

---

# Requirements and Use Cases

CS 321-005 Team 3 Project Deliverable 1

Baker, Guo, Oh, Syed



Department of  
**Computer Science**

Compiled on September 29, 2019 at 9:37pm

## Contents

<b>System Overview</b>	<b>2</b>
<b>Rationale for Building</b>	<b>2</b>
Problem and Motivation . . . . .	2
Solution . . . . .	2
<b>Methodology Used</b>	<b>3</b>
<b>Requirements</b>	<b>3</b>
Functional Requirements . . . . .	3
Fetch data from Twitter . . . . .	3
Extract meaningful content from Twitter . . . . .	4
Enrich content from Twitter with sentiment analysis . . . . .	4
Load enriched content into a data store . . . . .	4
Load data into multiple data stores . . . . .	5
Provide relevant analytics on the ingested information . . . . .	5
Non-Functional Requirements . . . . .	6
Scalable system . . . . .	6
Distributed system . . . . .	6
Guaranteed Delivery . . . . .	6
<b>Use Case Descriptions</b>	<b>6</b>
U1: View average historical sentiment . . . . .	7
U2: View specific user's historical sentiment . . . . .	8
U3: View average historical sentiment relating to a keyword . . . . .	9
U4: View a specific user's historical sentiment relating to a keyword . . . . .	10
<b>Contributions</b>	<b>11</b>

## System Overview

Provide a high-level overview of the system. This is the product vision and should be stated in the users' vernacular.

The Team Big Data (TBD) Analytics Package performs sentiment analysis of tweets. The sentiment analysis of a user's tweet is a group of numbers which can be thought of as representing the user's net emotional state (negative/neutral/positive). This information can be used to more effectively target a user with advertisements or gain insights to their desires.

Take your marketing to the next level with the TBD Analytics Package.

## Rationale for Building

Outline the problem, the motivation for building such a product, and your solution approach. Also, if there are other alternatives to your product in the market, discuss the unique characteristics of your solution and the shortcomings of the existing products.

## Problem and Motivation

Facebook and Google, the "F" and "G" of FANG, are two of the market's four best-performing tech stocks. Facebook and Google's primary sources of revenue comes from selling ads through their respective platforms. Part of why Facebook and Google have excelled where their competitors have stalled is their ability to acquire information about their users. With every post, click, or search, users of these platforms tell the monolithic companies behind them things about themselves. This information is sold to a third party or used by the company to increase the effectiveness of targeted advertisements.

Companies like Facebook and Google have a monopolistic control of their platform – they decide what information to share about their users and the prices to charge for that information. In part due to this monopolistic control, information has become a market. Numerous companies have sprung up in this "information market", each purporting to deliver analytics packages to businesses which would perform miracles like doubling their marketing reach or tripling their click-through rate. These offerings are particularly enticing to businesses without the fortune of controlling a platform which passively aggregates their target market's information. However, many of the benefits these companies offer never materialize.

## Solution

Team Big Data (TBD) seeks to fill a perceived gap in the market. Most companies lack a platform which passively aggregates information about their target market (à la Google and Facebook). These companies stand to benefit from analytics about the demographics they hope to reach with their advertisements.

TBD's Analytics Package uses semantics analysis in a novel way to provide relevant and actionable information ranging from specific members of a target market to well the market as a whole.

## Methodology Used

Describe the process by which you gathered, analyzed, and validated the requirements for the system under consideration. Briefly discuss the criteria used for evaluating the requirements.

Following the Scaled Agile Framework (SAFe) our team began the semester on a Program Increment Planning (IP) event which lasts for the duration of a single sprint.

During the IP, our team met with Dr. Soundararajan to run through a mock Program Increment event. Dr. Soundararajan and our team's System Architect, Connor, acted as the customer and described features that they would like the system to have. This conversation between the customer and the developers ensured that both are on the same page and that no infeasible requirements were left on the table. The team took those features and began to break them down into user stories so that they could be more accurately estimated.

With a backlog full of user stories, the team prioritized features by using Weighted Shortest Job First (WSJF) to enable higher velocity and an increased likelihood of producing a Minimum Viable Product (MVP).

## Requirements

### Functional Requirements

Provide a prioritized and approximately estimated list of functional requirements for the system. Describe each requirement briefly and state the rationale for including that requirement. Also, discuss the prioritization and estimation schemes used.

#### Fetch data from Twitter

Twitter is the main source of information that our analytics package draws from. As such, we need to be able to fetch data from Twitter.

**Dependencies** No dependencies.

**Priority** This requirement has an extremely high priority because all the other priorities depend on it.

**Estimate** The time estimate is about two weeks due to our team's low velocity. This estimation was derived from prior experience.

#### **Extract meaningful content from Twitter**

The data that Twitter returns is too verbose for the purposes of our analytics package. Since processing, storing, and analyzing data requires energy and money, we should aim to do as little of these things as possible. Therefore, it is necessary to extract only meaningful information from what Twitter returns.

