

## TABLE OF CONTENT

<b>CHAPTER 1 .....</b>	1
<b>INTRODUCTION.....</b>	1
1.1 Topic of the System .....	1
1.2 Purpose of the System.....	1
1.3 Target Reader.....	1
1.4 Topic Background.....	1
1.5 Problem Context .....	2
1.6 Rational .....	3
1.6.1 Tangible Benefits.....	3
1.6.2 Intangible Benefits.....	3
1.7 Objectives of the System .....	4
1.7.1 Project Objective .....	4
1.7.2 Project's Learning Objective .....	5
1.8 Project Scope .....	5
1.9 Functionality of the System .....	6
1.9.1 Core Functionalities.....	7
1.9.2 Enhance Functionalities.....	8
1.9.3 Special Features .....	10
1.10Conclusion .....	10
<b>CHAPTER 2 .....</b>	11
<b>PROBLEM DESCRIPTION.....</b>	11
2.1 Problem Identification and Description .....	11
2.2 Problem Importance and Justification .....	11
2.2.1 Importance of problem for the proposed system .....	11
2.2.2 Justification of problem for the proposed system.....	12
2.3 Feasibility Study .....	12
2.3.1 Technical Feasibility.....	12
2.3.2 Operational Feasibility .....	14

---

2.3.3 Economic Feasibility .....	16
2.3.4 Schedule Feasibility.....	16
2.4 Existing Constraint of the Problems .....	17
2.5 Problem Assumption.....	17
2.6 Challenges.....	17
2.7 Conclusion .....	18
<b>CHAPTER 3 .....</b>	<b>19</b>
<b>LITERATURE REVIEW.....</b>	<b>19</b>
3.1 Introduction to Literature Review.....	19
3.2 Advance Preliminary Research.....	19
3.2.1 Big Data.....	20
3.2.2 Critical Research on Big Data and Hadoop.....	21
3.2.3 Critical Research on MapReduce Analysis .....	22
3.3 Critical Research on Existing Systems .....	23
3.3.1 Market Research .....	23
3.3.2 Existing Product Analysis and Comparison .....	25
3.4 Secondary Research.....	26
3.4.1 Academic Research .....	26
3.4.2 Technical and Domain Research .....	29
3.4.3 How is Hadoop Different from Past Techniques? ( <i>Mapr.com, 2016</i> ) .....	30
3.4.4 Predictive Analysis .....	30
3.5 Conclusion .....	31
<b>CHAPTER 4 .....</b>	<b>32</b>
<b>RESEARCH.....</b>	<b>32</b>
4.1 Data Gathering Techniques.....	32
4.2 Primary Research .....	32
4.2.1 Questionnaire .....	32
4.2.2 Interview .....	39
4.3 Technical Research .....	42
4.3.1 MapReduce Algorithm .....	42
4.3.2 Database Research .....	42

---

---

4.3.3 Programming Language Research.....	43
4.3.4 Methodology Research .....	44
4.4 Development Plan.....	45
4.5 Conclusion.....	47
<b>CHAPTER 5 .....</b>	<b>48</b>
<b>ANALYSIS.....</b>	<b>48</b>
5.1 Analysis .....	48
5.2 Analysis of Questionnaire .....	48
5.2.1 Overall Conclusion of the analysis .....	49
5.3 Analysis of Interview .....	50
5.3.1 Overall Conclusion of the analysis .....	52
5.4 Conclusion.....	52
<b>CHAPTER 6 .....</b>	<b>53</b>
<b>SYSTEM DESIGN.....</b>	<b>53</b>
6.1 Introduction to system design .....	53
6.2 System Use Cases.....	54
6.2.1 Use Case 1: Earning Explorer Based On Hadoop Streaming .....	54
6.2.2 Use Case 2: Registration .....	56
6.2.3 Use Case 3: Referral Program.....	58
6.2.4 Use Case 4: Weekly Poll Specification .....	59
6.2.5 Use Case 5: Daily Goal.....	61
6.2.6 Use Case 6: Settings.....	62
6.3 System Activity Diagram .....	64
6.3.1 Activity 1: Earning Explorer based on Hadoop Streaming .....	64
6.3.2 Activity 2: Registration.....	65
6.3.3 Activity 3: Referral Program .....	66
6.3.4 Activity 4: Weekly Poll .....	67
6.3.5 Activity 5: Notification .....	68
6.3.6 Activity 6: Settings .....	69
6.4 System Sequence Diagram.....	70
6.4.1 Sequence 1: Earning Explorer based on Hadoop Streaming .....	70

---

---

6.4.2	Sequence 2: Registration.....	71
6.4.3	Sequence 3: Referral Program .....	72
6.4.4	Sequence 4: Weekly Poll .....	73
6.4.5	Sequence 5: Notification.....	74
6.4.6	Sequence 6: Settings .....	75
6.5	Class Diagram .....	76
6.6	Entity Relationship Diagram (ERD) .....	77
6.6.1	Entities Identified.....	77
6.6.2	Relationship Identified.....	77
6.6.3	System ERD.....	78
6.6.4	Mapping Cardinality and Participation Constraint .....	79
6.6.5	Mapping ERD to Relational Database .....	80
6.7	Normalization.....	82
1.	First Normal Form (INF).....	82
2.	Second Normal Form (2NF) .....	83
3.	Third Normal Form (3NF) .....	83
6.7.1	Tables Dependencies .....	84
6.8	Data Dictionary .....	84
6.8.1	Table registration .....	84
6.8.2	Table earning .....	85
6.8.3	Table refer .....	85
6.8.4	Table survey.....	86
6.8.5	Table surveyresponses .....	86
6.8.6	Table weeklypoll.....	87
6.8.7	Table pollresponse .....	87
6.9	Testing Template.....	88
6.10	Conclusion.....	88
<b>CHAPTER 7</b>	.....	89
<b>IMPLEMENTATION</b>	.....	89
7.1	Introduction .....	89
7.2	Implementation Plan .....	90

---

---

7.2.1	PERT Chart.....	90
7.2.2	PERT Chart for Earning Explorer.....	90
7.3	Tools Used.....	91
7.4	Implementation of Complex Feature.....	92
7.4.1	Referral Program.....	92
7.4.2	Daily Goal .....	93
7.4.3	Analysis.....	94
7.5	Complex Feature Tried But Not Completed .....	96
7.6	Complex Feature of Medium Used .....	97
7.6.1	Java Programming.....	97
7.6.2	Hadoop Technology.....	98
7.7	Technical Quality According to Software Engineering Standard.....	99
7.7.1	Memory Utilization.....	100
7.7.2	Performance .....	100
7.7.3	User Interface.....	102
7.7.4	Validation.....	106
7.8	Report Quality .....	107
7.9	Technical Manual.....	109
7.9.1	How to set up Earning Explorer in NetBeans IDE? .....	109
7.9.2	How to start file scheduling from Local System to Hadoop Distributed File System? .....	111
7.10	User Manual .....	115
7.11	Conclusion.....	119
<b>CHAPTER 8</b>	.....	120
<b>TESTING</b>	.....	120
8.1	Introduction .....	120
8.2	Test Plan.....	121
8.2.1	Test Duration .....	121
8.2.2	Test Cases to be Tested.....	121
8.2.3	Testers List.....	122
8.2.4	Testing Traceability .....	122

---

---

8.3	Unit Testing.....	128
8.3.1	Black Box Testing.....	128
8.3.2	White Box Testing .....	135
8.4	Integration Testing .....	141
8.5	System Testing .....	144
8.6	User Acceptance Testing.....	145
8.7	Usability Testing .....	149
8.8	Compatibility Testing.....	153
8.9	Runtime Testing .....	154
<b>CHAPTER 9</b>	.....	155
<b>CRITICAL EVALUATION</b>	.....	155
9.1	Problem Solved for User Benefits .....	155
9.1.1	Search Engine .....	155
9.1.2	Earning.....	155
9.1.3	Profit .....	155
9.1.4	Time Saving.....	155
9.2	Success Degree.....	155
<b>CHAPTER 10</b>	.....	157
<b>CONCLUSION</b>	.....	157
10.1	Limitations of the system .....	157
10.2	Future Enhancement.....	157
10.3	Computational Challenges .....	157
10.4	Learning Experience .....	158
10.5	Development Experience .....	158
10.6	Personal Experience .....	159
<b>REFERENCE</b>	.....	160
<b>APPENDIX</b>	.....	164
	PROJECT PROPOSAL FORM.....	165
	PROJECT SPECIFICATION FORM.....	169
	ANALYSIS REPORT .....	194
	IMPORTANT DOCUMENTS.....	205

---

**TABLE OF TABLE**

Table 1: Core Functionalities.....	7
Table 2: Enhance Functionalities.....	8
Table 3: Special Features .....	10
Table 4: Hardware Requirements for the project.....	13
Table 5: Software requirements for the project .....	13
Table 6: PIECES Report for the project's Operational Feasibility.....	14
Table 7: Journal Excerpt .....	21
Table 8: Research Excerpt on MapReduce .....	22
Table 9: Comparison of existing products with Earning Explorer .....	25
Table 10: Agile phases and methods .....	27
Table 11: Comparison between Hive and RDBMS .....	43
Table 12: Comparison between C++,C# and Java.....	43
Table 13: Comparison between Waterfall, Spiral and Agile .....	44
Table 14: Test Strategy Schedule .....	45
Table 15: Questionnaire Analysis Matrix .....	49
Table 16: Interview Question Analysis.....	50
Table 17: Use Case 1 .....	54
Table 18: Use Case 1: Earning Explorer Specification .....	56
Table 19: Use Case 2 .....	56
Table 20: Use Case 2: Registration Specification.....	57
Table 21: Use Case 3 .....	58
Table 22: Use Case 3: Referral Program Specification .....	59
Table 23: Use Case 4 .....	59
Table 24: Use Case 4: Weekly Poll Specification .....	60
Table 25: Use Case 5 .....	61
Table 26: Use Case 5: Daily Goal Specification .....	62
Table 27: Use Case 6 .....	62
Table 28: Use Case 6: Settings Specification .....	63

---

Table 29: Class Diagram Description .....	77
Table 30: Entities Identified.....	77
Table 31: Relationship Identified.....	77
Table 32: Mapping Cardinality .....	79
Table 33: Participation Constraint .....	80
Table 34: Mapping of Strong Entity .....	81
Table 35: Mapping of Weak Entity .....	81
Table 36: Mapping of 1: M.....	81
Table 37: Mapping of M: 1.....	82
Table 38: Mapping of M: N .....	82
Table 39: Database Table Registration .....	85
Table 40: Database Table Earning .....	85
Table 41: Database Table Refer.....	86
Table 42: Database Table Survey .....	86
Table 43: Database Table Surveyresponses.....	86
Table 44: Database Table Weeklypoll .....	87
Table 45: Database Table Pollresponse .....	87
Table 46: Testing Template .....	88
Table 47: Implementation plan .....	90
Table 48: Tools Used .....	91
Table 49: System Design Principles .....	102
Table 50: Report Quality Evaluation .....	107
Table 51: Steps to install.....	109
Table 52: Test Duration .....	121
Table 53: Test Cases .....	121
Table 54: Testers.....	122
Table 55: Registration Traceability .....	122
Table 56: Login Traceability .....	123
Table 57: Search Traceability .....	123
Table 58: Weekly Poll Traceability .....	123
Table 59: Survey Traceability.....	124

---

---

Table 60: Daily Goal Traceability .....	124
Table 61: Referral Program Traceability .....	125
Table 62: Notification Traceability.....	125
Table 63: Explorer TV Traceability.....	125
Table 64: Settings Traceability .....	126
Table 65: User Guide Traceability.....	126
Table 66: Visualization Traceability.....	126
Table 67: Login and User Home Traceability .....	127
Table 68: Weekly Poll and Visualization Traceability .....	127
Table 69: Refer and Registration Traceability.....	128
Table 70: Daily Goal Integration Traceability.....	128
Table 71: Registration Unit Testing.....	129
Table 72: Login Unit Testing.....	130
Table 73: Search Unit Testing .....	131
Table 74: Weekly Poll Unit Testing .....	131
Table 75: Survey Unit Testing.....	132
Table 76: Daily Goal Unit Testing.....	132
Table 77: Referral Program Unit Testing .....	133
Table 78: Notification Unit Testing.....	133
Table 79: Settings Updating Data Unit Testing .....	134
Table 80: Settings Report View Unit Testing.....	134
Table 81: User Guide Unit Testing.....	135
Table 82: Visualization Unit Testing.....	135
Table 83: MapReduce Analysis White Box Chart:.....	137
Table 84: Analysis Code.....	137
Table 85: Referral White Box Chart.....	140
Table 86: Referral Program Code .....	140
Table 87: Login with user Home Integration Testing.....	141
Table 88: Login with user Home Integration Testing.....	142
Table 89: Refer with Registration Integration Testing .....	142
Table 90: Daily Goal with Search, Survey, Weekly Poll and Referral Integration Testing .....	143

---

---

Table 91: System Testing.....	144
Table 92: Analysis Summary Chart.....	148
Table 93: Learnability.....	150
Table 94: Memorability .....	150
Table 95: Efficiency.....	151
Table 96: Error .....	152
Table 97: Satisfaction .....	152
Table 98: System Compatibility Testing .....	153
Table 99: System Compatibility Testing .....	154
Table 100: Success Degree Evaluation.....	156

## TABLE OF FIGURE

Figure 1: InboxDollars Screen View .....	23
Figure 2: GrabPoints Screen View .....	24
Figure 3: Software Project Management .....	29
Figure 4: MapReduce Algorithm .....	42
Figure 5: Earning Explorer Use Case .....	55
Figure 6: Registration Use Case .....	57
Figure 7: Registration Use Case .....	58
Figure 8: Weekly Poll Use Case .....	60
Figure 9: Daily Goal Use Case .....	61
Figure 10: Settings Use Case .....	63
Figure 11: Earning Explorer Activity Diagram .....	64
Figure 12: Registration Activity Diagram .....	65
Figure 13: Referral Program Activity Diagram .....	66
Figure 14: Weekly Poll Activity Diagram .....	67
Figure 15: Notification Activity Diagram.....	68
Figure 16: Settings Activity Diagram .....	69
Figure 17: Earning Explorer Sequence Diagram .....	70
Figure 18: Registration Sequence Diagram .....	71

Figure 19: Referral Sequence Diagram.....	72
Figure 20: Weekly Poll Sequence Diagram.....	73
Figure 21: Notification Sequence Diagram .....	74
Figure 22: Settings Sequence Diagram .....	75
Figure 23: System Class Diagram.....	76
Figure 24: Earning Explorer ERD .....	78
Figure 25: System inside View .....	89
Figure 26: Implementation PERT .....	91
Figure 27: Notification Code Snippet.....	97
Figure 28: Mail Code Snippet.....	98
Figure 29: Running MapReduce .....	98
Figure 30: Analysis Output.....	99
Figure 31: MapReduce Framework Performance .....	99
Figure 32: Array list Used.....	100
Figure 33: String Array .....	100
Figure 34: JSP Code Snippet .....	101
Figure 35: MapReduce Performance .....	101
Figure 36: Naming Convention .....	105
Figure 37: Password and Confirm Password Validation .....	106
Figure 38: Already Exist Email Validation .....	106
Figure 39: NB Icon .....	109
Figure 40: Starting NetBeans IDE .....	109
Figure 41: Importing Project.....	110
Figure 42: Browse Zip File .....	110
Figure 43: Project Navigatio.....	111
Figure 44: IP Address at HDFS .....	111
Figure 45: Icon .....	112
Figure 46: WinSCP network login.....	112
Figure 47: Authenticating the Host name .....	113
Figure 48: File Location.....	113
Figure 49: WinSCP Starts Synchronization.....	114

---

---

Figure 50: File at Local System and HDFS .....	114
Figure 51: Earning Explorer Home.....	115
Figure 52: Registration .....	115
Figure 53: User Home.....	116
Figure 54: Search .....	116
Figure 55: Daily Goal .....	117
Figure 56: Refer and Earn.....	117
Figure 57: Weekly Poll .....	118
Figure 58: Survey.....	118
Figure 59: Testing Types .....	120
Figure 60: MapReduce Analysis White Box Testing .....	136
Figure 61: Referral White Box Testing: .....	139
Figure 62: User Acceptance graph.....	149
Figure 63: Usability Features:.....	149
Figure 64: Build Success Report.....	154
Figure 65: Run tine testing.....	154

## CHAPTER 1

### INTRODUCTION

#### 1.1 Topic of the System

‘Earning Explorer based on Hadoop Streaming’: Earning cum Search Engine

#### 1.2 Purpose of the System

The main purpose of the system is to provide an efficient product of search engine coupled with earning feature in the market. There are various determinations of the project to make users understand the concept of online earning and through this project, user have can use search engine to earn. The purposes can be noted down as follows.

- ✓ The users can have such kind of system which offers them to earn in return of their searches done on the search engine.
- ✓ But not only this, they have facility to grab the points on different tasks like surveys and weekly poll. The other tasks have been explained further in detail.
- ✓ User should not face problem in searching updated content.
- ✓ User can accept the concept and feel comfortable to earn maximum amount from home. Only they need is internet.

#### 1.3 Target Reader

The proposed system is open for everyone whether they are student or teachers, employed or unemployed. Anyone who are willing to get money in return of whatever they do online. For example, search about anything on internet and receive cash. This system is capable in attracting people to interact with the system efficiently and work on it in order to make profit. It can produce large scale benefit for those user groups who are linked with the internet on daily basis.

#### 1.4 Topic Background

The idea of ‘**Earning Explorer**’ has been originated with a thought of enhancing the features of search engine along with earning facility. The researcher of the project have researched well and understand the need of earning in the country and same with the improvement of knowledge areas.

The proposed system is a new indication to explore the online earnings with search engine technology and other capacities from where we can highlight the everyday news.

Search engine and other processes of the system includes a large set of data and there is need of proper management of those data without any error. Hence, the developer has chosen the **Hadoop technology** to develop the proposed system as per the user requirements. Hadoop can efficiently and smartly offers the distributed processing of large data set stored in the **Hadoop Cluster** (storing and analyzing huge amount of unstructured data) by using simple programming models. It will handle the web results stored on different nodes in the cluster as per the users' search queries given to the system.

Coming to the earning part, the developer needs to calculate the each earning related to tasks done by each user. Hence, to provide a platform where users can explore their interest jobs through which they can earn money. The more they will work and the more they will earn. The system is allowing to get as much money the users can perform the tasks like **survey**, **weekly poll**, **searches**, and more.

### 1.5 Problem Context

The system is an approach to provide a platform where user can spend working time and can save their personal time. Today almost everyone is familiar with the word 'Search Engine' and practically using it also. We are searching anything starting from very small to very vast things on the search engine. But users are complaining that they are not sure whether the existing search engine is providing updated search results or not. They want the results should be appropriate according to the requirement.

In the same manner, the employability is the need for every individual in the world. The craze of earning online is also well known and everyone is already in the race of and have grabbed the benefits. We all are working hard for our expense and savings. People are searching for different jobs by using search engines. Hence, the developer has thought to develop an earning platform by combining both the topics through which anyone can get an opportunity of getting money. The project will act as a facilitator for them by providing different tasks to earn money.

## 1.6 Rational

After problem identification, the system comes to its implementation area that will provide solutions for the above mentioned problems. It comprises various functionalities based on user interaction and their benefits. The benefits that the users are going to make use of are highlighted below. The benefits has been categorized into two parts.

### 1.6.1 Tangible Benefits

The most important tangible benefit is to get money by using the system and user can get it by performing different activities that are mentioned below.

- **Searching:** The users are getting appropriate search results regarding their searched queries and paid for those searches with a definite amount of points.
- **Referral:** Users can get money on referring the application to their friends after a period of time.
- **Opinion poll:** Answering the question asked on weekly basis and earn one point on every answer.
- **Survey:** Comfortable users can attend available surveys in the system and after completing it successfully they will get plus point in their registered account.
- **Video:** The system is providing money to the users without any extra efforts. Hence, users are having a chance of earning simply by watching videos.

### 1.6.2 Intangible Benefits

Users will be able to use the system very smoothly and can easily recognize the elements and working process of the system. The certain intangible benefits have been listed below.

- **User friendly interface:** Users are provided with the full functional and friendly system. They can easily locate the things presented in the system.
- **Trustworthy:** The system has been managed in the way so that users can trust and feel free to share anything with the system.
- **User satisfaction:** Users will feel satisfied with their hopes. The system will meet the user's expectation level and deliver a full fetched version.

- **More efficient:** The users can run the system without any efforts or difficulties. The system is developed in a competent way and well organized.
- **Memorable:** The system elements are memorable in use. The users of the system will feel easy to remember the functioning of the system so that they can never face any problem in working with the system.
- **Error free:** The system is capable in running without generating any system related errors.

## 1.7 Objectives of the System

The main objective of the project is to facilitate the users with an opportunity of doing and getting paid for their own work. Moreover, they can work on several activities offered by the system for money.

### 1.7.1 Project Objective

The basic thing is to learn how to develop a better search engine using Hadoop streaming as the proposed system is totally based on Hadoop technology. The innovation done with the system is to award money in terms of points to users for operating on it.

The project's main objective lies in certain areas that are as follows.

- To provide a better searching facility with updated results. Using search bar, user can walk around whatsoever they want to hunt.
- To improve the users unemployment rate by offering money for their work done in the system.
- To interact users with different kind of surveys and perform those to get points.
- User can answer the daily asked questions in order to get points. The system will observe the percentage of user response for a particular question as a survey.
- Regular notifications for performing new activities to earn.
- Can refer the system to their friends in order to get advantage from their friend's earning.
- Watch videos of any choice.
- User's account settings are secured. They can make changes only after entering account password even though he or she is logged in to the system.

### 1.7.2 Project's Learning Objective

The developer is required to form new algorithms in order to compute the plus point for a particular user account. Aiming to accomplish the mentioned tasks, the developer need to study Hadoop technology and should be able to work proficiently under Hadoop. The developer must have strong knowledge of the technologies that are mentioned below:

- To understand Linux Operating System Commands.
- To learn JSP and Bootstrap to produce a user friendly interface.
- Applying HCIU principles to enhance the system's usability.
- To apply project management concepts in the process of integrating the whole system.
- Text mining to identify the most relevant information and its pattern from the large data set to evaluate and interpret the output.
- Hadoop Distributed computing to solve a problem or task in a very small time interval. Problem will be divided into many sub tasks and each of which is solved by one or more computers.
- Data migration to transfer data within the system in order to perform different processes.
- Visualization to display the generated prediction in the form of graphs.
- Learned different software that provides ease of working on files e.g. data transfer.

### 1.8 Project Scope

Earning Explorer is a web application through which user can search solutions for their queries and each member of the system will be capable in earning points from their jobs done in the system and can redeem those as money. The structure is talented in data gathering and analyzing. It consists various functionality such as user account, searching, surveys, and weekly poll to collaborate search engine with earnings.

The system is constitutes of three module that are as follows.

#### 1. Profile Module

In this module, the user can create their account to perform different tasks available in the system. One user can refer the system to other person so that he or she can be one of the user of the system.

## 2. Search Module

Basically the search module is defined as its user can search about anything in the system and they will receive an updated search result regarding the entered value. They can explore any type of data like text, images or videos.

## 3. Earn Module

The earning module of the system is providing facility to earn money by attending jobs assigned to a user or even by searching any queries. This module will keep record of earned points with respect to user ids.

## 4. Notification Module

The module is designed to manage the information that should be notified to the user. User will always be made remind about the tasks available for them, successful tasks done by them and about their earnings.

## 5. Visualization Module

The module includes data gathering and data analyzing. Data gathering is to be done by collecting answers of the system's users corresponding to the questions asked on weekly basis. Hence, analysis of those collected data will be done and visualized in the form of different graphs.

## 6. Settings Module

The module is featured with managing user account settings. User can view and change their account details or other account settings after going through the successful security so that no other invalid user can make any unwanted changes in the account.

### 1.9 Functionality of the System

Functionalities explains the working anatomy of the system so that one can easily understand the constituent procedures in an according manner. It differentiate the authorities and accessibility of the system modules. In this project the functionalities has been categorized in three parts.

#### 1. Core Functionalities

## 2. Enhanced Functionalities

### 3. Special Features

To make the Earning Explorer as a successful project, the system should be capable in completing the list of functionalities mention down.

#### 1.9.1 Core Functionalities

Core functionalities holds the essential functions without which the system's basic purpose is incomplete or the system cannot work.

**Table 1: Core Functionalities**

S.no.	Functionality	Description	Module
1.	<b>User Account</b>	Users are required to register themselves in order to create an account on Earning Explorer to perform search and earning events. For registration, they need to submit their personal details including email id and account password and a confirmation link will be sent to the user's email id. After confirmation, user can use the features of Earning Explorer.	Profile Module
2.	<b>Searching</b>	User can search about anything by using search bar. It hosts a custom search engine that will periodically reward registered users with random amounts of plus points (P+).	Search Module, Earn Module
3.	<b>Survey</b>	Survey has been included in the system to take user's criticisms on any topic. The members of Earning Explorer can do many surveys as they will be notified timely about new surveys to do. And hence they will be awarded with some specific plus points on completing the surveys successfully.	Earn Module

<b>4.</b>	<b>Weekly Poll</b>	User can earn points on weekly basis by voting for a question. The given answers will be analyzed to make prediction about future of an issue.	Earn Module, Visualization Module
<b>5.</b>	<b>Profile Settings</b>	User can view and change their account details or other account settings only when they are able to cross the security question successfully so that no undesirable user can access those settings.	Settings Module
<b>6.</b>	<b>Explorer TV</b>	Members can discover and watch videos on Earning Explorer, in a similar fashion to YouTube.	Search Module

### 1.9.2 Enhance Functionalities

Enhance functionalities holds the tasks that is designed to improve the efficiency of system. Without enhanced functionalities the system can work but these function will improve the quality of system.

**Table 2: Enhance Functionalities**

S.no.	Functionality	Description	Module
<b>1.</b>	<b>Updated Search</b>	<p>This functionality will let the system to show the updated results at the top. User can check that how many days back the search result has been updated or we can say when the link was updated.</p> <p>The developer is thinking and doing research to implement this functionality effectively to improve the quality of search results. This will assure the users that they will get appropriate and updated result for their queries.</p>	Search Module

2.	<b>Referral Program</b>	A current user can invite another person to become a member of the system. If referred user signs up as a new user by using the current user's referral link, the current user can earn points that will be equal to 10 percent of the P+ earned over time by the new user.	Profile Module
3.	<b>Daily Goal</b>	<p>A meter on the home page will display daily an amount of plus point that a user should earn on that particular day that will be the goal of the day. And if the goal is reached, bonus will be awarded.</p> <p>The developer is thinking and doing research to implement this functionality effectively to make the user more focused on their earnings.</p>	Earn Module
4.	<b>Notification</b>	<p>User will get notified every time about the new surveys, messages or any type of news.</p> <p>The developer is thinking and doing research to implement this functionality to make the system more efficient and user friendly.</p>	Notification Module
5.	<b>Result Visualization</b>	Weekly poll's answers will be collected as the structured form data and will be analyzed to make predictions on the statement. Users can check the ratio of favor and against on a particular topic. The visualization of analyzed data will be performed in the form of graphs so that user can understand the results more effectively.	Visualization Module

<b>6.</b>	<b>User Guide</b>	The user guide basically tells about how the system will work. Every user can access this facility whenever they want and it will help them to perform any task effortlessly.	Profile Module
-----------	-------------------	---	----------------

### 1.9.3 Special Features

Special features holds the tasks that is designed to make the developed in more demand. This will let the system to work differently in compare with the existing systems. The system will come up as an extraordinary system.

**Table 3: Special Features**

S.no.	Functionality	Description	Module
<b>1.</b>	<b>Bookmark</b>	<p>The user can save any link or search result in the bookmark section as some user tends to store some important pages for their future use. The system aims to allow user to access those pages even when they don't have internet.</p> <p>The developer wants to add this and doing research so that she can implement it to make the system approachable.</p>	Search Module

### 1.10 Conclusion

The chapter Introduction has successfully introduced the problem background and the proposed solution regarding the project. The searching and earning are the two major problem to be worked on in the system. The developer has identified the core, enhanced and special functionalities of the system as a solution for the mentioned problems. The functionalities have been categorized in the form of modules.

## CHAPTER 2

# PROBLEM DESCRIPTION

### 2.1 Problem Identification and Description

As we came to know that users are facing problems in two section one is searching results of search engine and secondly they wants to earn money. The solutions has been combined together to develop a system for user's benefits. The description and the solution for the highlighted problems has been described below.

**Search Engine:** Nowadays there are many search engines available in the market but still users are not satisfied with the existing results. And also many users are not aware about earning online facilities. Users wants to have appropriate and updated search for their queries. The proposed system is planned in a way to provide updated result at the top by showing the number of days back a fact has been updated. Hence, they will get reorganized and efficient information.

**Earning:** Everyone wants to earn money and in comparison with the existing online earning platforms, the suggested system is going to be developed with a thought of getting knowledge plus capital. Now a days there are several platforms which offers money for completing certain tasks. Even most of us are working on those kind of sites too, but the proposed system is not demanding for extra attention to work on it. It is just awarding money for the tasks that takes less amount to complete.

### 2.2 Problem Importance and Justification

#### 2.2.1 Importance of problem for the proposed system

The described problems holds weightage when we look around the small issues regarding the topic. The users are facing problem with several implemented technologies but still they are using those as they are getting benefit whatever the system is providing. The developer has proposed the system to provide a kind of structure that involves user for their advantage only without any issues or expectation failure. The system is important to develop the confidence of users towards the technology and solutions provided by the system.

## 2.2.2 Justification of problem for the proposed system

‘Earning Explorer’ is something that is not bounded to search results only, it is about to encourage people and award users for their searches and other performed activities. Although they were exploring things with existing search engines also, but the proposed system will let them feel more interested in exploring the world as they are getting paid for it. They can enjoy their earnings through searching and different interacting activities like survey, polls or referral programs.

There are several activities that are planned in the proposed system to be develop without any errors using Java and Hadoop technology.

- ✓ Java Mail architecture.
- ✓ Mining analysis on user responses.
- ✓ A Hadoop framework require to process the big data very easily and efficiently.
- ✓ Bootstrap and JSP to make a user friendly web application.
- ✓ Algorithms to calculate earnings for a customer.

Users can join the system for full time as well as for part time. With full time job there is no limitation on earning, customers will feel as an employed person. More they will work, more they will get. But, users with part time can enjoy the system and earn like to fulfill their pocket money or anything.

## 2.3 Feasibility Study

The feasibility study has been carried out to select the best to meet the system performance requirement. The developer has done the four types of feasibility study for the system to identify whether the planned project is technically and financially feasible to develop or not. The feasibility studies are as follows.

### 2.3.1 Technical Feasibility

In technical feasibility, the developer is required to examine whether the proposed project is technically feasible or not. To identify the technical feasible project the following concerns are need to be taken into consideration.

- ✓ Whether the developer has sufficient knowledge on the selected tools, techniques & methodology or not.
- ✓ Whether the required resources are available to the developer or not.
- ✓ Whether the techniques are practical to implement or not.

The analysis done by the developer for the possible hardware, software and operating environment that are required to be used in the implementation of the project. The analysis output has been showed in the below mentioned table.

## Hardware Requirements

**Table 4: Hardware Requirements for the project**

Sr. No.	Technical Requirements	Description
1.	Device	Laptop or Desktop PC
2.	Processor	At least Intel® Core™ i3-3110M
3.	CPU	At least 1.90GHz
4.	RAM	At least 7.5 GB free space in 8GB for faster response
5.	Monitor	Any standard size
6.	Other	USB, Dongle or Wi-Fi router for internet connection

## Software Requirements

**Table 5: Software requirements for the project**

Sr. No.	Technical Requirements	Description
1.	Operating System	Windows 10 (32 or 64 bit)
2.	Software/ Tools	<ul style="list-style-type: none"> <li>✓ NetBeans IDE 8.0.2</li> <li>✓ Oracle VM Virtual Box</li> <li>✓ Cloudera Manager</li> <li>✓ JRE 8, JDK 1.8</li> </ul>

3.	Technology	<ul style="list-style-type: none"> <li>✓ Hadoop framework-2.3.0</li> <li>✓ JSP</li> <li>✓ Bootstrap</li> </ul>
4.	Web Browser	Firefox / Google chrome

### 2.3.2 Operational Feasibility

The developer has to judge the expected responses of the customer whether the developed project will be used by the end users or not. The feasibility report will check whether the project has successfully accomplished its object or not. The developer has analyzed the facts by using PIECES framework to identify and solve operational feasibility.

**Table 6: PIECES Report for the project's Operational Feasibility**

PIECES	PIECES Terms	Description
<b>P</b>	<b>Performance</b>	<p>Whether the system is providing sufficient space for the functioning of each module.</p> <ul style="list-style-type: none"> <li>✓ A platform to increase efficiency rate of search.</li> <li>✓ Correct output of the given input in time.</li> <li>✓ Better ways to earn money.</li> <li>✓ Good usability of the system as per the requirement.</li> <li>✓ Management of each account of the system. Such as notifications, profile.</li> <li>✓ Record maintenance and system update.</li> </ul>
<b>I</b>	<b>Information</b>	<p>Whether the system is capable in delivering fruitful information for its intended users.</p> <ul style="list-style-type: none"> <li>✓ Accuracy of information.</li> <li>✓ Displaying the available data in a useful manner.</li> <li>✓ Bugs free information and rechecking of it.</li> </ul>

		<ul style="list-style-type: none"> <li>✓ Make available correct data to user according to input provided by the user.</li> <li>✓ Buffer free input and fast output.</li> <li>✓ Security of stored information.</li> </ul>
E	<b>Economic</b>	<p>Cost of accessing the website, uploading and downloading.</p> <ul style="list-style-type: none"> <li>✓ Provide cost effective information.</li> <li>✓ Make profit by doing membership.</li> <li>✓ Cost of team involvement and materials.</li> <li>✓ Avoid data redundancy.</li> </ul>
C	<b>Control</b>	<p>The system is fully protected from the frauds who can attempt to bring errors in the system.</p> <ul style="list-style-type: none"> <li>✓ Privacy steps has been implemented to gain high security.</li> <li>✓ Maintain confidentiality with documents available in each user accounts.</li> <li>✓ Error free decision making processes.</li> <li>✓ Protection of data from unauthorized access.</li> </ul>
E	<b>Efficiency</b>	<p>The system is developed efficiently without wasting time, money and energy.</p> <ul style="list-style-type: none"> <li>✓ It provides a better way of online recruitment.</li> <li>✓ Easy to surf the website.</li> <li>✓ Easy to handle problems while running the system.</li> </ul>
S	<b>Service</b>	<p>It offers reliable services to the users.</p> <ul style="list-style-type: none"> <li>✓ Easy to search and tie up.</li> <li>✓ It is compatible with other system.</li> <li>✓ It is a user friendly system.</li> </ul>

### 2.3.3 Economic Feasibility

In economic feasibility, the developer needs to find out the financial benefits and costs associated with the project. Basically, the developer is required to perform cost-benefit analysis. The project cost may lie in hardware and software purchase and installation, development cost and maintenance cost.

The developer should have their personal computers or laptops with the required software installed and hardware specification components in it. But, there is no need of purchasing new hardware devices for the development of the project as the developer is having her own system and required software are open source. The following points are highlighting some economic feasible for the project.

- ✓ **Cost of research:** In finding various information related to the project we will be getting cost of internet.
- ✓ **Cost of analysis:** Dependent on group member's awareness.
- ✓ **Cost of designing:** Installation of paint software and MS Office.
- ✓ **Cost of implementation:** Installation of the required environment for the development of web application.
- ✓ **Cost of maintenance:** Soft copy (CD) and hard copy (documentation).

### 2.3.4 Schedule Feasibility

The developer has done the schedule feasibility by making different modules as milestones of the project. There are allotted time for each phases and the modules of the system and it is mandatory to complete the project within the time duration. The final project is required to submit in the give deadline.

Hence, the developer has created the Gantt chart to schedule the tasks according the given deadline. Gantt chart is a type of bar chart that describes the start and finish dates of the main elements and sub elements of a project and shows the total duration of the project.

**Note:** *The Gantt chart has been attached in the appendix of the document.*

## 2.4 Existing Constraint of the Problems

Everything in this world have some limitations where we can't use that thing. The Earning Explorer also deals with certain limitation in its application. These are as follows.

1. **Interface:** The proposed system is web based application only. It will not be applicable as desktop application. Hence, it is interface specific.
2. **Internet:** User will be able to run the system with internet connection only.
3. **Search Term:** Users are required to know how to extract the appropriate search results that means the way of writing search query.
4. **Email Account:** Users will not be able to register their account in the system until they don't have any email account.

## 2.5 Problem Assumption

The system Earning Explorer has been framed by keeping some assumption for the successful implementation of the application. The assumptions made in the system are being shared below.

1. **Literacy:** The users of the system are computer literate and understand the simple English language.
2. **Internet:** The users must have internet connection.
3. **Email Security:** The user email ids must have enabled the option of security of receiving email from any source i.e. 'less secure app' feature so that the email sent by the system should be received in the email inbox.
4. **Email Account:** Users will not be able to register their account in the system until they have Gmail account.

## 2.6 Challenges

The system's domain and the main challenges of the system are to generate search results and plus point calculation with a high accuracy. The developer needs to keenly observe the algorithm construction and application throughout the system process. The system should be developed with the noteworthy quality procedures and user friendly.

There are various leading challenges in the development of the system:

- Sending mail to user email address through Java Mail Server (SMTP).
- Understanding the HDFS (Hadoop Distributed File System) cluster. As the system is based on Hadoop streaming so the developer should have vast knowledge and experience of working on it in order to have efficient file transfer processes.
- Data collection, mining and handling. The structured data needs to travel into the system and analyzed on various stages of the system.
- To have a check on precise and updated output of user's search. Users should receive updated and organized results related to the item searched.
- Performing prediction using user responses. The system will collect the user's response in a structured form and apply the Hadoop MapReduce analysis for identifying the user acceptance on the asked subject.
- Visualization of the analyzed data will be done by collaborating the analysis with the web application interface.
- The system is required to analyze the user search behavior and the related data mining should be done.
- Bootstrap and JSP to make a user friendly web application.
- Algorithms to calculate earnings for a customer.

## 2.7 Conclusion

The chapter Problem Description has delivered the deep explanation unambiguously on the problem area. For the successful implementation of the proposed project to solve the problem, the developer has done the feasibility study. The evaluation is on technical feasibility, operational feasibility, economic feasibility and schedule feasibility. The developer has schedules the project through Gantt chart to avoid project failure. There are various challenges of the system that have been discussed in the chapter as a checkpoint for the system.

## CHAPTER 3

# LITERATURE REVIEW

### 3.1 Introduction to Literature Review

Literature review is the process of reviewing the research works done by the researchers on a particular topic that has been published. Occasionally, every things is documented as a part of an essay or thesis related to the assigned subject. In the documentation of literature review of a topic, the knowledge and ideas regarding that topics has been conveyed to the expected readers. It guides the users on the problems, strengths and weaknesses regarding the matter. The reviewer enhances their skills like as follows. (*Taylor, 2016*)

- 1. Seeking information:** The reviewer can improve the ability to examine the literature efficiently. They are able to use different techniques to categorize the set of useful elements from the literature.
- 2. Analysis:** The reviewer can analyze the collected information by following the principles of analysis. Literature review helps in analyzing the topic in a deeper sense.

The developer has done this review by having massive research to find the basic and important terms and functions that has been used in different search engines or earning web applications. Through this literature review, the developer is going to covey the knowledge and the ideas established on the proposed system to the readers.

### 3.2 Advance Preliminary Research

The idea of search engine with earning facility has been explored by the developer after going through the problems regarding an existing search engine. The concept of earning with any basic system was only invented to provide the part time earnings to the users. Hence, the developer has understood the complications available in the existing system and solutions to those complications has been discovered on the basis of existing knowledge of the developer concerning various software development technologies. The developer attains the knowledge of Hadoop technology and Java Server Page with Bootstrap.

---

The research has been done for the development process that is to be included in searching engine with Hadoop streaming. The techniques that have been decided to overcome the topic's problem context are described below.

1. Information retrieval architecture from big data.
2. Test mining analysis on user responses.
3. A Hadoop framework require to process the big data very easily and efficiently.
4. Bootstrap and JSP to make a user friendly web application.
5. Algorithms to calculate earnings for a customer.

### 3.2.1 Big Data

**According to the research,** Big Data are the large set of data that are so complex and difficult to manage with the traditional approaches of data processing. There are various big data challenges such as storage, analysis, transfer, visualization, searching, updating and privacy.

These data sets have been classified into three categories:

**Structured data:** Structured data is very easy to predict and manage. It can be handled by storing the data in database with table rows and columns. They can be mapped by using relational keys into pre-designed fields. It represents 5 to 10 % of whole data existence. (*Ronk, 2016*)

**Semi Structured Data:** Semi-structured data are those kind of data which doesn't reside in a relational database management system. It have some structural properties due to which it is easier to analyze. With some managerial processes we can store these data in relation database. Semi structured data also represents 5 to 10 % of whole data existence. (*Ronk, 2016*)

**Unstructured Data:** Unstructured data exist around 80% of whole data. It includes text and multimedia contents for example mails, documents, videos, audio, images and web pages. These set of data may have some core structure but not understandable. They doesn't fit precisely in a relational database. (*Ronk, 2016*)

**In the proposed project,** all these types of data will be converted into a unique format which would be similar to structured data format. SQL queries can be easily performed on this unique data to retrieve any information from the given data set.

### 3.2.2 Critical Research on Big Data and Hadoop

After reading the below mentioned excerpt from a journal, the developer came to a decision to implement Predictive Analysis on the big set of data and do visualization of the analyzed data in the system. In the present scenario large amount of data is generating every second of the day and managing those data is becoming so difficult. The data is waste if it is not useful. Hence, to retrieve a meaningful information from the big data the developer has planned the system accordingly.

Resource citation: (*Bhosale and Gadekar, 2014*)

**Table 7: Journal Excerpt**

<b>A Review Paper on Big Data and Hadoop</b>	
<b>Harshawardhan S. Bhosale, Prof. Devendra P. Gadekar</b>	
<p><b><i>Mrigank Mridul, Akashdeep Khajuria, Snehasish Dutta, Kumar N. et.al</i></b> did the analysis of big data he stated that Data is generated through many sources like business processes, transactions, social networking sites, web servers, etc. and remains in structured as well as unstructured form . Today's business applications are having enterprise features like large scale, data-intensive, web-oriented and accessed from diverse devices including mobile devices. Processing or analyzing the huge amount of data or extracting meaningful information is a challenging task. The term "Big data" is used for large data sets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time. Big data sizes are a constantly moving target currently ranging from a few dozen terabytes to many Peta bytes of data in a single data set. Difficulties include capture, storage, search, sharing, analytics and visualizing. Typical examples of big data found in current scenario includes web logs, RFID generated data, sensor networks, satellite and geo-spatial data, social data from social networks, Internet text and documents, Internet search indexing, call detail records, astronomy, atmospheric science, genomics, biogeochemical, biological, and other complex and/or interdisciplinary scientific project, military. Surveillance, medical records, photography archives, video archives, and large-scale ecommerce.</p>	

### 3.2.3 Critical Research on MapReduce Analysis

After reading the below mentioned excerpt from a journal, the developer came to a decision to implement analysis through MapReduce. It is a framework which can handle a big data set file and produce output very quickly. It comprises map and reduce function that works parallel which allow the program to generate output in fewer time duration. The programming model works key value pair at the time of mapping and then merges the key and value in reducing. Hence, to utilize the resources the programmers can easily work on MapReduce.

Resource citation: (**Dean and Ghemawat, 2004**)

**Table 8: Research Excerpt on MapReduce**

<b>A Review Paper on MapReduce: Simplified Data Processing on Large Clusters</b>	
<b>Jeffrey Dean and Sanjay Ghemawat</b>	
<p>MapReduce is a programming model and an associated implementation for processing and generating large data sets. Users specify a <i>map</i> function that processes a key/value pair to generate a set of intermediate key/value pairs, and a <i>reduce</i> function that merges all intermediate values associated with the same intermediate key. Many real world tasks are expressible in this model, as shown in the paper. Programs written in this functional style are automatically parallelized and executed on a large cluster of commodity machines. The run-time system takes care of the details of partitioning the input data, scheduling the program's execution across a set of machines, handling machine failures, and managing the required inter-machine communication. This allows programmers without any experience with parallel and distributed systems to easily utilize the resources of a large distributed system. Our implementation of <i>MapReduce runs on a large cluster of commodity machines and is highly scalable</i>: a typical MapReduce computation processes many terabytes of data on thousands of machines. Programmers find the system easy to use: hundreds of MapReduce programs have been implemented and upwards of one thousand MapReduce jobs are executed on Google's clusters every day.</p>	

### 3.3 Critical Research on Existing Systems

#### 3.3.1 Market Research

Research on existing similar products is known as Market Research. The information gathered from market research are analyzed and these analyzed reports helps the organization or an individual to make appropriate and profitable decisions for them. It is the combination of social and opinion research. (*Esomar, 2016*)

The developer has critically researched on three other related system in order to compare those with the proposed system. The three existing systems are as follows.

#### 1. InboxDollars

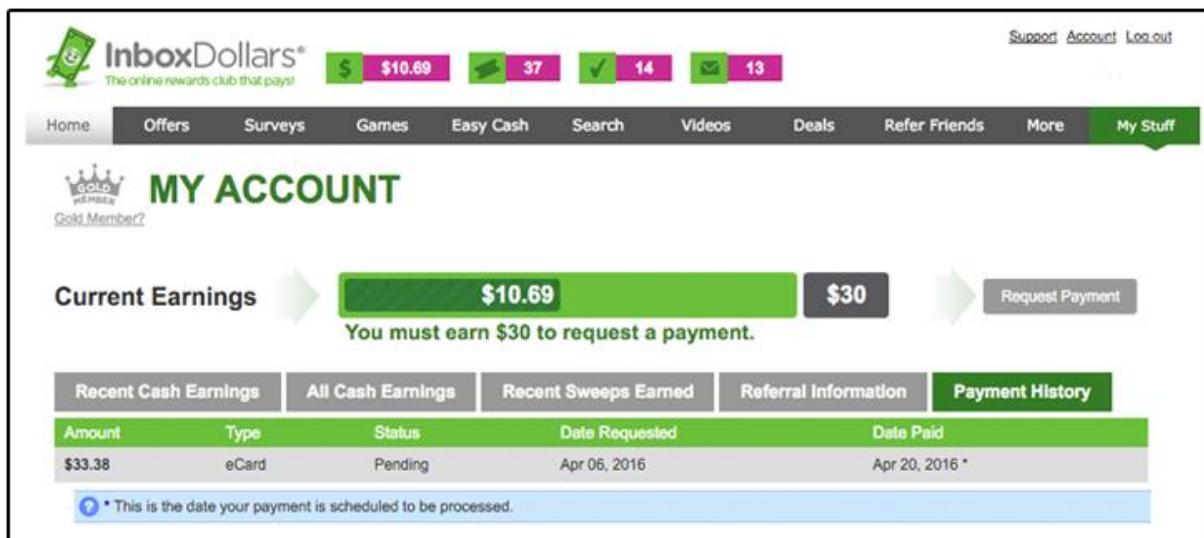


Figure 1: InboxDollars Screen View

Company Terms	Description
<b>Website</b>	<a href="http://www.inboxdollars.com/">http://www.inboxdollars.com/</a>
<b>Purpose</b>	InboxDollars is an earning website through which users can earn by performing several provided tasks in the system.
<b>Functionalities</b>	<ul style="list-style-type: none"> <li>✓ Read Email</li> <li>✓ Perform Surveys</li> <li>✓ Play Games</li> </ul>

	<ul style="list-style-type: none"> <li>✓ Play Videos</li> <li>✓ Search anything</li> <li>✓ Check Deals</li> <li>✓ Refer Friends</li> </ul>
<b>Customer Reviews</b>	<p> <b>214 Reviews</b>  <b>According to</b>  <a href="https://www.sitejabber.com/reviews/www.inboxdollars.com">https://www.sitejabber.com/reviews/www.inboxdollars.com</a></p> <p> <b>14 Reviews</b>  <b>According to</b> <a href="http://www.mouthshut.com/websites/InboxDollars-com-reviews-925775180">http://www.mouthshut.com/websites/InboxDollars-com-reviews-925775180</a></p>

## 2. GrabPoints

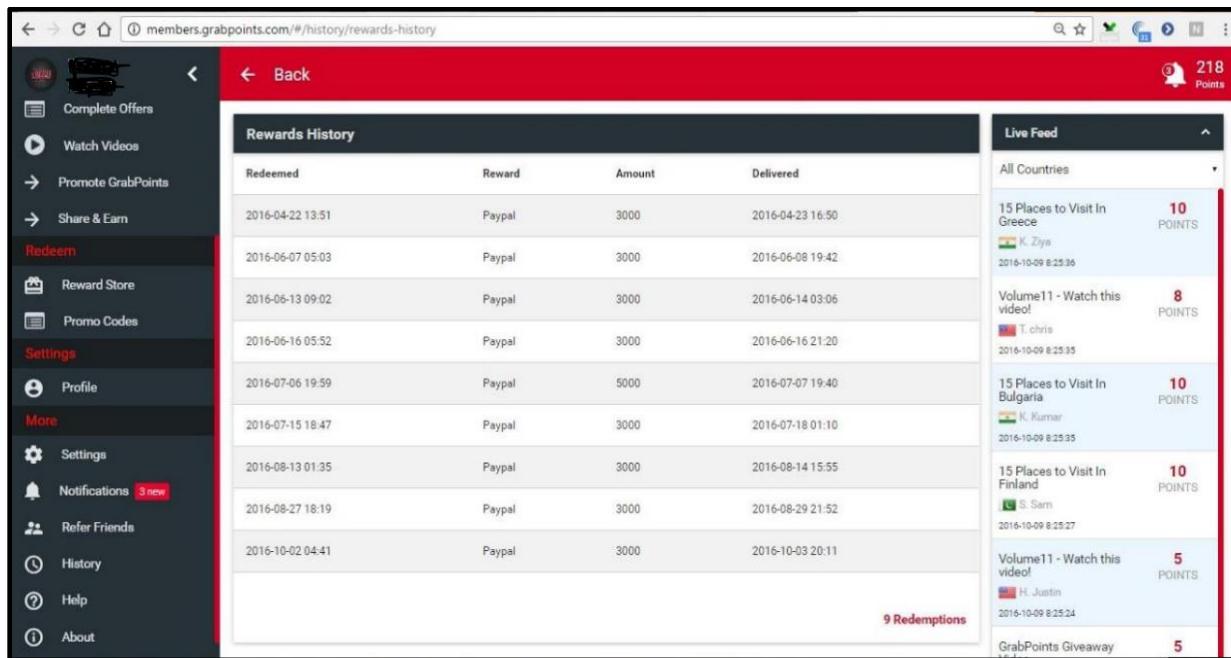


Figure 2: GrabPoints Screen View

Company Terms	Description
<b>Website</b>	<a href="http://www.grabpoints.com/">http://www.grabpoints.com/</a>
<b>Purpose</b>	GrabPoints is an earning website through which users can grab the opportunities to earn more and more.
<b>Functionalities</b>	<ul style="list-style-type: none"> <li>✓ Perform Surveys</li> </ul>

	<ul style="list-style-type: none"> <li>✓ Download Apps</li> <li>✓ Play Videos</li> <li>✓ Check Deals</li> <li>✓ Refer Friends</li> </ul>
<b>Customer Reviews</b>	 <b>According to <a href="http://www.mouthshut.com/review/GrabPoints-review-slpspqunut">http://www.mouthshut.com/review/GrabPoints-review-slpspqunut</a></b>

### 3.2.2 Existing Product Analysis and Comparison

The developer has compared the existing products with the proposed system ‘Earning Explorer’ on the basis of some valuable points of facilities provided by the systems. Results for the existing system and the proposed system has been defined on the basis of market reviews and system requirement specification respectively.

