Big Data Analytics Framework for Agriculture

Bradwin Danai Zinyoni

(R141458A)

Big Data Analytics Framework For Agriculture

By

Bradwin D Zinyoni

Submitted in partial fulfilment of the requirements for the degree of

BSc Honours Computer Science

Department of Computer Science and Information Systems

in the

Faculty of Science and Technology

at the

Midlands State University

Gweru

October 2017

Supervisor: T G Rebanowako

Abstract

Farming is undergoing a digital revolution. Farmers are gathering information passively collected by precision agricultural equipment and manually and many farmers are using information from large datasets and precision analytics to make on-farm decisions. Big data includes extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions. The use of large information sets and the digital tools for collecting, aggregating and analysing them together is referred to as big data. Compare a notebook wherein a farmer might log information about his or her crop performance with a computer used to predict and direct future production practices. Logging information using the application can be done more efficiently and the volume of information the farmer may access using profound agricultural management tools provides access to interacting with datasets that stretch way beyond the individual farm. The analysis was done successfully. Therefore, from the analysis the researcher proposed development of a big data analytics framework for agriculture that enables the farmers to assess and to predict the outcomes of the crops before they grow them by using the historical information. A detailed feasibility study was carried out and it resulted feasible to design the system and an in-house development solution was recommended. Various designing tools have been used which includes MYSQL and PHP servers. The system allows the farm worker to record the farm activities in order to be able to use that data to access and to analyse the crops behaviour. The system was successfully implemented and parallel changeover was the recommended changeover strategy due to its many advantages over other strategies. Maintenance was carried out using perfective maintenance strategy which allows for continual improvement of the system. It’s the view and aspirations of the researcher to have the system integrating the training modules which manages recommended training schedules in a bid to continuously cope with changing technological environment.

Declaration

I **Bradwin Danai Zinyoni** do hereby declare that I am the sole author of this dissertation; I authorize the Midlands State University to lend this dissertation to other institutions or individuals for the purpose of scholarly research.

Signature…………………………. Date …………………………

Approval

This dissertation entitled "**Big data analytics framework for agricultural services"** system by **Bradwin Danai Zinyoni** meets the regulations governing the award of the degree of BSc Computer Science Honours of the Midlands State University, and was approved for its contribution to knowledge and literary presentation.

Supervisor………………………………. Date ………………………………

Acknowledgements

Firstly, I would like to thank the almighty God for the source of strength that he has always given me throughout my research. I would like to express my profound acknowledgement to the following people: The project supervisor Mr T G Rebanowako who contributed immensely throughout the entire project development process. I would also like to extend my sincere gratitude to I.T director for his support during the research study.

Lastly, l would like to thank Higher life Foundation and my brother Blanden Zinyoni for their love and financial support. I could have never done it without them. I can safely say," May God continue to bless you".

Dedication

I dedicate this dissertation my late parents Mrs Zinyoni and Mr Zinyoni, my elder brother Blanden and my younger brother Budwel as way to thank you for such a lovely and incredible family. Though my source of inspiration is deceased with the grace of the Lord I decree that I am guided and I am not alone through this journey. May the good Lord always remember you.

Table of Contents

List of Acronyms

ANSI-SPARC……. American National Standards Institute Standards Planning and Requirements Committee
DBMS……………. Database Management System
DFD………………. Data Flow Diagram
GUI………………. Graphical User Interface
ICT………………...Information communication Technology
NVP……………….Net present value
ROI………………...Return on Investment
SQL………………. Structured Query Language

List of Tables

List of Figures

List of Appendices

Chapter 1: Introduction

1.1 Introduction

Big data analytics framework for agricultural services system is a solution that enables farmers and individuals to solve problems, track and trace, predict unprecedented trend in agriculture sector by using the quantity of generated and data stored from past decades and current information during daily operations.

1.2 Problem definition

Clifton (1977) defined problem definition as the process of picking the apparent gap between the current state and the desired state. Presently most farmers in Zimbabwe are going about their business without having the ability to track and to trace the behaviour and trends in agricultural sector. They face some difficulties when they try to predict and to solve the current problems that had occurred and solved some time ago.

- Farmers are not able to predict the trend in the agricultural sector because they don't have the historical databases or they have data but not structured in way that enabled them to use that data to solve current problems and to provide information
- Some of the farmers are not able to assess the behaviour of species and crops before they grow them and to plan ahead of the season because they really don't adequate information
- Some farming knowledge is not available to the farmers due to the fact that there is no big data systems available.
- Farmers use the farm diaries to record their daily activities each and every year, due to the use of manual system the farmer will be not be able to use all those diaries to correct his/her past season's mistakes.

