Foreword

The Big Data paradigm requires a new way of thinking in how data is stored, processed, and analyzed. This paradigm shift is crucially needed in all aspects of data processing, including data warehousing and business intelligence. Big Data paradigms offer researchers and practitioners opportunities to explore new techniques to address Big Data challenges.

It is often to address Big Data paradigms by addressing many different "V"s of Big Data. The main ones are Volume, Variety, Velocity, and Veracity. These will be outlined in the context of data processing, especially in business intelligence.

VOLUME

Big Data certainly has large volume of data. Data retrieval techniques that have been used to efficiently retrieve smaller volume of data might not be effective to retrieve data from Big Data. It is then natural to apply some parallelism techniques in Big Data Retrieval. Therefore, parallel operators must be specifically designed for big data retrieval, including the use of parallel indexing.

Traditionally, parallel data retrieval relies on horizontal data partitioning of relational tables. In other words, data parallelism is due to parallel I/O scans. Challenges occur in regard to parallelism of indexes, where indexes are not implemented as flat tables, but as indexing trees. Consequently, index tree partitioning is a new level of challenge in parallel data retrieval. The issue becomes more complex when the data is moving objects which is dynamic and frequently updates.

Parallelism in data warehousing and business intelligence to process big data volume is even more important, as building and processing data cubes is more complex than relational database processing. Data cubes are multi-dimensional, and this creates similar complexities to parallel indexing.

VARIETY

The data Big Data exists in a variety of data formats; not only flat relational tables. This new data format raises new challenges not only in data storage, but also data modelling and processing. In the past, there had been waves in different data formats and modelling, including object-oriented data format and modelling, object-relational, temporal databases, multimedia databases, etc. Although these waves are passing, many more new data format requirements have emerged. This is due to many raising and emerging applications, such as social networks, sensors, IoT, etc., which have become big data producers in many applications. Consequently, Big Data must accommodate this new wave of data formats. This raises challenges in both processing, as well as data design and modelling.

This new data format trends give even more impact to data warehousing and business intelligence, as business analysis must take into account the large volume of non-traditional data, especially social network data, and time-series data, both of which are highly dynamic.

VELOCITY

Data in the Big Data era is coming faster than ever. It is not about time-series data and how to handle or to process time-series data. It is even bigger than this. As data is ubiquitous, and ubiquitous data producers produce high-speed data, it is critical for the data repository be able to absorb the data as quickly as they are being produced. Data ingestion in Big Data is literally a big problem. New techniques to load and to ingest the data must be used in order to keep up with the high velocity of data being produced by data producers. Handling time-series data is more challenging than ever.

Traditionally data warehousing and business intelligence handle more static data. High velocity of data raises new requirements in how data warehousing and business intelligence operates. It is crucial for businesses to provide a faster feedback looks, and business intelligence must be able to cope with this when high velocity of data is concerned.

VERACITY

Veracity, which is often described as data quality and governance, is an important aspect in Big Data, as data not only comes from everywhere, but belongs to everyone.

Data quality traditionally is an important problem; it becomes even more important when Big Data is concerned, as there are many different data producers, as well as data ubiquitous which may not belong to one organization, and is often a public data source. Therefore, veracity is a new complexity, which does not naturally exist previously.

This book contains a rich collection of knowledge in Big Data especially in the context of business intelligence and its applications, including those in artificial intelligence, machine learning, imaging, etc. This book provides a good source of emerging techniques and applications in utilizing Big Data paradigm in business intelligence, as the title states.

Happy reading, and keep learning.

David Taniar
Monash University, Australia

David Taniar *is an Associate Professor at the Faculty of Information Technology, Monash University, Australia. His research is in the area of data management, particularly big data processing, parallel database processing, spatial and Geographical Information Systems, data warehousing and mining. He has authored two books on databases ("High Performance Parallel Database Processing and Grid Databases", Wiley 2008, and "Object-Oriented Oracle", 2006), and has received four best paper awards (from three IEEE conferences and one international journal). He is the founding editor-in-chief of three Science Citation Index Expanded (SCI-E) journals (Mobile Information Systems, Intl. J. of Data Warehousing and Mining, and Intl. J. of Web and Grid Services), a PC chair and a General chair of a number of international conferences, and has been invited to deliver keynote speeches and tutorials at various international events.*