#### Comparing Earning Explorer with InboxDollars and GrabPoints

**Table 9: Comparison of existing products with Earning Explorer**

Terms	InboxDollars	GrabPoints	Earning Explorer
1. Searching facility	Yes	Yes	Yes
2. Displaying update search results	No	No	Yes
3. Searching speed	Slow	No	Fast
4. User interface	Good	Good	Excellent
5. Earning facility	Yes	Yes	Yes
6. Earning rate	Low	Medium	High
7. Survey for earning	Yes	Yes	Yes
8. Watching video for earning	Yes	Yes	Yes
9. Daily meter showing earning goal	No	No	Yes
10. Notifying surveys	No	No	Yes

<b>11. Video user guide</b>	No	No	Yes
<b>12. Security in settings</b>	No	No	Yes
<b>13. Refer friends to earn</b>	Yes	Yes	Yes
<b>14. Redeem points</b>	Yes	Yes	Yes
<b>15. Visualization of results</b>	No	No	Yes

### 3.4 Secondary Research

Secondary research is a data gathering method that researches on other research works. We explore the already published research papers that includes journals, e-books, websites or any existing system. It should be done before primary research.

#### 3.4.1 Academic Research

In the academic research, the developer has gone through many modules related to the software development in order to gain the concepts of different phase included in the development from initial (Planning) to the final (Testing) stage. The below mentioned modules are helpful in managing and increasing the project works.

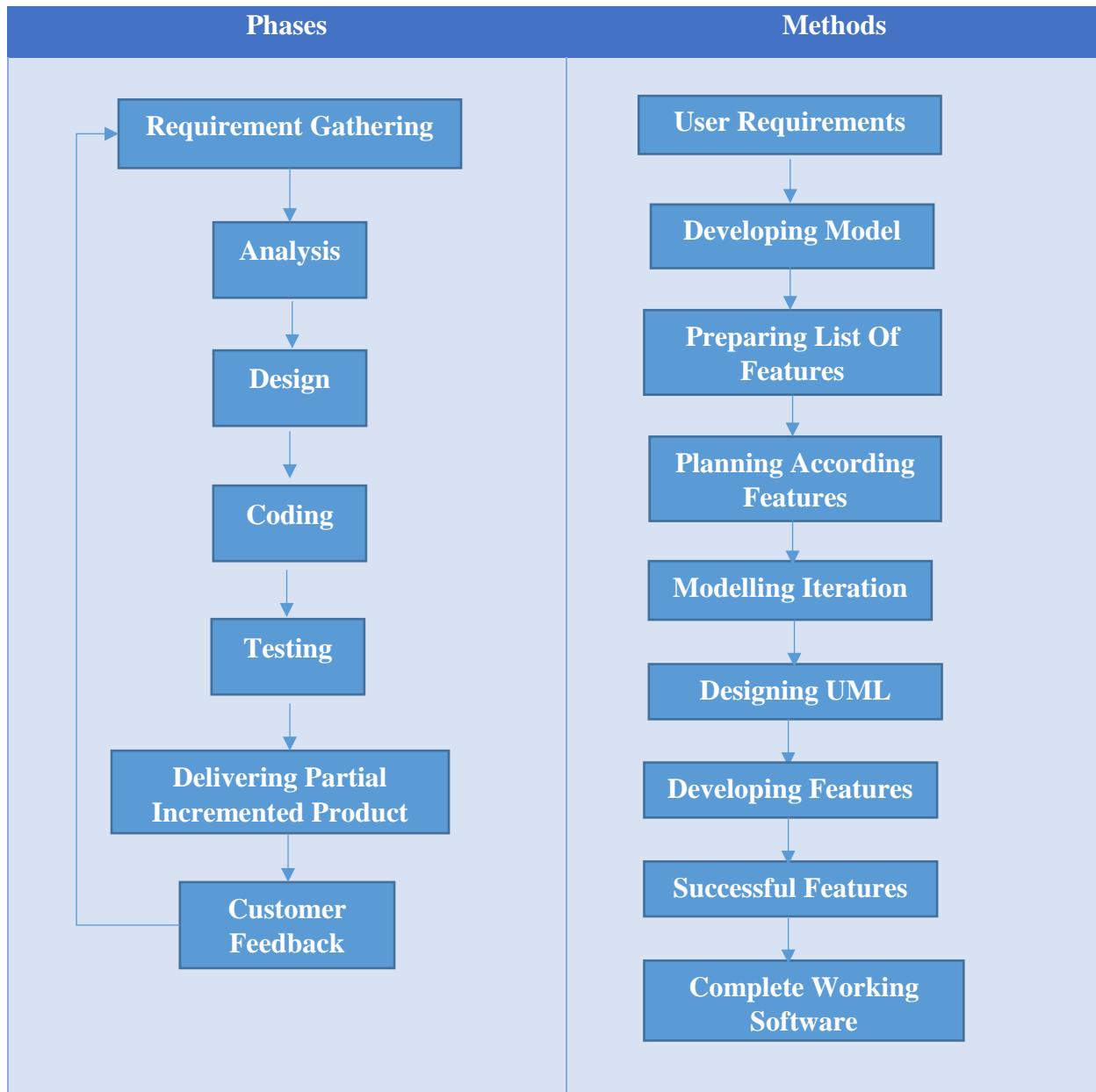
##### 3.4.1.1 Software Analysis and Development

The Software Analysis and Development module helps in identifying the suitable model to work with the proposed project. The developer has evaluated almost every methodologies by doing research and analysis of the software development methodologies. After reviewing the methods, the developer has come up with the decision of selecting Agile Methodology to be followed for the development of proposed system.

**According to the research,** the agile methodology follows an iterative as well as an incremental approach in which any kind of modifications can be done after planning as per client's feedback. The main concentration of agile method is the early release of software product by proper judgements and the changes required in the project's requirement specification. The agile project is much better than other software development process in terms of productivity, performance,

faster time cycles, risk analysis. Agile processes are implemented in several important applications such as web based, testing tools, etc. The below mentioned table is describing the seven phases of the methodology and their respective methods which have been shown with the help of flow diagram. (*Sharma, Sarkar and Gupta, 2012*)

**Table 10: Agile phases and methods**



---

The process includes iteration to deliver different working software anatomy after each iteration of planning, requirement analysis, designing, implementation and testing. The end product of each iteration is incremental in terms of features as the final product holds all the wanted features required by the customer.

The developer has opted for agile methodology for the following reasons.

1. The project's requirements are not clear at the initial stage, hence changes can be done on the basis of research accordingly. This models allows to modify the things unlike the waterfall methodology.
2. Because of iterations done in the process the product is tested very frequently so there is less chance of having any risk in project's future.
3. This model has been undertaken because the project is required to complete within a given time frame.
4. It provides flexibility to the developer to think independently.
5. It focuses on technical excellence and enhanced design for the project.
6. The methodology allows to have more user interaction while the development of the system.

### **3.4.1.2 Human Computer Interaction and Usability Principles (HCIU)**

It is a well-known fact that users love to work on the system's with which they can interact easily and which provides them an enjoyable experience. Hence, the developed product should contain the usability properties. The components of usability are Learnability, Efficiency, Memorability, Error Free, and User Satisfaction. The accessibility principle should be followed in the project that are as follows. (*Usabilityfirst, 2015*)

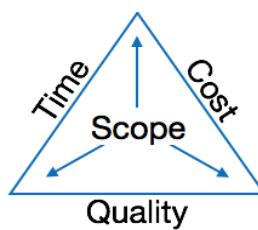
- ✓ Allow for flexibility.
- ✓ Be simple and intuitive.
- ✓ Minimize errors.
- ✓ Include redundant modality.
- ✓ Avoid side effects on user interaction.

**According to the research,** the system's usability testing should be performed to check the extent to which the system is user friendly. To analysis the usability of the system in terms of learnability, efficiency, memorability, error free and satisfaction.

### 3.4.1.3 Project Management

“Software is an intangible product to meet the user’s requirements regarding some problems.”

The most risky part of any software is that the technologies are frequently getting changed and advanced. The experience of any product are no longer in use. These all environments and constraints leads to different risks in the software development. Hence it is important to manage the software projects efficiently. (*tutorialspoint, 2016*)



**Figure 3: Software Project Management**

The image is explaining the triple constraints for any kind of software projects. Any software developer is essentially required to deliver a quality product with cost under client’s budget and should be delivered on time. Any of the three factor can severely impact the other two because of several factors either internal or external. Therefore, software project management is essential to incorporate user requirements along with budget and time constraints. (*tutorialspoint, 2016*)

**According to the research,** the whole planning and risk assessment of the system development should be very much effective so that project should not fail in management. The scheduling should be done in detail with following the time, cost and quality scope.

### 3.4.2 Technical and Domain Research

#### 3.4.2.1 Hadoop

Hadoop is an open-source framework to store and process big data in a distributed environment using simple programming representations. The Hadoop software has been introduced entirely new technology for storing and processing data. Users are able to install their complete hardware and software along with the OS and Hadoop software across the entire cluster and can manage the full cluster by sitting at one place through a single management interface. (*Mapr.com, 2016*)

### 3.4.3 How is Hadoop Different from Past Techniques? (*Mapr.com, 2016*)

- ✓ Hadoop can handle data in a very fluid way: Hadoop is working as a cheaper and faster database as well as analytics tool in the market. Unlike relational databases, users are allowed to upload the unstructured data into the framework.
- ✓ Hadoop has a simplified programming model: Hadoop is having simplified programming modelling features that provides its users a fast write and test software or any data in the distributed file systems. The scripting language that have been used on a large scale for data preparation work are Hive, Pig or Python.
- ✓ Hadoop is easy to administrate: Hadoop is capable in handling the data intangibly. It controls the issues in way that the every running computations are should not be lost.
- ✓ Hadoop is agile: Hadoop is rich with features to economically and reasonably process and analyze large amount of both data format structured and unstructured together. It is cheap in terms of cost and time both.

### 3.4.4 Java Server Pages (JSP) and Bootstrap

- ✓ Through Bootstrap developer can make the user interface of the system more interactive and responsive to the users by using CSS and JS files.
- ✓ JSP programing language will be used to make pages interact to each other efficiently. Java programing is more supportable with Hadoop technology.

### 3.4.5 Predictive Analysis

To perform Predictive Analysis in the system, the developer needs to go through different sub processes. The Predictive Analysis includes the process of data collection, data analysis, statistics, modeling, deployment and model monitoring. (*Predictive Analytics Today, 2016*)

- ✓ **Data Collection:** It is data mining to arrange data from multiple sources to do analysis on it.
- ✓ **Data Analysis:** It embraces the process of inspecting, cleaning, transforming and modelling data to make a conclusion from it.
- ✓ **Statistics:** It validates the assumption and test them by using standard statistical models.

- ✓ **Modeling:** It is the process of creating accurate predictive models about future.
- ✓ **Deployment:** It offers to deploy the analytical result in to the daily decision making process.  
The decisions are based on modeling.
- ✓ **Model Monitoring:** This process manages monitoring of model performance to ensure that it should provide the expected results.

### 3.5 Conclusion

The chapter Literature Review has enlightened the research work for key features of the proposed system by doing market review and secondary research. It has cleared the knowledge about the functionalities, selected methodologies and technologies. The agile project is much better than other software development process in terms of productivity, performance, faster time cycles, risk analysis.

The developer is expected to implement the system on mapping the requirements to the research from academic as well as from domain.

## CHAPTER 4

# RESEARCH

### 4.1 Data Gathering Techniques

Data gathering is the base of any kind of assignment. For nurturing an idea we need data and then establishment takes place. Major part of a project depends on information relative to a specific purpose. It helps in collecting data for initiating a task by applying different methods. Basically, there are two ways of data gathering methods:

1. **Primary Research:** It is an own research that is done by putting on own mind and skills.
2. **Secondary Research:** It is a research that is done on other research. We explore the already published research papers that includes journals, e-books, websites and any existing system. It should be done before primary research.

### 4.2 Primary Research

#### 4.2.1 Questionnaire

Questionnaire covers a large part of audience and helps in gathering more data in one time. It is easy to evaluate to seek more relevant facts. The questions asked in questionnaire are technical and domain based, user interface and user satisfaction. The questionnaire will be distributed to different occupation holder and age group candidates in order to get the most precise results. The result will be beneficial in further development phases.

**Justification:** This research technique allows the developer to capture a large number of user from the target audience. If the responses will from more and accurate respondents then the project can gain the positive results for the functionalities of the system. The developer will be able to perform the proper analysis from the gathered information over the topic.

The analysis of each question included in the questionnaire will help in identifying the requirements change. Questionnaires have been distributed to the customers who are actually going to be affected by the system. The sample of the questionnaire shown below.

**Note:** *The filed up questionnaire has been attached in the appendix of the document.*

## QUESTIONNAIRE

### EARNING EXPLORER BASED ON HADOOP STREAMING

Dear Respondent,

Questionnaire ID: EEQN001

I am Shivani Sinha, a final year student of BE from Asia Pacific Institute of Information Technology, wishes to improve customer satisfaction regarding search engine facility as a part of research work for Final Year Project. Through this brief survey on existing search engines, I as the student researcher and developer, desires to collect the potential user's responses about the topic so that better services can be provided to them. The resulting statistical data will support in developing the Search Engine as per usability goals and customer's need in terms of awareness and capital. If you are uncomfortable or you have any questions regarding the survey, you can drop a mail on **sinha.shivani02@apiit.edu.in**. Your response will only be used for the survey purposes. The developer is very thankful to you for spending your precious time and suggestions in this questionnaire.

The questionnaire will take 10 to 15 minutes to complete your responses for the given set of questions. You can return it back in distributor's hand after comprehending. Your personal details are not going to be disclosed anywhere (internet or any website). The extracted data will assist in research works and for developing a better product for the customer. Your private data are with us till the duration of Final Year Project gets completed.

#### SECTION A

The sections is asking for your personal details. It will assist us in checking who are the respondent and how much the received data are accurate.

<b>Gender</b>	<input type="radio"/> Male	<input type="radio"/> Female
<b>Age</b>		
<b>Occupation</b>		
<b>Email</b>		

## SECTION B

The section is comprised of objective type questions that allows you to answer the questions from the given set of options. Please select the most suitable option for your answer.

1. How often do you use internet? *Answer in an average usage.*

- Less than or equal to 2 times in a day
- Less than or equal to 5 times in a day
- Less than or equal to 2 time in a week
- Less than or equal to 5 times a week

**Justification:** If the user uses the internet maximum number of time in a month then he or she will be capable in providing more accurate reviews. Hence, the developer can do better analysis of user requirement.

2. When you are connected to internet, for how long do you use it? *Answer in an average usage.*

- 0–1 Hours
- 1-2 Hours
- 2-4 Hours
- More than 4 Hour

**Justification:** If the user are using for more hours then there is high probability that they are doing something productive. These users can understand the concepts asked in the question and can provide proper response. Hence, the developer can do better analysis of user requirement.

3. Are you aware about the facilities that provides you to earn online (On internet)?

- Yes
- No

**Justification:** This question will let the developer know how many users are knowing that they can earn by doing different type of jobs on internet.

4. Do you use search engine? *If yes, please specify the name(s) of the search engine and move to further questions else directly move to question number 10.*

- Yes
- No
- Don't Know

**Justification:** This question will let the developer know how many users are using search engine. The users who are using search engines can report better and true information.

5. What do you search mostly on search engine? *You can select more than one option.*

- Academic
- Online Shopping
- News
- Social Networking
- Videos or Music
- Gaming
- Others, please specify

**Justification:** The developer wants to know for how many purposes users are using search engine. More purpose then automatically more hits to search engine and if more hits then definitely they have more interacted with the search engine.

6. On which results do you click first from the presented set of results on search engine? *If you go beyond the first page, then please mention up to which page number you visit.*

- 1<sup>st</sup> result only
- 1-3 results only
- 1-6 results only
- 1-10 results only

**Justification:** The question will help in analyzing the customer's responses to the search. If they visit the links presented at the top only then the company should understand the customers and make their important links at the top for receiving good outcomes to the company's product.

7. How often do you search for images or videos?

- Less than 25 % of time
- Less than 50 % of time
- Less than 75 % of time
- More than 75 % of time

**Justification:** The responses will outline the percentage of demand of image and video search. If the demand is more then the developer should implement an efficient search ability in the system.

8. Are you satisfied with the search results provided on the search engine? *Please rate your satisfaction level on the scale of 5.*

- Outstanding
- Very Good
- Good
- Fair
- Marginal

**Justification:** The ratings will describe how much the users are satisfied with the existing search engines. Hence the developer can analyze the improvements needed in the existing system so that those improvements can be applied in the proposed system.

9. Do you feel that you receive updated search results? *If no, please specify the reason.*

- Yes
- No

**Justification:** If maximum no then there should be some indication or feedback about the updated search on the result pages of the search engine.

10. Are you satisfied with the bookmark utility provided in the web browser? *If no, please specify the reason.*

- Yes
  - No
- 

**Justification:** If users are satisfied then it's good to have bookmark utility with a search engine and if they are not satisfied then an upgraded version of bookmark should be implemented.

11. Have you ever gone through any type of surveys?

- Yes
- No
- Don't Know

**Justification:** Through this question the developer will find out, how many users are aware about the surveys and have taken part in it.

12. How will you feel if the analysis of survey will be displayed in an organized form of graphs?

- Outstanding
- Very Good
- Good
- Fair
- Marginal

**Justification:** Through this question the developer will get the reactions of the users about the visualization of any kind of analysis. Hence, the visualization can be implemented in a more advance way like charts or on world map.

13. How will you feel if you are able to earn money while you are searching queries or attending surveys on a search engine?

- Outstanding
- Very Good
- Good
- Fair
- Marginal

**Justification:** Through this question the developer will get the reactions of the users about the earning facility provided to them in a search engine.

## SECTION C

The section is comprised of subjective type questions. Questions are related to your personal experience or feelings.

14. Please share your views about the search engine?

---

---

---

**Justification:** It will help in serving the customized system according to prescribed choice.

15. Would you like to suggest any area through which users can earn online?

---

---

---

**Justification:** It will help in serving the customized system according to prescribed choice.

Thank you for participating in the survey and providing valuable answers. Please check whether you have answered every questions as per your choices.

Have a nice day!

#### 4.2.2 Interview

The interview is a technique of data gathering which involves two or more personal opinions. The persons are those who have active participation for a particular topic so that they can inform better about the subject. It is done with a high level security in between of interviewee and interviewer. The provided information will be collected by the developer for further discussion to analyze the user responses and no other person can see those information. It will help in enhancing the system efficiency.

**Justification:** The techniques has been selected to collect the more detailed data about the system requirements and issues. The developer has designed the set of question to be asked from specific peoples so that the system can get the proper idea and direction to get implemented. From one to one interaction there is a high chance of getting the actual reactions of the customer regarding the problem. The developer will create the comfort zone for the interviewee as an interviewer.

Hence, the technique may take a lot of time to be conducted as it includes interviewing many users in detail. The developer will get help in creating the plans for the development of the project. So the developer has organized the meetings with the targeted audience to fetch a concerned and profitable output from the meetings. The sample of the interview questions shown below.

**Note:** *The filed up interview question sheet has been attached in the appendix of the document.*

### INTERVIEW

#### EARNING EXPLORER BASED ON HADOOP STREAMING

Dear Respondent,

I am Shivani Sinha, a final year student of BE from Asia Pacific Institute of Information Technology, wishes to improve customer satisfaction regarding search engine facility as a part of research work for Final Year Project. Through this interview, I as the student researcher and developer, wishes to collect the potential user's responses about the topic so that better services can be provided to them. The resulting statistical data will support in developing the Search Engine as per usability goals and customer's need in terms of awareness and capital.

If you are uncomfortable or you have any questions regarding the survey, you can ask at any time during the interview. Your response and personal details will only be used for the research purposes. The developer is very thankful to you for spending your precious time and suggestions.

## QUESTIONS ASKED

### **1. What is your opinion about the existing search engine and their search results?**

**Justification:** This question has been asked for getting the percentage of acceptance among different age group of users in order to find out the different user ratings regarding these services provided by the existing search engine.

### **2. How do you search for videos? Do you face any problem in searching?**

**Justification:** The developer can judge the user satisfaction regarding the video results they are getting from any portal. The search results for videos can be organized in a search engine itself according to user's expectations.

### **3. How do you feel about the facilities that provides you to earn online?**

**Justification:** This question will let the developer know how many users are knowing that they can even earn by doing different type of jobs on internet and what is their opinion about it.

### **4. How is your experience about the online surveys?**

**Justification:** This question will let the developer know how many users are knowing about the surveys taken online on any of the products launched in the market.

### **5. How will you feel if the analysis of survey are displayed in an organized form of graphs?**

**Justification:** Through this question the developer will get the reactions of the users about the visualization of any kind of analysis. Hence, the visualization can be implemented in a more advance way like charts or on world map according to user's acceptance.

**6. How much you are satisfied with the bookmark utility provided in the web browser?**

**Justification:** If users are satisfied then it's good to have bookmark utility with a search engine and if they are not satisfied then an upgraded version of bookmark should be implemented as per the user's requirements.

**7. What do you want to say if you will get to earn money while you are searching queries or attending surveys on a search engine?**

**Justification:** Through this question the developer will get the reactions of the users regarding the earning facility provided to them in a search engine. Can figure out the opinions if they can earn money by performing online activities on a search engine.

**8. Would you like to give some suggestions for improving the existing system?**

**Justification:** The suggestions can help in developing a better system from the existing one by fulfilling all user's requirements. The developer can note down the important points to make those suggestion implemented in the system.

Thank you for participating in the survey and providing valuable answers.

Have a nice day!

## 4.3 Technical Research

### 4.3.1 MapReduce Algorithm

**According to the research**, the main challenge is to analyze the big data. For analyzing such a big amount of data the developer is required to apply MapReduce algorithm or framework. MapReduce is a kind of programming model through which developer can efficiently develop the accessible parallel applications in order to process big data on large clusters of service machines. (*Dean and Ghemawat, 2004*)

MapReduce jobs are divided into two parts. The Map function and the Reduce function. The Mapper class splits the request into multiple division and then process data at the node level. The Reducer class aggregates the output of the Mapper class. So, MapReduce includes the key value pairs. (*Dean and Ghemawat, 2004*)

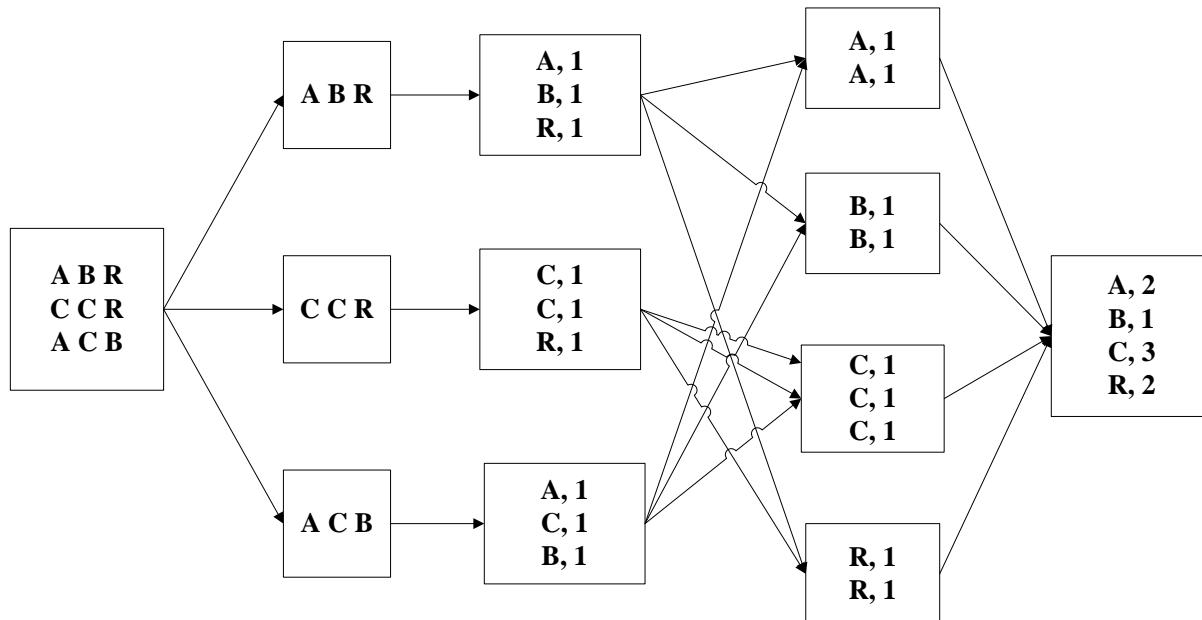


Figure 4: MapReduce Algorithm

### 4.3.2 Database Research

For analyzing the suitable data for the evaluation on big data, the developer has compared the Hadoop with the Relational Database Management System (RDBMS) in context of the project that is based on search engine with different algorithms for earning calculation. The table mentioned below is showing the comparison. (*bigdatanerd, 2016*)

**Table 11: Comparison between Hive and RDBMS**

Terms	HADOOP	RDBMS
<b>1. Secure the failed data stored in the system</b>	Yes	No
<b>2. Process the data faster</b>	Yes	Yes
<b>3. Processing semi-structure data</b>	Yes	No (works with structured data)
<b>4. Normalization</b>	No (Due to duplicate data at several nodes)	Yes
<b>5. Linear scalability</b>	Yes	No
<b>6. Read and Write</b>	Read often and write one time a while	Read and write often
<b>7. Integrity</b>	Low	High
<b>8. Recommendation</b>	Hadoop is good when we have a Peta bytes of semi-structure or unstructured	RDMBS is good when we have a Gigabytes of structured data.

**Justification:** After the comparison, the developer has confirmed that the Hadoop technology will be best suited with the proposed project as the project requires a large set of data. The data may be of any type structured, semi-structured or unstructured formats. The technology is also supporting with the security and the processing of data. So, the developer has concluded the database research with the result of using Hadoop for developing the system.

#### 4.3.3 Programming Language Research

For analyzing the suitable language for the creating working pages of the project and linking them together, the developer has compared the C++, C# and JSP programming languages in context of the project requirement. The table mentioned below is showing the comparison. (*Alshamrani and Bahattab, 2015*)

**Table 12: Comparison between C++,C# and Java**

Terms	C++	C#	Java
<b>1. Object Oriented</b>	Yes	Yes	Yes

<b>2. Classes/ Methods</b>	Variable/	Yes	Yes	Yes
<b>3. Multithreading</b>		Libraries	Yes	Yes
<b>4. Language Integration</b>		C and Assembler	All .Net languages	C and some feature of C++
<b>5. Hadoop</b>		No	No	Yes
<b>6. Built in security</b>		No	Yes	Yes
<b>7. Data Storage</b>		File System	Database Management	Database Management

**Justification:** After the comparison, the developer has got the clear result that Java is the best suited programming language for the proposed project implementation. By using different classes and methods the project functionalities will be working to get the input and output from the users. Advance Java concept will also be used in order to develop web application as Core Java is meant for desktop application only. So, the developer has concluded the core java and advance java concepts (Java Server Pages) should be implemented and checked for the compatibility.

#### 4.3.4 Methodology Research

For analyzing the suitable methodology to be followed throughout the project development, the developer has research about different methodology and did comparison on them. The table mentioned below is showing the comparison.

Table 13: Comparison between Waterfall, Spiral and Agile

Features	Waterfall	Spiral	Agile
<b>1. Requirements Specification</b>	In the beginning.	Frequently changes according to users.	Frequently changes according to the users.
<b>2. Project Duration</b>	Best suited for short term.	Best suited for long term.	Best suited for both short and long term.
<b>3. Complexity</b>	Easy to understand	Process is complex.	Easy to manage.
<b>4. Flexibility</b>	Not flexible.	Less flexible.	Flexible.

<b>5. Cost</b>	Not costly.	Costly.	Costly.
<b>6. Risky Projects</b>	Doesn't support.	Supports high risk projects.	Supports high risk projects.
<b>7. Success Rate</b>	Low	High	High
<b>8. Customer Involvement</b>	At the end of the project.	After each iteration but less involvement.	Customer evaluation, after each iteration.
<b>9. Testing</b>	At final stage.	After every phase.	After every iteration.
<b>10. Maintenance</b>	Hard to maintain.	Maintenance is required.	Easy to maintain.
<b>11. Implementation</b>	Easy.	Complex.	Easy.

**Justification:** The Agile methodology has been examined well with respect to the project's requirements and it is found that it can associate with the constraints of the project. This methodology follows an iterative as well as an incremental approach in which any kind of modifications can be done after planning as per client's feedback. The main concentration of agile method is to early release of the software product by responding to adopt the changes required in specification.

#### 4.4 Development Plan

Calculating the duration for each phase of the development has been allotted as per the iterations followed in Agile Methodology.

**Table 14: Test Strategy Schedule**

TEST STRATEGY SCHEDULE			
Starting Date		1, August, 2016	
Ending Date		18, April, 2017	
Total Duration		33 Weeks	
S.no.	Phases	Tasks	Duration
1.	<b>Project Definition</b>	1. Search a topic 2. Idea Generation	42 Days

		3. Project Title Selection 4. Draft Proposal Form	
2.	<b>Project Planning (Iteration)</b>	1. Schedule and Time Estimation 2. Gantt Chart 3. Project Proposal Form	6 Days
3.	<b>Requirement Analysis (Iteration)</b>	1. Identify project specifications <ul style="list-style-type: none"> <li>✓ Project Background</li> <li>✓ Project Limitation and Assumption</li> <li>✓ Techniques to be learnt</li> <li>✓ Goals, Objectives and Scope</li> <li>✓ System Functionalities</li> <li>✓ Methodology to be chosen</li> </ul> 2. Project Specification Form 3. Feasibility Study 1. Investigation and Research <ul style="list-style-type: none"> <li>✓ Academic Research</li> <li>✓ Primary Research</li> <li>✓ Secondary Research</li> <li>✓ Questionnaire</li> <li>✓ Interview</li> </ul> 2. Analysis <ul style="list-style-type: none"> <li>✓ Questionnaire</li> <li>✓ Interview</li> <li>✓ Overall conclusion</li> </ul> 3. Mid-point submission.	41 Days
4.	<b>System Design (Iteration)</b>	1. UML Design <ul style="list-style-type: none"> <li>• Use Case</li> <li>• Data flow Diagram</li> <li>• Class Diagram</li> </ul>	38 Days

		<ul style="list-style-type: none"> <li>• Sequence Diagram</li> </ul> <p>2. Database Design</p> <ul style="list-style-type: none"> <li>• ERD</li> <li>• Normalization</li> </ul>	
5.	<b>Implementation (Iteration)</b>	<p>1. Code generation</p> <p>2. Module integration</p>	50 Days
6.	<b>Testing (Iteration)</b>	<p>1. Prototype Evaluation</p> <p>2. Unit Testing</p> <p>3. Integration Testing</p> <p>4. System Testing</p> <p>5. Functional Testing</p> <p>6. User Acceptance Testing</p> <p>7. Compatibility Testing</p> <p>8. Runtime Testing</p> <p>9. Configuration Testing</p>	32 Days
7.	<b>Project End</b>	<p>1. Documentation</p> <p>2. Finalizing Hard Copy and Soft Copy</p> <p>3. Presentation</p>	23 Days

#### 4.5 Conclusion

The chapter Research has been done to collect a huge amount of information from the expected users about the system. The information has been collected by applying two data gathering methods that are Questionnaire and Interview. The developer has made a sample of questionnaire and interview question sheet do that that can be examined by distributing among the appropriate users.

## CHAPTER 5

# ANALYSIS

### 5.1 Analysis

The analysis is the important phase for any software development project in order to evaluate the user responses for the system so that requirements should be achieved. The researcher has done the primary research by conducting questionnaire and interview process to gather information regarding existing systems and get suggestions for the proposed system. The overall conclusion for the analysis of the questionnaire and interview collected data has been provided in this chapter.

**Note:** *The analysis of the questions of questionnaire data and interview data has been done and attached in in the appendix.*

### 5.2 Analysis of Questionnaire

The analysis has been performed on the responses given by the customers in the form chart to make it better to understand. The questionnaire has been attended maximum by the students of age group 12 to 24 which results in an accurate data as they are the maximum user of internet and search engine. After doing analysis, the researcher has planned to develop a full functional and efficient search engine with different tasks to earn money as the users are not satisfied with the functions of existing systems. The proposed system should cover every feature in the system to cover the user requirements and user satisfaction.

Each question response itself explaining the acceptance level of the user. The following table is representing the summarized analysis of the collected data for each questions. The matrix is showing the rating given by each user for each question to conclude the analysis of he asked questions. The ratings has been done on the scale of 10. It includes the responses of 10 précised respondents.

The analysis of the ratings are done on the basis of the examination done on the questions of the questionnaire i.e. attached in the appendix.

**Table 15: Questionnaire Analysis Matrix**

Questionnaire Analysis Matrix														
Attend	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	
1	8/10	5/10	6/10	7/10	8/10	7/10	5/10	3/10	6/10	9/10	7/10	7/10	7/10	
2	7/10	6/10	6/10	8/10	9/10	7/10	6/10	4/10	7/10	8/10	9/10	8/10	10	
3	9/10	4/10	6/10	7/10	9/10	6/10	5/10	5/10	7/10	8/10	8/10	8/10	9/10	
4	7/10	6/10	6/10	6/10	9/10	8/10	7/10	3/10	6/10	7/10	9/10	7/10	8/10	
5	8/10	5/10	6/10	9/10	8/10	7/10	7/10	4/10	5/10	10	9/10	6/10	8/10	
6	8/10	7/10	6/10	9/10	7/10	5/10	5/10	4/10	8/10	8/10	6/10	6/10	6/10	
7	6/10	7/10	6/10	7/10	7/10	7/10	6/10	2/10	8/10	9/10	5/10	9/10	6/10	
8	8/10	6/10	6/10	8/10	9/10	6/10	6/10	5/10	6/10	7/10	8/10	5/10	7/10	
9	7/10	4/10	6/10	8/10	8/10	8/10	5/10	3/10	7/10	9/10	7/10	7/10	7/10	
10	9/10	5/10	6/10	7/10	6/10	8/10	5/10	2/10	6/10	8/10	6/10	8/10	8/10	

### 5.2.1 Overall Conclusion of the analysis

The following points need to be considered during the development of the system.

- ✓ Should build the system in a way that the important updated links should come at the top according to searched query.
- ✓ Required to understand the feeling of the users regarding the search results, what is to be needed to implement so that users can gain trust on the system for receiving updated search results.
- ✓ Should add the bookmark with the search engine.
- ✓ Should do the visualization in a more advance way like charts or on world map.
- ✓ Should develop an efficient system in order to reach the customer satisfaction.

### 5.3 Analysis of Interview

After doing interview with the users, it was clear that they are excited to go through the system and will accept the system so that they can earn while they are performing searches or we can say earning while learning. It would be a unique approach for the students or who are in need to earn money. For collecting appropriate data the interview was conducted with the peoples of age group 12 – 50 because generally they are more connect with the internet and search engine.

The following table is showing the sample of recorded answer of the interview with Ms. Shubhangi Sinha. She is 22 years old and pursuing MBA from SEIS College of Management, Navi Mumbai.

**Table 16: Interview Question Analysis**

QUESTIONS ASKED
<b>What is your opinion about the existing search engine and their search results?</b>
<b>Answer:</b> I am using google search engine and it's quite good. It shows the proper results for whatever we enter in the search bar. It has good way of displaying the results too. But sometimes we will get irrelevant result maybe it's because of word selection for the search. Overall it's an excellent search engine.
<b>How do you search for videos? Do you face any problem in searching?</b>
<b>Answer:</b> I use YouTube to search videos of anything and I am totally satisfied with the results of YouTube.
<b>How do you feel about the facilities that provides you to earn online?</b>
<b>Answer:</b> Sometimes I feel trustworthy and sometimes it seems like fake services. They demand a lot of time to work on it for earning. At last they are paying less money to us according to the time consumed.

### **How is your experience about the online surveys?**

**Answer:** Surveys are good to attend. We get to know about what we were not knowing. Sometimes it takes a long time to complete a survey. If we talk about the surveys from which we can earn, these are also taking a long time to complete and at last paying less money. Most of the earning websites services are like this only.

### **How will you feel if the analysis of survey are displayed in an organized form of graphs?**

**Answer:** It will take less time to understand proficiently rather than proving percentage with the text.

### **How much you are satisfied with the bookmark utility provided in the web browser?**

**Answer:** Bookmarks are working good to save pages for future use. I can easily maintain the desired links without spending time in copying the link and pasting it in a document file.

### **What do you want to say if you will get to earn money while you are searching queries or attending surveys on a search engine?**

**Answer:** It will be a good experience only if the system saves the user time. As these type of services are poor at user's expectation and satisfaction level. It's appreciative for providing money for searching. Users are getting paid for their own work.

### **Would you like to give some suggestions for improving the existing system?**

**Answer:** The existing systems of earning website should provide facilities for user benefits only as they are advertising to the world.

### 5.3.1 Overall Conclusion of the analysis

The following points need to be considered during the development of the system.

- ✓ Should work on information retrieval architecture for heterogeneous big data.
- ✓ Should implement effective algorithms to calculate earnings for a customer.
- ✓ Should add the bookmark with the search engine.
- ✓ Should do the visualization in a more advance way like charts or on world map.
- ✓ Should develop an efficient system in order to reach the customer satisfaction.

### 5.4 Conclusion

The chapter Analysis has documented the detailed analysis of both the used data gathering techniques. From the opinions of the customers given in the questionnaire and the interview, the result has come in the favor of the proposed system. The researcher has done the critical analysis of each questions asked in the in the questionnaire as well in the interview to extract the accurate response of the users.

The user are unsatisfied with the existing system, hence there is a need of efficient system to solve the problems and should reach the level of customer satisfaction. The developer will now take this analysis into consideration in order to develop the proposed system successfully by delivering the application as per the user's requirement and the problem background.

## CHAPTER 6

# SYSTEM DESIGN

### 6.1 Introduction to system design

The system design holds main role in establishing the base interface according to user's requirement. It produces the prototype for the proposed the system. For, Earning Explorer the developer is required to build different Unified Modeling Language (UML) Diagrams to provide the proper visualization to the object oriented system design and Data Flow Diagrams (DFD) to specify the data flow through the system. Hence, the following diagram will be followed to represent the system interaction process with its end users.

#### 6.1.1 Use Case Diagram

The Use Case Diagram defines behavior of the activities or functionalities that one or more user of the system can perform during interaction. It includes the dependencies in between the functionalities and also extends the number of ways by which the functionality can be done. Overall this diagram is used to show the user part actions.

#### 6.1.2 Activity Diagram

The Activity Diagram is used to elaborate the flow of processes between the activities of the system. It starts from the beginning and move through activities or operations to reach till last. It also shows the different conditions included in the system flow.

#### 6.1.3 Sequence Diagram

The Sequence Diagram is an appropriate description of the interactions between the user and the system objects and within objects. It shows the exchange of information in order of proper message and time sequence.

#### 6.1.4 Class Diagram

The Class Diagram is a proper representation of the system classes, attributes, methods and relationship among the instances of the classes. The diagram is also capable in illustrating the constraints enforced in the system.

### 6.1.5 Entity Relationship Diagram

The Entity Relationship Diagram is a graphical display of logical structure of database of the system. It shows relationship between the entities stored in the database. The relationships maintained among the entities are one to one, one to many, many to one and many to many. Entities are the information receiver or sender.

### 6.1.6 Data Dictionary

The Data Dictionary is the logical data modeling to represent textually the components of data flow concisely. It includes both entity and process with their interactive components. It helps in maintaining the database system requirements.

## 6.2 System Use Cases

**Actors of the system:** The persons who are active on internet i.e. almost everyone can interact with the system use cases.

**Use Cases of the system:** The functionalities and their processes are the use cases of the system.

### 6.2.1 Use Case 1: Earning Explorer Based On Hadoop Streaming

Table 17: Use Case 1

EARNING EXPLORER BASED ON HADOOP STREAMING	
Use Case ID	UCEE-01
Description	This use case describes the whole system's functionalities interacting with the user.

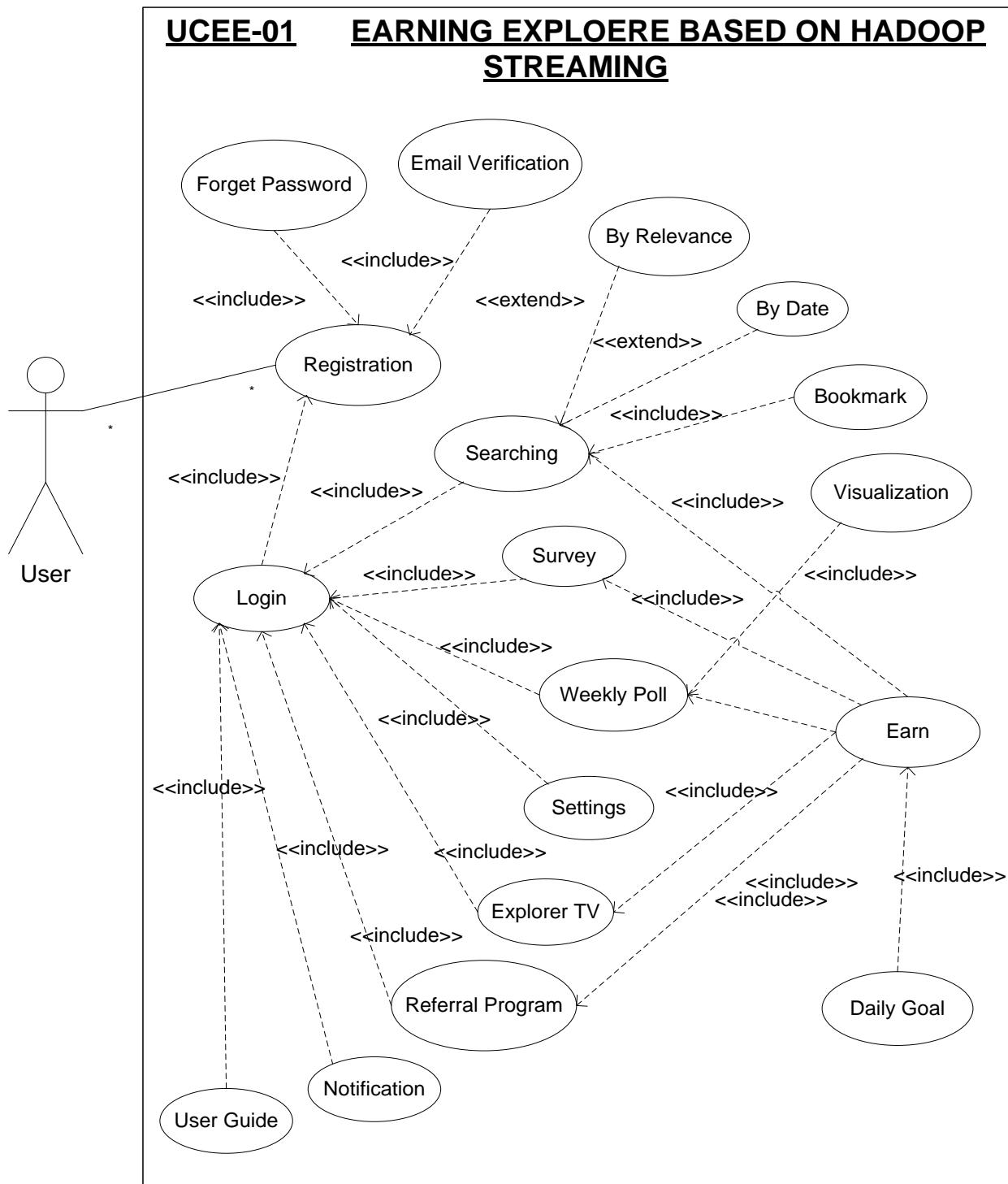


Figure 5: Earning Explorer Use Case

## Use Case 1: Earning Explorer Based On Hadoop Streaming Specification

**Table 18: Use Case 1: Earning Explorer Specification**

EARNING EXPLORER BASED ON HADOOP STREAMING	
<b>Use Case ID</b>	UCEE-01
<b>Primary Actors</b>	Internet Users, Anyone
<b>Preconditions</b>	<ul style="list-style-type: none"> <li>• Internet connection should be established on the device.</li> <li>• Should have installed any web browser on the user's device.</li> </ul>
<b>Assumptions</b>	Should understand basic English language.
<b>Priority</b>	High
<b>Event Occurrence</b>	Continuous
<b>Basic Flow</b>	<ul style="list-style-type: none"> <li>• User will register and email verification will be done.</li> <li>• User will login and perform the relevant functionalities.</li> <li>• Earnings will be updated in the user's account accordingly.</li> </ul>
<b>Substitute Flow</b>	<ul style="list-style-type: none"> <li>• If login fails, need to go to forget password.</li> <li>• Change password.</li> <li>• Again login.</li> </ul>

### 6.2.2 Use Case 2: Registration

**Table 19: Use Case 2**

REGISTRATION	
<b>Use Case ID</b>	UCEE-02
<b>Description</b>	This use case describes the user registering themselves to the system.

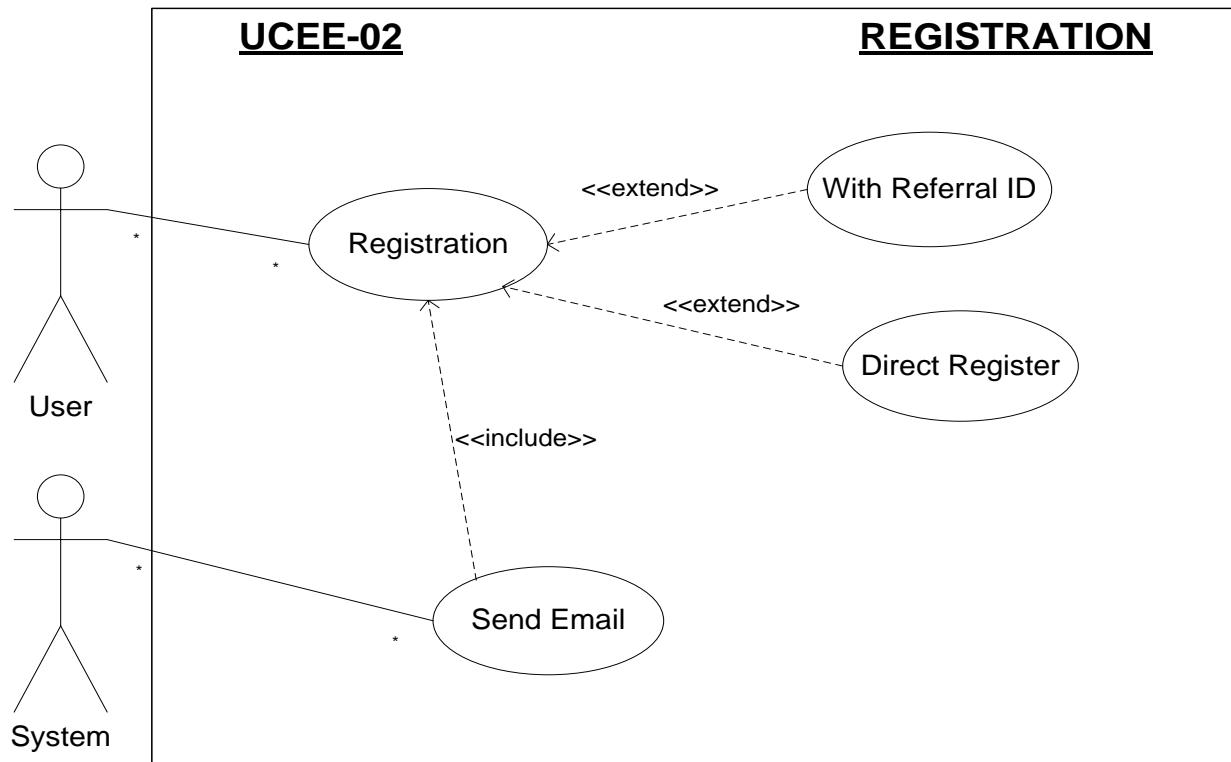


Figure 6: Registration Use Case

### Use Case 2: Registration Specification

Table 20: Use Case 2: Registration Specification

REGISTRATION	
Use Case ID	UCEE-02
Primary Actors	Internet Users, Anyone
Preconditions	<ul style="list-style-type: none"> <li>Internet connection should be established on the device.</li> <li>Should have installed any web browser on the user's device.</li> </ul>
Assumptions	Should understand basic English language.
Priority	High
Event Occurrence	Once
Basic Flow	<ul style="list-style-type: none"> <li>User will enter their details and do registration.</li> </ul>

	<ul style="list-style-type: none"> <li>• Registration can be done by using referral id too. Referral Id is the reference id of the friend user.</li> <li>• Details will be verified.</li> <li>• Details will get registered and become a user.</li> </ul>
<b>Substitute Flow</b>	<ul style="list-style-type: none"> <li>• If invalid details then registration process fails.</li> <li>• If through referral program, use referral code after entering details at the time of registration.</li> </ul>

### 6.2.3 Use Case 3: Referral Program

Table 21: Use Case 3

REFERRAL PROGRAM	
<b>Use Case ID</b>	UCEE-03
<b>Description</b>	This use case describes the user registration through referral code in the system.

