Chapter 7 and 8

[Author Name(s), First M. Last, Omit Titles and Degrees]

[Institutional Affiliation(s)]

Chapter 7 and 8

# Chapter 7

## Assignment

### Response to Question 1

A data lake is like a storage depositary and it comprises of a large amount of data in its original or raw form until the data is needed by the business. On the other hand, a traditional warehouse stores data in a logical format, that comes from various operational databases and it is needed to support business decisions and business analysis activities. The data lake utilizes a flat structure to store the data. Whenever, a query arises, the data lake is queried for all the relevant information. Zillow must use a data lake to obtain and reserve a large amount of information, in this way Zillow may not need to pay for the upfront costs for data absorption and transformation, in the housing market, as in the data warehouse. The data in the data lake takes time because it is in the raw form.

### Response to Question 2

As the name suggests, dirty data is not good for organizations, as it is flawed and erroneous. Dirty data takes many forms such as incomplete data, incorrect data, and inaccurate or inconsistent data. The removal of dirty data is virtually impossible and it is the problem of business. Dirty data deters business success and efficacy and results in losses to an organization. Since business decisions are based on the data therefore inaccurate and flawed data results in loss of revenue, and increases costs. Managers may make unnecessary and useless investments which not only leads to a loss in revenue but also makes managers accountable for the high cost. It also results in ineffective marketing activities and results in marketers taking wrong decisions. For instance, in this case, if Zillow's database contains unreliable data, thousands of investments decisions will have an impact on this dirty data.

### Response to Question 3

Source data refers to the primary source or location where the data exists. This may include transactions, invoices, time-sheets, and other online sources of data and online databases. Managers in the need for information send requests to the MIS department for the needed information. The source data of Zillow comes from a variety of sources mainly public sources. All the information comes from various state-owned websites and databases, for instance, the company can take data freely from the U.S. Census Bureau to collect data relevant to construction and housing vacancy surveys and different other surveys. In addition, data can also be obtained from the records such as a country record of sales, data from the database of the Federal Housing Finance Agency and House Price Index.

## Discussion

The logical collection of information compiled and gathered from a range of operational databases with the purpose of analysis and decision making is referred to as a data warehouse. The basic value of a data warehouse arises from combining strategic information throughout an enterprise into a solo repository enabling managers to make analysis based decisions.

# Chapter 8

## Discussion

Big Data refers to the massive volume of data, in structured and unstructured form. The four common characteristics of big data include high volume, variety of information, the velocity of the frequency of the upcoming information and the value obtained from analyzing big data. Examples of big data include petabytes of data of a public organization containing records of people and big organizations like Amazon (Chen, Chiang, & Storey, 2012).

## Assignment

### Response to Question 1

The success of a company like Ben & Jerry’s is highly dependent on the business analytical tools, in this highly saturated market. Information obtained from customers is analyzed and used to ensure that the ingredients used in each ice cream meet the required standard of quality. Utilizing Consumer complaints calls are utilized for the purpose of matching pint with the suppliers’ supplies to assess what did not meet with the quality standards of the company. In addition, with the help of analysis, they also assess which supplier's supplies are high in demand and loved by customers. The data warehouse is used by the company to track the life of each product that has been sent to shelves for selling. The business intelligence tools are not only utilized to track customer complaints but also collects all the information from different departments. In the year 2005, the company was able to track more than 12,500 consumers and obtained customers' insight into the social causes supported by the company. Based on the customers' feedback the company has been able to maintain the excellent quality of its products and services.

### Response to Question 2

Information cleansing is the process by which inconsistent, incorrect and incomplete information is discarded. Information cleansing ensures that the information of data a company possesses is free of errors and it is the most relevant information needed. This method is critical to eliminate dirty data from the repository of data. Taking the case of California Pizza Kitchen (CPK), information cleansing assisted the management of the company in weeding out the inconsistent or incorrect information. This was necessary to do before the implementation of a business intelligence tool. The management of the company utilized spreadsheets to plan and track financial information. It was not easy to track hundreds and thousands of transactions and complex calculations along with the constant variations in the spreadsheets. Owing to the inconsistent data, the financial team used to work full time for weeks to generate one report for forecasting and decision making. The company launched the BI tool, Cognos, to reduce this time period of weeks to two days and allowed the finance team to spend more time reviewing the data. This information cleaning method provided a financial team with the required information essential for decision making. It is critical to saving the time and resources of the company that ultimately leads to business process effectiveness.

# References

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4).