PhD Examination Report

Name: Nour Oweis

University: VSB Technical University of Ostrava

Title: Parallel Association Rule Mining Algorithm Based on MapReduce by Using Lift Interestingness Measure for Big Data

Big Data mining is an analytic process utilized to discover the hidden knowledge and patterns from a massive, complex, and multidimensional dataset. Single processors memory and CPU resources are very limited in this aspect, which makes the algorithm performance ineffective. Association rule mining is traditionally used to uncover hidden knowledge in data sets. To use with big data, a scalable and parallel strategies for are needed. This thesis proposed two algorithms for data mining and optimization. The first is parallel association rule mining algorithm based on MapReduce by using LIM (MapReduce Lift Association Rule (MRLAR)), to provide high scalability over parallel execution. The second is to reduce dimensionality by using multiple data reduction techniques including principle component analysis (PCA), singular value decomposition (SDD), semi-discrete decomposition (SVD), and applied to reduce the data into fewer dimensions as pre-processing techniques for data optimization. The thesis is organized as 6 chapters as follows:

After a gentle introduction, Chapter Two Consists of the literature review involving big data and analytic tools, data mining types, and data reduction with multiple techniques including SDD, SVD, and PCA.

Chapter Three is focused on parallel association rule mining and reducing dimensionality by utilizing several methods, measures, and techniques. The two proposed algorithms including the prototype implementation, Graphic User Interface (GUI) used in the MapReduce methodologies, dataset collections, software used, algorithms steps and experimental results are presented in Chapters four and five. Dataset provided by the USA domestic airline flights between 1987 and 2008 and Groceries dataset of real-world point-of-sale transaction were used in the experiments. Finally conclusions are provided towards the end.

I am impressed by the quality of the work, which is indeed very tough. I believe that this thesis has made a good scientific contribution for the award of a PhD degree and I take this opportunity to congratulate the student and his supervisor.
May 30, 2016