**Sri Lanka Institute Of Information Technology**

****

Comprehensive Design/Analysis Projects

Project Proposal

**Vigraha High Volume CDR Analyzing Framework**

Title of the project :Vigraha High Volume CDR Analyzing Framework

Project ID :

Student name :ThejaniDineshikaWedikkaraArachchi

Supervisor name :Mrs. SujaniSumanadasa

Co - Supervisor name :

Supervisor signature :

Co - Supervisor signature :

**Declaration**

I hereby declare that the proposed project work entitled” **Vigraha High Volume CDR Analyzing Framework**”submittedto the Sri Lanka Institute of Information Technology, is a record of an original workdone by me.

Date Signature

**Abstract**

In this project proposal it concerns the***“Vigraha High Volume CDR Analyzing Framework”***. In here it mainly focuses on distributed computing. In the sense communication is a major factor in today world. Therefore here it mainly focuses on brand company’s announcements through mobile companies to their customers. The major contribution of this is analyzing data.

In this application it takes all brand companies requirements as inputs and according to requirements it generates the messages for customers through SMS. To select the target customer group here it uses some log files which already available at mobile companies. It captures all customer details under these files. Here it mainly focuses on five log files to gather customer information since the current mobile company’s only concern about one log file when they provide messages to customers according to brand company’s requirements. It keeps all these records using a single database and analyze when brand companies requirements available. Then it provides a SMS for all the necessary customers and sends the messages. If the server has some failure it sends a SMS to the person who is responsible at that moment.

As the outcome of this system it generates SMS for selected customer group by analyzing millions of records. Also the system is providing reports as the brand companies’ wish.Assume the brand company needs to know to whom mobile company sends SMS obviously the system need to provide a report for them.So this going to be more useful area for future as well as today hence the communication is the most powerful area in today.

**Table of content**

**Chapter 01**

**Introduction**

The proposed project is about application for distributed computing. In the world most important fact is communication. To do that now uses many methods such as mobile phones, internet and so on. Here it mainly focuses on mobile phones. In mobile phones there are many features such as SMS, voice calls, LBS, GPRS and USSD. Mobile companies they use log files to record customer information regarding these. It has customer’s phone number, receiver’s phone number, location and some other information. Log files record the activity of MVS server components and can be used to troubleshoot issues with the components[1].Suppose any other brand company need to inform their customers that they have a promotion or some other special event. So they tell it to the mobile company under some requirements. Then the system inserts those details and generates the rule. Then it executes and selects the necessary customers to send SMS. Here it analyzes millions of records per day.

We divide the system into four major independent components to develop.

* **Data loader**
* **Admin UI**
* **Reporting UI**
* **Rule Engine**

Data loader **–** This takes records from log files and load it to database.

Admin UI – Here this generates the rule and it includes some other functionality such as registration, login to the system, change password.

Reporting UI –Here it generates reports. Under this it has to generate thirty reports including drill down report.

Rule Engine – It execute rules by rule engine and analyze records by apache hadoop map reduce programme.

The project is worth, because in existing systems which are in mobile companies they only consider about only one log file to select suitable customer groups, because there is no system to analyze more log files since it contains huge amount of records. But here it considers about five log files to analyze data. So the project is very useful when brand companies need to inform their special events by using mobile companies. The system analyzes millions of records per day including special days. Also it generates reports according to brand Company’s wish. So this is worth for brand companies. Then the mobile company can increase their income by this, because has five log files to analyze. So this is worth for mobile companies. Also customers can know about brand company’s promotions. So it is worth for mobile users also. So the system is very useful area for mobile companies, brand companies and mobile users.

There are many objectives visible with the development of the proposed system, because it will be a new chapter for mobile companies since their existing system support for one log file and this analyze data by five log files. So there have huge number of records to handle for each of the day. In many ways the system helps to their clients.

The system contains several objectives.