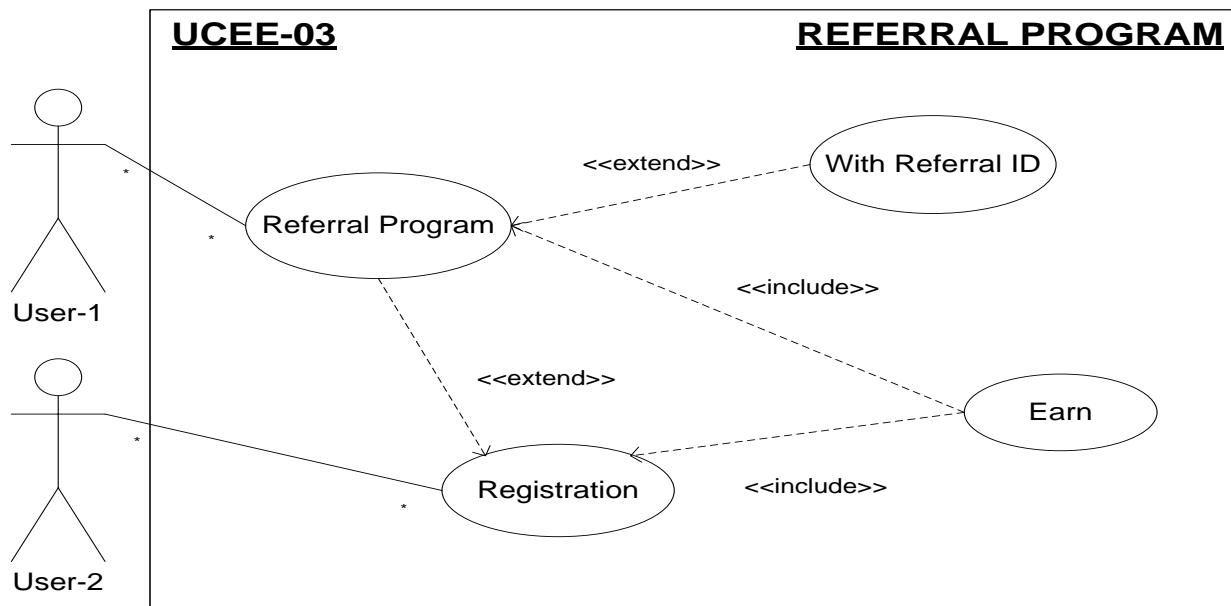


Figure 7: Registration Use Case

### Use Case 3: Referral Program Specification

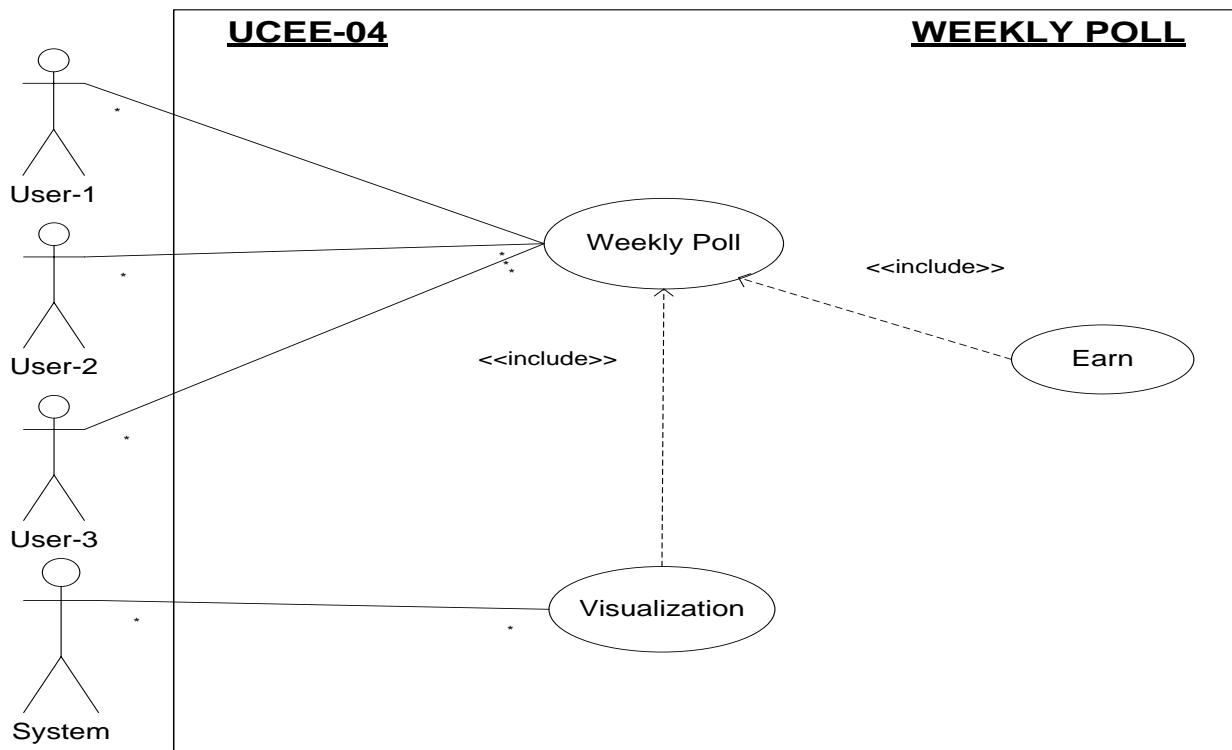
Table 22: Use Case 3: Referral Program Specification

REFERRAL PROGRAM	
<b>Use Case ID</b>	UCEE-03
<b>Primary Actors</b>	Internet Users, Anyone
<b>Preconditions</b>	<ul style="list-style-type: none"> <li>• Internet connection should be established on the device.</li> <li>• Should have installed any web browser on the user's device.</li> </ul>
<b>Assumptions</b>	Should understand basic English language.
<b>Priority</b>	Medium
<b>Event Occurrence</b>	Once
<b>Basic Flow</b>	<ul style="list-style-type: none"> <li>• Open email and click on referral link.</li> <li>• Enter user details.</li> <li>• Details will be verified.</li> <li>• Details will get registered and become a user.</li> </ul>
<b>Substitute Flow</b>	<ul style="list-style-type: none"> <li>• If invalid details then registration process fails.</li> </ul>

#### 6.2.4 Use Case 4: Weekly Poll Specification

Table 23: Use Case 4

WEEKLY POLL	
<b>Use Case ID</b>	UCEE-04
<b>Description</b>	This use case describes a functionality of the system in which user can do polls on weekly basis and earn money on the completion of the process.



**Figure 8: Weekly Poll Use Case**

## **Use Case 4: Weekly Poll Specification**

**Table 24: Use Case 4: Weekly Poll Specification**

WEEKLY POLL	
Use Case ID	UCEE-04
Primary Actors	Internet Users, Anyone
Preconditions	<ul style="list-style-type: none"><li>• Internet connection should be established on the device.</li><li>• Should have installed any web browser on the user's device.</li></ul>
Assumptions	Should understand basic English language.
Priority	High
Event Occurrence	Once in a week
Basic Flow	<ul style="list-style-type: none"><li>• User will login with their registered account.</li><li>• Go to weekly poll section.</li><li>• Select an option to answer the poll and click submit.</li></ul>

	<ul style="list-style-type: none"> <li>Result will be visualized in charts and point will be added to the user's account.</li> </ul>
Substitute Flow	None

### 6.2.5 Use Case 5: Daily Goal

Table 25: Use Case 5

DAILY GOAL	
Use Case ID	UCEE-05
Description	This use case describes a functionality in which all the earnings gets calculated and a goal will be set for the user benefit.

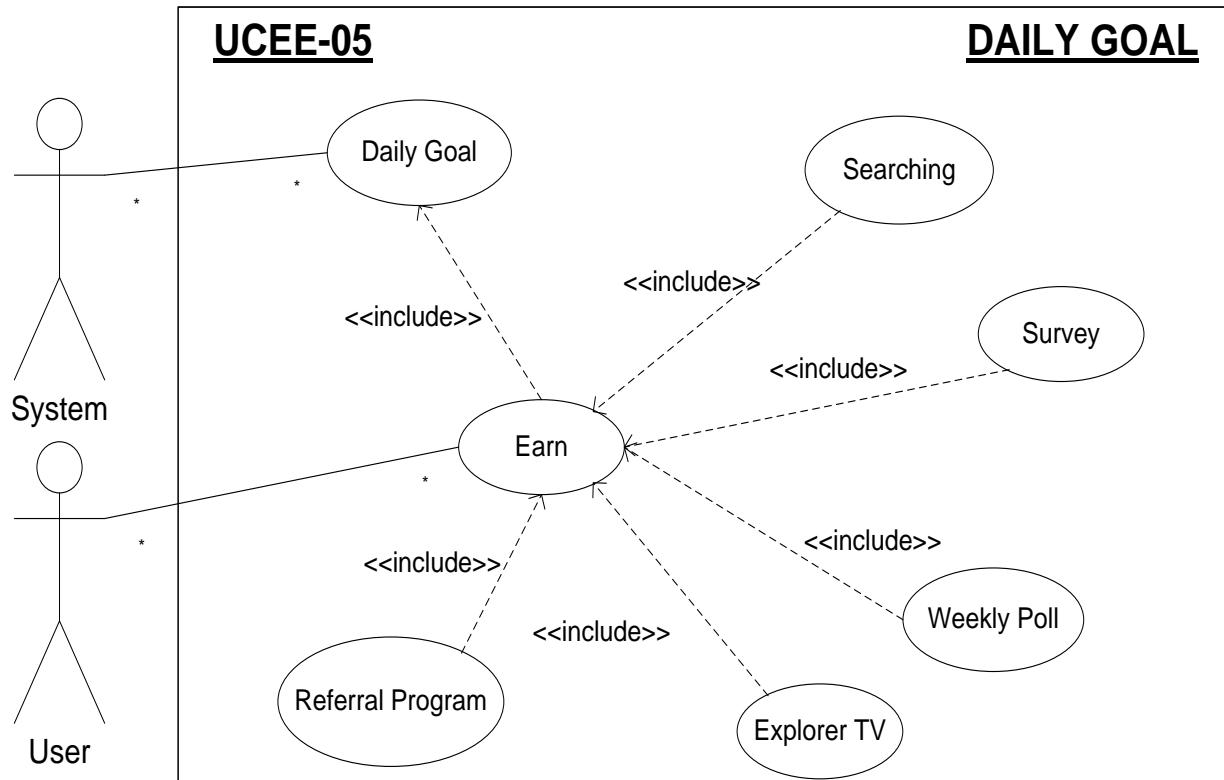


Figure 9: Daily Goal Use Case

## Use Case 5: Daily Goal Specification

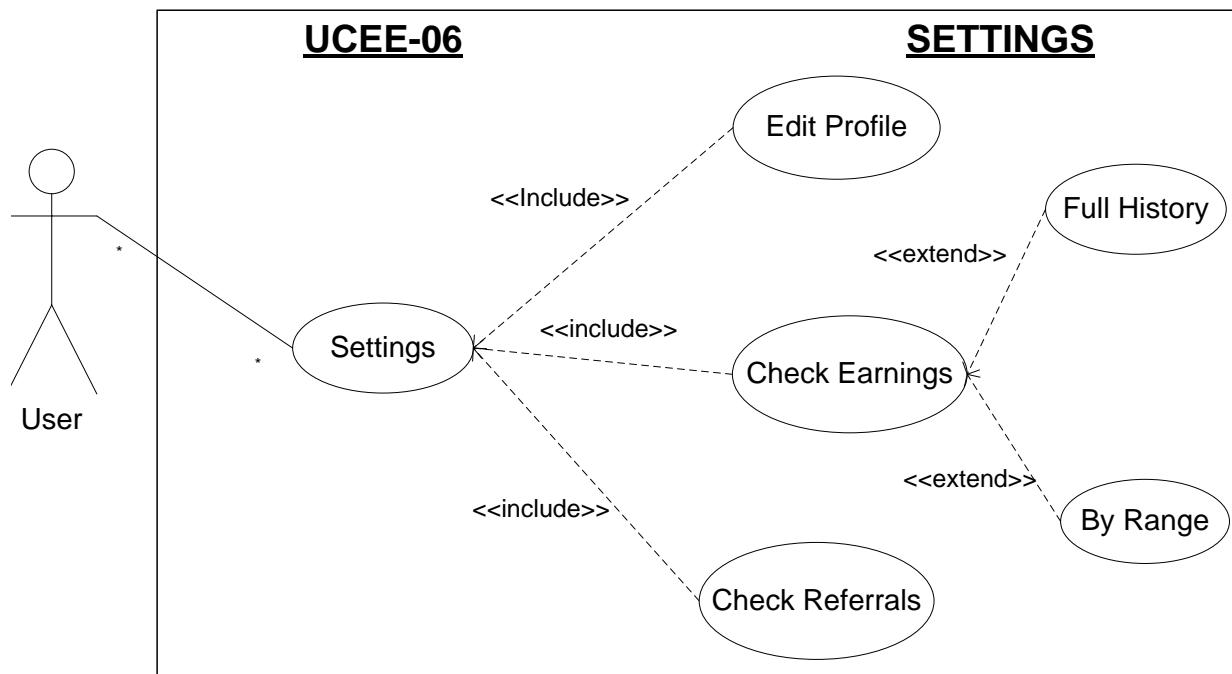
**Table 26: Use Case 5: Daily Goal Specification**

DAILY GOAL	
<b>Use Case ID</b>	UCEE-05
<b>Primary Actors</b>	Internet Users, Anyone
<b>Preconditions</b>	<ul style="list-style-type: none"> <li>• Internet connection should be established on the device.</li> <li>• Should have installed any web browser on the user's device.</li> </ul>
<b>Assumptions</b>	Should understand basic English language.
<b>Priority</b>	High
<b>Event Occurrence</b>	Continuous
<b>Basic Flow</b>	<ul style="list-style-type: none"> <li>• User will login with their registered account.</li> <li>• Go to daily goal section and the meter will show the users to achieve the total points to get bonus.</li> </ul>
<b>Substitute Flow</b>	None

### 6.2.6 Use Case 6: Settings

**Table 27: Use Case 6**

SETTINGS	
<b>Use Case ID</b>	UCEE-06
<b>Description</b>	This use case describes a functionality of the system through which the users can edit their profile and check the record of their earnings.



**Figure 10:** Settings Use Case

### Use Case 6: Settings Specification

**Table 28: Use Case 6: Settings Specification**

SETTINGS	
<b>Use Case ID</b>	UCEE-06
<b>Primary Actors</b>	Internet Users, Anyone
<b>Preconditions</b>	<ul style="list-style-type: none"> <li>• Internet connection should be established on the device.</li> <li>• Should have installed any web browser on the user's device.</li> </ul>
<b>Assumptions</b>	Should understand basic English language.
<b>Priority</b>	Medium
<b>Event Occurrence</b>	Continuous
<b>Basic Flow</b>	<ul style="list-style-type: none"> <li>• User will login with their registered account.</li> <li>• Go to settings to change settings and view earning report.</li> </ul>
<b>Substitute Flow</b>	None

### 6.3 System Activity Diagram

#### 6.3.1 Activity 1: Earning Explorer based on Hadoop Streaming

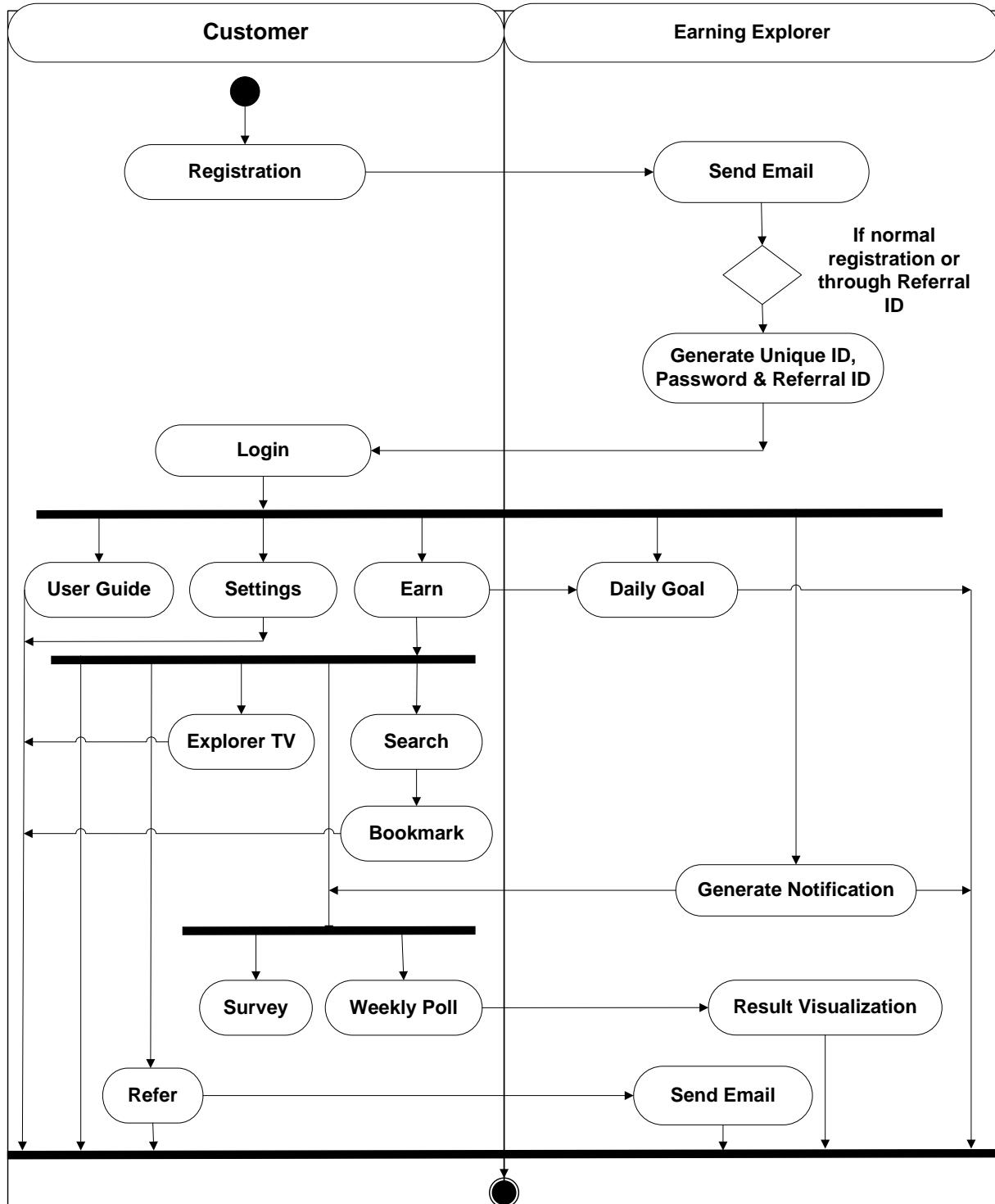
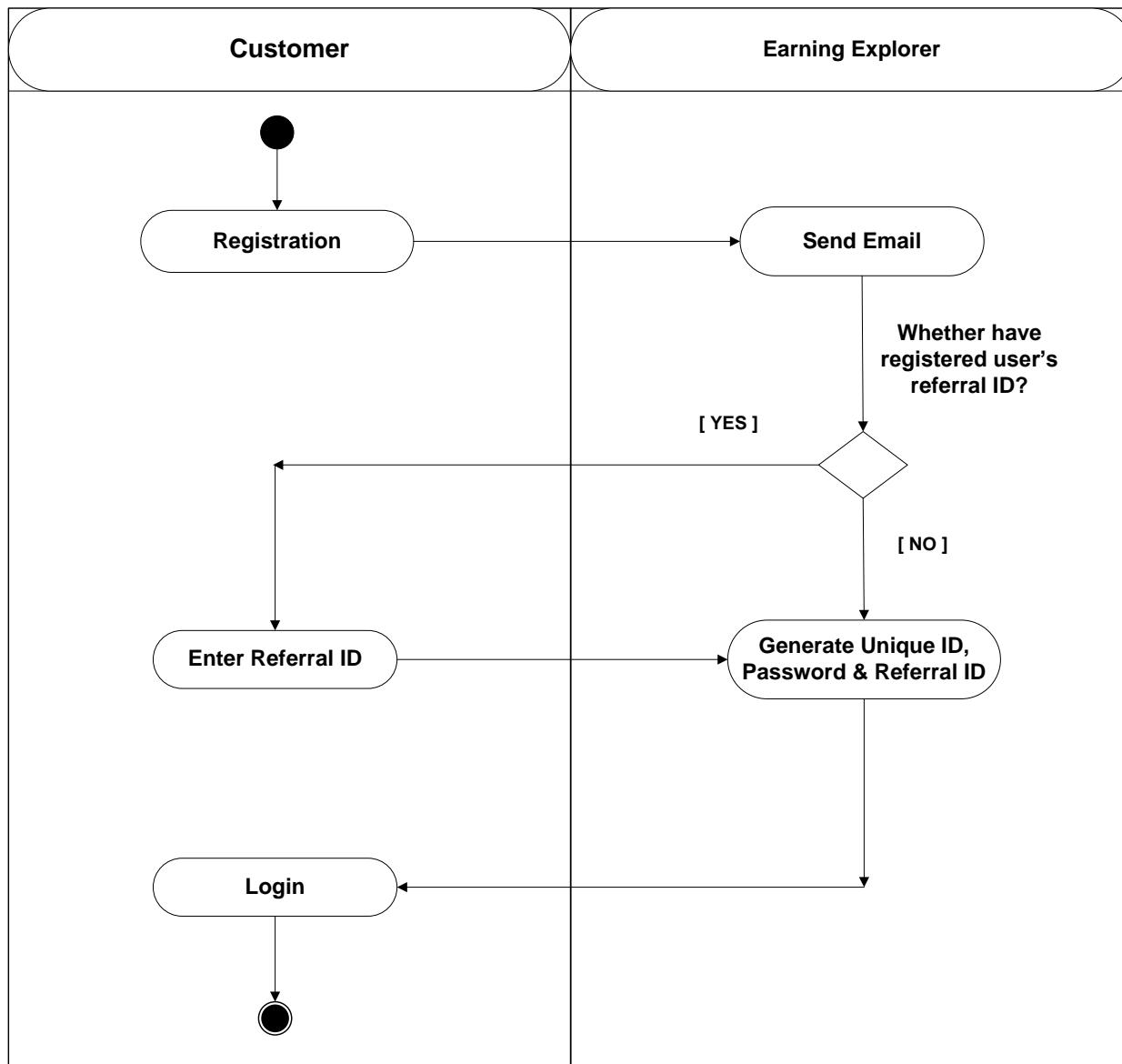


Figure 11: Earning Explorer Activity Diagram

### 6.3.2 Activity 2: Registration

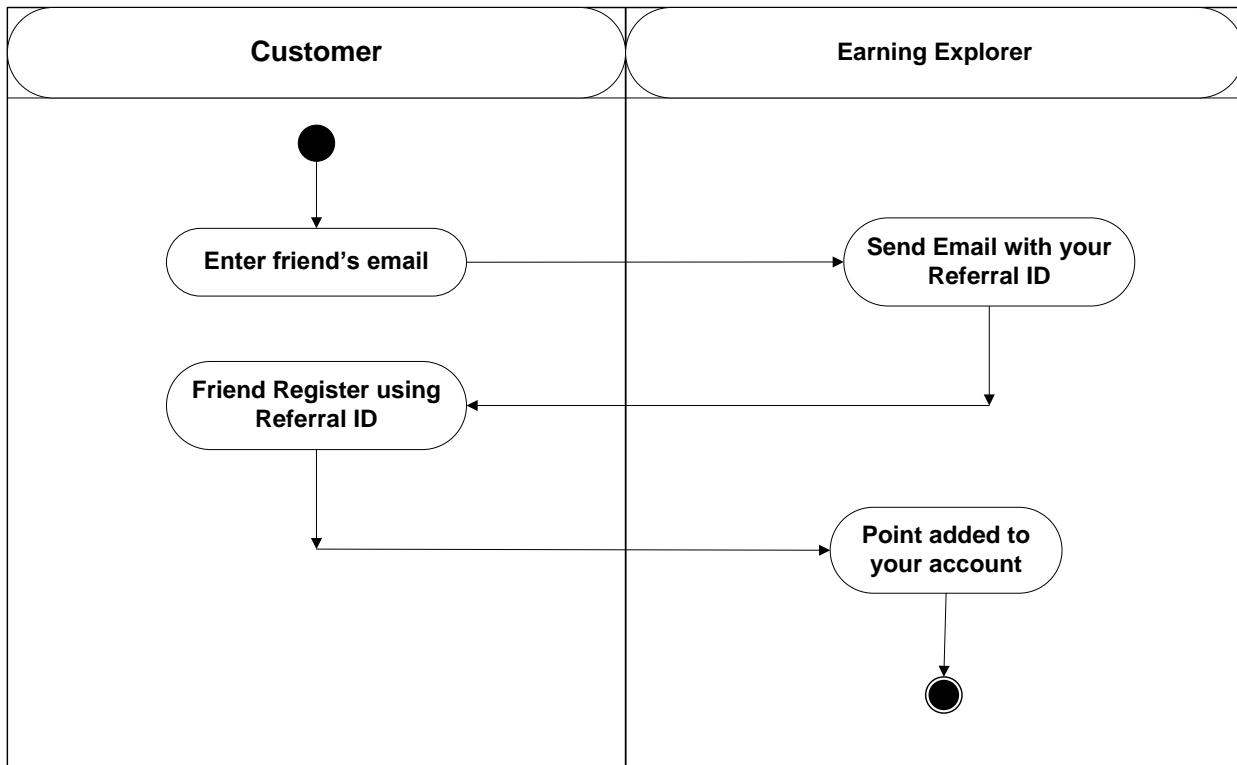
The registration activity diagram shows the flow of processes involved in completing the activity of registration functionality. The users are required to have a Gmail account to register themselves in the system. The system will email the registration link to that account. Then at the final stage user will click on the link to accept the registration process and become a member of the system. Also, the system will generate the unique referral code for the account.



**Figure 12: Registration Activity Diagram**

### 6.3.3 Activity 3: Referral Program

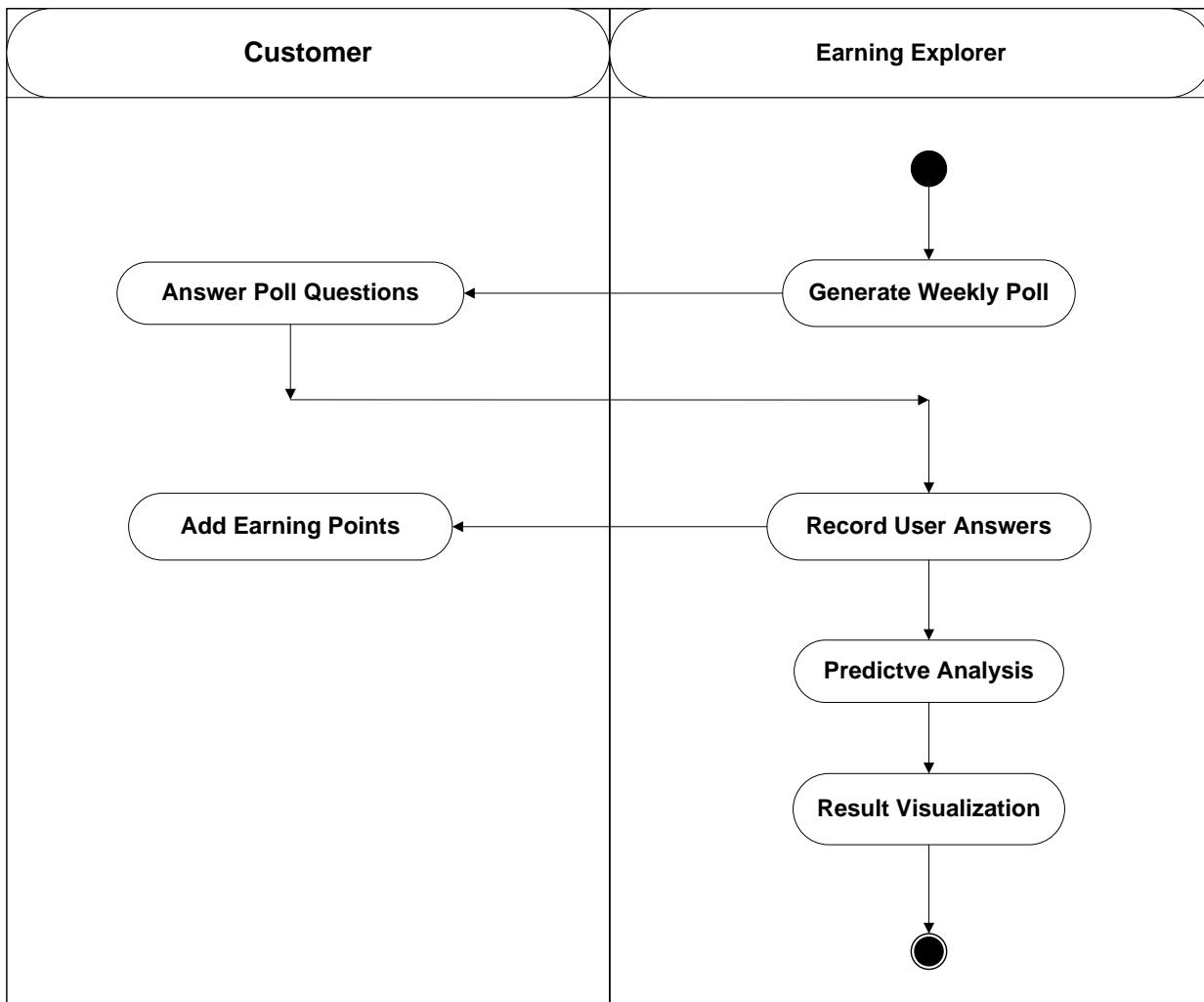
The referral program activity diagram shows the flow of processes involved in completing the activity of referral program functionality. One user can refer the system by sending its unique code to other user through email and hence earn money after that user register to the system by using that code. After this, the first user will get some weightage of referred user earning to the first user.



**Figure 13: Referral Program Activity Diagram**

### 6.3.4 Activity 4: Weekly Poll

The weekly poll activity diagram shows the flow of processes involved in completing the activity of weekly poll functionality. The users can attend the polls generated on weekly basis by the system. After attending the polls, user account will get added by some points. At the last stage the system will visualize the total result to represent the number of user favor and against on the subject.



**Figure 14: Weekly Poll Activity Diagram**

### 6.3.5 Activity 5: Notification

The notification activity diagram shows the flow of processes involved in completing the activity of notification functionality. The system generates the notification when new survey and weekly poll are saved in database. Hence, the users will get the earning points after completing the survey and polls.

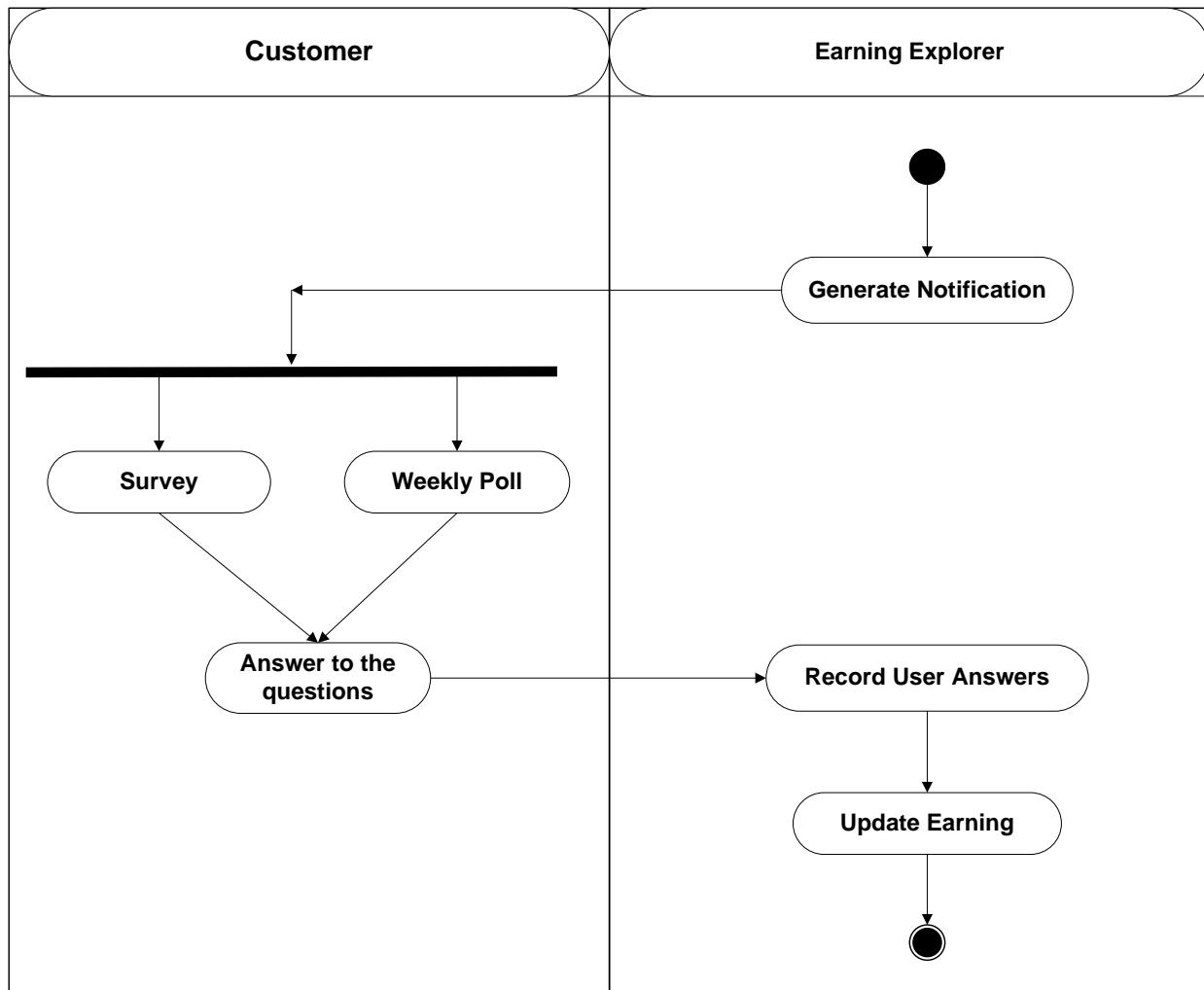
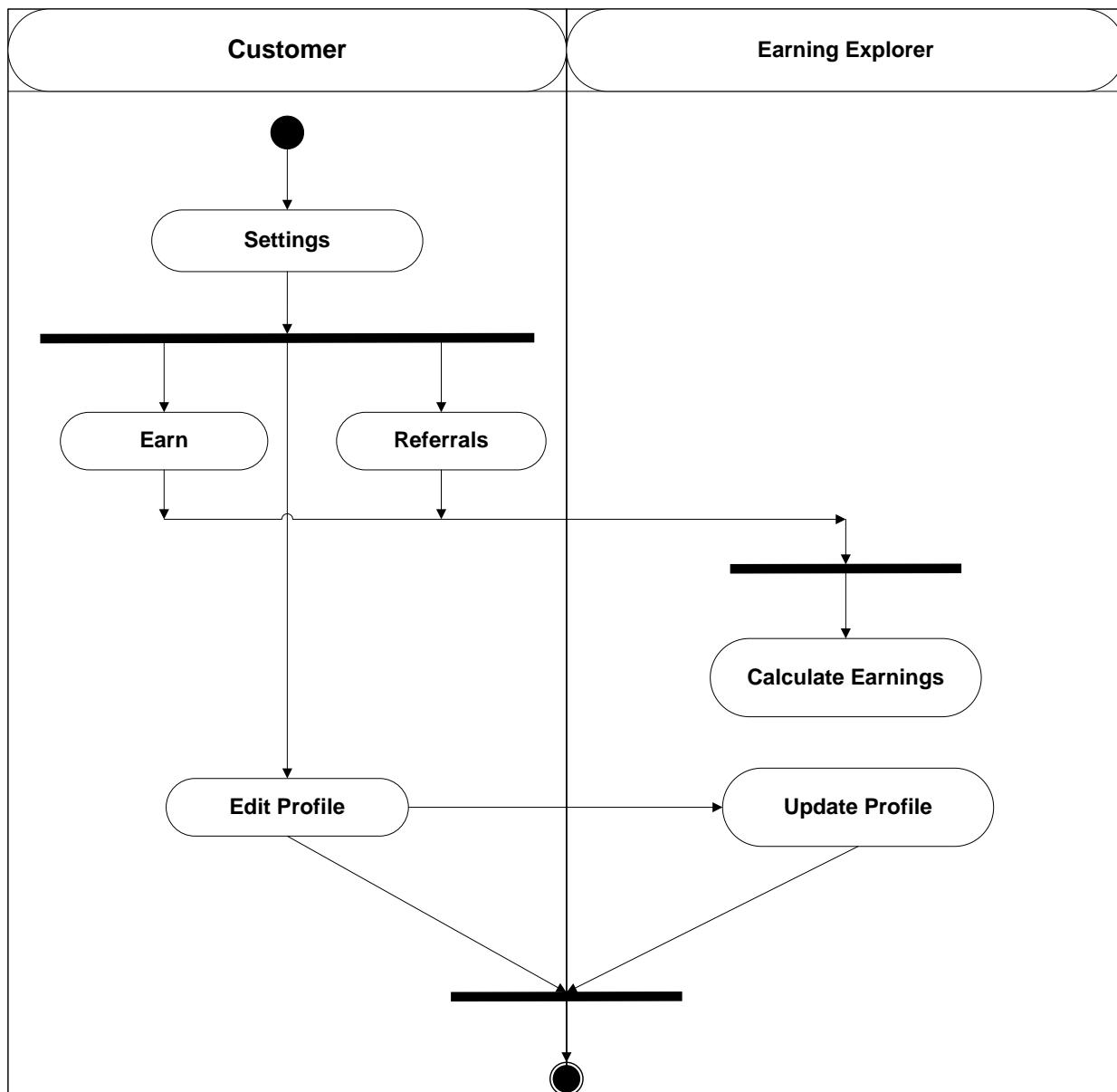


Figure 15: Notification Activity Diagram

### 6.3.6 Activity 6: Settings

The settings activity diagram shows the flow of processes involved in completing the activity of settings functionality. It includes the report for earn and referral. And also user can edit their profile information.



**Figure 16: Settings Activity Diagram**

## 6.4 System Sequence Diagram

### 6.4.1 Sequence 1: Earning Explorer based on Hadoop Streaming

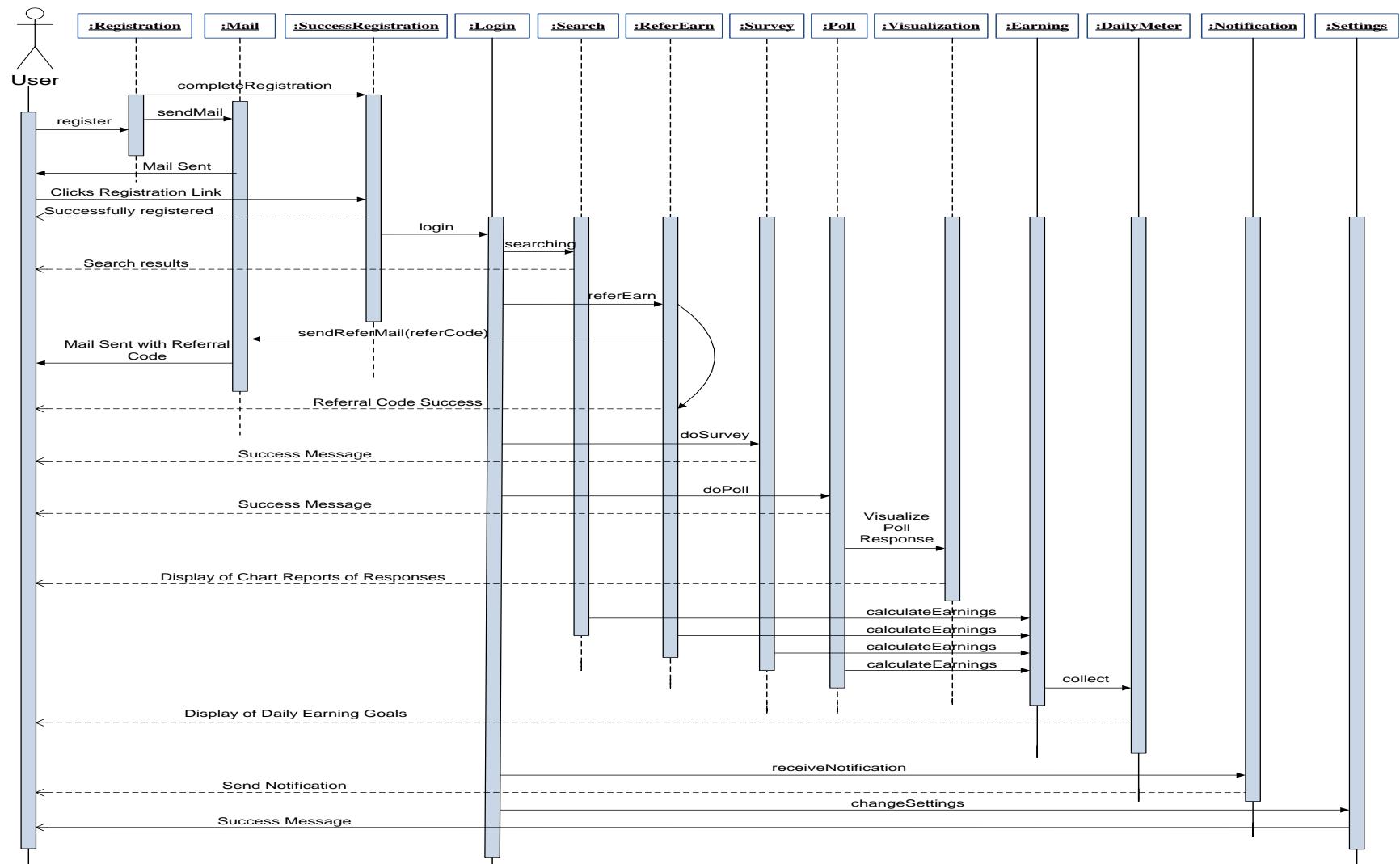
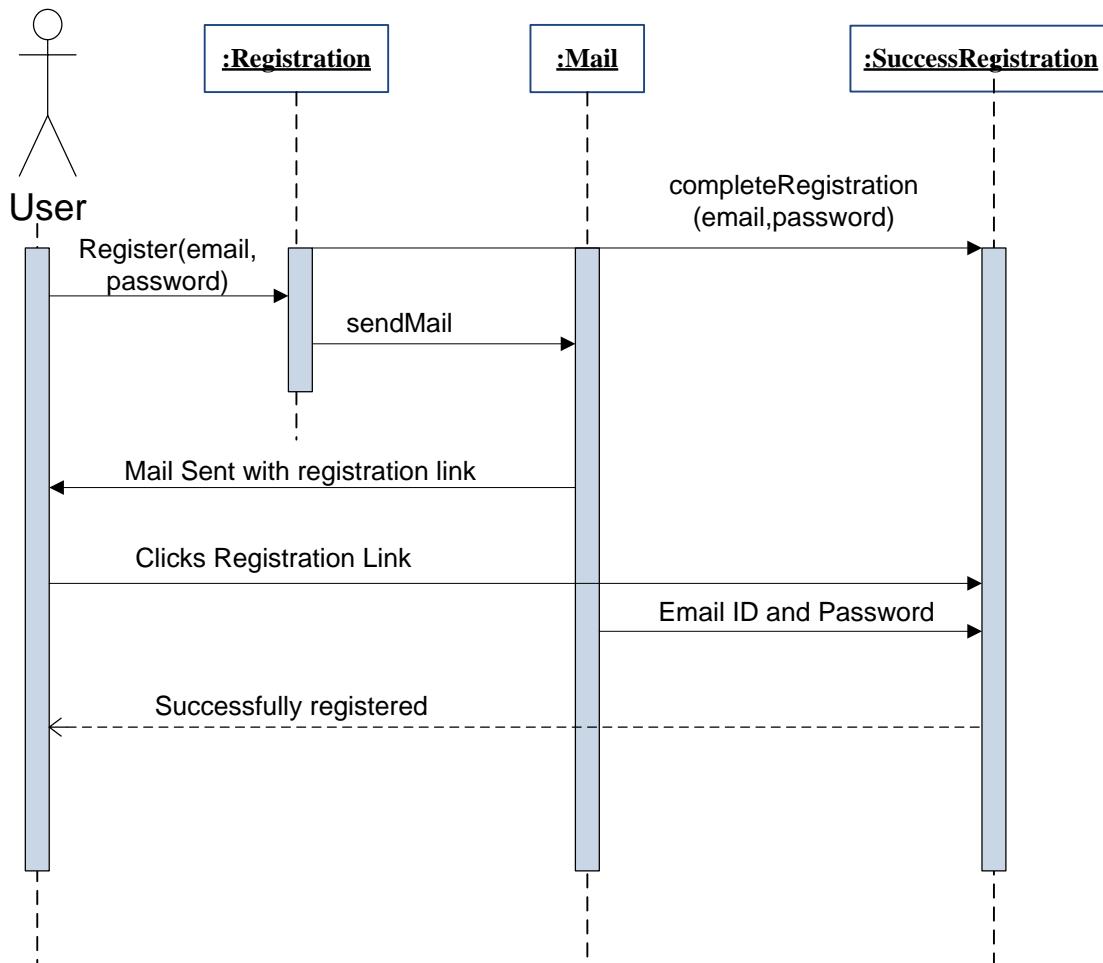


Figure 17: Earning Explorer Sequence Diagram

This is showing the whole system's sequential diagram to represent the system flow from registration to each functionality after logging in to the system. It explains how one user is going to interact with the users from one point to other within the system. Every task awards earning point to the users.

#### 6.4.2 Sequence 2: Registration

The registration sequence diagram describes the flow of data between the objects of the required classes for performing registration functionality. The sequence is saying that the users are required to have a Gmail account to register themselves in the system. The system will email the registration link to that account. Then at the final stage user will click on the link to accepts the registration process and become a member of the system.



**Figure 18: Registration Sequence Diagram**

### 6.4.3 Sequence 3: Referral Program

The referral program sequence diagram describes the flow of data between the objects of the required classes for performing referral program functionality. The sequence is saying that the user will login first then in the referral panel, the first user will refer the system to his friend user by sending referral code. The referred user will register with that code and allow the first user to get some weightage of his earning to the first user.

