Financial Statement Analyses and Visualizing it as a Graph

Eyup Tekkanat

September 2016

Dissertation submitted in partial fulfillment for the degree of

Master of Science in *Big Data*
Computing Science and Mathematics

University of Stirling

Abstract

The financial data analysis has become increasingly more relevant for the enterprises because the efficient handling of financial information represents a unique competitive advantage, being its application as diverse as the nature of the data, for example, trend analysis, defining spending habits, fraud detection, etc.

A financial graph is a graph of financial transactions. In this case, the nodes are users, banks, categories, transaction method, date and any other information. The edges that connect them show how the data relates to each other. Such as users being connected to different categories that they spend in. There are similarities between topic trending based on classification and aggregating.

Apache Spark is an engine for large-scale data processing, intended to be a drop-in replacement for Hadoop MapReduce providing the benefit of improved performance; the main goal of this project is meaning of financial bank statement data, through the development and implementation of a distributed pipeline for processing and visualizing high-speed financial data streams coming from bank statement and detect trends and spending habits. Spark-related projects and libraries: Spark Streaming and Spark MLlib used for this purpose. To verify the effectiveness of the algorithm, different benchmarks (with different configurations) will be performed, these results will be analyzed and visualized with data visualization tool and libraries.