1. Provide service for brand companies - This system useful for brand companies such as pizza hut, ODEL and so on. To announce their promotions, events, occasions or any other special functions they can inform their customers through mobile companies. They only need to tell about their requirements. Then the all tasks will handle by mobile company.
2. Inform customers - When the mobile company sends SMS for necessary customers they can know about what are the things they can consume from brand companies. Assume mobile company is having requirements from pizza hut and they announce about pizza promotion to customers and at that moment a customer needs to it, actually the message will help him to fulfill his need.
3. Handling millions of records - In this system it has to deal with millions of records, since it receives millions of customer details per day by voice calls, SMS, LBS, GPRS and USSD. So the system has to maintain each and every record in proper manner and it able to select necessary records according to their customer requirements. Then generate the SMS for customers. Also for special days such as new year, year end and so on the mobile companies receives more customer information than regular days.
4. High availability - In here the server should up and running each and every day, since it deal with customer records. So maintain system’s availability is more important. Also if the server face to a problem it inform the person who responsible the system at that moment. By this it maintains the system availability.
5. Generate reports - In order to providing the summary report to the brand companies is another objective, because if the brand company need to know to whom mobile company sends SMS obviously the system need to provide a report for them.

This kind of similar project has done in early as industrial project. They only concern about one log file to analyze data and to select customers, because it is hard to analyze records by considering more than one log file since it contains huge number of records.Alsoto analyze data they do not use Hadoopintegration technology. So the existing system cannot analyze huge amount of records as the proposed system. But proposed system it concern about five log files and it receives millions of records. So to maintain them it uses Hadoopintegration.

To do the proposed project has to be mastered on java IDE. Here it uses the NetBeans environment. Also need to familiar with java language. As the database it uses MYSQL, it selected because of that database support for java language. Also to do the data analyzing part need to master about apache hadoop map reduce programme. To do my contribution it requires sprint or iteration which relates with scrum model also.In the Scrum method of agile software development, work is confined to a regular, repeatable work cycle, known as a sprint or iteration [2].Except these techniques the system requires the specification of log files to prepare them, because mobile companies do not provide their customer details since it cause to be a security problem. So we have to study the methods to create simulators to create customer records also.

**Chapter 02**

**Objectives**

There are many objectives visible with the development of the proposed system, because it will be a new chapter for mobile companies since their existing system support for only one log file and this system analyze data by five log files. So there have huge number of records to handle per day. By system’s objectives it helps to their clients in many ways.

Main objectives:

* + To analyze millions of records.
  + To provide service to mobile companies, brand companies.

Except major objectives there are some other objectives also:

* To introduce a system which has more features than available systems.
* To inform mobile customers by SMS about promotions and special events of brand companies.
* To select suitable customer groups from millions of records and send SMS for them.
* To generate reports for companies.
* To categorize the mobile users by log files such as voice call, SMS, GPS, LBS and USSD.
* To perform high availability.
* To generates the rule to execute.
* To maintain log files, because for each log file has two separate files.

Research questions

1. When analyze millions of records in the database what are the methods use to perform that task and how it involve to the system ?
2. What are the problems that occur when analyze data since the database is receiving millions of records per day ?
3. How the log files affect when select the suitable customer groups to send SMS ?
4. If the main server crash how it affect to the system and how it recover ?
5. When send SMS what are the sections and sub sections that consider to categorize customers ?
6. What are the limitations when take brand companies requirements ?
7. What are the supporting databases when analyze data and how the selected database involve with the system ?
8. What is the model that take to carry on the project and what are the limitations if select another model ?

**Chapter 03**

**Methodology**

Project overview

The proposed project is mainly focus on handling and analyzing the millions of records which receive from mobile customers to mobile companies. The system keeps all records which come from customers then analyze them and according to brand company requirement select appropriate customers and inform the message by SMS.