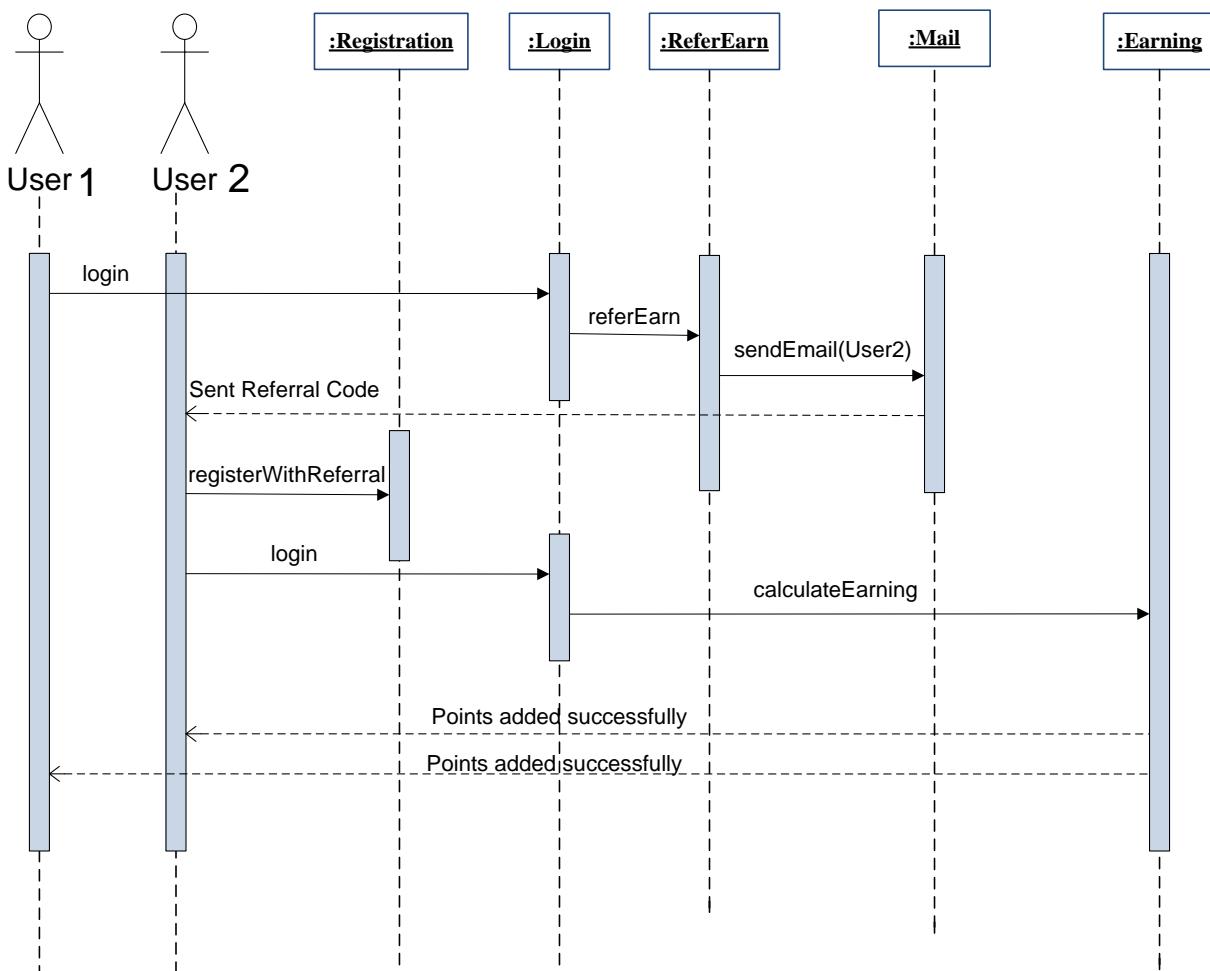
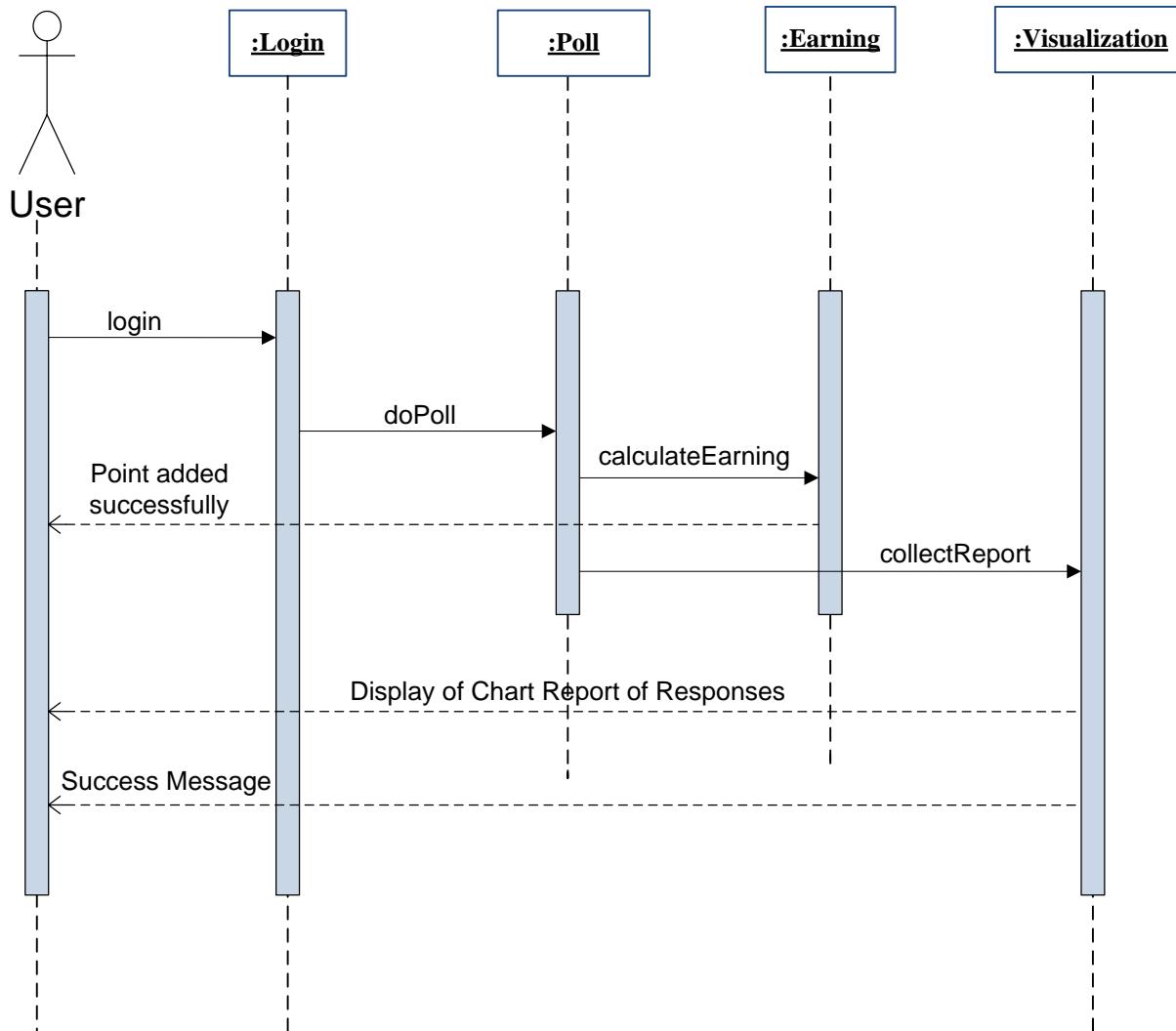


Figure 19: Referral Sequence Diagram

#### 6.4.4 Sequence 4: Weekly Poll

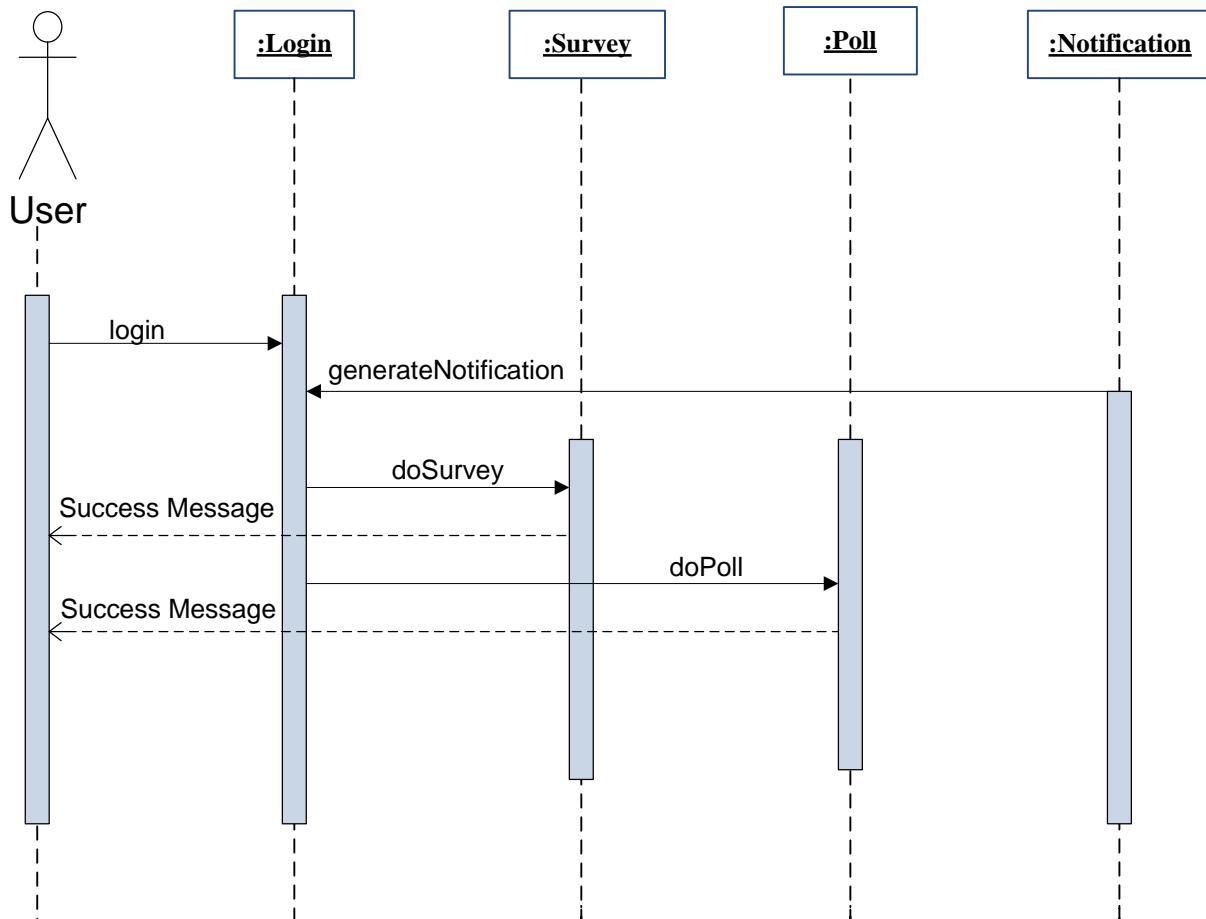
The weekly poll sequence diagram describes the flow of data between the objects of the required classes for performing weekly poll functionality. The sequence is saying that the user will login first then in the weekly poll panel and they will answer the poll. Then the system will show the user acceptance regarding the subject by visualizing the result and points will get added to earning module.



**Figure 20: Weekly Poll Sequence Diagram**

#### 6.4.5 Sequence 5: Notification

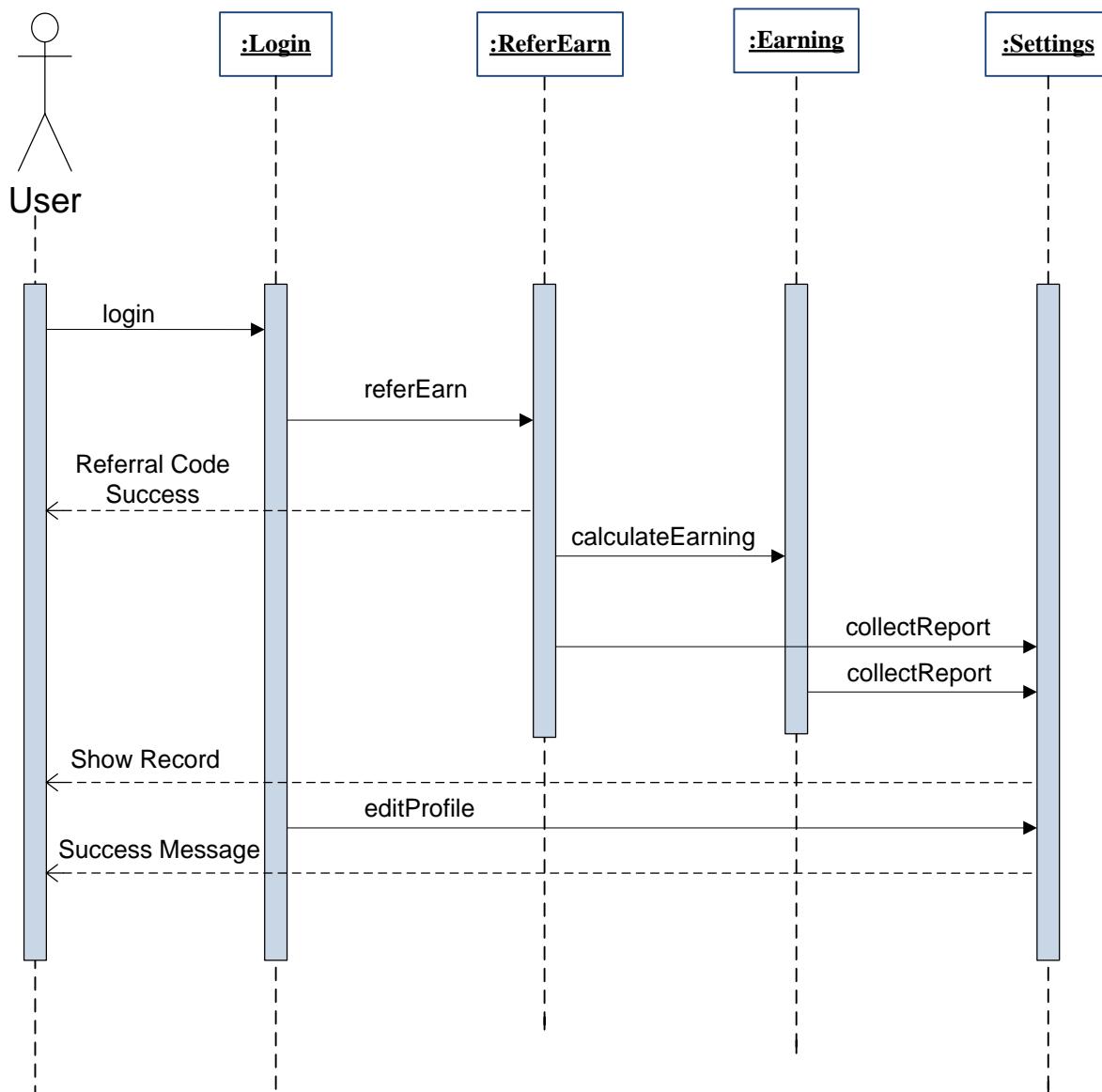
The notification sequence diagram describes the flow of data between the objects of the required classes for performing notification functionality. The sequence is saying that the user will login first then in the notification panel, they can get the response from the system on any occurrence of task for earning.



**Figure 21: Notification Sequence Diagram**

#### 6.4.6 Sequence 6: Settings

The settings sequence diagram describes the flow of data between the objects of the required classes for performing settings functionality. The sequence is saying that the user will login first then in the settings panel, they can view the record of points earned by them. Also, they can edit their profile settings.



**Figure 22: Settings Sequence Diagram**

## 6.5 Class Diagram

Class Diagram represents the static structure diagram to describe the system's functional arrangement. The below class diagram is of Earning Explorer which includes various classes and interfaces inside packages.

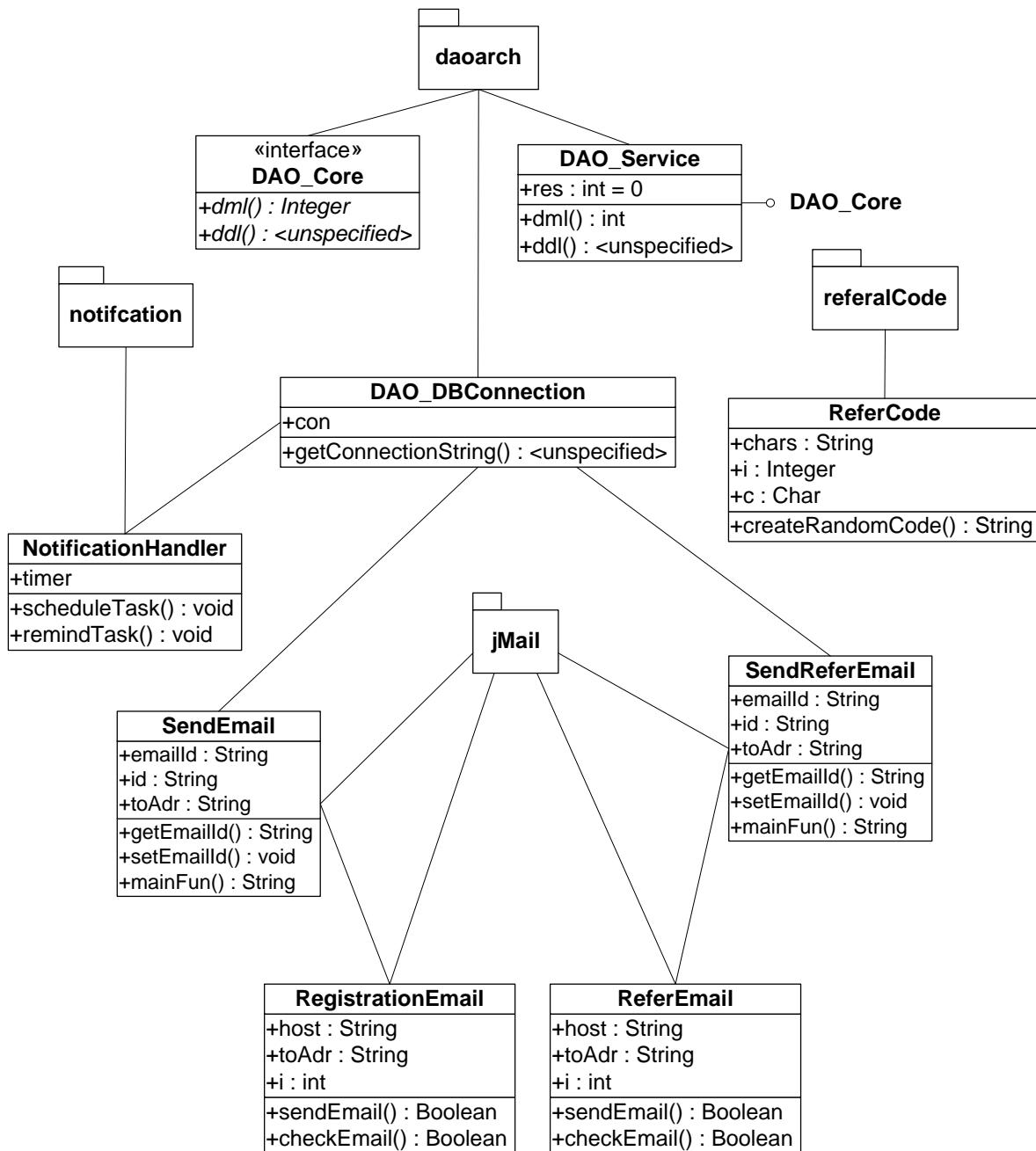


Figure 23: System Class Diagram

**Table 29: Class Diagram Description**

S. no	Includes	Description
1	Package	Daoarch, notification, referalcode, jMail
2	Class	DAO_DBConnection, DAO_Service, NotificationHandler, ReferCode, SendEmail, RgistrationEmail, ReferEmail, SendReferEmail
3	Interface	DAO_Core

## 6.6 Entity Relationship Diagram (ERD)

The Entity Relationship Diagram shows the relationship among the entity involved in the system. It is basically representing the backend data management of the system database. The following tables are listing the entities and relationships involved among them.

### 6.6.1 Entities Identified

**Table 30: Entities Identified**

S. no	Entity	Attributes
1	registration	<u>userid</u> , <u>password</u> , <u>fname</u> , <u>lname</u> , <u>refercode</u> , <u>usedcode</u> , <u>regdate</u>
2	earning	<u>earnid</u> , <u>userid</u> , <u>earning</u> , <u>earningtype</u> , <u>earningdate</u>
3	survey	<u>Surver_id</u> , <u>Survey_name</u> , <u>Question_id</u> , <u>Question_name</u> , <u>Points</u>
4	surveyresponses	<u>Response_id</u> , <u>Survey_id</u> , <u>User_id</u> , <u>Response</u> , <u>Respondedate</u>
5	weeklypoll	<u>pollid</u> , <u>pollquest</u> , <u>polldate</u>
6	pollresponse	<u>responseid</u> , <u>pollid</u> , <u>userid</u> , <u>response</u> , <u>respondedate</u>
7	refer	<u>refercode</u> , <u>userid</u> , <u>sento</u> , <u>status</u>

### 6.6.2 Relationship Identified

**Table 31: Relationship Identified**

S. no	Entities Involved	Relationship
1	registration, earning	Allows
2	earning, surveyresponses, pollresponse, refer	Receive points
3	surveyresponses, pollresponse, survey, weeklypoll	Taken on

### 6.6.3 System ERD

The system ERD is carrying the tables which have been created for the storage of data received on user input in the system. It explains the participation ratio in the relationship of each entity and also identifies the weak and strong entity in the system.

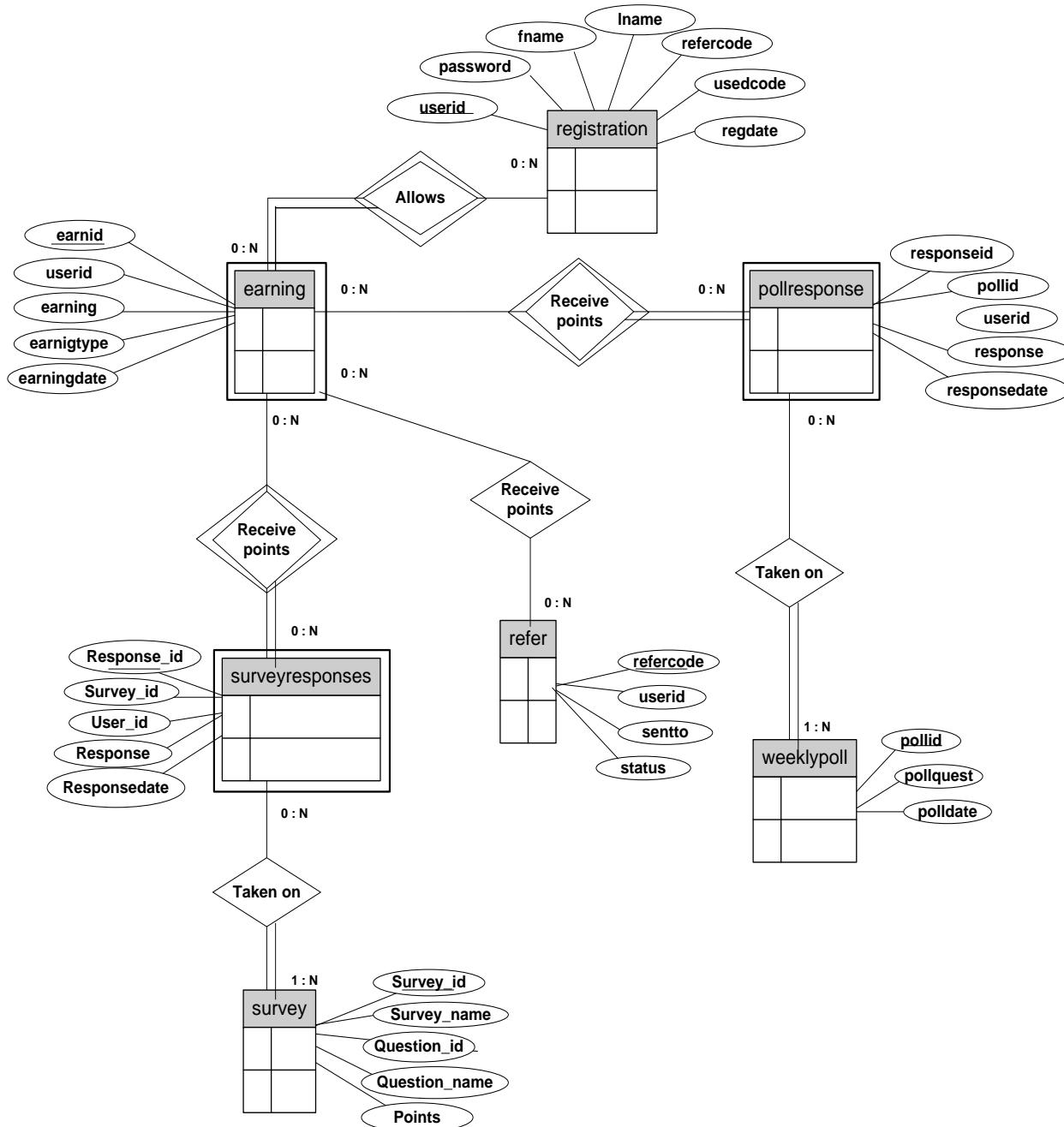


Figure 24: Earning Explorer ERD

## 6.6.4 Mapping Cardinality and Participation Constraint

### 6.6.4.1 Mapping Cardinality

**Table 32: Mapping Cardinality**

S. no	Entity	Relation	Entity
1	registration	Minimum: 0 Maximum: N <b>One To Many:</b> One registration to many earning. <b>Many To Many:</b> Many user can register and can earn many points accordingly.	earning
2	earning	Minimum: 0 Maximum: N <b>One To Many:</b> One user can earn from many refers or survey and poll responses.	surveyresponses, refer, pollresponse
3	survey	Minimum: 1 Maximum: N <b>One To Many:</b> One survey can have many responses. <b>Many To One:</b> Many user can respond to one survey.	surveyresponses
4	surveyresponses	Minimum: 0 Maximum: N <b>One To Many:</b> One user can respond to many surveys they want to and can earn accordingly.	earning, survey
5	weeklypoll	Minimum: 1 Maximum: N <b>One To Many:</b> One weekly poll can have many responses. <b>Many To One:</b> Many user can respond to one poll.	pollresponse

<b>6</b>	pollresponse	Minimum: 0  Maximum: N  <b>One To Many:</b> One user can respond to as many polls they want to and can earn accordingly.	earning, weeklypoll
<b>7.</b>	refer	Minimum: 0  Maximum: N  <b>One To Many:</b> One user can refer to as many friends they want to and can earn accordingly.	earning

#### 6.6.4.2 Participation Constraint

**Table 33: Participation Constraint**

S. no	Entity	Relation	Entity
<b>1</b>	registration	Partial Participation	earning
<b>2</b>	earning	Total Participation	registration
<b>3</b>	survey	Total Participation	surveyresponses
<b>4</b>	surveyresponses	Total Participation	earning
<b>5</b>	weeklypoll	Total Participation	pollresponse
<b>6</b>	pollresponse	Total Participation	earning
<b>7</b>	refer	Partial Participation	earning

#### 6.6.5 Mapping ERD to Relational Database

The following tables are explaining the features of entities involved in ERD model of Earning Explorer according to the database. Hence, the mapping entity type and cardinality ratio has been represented below.

### 6.6.5.1 Mapping of Strong Entity Type

**Table 34: Mapping of Strong Entity**

S.no	Entity Type	Relation
1	registration	<u>userid</u> , password, fname, lname, refercode, usedcode, regdate
2	survey	<u>Survey_id</u> , Survey_name, Question_id, Question_name, Points
3	weeklypoll	<u>pollid</u> , pollquest, polldate
4	refer	<u>refercode</u> , userid, sentto, status

### 6.6.5.2 Mapping of Weak Entity Type

**Table 35: Mapping of Weak Entity**

S.no	Entity Type	Relation
1	earning	<u>earnid</u> , userid, earning, earningtype, earningdate
2	surveyresponses	<u>Response_id</u> , Survey_id, User_id, Response, Respondedate
3	pollresponse	<u>responseid</u> , pollid, userid, response, respondedate

### 6.6.5.3 Mapping of 1: 1 Relationship

The system doesn't include the one to one relationship because any user can perform any number of earnings, survey, refers and weekly poll.

### 6.6.5.4 Mapping of 1: M Relationship

In the system, one user can respond to many survey or weekly poll and can earn points accordingly.

**Table 36: Mapping of 1: M**

Relation Name	Entities Involved	Approach	Relation
Taken on	survey, surveyresponses	Foreign Key	survey.Survey_id = PK surveyresponses.Survey_id = FK
Taken on	weeklypoll, pollresponse	Foreign Key	weeklypoll.pollid = PK pollresponse.pollid = FK

### 6.6.5.5 Mapping of M: 1 Relationship

In the system, many user can respond to one survey or weekly poll and can earn points accordingly.

**Table 37: Mapping of M: 1**

Relation Name	Entities Involved	Approach	Relation
<b>Taken on</b>	surveyresponses, survey	Foreign Key	survey.Survey_id = PK surveyresponses.Survey_id = FK
<b>Taken on</b>	Pollresponse, weeklypoll	Foreign Key	weeklypoll.pollid = PK pollresponse.pollid = FK

### 6.6.5.6 Mapping of M: N Relationship

In the system, many user can register themselves and can earn many points by doing several activity provided by the system.

**Table 38: Mapping of M: N**

Relation Name	Entities Involved	Approach	Relation
<b>Allows</b>	registration, earning	Foreign Key	registration.userid = PK earning.userid = FK

## 6.7 Normalization

Normalization is the process of making a database design perfect by avoiding different anomalies like update, delete and insert. It modifies the relational schema based its functional dependencies and primary key.

### 1. First Normal Form (INF)

This form defines that all the attributes of a relation should have atomic value, no composite. Each column must have unique name and must contain same kind of value (data type). The first normal form achieved in the system has been shown below. Every column contains unique value of same data type.

earnid	userid	earning	earningtype	earningdate
--------	--------	---------	-------------	-------------

## 2. Second Normal Form (2NF)

Firstly, it should satisfy the 1NF. This normal form defines that non-prime attributes should be fully functional dependent on any key of the relation.

### Before Normalization

The weekly poll was including poll data and poll response data in one table.

<u>pollid</u>	pollquest	polldate	userid	response
---------------	-----------	----------	--------	----------

### After Normalization

Now the table has been separated into two tables that are “weeklypoll” and “pollresponse” on the basis of poll id (primary key) to make it in 2NF.

<u>pollid</u>	pollquest	polldate
---------------	-----------	----------

pollid	<u>responseid</u>	userid	response	respondedate
--------	-------------------	--------	----------	--------------

## 3. Third Normal Form (3NF)

3NF should fulfill 2NF. Then, no non-prime attributes should be transitively dependent on candidate key. All attributes should be non-transitively dependent on primary key.

### Before Normalization

The survey was including survey data and response data in one table.

<u>pollid</u>	pollquest	polldate	userid	response
---------------	-----------	----------	--------	----------

### After Normalization

Now the table has been separated into two tables that are “weeklypoll” and “pollresponse” on the basis of poll id (primary key) for first table and response id (primary key) for second table to make it in 3NF.

Survey_id	Survey_name	Question_id	Question_name	Points	userid	response
-----------	-------------	-------------	---------------	--------	--------	----------

<u>Responseid</u>	Survey_id	Userid	Question_id	Response	Respondedate
-------------------	-----------	--------	-------------	----------	--------------

### 6.7.1 Tables Dependencies

The ER-diagram of the proposed system “Earning Explorer” has been normalized up to 3<sup>rd</sup> NF.

Hence, the relationship between the tables after 3<sup>rd</sup> NF is same as the proposed ER-Diagram.

#### Registration

<u>userid</u>	password	Fname	Lname	refercode	usedcode	regdate
---------------	----------	-------	-------	-----------	----------	---------

#### Earning

<u>earnid</u>	userid	earning	earningtype	earningdate
---------------	--------	---------	-------------	-------------

#### Survey

<u>Survey_id</u>	Survey_name	Question_id	Question_name	Points
------------------	-------------	-------------	---------------	--------

#### Surveyresponses

<u>Responseid</u>	Survey_id	Userid	Question_id	Response	Respondedate
-------------------	-----------	--------	-------------	----------	--------------

#### Weeklypoll

<u>pollid</u>	pollquest	polldate
---------------	-----------	----------

#### Pollresponse

<u>responseid</u>	pollid	userid	response	respondedate
-------------------	--------	--------	----------	--------------

#### Refer

<u>refercode</u>	userid	sentto	status
------------------	--------	--------	--------

### 6.8 Data Dictionary

#### 6.8.1 Table registration

It stores the details related to registration process of a new user to become a member of the system and hence, can access the authorized part of the system after successful registration.

**Table 39: Database Table Registration**

Information	Data Type	Description	Null/Not Null
<b>userid</b>	varchar(50)	Primary Key	Not Null
<b>password</b>	varchar(45)	Store password of the user.	Not Null
<b>fname</b>	varchar(45)	Stores first name of users.	Null
<b>lname</b>	varchar(45)	Stores last name of users.	Null
<b>refercode</b>	varchar(10)	Stores referral code for the profile of users.	Not Null
<b>usedcode</b>	varchar(10)	Stores referral code used by the user at the time of registration.	Null
<b>regdate</b>	Date	Stores date value for the date of registration done by the users.	Not Null

### 6.8.2 Table earning

It stores the details related to the earnings involved for a user in the system. It has relation with the survey and weekly poll response table.

**Table 40: Database Table Earning**

Information	Data Type	Description	Null/Not Null
<b>earnid</b>	Int(11)	Primary Key, Auto increment	Not Null
<b>userid</b>	varchar(45)	Store user id reference from registration table.	Not Null
<b>earning</b>	Decimal(10,0)	Stores the points earned by the user	Not Null
<b>earningtype</b>	varchar(20)	Stores the process type of earning.	Not Null
<b>earningdate</b>	Date	Stores date value for the date of earning received by the users.	Not Null

### 6.8.3 Table refer

It stores the details related to referral code send by one user to other. Its status is updated once the receiver accepts the referral code at the time of registration.

**Table 41: Database Table Refer**

Information	Data Type	Description	Null/Not Null
<b>refercode</b>	varchar(45)	Primary Key	Not Null
<b>userid</b>	varchar(45)	Store user id who is sending the referral code.	Not Null
<b>sentto</b>	varchar(45)	Store user id who is receiving the referral code.	Not Null
<b>status</b>	varchar(45)	Stores the process type of earning.	Not Null

#### 6.8.4 Table survey

It stores the details related to surveys send to the users by the system.

**Table 42: Database Table Survey**

Information	Data Type	Description	Null/Not Null
<b>Survey_id</b>	Int(11)	Composite Primary Key	Not Null
<b>Survey_name</b>	varchar(250)	Stores the subject of the survey.	Not Null
<b>Question_id</b>	Int(11)	Composite Primary key and stores question numbers for a survey.	Not Null
<b>Question_Name</b>	varchar(250)	Stores the question statement.	Not Null
<b>Points</b>	Float	Stores points for a survey.	Not Null

#### 6.8.5 Table surveyresponses

It stores the details related to surveys responses send by the users to the system. For every survey response point has been calculated.

**Table 43: Database Table Surveyresponses**

Information	Data Type	Description	Null/Not Null
<b>Responseid</b>	Int(11)	Primary Key	Not Null
<b>Survey_id</b>	Int(11)	Stores survey identity number.	Not Null
<b>User_id</b>	varchar(50)	Stores user id who have submitted responses for the survey	Not Null
<b>Question_id</b>	Int(11)	Stores question identity number.	Not Null

<b>Response</b>	Varchar(100)	Stores text value of responses.	Not Null
<b>Respondedate</b>	Date	Stores date of response given by the user.	Not Null

### 6.8.6 Table weeklypoll

It stores the details related to weekly poll send to the users by the system. For every weekly poll question it has points which system will award to the users.

**Table 44: Database Table Weeklypoll**

Information	Data Type	Description	Null/Not Null
<b>pollid</b>	Int(11)	Primary Key, Auto increment	Not Null
<b>pollquest</b>	varchar(100)	Stores the question statement.	Not Null
<b>polldate</b>	Date	Stores date on which poll has been created by the system.	Not Null

### 6.8.7 Table pollresponse

It stores the details related to weekly poll send to the users by the system. For every weekly poll question it has points which system will award to the users.

**Table 45: Database Table Pollresponse**

Information	Data Type	Description	Null/Not Null
<b>responseid</b>	Int(11)	Composite Primary Key, Auto increment	Not Null
<b>pollid</b>	Int(11)	Composite Primary Key that stores the weekly poll identity number.	Not Null
<b>userid</b>	varchar(50)	Stores user id who have submitted responses for the poll	Not Null
<b>response</b>	Varchar(10)	Stores text value of responses.	Not Null
<b>respondedate</b>	Date	Stores date of response given by the user.	Not Null

## 6.9 Testing Template

The below table has been designed to figure out the parameters required for each type of testing. It will help in the accuracy of the expected results for each module of the system.

**Table 46: Testing Template**

Test Case ID		Test Case Name			
Testing Type		Tester Name			
S.no.	Input	Expected Output	Actual Output	Success/ Failure	Remarks

## 6.10 Conclusion

The chapter system design has been successfully implemented which will help in producing the expected configured system for the users. Different modelling designs for each module of the system has been implemented very precisely to develop an error free system. Also, the designing of the system is intended to get the high user acceptance rate.

## CHAPTER 7

# IMPLEMENTATION

### 7.1 Introduction

The significance of project's implementation phase is of stronger weightage as it brings the idea to real time execution. Execution is the only way through which one developer can convince their users to spend their time on the developed system.

For the development of proposed system, several tools and techniques have been used for each functions to be successful. The basic requirement would be knowledge of programming language and functioning of tools used. The programming work experience of developer is in HTML 5, Java, Java Server Pages, Bootstrap, MySQL and Hadoop and have developed various web application on NetBeans IDE 8.0.2. Hence, the developer has figured out that these experience would be worth to develop the Earning Explorer web application combining with Hadoop technology. Also, there will be use of WinSCP-5.9.4 software for creating communication of files (table data files) between local file system where web application is stored and Hadoop file system where whole analysis will be performed.

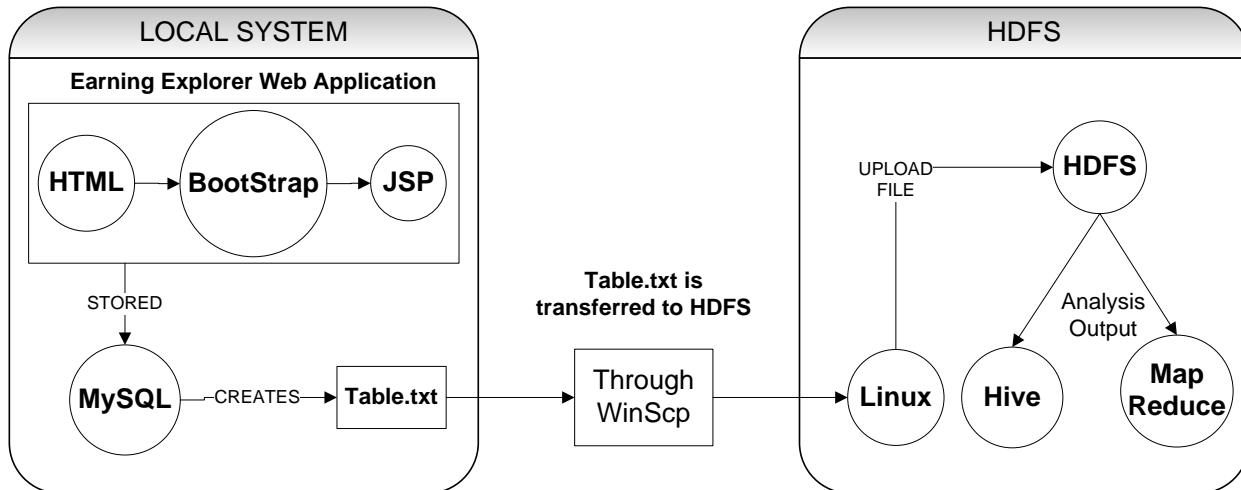


Figure 25: System inside View

After the commencement of application, there are many parameter which generally does not match with the real time proposals. Hence, the implementation phase will overcome the drawback of design phase decisions.

## 7.2 Implementation Plan

### 7.2.1 PERT Chart

PERT stands for Program Evaluation Review Technique. It is a tool designed for project management planning to estimate the time period that it will take to final the project. The main benefits of preparing a PERT chart in the project are as follows.

1. It maintains the work according to deadline.
2. It identifies the possibility of completing a task before the decided end date.
3. The critical path actions can affect the projects duration.
4. It displays the activity's start dates and end dates.

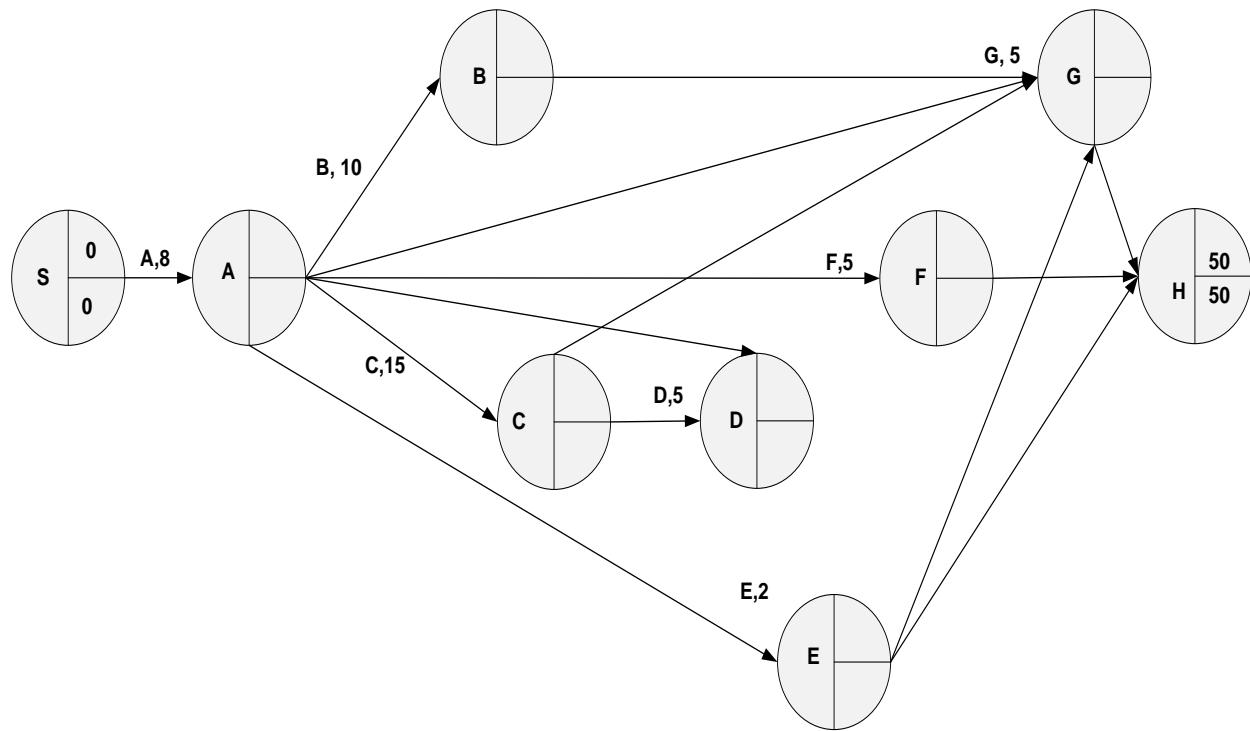
### 7.2.2 PERT Chart for Earning Explorer

The below mentioned table is representing the variables used for tasks included in the proposed System with the dependencies among each other. According to **Agile Methodology**, implementation phases in the development will be on continuous analysis to develop a best prototype of the system. Hence, the developer has judged the phase minutely and explored the best possible way of working on the project to deliver it on time.

**Table 47: Implementation plan**

IMPLEMENTATION PLAN			
Starting Date		25, December, 2016	
Ending Date		20, February, 2017	
Total Duration		50 days	
Task	Description	Predecessor	Duration
S	Start	-	-
A	Registration , Login	S	8 Days
B	Searching	A	10 Days
C	Survey, Weekly Poll	A	15 Days
D	Visualization	A, C	5 Days
E	Referral Program	A	2 Days

<b>F</b>	<b>Settings, User Guide</b>	<b>A</b>	<b>5 Days</b>
<b>G</b>	<b>Notification</b>	<b>A, B, C, E</b>	<b>5 Days</b>
<b>H</b>	<b>End</b>	-	-



**Figure 26: Implementation PERT**

### 7.3 Tools Used

The below mentioned table shows the basic tools required to successfully execute the system on any other laptop device.

**Table 48: Tools Used**

Sr. No.	Technical Requirements	Description
1.	Operating System	Windows 10 (32 or 64 bit)
2.	Software/ Tools	<ul style="list-style-type: none"> <li>✓ NetBeans IDE 8.0.2</li> <li>✓ Oracle VM Virtual Box (Linux)</li> <li>✓ Cloudera Manager</li> <li>✓ WinSCP-5.9.4</li> <li>✓ JRE 8, JDK 1.8</li> </ul>

<b>3.</b>	Web Browser	Firefox / Google chrome.
<b>4.</b>	Documentation Tools	<ul style="list-style-type: none"> <li>✓ Microsoft Word 2013</li> <li>✓ Microsoft Visio 2013</li> <li>✓ Snipping Tool</li> </ul>

## 7.4 Implementation of Complex Feature

### 7.4.1 Referral Program

The module referral program allow the users to earn money through unique referral code. The user can send that code to another friend user and once the friend user will register through the same referral code then both will get money.

#### Pseudo Code

```

BEGIN
DECLARE UniqueReferralCode /* It will be auto generated by java code for USER 1 */
DECLARE User2Emailid
DECLARE Message
READ User2Emailid
READ Message
READ UniqueReferralCode
IF User2Emailid registered /* Check in registration table */
  DISPLAY "Email id already registered"
ELSE
  DISPLAY "Referral code sent successfully"
END IF
END
BEGIN
PROMT Registration (Friend User)
DECLARE bonus = 50      /* To award bonus to USER 1 after registration of USER 2*/
DECLARE regularBonus = 0 /* To award bonus to USER 1 regularly of USER 2 earning*/
DECLARE EarnedPoints    /* For calculating earning of USER 2 */

```

```

READ Referral code from email link.

READ Email id

READ Password

READ ConfirmPassword

READ bonus

READ Referral Code sender's Email id

UPDATE Earning of USER 1 + bonus /* Only on same date */

READ EarnedPoints

LOOP READ EarnedPoints TO end of rows STEP 1 /* Reading total earning of the USER2
from database till the date */

earnings = earning + EarnedPoints

NEXT USER2

END LOOP

READ earnings

SET regularBonus = round ( earning * 0.1 ) /* Award bonus to USER 1 regularly*/

END

```

#### 7.4.2 Daily Goal

The module Daily Goal allow the users to see their daily total earned points. This module calculates the whole earning and compare with a range of bonus. If the user will cross that range of bonus then they will be awarded with some extra bonus points. Through calendar they can track their total pays of any day.

#### Pseudo Code

```

BEGIN

DO Login

READ Userid

GO TO Daily Goal Module

DECLARE earnings = 0

DECLARE rangeOfBonus = 0

DECLARE bonus = 0

```

### READ EarnedPoints

LOOP READ EarnedPoints TO end of rows STEP 1 /\* Reading total earning of the user from database till the date\*/

SET earnings = earning + EarnedPoints

NEXT Userid

END LOOP

READ earnings

SET rangeOfBonus = round ( earning \* 0.1 )

READ rangeOfBonus

SET bonus = round ( rangeOfBonus / 10 )

IF earning >= rangeOfBonus /\* For the current date \*/

    earnings = earnings + bonus

END IF

DISPLAY "Earn" + range of bonus + "Get" + bonus

DISPLAY earnings

END

### 7.4.3 Analysis

The module Analysis is done on user responses by fetching the record and transferring it to HDFS from Local system. The analysis is done by calculating the total occurrence of positive and negative words in the record. Hence, by seeing the result on an average it will return the output as how many users are in favor of the subject and how many are not. It will predict the changing attitude of the society.

### Pseudo Code

\*\*\*\*\* Mapper \*\*\*\*\*

BEGIN

DECLARE IntWritable one

DECLARE Text word

SET one = new IntWritable(1) /\* 1 will be printed in output \*/

SET word = new Text()

```

DECLARE map(Object key, Text value, Context context)

DECLARE StringTokenizer itr

SET itr = new StringTokenizer(value.toString())

WHILE (itr.hasMoreTokens())      /* To the length of the file */

    word.set(itr.nextToken());   /* For every word */

    context.write(word, one);    /* places one for every word count */

    NEXT itr

END WHILE

END

```

*\*\*\*\*\* Reducer \*\*\*\*\**

```

BEGIN

DECLARE IntWritable result

SET result = new IntWritable();

DECLARE reduce(Text key, Iterable<IntWritable> values, Context context)

DECLARE int sum = 0

FOR (IntWritable val : values)

    SET sum += val.get()      /* adding 1s of each word to show total occurrence of a
word in the file */

END FOR

result.set(sum);

context.write(key, result)  /* writes the result */

END

```

*\*\*\*\*\* Job Calling in Main \*\*\*\*\**

```

DECLARE Configuration conf

SET conf = new Configuration();

DECLARE String[] otherArgs = new GenericOptionsParser(conf, args).getRemainingArgs();

IF (otherArgs.length != 2)

```

```

DISPLAY ("Usage: wordcount <in> <out>");
System.exit(2);

DECLARE Job job
SET job = new Job(conf, "word count");
job.setJarByClass(WordCount.class);
job.setMapperClass(TokenizerMapper.class);
job.setCombinerClass(IntSumReducer.class);
job.setReducerClass(IntSumReducer.class);
job.setOutputKeyClass(Text.class);
job.setOutputValueClass(IntWritable.class);
FileInputFormat.addInputPath(job, new Path(otherArgs[0])); /* file input path */
FileOutputFormat.setOutputPath(job, new Path(otherArgs[1])); /* file output path */
System.exit(job.waitForCompletion(true) ? 0 : 1);

END
  
```

Creates the class file

## 7.5 Complex Feature Tried But Not Completed

The developer has tried to develop many codes for successfully implementing the notification module in the system but unfortunately not able to complete. The code runs the interface but not able to ping the actions on it i.e. total earning should be updated and automatically reported timely. The developer has applied different ways to write the code to fulfill the module requirement but again then receives failed output. It was difficult to get the problem that may be related server compatibility or somewhere in the code or jar files.

The code imports various packages related to “applicationnotifier” and “springframework” in order to generate notification framework combining with spring framework beans factory. It will then connect with the glassfish server to broadcast and bound the events. But regrettably, the developer was not able to catch the issue. The below mentioned code snippet shows the unproductive implementation of the notification module. It contains a “NotificationHandler” class having a responsive method to get the summary of an event occurrence.

```

public class NotificationHandler {

    final static Logger logger = LoggerFactory.getLogger(NotificationHandler.class);
    @Autowired
    @Qualifier("notificationDaoImpl")
    NotificationDao notificationDao;

    Map<String, SseBroadcaster> notificationBroadcasterMap = new HashMap<String, SseBroadcaster>();

    @Path("/register/{userName}")
    @Produces(SseFeature.SERVER_SENT_EVENTS)
    @GET
    public @ResponseBody EventOutput registerForAnEventSummary(
        @PathParam("userName") String userName) {
        try {
            NotificationFrameworkFactory factory = new NotificationFrameworkFactory();

            EventOutput eventOutput = new EventOutput();

            List notificationTypes = getAllNotificationTypes();

            for (String notificationType : notificationTypes) {
                NotificationFrameworkInterface notificationInterface = factory
                    .getNotifieir(notificationType);
                String keyVal = getKeyVal(notificationType, userName);
                if (!notificationBroadcasterMap.containsKey(keyVal)) {
                    notificationBroadcasterMap.put(keyVal,
                        notificationInterface.getBroadcaster());
                }
                notificationBroadcasterMap.get(keyVal).add(eventOutput);
            }
            return eventOutput;
        } catch (NullPointerException exception) {
            logger.error("Exception Occurred: ", exception);
        }
        return null;
    }
}
  
```

**Figure 27: Notification Code Snippet**

## 7.6 Complex Feature of Medium Used

In the development of Earning Explorer, the developer has done the user registration module and analysis module by using the complex feature of Java and Hadoop technology.

### 7.6.1 Java Programming

In Java, the system includes the java mail server (enhanced SMTP feature) to send the secure email to user's email address. SMTP stands for Simple Mail Transfer Protocol and it is an internet standard for electronic mail (email) transmission. It creates a "JavaMail" session and "MimeMessage" object with proper handling "MessagingException". These all together creates an environment to send email by passing the sender and email address and it also verifies whether the receiver already registered or not.

```

msg.setSubject("Earning Explorer: Confirm Registration");
StringBuilder sb = new StringBuilder();
for (int i = 0; i < to.length; i++) {
    String item = to[i];
    sb.append(item);
}
String toAdrs = sb.toString();
System.out.println(toAdrs);
//EmailDemo ed = new EmailDemo();
msg.setContent("<html>\n" +
    "<body>\n" +
    "\n" +
    "<a href=\"http://localhost:8080/Earning_Explorer_using_Hadoop_Streaming/successRegistration.jsp\">" +
    "</a>\n\n Username: " +
    "\n"+toAdrs+"\nPassword: "+userpass+
    "</body>\n" +
    "</html>", "text/html");
Transport transport = session.getTransport("smtp");
transport.connect(host,from,password);
transport.sendMessage(msg, msg.getAllRecipients());
transport.close();

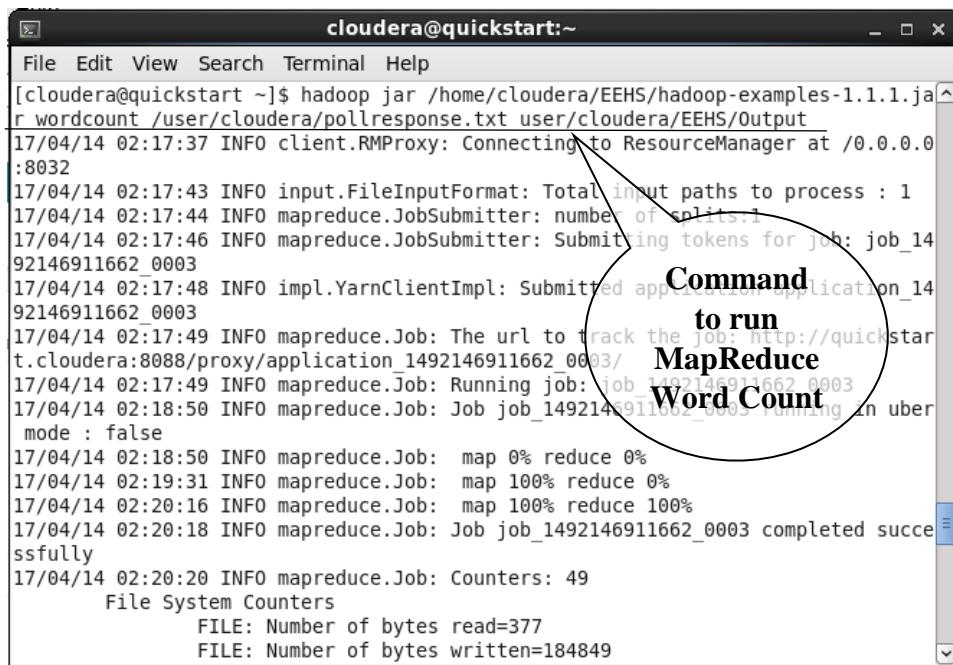
```

Sending registration link in email

Figure 28: Mail Code Snippet

## 7.6.2 Hadoop Technology

In Hadoop, the developer has done the analysis of user responses by using its major components that are HDFS for response data storage and Hadoop MapReduce Framework. The file of which analyses is to be done is placed in HDFS by Linux command and from there it will be used by the MapReduce framework to run the word count program on it. The program has been written in Java and will run by Linux command on HDFS. To run the commands we will use Cloudera terminal.

