ABSTRACT

Name of Candidate:	**Imran Khan**
Name of Supervisor:	**Dr. S. K. Naqvi**
Name of Co-supervisor:	**Dr. Mansaf Alam**
Name of Co-supervisor:	**Dr. S. N. A. Rizvi**
Name of Department:	**FTK- Centre for Information Technology**
Title:	**A Framework for Analysis of Big Data in Cloud based Environemnt**

As the field of computer science and information technology has grown rapidly there is a great enhancement in the traditional systems which makes a revolutionizing increase in the available data and created a new discipline that we call Data Science. This outburst of huge data can change the scenario for development of a number of disciplines including science, engineering mathematics etc. As the Internet of things (IoT) applications has gained popularity and the cost of Information and Communications technology (ICT) has reduced, a huge data oceans has been generated from a number of devices and many valuable information can be mined from it. The rate at which data is increasing is very high that brings a new era in data processing 'The Big Data Processing'. Majority component of Big Data's output on the digital forefront is unstructured and analyzing it to draw meaning out of it is a challenging task. As social networking sites play a very important role in Big Data generation, analyzing their data volumes is highly worth-full and can lead to venturing into meaningful insights pertinent to specific industries. Such analysis not only helps other businesses to benefit from but also enhances the quality and satisfaction of user experience on the same social networking channels as well. Twitter is one of the most popular social networking platforms where daily in and out users come and speak their mind out on trending

topics in national as well as international space. The audience sentiment and opinion on any trending topic can easily be gauged by the tone of their arguments and opinions that they post on their twitter handles. The consequences on the social and political issues can be predicted in advanced and respective actions can be taken. Analyzed data of Twitter could be used by many other secondary businesses to benefit for their growth and create a cutting edge space for themselves over their competitors in their business domain. For instance, in order to find the popularity of any product, or to find the sentimental inclination of people towards a particular political leader etc, the trending twitter handles on the topic can be conjoined and analyzed for insights. A number of tools and techniques are available for analyzing such data but they are well suited for the organizations that can spend ample amount of money on the infrastructure needed to implement the analyzing platforms. Mid and small scale organizations are still in need for a solution that they can implement to harness Big Data.

In this research work, we have done exhaustive study of the state of the art and proposed a unified cloud based model for mid-scale organizations so that they can leverage Big Data analysis and optimizations at their organizational cost. We propose two algorithmic frameworks for real time tweets analysis and a unified cost effective model to implement those algorithmic frameworks in the real world. The proposed framework can be very useful for the organizations that have limited resources as it provide solution to leverage Big Data at a very affordable cost.