1.3 Aims

Implementation of big data analytics framework for agricultural services.

1.4 Objectives

Hoffer (2006) explained an objective as measurable, specific result that a project target to accomplish within a space of time and with the resources available. These are basic tools that inspire all the strategic activities and all planning and they serve as the foundation for constructing the policies and evaluation of performance.

- To enable to assess the best crops to grow by analysing their farm conditions and the crops conditions.
- To enable farmers to diagnose the crop diseases and be able to deal with crop diseases by using recommended way or chemicals.
- To enable farmers to maximise productivity by comparing assessment before they grow crops and the yields after harvesting
- To enable companies and individual farmers to get the approved and verified information to use for farming
- It enables farmers to use the history for crop rotation.

1.5 Hypothesis

This system will be developed by the following tools:

- PHP for good Web interface.
- Office 2016 for editing and writing text.
- MySQL database is one of the open source relational database management system available on the market with many advanced features. MySQL can work with different languages, it is easy to use and to install and it is interoperable.
- Maven- project management and build tool which provides the platform for unit test, modular programming and loose coupling.

1.6 Justification

The development of the big data analytics framework for agriculture enhances the farmers and individual to be able to trace and track the trend of farming by using the stored data to solve the problems. Through capturing the farm details and daily activities data, it will be easy to analyse the trends and to predict the outcomes. The following is the rationale for the study

- Proposed framework will support online service to all the users to enable them to access the system anywhere
- It enables to track and trace unprecedented trend in agricultural sector by using the quantity of generated and data stored
- It enables companies and individual farmers to get the approved and verified information to use for farming.

- It enables farmers to assess the best crops to grow by analysing their farm conditions and favourable crop conditions.
- It enables farmers to use the history for crop rotation.
- It enables farmers to diagnose the crop diseases and be able to deal with crop diseases by using recommended way or chemicals.

1.7 Conclusion

The project objectives and the justification has been clearly listed and understood for necessary managerial purposes and objectives pitched from problem defined. Since the problem is well defined we are now proceeding to analysis and designing of the system.

Chapter 2: Planning Phase

2.1 Introduction

The planning stage is important in the development of software project, the work plan of the project will be drafted to result with the task starting dates and fining dates, the business values will be listed out to see the real value of the project and the feasibility study will be taken into consideration to see if the software project is really economical, technical and operationally feasible (Swanson, 2004).

2.2 Business value

Bennatan (1995) defined business value as the principles and the standards that shows the channel in which the system must perform, mean and what makes it valuable to the business. It also refers to the overall functioning way of the system which gives gain to the business.

The system will give the following to the business.

- Farmers will be able to predict the trend in the agricultural sector because they don't have the historical databases or they have data but not structured in way that enabled them to use that data to solve current problems and to provide information
- farmers will able to assess the behaviour of species and crops before they grow them and to plan ahead of the season.
- Farmer will be able to access the adequate information based on farm location.
- Farmers will be able to use computerised system to record their daily activities other than to use farm diaries to record their daily activities each and every year and the will use the system to correct his/her past season's mistakes.

2.3 Feasibility Study

According to Clifton (1997) the main reason to conduct the feasibility study is to find out whether the new system is reasonable and worth to use given the schedule and the available budget. The feasibility study takes into consideration the business requirements, the system description and the way in which the system supports the processes of the business.

Purpose of the feasibility study is:

- To find out whether the software project is technically, socially, operationally and economically feasible.
- To gather the specifications
- To notice the weakness and the strengths of the system.
- To determine whether the advantages outweigh costs.

2.3.1 Technical Feasibility

Kendell (2002) explained technical expertise as the ability to build the system given the available resources and time. The researcher has to consider the following in the feasibility study

- Hardware needed
- Software needed
- technical expertise

2.3.1.1 Hardware needed

The system will be operated on certain environment, in order to set up the environment the following hardware will be required:

Table 2.1 Database Server

Unit	Minimum Specifications	Recommended Specifications	Available
Disk drive	250GB SSD	500GB SSD	500GB HHD
Printer	Laser Printer	Laser Printer	Laser printer

Table 2.2 User Computer Specifications

Processor	2GHz	2.8GHz Intel P4	2.3GHz
Memory	4G	8G	6GHz
HDD	60 Gb	600 Gb	500 Gb
Network Cards	10/100	10/100	None

Table 2.3 Network Requirements

HUB	32 Port	32 Port	16 Port
Patch Panel	32 Port	32 port	6 port
	UTP CAT 35 Fly	UTP CAT 35 Fly	UTP CAT 35 Fly
Cables	leads patch codes	leads patch codes	leads patch codes
UPS	3u	3u	3u

2.3.1.2 Software Specifications

In order to develop this system, so many softwares are needed but most of the are open source softwares which means there no cost incurred to acquire licences. The crucial software required to develop this system are listed in the table below:

Table 2.4 Software specifications

Windows Operating System	10
Adobe Phantom	4
UC browser	6.4
Anti-virus	Nicol
Microsoft Office	2016

2.3.1.3 Technical Expertise

- **Developers expertise**

The software developers of the system are well skilled and they really work to brought the system on time. All the programmers are certified with well-known boards.