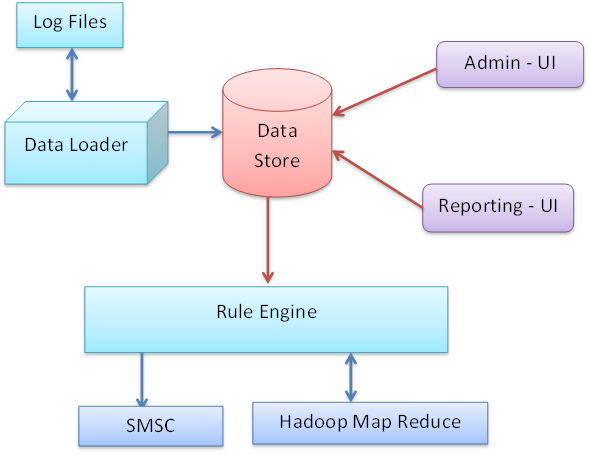
In this world most important fact is communication. To do that now uses many methods such as mobile phones, internet and so on. Here it mainly focuses on mobile phones. In mobile phones there are many features such as SMS, voice calls, LBS, GPRS, and USSD. Mobile companies they use log files to record customer information regarding these. It has customer’s phone number, receiver’s phone number, location and so on. As an example if someone gets a voice call to someone that details record in voice call log. Same as the other log files also.

Suppose any other brand company such as odel, Pizza hut need to inform their customers that they have a promotion or some other special event. So they tell it to the mobile company under some rules. As an example if the pizza company needs to inform about their promotion to customers who are currently living in Colombo area, with age limitation, at least the customer visit that place at one time and so on. So the mobile company has to consider these requests, generate the rule according to them and send it to database. Then analyze records according to generated rule and then select the matching records. Then they send the messages to customers through SMS. In existing systems which are in mobile companies they only consider about only one log file to do this task. But here it considers about five items to analyze data.

In this case it considers all log files and then selects the items to send SMS. Here it takes all records from log files and send them to database. There are two files for each item. In one it has customer records. For that it has given some time period. After that time period it checks the file and if have records it send to the database and those records send to another file which belong to that item.This happens for all log files. Here it has millions of records in log files for a day. In special days such as New Year, year end, Christmas day the record count will be increased. So here the research part is **handling millions of records**. It is not an easy task. So to deal with records the server has to up and running all the time.

Here it has separate user interface for administration and for generate reports. In admin level here it defines the rules according to customer requirements under some categories such as promotions, loyalty programs, tenure program. Then it selects the relevant records to provide SMS. In report generation part it uses to know about the summary. As an example assumes the brand company needs to know to which customers the mobile company sends SMS. Then the system has to provide the report with necessary information.

Assume server has crashed, then it sends a message to the person who is responsible for the system at that time, because the server has to up and running all the time since this relate with customer records.



**Figure 1**

My contribution

My contribution towards the project will be the development of Graphical User Interface to take the requirements from brand companies and generates the rules for the system. Then the rule sends for execution. As an example if the pizza company needs to inform about their promotion to customers they tell it to mobile company with some requirements such as who are currently living in Colombo area, age limitation (eg: age between twenty to twenty five) , at least the customer visit that place at one time and so on. So the mobile company has to consider these requests and insert them to the system. So the system willgenerate the rule according to these requirements and send them for execution to the database. Execution may be done in next day. Then the system sends the messages to customers through SMS. In Graphical User Interface have some options and sub options such as,

* Promotions –
* LBS (Location Base System) promotions. Here according to the customer’s current location and by some other information such as age limitation, it includes the details to the system. Then it generates the rule for it.
* Loyalty programs –
* Here it selects the customer who reloads the maximum amount. Then the system announces him as the winner and allocates the price for him. For that it generates a rule.
* Here it concerns the star points also to select customers to generate the rule.
* Tenure programs –
* Here it concerns the first customer who sends SMS to someone, as an example it selects the first customer who sends SMS to someone in New Year. To do that purpose the system generates a rule.
* Here it selects also the customer who sends the maximum SMS amount within given period of time. To select it by the database it generates a rule.

All these work has done by administrator. So I have to give access permission for his. For that the administrator has to register to the system, then he has to log to the system each and every time when he visits the system for security purpose. Also here has to handle some basic functions such as change password.

So the objective of my contribution is creating admin functions and generates rules for the appropriate customer groups by given requirements to execute.

The tasks to achieve my objectives,

* Has to create Graphical User Interface for administrator to register and log to the system.
* Has to create Graphical User Interface to take options - For this need to know requirements that brand companies expect.
* Has to generate rules according to brand company requirement to execute and send them to database.



