Analytics it is possible for many organizations/institutions to enhance their Business Analytics practices, which refer to the broad use of data and quantitative analysis to make business decisions in corporations. According to Turban, Sharda and Delen (2011) and also as shown in Figure 1, Business Analytics, which is a part of Business Intelligence, consists of three categories. The first category is information and knowledge discovery that includes OLAP (Online Analytical Processing), ad-hoc queries, data mining, text mining, and Web mining. The second category is decision support systems such as automated decision support, predictive analytics, performance management, etc. The third category is visualization that includes scorecards and dashboards. As a part of Business Analytics, Big Data Analytics also empowers people in the organization to make better decisions, improve processes and achieve desired outcomes. It combines the best of data management, analytic methods, and the presentation of results—all in a closed-loop cycle for continuous learning and improvement (Davenport, 2010).