- **Workers**

All the workers are able to use the computer system and to solve tasks using computers, and they are able to adapt the new technology and system and times goes on. The Farmer will be given the manual and they will be invited to the workshop so that they will be able to use the system.

2.3.1.4 Conclusion on Technical Feasibility Study

The study was done very well by taking into account all the hardware and software needed, affordability and availability to develop the system. Since most of the softwares needed to develop the system are open source, this helps the system developer to finish tasks on time since the resources will be available on time.

2.3.2 Economic Feasibility

According to Jeffrey (2001) economic feasibility is the profit test to find out whether the system will give back enough, less or no profit. The current system will be compared to see if there is benefit of the new system in monetary value, if the new system gives enough profit, it will be taken as the best solution. Economic feasibility takes into consideration the following:

- Cost benefit analysis
- Return on investment
- Period of Payback
- Net book value

2.3.2.1 Cost benefit analysis

Whitten and Bentley (2008) explained the cost benefit analysis as the tool used to analyse the benefits and the cost of the proposals. The cost benefit analysis consists of the following two steps:

- Estimation and identification all the benefits and the costs of carrying out the project from start to the end, the operating costs, the cost of development and all other small costs.
- Cost and the benefits need to be expressed in common units, evaluation of the net benefits.

Each cost and benefit are expressed in monetary form to establish whether the system was feasible.

2.3.2.2 Cost

Steward (1987) explained cost as the charges that are related to development and operational of the system. These costs are categorised according to their origination in the system development life cycle and these costs are identified and quantified in the approximate monetary value. The cost is grouped into either developmental cost or operational costs. The development cost consists of the following:

- Cost of user training
- Setup and equipment cost

- Expert hiring

The cost incurred during the system development only and are being estimated at the onset of the system development and should be refined on each and every project phase.

Operational cost is categorised as fixed or variable and these costs incurred during the development of the big data project.

- Fixed Costs: these costs occur at regular intervals and at rates which are relatively fixed.
- Variable costs: These are cost that occur depending on the usage factor. Good example is supplies such as DVDs and USB flash drives.

2.3.2.3 Benefits

Benefits come along with the system were grouped as either tactical benefits or strategic benefits.

- Tactical benefits are the long-term benefits that are realised by the management team. These benefits can help to improve the working environment or socially.
- Strategic benefits that help system to perform better at a lower cost and usually are regarded as intangible since they are not seen clearly by anyone

Steward (1987) grouped benefits as tangible and intangible. The tangible benefits are the benefits or costs that can be measured. These costs are known in the project development and they can be estimated accurately. Intangible benefits are not easily quantified for example improvement of the customers service through the use of the system. During the evaluation process the intangible and tangible benefits must be taken into consideration. Some of the benefits that can be noticed after implementation of the project includes the following:

2.3.2.4 Tangible benefits

These benefits are hard to estimate but the farmers will enjoy these benefits in the future.

- Reduction of the operations
- Easy prediction
- Process automation

2.3.2.5 Intangible benefits

- Time saving
- Asset utilization
- Proper farming

Table 2.5 Development cost

Unit	Cash($)
Hp Pavilion 600 series	1300
Hp z book 400b x3	2100
Sum	**3400**

Table 2.6 Operational costs

Activity	Year 1 ($)	Year 2($)	Year 3($)	Sum($)
System implementation	400	-	-	400
Farmer training session	1000	200	200	1400
System upgrade and adaptation	200	300	300	700
Sum	**1600**	**500**	**500**	**2500**

Table 2.7 Tangible benefits

Unit	Cash($)
Reduced cost of Operation	900
Improved quality of Data	900
Time saving and Reports Consistent	900
Ease prediction	900
Sum	**3600**

Table 2.8 Cost benefit analysis

Cost and Benefits	Estimated Value ($)	Total ($)
Tangible Benefits	3600	
Intangible benefits	4000	
Total benefits		7600
Development costs	3400	
Operational costs	2500	
Total costs		(5900)
Net Benefits		1700

2.3.2.6 Return on investment

Bennatan (1995) explained R.O.I is a calculation technique used to compare net profit on the investment. The R.O.I technique was used to estimate the viability of the system. The following formulae was used to calculate R.O.I. and as expressed as %.

$$\text{R.O. I} = \frac{\text{Net Benefits X 100}}{\text{Total Costs}}$$

$$= \frac{1700 \text{ x } 100\%}{5900}$$

$$= 28.813\%$$

2.3.2.7 Conclusion: Economic feasibility

By considering the net benefit as well as the favourable R.O.1 which was 28.813% the project is economically feasible, therefore the project can procced.