Administrator

Administrator Registration

Log to the System

**Options**

* Promotions
* Loyalty programs
* Tenure programs

Take requirements

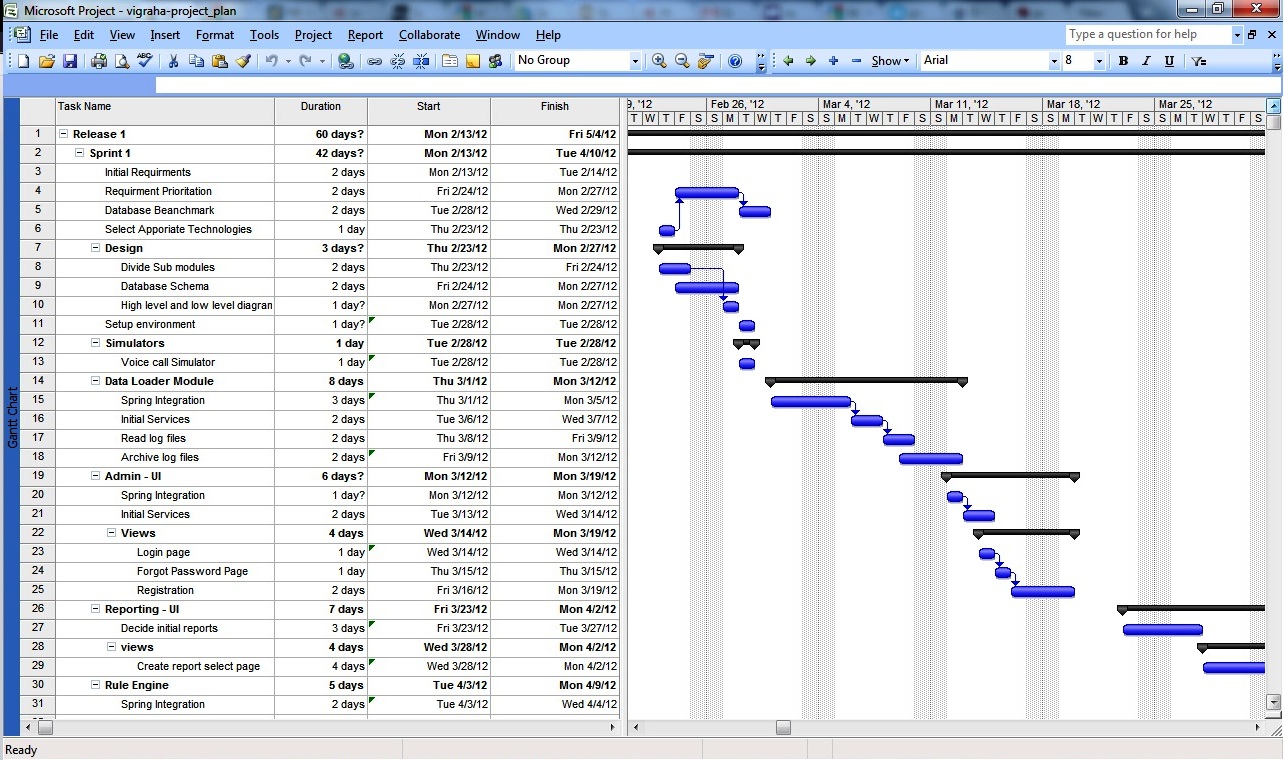
Generate rule

Database

**Figure 2**

Carry out the project

All four members take clear idea about the system and the system divided into four major components to carry out the project within one year. First we gather all information to develop the system. For that do back ground reading and do literature reviews. Then according to given time period we develop the work brake structure to get clear idea about the proposed project. Then according to the plan we build our components in parallel way.

****

Work Breakdown Structure

**Vigraha High Volume CDR Analyzer**

Analyzer

Administrator UI

Maintain log files

Report UI

Create stimulators

Send records to database

Register and log administrator

Get requirements

Generate rules

Execute rule

Analyze records

Send SMS

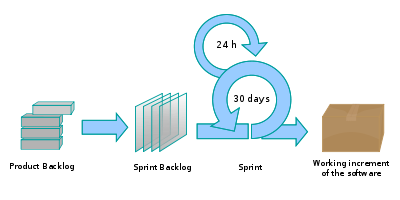
Get requirements

Generate report

**Figure 4**

According to my contribution first I am planning to be familiar with the technology we are using. Then start to create the Graphical User Interface for administrator and then for the options. Later on build the coding for the system including rule generation.