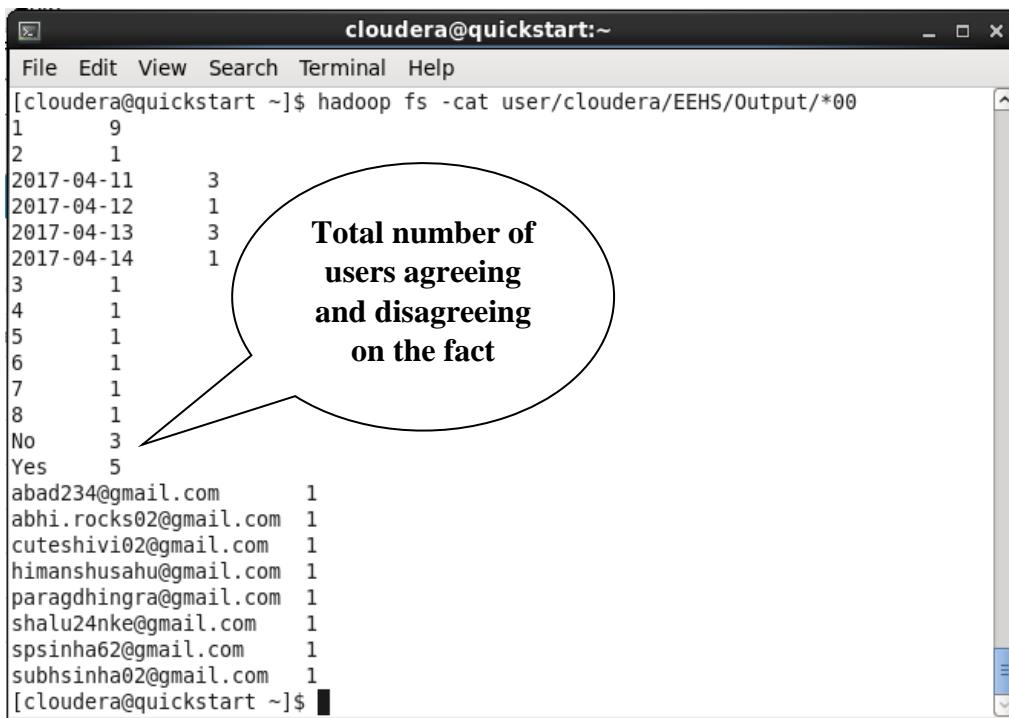


The terminal window title is "cloudera@quickstart:~". The command entered is "hadoop jar /home/cloudera/EEHS/hadoop-examples-1.1.1.jar wordcount /user/cloudera/pollresponse.txt user/cloudera/EEHS/Output". The output log shows:

- INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
- INFO input.FileInputFormat: Total input paths to process : 1
- INFO mapreduce.JobSubmitter: number of splits:1
- INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1492146911662\_0003
- INFO impl.YarnClientImpl: Submitted application application\_1492146911662\_0003
- INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application\_1492146911662\_0003/
- INFO mapreduce.Job: Running job: job\_1492146911662\_0003
- INFO mapreduce.Job: Job job\_1492146911662\_0003 running in uber mode : false
- INFO mapreduce.Job: map 0% reduce 0%
- INFO mapreduce.Job: map 100% reduce 0%
- INFO mapreduce.Job: map 100% reduce 100%
- INFO mapreduce.Job: Job job\_1492146911662\_0003 completed successfully
- INFO mapreduce.Job: Counters: 49
- File System Counters
  - FILE: Number of bytes read=377
  - FILE: Number of bytes written=184849

Command to run MapReduce Word Count

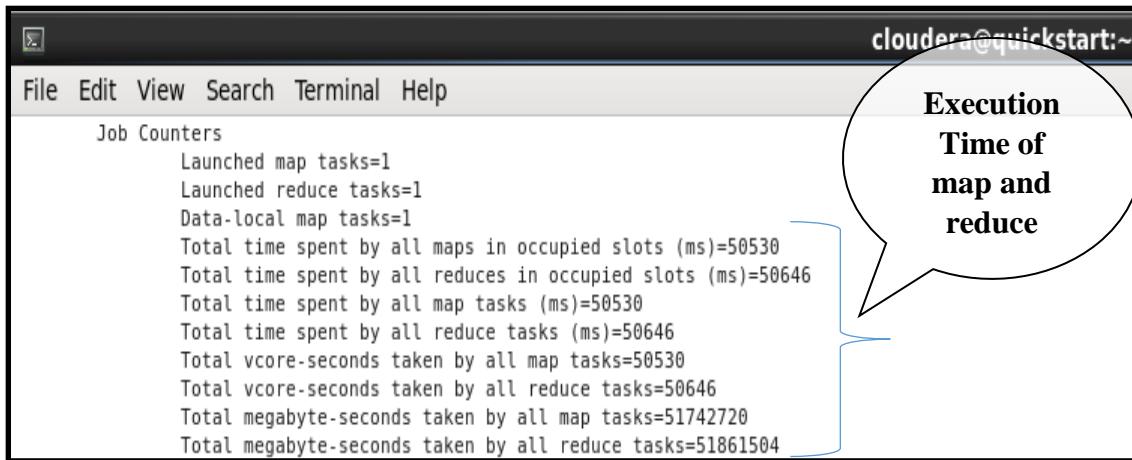
Figure 29: Running MapReduce



```

cloudera@quickstart:~$ hadoop fs -cat user/cloudera/EEHS/Output/*00
1      9
2      1
2017-04-11   3
2017-04-12   1
2017-04-13   3
2017-04-14   1
3      1
4      1
5      1
6      1
7      1
8      1
No     3
Yes    5
abad234@gmail.com 1
abhi.rocks02@gmail.com 1
cuteshivi02@gmail.com 1
himanshusahu@gmail.com 1
paragdhingra@gmail.com 1
shalu24nke@gmail.com 1
spsinha62@gmail.com 1
subhsinha02@gmail.com 1
[cloudera@quickstart ~]$
  
```

Figure 30: Analysis Output



```

cloudera@quickstart:~$ 
File Edit View Search Terminal Help
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=50530
  Total time spent by all reduces in occupied slots (ms)=50646
  Total time spent by all map tasks (ms)=50530
  Total time spent by all reduce tasks (ms)=50646
  Total vcore-seconds taken by all map tasks=50530
  Total vcore-seconds taken by all reduce tasks=50646
  Total megabyte-seconds taken by all map tasks=51742720
  Total megabyte-seconds taken by all reduce tasks=51861504
  
```

Figure 31: MapReduce Framework Performance

## 7.7 Technical Quality According to Software Engineering Standard

The technical quality is required to get the assurance of systems technical requirements and expectations. The extent to which the project is specific to its functionalities. It will done by examining the functional quality and interface designing quality. The following description includes the report of the quality investigation done on the system.

### 7.7.1 Memory Utilization

It measure the style of code and programing concepts used in the operations of the system functions in order to check how much memory space the program requires for its execution. The developer has used the concept of arrays for utilizing less memory space as it utilizes the memory according to the requirement.

```

try
{
  msg.setFrom(new InternetAddress(from));
  InternetAddress[] toAddress = new InternetAddress[to.length];
  for(int i =0; i< to.length ;i++)
  {
    toAddress[i] = new InternetAddress(to[i]);
  }
  for(int i =0; i< toAddress.length ;i++)
  {
    msg.addRecipient(Message.RecipientType.TO, toAddress[i]);
  }
}

```

One  
dimensional  
array

Figure 32: Array list Used

```

public String mainFun(String emailid, String pass)
{
  String id = emailid;
  String[] to = {id};
  DAO_Service ds = new DAO_Service();
  StringBuilder sb = new StringBuilder();
  for (int i = 0; i < to.length; i++) {
    String item = to[i];
    sb.append(item);
  }
}

```

String  
Array

Figure 33: String Array

### 7.7.2 Performance

The system's perform is measured to check the execution time taken on starting the application or how much time it is consuming for completing a function. The developer has used JSP which is suggestively better because it allow embedding Dynamic Elements in HTML pages itself instead of using Common Gateway Interface (CGI). It uses less time in data processing between the pages.

```

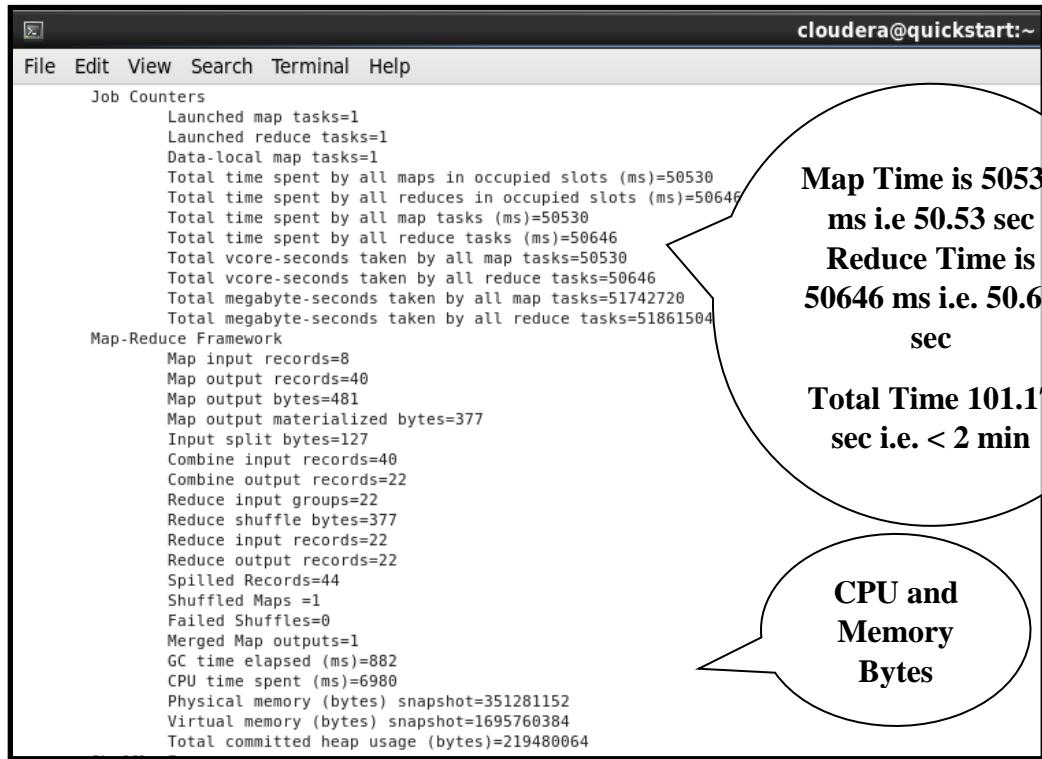
<sql:setDataSource var="dbsource" driver="com.mysql.jdbc.Driver"
    url="jdbc:mysql://localhost:3306/eehs"
    user="root" password="mysql"/>
<sql:query dataSource="${dbsource}" var="result1">
    select refercode from registration where userid = "${sessionScope.userID}" ;
</sql:query>

<c:forEach var="row" items="${result1.rows}">
<c:set var="code" value="${row.refercode}" scope="page"/>
<h3 class="section-heading"><b>Refer Your Friends Through Email.</b></h3>
<hr class="light">
<div class="col-md-12">
    <form data-toggle="validator" name="login" class="form" role="form" method="post" action="sendingCode.jsp" accept-charset="UTF-8">
        <div class="form-group">
            <span class="input-group-addon">Friend's Address</span>
            <br>
            <input type="email" style="width: 300px;" id="exampleInputEmail2" name="femail" placeholder="Email address" required="required" />
            <br><br>
            <textarea rows="5" cols="86" id="message" name="message" placeholder="Add a custom note(Optional)"/>
            <div class="help-block with-errors"></div>
        </div>
        <div class="form-group">
            <span class="input-group-addon">Referral Code</span>
            <br>
            <input type="text" value=<c:out value="${code}" /> class="form-control" id="exampleInputPassword2" name="rcode" required="required" />
        </div>
    </form>
</c:forEach>
  
```

**Embedded  
JSTL  
Tags**

Figure 34: JSP Code Snippet

The other performance effective technology used in the system is Hadoop MapReduce framework in which the Map function filters data on conditions, whereas the Reduce function combines data from the Map function to create the final output. It has assured essentials (Counters, Combiners and Partitioners) that improves the performance of data processing.



cloudera@quickstart:~

File Edit View Search Terminal Help

Job Counters

- Launched map tasks=1
- Launched reduce tasks=1
- Data-local map tasks=1
- Total time spent by all maps in occupied slots (ms)=50530
- Total time spent by all reduces in occupied slots (ms)=50646
- Total time spent by all map tasks (ms)=50530
- Total time spent by all reduce tasks (ms)=50646
- Total vcore-seconds taken by all map tasks=50530
- Total vcore-seconds taken by all reduce tasks=50646
- Total megabyte-seconds taken by all map tasks=51742720
- Total megabyte-seconds taken by all reduce tasks=51861504

Map-Reduce Framework

- Map input records=8
- Map output records=40
- Map output bytes=481
- Map output materialized bytes=377
- Input split bytes=127
- Combine input records=40
- Combine output records=22
- Reduce input groups=22
- Reduce shuffle bytes=377
- Reduce input records=22
- Reduce output records=22
- Spilled Records=44
- Shuffled Maps =1
- Failed Shuffles=0
- Merged Map outputs=1
- GC time elapsed (ms)=882
- CPU time spent (ms)=6980
- Physical memory (bytes) snapshot=351281152
- Virtual memory (bytes) snapshot=1695760384
- Total committed heap usage (bytes)=219480064

**Map Time is 50530  
ms i.e 50.53 sec  
Reduce Time is  
50646 ms i.e. 50.64  
sec**

**Total Time 101.17  
sec i.e. < 2 min**

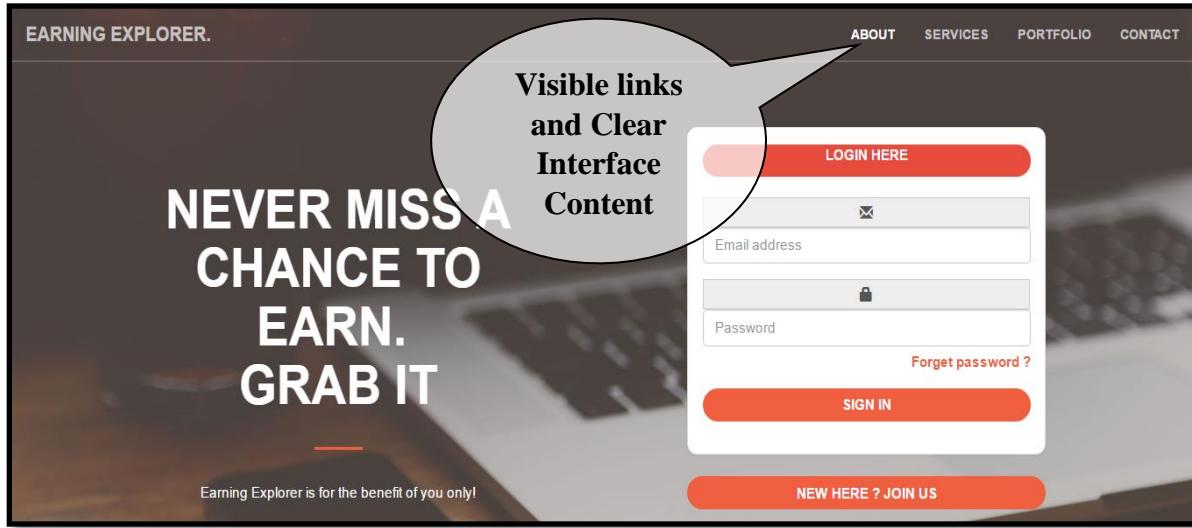
**CPU and  
Memory  
Bytes**

Figure 35: MapReduce Performance

### 7.7.3 User Interface

The design principles of Human Computer Interaction (HCI) have been implemented in the system that are Visibility, Feedback, Constraints, Mapping, Consistency and Affordance.

**Table 49: System Design Principles**

Design Principles Followed	
<b>Visibility</b>	
<p>The visibility of the system is determined by the contents available on the interface that should be clear and obvious. The users should be able to understand the use of controls easily. By achieving this principle all quality components are also achieved.</p>	
	
<b>Feedback</b>	
<p>Feedback is defined as the information which are provided by the system to the users in response to actions performed by the users. It increases the user confidence in system running state. We have achieved feedback principle in our system by providing the appropriate messages at every actions of the users. For example, the system have different colors for active and non-active links, mandatory fields.</p>	

**EARNING EXPLORER.**

ABOUT    **SERVICES**    PORTFOLIO    CONTACT

At Your Service

**Feedback of active link**

  
**Sturdy Work**  
Our system are updated regularly so they don't break trust.

  
**Ready to Ship Money**  
You can use this system, and earn !

  
**Up to Date**  
We update dependencies to keep things fresh.

  
**Made with Love**  
We care for our customers!

**LOGIN HERE**

✉

!

**Feedback of mandatory fields**

Please fill out this field.

[Forget password ?](#)

SIGN IN

**EARNING EXPLORER.**

**PLEASE CHECK YOUR EMAIL. IF NOT RECEIVED, PLEASE REGISTER. MAY BE INVALID EMAIL ID.. THANK YOU!**

**Success message of Mail**

### Constraint and Mapping

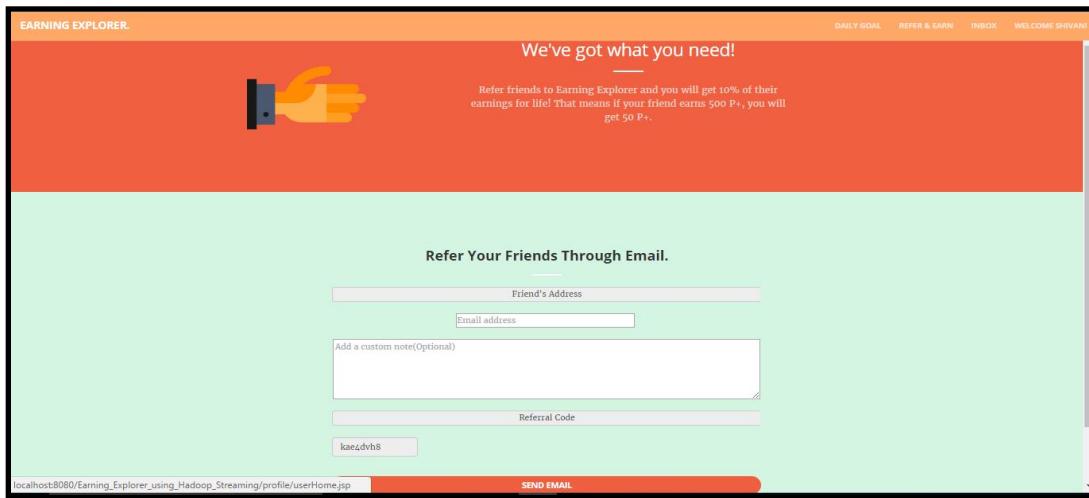
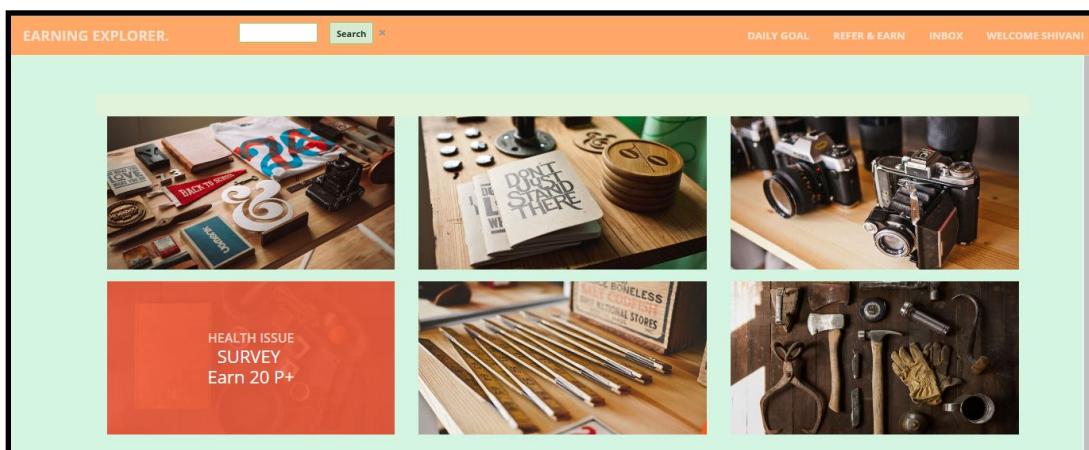
Constraints are the features that are applied in the system to reduce the chance of error occurrence. The property of an object that restrict the approach in which it can be done. In the system, the controls have been designed informatively so that user can understand how to use it.

Mapping defines the positions and sequence of controls. Controls should be relative to each other. We have achieved this by using icons that are universally accepted.



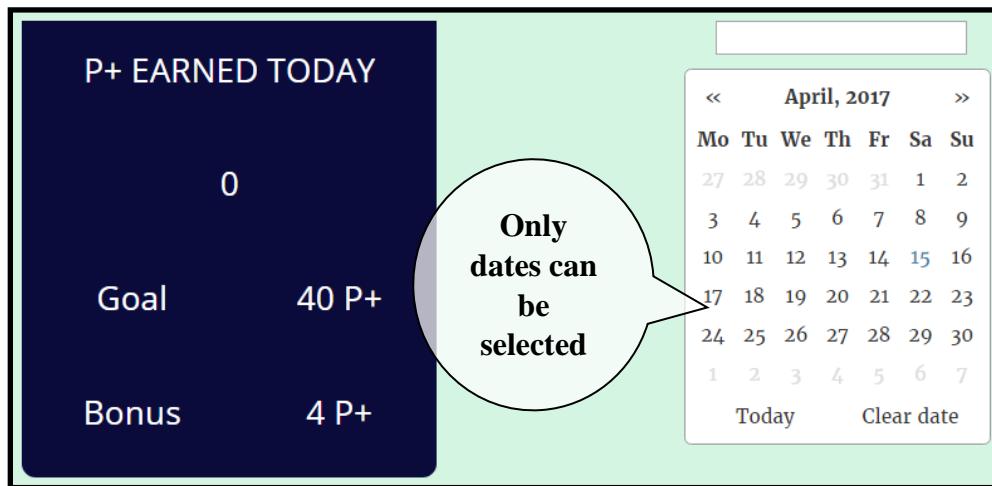
### Consistency

Consistency means same pattern. Interface design have similar operations and use similar elements for similar task. In the system, same design across the screen within same application.



## Affordance

It explains the property of object that automatically indicates how it should be used. We have achieved it by using date picker controls in the system.



### 7.1.1 Proper Naming Convention, Comments and Indentation

It emphasizes on the names used for classes, methods and variables in the programs which should be meaningful and testable. The names should highlight the use of the function. The code should include the comments and indentation to increase the code readability.

```
package referralCode; //userdefined package for the java file

import java.security.SecureRandom;
import java.util.Random; // importing predefined package to create random number

public class ReferCode {
    public String createRandomCode(int codeLength){
        //chars contain all the alphabets and digits (0-9)
        char[] chars = "abcdefghijklmnopqrstuvwxyz1234567890".toCharArray();
        StringBuilder sb = new StringBuilder();
        Random random = new SecureRandom();
        for (int i = 0; i < codeLength; i++) {
            //variable c will contain the random values
            char c = chars[random.nextInt(chars.length)];
            //variable sb of StringBuilder will get unique code on appending
            sb.append(c);
        }
        String output = sb.toString();
        System.out.println(output);
        return output ;
    }
}
```

**Indent**

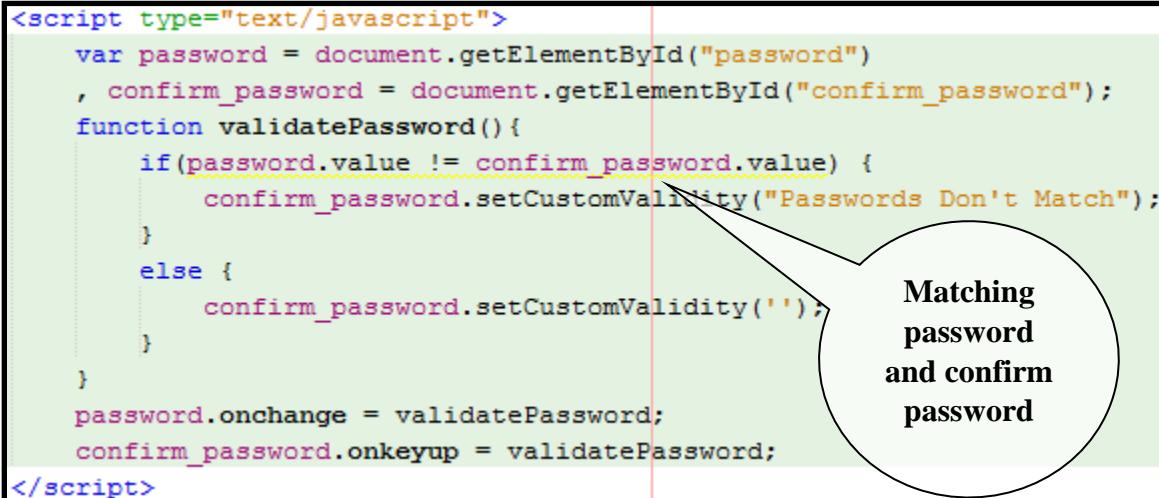
**Comment**

Figure 36: Naming Convention

#### 7.7.4 Validation

The validations are done to authenticate the date inputting and the data processing in the system. In Earning Explorer, the validation of user input and matching of values with database has been done. The below mentioned figures are the example of the system validation.

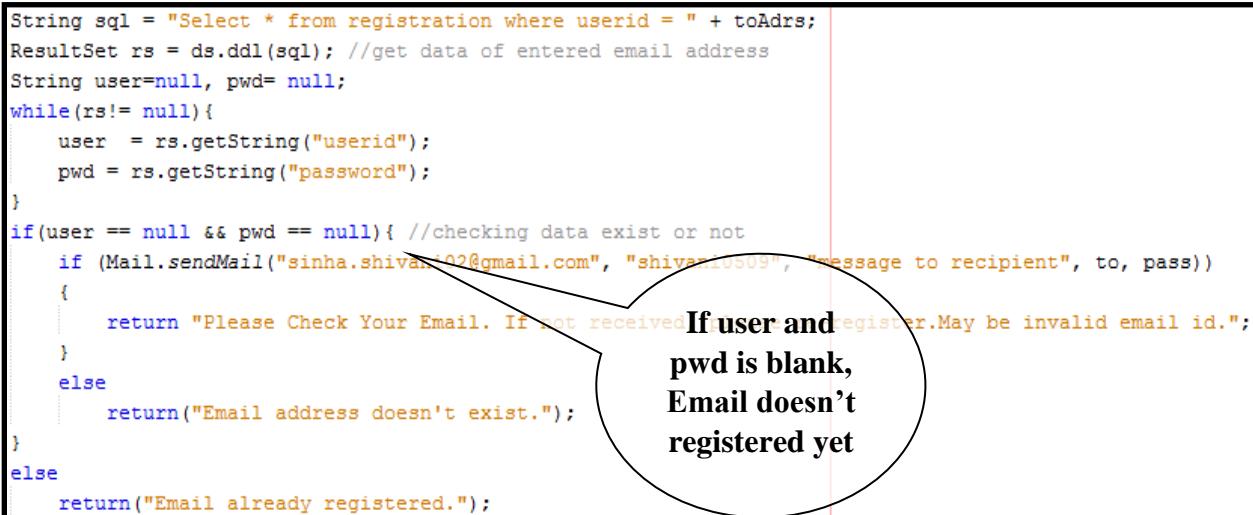
```
<script type="text/javascript">
  var password = document.getElementById("password");
  , confirm_password = document.getElementById("confirm_password");
  function validatePassword() {
    if(password.value != confirm_password.value) {
      confirm_password.setCustomValidity("Passwords Don't Match");
    }
    else {
      confirm_password.setCustomValidity('');
    }
  }
  password.onchange = validatePassword;
  confirm_password.onkeyup = validatePassword;
</script>
```



**Matching password and confirm password**

Figure 37: Password and Confirm Password Validation

```
String sql = "Select * from registration where userid = " + toAdrs;
ResultSet rs = ds.ddl(sql); //get data of entered email address
String user=null, pwd= null;
while(rs!= null){
  user = rs.getString("userid");
  pwd = rs.getString("password");
}
if(user == null && pwd == null){ //checking data exist or not
  if (Mail.sendMail("sinha.shivam02@gmail.com", "shivam10909", "message to recipient", to, pass))
  {
    return "Please Check Your Email. If not received register. May be invalid email id.";
  }
  else
    return("Email address doesn't exist.");
}
else
  return("Email already registered.");
```



**If user and pwd is blank, Email doesn't registered yet**

Figure 38: Already Exist Email Validation

## 7.8 Report Quality

The report quality is required to check the extent of project report's clarity and readability in terms of style, structure and language of the report. The following table elaborates the clear summary to the quality of which the project report has been made.

**Table 50: Report Quality Evaluation**

Quality Checklist For Evaluation Report		
Title Page		
<b>Includes</b>	<ul style="list-style-type: none"> <li>✓ Project Title</li> <li>✓ Submitted By</li> <li>✓ Supervisor Details</li> <li>✓ University Details with logo</li> </ul>	
<b>Comments</b>	<p>The project document includes a well formatted title page. The spacing between the informative lines are appropriate. Overall look of the title page is professional.</p>	
Project Preface		
<b>Includes</b>	<ul style="list-style-type: none"> <li>✓ Acknowledgement</li> <li>✓ Abstract</li> <li>✓ Certificate</li> <li>✓ Table of Content, Tables and Figures</li> </ul>	
<b>Comments</b>	<p>The document has included the details related to contributors and guiders of the project. A good acknowledgment has been shown for them. And the proper table navigation has been used for the contents, tables and figures included in the documentation.</p>	
Introduction and Background		
<b>Includes</b>	<ul style="list-style-type: none"> <li>✓ Topic of the system</li> <li>✓ Problem background and Objective of the system</li> <li>✓ Target reader</li> <li>✓ Functionality and scope</li> </ul>	

	<ul style="list-style-type: none"> <li>✓ Learning objectives and challenges</li> <li>✓ Feasibility studies</li> </ul>
<b>Comments</b>	The document has precisely comprised the information regarding project scope, functionalities and background studies. It has also quoted the learnings and challenges faced by the developer which is good to explain to show the technical knowledge required to develop the project.
<b>Main Body</b>	
<b>Includes</b>	<ul style="list-style-type: none"> <li>✓ Literature Review</li> <li>✓ Methodology</li> <li>✓ Analysis and Design</li> <li>✓ Implementation</li> <li>✓ Testing</li> </ul>
<b>Comments</b>	A suitable documentation has been made with the proper research work and explain the methodology chosen for the development of the project. A detailed analysis report has been represented followed by the designs of the system functionalities and the database. Finally testing module embraces test cases performed on the system from unit to the system in order to provide the full log of success and error reports.
<b>Wrap Up</b>	
<b>Includes</b>	<ul style="list-style-type: none"> <li>✓ Critical Evaluation</li> <li>✓ Conclusion</li> </ul>
<b>Comments</b>	A good explanation has been given to the system benefits for the users. Conclusion has thoroughly described the system limitations and future enhancement. At the last, experience of the developer has been quoted that shows how much efforts has been put into this project to make it to the success.

## 7.9 Technical Manual

To run the project on any other computer or laptop, need to install the below mentioned list of software on that device to make it compatible with the project's basic requirement. Follow the following steps.

**Table 51: Steps to install**

Steps	Instructions	Time
<b>1</b>	Install Java	5 min
<b>2</b>	Set Path of JAVA_HOME	5 min
<b>3</b>	Install NetBeans IDE 8.0.2	20 min
<b>4</b>	Install MySQL Workbench	10 min
<b>5</b>	Install Virtual Box	15 min
<b>6</b>	Install Cloudera Manager	5 min
<b>7</b>	Install WinScp-5.9.4	10min

### 7.9.1 How to set up Earning Explorer in NetBeans IDE?

#### STEP 1: Start NetBeans IDE 8.0.2



Click on the icon to start the NetBeans and wait for the process to get completed i.e. Starting all the modules.

*Figure 39: NB Icon*



**Figure 40: Starting NetBeans IDE**

## STEP 2: Import File

Follow the process File > Import Project > From ZIP.

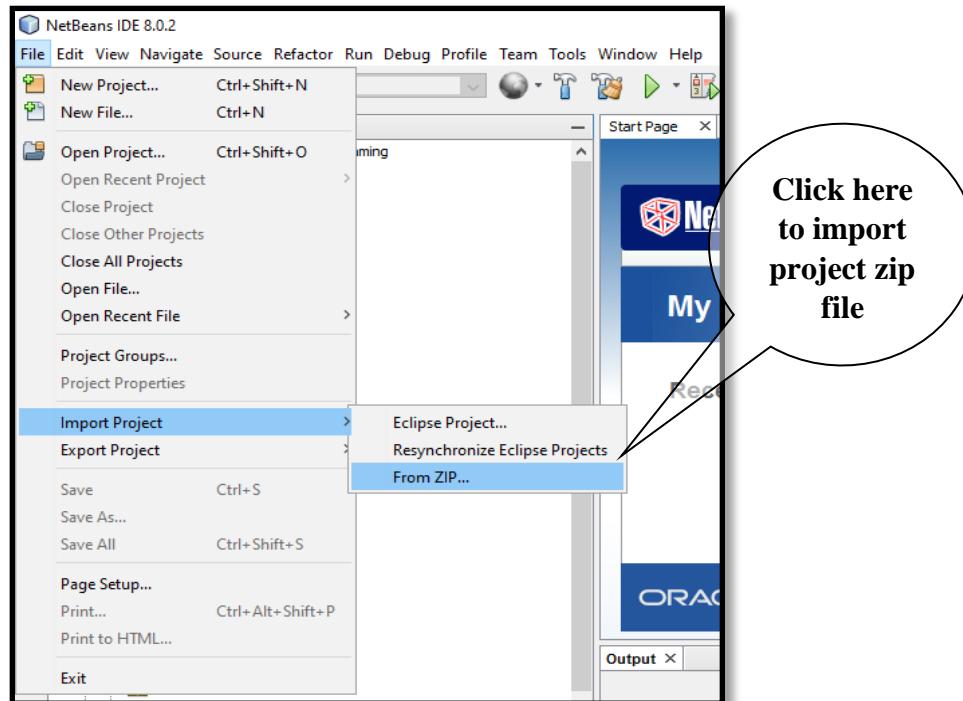


Figure 41: Importing Project

## STEP 3: Browse Zip File Location

Browse the location of Earning Explorer project file (named as Earning\_Explorer\_using\_Hadoop\_Streaming.zip) and click on import. It will take a little time and then the project will be available in projects section at the top left corner of the window.

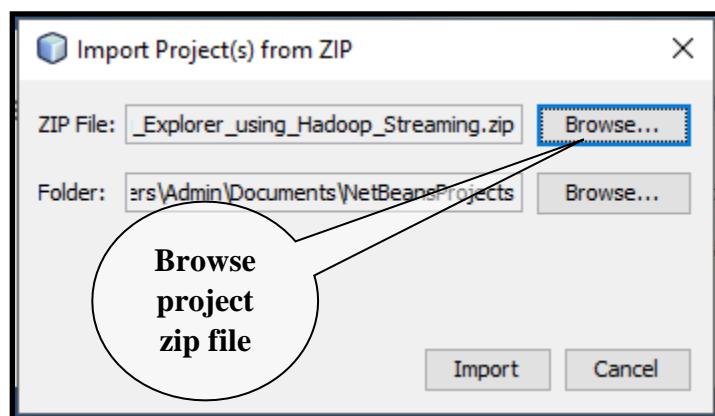


Figure 42: Browse Zip File

## STEP 4: Check Project Window

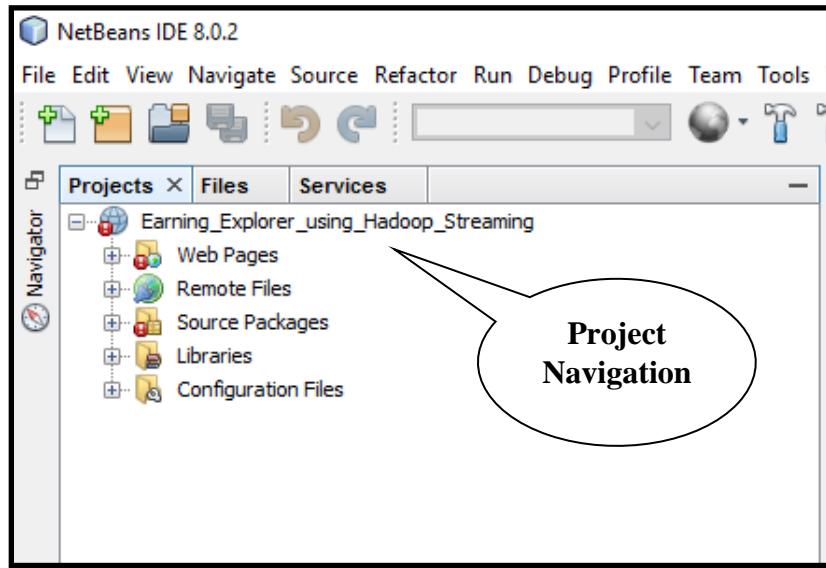
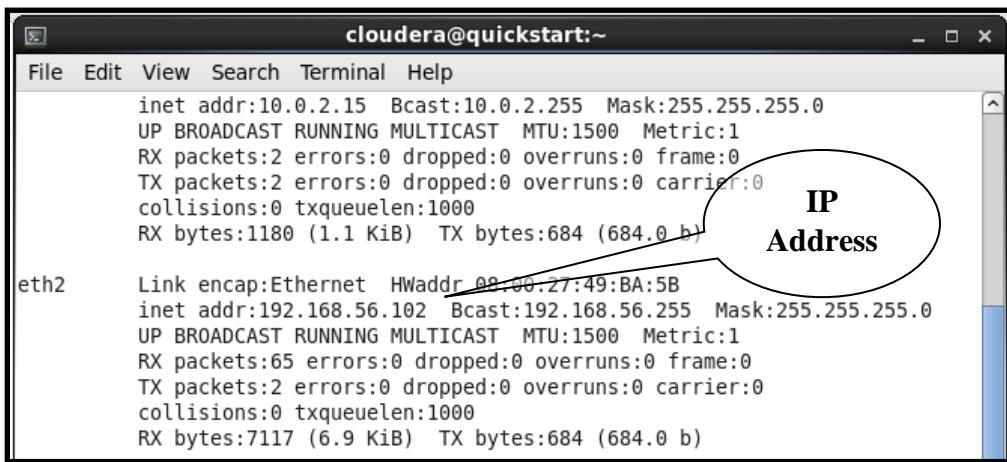


Figure 43: Project Navigation

### 7.9.2 How to start file scheduling from Local System to Hadoop Distributed File System?

#### STEP 1: Ping IP address of HDFS

To transfer file we need to fetch the IP address of HDFS and ping the same IP address from the software named WinSCP installed on local machine. For checking IP of the Linux machine i.e. Oracle VM Virtual Box, open the terminal of the machine and put command “ifconfig” and press enter key. It will return you the working IP address. Please check out the following image showing the IP address followed in Ethernet (eth2).



```
cloudera@quickstart:~
```

File Edit View Search Terminal Help

```
inet addr:10.0.2.15 Bcast:10.0.2.255 Mask:255.255.255.0
UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1
RX packets:2 errors:0 dropped:0 overruns:0 frame:0
TX packets:2 errors:0 dropped:0 overruns:0 carrier:0
collisions:0 txqueuelen:1000
RX bytes:1180 (1.1 KiB) TX bytes:684 (684.0 b)

eth2      Link encap:Ethernet HWaddr 08:00:27:49:BA:5B
          inet addr:192.168.56.102 Bcast:192.168.56.255 Mask:255.255.255.0
          UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1
          RX packets:65 errors:0 dropped:0 overruns:0 frame:0
          TX packets:2 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:7117 (6.9 KiB) TX bytes:684 (684.0 b)
```

A callout bubble with the text "IP Address" points to the IP address listed for the eth2 interface.

Figure 44: IP Address at HDFS

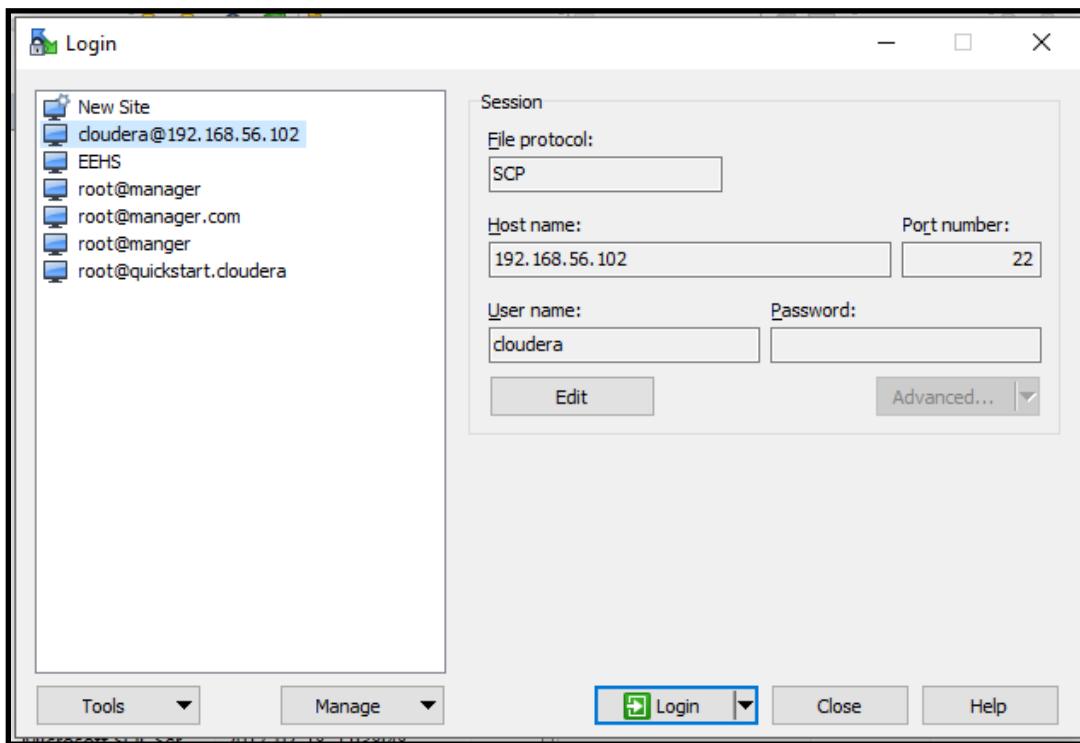
## STEP 2: Start the software WinSCP and Login through Host name



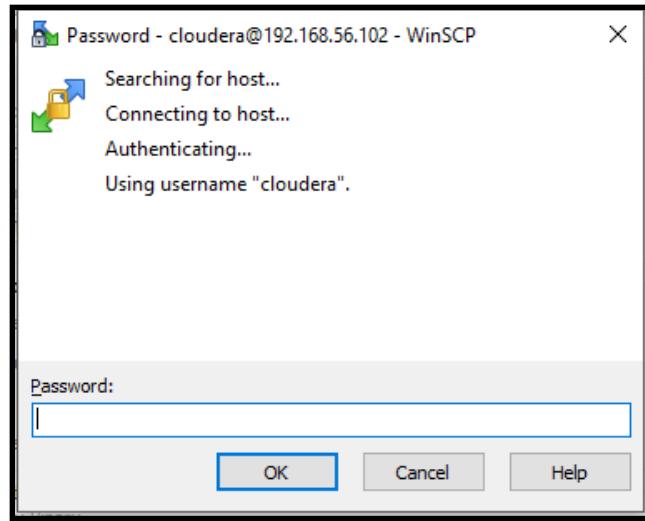
Click on the icon and wait to initiate the software. Once it got opened, it will come up with a login screen where we need to enter the details related to File Protocol, Host name, Port number, User name and Password. Please come in with the following mentioned values.

- Figure 45: Icon**
- Select SCP (Secure Copy Protocol) in File Protocol as it is based on Secure Shell (SSH) protocol. As most SSH and FTP servers run on Linux system.
  - Enter the IP 192.168.56.102 in Host name with Port number as 22.
  - Username = cloudera and Password = cloudera.

If we will not enter the password at the time of login, the software will ask for password once the host name connection has been found.



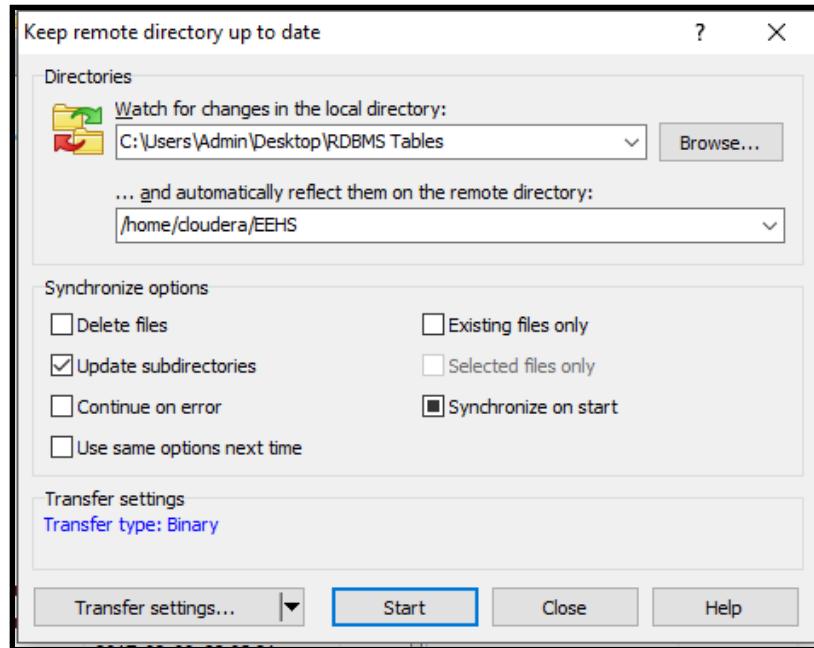
**Figure 46: WinSCP network login**



**Figure 47: Authenticating the Host name**

### STEP 3: Provide the file location.

We are required to provide the location of the folder available at local system, which we want to transfer to a particular location in HDFS. In the following figure “RDMS Tables” is located at local and “EEHS” is at HDFS.



**Figure 48: File Location**

We can change the synchronization settings as per the requirement, but in this case we only require to update the file. Click on start and begin the full synchronization.

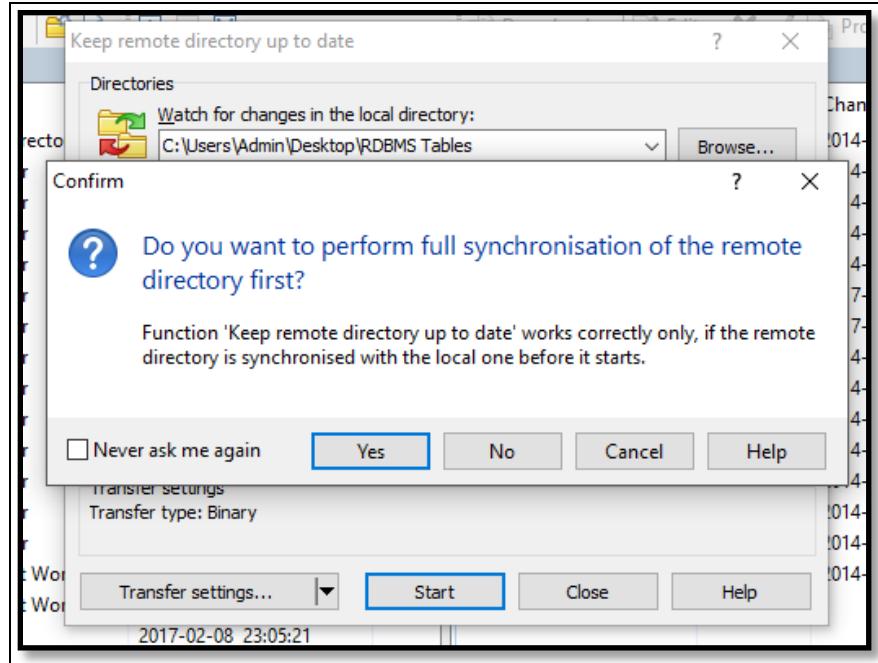


Figure 49: WinSCP Starts Synchronization

#### STEP 4: Check out the synchronized file update

The files will be automatically updated in HDFS on every action triggered on Local System. The synchronization will be continued until the system well get turn off.