2.3.3 Social Feasibility

Hoffer et al (2006) defines the social feasibility study as the way in which the project benefits the society and other people at large other than the organization and.

- New system will unite the workers and its branches through web communication.

- The new system will enable workers to learn new technology during guide training
- The people in the community will be able to communicate with the organisation using customer care on website.

2.3.4 Operational Feasibility

Bentley et al (2006) defined Operational feasibility as the assessment on how the new system can tackle the challenges and fits the existing business environment and how it can affect the existing business processes. The researcher found the following during the feasibility study:

- Farmer were able to use the system easily which means the system is user friendly
- The system can run of many devices with different operating system
- The user interface is clear and there is a help menu.

2.3.4.1 Conclusion on operational feasibility

The big data system is operationally feasible, the farmers are able to use the system without reading the manual and the system can run on different devices.

2.3.4 Conclusion on Feasibility Study

Feasibility study was carried out successfully by taking not of the economic feasibility, technical feasibility and the operational feasibility ,the proposed system is feasible.

2.4 Risk Analysis

According to Bennatan (1999) risk analysis is a technique used to assess and to pick several factors which may hinder the success of the system. The risk analysis takes into consideration about threats vulnerable to the completion of the project. Risk analysis consists of identifying the possible risks and the possible solution to avoid those risks.

2.4.1 Technical risks

The technical risk analysis is done in the project development in order to produce high quality products and to avoid poor outcomes. The technical risk consists of the following:

- Giving task to the personnel with appropriate skills to avoid poor software.
- Formation of teams with people with different skills to allow modular programming, experts will be doing hard tasks while beginners will be doing simpler tasks.
- Be able to track and trace the error during the development of a software.

The writer has skills for developing the big data system.

2.4.2 Risks related to Software development process

During the software development process, the engineers may face several threats that can hinder the successful of the project if they do not take of those development risk. The following checklist and guidelines help to notice the software development related threats and the possible counter measures:

- The writer gathers and ensure that all the processes are planned and well documented to avoid confusion.
- Ensure that the stockholders are following the planned and documented process.
- Ensure that each and every stage is tested before proceeding to the next stage.
- To ensure that all the software bugs are successfully removed and to avoid the same problem to occur during development.

2.4.3 Tight Schedule

Most of the software projects fail due to the tight schedules which result in poor quality product. To avoid that risk the writer planned the work schedule and resource schedule in a way that no delays will take place in order to meet all the deadlines. The writer worked so hard, sometimes the writer over the night in order to deliver high quality product on time.

2.5 Work Plan

According to Rajaraman (2004) defined work plan as the outline of the small and major tasks to be done in the software development process. The work plan consists of the team members and the corresponding task the member should perform. The work plan consists of the following:

- The list of the members and their corresponding task to perform in the project.
- The equipment list needed to complete the project.
- The project breakdown into tasks in their completion order.
- The schedule showing the activities with their start date, duration and end date.

2.5.1 Work Breakdown

The work breakdown groups and it defines the project’s tasks and its elements in a manner that helps to define and help to organise the work to be done in the project.

Table 2.9 Work Breakdown

Task	Start Date	End Date	Duration
Project Introduction	18/06/17	25/06/17	1 week
Project Planning	25/06/17	09/07/17	2 weeks
System Analysis	09/07/17	16/07/17	1 week
System Design	16/07/17	30/07/17	2 weeks
System Maintenance	30/08/17	06/10/17	5 weeks
Project Documentation	18/06/17	29/09/17	10 weeks

2.5.2 Gantt Chart

The tasks and activities in the project management is commonly shown by using the grant chart. The grant chart consists of the activities on the left side and on the right side consist of the time in weeks or in days and there is a coloured bar which is used to show the duration (start date and end date). The writer used the Gantt chart to take note of the following:

- The list of the activities and their corresponding duration.
- The order of each and every activity during the project development.
- The project's start and end date.

