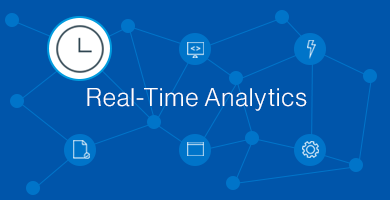
LARGE SCALE DATA PROCESSING PROJECT

(CSE 3025)

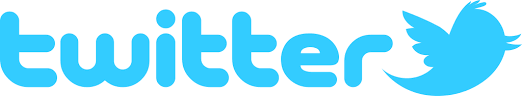
DR R. BHARGAVI



WINTER SEMESTER 2016-2017

|  |
| --- |
| Real-Time Analytics with Apache Storm  using data provided by Twitter |
|  |
|  |
|  |
|  |





KASHISH MIGLANI (15BCE1003)

OSHO AGYEYA (15BCE1326)

TERMINOLOGY

Apache Storm

Apache Storm is a free and open source distributed real-time computation system. Storm makes it easy to reliably process unbounded streams of data, doing for real-time processing. Moreover Storm can be used with any programming language.

Real Time Data Analytics

Real-time analytics is the use of or the capacity to use, data and related resources as soon as the data enters the system.

It refers to a level of computer responsiveness that a user senses as immediate or nearly immediate.

Twitter

Twitter is an online news and social networking service where users post and interact with messages, "tweets," restricted to 140 characters.

ABSTRACT

**PLATFORM:** APACHE STORM (VERSION-1.0.3)

This project is aimed at the following targets:

* To learn Apache Storm, fundamental concepts involved in real time analytics as well as basic Storm terminology.
* To understand and implement Storm-Twitter connectivity.
* To use the above connectivity in order to obtain top-n trending worldwide hash tags.
* To retrieve tweets related to the above top-n trending worldwide hash tags.