Chapter 14
Recognizing Threats From Unknown Real-Time Big Data System Faults

William H. Money
Baker School of Business, The Citadel, USA

Stephen J. Cohen
Microsoft, USA

ABSTRACT

Processing big data in real time creates threats to the validity of the knowledge produced. This chapter discusses problems that may occur within the real-time data and the risks to the knowledge pyramid and decisions made based upon the knowledge gleaned from the volumes of data processed in real-time environments. The authors hypothesize that not yet encountered faults may require fault handling, analytic models and an architectural framework to manage the faults and mitigate the risks of correlating or integrating otherwise uncorrelated big data and to ensure the source pedigree, quality, set integrity, freshness, and validity of the data. This chapter provides a number of examples to support the hypothesis. The objectives of the designers of these knowledge management systems must be to mitigate the faults resulting from real-time streaming processes while ensuring that variables such as synchronization, redundancy, and latency are addressed. This chapter concludes that with improved designs, real-time big data systems may continuously deliver the value of streaming big data.

DOI: 10.4018/978-1-7998-2189-2.ch014

KNOWLEDGE MANAGEMENT CHALLENGES FROM THE THIRD GENERATION PROCESSING OF REAL-TIME BIG DATA

This chapter agrees with the basic proposed reconceptualization of the data underpinning the knowledge pyramid as a better representation that documents (and records) added aspects of reality because of the inclusion of Big Data. The arguments presented in this chapter extend our understanding of the intricacy and evolution of the knowledge pyramid with an in-depth analysis of the risks of Big Data when the Big Data are incorporated into the data, information, knowledge, and wisdom hierarchy without standards that recognize the characteristics and control or mitigate potential risks associated with such data.

The new proposition and basic premise offered by Jennex and Bartczak (2013) is that Big Data represents an unpresented data expansion via massive increases in the number of different views, availability and depth of the reality images, and an increased resolution of points depicting reality. The data expansion creates additional data layers through greatly expanded transaction/operational data, and requires the addition of a lower or bottom layer of Big Data. This layer is acquired from sources such as sensors and social media to better depict the knowledge pyramid. Our agreement with this new pyramid and its conceptualization can best be described with an analogy that compares the Big Data knowledge age we are entering to the launch of the Hubble telescope over 25 years ago. The Hubble was a huge leap for astronomy because it removed Earth's atmospheric changes and light blocking distortions from our view of the stars, and employed digital cameras to construct our views of the universe (from the collected data) through visible, infrared and ultraviolet light. However, Hubble's data had initial imaging defects and the telescope received ongoing servicing with improvements for many years to make the data usable via the additions of a number of improvements such as corrective lens, cameras, gyroscopes, stabilizing arrays, and corresponding software modification to correctly account for the objective sensory improvements. This chapter correspondingly analyzes possible faults or misconceptions with the Big Data that are now being incorporated into the revised knowledge-KM pyramid. This new conceptualization theorizes that there is more knowledge than data, and illustrates that knowledge management is an organizational learning extraction of the knowledge pyramid. Our contribution is to begin to describe how the revised knowledge pyramid has been enhanced and what is still unknown. We address when Big Data, composed of data sets from sources including the Internet of Things, sensors, and social data sources, introduces risks and as yet unknown dependencies and/or possible errors because these Big Data sets may or may not have unknown or mixed provenance, historical biases, well known but unmanaged errors or well understood but detached contexts.