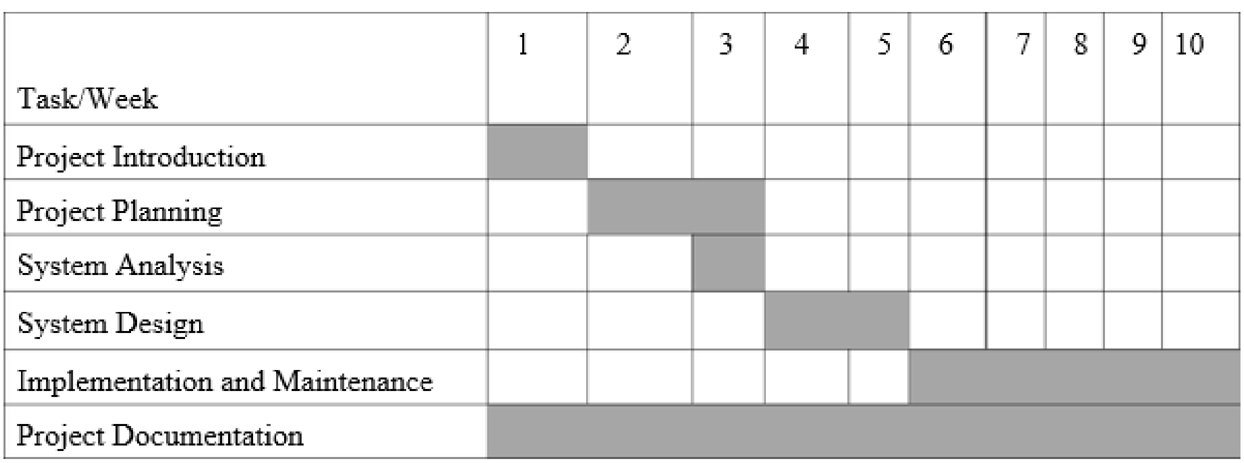


Fig 2.1 Gantt chart

2.6 Conclusion

The big data project was successfully approved, the schedule, each and every task established, the project work plan was laid down and the writer managed to complete the analysis of the existing system. Planning phase was done well and all the human and the financial resources allocation was well stated. The planning phase was completed well in two weeks and the writer is now proceeding to the analysis stage.

Chapter 3: Analysis Phase

3.1 Introduction

Rajaraman (2004) explained system analysis as the process of gathering the factual data and noting the problems within the processes and give the best suggestions that improves the performance and functioning of the system. During the analysis phase the operational data will be gathered taking note of the business processes and understanding the business flow, looking for the systems weaknesses and the possible solutions to solve them in order to achieve the project goals. During the system analysis stage, the main objective is to:

- Study and take note on how the current system functions
- Finding the ways to improve the current system.
- Develop the new system’s business model or logical model

3.2 Literature Review

“Big data includes extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions. Farming is undergoing a digital revolution. For example, even small-scale farmers are gathering information passively collected by precision agricultural equipment, and many farmers are using information from large datasets and precision analytics to make on-farm decisions. The use of large information sets and the digital tools for collecting, aggregating and analysing them together is referred to as Big Data” Wolfert(2016)

“Arguably, farming has been empirically driven for over a century but the data collected was not digital. Big Data is different from this historic information gathering in terms of the volume and the analytical potential embedded in contemporary digital technologies. Big Data proponents promise a level of precision, information storage, processing and analysis that was previously impossible due to technological limitations “ Datafloq (2015). Compare a notebook wherein a farmer might log information about his or her crop performance with a digital phone ‘app’ used to predict and direct future production practices. Logging information using the application can be done more efficiently (even by voice recognition) and the volume of information the farmer may access using profound: agricultural management tools provides access to interacting with datasets that stretch way beyond the individual farm. Systematically tracing the digital revolution in agriculture as well as the limitations of Big Data-driven decisions should be a broad research goal for Big Data scholarship in the realm of food and agriculture.

3.3 Weaknesses of the existing systems

- Farmers ae not able to user the stored information to assess the growth of crops before growing them.
- Farmers use the farm diaries to capture and store their information, they find difficulties to trace and track the tends of farming.
- Data inconsistency: Data is misplaced during capturing and filing hence the information is prone to errors.
- Information is lost due to excess paper work.