Here we are using agile methodology. Under this methodology it contains many sub categories. From those we are selecting scrum method.Scrum is an agile approach to software development. Rather than a full process or methodology, it is a framework. So instead of providing complete, detailed descriptions of how everything is to be done on the project, much is left up to the software development team. This is done because the team will know best how to solve the problem they are presented. This is why, for example, a sprint planning meeting is described in terms of the desired outcome [3].In this method we are developing the system and after within two weeks or one month we show it to the client to take feedbacks. So if the created part has some problems he can inform that at the moment. Then we able to fix up the errors and then continue the rest of the parts. Likewise it will happen throughout the system. The relevant time period known as sprint or iteration. Here we do not use any presentations to describe our contribution of the system to the client. Instead of that we present the actual system to get feedbacks and then continue to develop the same system. But if we are using some other methods such as spiral model, waterfall model, prototyping model they consist of many problems. Sometimes able to repeat the things what we have already done or may be those models cause to waste resources. But scrum model solve these entire problem, because time to time it takes feedback from the client.

[](http://en.wikipedia.org/wiki/File:Scrum_process.svg)

The scrum process

**Figure 5**

**Initial User Stories**

In scrum entire system need to describe with user stories. Following are the user stories what we identified.

1. Read log files
2. Archive Log files
3. Summery table uploading schedules
4. Spring Integration
5. Initial Services
6. CAS integration with Single Sign On
7. Open ID Integration
8. Create rules based on programs
9. Search rule
10. View and edit created rules
11. Data analyze with map reduce
12. Configure properties
13. Initial Reports
14. Identify Drill down reports and charts
15. Identify necessary indexes
16. Spring Birt integration
17. Test cases
18. Concurrent access
19. Security
20. SSL for User interfaces

Here are the initial user stories. After Identify user stories we need to estimate the above stories.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| User Stories | | | | | | | | | | | | | | | | | | | | |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Rajith | 3 | 3 | 5 | 8 | 13 | 8 | 5 | 3 | 3 | 3 | 13 | 5 | 8 | 5 | 3 | 13 | 5 | 0 | 0 | 0 |
| Lasantha | 3 | 3 | 3 | 8 | 13 | 5 | 8 | 5 | 3 | 3 | 13 | 5 | 5 | 5 | 5 | 13 | 5 | 0 | 0 | 0 |
| Thejani | 3 | 3 | 3 | 13 | 13 | 8 | 8 | 8 | 3 | 3 | 13 | 5 | 8 | 5 | 3 | 13 | 5 | 0 | 0 | 0 |
| Naveen | 3 | 3 | 5 | 8 | 13 | 8 | 8 | 5 | 3 | 3 | 13 | 5 | 8 | 5 | 3 | 13 | 5 | 0 | 0 | 0 |
| **Final Extimation** | 3 | 3 | 5 | 8 | 13 | 8 | 8 | 5 | 3 | 3 | 13 | 5 | 8 | 5 | 3 | 13 | 5 | 0 | 0 | 0 |

Software Requirements for Development

* Net Beans – Here it uses Net Beans 7.0.1 version. The system develops by using java language.
* My SQL

Graphics

* Microsoft project2007 - This use to develop Gann chart.
* Corel Draw 12 – Use to draw all diagrams of the report.
* Photoshop - To design the interfaces of the system

Technology

* To analyze data use hadoop integration. It has to customize according to our system.

Methods

* Agile model – under this use scrum model and sprint.

How collect data?

To categorize customer records base on log files it requires the specification from mobile company, because each log file contains specific customer information such as mobile number of the sender, mobile number of the receiver, the contact location and so on. So we collected it from a mobile company. According to their specification it categorized the mobile users in voice calls, SMS, GPS, LBS and USSD. To do that they used log files. Here we selected five categories to gather customer records but existing systems they use only one log file to do this task. The company did not provide the customer records since if they provide them it will be a security problem. So we have to simulate those records. For that we use Perl language. For all five log files we have to create simulates.Also to analyze records has to have a method. So the system uses **apache hadoop map reduce programme** to do that task. We have to take this and customize according to our system. **Apache hadoop** is a platform for analyzing large data sets that consists of a high-level language for expressing data analysis programs, coupled with infrastructure for evaluating these programs [4].