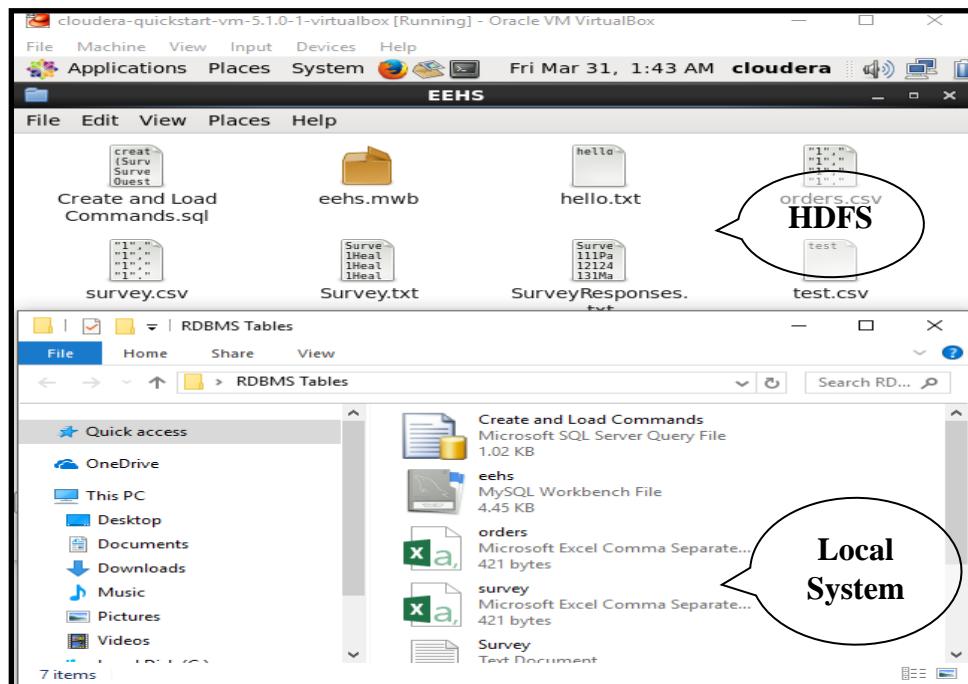


Figure 50: File at Local System and HDFS

## 7.10 User Manual

The user manual will help the users to understand the directions to use the system without any problem. The following steps should be followed by the users to become a part of the system and enjoy the tasks.

### STEP 1: Earning Explorer Home

This is the main page of the system through this page user can register themselves and login with that registered user id and password.

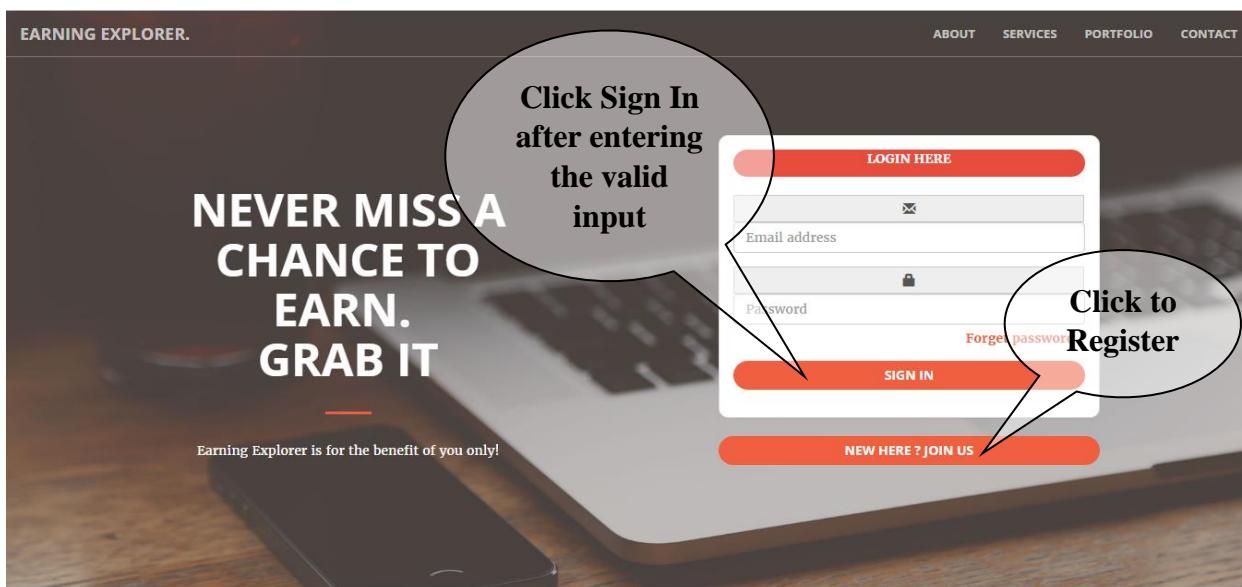


Figure 51: Earning Explorer Home

### STEP 2: Registration

Users are required to register by entering the following details and click on register button.

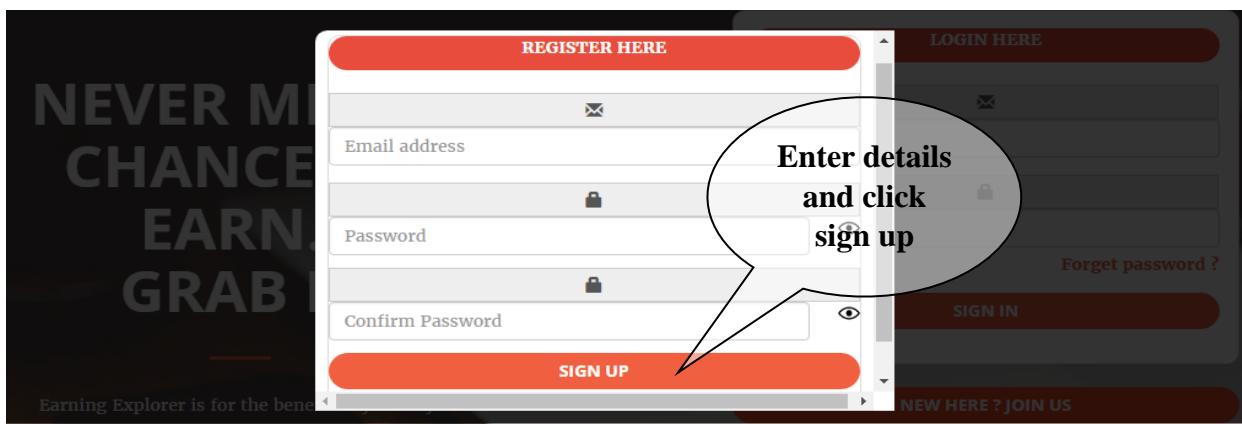
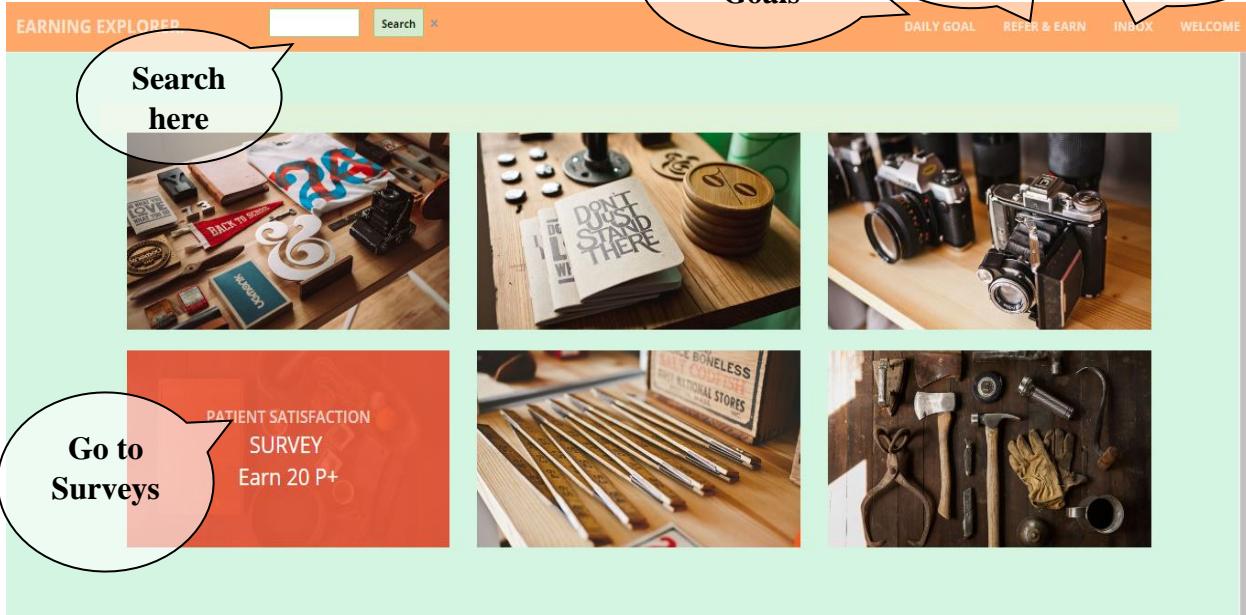


Figure 52: Registration

### STEP 3: User Home

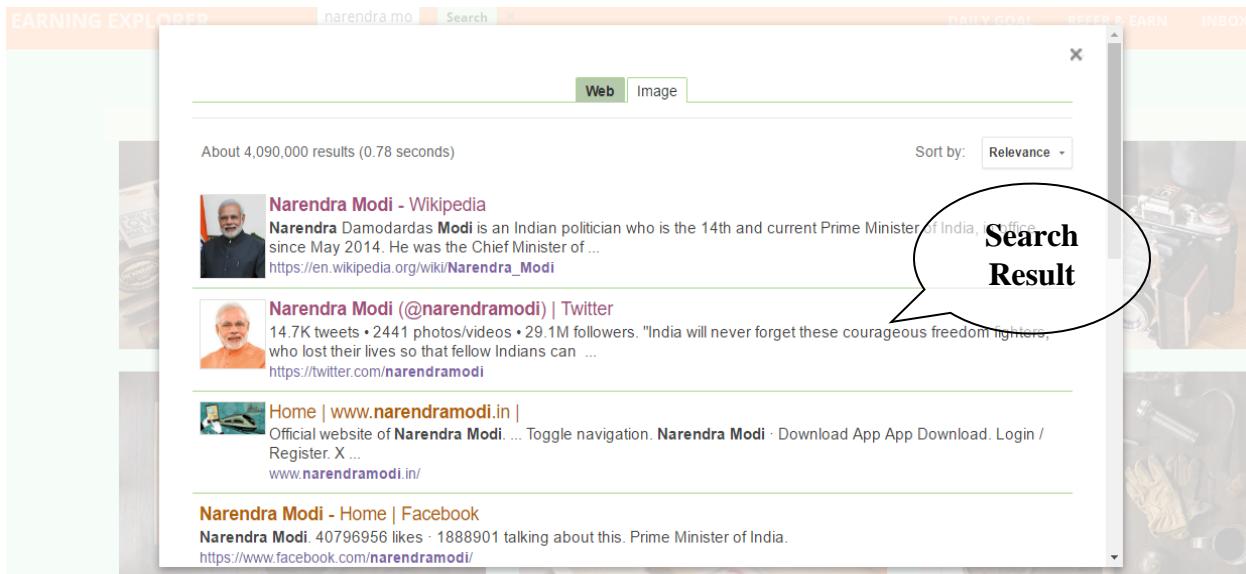
This is the user panel from where users can do several tasks by visiting different links presented for their earnings.



**Figure 53: User Home**

### STEP 3: Search

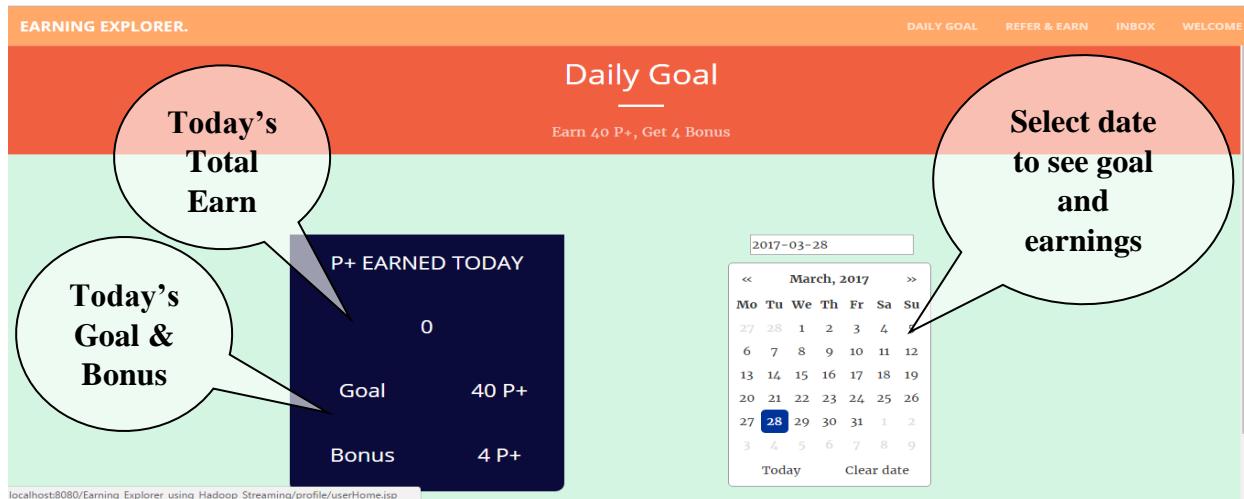
This panel allows the user to search on any query and see the result provided by the system.



**Figure 54: Search**

## STEP 4: Daily Goal

This panel allows the user to check their daily earnings and see how much more amount they need to earn to get extra bonus points.

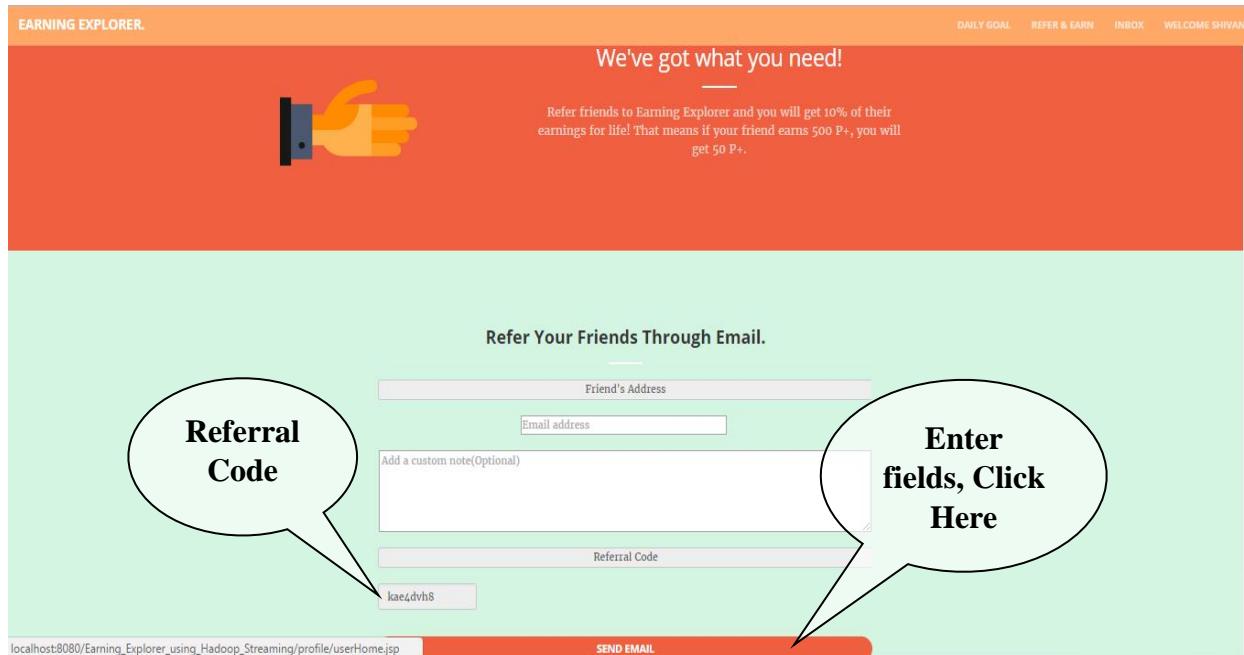


The screenshot shows the 'EARNING EXPLORER' interface. At the top, there are navigation links: DAILY GOAL, REFER & EARN, INBOX, and WELCOME. Below the header, a large orange section titled 'Daily Goal' displays the message 'Earn 40 P+, Get 4 Bonus'. A central dark blue box shows 'P+ EARNED TODAY' (0), 'Goal' (40 P+), and 'Bonus' (4 P+). Two speech bubbles point to this area: one labeled 'Today's Total Earn' and another labeled 'Today's Goal & Bonus'. To the right, a large green section contains a date picker for 'Select date to see goal and earnings', showing the month of March 2017. The URL at the bottom is 'localhost:8080/Earning\_Explorer\_using\_Hadoop\_Streaming/profile/userHome.jsp'.

**Figure 55: Daily Goal**

## STEP 5: Refer and Earn

From here one user can refer the system to other system and can earn money parallel from the friend's earnings.

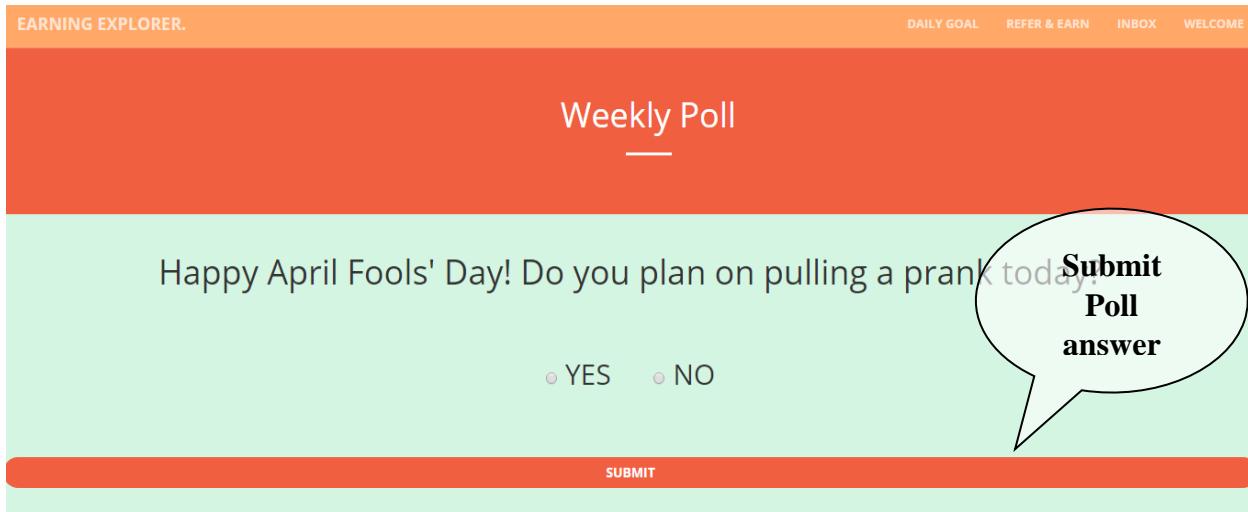


The screenshot shows the 'EARNING EXPLORER' interface. At the top, there are navigation links: DAILY GOAL, REFER & EARN, INBOX, and WELCOME SHIVANI. A message 'We've got what you need!' is displayed above a hand icon. Below it, text explains that referring friends will earn 10% of their friend's earnings. A green section titled 'Refer Your Friends Through Email.' contains a form with fields for 'Friend's Address' (Email address), 'Add a custom note(Optional)', 'Referral Code' (with placeholder 'kae4dvh8'), and a 'SEND EMAIL' button. Two speech bubbles point to the form: one labeled 'Referral Code' pointing to the Referral Code field, and another labeled 'Enter fields, Click Here' pointing to the 'SEND EMAIL' button. The URL at the bottom is 'localhost:8080/Earning\_Explorer\_using\_Hadoop\_Streaming/profile/userHome.jsp'.

**Figure 56: Refer and Earn**

## STEP 6: Weekly Poll

From here user can earn plus points (P+) by simply answering the asked poll question.

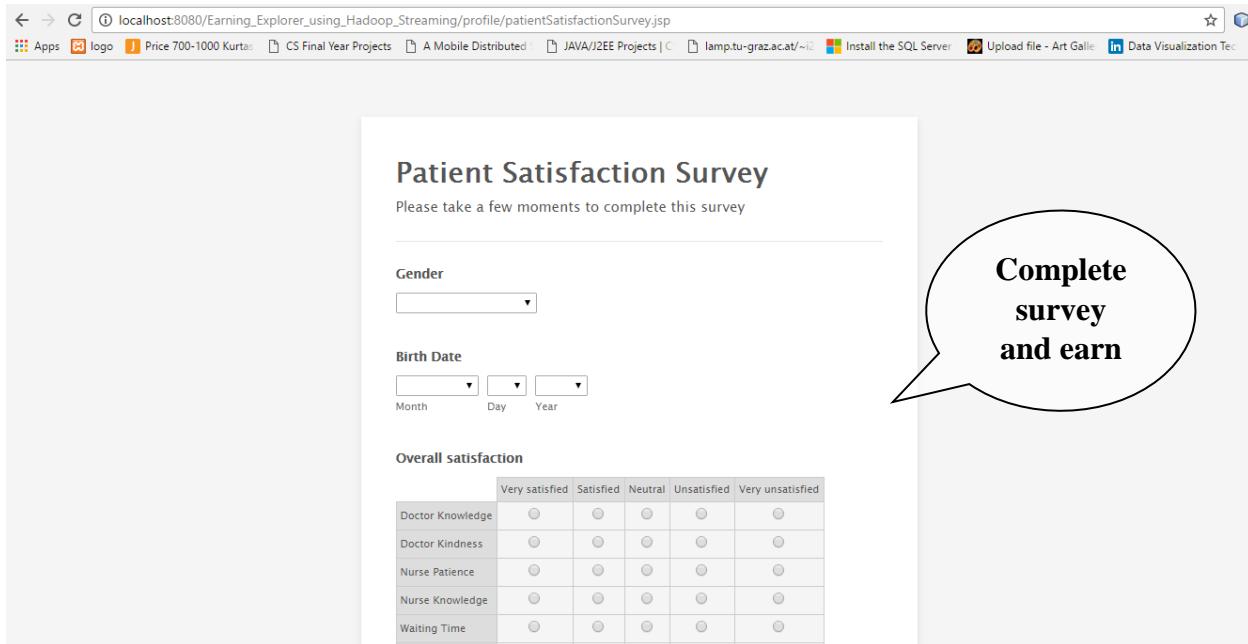


The screenshot shows a web-based poll interface. At the top, there's a navigation bar with links for 'DAILY GOAL', 'REFER & EARN', 'INBOX', and 'WELCOME'. Below this is a red header bar with the text 'Weekly Poll'. The main content area has a light green background. It displays a poll question: 'Happy April Fools' Day! Do you plan on pulling a prank *today*?'. Below the question are two radio button options: 'YES' and 'NO'. At the bottom right of the main area is a large white speech bubble containing the text 'Submit Poll answer'. A red footer bar at the bottom contains a 'SUBMIT' button.

**Figure 57: Weekly Poll**

## STEP 7: Survey

By clicking on any of the survey, from home page the users will get directed to that particular survey page provided by the facilitator. In the below screen we are on patient satisfaction survey form. After completing the survey user will be awarded with the associated Plus Points (P+).



The screenshot shows a web-based survey interface titled 'Patient Satisfaction Survey'. The URL in the browser is 'localhost:8080/Earning\_Explorer\_using\_Hadoop\_Streaming/profile/patientSatisfactionSurvey.jsp'. The page includes instructions: 'Please take a few moments to complete this survey'. It features several input fields: 'Gender' (dropdown), 'Birth Date' (dropdown with Month, Day, Year fields), and an 'Overall satisfaction' section. This section contains a grid where users can rate various factors on a scale from 'Very satisfied' to 'Very unsatisfied'. A large white speech bubble on the right side of the form contains the text 'Complete survey and earn'.

	Very satisfied	Satisfied	Neutral	Unsatisfied	Very unsatisfied
Doctor Knowledge	<input type="radio"/>				
Doctor Kindness	<input type="radio"/>				
Nurse Patience	<input type="radio"/>				
Nurse Knowledge	<input type="radio"/>				
Waiting Time	<input type="radio"/>				

**Figure 58: Survey**

## 7.11 Conclusion

The chapter has successfully shown the implementation of each module of the proposed project with detailed explanation of the achievements. It has included all the report of technical quality of the system, technical manual for the deployment of the project and screenshot of the user interfaces of the system. Every part done in the implementation phase shows the developer efforts to make the project successful.

The whole project work got completed on time because the developer has strictly followed the implementation plan shown in the chapter. After covering whole module the chapter also elaborated the complex features of the system and the medium used (Java and Hadoop).

## CHAPTER 8

# TESTING

### 8.1 Introduction

Testing is the process of executing a program and finding out and solving all possible bugs. It is the presence of error in the project but it will not ensure the absence of errors in the project. It helps in exploring the product quality. Testing includes: (*Softwaretestingclass.com, 2016*)

- Verification (Step by step checking of milestones)
- Validation (Final check of the user's requirements)

### Testing Processes

**Testing**      **Unit Testing** (On login click, passing username and password)

**Integration Testing** (Login module)

**System Testing** (User and System functionality)

**White Box Testing** (Validate functional and non-functional requirement)

**User Acceptance Testing** (User friendly system)

**Usability Testing** ( Website usability)

**Compatibility Testing** (Software Compatibility)

**Runtime Testing** (Website Efficiency)

**Figure 59: Testing Types**

Software development process includes testing in order to remove errors before the delivery of the project. It avoids errors by detecting and correcting the problems occurred during testing of each and every module so that user can efficiently use the system. The test processes, techniques and tools are major contributors to effective and efficient testing and quality assurance. It helps in investigating the whole system from deep (functions and validations) to top (User Interface).

## 8.2 Test Plan

### 8.2.1 Test Duration

**Table 52: Test Duration**

Test ID	Test Strategy	Start Date	End Date
T1	Unit Testing	21 <sup>st</sup> Feb 2017	27 <sup>th</sup> Feb 2017
T2	Integration Testing	28 <sup>th</sup> Feb 2017	6 <sup>th</sup> March 2017
T3	System Testing	7 <sup>th</sup> March 2017	9 <sup>th</sup> March 2017
T4	White Box Testing	10 <sup>th</sup> March 2017	14 <sup>th</sup> March 2017
T5	User Acceptance Testing	15 <sup>th</sup> March 2017	20 <sup>th</sup> March 2017
T6	Usability Testing	15 <sup>th</sup> March 2017	20 <sup>th</sup> March 2017
T7	Compatibility Testing	21 <sup>st</sup> March 2017	22 <sup>nd</sup> March 2017
T8	Runtime Testing	23 <sup>rd</sup> March 2017	24 <sup>th</sup> March 2017

### 8.2.2 Test Cases to be tested

**Table 53: Test Cases**

Case	Test Cases	Description
TC1	Registration	Registration is used to enroll new user account in the system.
TC2	Login	Registered users can login to the system and can work on the interfaces provided to them by the system.
TC3	Search	Search allows the user to get the result for every search query they will enter in search bar.
TC4	Weekly Poll	Weekly poll allow the users to respond to the poll and earn money for their every responses on weekly basis.
TC5	Survey	The users can earn plus points for the surveys they will complete.
TC6	Daily Goal	Daily goal is the feature that tracks the whole earning of a user and set a goal. It provides bonus to them who reaches the goal.
TC7	Referral	User can get pays by referring the system to their friends.

<b>TC8</b>	Notification	The latest survey or weekly poll or any other message will get notify to the user's account.
<b>TC9</b>	Explorer TV	Explorer TV provides the facility to earn money by videos.
<b>TC10</b>	Settings	Users can edit their profiles by making changes from settings.
<b>TC11</b>	User Guide	Allow users to get the easy and usable guide to use the system well.
<b>TC12</b>	Visualization	Visualization is the average result of user responses in graphical form.

### 8.2.3 Testers List

**Table 54: Testers**

Tester ID	Tester Name	Occupation	Age Group
<b>UT1</b>	Shivani Sinha	Developer of the system	18-25
<b>UT2</b>	Shilu Kumari	House Wife	25-30
<b>UT3</b>	Dhananjay Kumar	Android Developer	30-35
<b>UT4</b>	Shubhangi Sinha	MBA Student	18-25
<b>UT5</b>	Parag Dhingra	Business Analyst	25-30

### 8.2.4 Testing Traceability

**Table 55: Registration Traceability**

Testing Process: UNIT TESTING (T1)	
<b>Test Case:</b>	TC1
<b>Tested Item:</b>	Registration (home.jsp)
<b>Performed By Tester ID:</b>	UT1, UT2, UT3, UT4, UT5
<b>Requirement Traceability:</b>	User will enter the details and clicks on registration button. Information will be transferred to the server.
<b>Testing Schedule:</b>	21 <sup>st</sup> Feb 2017
<b>Test Recording Procedure:</b>	Inputs are provided and execute for expected output.
<b>Hardware/ Software Requirement:</b>	Browser, NetBeans IDE, Glassfish Server and MySQL.
<b>Constraint:</b>	Enter only valid information. Internet Connection

**Table 56: Login Traceability**

<b>Testing Process: UNIT TESTING (T1)</b>	
<b>Test Case:</b> TC2	
<b>Tested Item:</b> Login (home.jsp)	
<b>Performed By Tester ID:</b> UT2, UT3, UT4, UT5	
<b>Requirement Traceability:</b> User will enter the details and clicks on registration button. Information will be transferred to the server.	
<b>Testing Schedule:</b> 21 <sup>st</sup> Feb 2017	
<b>Test Recording Procedure:</b> Inputs are provided and executed for taking the users to their respective home page.	
<b>Hardware/ Software Requirement:</b> Browser, NetBeans IDE, Glassfish Server and MySQL.	
<b>Constraint:</b> Authorized inputs are only allowed to successfully login.	

**Table 57: Search Traceability**

<b>Testing Process: UNIT TESTING (T1)</b>	
<b>Test Case:</b> TC3	
<b>Tested Item:</b> Search (userHome.jsp)	
<b>Performed By Tester ID:</b> UT2, UT3, UT4	
<b>Requirement Traceability:</b> Search query traced and display results with earning calculation. Check the options to sort the results i.e. by data or relevance.	
<b>Testing Schedule:</b> 22 <sup>nd</sup> Feb 2017	
<b>Test Recording Procedure:</b> Inputs are provided and execute for expected output.	
<b>Hardware/ Software Requirement:</b> Browser, NetBeans IDE, Glassfish Server and MySQL.	
<b>Constraint:</b> Internet Connection	

**Table 58: Weekly Poll Traceability**

<b>Testing Process: UNIT TESTING (T1)</b>	
<b>Test Case:</b> TC4	
<b>Tested Item:</b> Weekly Poll (weeklyPoll.jsp)	
<b>Performed By Tester ID:</b> UT4, UT5	

<b>Requirement Traceability:</b> User will select the option for their response for the poll and clicks on submit. After submit, calculate plus point for the account.
<b>Testing Schedule:</b> 22 <sup>nd</sup> Feb 2017
<b>Test Recording Procedure:</b> Inputs are provided and execute for expected output.
<b>Hardware/ Software Requirement:</b> Browser, NetBeans IDE, Glassfish Server and MySQL.
<b>Constraint:</b> None

**Table 59: Survey Traceability**

<b>Testing Process: UNIT TESTING (T1)</b>
<b>Test Case:</b> TC5
<b>Tested Item:</b> Survey (survey.jsp)
<b>Performed By Tester ID:</b> UT3, UT4, UT5
<b>Requirement Traceability:</b> User will do the survey and clicks on submit. After submit, calculate plus point for the account.
<b>Testing Schedule:</b> 22 <sup>nd</sup> Feb 2017
<b>Test Recording Procedure:</b> Inputs are provided and execute for expected output.
<b>Hardware/ Software Requirement:</b> Browser, NetBeans IDE, Glassfish Server and MySQL.
<b>Constraint:</b> None

**Table 60: Daily Goal Traceability**

<b>Testing Process: UNIT TESTING (T1)</b>
<b>Test Case:</b> TC6
<b>Tested Item:</b> Daily Goal (dailyGoal.jsp)
<b>Performed By Tester ID:</b> UT2, UT3, UT4, UT5
<b>Requirement Traceability:</b> In daily goal, see the total earning of a day with range of bonus.
<b>Testing Schedule:</b> 23 <sup>nd</sup> Feb 2017
<b>Test Recording Procedure:</b> Execute for expected output.
<b>Hardware/ Software Requirement:</b> Browser, NetBeans IDE, Glassfish Server and MySQL.
<b>Constraint:</b> None

**Table 61: Referral Program Traceability**

<b>Testing Process: UNIT TESTING (T1)</b>
<b>Test Case:</b> TC7
<b>Tested Item:</b> Referral Program (referEarn.jsp)
<b>Performed By Tester ID:</b> UT2, UT3, UT4, UT5
<b>Requirement Traceability:</b> Enter the friend's email id and click submit. It shows success message of sending email to friend's id.
<b>Testing Schedule:</b> 23 <sup>nd</sup> Feb 2017
<b>Test Recording Procedure</b> Inputs are provided and execute for expected output.
<b>Hardware/ Software Requirement:</b> Browser, NetBeans IDE, Glassfish Server and MySQL.
<b>Constraint:</b> Internet Connection

**Table 62: Notification Traceability**

<b>Testing Process: UNIT TESTING (T1)</b>
<b>Test Case:</b> TC8
<b>Tested Item:</b> Notification (userHome.jsp)
<b>Performed By Tester ID:</b> UT1, UT5
<b>Requirement Traceability:</b> Should timely fetch the update in the values.
<b>Testing Schedule:</b> 23 <sup>nd</sup> Feb 2017
<b>Test Recording Procedure</b> Execute for expected output.
<b>Hardware/ Software Requirement:</b> Browser, NetBeans IDE, Glassfish Server and MySQL.
<b>Constraint:</b> None

**Table 63: Explorer TV Traceability**

<b>Testing Process: UNIT TESTING (T1)</b>
<b>Test Case:</b> TC9
<b>Tested Item:</b> Explorer TV (userHome.jsp)
<b>Performed By Tester ID:</b> UT1, UT2, UT5
<b>Requirement Traceability:</b> Enter the video query and system generates result of that query.
<b>Testing Schedule:</b> 23 <sup>nd</sup> Feb 2017
<b>Test Recording Procedure</b> Inputs are provided and execute for expected output.

**Hardware/ Software Requirement:** Browser, NetBeans IDE, Glassfish Server and MySQL.

**Constraint:** Internet Connection

**Table 64: Settings Traceability**

<b>Testing Process:</b> UNIT TESTING (T1)
<b>Test Case:</b> TC10
<b>Tested Item:</b> Settings (userSettings.jsp)
<b>Performed By Tester ID:</b> UT1, UT2, UT5
<b>Requirement Traceability:</b> Enter the values and system generates make changes successfully or display the result of that value.
<b>Testing Schedule:</b> 23 <sup>nd</sup> Feb 2017
<b>Test Recording Procedure:</b> Inputs are provided and execute for expected output.
<b>Hardware/ Software Requirement:</b> Browser, NetBeans IDE, Glassfish Server and MySQL.

**Table 65: User Guide Traceability**

<b>Testing Process:</b> UNIT TESTING (T1)
<b>Test Case:</b> TC11
<b>Tested Item:</b> User Guide (userGuide.jsp)
<b>Performed By Tester ID:</b> UT1, UT2, UT3, UT4, UT5
<b>Requirement Traceability:</b> It should show the user manual.
<b>Testing Schedule:</b> 23 <sup>nd</sup> Feb 2017
<b>Test Recording Procedure:</b> Execute for expected output.
<b>Hardware/ Software Requirement:</b> Browser, NetBeans IDE, Glassfish Server and MySQL.

**Table 66: Visualization Traceability**

<b>Testing Process:</b> UNIT TESTING (T1)
<b>Test Case:</b> TC12
<b>Tested Item:</b> Visualization (visualization.jsp)
<b>Performed By Tester ID:</b> UT4, UT5
<b>Requirement Traceability:</b> Fetch the inputs from database and represent the graphical representation of the values.

**Testing Schedule:** 23<sup>nd</sup> Feb 2017

**Test Recording Procedure:** Execute for expected output.

**Hardware/ Software Requirement:** Browser, NetBeans IDE, Glassfish Server and MySQL.

**Constraint:** Internet Connection

**Table 67: Login and User Home Traceability**

**Testing Process: INTEGRATION TESTING (T2)**

**Test Case:** TC2 and TC3

**Tested Item:** Registration and Login

**Performed By Tester ID:** UT4, UT5

**Requirement Traceability:** Inserting login details through login module and taking user to home page.

**Testing Schedule:** 28<sup>th</sup> Feb 2017

**Test Recording Procedure:** Inputs are provided and execute for expected output.

**Hardware/ Software Requirement:** Browser, NetBeans IDE, Glassfish Server and MySQL.

**Constraint:** Internet Connection

**Table 68: Weekly Poll and Visualization Traceability**

**Testing Process: INTEGRATION TESTING (T2)**

**Test Case:** TC4 and TC12

**Tested Item:** Weekly Poll and Visualization

**Performed By Tester ID:** UT4, UT5

**Requirement Traceability:** Inserting user responses through weekly poll module and fetching responses in visualization module to create a graph of the results after analysis.

**Testing Schedule:** 2<sup>nd</sup> March 2017

**Test Recording Procedure:** Inputs are provided and execute for expected output.

**Hardware/ Software Requirement:** Browser, NetBeans IDE, Glassfish Server and MySQL.

**Constraint:** Internet Connection

**Table 69: Refer and Registration Traceability**

<b>Testing Process: INTEGRATION TESTING (T2)</b>
<b>Test Case:</b> TC7 and TC1
<b>Tested Item:</b> Refer and Registration
<b>Performed By Tester ID:</b> UT4, UT5
<b>Requirement Traceability:</b> In refer module, User 1 sends referral code through email to User 2 and the in registration module User 2 will click on the link and register with referral code.
<b>Testing Schedule:</b> 4 <sup>th</sup> March 2017
<b>Test Recording Procedure:</b> Inputs are provided and execute for expected output.
<b>Hardware/ Software Requirement:</b> Browser, NetBeans IDE, Glassfish Server and MySQL.
<b>Constraint:</b> Internet Connection

**Table 70: Daily Goal Integration Traceability**

<b>Testing Process: INTEGRATION TESTING (T2)</b>
<b>Test Case:</b> TC6 and TC3, TC4, TC5, TC7
<b>Tested Item:</b> Daily Goal and Search, Weekly poll, Survey, Referral Program.
<b>Performed By Tester ID:</b> UT4, UT5
<b>Requirement Traceability:</b> In daily goal module, the earnings from every module will be calculated for a user and display a bonus range to get the earning bonus.
<b>Testing Schedule:</b> 5 <sup>th</sup> March 2017
<b>Test Recording Procedure:</b> Inputs are provided and execute for expected output.
<b>Hardware/ Software Requirement:</b> Browser, NetBeans IDE, Glassfish Server and MySQL.
<b>Constraint:</b> Internet Connection

## 8.3 Unit Testing

### 8.3.1 Black Box Testing

Black box testing is used to test the software from user point of view, there is no internal code view. It is done on each module of the system. It includes of unit, integration, system and acceptance testing. It is performed by developers only.

### 8.3.1.1 Registration

**Table 71: Registration Unit Testing**

Test Case ID		TC1		Test Case Name	Registration	
Testing Type		Unit Testing		Tester Name	UT5	
S.no.	Input	Expected Output	Actual Output	Success/ Failure	Remarks	
1.	Email ID = 'cuteshivi02@gmail.com'  Password = 'shivani'  Confirm Password = 'shivani'	User Home page.	User Home page	Success		
2.	Email ID = 'cuteshivi02.com'  Password = 'shivani'  Confirm Password = 'shivani'	Email:  Please match the requested format.	Email: Please match the requested format.	Success	Email format.	
3.	Email ID = 'cuteshivi02@gmail.com'  Password = 'shivani12'  Confirm Password = 'shivani'	Password doesn't match.	Password doesn't match.	Success		
4.	Email ID = 'cuteshivi02@gmail.com'  Password = 'shivani12kjj'  Confirm Password = 'shivani'	Please Match the requested format.	Please Match the requested format	Success	Password length (5 to 10)	
4.	Email ID = “ ”  Password = “ ”  Confirm Password = “ ”	Please fill out the fields.	Please fill out the fields.	Success		

### 8.3.1.2 Login

**Table 72: Login Unit Testing**

Test Case ID		TC2	Test Case Name	Login	
Testing Type		Unit Testing	Tester Name	UT3, UT4, UT5	
S.no.	Input	Expected Output	Actual Output	Success/ Failure	Remarks
1.	User ID = 'cuteshivi02@gmail. com'  Password = 'shivani'	User Home page.	User Home page	Success	
2.	User ID = 'cuteshivi02@gmail. com'  Password = 'shivi'	Password incorrect	Unauthenticated User	Failure	System should show user understandable message.
3.	User ID = 'shivisinha'  Password = 'shivi123'	Unauthenticated User	Unauthenticated User	Success	
4.	User ID = ''  Password = ''	Please fill out required fields.	Please fill out the required fields.	Success	

### 8.3.1.3 Search

**Table 73: Search Unit Testing**

Test Case ID		TC3		Test Case Name	Search
Testing Type		Unit Testing		Tester Name	UT2, UT3, UT4, UT5
S.no.	Input	Expected Output	Actual Output	Success/ Failure	Remarks
1.	Search = valid query statement	Search Results	Search Results	Success	Need internet connection.
2.	Search = Invalid query statement (E.g. Searching *)	No Results	No Results	Success	
3.	Search = ‘ ’	No Results	No Results	Success	

### 8.3.1.4 Weekly Poll

**Table 74: Weekly Poll Unit Testing**

Test Case ID		TC4		Test Case Name	Weekly Poll
Testing Type		Unit Testing		Tester Name	UT3, UT4, UT5
S.no.	Input	Expected Output	Actual Output	Success/ Failure	Remarks
1.	Select = Yes / No	Thank You with result visualization.	Thank You with result visualization.	Success	
2.	Select = ‘ ’	Please select.	Please select.	Success	

### 8.3.1.5 Survey

**Table 75: Survey Unit Testing**

Test Case ID		<b>TC5</b>		Test Case Name	<b>Survey</b>	
Testing Type		<b>Unit Testing</b>		Tester Name	<b>UT2, UT4, UT5</b>	
S.no.	Input		Expected Output	Actual Output	Success/ Failure	Remarks
1.	Survey loaded.		Survey form	Survey form	Success	
2.	Survey inputs		Responses inserted successfully.	Responses inserted successfully.	Success	

### 8.3.1.6 Daily Goal

**Table 76: Daily Goal Unit Testing**

Test Case ID		<b>TC6</b>		Test Case Name	<b>Daily Goal</b>	
Testing Type		<b>Unit Testing</b>		Tester Name	<b>UT2, UT4, UT5</b>	
S.no.	Input		Expected Output	Actual Output	Success/ Failure	Remarks
1.	Fetching Earned Points		Display today's goal and bonus.	Display today's goal and bonus.	Success	
2.	Failed fetching Earned Points		Unfortunately system has occurred problem.	Blank	Failure	Database connection failure.

### 8.3.1.7 Referral Program

**Table 77: Referral Program Unit Testing**

Test Case ID		TC7		Test Case Name	Referral Program	
Testing Type		Unit Testing		Tester Name	UT2, UT3, UT4, UT5	
S.no.	Input	Expected Output	Actual Output	Success/ Failure	Remarks	
1.	Email ID = ‘suvisinha@gmail.com’  Message = “ Hello”  Code = kae4dvh8	Referral code has been sent.	Code has been sent. Wait and enjoy your earning.	Success		
2.	Email ID = ‘suvisinha@gmail.com’  Message = “ Hello”  Code = kae4dvh8	Incorrect Email ID	Please match the email pattern.	Success		
3.	Email ID = ‘ ’  Message = “ Hello”  Code = kae4dvh8	Please fill the field.	Please fill the field.	Success		

### 8.3.1.8 Notification

**Table 78: Notification Unit Testing**

Test Case ID		TC8		Test Case Name	Notification	
Testing Type		Unit Testing		Tester Name	UT2, UT3, UT4, UT5	
S.no.	Input	Expected Output	Actual Output	Success/ Failure	Remarks	
1.	Notifying updated data on the interface	Displays updated data.	Need to refresh the page.	Failure	Code not worked.	

### 8.3.1.9 Settings (Updating User data)

**Table 79: Settings Updating Data Unit Testing**

Test Case ID		TC9		Test Case Name	Settings (Updating User data)	
Testing Type		Unit Testing		Tester Name	UT2, UT3, UT4, UT5	
S.no.	Input	Expected Output	Actual Output	Success/ Failure	Remarks	
1.	Password: Valid entry Update: Entered value	Successfully updated.	Successfully updated.	Success		
2.	Password: Wrong entry	Please enter correct password.	Please enter correct password.	Success		
3.	Password: “ ”	Please fill the field.	Please fill the field.	Success		

### 8.3.1.10 Settings (View Report)

**Table 80: Settings Report View Unit Testing**

Test Case ID		TC9		Test Case Name	Settings (View Report)	
Testing Type		Unit Testing		Tester Name	UT2, UT3, UT4, UT5	
S.no.	Input	Expected Output	Actual Output	Success/ Failure	Remarks	
1.	Password: Valid entry View report: Select date of which report want to see.	Display report.	Display report.	Success		
2.	Password: Wrong entry	Please enter correct password.	Please enter correct password.	Success		

### 8.3.1.11 User Guide

**Table 81: User Guide Unit Testing**

Test Case ID		TC11		Test Case Name	User Guide	
Testing Type		Unit Testing		Tester Name	UT2, UT3, UT4, UT5	
S.no.	Input	Expected Output	Actual Output	Success/ Failure	Remarks	
1.	-	Display whole user manual.	Display whole user manual.	Success		

### 8.3.1.12 Visualization

**Table 82: Visualization Unit Testing**

Test Case ID		TC12		Test Case Name	Visualization	
Testing Type		Unit Testing		Tester Name	UT2, UT5	
S.no.	Input	Expected Output	Actual Output	Success/ Failure	Remarks	
1.	Poll responses	Graphical Representation of Analysis Result.	Graphical Representation of Analysis Result.	Success		

## 8.3.2 White Box Testing

White box testing is a set of testing procedures that evaluate the code and internal structure of a program. It is done at complex modules of the system. It is done by both developers and testers. It gives the clear view of errors like on which line of code fault has occurred. It shows the line by line execution.