3.4 Rationale of the proposed system

"Big data analytics framework for agricultural services system is a solution that enables farmers and individuals to solve current problems, track and trace, predict unprecedented trend in agricultural sector and the food industry by using the quantity of generated and data stored from past decades and current information stored daily. Though the use of big data farmers will be able to predict outcomes for agricultural sector by analysing the trend using stored information and they will be able to get the approved and verified information to use for farming. The big data system will allow the farmers access to information that can help them decide, for example, where to plant crops and to solve the farming problems using the volumes of data stored in the system." Bogaardt (2016)

3.5 Evaluation of Alternatives

The viability of the project was stated in the planning stage under the feasibility study and it was proven that the project is economically, technically and socially feasible. The writer takes into consideration the different available before choosing the best alternative to achieve optimal results. The following are the alternative which are available:

- Current system improvement.
- Outsourcing
- In house Development

3.5.1 Outsourcing

The outsourcing option consists of buying the complete software package or give the contract to external company to develop the system for the company (William, 1998). With reference to the feasibility study, the resources that are needed to develop the big data system were said to be available hence there is no need for outsourcing. The range of software and hardware that is needed is readily available on place. The big data system should not be out sourced due to the following reasons:

- The outsourced package may not meet the required standards.
- The cost of outsourcing is very high since there will be some overhead cost to hire the specialist to operate and to install the package.
- The maintenance cost will be high since the specialist will be hired.
- May not be possible to integrate the package with the existing systems, since the application programming interface(API) will be not available.
- The resources needed to develop the big data system is readily available.

3.5.2 Improvement of the current system

From the results of the feasibility study conducted earlier, current system improvement is the option which has less cost, the following are the demerits of improving the system:

- The current system interfaces do not meet global standards.
- The current system is not interoperable, it can only run on windows operating system.
- The backend of the system was developed using Fotran language which is not commonly used these days.

Based on the above outlined reasons it is not the optimal solution.

3.5.3 In-house Development

The in-house developers can use the requirements specifications to create the big data system. The big data system to be developed should solve the data related problems with the agricultural sector. The writer has the capabilities required to create the big data system. The following reasons address the question why the big data system was developed within the firm:

- All the requirements that are needed are available with the form.

- It requires less money to install and maintain the system.
- All the firm's short term and long-term goals are being meet.
- It becomes easy to integrate the big data system and other payment gate ways since the in-house developers will be able to provide APIs
- The corrective maintenance and continuous testing will be done easily since the developers will be always available in house.

3.5.4 Recommended alternative

By comparing the advantages and disadvantages of the alternatives, the in-house development was recommended by the writer.

3.6 Requirements Analysis

Wiegers and Karl E (2003) explained requirements analysis as the key factor that contribute to the successful of the software project development. The system requirements should be measurable, testable, well documented, traceable and stated in greater detail sufficient for the development of the system. The requirements analysis involves the following types of activities:

- Eliciting requirements-the different kinds of requirements from all the sources will be identified.
- Recording requirements: all the identified requirements are documented in different formats such as summary list, process specifications and other formats.
- Analysing requirements-the analyst have to determine that the system requirements are complete, straight forward, unambiguous and consistent.

The requirements analysis takes time to process and it also involves the delicate skills. The development of the new system changes the relationships with the work groups and the environment hence it is important to take note of the stakeholders needs and they should understand the new system's implications.

3.6.1 Functional Requirements

Kendell (2002) defined functional requirements as the division of the system features which consist of the raw data entering the system, all the system processes and the system's outputs which satisfy most of the system's objectives.

The following are the big data systems functional requirements:

- The system supports online service to all the users to enable them to access the system anywhere. The new farmers are being registered online.
- Assess the plants to crow ahead of the season.
- The big data system track and trace unprecedented trend in agricultural sector by using the quantity of generated and data stored
- Records daily activities.
- Companies and individual farmers get the approved and verified information to use for farming.
- The system predicts the outcomes for agricultural sector by analysing the trend using stored information
- farmers use the history for crop rotation.
- The farmers can generate seasonal reports.

3.6.1.1 Use Case Diagram

The use case technique is mostly used in the software projects for documenting the new system requirements or any software change. The individual use case shows scenarios that convey the way in which system should interact with other systems or the system end users to meet certain business goal (Sommerville, 1998). The use case diagrams are the simplest tool for showing the system behaviour. The use case diagrams external overview of the system and simply shows how the user perform specific task.