Anticipated conclusions

In real world the project is more useful for mobile companies. It is use for them, because of brand companies. According to the brand company’s requirements the actions of the system will perform by mobile companies. All of them provide services to mobile users. So if consider the real world mainly three parties benefited from this system. The major task of this system is handling and analyzing customer records. It receives millions of records per day. Except the main result providing from the system it will generates reports for mobile companies, brand companies and sends SMS for necessary mobile users.

**Description of Personal and Facilities**

Considering each fact of the project, distribute the responsibilities to each member. For each member should be allocated with similar weighted tasks. To accomplish this requirement the project split on to components. So each member assigns with each component and has to complete theircomponents from the beginning to the end. Then at the end all four components should be merged together to perform the final outcome.

According to the project plan, the components of each group member can be categorized as follows.

|  |  |
| --- | --- |
| Name of the member | Tasks |
| Naveen | * Has to create stimulators for log files * Develop the database to store daily customer details from stimulators. * Has to maintain two files to each log file. * To collect records. * After sending records to database   after some time period same  records send to second file. |
| ThejaniDineshika | * Develop Graphical User Interface for administrator. * To register to the system. * To log to the system * To change pass word * Develop Graphical User Interface for the options. * Promotions. * Loyalty programs. * Tenure program. * Generate the rules. |
| RajithDelantha | * Analyze the customer records by hadoop integration. * Execute the rules when necessary time. * At same time when the rule generates may not need to execute it. It may be done in next day. |
| LasanthaPerera | * Develop to generate reports * If the brand companies or mobile company need the current reports the system able to provide them. * The system able to provide history reports when necessary. Here uses the drill down reports.   To generate reports has to filter them according to the requirements. Has to check whether the SMS receive to the mobile users. |

**Budget**

# References

# [1]” Managing log files for BlackBerry Mobile Voice System” , 10-06-2011 , <http://btsc.webapps.blackberry.com/btsc/search.do?cmd=displayKC&docType=kc&externalId=KB17955>

# [2] “[Scrum Methodology](http://scrummethodology.com/)” , 10-10-2011, <http://scrummethodology.com/scrum-sprint/>

# [3] “Introduction to Scrum – An Agile process” , 13-12-2011 , <http://www.mountaingoatsoftware.com/topics/scrum>

# [4] “Welcome to Apache Hadoop!”, 22-01-2010 , <http://pig.apache.org/>

Appendices

# *Appendix A:* *List of Acronyms and Abbreviations*

CDR – Caller Detail Record

SMS - Short Message Service

LBS - Location Base System

GPRS - General Packet Radio Services

USSD - Unstructured Supplementary Service Data

WBS – Work Break Structure

MVS - Mobile Voice System

# *Appendix B: Diagrams and Figures*

**Figure 1**

This represent how the proposed project is carry on. It describes how each component involves with the system and it shows the diagram for entire system. This is the high level diagram.

* It collects customer records from log files to the database. Here it shows only one log file and its two files. Likewise it has five log files and each log file contains two files.
* When gather requirements it generates the rules.
* Then according to the rule it executes to select matching customers from database to send SMS.
* Then print the reports.

**Figure 2**

Here it describes about my contribution of the project and how it carry on.

* Register the administrator.
* Log to the system.
* Enter requirements to options.
* Generate the rule.
* Send to database for execution.

**Figure 3**

Here it is the plan of the system for one year. It contains all activities regarding with all four members the proposed system and the time period allocate for them.

**Figure 4**

The Work Breakdown Structure represents the scope of the system. It shows the major contributions of the system and the functions of them. Under those functions it represents the activities relate with each component.

**Figure 5**

The scrum process

This regards with agile methodology. Under this the proposed project is using scrum model. Here develop the system and after two or four weeks it represents to the client. Then get the feedback from the client. This is known as sprint or iteration. Then if has any deviations at the same stage fix them and develop the same system further. This process happensthroughout the system.