### 8.3.2.1 MapReduce Analysis

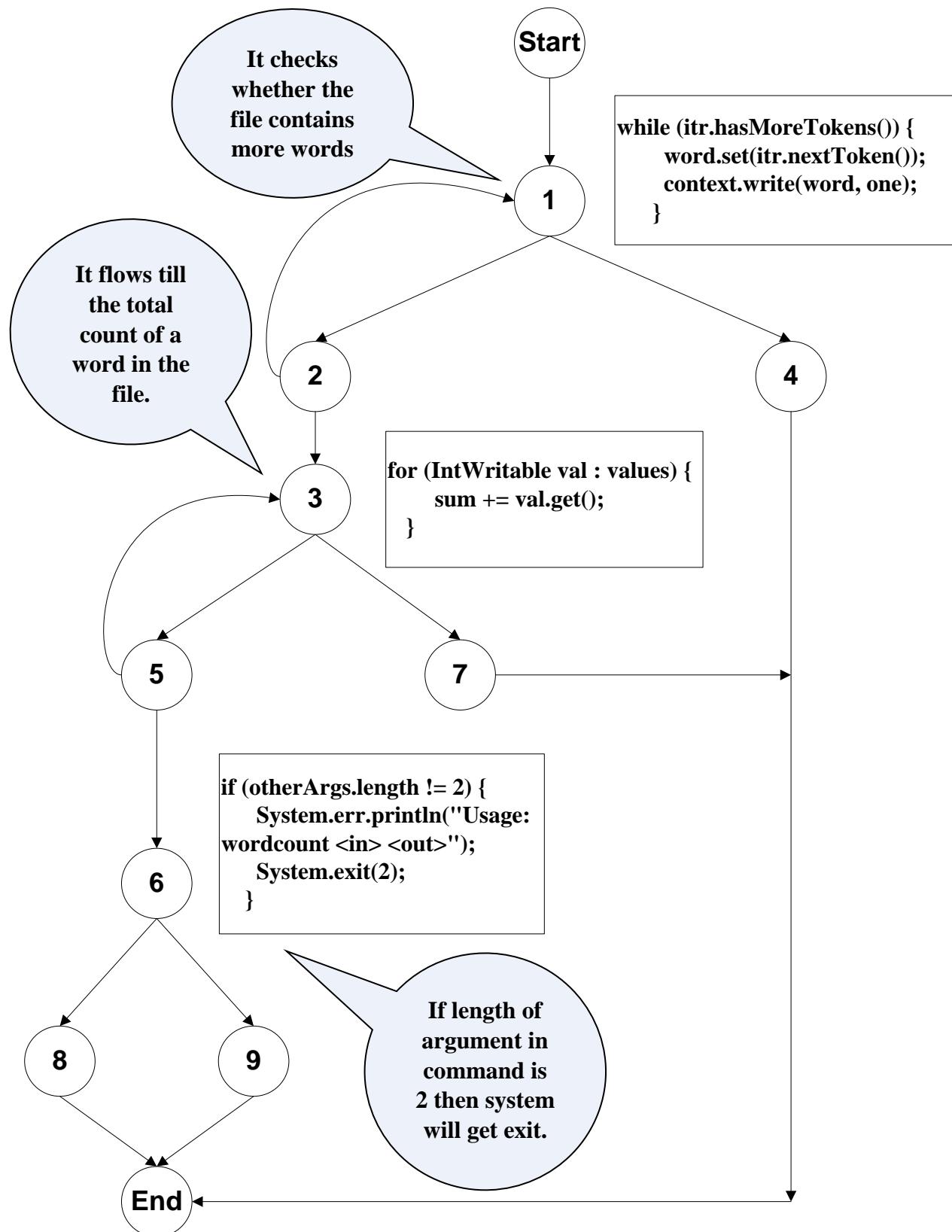


Figure 60: MapReduce Analysis White Box Testing

**Table 83: MapReduce Analysis White Box Chart:**

Test Case ID		TWB1	Test Case Name	MapReduce Analysis	
Testing Type		White Box Testing	Tester Name	UT1	
S.no.	Input	Expected Output	Actual Output	Success/ Failure	Remarks
1.	Data set is transferred to HDFS	Data set in HDFS	Data set in HDFS	Success	
2.	Run Linux Command in the Cloudera terminal to create an output directory.	Directory created in HDFS	Directory created in HDFS	Success	
3.	Run Linux Command in the Cloudera terminal to fetch the Data set.	Locate Data set.	Locates Data set.	Success	
4.	Run Linux Command in the Cloudera terminal to start MapReduce Analysis.	Start MapReduce Analysis.	Start MapReduce Analysis.	Success	
5.	Display the output	Analysis result.	Analysis result.	Success	

**Table 84: Analysis Code**

### ANALYSIS CODE

```

public class WordCount {
  public static class TokenizerMapper extends Mapper<Object, Text, Text, IntWritable>{
    private final static IntWritable one = new IntWritable(1);
  }
}
  
```

```

private Text word = new Text();
public void map(Object key, Text value, Context context) throws IOException,
InterruptedException {
  StringTokenizer itr = new StringTokenizer(value.toString());
  while (itr.hasMoreTokens()) {
    word.set(itr.nextToken());
    context.write(word, one);
  }
}

public static class IntSumReducer extends Reducer<Text,IntWritable,Text,IntWritable> {
  private IntWritable result = new IntWritable();
  public void reduce(Text key, Iterable<IntWritable> values, Context context) throws
IOException, InterruptedException {
    int sum = 0;
    for (IntWritable val : values) {
      sum += val.get();
    }
    result.set(sum);
    context.write(key, result);
  }
}

public static void main(String[] args) throws Exception {
  Configuration conf = new Configuration();
  String[] otherArgs = new GenericOptionsParser(conf, args).getRemainingArgs();
  if (otherArgs.length != 2) {
    System.err.println("Usage: wordcount <in> <out>");
    System.exit(2);
  }
  Job job = new Job(conf, "word count");
  job.setJarByClass(WordCount.class);
  job.setMapperClass(TokenizerMapper.class);
  job.setCombinerClass(IntSumReducer.class);
  job.setReducerClass(IntSumReducer.class);
  job.setOutputKeyClass(Text.class);
}
  
```

1

3

6

```

job.setOutputValueClass(IntWritable.class);
FileInputFormat.addInputPath(job, new Path(otherArgs[0]));
FileOutputFormat.setOutputPath(job, new Path(otherArgs[1]));
System.exit(job.waitForCompletion(true) ? 0 : 1);
}
}
  
```

### 8.3.2.2 Referral Program

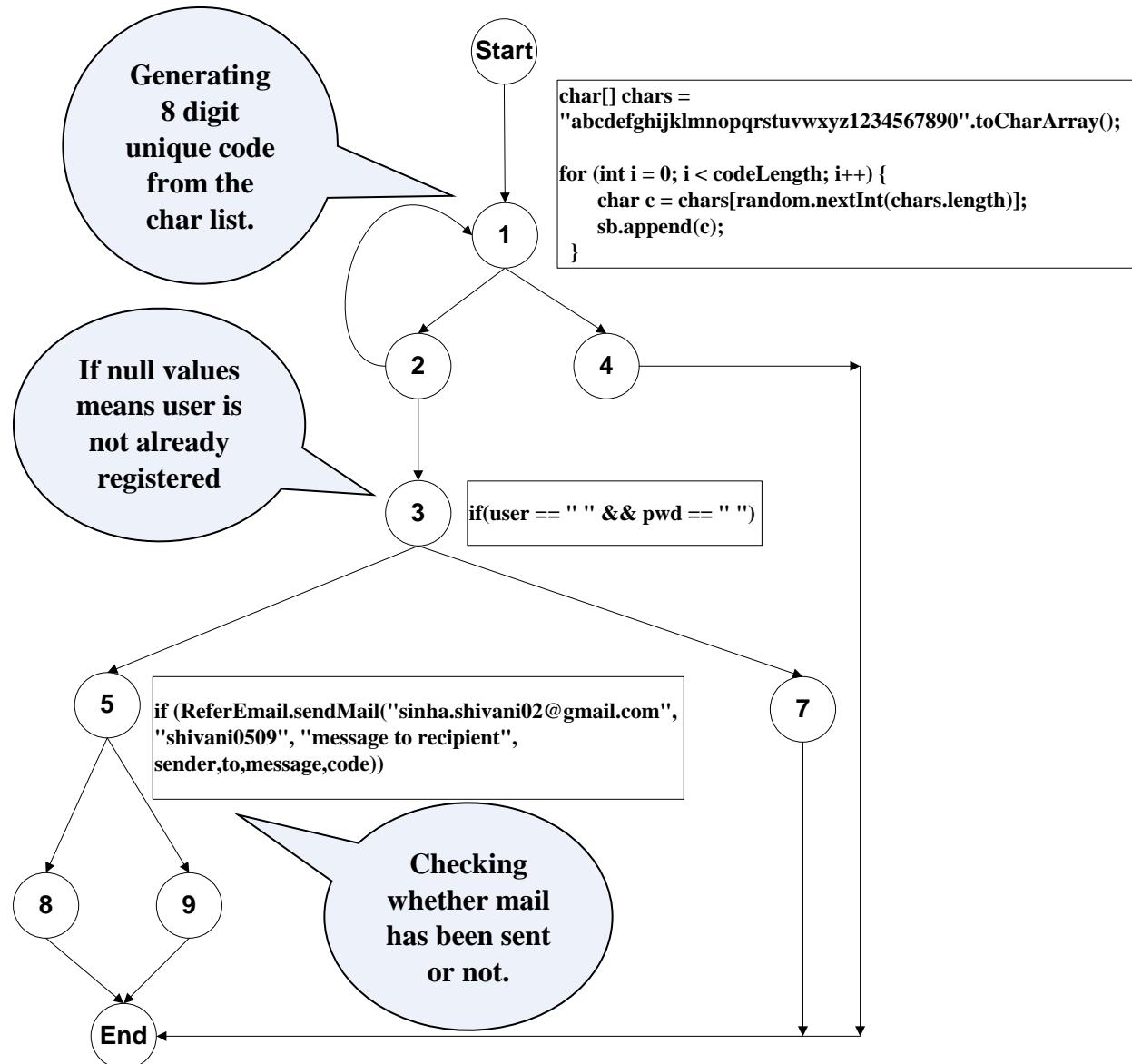


Figure 61: Referral White Box Testing:

**Table 85: Referral White Box Chart**

Test Case ID		TWB2	Test Case Name	Referral Program	
Testing Type		White Box Testing	Tester Name	UT1	
S.no.	Input	Expected Output	Actual Output	Success/ Failure	Remarks
1.	Referral Code generated	Unique Code	Unique Code	Success	
2.	Check Email id already registered or not.	If yes, User already exist.	If yes, User already exist.	Success	
3.	If not register, send code through email.	Code sent successfully	Code sent successfully	Success	

**Table 86: Referral Program Code**

### REFERRAL PROGRAM CODE

```

public class ReferCode {
    public String createRandomCode(int codeLength){
        char[] chars = "abcdefghijklmnopqrstuvwxyz1234567890".toCharArray();
        StringBuilder sb = new StringBuilder();
        Random random = new SecureRandom();
        for (int i = 0; i < codeLength; i++) {
            char c = chars[random.nextInt(chars.length)];
            sb.append(c); } } 1
        String output = sb.toString();
        return output;
    }
    if(user == " " && pwd == " ") { } 3
        if (ReferEmail.sendMail("sinha.shivani02@gmail.com", "shivani0509", "message to recipient", sender,to,message,code)) { } 5
    
```

```

{
  return "Please Check Your Email. If not received, please re-register.";
}
else
  return ("Email address doesn't exist.");
}
else
  return ("Email already registered.");

```

## 8.4 Integration Testing

### 8.4.1 Login with User Home

**Table 87: Login with user Home Integration Testing**

Test Case ID		TC1, TC2		Test Case Name	Login with User Home	
Testing Type		Integration Testing		Tester Name	UT2, UT3, UT5	
S.no.	Input	Expected Output	Actual Output	Success/ Failure	Remarks	
1.	Open login page	Login page should get opened.	Login page was opened.	Success		
2.	User ID = 'cuteshivi02@gmail.com' Password = 'shivani'	Valid User ID and Password Entered	Valid User ID and Password Entered	Success		
3.	Click on login button	User Home should get opened.	User Home was opened.	Success		

#### 8.4.2 Weekly Poll with Visualization

**Table 88: Login with user Home Integration Testing**

Test Case ID		TC4, TC12		Test Case Name	Weekly Poll with Visualization	
Testing Type		Integration Testing		Tester Name	UT2, UT3, UT5	
S.no.	Input	Expected Output	Actual Output	Success/ Failure	Remarks	
1.	Open weekly poll page	Page should get opened.	Page was opened.	Success		
2.	Answer the poll and click submit.	Thank you message and transfer to Visualization	Thank you message and transfer to Visualization	Success		
3.	Check the Graph of the poll result.	Page should contain a graph.	Page displays a graph of poll result.	Success		

#### 8.4.3 Refer with Registration

**Table 89: Refer with Registration Integration Testing**

Test Case ID		TC7, TC1		Test Case Name	Refer with Registration	
Testing Type		Integration Testing		Tester Name	UT2, UT3, UT5	
S.no.	Input	Expected Output	Actual Output	Success/ Failure	Remarks	
1.	Login	User Home Page	User Home Page	Success		
2.	Open referral page	Page should get opened.	Page was opened.	Success		

3.	Email ID = 'suvisinha@gmail.com' Message = "Hello" Code = kae4dvh8	Referral code has been sent.	Code has been sent. Wait and enjoy your earning.	Success	
4.	Registration of suvisinha@gmail.com" with code.	Successfully registered.	Successfully registered.	Success	

#### 8.4.4 Daily Goal with Search, Survey, Weekly Poll and Referral

Table 90: Daily Goal with Search, Survey, Weekly Poll and Referral Integration Testing

Test Case ID		TC6, TC3, TC4, TC5, TC7	Test Case Name	Daily Goal with Earning Modules	
Testing Type		Integration Testing	Tester Name	UT2, UT3, UT5	
S.no.	Input	Expected	Actual Output	Success/Failure	Remark
1.	Login	User Home	Page was opened.	Success	
2.	Open Daily Goal	Daily Goal	Page was opened.	Success	
3.	Fetch earned points of the user from all earning modules.	Display today's goal and bonus points.	Display today's goal and bonus points.	Success	

## 8.5 System Testing

System testing is a type of testing that is conducted on a complete or integrated modules to assess the system's compliance with its specified requirements. Earning Explorer has efficiently responded to the data processing between the system's modules on any browser.

*Table 91: System Testing*

S.no.	Input	Expected Output	Actual Output	Success/ Failure	
				Success	Remarks
1.	Run application on Google Chrome.	Application should work accurately and layout design should remain continuous.	Application has worked accurately and layout remained continuous.	Success	
2.	Run application on Mozilla Firefox.	Application should work accurately and layout design should remain continuous.	Application has worked accurately and layout remained continuous.	Success	Improvement is required.
3.	Run the web application in the Microsoft Edge.	Application should work accurately and layout design remain same.	Application has worked accurately and layout remained same.	Success	

## 8.6 User Acceptance Testing

User Acceptance Testing has been done by conducting questionnaires with the different age group peoples just after the release of final prototype. The developer has recorded and investigated the responses minutely and then came to the made a graphical representation of the results of each questions asked in the questionnaire.

### QUESTIONNAIRE

#### EARNING EXPLORER BASED ON HADOOP STREAMING

Dear Respondent,

Questionnaire ID: EEU001

I am Shivani Sinha, a final year student of BE from Asia Pacific Institute of Information Technology, wishes to get the customer satisfaction regarding Earning Explorer based on Hadoop Technology for Final Year Project. Through this brief survey on the developed system, I as the student researcher and developer, desires to collect the potential user's responses about the topic. If you are uncomfortable or you have any questions regarding the survey, you can drop a mail on [sinha.shivani02@apiit.edu.in](mailto:sinha.shivani02@apiit.edu.in). Your response will only be used for the survey purposes. The developer is very thankful to you for spending your precious time and suggestions in this questionnaire.

The questionnaire will take 10 to 15 minutes to complete your responses for the given set of questions. You can return it back in distributor's hand after comprehending. Your personal details are not going to be disclosed anywhere (internet or any website). Your private data are with us till the duration of Final Year Project gets completed.

#### SECTION A

The sections is asking for your personal details. It will assist us in checking who are the respondent and how much the received data are accurate.

Gender	<input type="radio"/> Male	<input type="radio"/> Female
Age		

<b>Occupation</b>					
<b>Email</b>					

## SECTION B

The section is comprised of objective type questions that allows you to answer the questions from the given set of options. Please select the most suitable option for your answer.

1: Strongly Disagree    2: Disagree    3: Neutral    4: Agree    5: Strongly Agree

Q1. Are you satisfied with the details asked in registration process of the system?

1  2  3  4  5

**Justification:** To check how many users are agreeing that sufficient details have been asked in the process

Q2. Did you receive an understandable email content when you register yourself?

1  2  3  4  5

**Justification:** Whether the user can identify the email sent by the system or the contents are reasonable.

Q3. Are you able to comprehend the format of survey form?

1  2  3  4  5

**Justification:** Whether users are comfortable with the survey form that they can get the meaning of the questions of the survey.

Q4. Are you satisfied with the earning points provided from the survey?

1  2  3  4  5

**Justification:** Whether the user can satisfied with amount of points awarded or want more so that information can be discussed with the survey taker.

Q5. Are you comfortable with the referral earnings?

1  2  3  4  5

**Justification:** Whether the user can satisfied with amount of points awarded by referring the system to their friends.

Q6. Do you feel that you receive updated search results?

1  2  3  4  5

**Justification:** If maximum disagreeing then there should be some indication or feedback about the updated search on the result pages of the search engine.

Q7. How much you are satisfied with the format of the result shown in the system?

1  2  3  4  5

**Justification:** To check whether user able to see distinct results clearly or not.

Q8. Are you satisfied with the graphical representation of weekly poll result?

1  2  3  4  5

**Justification:** Whether the graphs are helpful in demonstrating the response result analysis for the users.

Q8. Are you satisfied with the Daily Goal provided by the system?

1  2  3  4  5

**Justification:** Whether the users are satisfied with the bonus points provided by the system.

## SECTION C

The section is comprised of subjective type questions. Questions are related to your personal experience or feelings.

Q.9. Would you like to advise something for any kind of system improvement?

---

---

---

**Justification:** It will help in serving the customized system according to recommended choice.

Thank you for participating in the survey and providing valuable answers. Please check whether you have answered every questions as per your choice.

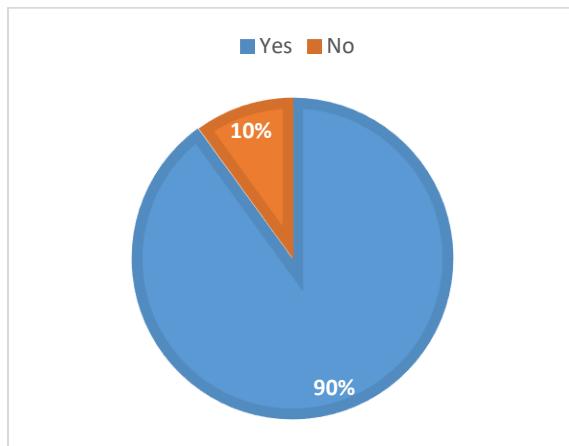
### 8.6.1 Questionnaire Analysis

The questionnaire has been attended maximum by the students of age group 12 to 24 which results in an accurate data as they are the supreme user of internet and search engine. After doing analysis, the researcher has received a positive response from the user. The developer has found that everything is correct and accordance with the user's review which was taken in prior during analysis phase. The proposed system have covered every feature in the system according to the user requirements and user satisfaction.

**Table 92: Analysis Summary Chart**

Summary Chart	
Question 1	8/10
Question 2	8/10
Question 3	8/10
Question 4	7/10
Question 5	8/10
Question 6	9/10
Question 7	7/10
Question 8	7/10

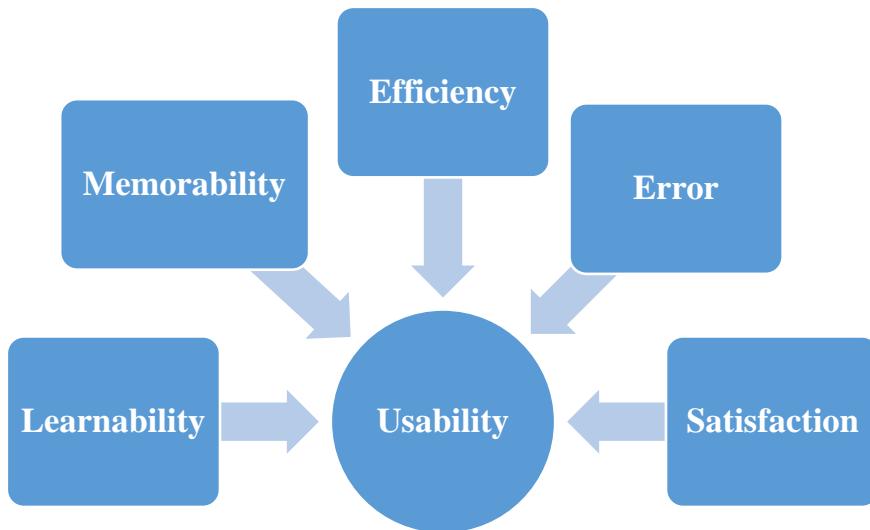
Hence, we can see the overall conclusion of user acceptance testing in the pie chart mentioned below. It shows that almost 90% of users have appreciated the system effectiveness and efficient.



**Figure 62: User Acceptance graph**

## 8.7 Usability Testing

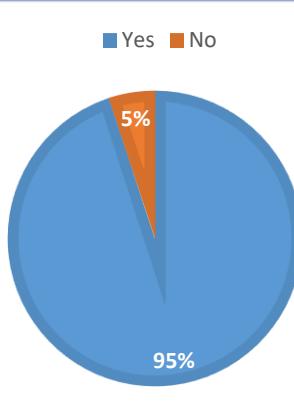
The system usability testing has been done by checking for each usability paradigms. The developer has recorded the user feedback after working on the system in terms of Learnability, Memorability, Efficiency, Error and Satisfaction.



**Figure 63: Usability Features:**

**Table 93: Learnability**

LEARNABILITY				
Are you able to identify the controls available on the interface?				
User	Response	Target	Current	Unacceptable
<b>User 1 (UT2)</b>	Yes	1 min	10 sec	> 2 min
<b>User 2 (UT4)</b>	Yes	1 min	15 sec	> 2 min
<b>User 3 (UT5)</b>	Yes	1 min	5 sec	> 2 min



A pie chart titled 'LEARNABILITY' showing the distribution of responses. The chart is divided into two segments: a large blue segment representing 'Yes' at 95% and a small orange segment representing 'No' at 5%. A legend at the top indicates that blue represents 'Yes' and orange represents 'No'.

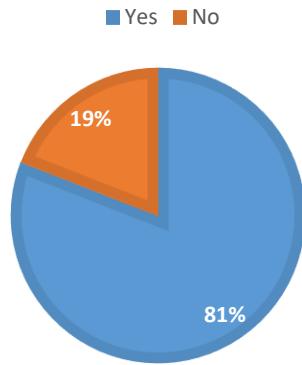
Response	Percentage
Yes	95%
No	5%

**Justification:** To identify whether the system has included the user friendly control or not by the user time taken in using those controls.

**Conclusion:** The developer has concluded that the most of the users were comfortable with the controls but few are still not able to perform in time as they are slow reader.

**Table 94: Memorability**

MEMORABILITY				
Are you able to remember the flow of the system navigation?				
User	Response	Target	Current	Unacceptable
<b>User 1 (UT2)</b>	Yes	1 day	Half day	> 2 day
<b>User 2 (UT4)</b>	No	1 day	1 day	> 2 day
<b>User 3 (UT5)</b>	Yes	1 day	On first interaction	> 2 day

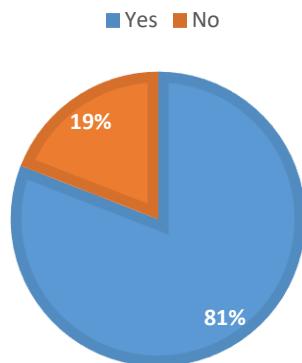


**Justification:** To find out whether the users are able remember the links and interface navigations on the first interaction with the system or not.

**Conclusion:** The developer has concluded that the most of the users were comfortable with the system memorability.

Table 95: Efficiency

EFFICIENCY				
Are you able to register your details efficiently in the system?				
User	Response	Target	Current	Unacceptable
User 1 (UT2)	Yes	1 min	30 sec	> 2 min
User 2 (UT4)	Yes	1 min	40 sec	> 2 min
User 3 (UT5)	No	1 min	2 min 20 sec	> 2 min



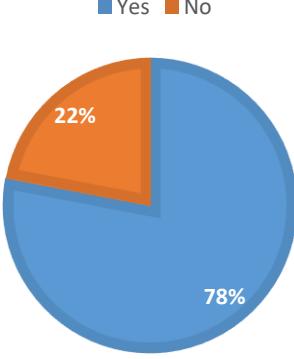
Response	Percentage
Yes	81%
No	19%

**Justification:** To find out the system efficiency. How many users are able to register details proficiently.

**Conclusion:** The developer has concuded that the most of the users were comfortable at the time of registration and have done efficiently. Some of the users have faced probem because of deploymet of the project on the first run of the software.

**Table 96: Error**

ERROR				
Have you faced any problem during system feedback?				
User	Response	Target	Current	Unacceptable
User 1 (UT2)	No	1 min	30 sec	> 2 min
User 2 (UT4)	No	1 min	40 sec	> 2 min
User 3 (UT5)	No	1 min	50 sec	> 2 min



■ Yes ■ No

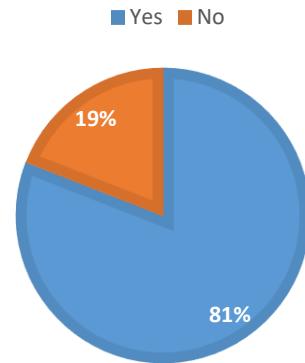
Response	Percentage
Yes	78%
No	22%

**Justification:** To get the users who have face dificulty in executing any functionality that means they will take more than unacceptable timing.

**Conclusion:** The developer has concuded that the system is capable in even flow from one function to other. But notification is not working.

**Table 97: Satisfaction**

SATISFACTION	
Are you satisfied with overall system performance?	
According to above all usability testing we analyze that our system have learnability, able to memorize the things means memorability, more efficiency and minimum error then we can say that our system given more satisfaction to the users.	



**Justification:** To find out the user satisfaction rate on system usability.

**Conclusion:** The developer has concluded that all this usability testing we can say that this system are given overall satisfaction to all user who already user of the system, new user and intermittent user.

## 8.8 Compatibility Testing

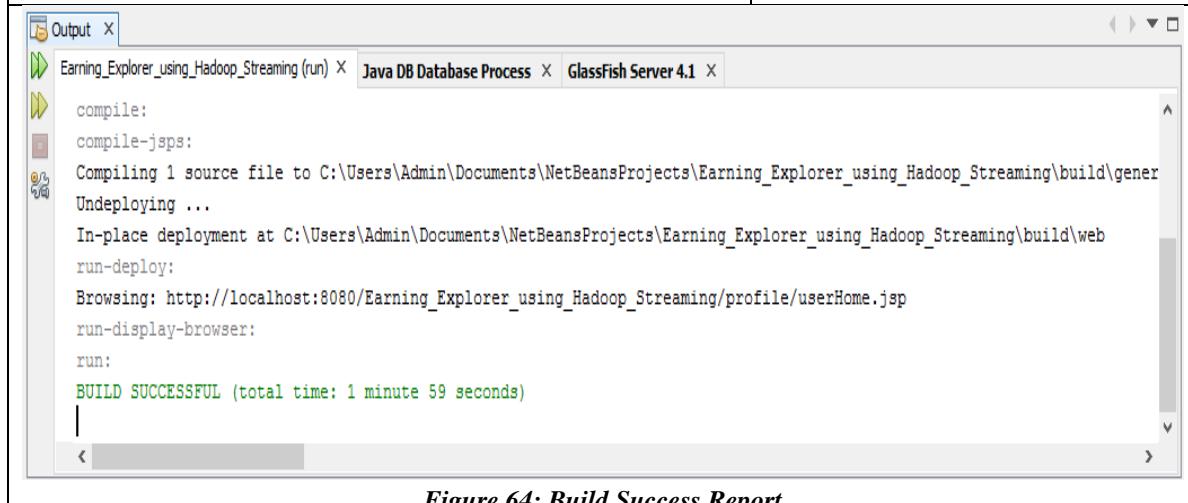
Compatibility testing was conducted on the application to evaluate the application's compatibility and reliability on different platforms. The system's compatibility testing has been performed by the developer of the system and the users.

Table 98: System Compatibility Testing

<b>Project Title</b>	<b>Earning Explorer</b>	<b>Testing Date</b>	22 <sup>st</sup> March 2017
<b>Test Case Name</b>	<b>Compatibility Testing</b>	<b>Test Case ID</b>	T6
<b>Conducted By</b>	<b>Parag Dhingra</b>	<b>Tester ID</b>	UT6
<b>Description</b>	It will check whether the system can run or responding properly on other operating system or not.		
<b>Test Performed On</b>		<b>Result</b>	
Windows 7, NetBeans IDE 8.1, MySQL Workbench		Failed (Software compatibility)	
Windows 8.1, NetBeans IDE 8.0.2, MySQL Workbench		Success	
Google Chrome		Success	
Mozilla Firefox		Success	
Microsoft Edge		Success	

**Table 99: System Compatibility Testing**

<b>Project Title</b>	<b>Earning Explorer</b>	<b>Testing Date</b>	21 <sup>st</sup> March 2017
<b>Test Case Name</b>	<b>Compatibility Testing</b>	<b>Test Case ID</b>	T6
<b>Conducted By</b>	<b>Shivani Sinha</b>	<b>Tester ID</b>	UT1
<b>Description</b>	It will check whether the system can run or responding properly on other operating system or not.		
<b>Test Performed On</b>	<b>Result</b>		
Windows 8, NetBeans IDE 8.1, MySQL Workbench	Success		
Windows 10, NetBeans IDE 8.0.2, MySQL Workbench	Success		
Google Chrome	Success		
Mozilla Firefox	Success		
Microsoft Edge	Success		



The screenshot shows the NetBeans IDE's Output window. It displays the build process for the project 'Earning\_Explorer\_using\_Hadoop\_Streaming'. The output log shows the following steps:

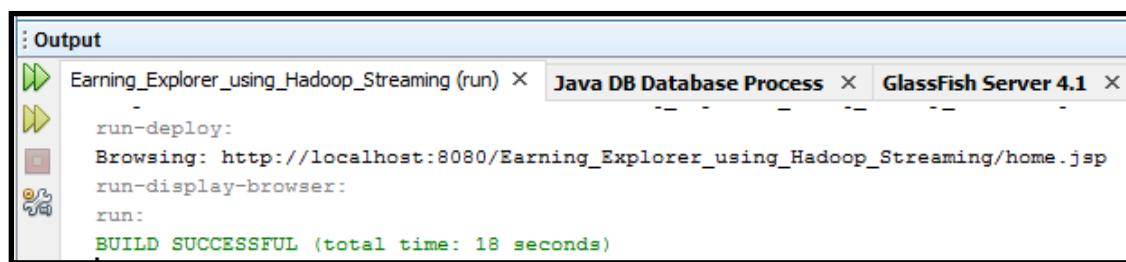
```

Output X
Earning_Explorer_using_Hadoop_Streaming (run) X Java DB Database Process X GlassFish Server 4.1 X

compile:
compile-jsp:
Compiling 1 source file to C:\Users\Admin\Documents\NetBeansProjects\Earning_Explorer_using_Hadoop_Streaming\build\gener
Undeploying ...
In-place deployment at C:\Users\Admin\Documents\NetBeansProjects\Earning_Explorer_using_Hadoop_Streaming\build\web
run-deploy:
Browsing: http://localhost:8080/Earning_Explorer_using_Hadoop_Streaming/profile/userHome.jsp
run-display-browser:
run:
BUILD SUCCESSFUL (total time: 1 minute 59 seconds)
  
```

**Figure 64: Build Success Report**

## 8.9 Runtime Testing



The screenshot shows the NetBeans IDE's Output window. It displays the runtime testing process for the project 'Earning\_Explorer\_using\_Hadoop\_Streaming'. The output log shows the following steps:

```

Output
Earning_Explorer_using_Hadoop_Streaming (run) X Java DB Database Process X GlassFish Server 4.1 X

run-deploy:
Browsing: http://localhost:8080/Earning_Explorer_using_Hadoop_Streaming/home.jsp
run-display-browser:
run:
BUILD SUCCESSFUL (total time: 18 seconds)
  
```

**Figure 65: Run tine testing**

## CHAPTER 9

# CRITICAL EVALUATION

### 9.1 Problem Solved for User Benefits

The overall conclusion drawn from the system's development lies in two sections of wide internet usage. Search and earn are two basic tasks that have been done to be good at wealth. The developer has worked on these problem areas of the proposed system and provided a combined solution for the users.

- 9.1.1 Search Engine:** The system has developed a built-in search engine through which any user can find solutions to their queries. The system is more efficient to provide the updated results in less time.
- 9.1.2 Earning:** Well, the source of income is a big issue for every person in this era. The system has successfully designed a platform with a proper and required equation of earning money. It is planned to provide most of the tasks to earn money in the form of points.
- 9.1.3 Profit:** The tasks have been intelligently selected so that the backend service provided can get most of it. For example, survey providers are getting a plenty of data regarding their subject of interest for their business advantage.
- 9.1.4 Time Saving:** Not on very big scale but still the system can save the user's time because it retains the feature of parallel tasking and hence can earn money as well as knowledge from the same platform at the same time.

### 9.2 Success Degree

- 1. Requirement:** To meet user requirements
- 2. Research:** Researches regarding the subject to obtain an accountable information.
- 3. Analysis:** Analysis of the research.
- 4. Functionality:** Functionalities implementation according to the requirements.
- 5. Usability:** Usability testing to check the system efficiency.
- 6. Management:** Project management to handle the project.
- 7. Documentation:** Proper documentation to be done.

**Table 100: Success Degree Evaluation**

SUCCESS DEGREE EVALUATION	
DEGREE	EVALUATION
<b>1. Requirement</b>	The project Earning Explorer requirement specification has been planned and implemented by the developer as mentioned in Project Specification Form (PSF).
<b>2. Research</b>	The developer has researched deeply about the topic and have successfully implemented the project by following the researched knowledge.
<b>3. Analysis</b>	The analysis of the secondary research has been done in order to develop the system according to the market reviews, technical analysis and user acceptance.
<b>4. Functionality</b>	The developer has successfully implemented the functionality with suitable efficiency rate. The functionalities have also been integrated well.
<b>5. Usability</b>	The developer has tested the usability of the system by recording the user interaction with the system. The result of the testing have proved that the system carries a good usability rate.
<b>6. Management</b>	The developer has managed all the phase of Agile methodology for the development of the project so well that it has successfully delivered.
<b>7. Documentation</b>	A descriptive and informative documentation of all the phases of the project development has been done by the developer.

## CHAPTER 10

# CONCLUSION

### 10.1 Limitations of the system

The Earning Explorer also deals with certain limitation in its application. These are as follows.

**10.1.1 Interface:** The proposed system is web based application only.

**10.1.2 Internet:** User will be able to run the system with internet connection only.

**10.1.3 Search Term:** Users are required to know how to write the query to extract the appropriate search results that is the way of writing search query.

**10.1.4 Email Account:** Email account is required for registration.

### 10.2 Future Enhancement

The ‘Earning Explorer’ also deals with certain limitation in its application. These are as follows.

**10.2.1 Mobile Application:** The system would be developed for the smart phone users too for their ease of accessing the system.

**10.2.2 Survey Provider Account:** Survey provider can directly upload their files or surveys by themselves. They are not needed to be dependent on third party.

**10.2.3 Online Shopping:** The system would provide the earning feature with shopping sites like coupon codes.

### 10.3 Computational Challenges

There are various leading challenges in the development of the system:

**10.3.1** Understanding the HDFS (Hadoop Distributed File System) cluster.

**10.3.2** Implementing the mail services in the system.

**10.3.3** Applied algorithms should produce precise and updated output of user’s search and earning calculations.

**10.3.4** Preforming prediction using predictive analysis.

**10.3.5** Visualization of the analyzed data.

**10.3.6** The system is required to analyze the user search behavior and the related data mining should be done.

**10.3.7** Bootstrap and JSP to make a user friendly web application.

#### 10.4 Learning Experience

As a researcher, analyzer and developer of the project '**Earning Explorer based on Hadoop Streaming**', it was a great challenge but equally have developed the skills of finding out the correct and actual data for one project.

The developer was required to form fully functional application in order to compute the plus point for a particular user account throughout its lifecycle. Aiming to accomplish the mentioned tasks, the developer had studied the Hadoop technology thoroughly and became able to work proficiently under Hadoop environment.

The developer wants to mention some notes regarding the learnings gained from doing the Final Year Project of her Bachelor Degree.

1. While studying the Hadoop technology, the developer has learnt the whole architecture of Hadoop cluster and how it maintains the integrity of jobs performed inside the **Hadoop Distribute File System (HDFS)**.
2. Have learnt the whole architecture of HDFS from task assigned to the nodes of the cluster to the delivery of the task solution and also the storage of those tasks performance.
3. Have gained the knowledge related to the communication between the Local file system and the Hadoop file system.
4. Have recalled all the principles related to **System Analysis and Development (SAD)**, **Human Computer Interaction (HCIU)** and **Project Management (PM)**.
5. Have understood the concept of Market Research and Analysis.

#### 10.5 Development Experience

During the development of the project the developer has faced many worthless experience but that was fully fruitful for the project development as "**good experience comes from bad experiences**".

The project planning of the developer has played a very important role in the project progress report. The developer has done its all tasks according to the ‘Software Engineering Life Cycle’ and made most of it. The best experiences from the development phase for the developer are following the **guidelines** Final Year Project and getting the knowledge of practical implementation of **Hadoop** environment to see the real world data management.

The developer has improved her capabilities in developing a web application using **Bootstrap** and **Java Server Pages (JSP)**, implementing the **Mail** services and most important to **transfer the database file** from Local file system to Hadoop System. Other than these technical skills the developer has productively applied the modules of **HCIU, SAD and PM** guidelines.

Overall, the developer has successfully increased her level of developer skills and now confident to develop any web based project within given time. Can effectively identify the business requirement and result in an advantage from it.

## 10.6 Personal Experience

As a researcher, I have come to know about the new era of online money making facilities and experience how the ideas are growing fast in internet technologies. By doing survey regarding the topic of online earning, it is very clear that less users are aware about the ongoing procedures but they want to join these services.

This **Final Year Project** was a full package of experience of working in real time project. It is very useful and very correct way for becoming a real time developer. The development of the project leads to an innovative personality under proper guidance of **Mrs. Gurpreet kaur Ma'am** and **Mr. Ravi Sachdeva Sir**. The proper throughout supervision of them has made me confident in completing the project.

## REFERENCE

### Website

- Base36.com. (2016). *Agile & Waterfall Methodologies – A Side-By-Side Comparison* / Base36. [online] Available at: <http://www.base36.com/2012/12/agile-waterfall-methodologies-a-side-by-side-comparison/> [Accessed 17 Oct. 2016].
- Bappalige, S. (2017). *An introduction to Apache Hadoop for big data*. [online] Opensource.com. Available at: <https://opensource.com/life/14/8/intro-apache-hadoop-big-data> [Accessed 16 Jan. 2017].
- bigdatanerd. (2016). *Hadoop vs RDBMS. Where Hadoop scores over RDBMS*. [online] Available at: <https://bigdatanerd.wordpress.com/2012/02/12/hadoop-vs-rdbms-where-hadoop-scores-over-rdbms/> [Accessed 18 Oct. 2016].
- Db-engines.com. (2016). *Elasticsearch vs. Oracle vs. Solr Comparison*. [online] Available at: <http://db-engines.com/en/system/Elasticsearch%3BOracle%3BSolr> [Accessed 17 Nov. 2016].
- Esomar. (2016). *Market research explained - ESOMAR*. [online] Available at: <https://www.esomar.org/knowledge-and-standards/market-research-explained.php> [Accessed 22 Oct. 2016].
- Free Code Camp. (2016). *Learn to code and help nonprofits*. [online] Available at: <https://www.freecodecamp.com/challenges/use-responsive-design-with-bootstrap-fluid-containers> [Accessed 26 Dec. 2016].
- Krishna, V. (2017). *Agile and Other Development Methods - CodeProject*. [online] Codeproject.com. Available at: <https://www.codeproject.com/articles/25278/agile-and-other-development-methods> [Accessed 17 Oct. 2016].
- Lewis, S., Csordas, A., Killcoyne, S., Hermjakob, H., Hoopmann, M., Moritz, R., Deutsch, E. and Boyle, J. (2017). *Hydra: a scalable proteomic search engine which utilizes the Hadoop distributed computing framework*. [online] Available at: <https://www.ncbi.nlm.nih.gov/pubmed/23216909> [Accessed 12 Jan. 2017].

- Mangal, A. (2017). *SendMailUsingJavaFromNetBeans* - NetBeans Wiki. [online] Available at: <http://wiki.netbeans.org/SendMailUsingJavaFromNetBeans> [Accessed 10 Jan. 2017].
- Mapr.com. (2016). *Apache Hadoop / Big Data Software - MapR*. [online] Available at: <https://www.mapr.com/products/apache-hadoop> [Accessed 16 Nov. 2016].
- Predictive Analytics Today. (2016). *What is Predictive Analytics ? - Predictive Analytics Today*. [online] Available at: <http://www.predictiveanalyticstoday.com/what-is-predictive-analytics/> [Accessed 13 Nov. 2016].
- Ronk, J. (2016). *Structured, semi structured and unstructured data*. [online] Jeremy Ronk. Available at: <https://jeremyronk.wordpress.com/2014/09/01/structured-semi-structured-and-unstructured-data/> [Accessed 16 Nov. 2016].
- Softwaretestingclass.com. (2017). *System Testing: What? Why? & How? / Software Testing Class*. [online] Available at: <http://www.softwaretestingclass.com/system-testing-what-why-how/> [Accessed 7 Mar. 2017].
- Sequoia Blogs. (2017). *Hadoop : Lucene Indexing with MapReduce*. [online] Available at: <https://blogs.sequoiainc.com/hadoop-lucene-indexing-with-mapreduce/> [Accessed 18 Jan. 2017].
- Taylor, D. (2016). *The Literature Review: A Few Tips On Conducting It*. [online] Writing.utoronto.ca. Available at: <http://www.writing.utoronto.ca/advice/specific-types-of-writing/literature-review> [Accessed 20 Oct. 2016].
- 1000hz.github.io. (2017). *Validator, for Bootstrap 3.* [online] Available at: <https://1000hz.github.io/bootstrap-validator/> [Accessed 6 Jan. 2017].
- Wired.com. (2017). *OPEN SOURCES BUILD ‘GOOGLE SEARCH FOR BIG DATA’*. [online] Available at: <https://www.wired.com/2013/06/cloudera-search/> [Accessed 22 Jan. 2017].
- www.tutorialspoint.com. (2017). *Hadoop Tutorial*. [online] Available at: <http://www.tutorialspoint.com/hadoop/> [Accessed 20 Jan. 2017].

www.tutorialspoint.com. (2016). *JSP Tutorial.* [online] Available at:  
<http://www.tutorialspoint.com/jsp/> [Accessed 29 Dec. 2016].

W3schools.com. (2016). *Bootstrap 3 Tutorial.* [online] Available at:  
<http://www.w3schools.com/bootstrap/> [Accessed 28 Dec. 2016].

## Journal

Alshamrani, A. and Bahattab, A. (2015). A Comparison Between Three SDLC Models Waterfall Model, Spiral Model, and Incremental/Iterative Model. *IJCSI International Journal of Computer Science Issues*, 12(1694-0784), pp.106-111[Accessed 18 Oct. 2016].

Bhosale, H. and Gadekar, P. (2014). A Review Paper on Big Data and Hadoop. *International Journal of Scientific and Research Publications*, 4(2250-3153) [Accessed 20 Oct. 2016].

Cs.cmu.edu, (2015). *A Brief History of Human Computer Interaction Technology.* [online] Available at: <http://www.cs.cmu.edu/~amulet/papers/uistory.tr.html> [Accessed 24 Oct. 2016].

Dean, J. and Ghemawat, S. (2004). MapReduce: Simplified Data Processing on Large Clusters. *Google, Inc..* [online] Available at: <https://static.googleusercontent.com/media/research.google.com/en//archive/mapreduce-osdi04.pdf> [Accessed 19 Oct. 2016].

Dean, J. and Ghemawat, S. (2008). MapReduce. *Communications of the ACM*, 51(1), p.107 [Accessed 7 Feb. 2017].

Hadoop and HDFS: Storage for next generation data management. (2016). 105 [Accessed 29 Dec. 2016].

Jain, A. and Kaushik, M. (2014). Performance Optimization in Big Data Predictive Analytics. *International Journal of Advanced Research in Computer Science and Software Engineering*, [online] 4(2277 128X), pp.126-129. Available at: [http://www.ijarcsse.com/docs/papers/Volume\\_4/8\\_August2014/V4I8-0247.pdf](http://www.ijarcsse.com/docs/papers/Volume_4/8_August2014/V4I8-0247.pdf) [Accessed 10 Feb. 2017].

Phadatare, G., Rathod, K., Shukla, V. and Salunkhe, M. (2014). GooDoop: Local Search Engine using Hadoop. *International Journal of Computer Science and Information Technologies*, 5 (3)(0975-9646), pp.3508-3510 [Accessed 24 Jan. 2017].

Sharma, S., Sarkar, D. and Gupta, D. (2012). Agile Processes and Methodologies: A Conceptual Study. *International Journal on Computer Science and Engineering (IJCSE)*, [online] 4(0975-3397), pp.892-898. Available at: <http://www.enggjournals.com/ijcse/doc/IJCSE12-04-05-186.pdf> [Accessed 16 Oct. 2016].

Webcache.googleusercontent.com, (2015). *Human-Computer Interaction: An Overview*. [online] Available at: <http://webcache.googleusercontent.com/search?q=cache:3YFbMeyi814J:www.s2is.org/Issues/v1/n1/papers/paper9.pdf+&cd=2&hl=en&ct=clnk&gl=in> [Accessed 24 Oct. 2016].

## Books

Basham, B., Sierra, K. and Bates, B. (2008). *Head First Servlets and JSP*. 2nd ed. O'Reilly Media.

Burke, R 2009, Project Management : Planning and Control Techniques, Wiley India Pvt Ltd., INDIA.

Faulkner, X. (2000). Usability engineering. 1st ed. Hounds mills, Basingstoke, Hampshire: Palgrave.

Nielsen, J. (1993). Usability engineering. 1st ed. Boston: Academic Press.

Pujari, A. (2001). *Data mining techniques*. Hyderabad: Universities Press.

White, T. (2012). *Hadoop: The Definitive Guide, 3rd Edition*. 3rd ed. O'Reilly Media / Yahoo Press.

Zacarias, M. and Oliveira, J. (2012). Human-computer interaction. 1st ed. Berlin: Springer.

# APPENDIX





# PROJECT PROPOSAL FORM

## **Supervisor**

Mrs. Gurpreet Kaur

## **Advisor**

Mr. Ravi Sachdeva

## **Project Title: "Earning Explorer based on Hadoop Streaming"**

### **Please describe the type of the system being developed.**

Today we all are familiar with the word 'Search Engine'. We are searching anything starting from very small to very vast things on it. We understand the concept of the search engine. But, Earning Explorer is something that is not bounded to search results only, it is about to award users for their search and performed activities. This system is an attempt to encourage people to make money even for their searches too. Now a days each one of us wants to earn money and even most of us tries to find several online platforms to get paid. Hence, although they were exploring things with existing system also, but the proposed system will let them feel more interested in exploring the world as they are being paid for it.

End users will be able to earn points (called Plus Points or P+) by performing a variety of online activities, including searching, taking surveys, and discovering online content, and then to redeem those points for retail gift cards, PayPal cash, or other rewards.

This web application will allow them to execute the tasks in the given flow.

- Create account.
- After creating account successfully, user can easily go through different doings that are available on the interface.
- Using search bar, user can walk around whatsoever they want.
- See the surveys available and perform several surveys to get Plus Points.
- Answer the daily asked questions in order to get respective P+. Perceive the percentage of user response for a particular question.
- Check the notifications for performing new activities to earn.
- User can effortlessly find many stuffs added in the system by exploring it.

### **Who is system being developed for?**

The proposed system is open for everyone whether they are student or teachers, employed or unemployed, old age or children. Anyone who are willing to receive currency and to connect with the social world. This system is capable in attracting people to join with it and work on it efficiently

to make advantage of them. The system is mostly beneficial for those user groups who are daily linked with the internet. Coming to different type of user in context of the proposed system. There are three type of users are as follows.

- **Expert Users:** They come under expert user who can easily adapt the system.
- **Intermittent Users:** They can also use it as this system contains remarking features whose pattern they can easily memorize.
- **Novice User:** The working process of the system is so easy and similar to existing search engine that they can flexibly get comfortable with it.

### Why is such a system required?

In comparison with the existing system, the suggested system is same as the current scheme in term of searching facility. Nevertheless, this system is providing some kind of structure that involves user for their advantage in terms of wealth. With existing system users are receiving knowledge only but the future system will be developed with a thinking that its end users should get knowledge plus capital.

It is an effort to attract everyone and make them aware about the happenings of the world. User will come to know that they can even get paid for the searches they are performing on the search engines. It is like user will obviously feel by their own to search deeply about several terms and stay connected with the system as they are feeling employed.

### What will be the main challenge for you in building the system?

The main challenge of the system is to generate search results and plus point calculation with a high accuracy. The developer needs to keenly observe the algorithm construction and application throughout the system process. The system should be developed with the noteworthy quality procedures and user friendly.

There are various leading challenges in the development of the system:

- Understanding the HDFS (Hadoop Distributed File System) cluster.

- Data collection, mining and handling. The structured data needs to travel into the system and analyzed on various stages of the system.
- To have a check on the applied algorithm to produce precise and updated output of user's search.
- Preforming prediction using predictive analysis.
- Visualization of the predicted data by integrating Hadoop with Power BI.

### **What new ideas and theory will you need to learn to build the system?**

The basic thing is to learn how to develop a search engine using Hadoop streaming and innovative is to award money in terms of points to the system users. The developer is required to form new algorithm in order to compute the plus point for a particular user account. For the accomplishment of the mentioned tasks, the developer has to study Hadoop technology and should be able to work proficiently under Hadoop.

As well the developer must have strong knowledge of the technologies mentioned below:

- Machine learning and text mining to identify the most relevant information and its pattern from the large data set to evaluate and interpret the output.
- Distributed computing to make solve a problem or task in very small time interval.
- Data migration to transfer data within the system in order to perform different processes.
- Predictive analysis is to analyze the collected facts and generate a prediction over it or we can say user acceptance regarding any topic.
- Visualization to display the generated prediction in the form of graphs to make user more clear about the ratio.



# PROJECT SPECIFICATION FORM

## **Supervisor**

Mrs. Gurpreet Kaur

## **Advisor**

Mr. Ravi Sachdeva

## SECTION A

**Project Title:** Earning Explorer based on Hadoop Streaming

## SECTION B

**Brief description on project background.**

- **Problem Context**

Today almost everyone is familiar with the word 'Search Engine' and practically using it also. We are searching anything starting from very small to very vast things on it. We understand the concept of the search engine but users are complaining that they are not sure whether the existing search engine is providing updated search results or not. They want to the results should be appropriate according to the requirement.

As we got to know about the issues related to the existing search engines, in the same manner the employment is a need for every individual in the world. The craze of earning online is also well known and everyone is in already in the race of online earnings. We all are working hard for our expense and savings. People are searching for different jobs by using search engines. Hence, the developer has thought to develop an earning platform by combining both the topics through which anyone can get an opportunity of getting money. The project will act as a facilitator for the thought of search and earn together.

- **Rationale**

After problem identification, the system comes to its implementation area. It comprises various functionalities based on user interaction and their benefits. The benefits that the users are going to make use of are highlighted below. The benefits has been categorized into two parts.

### Tangible Benefits

The most important tangible benefit is to get money by using the system and user can get it by performing different activities that are mentioned below.

- **Searching:** The users are getting appropriate search results regarding their searched queries and paid for those searches with a definite amount of points.

- **Referral:** Users can get money on referring the application to their friends after a period of time.
- **Opinion poll:** Answering the question asked on weekly basis and earn one point on every answer.
- **Survey:** Comfortable users can attend available surveys in the system and after completing it successfully they will get plus point in their registered account.
- **Video:** The system is providing money to the users without any extra efforts. Hence, users are having a chance of earning simply by watching videos.

### Intangible Benefits

Users will be able to use the system very smoothly and can easily recognize the elements and working process of the system. The certain intangible benefits have been listed below.

- **User friendly interface:** Users are provided with the full functional and friendly system. They can easily locate the things presented in the system.
- **Trustworthy:** The system has been managed in the way so that users can trust and feel free to share anything with the system.
- **User satisfaction:** Users will feel satisfied with their hopes. The system will meet the user's expectation level and deliver a full fetched version.
- **More efficient:** The users can run the system without any efforts or difficulties. The system is developed in a competent way and well organized.
- **Memorable:** The system elements are memorable in use. The users of the system will feel easy to remember the functioning of the system so that they can never face any problem in working with the system.
- **Error free:** The system is capable in running without generating any system related errors.

### Description of Problem Area

As we came to know that users are facing problem in two section one is searching results of search engine and secondly they wants to earn money. The solutions has been combined together to develop a system for user's benefits. The description and the solution for the highlighted problems has been described below.

- **Search Engine:** Nowadays there are many search engines available in the market but still users are not satisfied with the existing results. And also many users are not aware about earning online facilities. Users wants to have appropriate and updated search for their queries. The proposed system is planned in a way to provide updated result at the top by showing the number of days back a fact has been updated. Hence, they will get reorganized and efficient information.
- **Earning:** Everyone wants to earn money and in comparison with the existing online earning platforms, the suggested system is going to be developed with a thought of getting knowledge plus capital. Now a days there are several platforms which offers money for completing certain tasks. Even most of us are working on those kind of sites too, but the proposed system is not demanding for extra attention to work on it. It is just gifting money for their search.

The system has been proposed to provide a kind of structure that involves user for their advantage only. ‘Earning Explorer’ is something that is not bounded to search results only, it is about to encourage people and award users for their searches and other performed activities. Although they were exploring things with existing search engines also, but the proposed system will let them feel more interested in exploring the world as they are being paid for it. They can enjoy their earnings through searching and interacting activities.

Users can join the system for full time as well as for part time. With full time job there is no limitation on earning, customers will feel as an employed person. More they will work, more they will get. But, users with part time can enjoy the system and earn like to fulfill their pocket money or anything.

- **Nature of Challenge**

The system’s domain and the main challenges of the system are to generate search results and plus point calculation with a high accuracy. The developer needs to keenly observe the algorithm construction and application throughout the system process. The system should be developed with the noteworthy quality procedures and user friendly.

There are various leading challenges in the development of the system:

- Understanding the HDFS (Hadoop Distributed File System) cluster. As the system is totally based on Hadoop streaming so the developer should have vast knowledge and experience of working on it.
- Data collection, mining and handling. The structured data needs to travel into the system and analyzed on various stages of the system.
- To have a check on the applied algorithm (elastic search or solr) to produce precise and updated output of user's search. Users should receive updated and organized results related to the item searched.
- Performing prediction using predictive analysis. The system will collect the user's response in a structured form and apply the analysis for identifying the future performance.
- Visualization of the analyzed data will be done by using Apache Zeppelin and then collaborated with the web application of 'Earning Explorer'.

