**DOKUZ EYLUL UNIVERSITY**

**ENGINEERING FACULTY**

**COMPUTER ENGINEERING DEPARTMENT**



**Social Media Analyse Using Hadoop**

**Report#1**

**SociAnalyzer**

2008510057 Ferhat YENIGUN([ferhatyenigun1@gmail.com](mailto:ferhat_yenigun@hotmail.com))

20085100107 Halil SATIK([halilsatik@outlook.com](mailto:halilsatik@outlook.com))

2010510004 Bilal AKSOY([bilal\_roda@hotmail.com](mailto:bilal_roda@hotmail.com))

2010510046 M. Hanifi KOC([muha\_koc@gmail.com](mailto:muha_koc@gmail.com))

### 

### Advisor: Asst. Prof. Dr. Adil ALPKOCAK

# 1. Project Description

Big data is a popular term used to describe the exponential growth and availability of data, both structured and unstructured. And big data may be as important to business – and society – as the Internet has become. More data may lead to more accurate analyses.

More accurate analyses may lead to more confident decision making. And better decisions can mean greater operational efficiencies, cost reductions and reduced risk. Big data usually occupy large space like terabayte or petabytes.Processing Big data take a long time due to this, controlling such a large data become harder. Meanwhile holding large data is imposible for a single space

Cluster systems are used for tackling such circumtances.This systems are required strong CPU capacity, large storage area. This large data are distrubuted to different cpu according to algorithm and are processed.The management of this cluster system is done by Haddop System.

**2.Motivation**

We wonder How to distributed Big Data on cluster system and How to procecessing,. Also How to parallel processing with cpu.

Have to use Cluster System behind internet and How to Controling big Data.

**3. Issue and Scope**

SociAnalyzer aims to develop a solution for analyzing tweets using Hadoop and related Technologies. This project currently contains fetching the Twitter Api of those users whose statuses have been retweeted the most by the user whose tweets are being analyzed.

Project include That developed a program with Pig. Developed a database management with Apache Hive, Simple Interface to query for users.

**4. Outcome**

We will build an end-to-end social media aggregation service capable of handling large volumes of Twitter data.

A Program that to proceed from at the and of SociAnalyzer Project.

Program to be used by everyone. Doctor could be research Disease-related words, Politicians could be look that words related to politics.

**5. Standarts and Legal Conditions**

People have a private life. In the meantime project, private life of the people will be considered

**6.Keywords**

Big data,Paralel Processing, Hadoop, Social Media,Social Media Analyse

Similar Projet:Hadoop Twitter Analytics

**7. References**

<http://www.codeplex.com/>

<http://www.microsoft.com/en-us/openness/default.aspx#home>

<http://www.michael-noll.com/tutorials/running-hadoop-on-ubuntu-linux-single-node-cluster/>

<https://www.wiziq.com/course/23733-hadoop-big-data-training>

<http://www.cloudera.com/>

<http://tutorialshadoop.com/big-data-hadoop-mapreduce-java-programs/>

Hadoop.The.Definitive.Guide.3rd.Edition(Tom White)