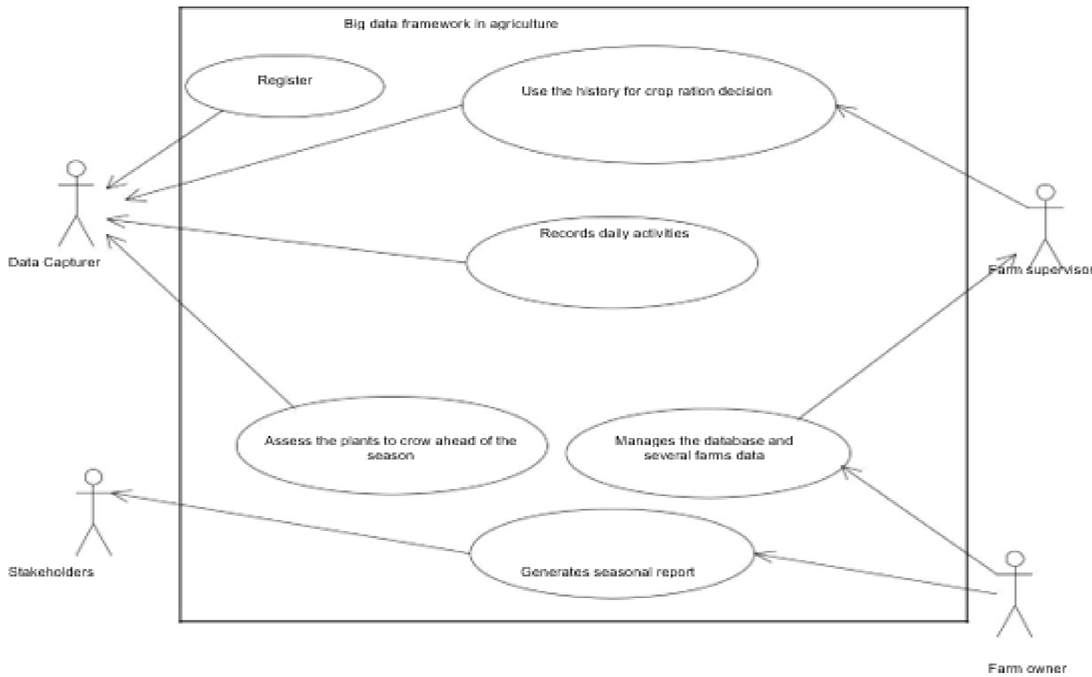
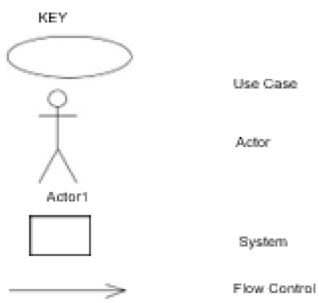


Fig 3.1 Big data system use case diagram



3.6.2 Non-functional requirements

The non-functional requirements specify the method used to judge the operation of the system unlike the specific behaviours. The non-functional requirements consist of the following:

- Efficiency: Shows how the system uses the resources: disk space, CPU cycles. The system requires powerful CPU in order to run smoothly.
- Availability: The system uptime is the time that the users will be able to use the system. The system will be not available during the backup, upgrade and update process.
- Interoperability: The ability of the system to run on different devices and various platforms. The big data system does not run mobile devices.
- Robustness: The system will be able to handle failure situations without failure. This includes the software defects, unexpected operating conditions and invalid data tolerance.

The non-functional requirements generally impose the constraints on implementation or the system design such as the quality standards, performance requirements.

3.7 Conclusion

The current system analysis was completed successfully and the writer managed to consider all the alternatives. In house development was chosen since developing a unique software package is the ideal solution for the problem and as a result, all the functional and non-functional requirements have been identified. The user requirements were clearly stated; therefore, the writer move over to the designing stage of the big data analysis.

Chapter 4: Design Phase

4.1 Introduction

Ulrich (2000) explained design stage as the art of laying out of the components of the system and the components' interrelationships in the right way to solve a stated problem. The functional requirements of the system are mapped in the design stage to the hardware and the software. In the design stage, the design plan is created of the projected development sequence through the rest of the system development life cycle process. Design phase concentrates on the manipulation of the maintainable and dependable system by considering the layout of the system and elementary design of the input and output forms interface.

4.2 System Design

Jones (1998) defined system design as the application process of different principles and techniques for the aim of defining the system in greater detail to allow its physical realisation. At this stage the system architectural design will be crafted. System design consists of different elements ranging from processing to entity communication (Steward, 1987). The system developers analyse and take note of the proposed system by taking into consideration the user requirements documentation and they find the methods and possibilities to implement them. If the user requirements are not clear, the system developers refer back to the users to find the resolution and to edit the document accordingly.

The following characteristics should be found in a well-designed system:

- **Security:** The system should be very secure in order to block hackers from access and tempering with the information and system should keep the user confidentiality and control the user access levels.
- **Efficiency:** The system's ability to execute all its functions at a faster rate without crushing which will enable the system users to process many tasks and reduce unnecessary delays.
- **Reliability:** The new system was proposed to solve the current problems being faced, the well-developed system should be able to reduce and eradicate the problems. The system should be able to restore the previous settings in case of crush.
- **User friendliness:** The system's graphical user interface allows the users to perform tasks with little or no supervision and support. The system user friendliness is a key factor in the

success of the software development since the system will be easy to operate hence the users will not take time to execute tasks.