## SECTION C

### Brief description on project objectives.

- **Project Goal**

The main aim of the project is to develop a web based application on which user can perform search on any topic and earn by working on this system.

- **Project Objective**

The main objective of the project is to facilitate the users with an opportunity of doing and getting paid for their own work. Moreover, they can work on several activities offered by the system for money. The basic thing is to learn how to develop a better search engine using Hadoop streaming as the proposed system is totally based on Hadoop technology. The innovation done with the system is to award money in terms of points to users for operating on it.

The project's main objective lies in certain areas that are as follows.

- To provide a better searching facility with updated results. Using search bar, user can walk around whatsoever they want to hunt.
- To improve the users unemployment rate by offering money for their work done in the system.

- To interact users with different kind of surveys and perform those to get points.
- User can answer the daily asked questions in order to get points. The system will observe the percentage of user response for a particular question as a survey.
- Regular notifications for performing new activities to earn.
- Can refer the system to their friends in order to get advantage from their friend's earning.
- Watch videos of any choice.
- User's account settings are secured. They can make changes only after entering account password even though he or she is logged in to the system.
- **Learning Objective**

The developer is required to form new algorithms in order to compute the plus point for a particular user account. Aiming to accomplish the mentioned tasks, the developer need to study Hadoop technology and should be able to work proficiently under Hadoop.

As well as the developer must have strong knowledge of the technologies that are mentioned below:

- To understand Linux Operating System.
- To learn JSP and Bootstrap to produce a user friendly interface.
- Applying HCIU principles to enhance the system's usability.
- To apply project management concepts in the process of integrating the whole system.
- Machine learning and text mining to identify the most relevant information and its pattern from the large data set to evaluate and interpret the output.
- Distributed computing to solve a problem or task in a very small time interval. Problem will be divided into many sub tasks and each of which is solved by one or more computers.
- Data migration to transfer data within the system in order to perform different processes.
- Predictive analysis is to analyze the collected facts and generate a prediction over it or we can say user acceptance regarding any topic.
- Visualization to display the generated prediction in the form of graphs to make user clearer about the ratio.

## Target Audience

The proposed system is open for everyone whether they are student or teachers, employed or unemployed. Anyone who are willing to search about anything on internet and to receive currency. This system is capable in attracting people to join with it and work on it efficiently to make advantage of them. The system is mostly beneficial for those user groups who are daily linked with the internet.

### • Project Scope

Earning Explorer is a web application through which user can search solutions for their queries and each member of the system will be capable in earning points from their jobs done in the system and can redeem those as money. The structure is talented in data gathering and analyzing. It consists various functionality such as user account, searching, surveys, and weekly poll to collaborate search engine with earnings.

The system is constitutes of three module that are as follows.

### Profile Module

In this module, the user can create their account to perform different tasks available in the system. One user can refer the system to other person so that he or she can be one of the user of the system.

### Search Module

Basically the search module is defined as its user can search about anything in the system and they will receive an updated search result regarding the entered value. They can explore any type of data like text, images or videos.

### Earn Module

The earning module of the system is providing facility to earn money by attending jobs assigned to a user or even by searching any queries. This module will keep record of earned points with respect to user ids.

## **Notification Module**

The module is designed to manage the information that should be notified to the user. User will always be made remind about the tasks available for them, successful tasks done by them and about their earnings.

## **Visualization Module**

The module includes data gathering and data analyzing. Data gathering is to be done by collecting answers of the system's users corresponding to the questions asked on weekly basis. Hence, analysis of those collected data will be done and visualized in the form of different graphs.

## **Settings Module**

The module is featured with managing user account settings. User can view and change their account details or other account settings after going through the successful security so that no other invalid user can make any unwanted changes in the account.

### **• Functionalities of the System**

Functionalities explains the working anatomy of the system so that one can easily understand the constituent procedures in an according manner. It differentiate the authorities and accessibility of the system modules. In this project the functionalities has been categorized in three parts that are as follows.

- 4. Core Functionalities**
- 5. Enhanced Functionalities**
- 6. Special Features**

To make the Earning Explorer as a successful project, the system should be capable in completing the list of functionalities mention down.

### **1. Core Functionalities**

Core functionalities holds the essential functions without which the system's basic purpose is incomplete or the system cannot work.

S.no.	Functionality	Description	Module
1	<b>User Account</b>	Users are required to register themselves in order to create an account on Earning Explorer to perform search and earning events. For registration, they need to submit their personal details including email id and account password and a confirmation link will be sent to the user's email id. After confirmation, user can use the features of Earning Explorer.	Profile Module
2	<b>Searching</b>	User can search about anything by using search bar. It hosts a custom search engine that will periodically reward registered users with random amounts of plus points (P+).	Search Module, Earn Module
3	<b>Survey</b>	Survey has been included in the system to take user's criticisms on any topic. The members of Earning Explorer can do many surveys as they will be notified timely about new surveys to do. And hence they will be awarded with some specific plus points on completing the surveys successfully.	Earn Module
4	<b>Weekly Poll</b>	User can earn points on weekly basis by voting for a question. The given answers will be analyzed to make prediction about future of an issue.	Earn Module, Visualization Module
5	<b>Profile Settings</b>	User can view and change their account details or other account settings only when they are able to cross the security question successfully	Settings Module

		so that no undesirable user can access those settings.	
<b>6</b>	<b>Explorer TV</b>	Members can discover and watch videos on Earning Explorer, in a similar fashion to YouTube.	Search Module

## 2. Enhanced Functionalities

Enhance functionalities holds the tasks that is designed to improve the efficiency of system. Without enhanced functionalities the system can work but these function will improve the quality of system.

S.no.	Functionality	Description	Module
7.	<b>Updated Search</b>	<p>This functionality will let the system to show the updated results at the top. User can check that how many days back the search result has been updated or we can say when the link was updated.</p> <p>The developer is thinking and doing research to implement this functionality effectively to improve the quality of search results.</p>	Search Module
8.	<b>Referral Program</b>	A current user can invite another person to become a member of the system. If referred user signs up as a new user by using the current user's referral link, the current user can earn points that will be equal to 10 percent of the P+ earned over time by the new user.	Profile Module

9.	<b>Daily Goal</b>	<p>A meter on the home page will display daily an amount of plus point that a user should earn on that particular day that will be the goal of the day. And if the goal is reached, bonus will be awarded.</p> <p>The developer is thinking and doing research to implement this functionality effectively to make the user more focused on their earnings.</p>	Earn Module
10.	<b>Notification</b>	<p>User will get notified every time about the new surveys, messages or any type of news.</p> <p>The developer is thinking and doing research to implement this functionality to make the system more efficient and user friendly.</p>	Notification Module
11.	<b>Result Visualization</b>	<p>Weekly poll's answers will be collected as the structured form data and will be analyzed to make predictions on the statement. The visualization of analyzed data will be performed in the form of graphs so that user can understand the results more effectively.</p>	Visualization Module
12.	<b>User Guide</b>	<p>The user guide basically tells about how the system will work. Every user can access this facility whenever they want and it will help them to perform any task effortlessly.</p>	Profile Module

### 3. Special Features

Special features holds the tasks that is designed to make the developed in more demand. This will let the system to work differently in compare with the existing systems. The system will come up as an extraordinary system.

S.no.	Functionality	Description	Module
2.	<b>Bookmark</b>	<p>The user can save any link or search result in the bookmark section as some user tends to store some important pages for their future use. The system aims to allow user to access those pages even when they don't have internet.</p> <p>The developer wants to add this and doing research so that she can implement it to make the system approachable.</p>	Search Module

- **Project Limitations**

Everything in this world have some limitations where we can't use that thing. The Earning Explorer also deals with certain limitation in its application. These are as follows.

5. **Interface:** The proposed system is web based application only. It will not be applicable as desktop application. Hence, it is interface specific.
6. **Internet:** User will be able to run the system with internet connection only.
7. **Search Term:** Users are required to know how to extract the appropriate search results that means the way of writing search query.
8. **Email Account:** Users will not be able to register their account in the system until they have any email account.

- **Project Assumptions**

The system Earning Explorer has been framed by keeping some assumption for the successful implementation of the application. The assumptions made in the system are being shared below.

- 6 **Literacy:** The users of the system are computer literate and understand the simple English language.
- 7 **Internet:** The users must have internet connection.
- 8 **Email Security:** The user email ids must have enabled the option of ‘security of receiving email from any source’ so that the email sent by the system should be received in the email inbox.
- 9 **Email Account:** Users will not be able to register their account in the system until they have any email account.

- **Project Deliverables**

The project will deliver a full functional system for its end users to hunt the solutions for their problems and let them earn from the system. The project’s purpose is to develop a system for meeting the user’s requirements in two areas that are as follows.

1. **Searching:** The proposed system is going to provide updated result for the searches made in the system at the top by indicating the number of days back a fact has been updated. Hence, they will get reorganized and efficient information.
2. **Earning:** The suggested system is going to be developed with a thought of providing income to the users for their every tasks associated with the system. The users will get chance to earn money for doing searches, surveys, weekly polls and watching videos.

At the end of this project, the project will be delivered in the below mentioned ways.

1. A prototype of the proposed system application in a DVD.
2. Soft copy of the project’s documentation.
3. Hard copy of the project’s documentation.

## SECTION D

- **Hardware Requirements**

The hardware specification for the project development has been mentioned below.

1. Device: Laptop or Desktop PC
2. Processor: At least Intel® Core™ i3-3110M
3. CPU: At least 1.90GHz
4. RAM: At least 7.5 GB free space in 8GB for faster response
5. Monitor: Any standard size
6. Other : USB, Dongle or Wi-Fi router for internet connection

- **Software Requirements**

The software specification for the project development has been mentioned below.

Operating System: Linux (RedHat, Ubuntu), Windows 10

Software/ Tools:

- ✓ NetBeans IDE 8.0.2
- ✓ Oracle VM Virtual Box
- ✓ Cloudera Manager
- ✓ JRE 8, JDK 1.8

Technology:

- ✓ Hadoop framework-2.3.0
- ✓ Hadoop2x-eclipse-plugin-master
- ✓ JSP
- ✓ Bootstrap

For running the system: Firefox / Google chrome.

- **Data Gathering Techniques**

Data gathering is the base of any kind of assignment. For nurturing an idea we need data and then establishment takes place. Major part of a project depends on information relative to a specific

purpose. It helps in collecting data for initiating a task by applying different methods. Basically, there are two ways of data gathering methods:

**Primary Research:** It is an own research that is done by putting on own mind and skills.

**Secondary Research:** It is a research that is done on other research. We explore the already published research papers that includes journals, e-books, websites and any existing system. It should be done before primary research.

In primary research, the developer has used two techniques for data gathering:

### 1. Interview

The interview is a technique of data gathering which involves two or more personal opinions. The persons are those who have active participation for a particular topic so that they can inform better about the subject. It is done with a high level security in between of interviewee and interviewer. The provided information will be collected by the developer for further discussion to analyze the user responses and no other person can see those information. It will help in enhancing the system efficiency.

### 2. Questionnaire

Questionnaire covers a large part of audience and helps in gathering more data in one time. It is easy to evaluate to seek more relevant facts. The questions asked in questionnaire are technical and domain based, user interface and user satisfaction. The different questionnaire will be distributed to different occupation holders and age groups in order to get the most precise results. The result will be beneficial in further development phases.

## SECTION E

### • Domain Research and Technical Research

Research is a careful and detailed study into a specific problem, concern, or issue using the scientific method. This is best accomplished by turning the issue into a question, with the intent of the research to answer the question. This project needs extensive understanding and research work on different concepts for providing effective and efficient solutions. Since the project is a web based application which will be implemented using Hadoop technology and Java Server Pages

(JSP) and its application development methods, it needs research on each field. Some of the research areas worth mentioning are listed below.

### **Java Server Pages (JSP) and Bootstrap**

Through Bootstrap developer can make the user interface of the system more interactive and attractive to the users. And JSP programming language will be used to make pages interact to each other efficiently.

#### **Books:**

- ✓ Basham, B., Sierra, K. and Bates, B. (2008). *Head First Servlets and JSP*. 2nd ed. O'Reilly Media.

#### **Websites:**

- ✓ W3schools.com. (2016). *Bootstrap 3 Tutorial*. [online] Available at: <http://www.w3schools.com/bootstrap/>.
- ✓ Free Code Camp. (2016). *Learn to code and help nonprofits*. [online] Available at: <https://www.freecodecamp.com/challenges/use-responsive-design-with-bootstrap-fluid-containers>.
- ✓ www.tutorialspoint.com. (2016). *JSP Tutorial*. [online] Available at: <http://www.tutorialspoint.com/jsp/>.

### **Human Computer Interaction and Usability Principles (HCIU)**

It is a well-known fact that users love to work on the system's with which they can interact easily and which provide them enjoyable experience.

#### **Books:**

- ✓ Faulkner, X. (2000). Usability engineering. 1st ed. Basingstoke, Hampshire: Palgrave.
- ✓ Nielsen, J. (1993). Usability engineering. 1st ed. Boston: Academic Press.
- ✓ Zacarias, M. and Oliveira, J. (2012). Human-computer interaction. 1st ed. Berlin: Springer.

### Websites:

- ✓ Hcibib.org, (2014). HCI Bibliography: Human-Computer Interaction Resources. [online] Available at: <http://hcibib.org/>.
- ✓ Cockton, G. (2013). Usability Evaluation. The Encyclopedia of Human-Computer Interaction, 2nd Ed., [online] p.-. Available at: [http://www.interaction-design.org/encyclopedia/usability\\_evaluation.html](http://www.interaction-design.org/encyclopedia/usability_evaluation.html).
- ✓ Usabilityfirst.com, (2014). Usability First: Usability in Website and Software Design. [online] Available at: <http://www.usabilityfirst.com/>.

### Research papers:

- ✓ Cs.cmu.edu, (2015). *A Brief History of Human Computer Interaction Technology*. [online] Available at: <http://www.cs.cmu.edu/~amulet/papers/uistory.tr.html>.
- ✓ Webcache.googleusercontent.com, (2015). *Human-Computer Interaction: An Overview*. [online] Available at: <http://webcache.googleusercontent.com/search?q=cache:3YFbMeyi814J:www.s2is.org/Issues/v1/n1/papers/paper9.pdf+&cd=2&hl=en&ct=clnk&gl=in>.

### Hadoop Domain

Hadoop is an open-source framework to store and process big data in a distributed environment using simple programming representations. Various technologies will be used with Hadoop to perform search functionality of the system.

### Books:

- ✓ Pujari, A. (2001). *Data mining techniques*. Hyderabad: Universities Press.
- ✓ White, T. (2012). *Hadoop: The Definitive Guide, 3rd Edition*. 3rd ed. O'Reilly Media / Yahoo Press.

### Websites:

- 
- ✓ www.tutorialspoint.com. (2016). *Hadoop Tutorial*. [online] Available at: <http://www.tutorialspoint.com/hadoop/>.
  - ✓ Intellipaat Blog. (2016). *Data Visualization via Apache Zeppelin - Intellipaat Blog*. [online] Available at: <https://intellipaat.com/blog/data-visualization-zeppelin/> [Accessed 13 Oct. 2016].
  - ✓ Hortonworks. (2016). *Apache Zeppelin - Hortonworks*. [online] Available at: <http://hortonworks.com/apache/zeppelin/#> [Accessed 13 Oct. 2016].
  - ✓ Dean, J. and Ghemawat, S. (2004). MapReduce: Simplified Data Processing on Large Clusters. *Google, Inc.* [online] Available at: <https://static.googleusercontent.com/media/research.google.com/en//archive/mapreduce-osdi04.pdf> [Accessed 14 Nov. 2016].

### Research Papers:

- ✓ Hadoop and HDFS: Storage for next generation data management. (2016). 105.
- ✓ Phadatare, G., Rathod, K., Shukla, V. and Salunkhe, M. (2014). GooDoop: Local Search Engine using Hadoop. *International Journal of Computer Science and Information Technologies*, 5 (3)(0975-9646), pp.3508-3510.
- ✓ Dean, J. and Ghemawat, S. (2008). MapReduce. *Communications of the ACM*, 51(1), p.107.
- ✓ Jain, A. and Kaushik, M. (2014). Performance Optimization in Big Data Predictive Analytics. *International Journal of Advanced Research in Computer Science and Software Engineering*, [online] 4(2277 128X), pp.126-129. Available at: [http://www.ijarcsse.com/docs/papers/Volume\\_4/8\\_August2014/V4I8-0247.pdf](http://www.ijarcsse.com/docs/papers/Volume_4/8_August2014/V4I8-0247.pdf).

### Project Management

For successful running of the project the concepts of project management must be clear.

### Books:

- ✓ Burke, R 2009, Project Management : Planning and Control Techniques, Wiley India Pvt Ltd., INDIA.

- ✓ O'Connell,F 2001, How to Run Successful Projects : The Silver Bullet, Pearson Education, UK.

## SECTION F

The developer has evaluated almost every methodologies by doing research and analysis of the software development methodologies. After reviewing about the methods, the developer has come up with the decision of selecting Agile Methodology to be followed for the proposed system development.

The Agile methodology has been examined well with respect to the project's requirements and it is found that it can associate with the constraints of the project. This methodology follows an iterative as well as an incremental approach in which any kind of modifications can be done after planning as per client's feedback. The main concentration of agile method is to early release of the software product by responding to adopt the changes required in specification.

- **Agile Methodology**

The process includes iteration to deliver different working software anatomy after each iteration of planning, requirement analysis, designing, implementation and testing. The end product of each iteration is incremental in terms of features as the final product holds all the wanted features required by the customer.

The developer has opted for agile methodology for the following reasons.

7. The project's requirements are not clear at the initial stage, hence changes can be done on the basis of research accordingly. This models allows to modify the things unlike the waterfall methodology.
8. Because of iterations done in the process the product is tested very frequently so there is less chance of having any risk in project's future.
9. This model has been undertaken because the project is required to complete within a given time frame.
10. It provides flexibility to the developer to think independently.

- 11.** It focuses on technical excellence and enhanced design for the project.
- 12.** The methodology allows to have more user interaction while the development of the system.

- **Methodology Comparison**

The selected methodology i.e. Agile Methodology has been compared with other two methodologies to justify the selection of Agile method. The comparison has been performed with Waterfall and Spiral model.

Features	Waterfall	Spiral	Agile
<b>Requirements Specification</b>	In the beginning.	Frequently changes according to users.	Frequently changes according to the users.
<b>Project Duration</b>	Best suited for short term.	Best suited for long term.	Best suited for both short and long term.
<b>Complexity</b>	Easy to understand	Process is complex.	Easy to manage.
<b>Flexibility</b>	Not flexible.	Less flexible.	Flexible.
<b>Cost</b>	Not costly.	Costly.	Costly.
<b>Risky Projects</b>	Doesn't support.	Supports high risk projects.	Supports high risk projects.
<b>Success Rate</b>	Low	High	High
<b>Customer Involvement</b>	At the end of the project.	After each iteration but less involvement.	Customer evaluation, after each iteration.
<b>Testing</b>	At final stage.	After every phase.	After every iteration.
<b>Maintenance</b>	Hard to maintain.	Maintenance is required.	Easy to maintain.
<b>Implementation</b>	Easy.	Complex.	Easy.

- **Development Plan**

Calculating the duration to be allotted for each phase of the Agile Methodology.

## TEST STRATEGY SCHEDULE

TEST STRATEGY SCHEDULE			
Starting Date		1, August, 2016	
Ending Date		18, April, 2017	
Total Duration		33 Weeks	
S.no.	Phases	Tasks	Duration
1.	<b>Project Definition</b>	Search a topic Idea Generation Project Title Selection Draft Proposal Form	42 Days
2.	<b>Project Planning (Iteration)</b>	Schedule and Time Estimation Gantt Chart Project Proposal Form	6 Days
3.	<b>Requirement Analysis (Iteration)</b>	Identify project specifications <ul style="list-style-type: none"> <li>• Project Background</li> <li>• Project Limitation and Assumption</li> <li>• Techniques to be learnt</li> <li>• Goals, Objectives and Scope</li> <li>• System Functionalities</li> <li>• Methodology to be chosen</li> </ul> Project Specification Form Feasibility Study Investigation and Research <ul style="list-style-type: none"> <li>• Academic Research</li> <li>• Primary Research</li> <li>• Secondary Research</li> <li>• Questionnaire</li> <li>• Interview</li> </ul>	41 Days

		<b>Analysis</b> <ul style="list-style-type: none"> <li>• Questionnaire</li> <li>• Interview</li> <li>• Overall conclusion</li> </ul> <p>Mid-point submission.</p>	
4.	<b>System Design (Iteration)</b>	1. UML Design <ul style="list-style-type: none"> <li>• Use Case</li> <li>• Data flow Diagram</li> <li>• Class Diagram</li> <li>• Sequence Diagram</li> </ul> 2. Database Design <ul style="list-style-type: none"> <li>• ERD</li> <li>• Normalization</li> </ul>	38 Days
5.	<b>Implementation (Iteration)</b>	1. Code generation 2. Module integration	50 Days
6.	<b>Testing (Iteration)</b>	1. Prototype Evaluation 2. Unit Testing 3. Integration Testing 4. System Testing 5. Functional Testing 6. User Acceptance Testing 7. Compatibility Testing 8. Runtime Testing 9. Configuration Testing	36 Days
7.	<b>Project End</b>	Documentation  Finalizing Hard Copy and Soft Copy  Presentation	23 Days

## • Description of Hardest Function

The developer will have to be very focused and durable in implementation phase as it includes the coding generation and module integration. The technology in which project is going to be developed is a big challenge for the developer. There are two area in which developer needs to work wisely are as follows.

### 1. Distributed Computing

Hadoop supports distributed file system that operates on cluster workstations to expedite Big Data. The data are processed by using MapReduce which stores these data in a Hadoop file system. The form of data is not in structures as in a relational database, we need to convert the unstructured form into a proper arrangement of data.

### 2. Predictive Analysis

To perform Predictive Analysis in the system, the developer needs to go through different sub processes. The Predictive Analysis includes the process of data collection, data analysis, statistics, modeling, deployment and model monitoring.

- ✓ Data Collection: It is data mining to arrange data from multiple sources to do analysis on it.
- ✓ Data Analysis: It embraces the process of inspecting, cleaning, transforming and modelling data to make a conclusion from it.
- ✓ Statistics: It validates the assumption and test them by using standard statistical models.
- ✓ Modeling: It is the process of creating accurate predictive models about future.
- ✓ Deployment: It offers to deploy the analytical result in to the daily decision making process. The decisions are based on modeling.
- ✓ Model Monitoring: This process manages monitoring of model performance to ensure that it should provide the expected results.

## SECTION G

### Evaluation and Test Plans

Testing is the process of executing a program and finding out and solving all possible bugs.

It is the presence of error in the project but it will not ensure the absence of errors in the project. It helps in exploring the product quality. Testing includes:

- Verification (Step by step checking of milestones)
- Validation (Final check of the user's requirements)
- **Basic Testing:** The testing has been categorized basically in two types.

### **1. Black Box Testing**

Black box testing is used to test the software from user's point of view, there is no internal code view. It is done on each module of the system. It focuses on the output generating against any input occurred or execution happened in the system. It is performed by developers only.

### **2. White Box Testing**

White box testing is a set of testing procedures that evaluate the code and internal structure of a program. It is done at complex modules of the system. It is done by both developers and testers. It gives the clear view of errors like on which line of code fault has occurred. It shows the line by line execution.

- **Testing Processes**

S.no.	Testing	Description	Tested By
1.	<b>Unit Testing</b>	Unit testing is done at every stage of the system functions. The system has been divided into modules or units and testing will be done on these units independently. For example: On login button click, passing username and password	Developer
2.	<b>Integration Testing</b>	Module (Profile, Search, Earn, Notification, Visualization, Settings) are tested in a group. Hence the modules are put together for the integration testing.	Developer

3.	<b>System Testing</b>	It verifies that all modules of the system should respond to each other properly and the overall system performance is achieved. It validates tangible and intangible requirements e.g. time, searching.	Developer
4.	<b>User Acceptance</b>	The user acceptance testing will help the developer to produce a user friendly system.	End users
5.	<b>Usability Testing</b>	The usability testing will help the developer to produce a usable and efficient system	End users
6.	<b>Compatibility Testing</b>	It will check whether the system can run or responding properly on other operating system or not.	Developer, End users
7.	<b>Runtime Testing</b>	It will check the efficiency of the system in context of loading time, execution problem and other run time issues.	Developer, End users
8.	<b>White Box Testing</b>	The system will be tested for complex features to make it compatible with the desired output.	Developer, End users

- **Success Criteria**

For the successful working of the system, the success criteria need to be evaluated are.

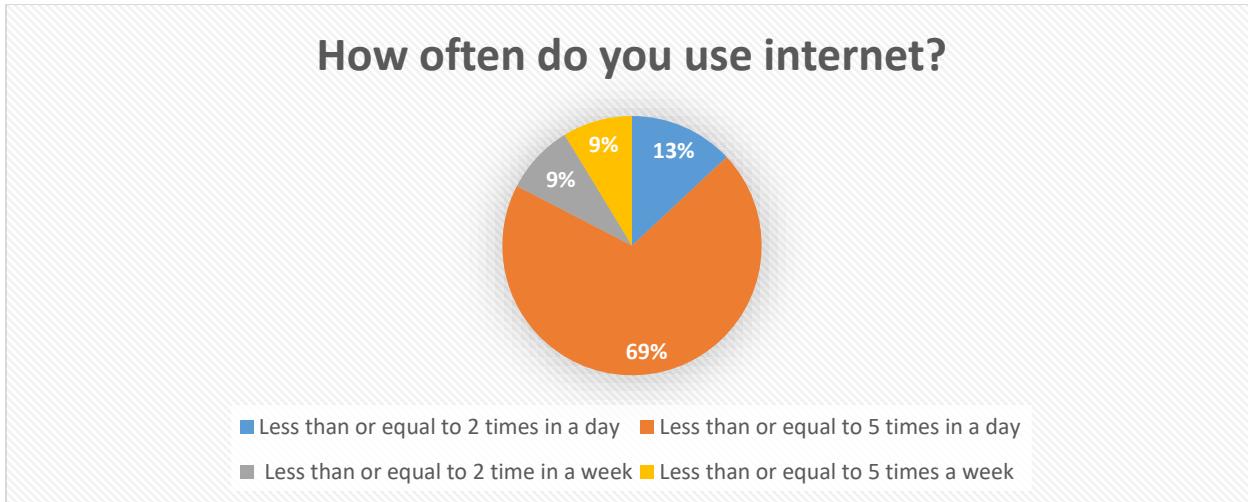
1. To meet user requirements
2. Researches regarding the subject to obtain an accountable information.
3. Analysis of the research.
4. Functionalities implementation according to the requirements.
5. Usability testing to check the system efficiency.
6. Project management to handle the project.
7. Proper documentation to be done.

# ANALYSIS REPORT

## QUESTIONNAIRE ANALYSIS

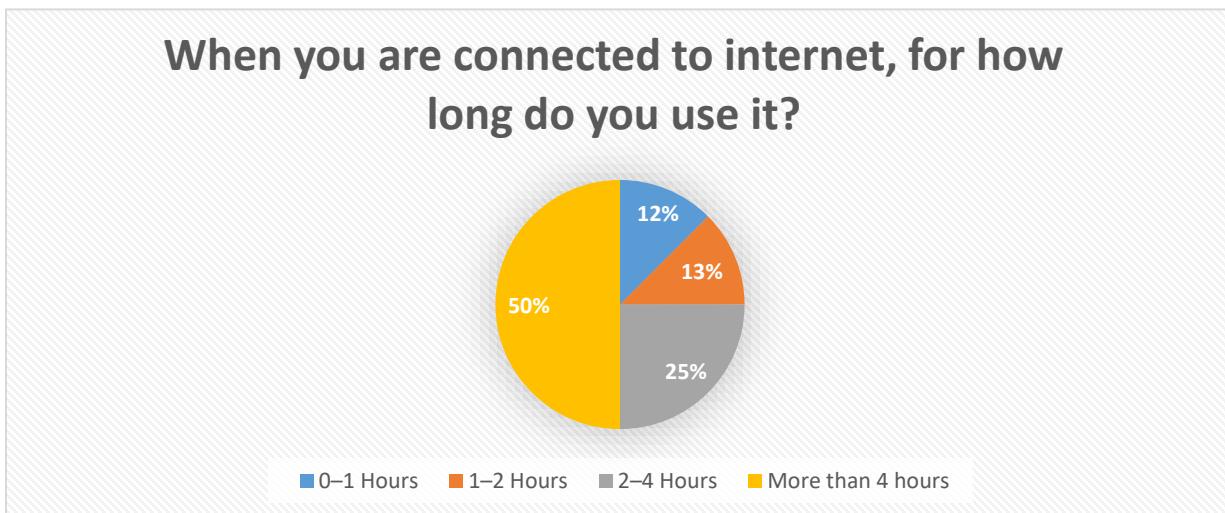
After conducting and evaluating the different questions provided in the questionnaire the developer has concluded the respective results graphically (Pie chart).

- **Question 1**



**Justification:** There are almost 70 % of the users who are using the internet very often that means maximum they are using for 5 times per day. The data will be helpful in deciding on average that how many can use search engine among those time. If maximum number of users are there then this survey will result in accurate responses.

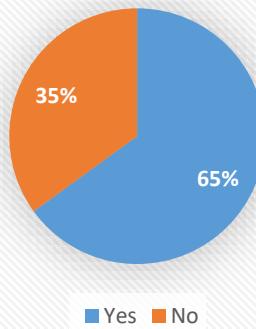
- **Question 2**



**Justification:** There are almost 50 % of the users who are using the internet for more than 4 hours and we have 1/4<sup>th</sup> users who are using internet for 2-4 hours. This shows that there are many people who are using internet frequently and for longer time. Hence, they are more connected with the internet and knowing many terms that are going to be asked further. If maximum number of appropriate users are there in the survey then they can respond seriously for the further questions.

- **Question 3**

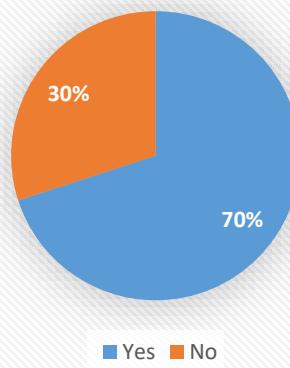
**Are you aware about the facilities that provides you to earn online (On internet)?**



**Justification:** There are almost more than 60 % of the users who are already aware about the earning facility provided by the different organizations online. Hence, the user can easily understand the system working if they have gone through these websites.

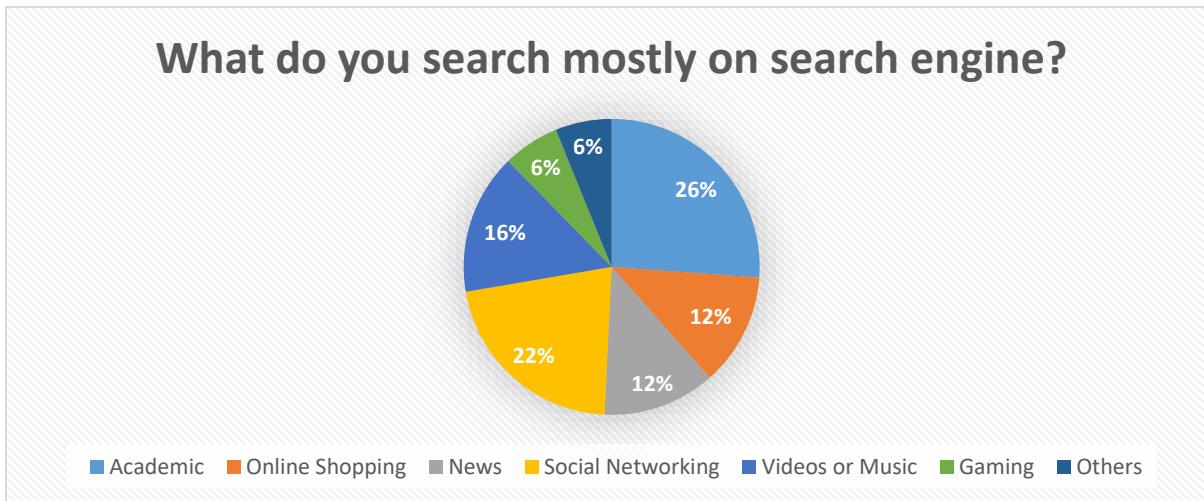
- **Question 4**

**Do you use search engine?**



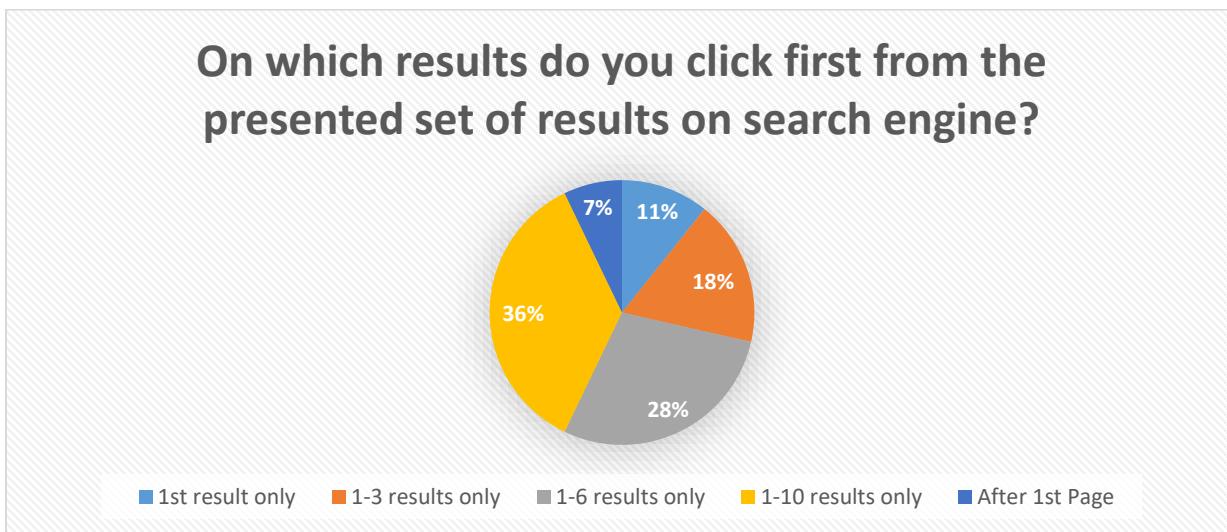
**Justification:** 7/10 have searched on the search engine. Therefore, the developer has analyzed that if 70 % users are using it then the system can get proper feedback about the exiting systems and can suggest better for the proposed system. The developer is not required to convey the meaning of search engine to a large group.

- **Question 5**



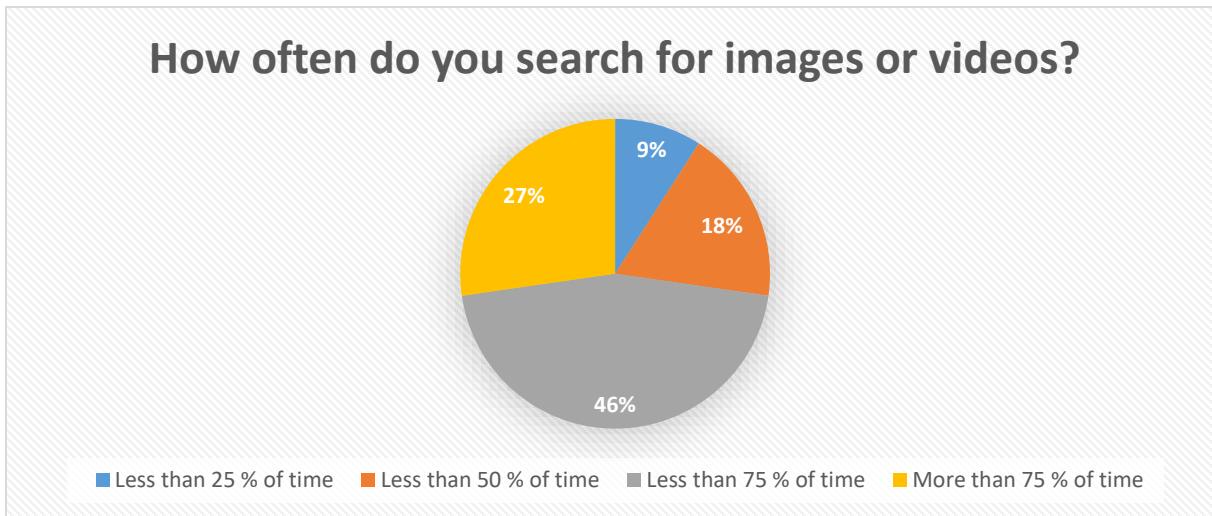
**Justification:** Maximum number of users are using search engine for searching academic purposes, social networking details and video and music. Hence, the developer has analyzed that the maximum users are using search engine for several purposes so they can give proper feedback about the service of the exiting systems and can suggest better for the proposed system.

- **Question 6**



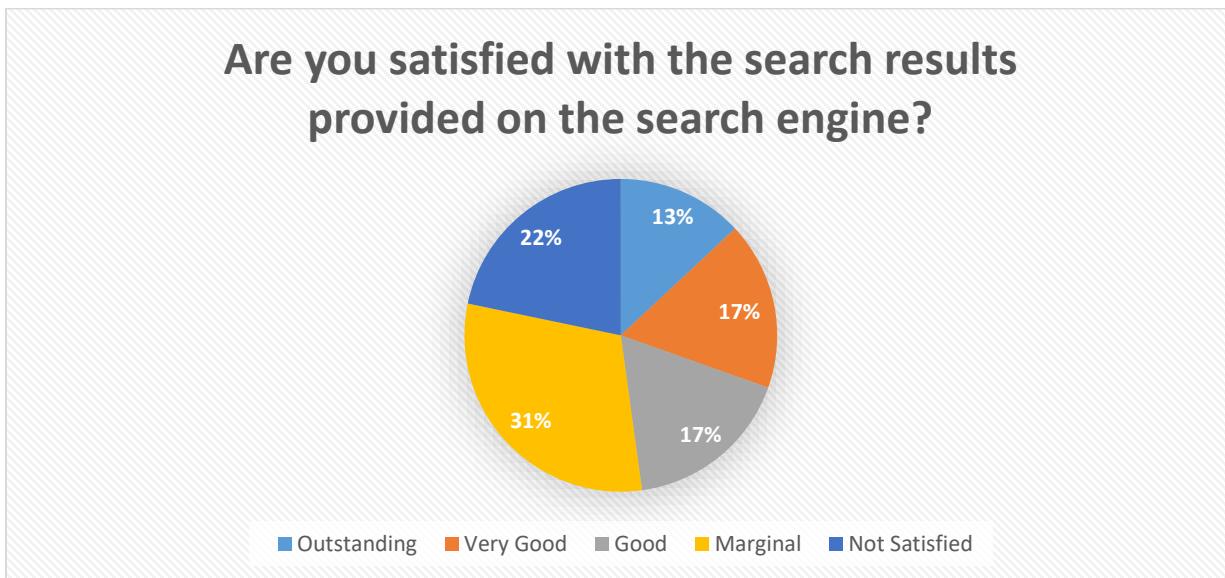
**Justification:** According to the survey, maximum users are not going to next page for the results. There are users who prefer to visit the links from 1 to 10 only. Hence, the developer should build the system in a way that the important updated links should come at the top according to searched query.

- **Question 7**



**Justification:** Approx. 50 % of the users are searching many times for the videos or images. Hence, these users can give proper feedback about the exiting systems and can suggest better for the proposed system to implement efficient search facility.

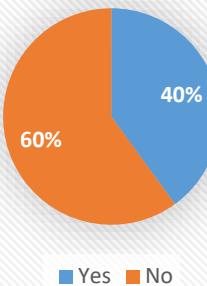
- **Question 8**



**Justification:** The 30 % users who are saying the search result is marginal and 20 % users are not satisfied with the existing search engine results. Hence, the developer should consider the data on a serious note and understand the user's requirements deeply to develop a successful project.

- **Question 9**

### Do you feel that you receive updated search results?

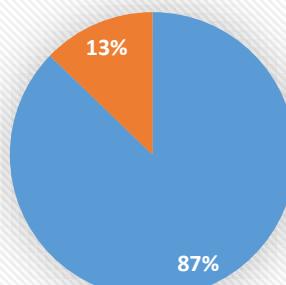


■ Yes ■ No

**Justification:** There are 60 % users who are saying that they do not feel that they are getting updated search result from the existing system. The developer is required to understand the feeling of the users regarding the results, what is to be needed to implement so that users can gain trust on the system. Hence, the developer should consider the data on a serious note and understand the user's requirements deeply to develop a successful project.

- **Question 10**

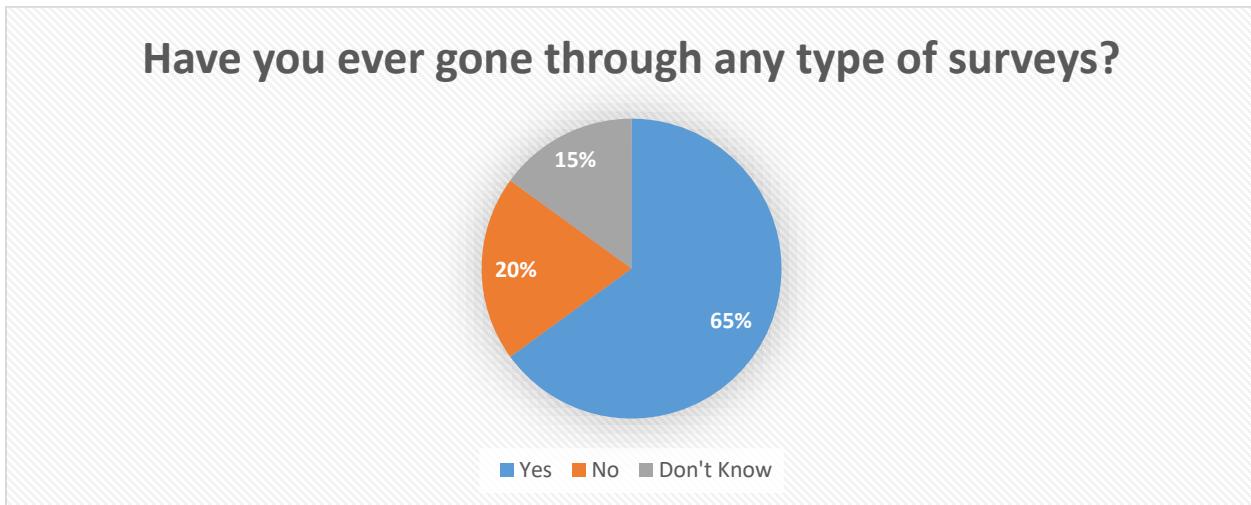
### Are you satisfied with the bookmark utility provided in the web browser?



■ Yes ■ No

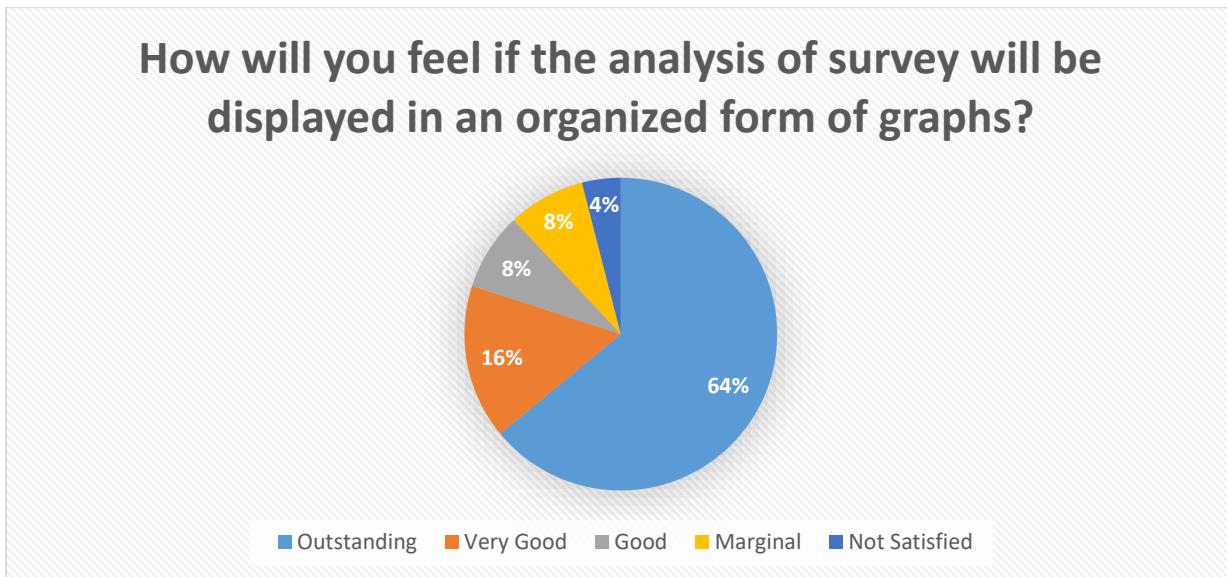
**Justification:** Almost 9/10 members have totally accepted the bookmark utility. Hence, it is clear that if the developer will add the bookmark with the search engine the users will feel good about it and it will be a successful function of the system.

- **Question 11**



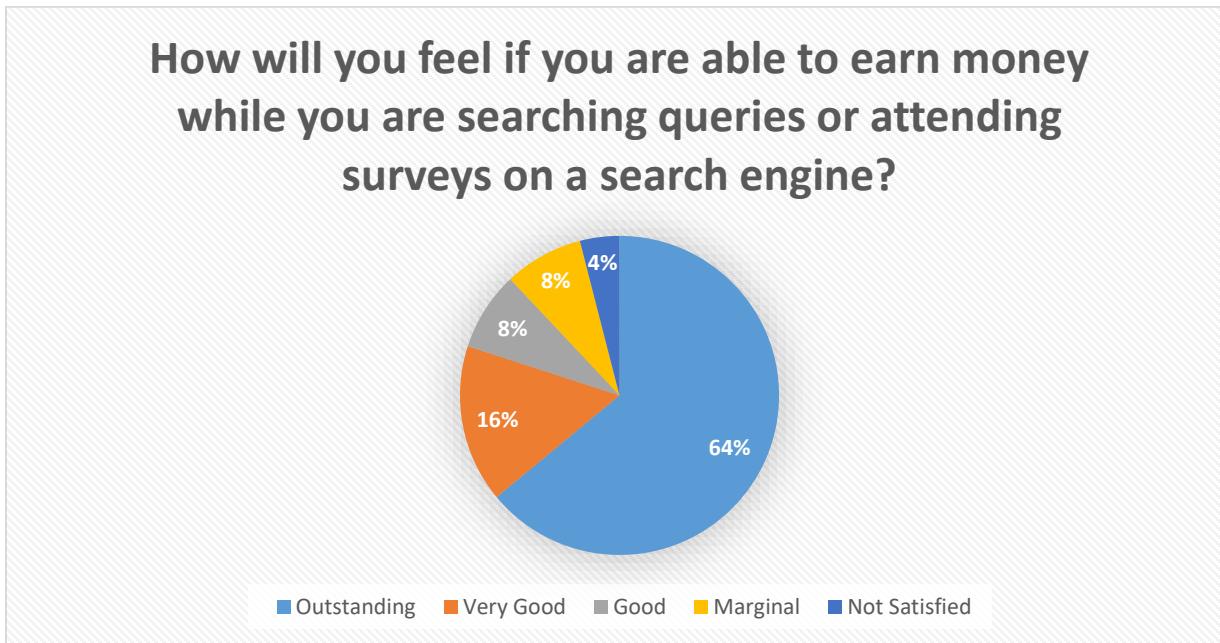
**Justification:** Almost 7/10 users are attending the survey and know how to perform a survey. Hence, they can easily understand the system purpose and earn money by performing many successful surveys.

- **Question 12**



**Justification:** Almost 70 % users are satisfied and feel good about having visualization for the survey analysis. Hence, the developer should do the visualization in a more advance way like charts or on world map.

- **Question 13**



**Justification:** Maximum users are in the favor of availing the facility of earning while searching or attending different activities online. Hence, the developer should develop an efficient system in order to reach the customer satisfaction.

## INTERVIEW ANALYSIS

The below mentioned matrix shows the user responses in terms of positive and negative to check the overall positive and negative outcome.

RESPONSE NOTATION	
	Strongly Positive
	Positive
	Neutral
	Negative
	Strong Negative

Interview Analysis Matrix									
Attend	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									

## QUESTIONS ASKED

### **Q1. What is your opinion about the existing search engine and their search results?**

**Conclusion:** Almost every age group users uses search engines and are many of them are not able to get the appropriate output. They are quoting that it takes a lot of time to give the specific result on any subject may be because of improper query. Sometimes they feel that they should be able to see the updated results according to date wise.

### **Q2. How do you search for videos? Do you face any problem in searching?**

**Conclusion:** Most user opens the YouTube for video search as it is a great platform for videos. But, the users of age group 40-60 are still giving query to the search engine. Hence, they are facing problem in finding the desired videos. The proposed system should give video result in searching web if applicable.

### **Q3. How do you feel about the facilities that provides you to earn online?**

**Conclusion:** Almost 60 % teenage and young customers are excited to look after the system of online earning and few have created accounts and working on these facility providers. These users are claiming that the online earning projects provide very slow income.

### **Q4. How is your experience about the online surveys?**

**Conclusion:** Almost 50 % of teenage users are attending the online surveys and saying that surveys are good to increase the knowledge regarding the subject asked in the survey. It let you know about the company expectations. Some surveys are very long and are not specific.

### **Q5. How will you feel if the analysis of survey are displayed in an organized form of graphs?**

**Conclusion:** Well, every user are expecting a good report of whatever they are submitting. A user can better understand the graphical representation of the results.

**Q6. How much you are satisfied with the bookmark utility provided in the web browser?**

**Conclusion:** Every user are satisfied to have bookmark utility with a search engine as if they save any page for accessing it in future.

**Q7. What do you want to say if you will get to earn money while you are searching queries or attending surveys on a search engine?**

**Conclusion:** Through this question the developer has concluded that users are feeling advantage for them as they will get paid for these regular kind of works. A large number of users are willing to be a part of the system

**Q8. Would you like to give some suggestions for improving the existing system?**

**Conclusion:** User suggestions regarding the existing systems of earning website are :

It should provide facilities for user benefits only as they advertising to the world.

It should not be efficient consume less time of users

# IMPORTANT DOCUMENTS

- **Filled Questionnaire**
- **Interview Record**
- **Ethical Form**
- **Log Sheet**
- **Gantt Chart**