#### **Dependencies**

- Fetch data from Twitter

**Priority** This requirement has a medium priority as it serves only to help TBD minimize the utilization of its pipeline and avoid retaining potentially sensitive information.

**Estimate** The time estimate is about two weeks due to our team's low velocity. This estimation was derived from prior experience.

#### **Enrich content from Twitter with sentiment analysis**

Sentiment analysis is the cornerstone of the proposed analytics package. As such, we must ensure that our pipeline can perform that analysis.

#### **Dependencies**

- Fetch data from Twitter

**Priority** This requirement has an extremely high priority so that we can enable analytics and visualizations.

**Estimate** The time estimate is about two weeks due to our team's low velocity. This estimation was derived from prior experience.

#### **Load enriched content into a data store**

Excluding streaming analytics, data must be at rest to be analyzed. As such, our pipeline must have the ability to load data into a data store.

**Dependencies**

- Fetch data from Twitter
- Enrich content from Twitter with sentiment analysis

**Priority** This requirement has an extremely high priority as it enables analytics and visualizations.

**Estimate** The time estimate is about one week due to our team's low velocity. This estimation was derived from prior experience.

**Load data into multiple data stores**

To avoid vendor-lock in and promote the health of the product, our pipeline should be able to ingest into several different data stores.

**Dependencies** No dependencies.

**Priority** This requirement has a medium priority as it only serves to prevent vendor lock-in.

**Estimate** The time estimate is about one week due to our team's low velocity. This estimation was derived from prior experience.

**Provide relevant analytics on the ingested information**

The proposed package should provide meaningful analytics which analyze not only a user's emotional sentiment at a single instance in time but also some arbitrary span of time.

**Dependencies**

- Fetch data from Twitter
- Enrich content from Twitter with sentiment analysis

**Priority** This requirement has an extremely high priority as it provides the analytics component of the TBD Analytics Package.

**Estimate** The time estimate is about two weeks due to our team's low velocity. This estimation was derived from prior experience.

## Non-Functional Requirements

Identify the non-functional requirements for the system. Include the measurement criteria for each of the non-functional requirements. Record any constraints that these requirements may place on the system to be built.

### Scalable system

A scalable system is one which can run on many different threads on a single machine. A scalable system is a necessary for non-trivial big-data packages.

**Metric** System has the capability to scale across multiple threads.

### Distributed system

A distributed system is one which can run on many different machines. A distributed system is a necessary for non-trivial big-data packages: it is not usually possible to "scale up" a machine (that is, make it more powerful), however, it is usually trivial to "scale out" an application to more machines.

**Metric** System has the capability to scale across multiple machines.

### Guaranteed Delivery

The quality of an analytics package is dependent upon the quality of the data it analyzes. Guaranteed Delivery ensures that all the data that enters the pipeline is eventually ingested into a data store, removing the possibility that data is lost along the way.

**Metric** Data Flow Pipeline provides Guaranteed Delivery.

## Use Case Descriptions

Identify four use cases and develop a use case description for each by applying the use case template. Each

use case description must include a description of the main sequence of interactions between the actor(s) and the system, as well as description of the alternative sequences.

## U1: View average historical sentiment

1. **Objective:** An analyst observes average historical sentiment over some interval of time
2. **Priority:** High
3. **Source:** A1 (an analyst, not a developer)
4. **Actors:** analyst, Kibana
5. **Flow of Events**
  - a. Basic Flow
    - i. Initiated when A1 navigates to Kibana
    - ii. The analyst selects the primary dashboard
    - iii. The analyst selects the time frame if the default time frame is not suitable
    - iv. Kibana displays the average sentiment over the provided interval of time
  - b. Alternative Flow 1
    - v. Initiated at the third step of Basic Flow if there is no data in the specified interval of time
    - vi. Kibana displays a message indicating that there is no data within that interval
    - vii. The analyst views the error message
    - viii. Return to step one of Basic Flow
  - c. Exception Flow 1
    - ix. Initiated at the first step of Basic Flow if Kibana is not available
    - x. The system displays an error message of Kibana being unavailable
    - xi. The analyst views the error message
  - d. Exception Flow 2
    - xii. Initiated at the second step of Basic Flow if the primary dashboard is unavailable
    - xiii. The system displays an error message of the object at that URL no long being unavailable
    - xiv. The analyst views the error message
6. **Includes:** None



7. **Preconditions:** The analyst is logged in to Kibana
8. **Post conditions:** The analyst can observe the average emotional sentiment over some interval of time
9. **Notes/Issues:**
  - e. None

## U2: View specific user's historical sentiment

1. **Objective:** An analyst observes a specific user's historical sentiment over some interval of time
2. **Priority:** High
3. **Source:** A1 (an analyst, not a developer)
4. **Actors:** analyst, Kibana
5. **Flow of Events**
  - a. Basic Flow
    - i. Include U1: View average historical sentiment
    - ii. The analyst enters a query for a specific user
    - iii. Kibana updates all visualizations to display only information pertaining to the user entered
  - b. Alternative Flow 1
    - iv. Initiated at the third step of Basic Flow if there is no data in the specified interval of time or the user does not exist
    - v. Kibana displays a message indicating that there is no data matching the provided criteria
    - vi. The analyst views the error message
    - vii. Return to step one of Basic Flow
  - c. Exception Flow 1
    - viii. Initiated at the first step of Basic Flow if Kibana is not available
    - ix. The system displays an error message of Kibana being unavailable
    - x. The analyst views the error message
  - d. Exception Flow 2
    - xi. Initiated at the second step of Basic Flow if the primary dashboard is unavailable
    - xii. The system displays an error message of the object at that URL no longer being unavailable

xiii. The analyst views the error message

6. **Includes:** U1: View average historical sentiment

7. **Preconditions:** The analyst is logged in to Kibana

8. **Post conditions:** The analyst can observe a specific user's historical sentiment over some interval of time

9. **Notes/Issues:**

e. None

### U3: View average historical sentiment relating to a keyword

1. **Objective:** An analyst observes average historical sentiment over some interval of time which pertains to some keyword

2. **Priority:** Medium

3. **Source:** A1 (an analyst, not a developer)

4. **Actors:** analyst, Kibana

5. **Flow of Events**

a. Basic Flow

i. Include U1: View average historical sentiment

ii. The analyst enters a query for a specific keyword

iii. Kibana updates all visualizations to display only information pertaining to the keyword entered

b. Alternative Flow 1

iv. Initiated at the third step of Basic Flow if there is no data in the specified interval of time or the keyword was not found

v. Kibana displays a message indicating that there is no data matching the provided criteria

vi. The analyst views the error message

vii. Return to step one of Basic Flow

c. Exception Flow 1

viii. Initiated at the first step of Basic Flow if Kibana is not available

ix. The system displays an error message of Kibana being unavailable

x. The analyst views the error message

d. Exception Flow 2

- xi. Initiated at the second step of Basic Flow if the primary dashboard is unavailable
  - xii. The system displays an error message of the object at that URL no long being unavailable
  - xiii. The analyst views the error message
6. **Includes:** U1: View average historical sentiment
7. **Preconditions:** The analyst is logged in to Kibana
8. **Post conditions:** The analyst can observe a specific user's average emotional sentiment over some interval of time
9. **Notes/Issues:**
- e. None

#### **U4: View a specific user's historical sentiment relating to a keyword**

1. **Objective:** An analyst observes a specific user's historical sentiment over some interval of time which pertains to some keyword
2. **Priority:** Medium
3. **Source:** A1 (an analyst, not a developer)
4. **Actors:** analyst, Kibana
5. **Flow of Events**
  - a. Basic Flow
    - i. Include U2: View specific user's historical sentiment
    - ii. The analyst enters a query for a specific keyword
    - iii. Kibana updates all visualizations to display only information pertaining to the keyword entered
  - b. Alternative Flow 1
    - iv. Initiated at the third step of Basic Flow if there is no data in the specified interval of time or the keyword was not found
    - v. Kibana displays a message indicating that there is no data matching the provided criteria
    - vi. The analyst views the error message
    - vii. Return to step one of Basic Flow
  - c. Exception Flow 1
    - viii. Initiated at the first step of Basic Flow if Kibana is not available

- ix. The system displays an error message of Kibana being unavailable
- x. The analyst views the error message
- d. Exception Flow 2
  - xi. Initiated at the second step of Basic Flow if the primary dashboard is unavailable
  - xii. The system displays an error message of the object at that URL no long being unavailable
  - xiii. The analyst views the error message
- 6. **Includes:** U2: View specific user's historical sentiment
- 7. **Preconditions:** The analyst is logged in to Kibana
- 8. **Post conditions:** The analyst can observe a specific user's average emotional sentiment over some interval of time
- 9. **Notes/Issues:**
  - e. None

## Contributions

Include information about the following: 1) the name of the person who managed the work for this deliverable and 2) individual contributions.

Connor managed the work for this deliverable.

Connor worked as the System Architect to answer questions the team had and acted as the customer to help deliver to the team a list of desired features. He made himself available for Q&A with the rest of the team as they created requirements and use cases.

Ghousia, Shin, and Ziyen held a mock Increment Planning event with Connor and Dr. Soundararajan acting as the customer to begin making notes of desired features, allowing them to formalize requirements. They also created the different use-cases. In addition, they did work outside of this deliverable to get up to speed with a technology critical to the TBD Analytics Package, producing two artifacts as they did so.