- **Maintainability:** The system should be able to allow the changes and the modifications to align the user needs. If there is need to change the system from a single user to a multi user, system it should be able to allow that change.

4.2.1 Big data system context diagram

Context diagram is a pictorial depiction of a system which contains the data processes and data flows of system (Kendall, 2002). It shows a high-level design of a system, the common inputs, processes and the output links are shown. The context diagram below shows how the entities are related and how they are linked and the way data is exchanged.

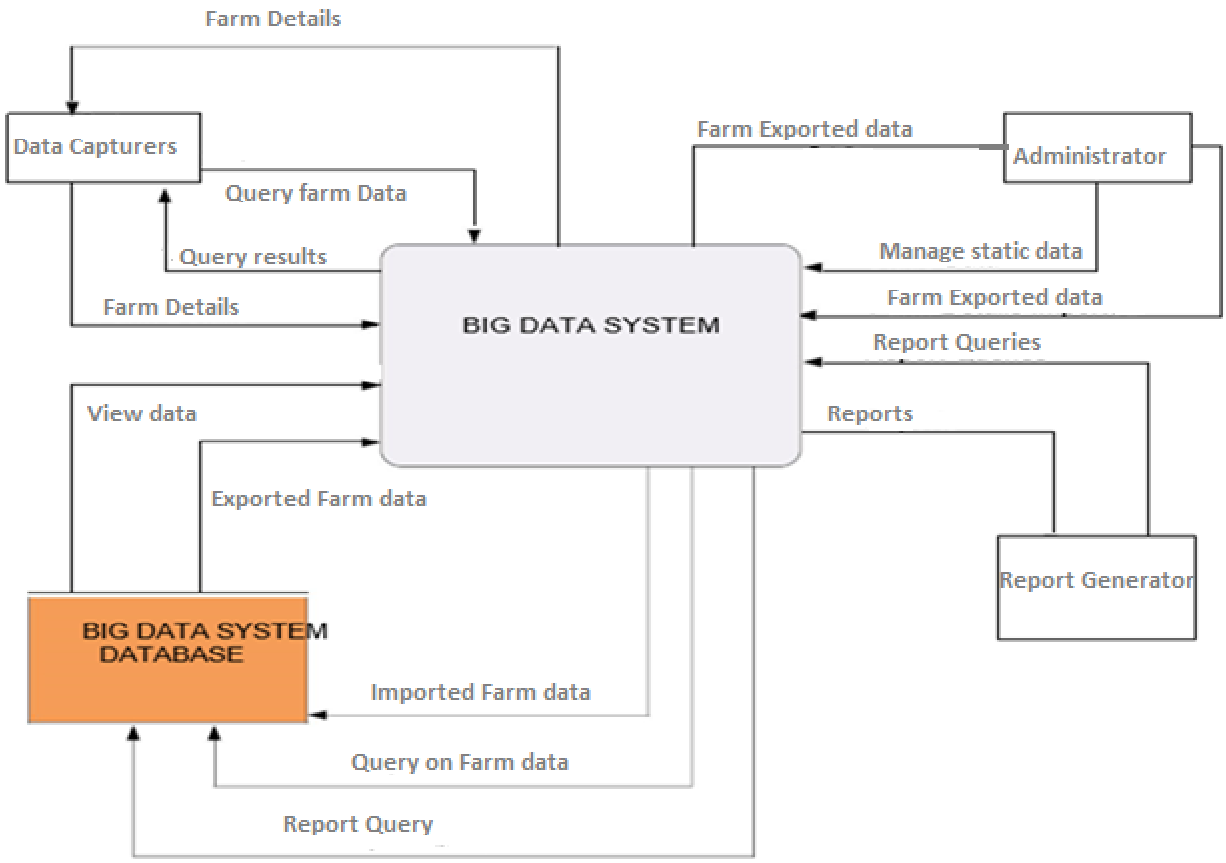
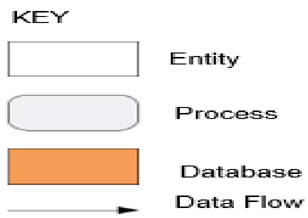


Fig 4.1 Big data system context diagram



4.2.2 Data flow diagram for big data system

Ulrich (2004) explained the DFD as the graphical depiction of the flow of data within the system. It shows the system's sources, data stores, processes and the data flow links form one entity to another. The DFD gives the way to the physical design, the procedures and programs are highlighted by the processes and database, files and data entities are suggested by the data stores.

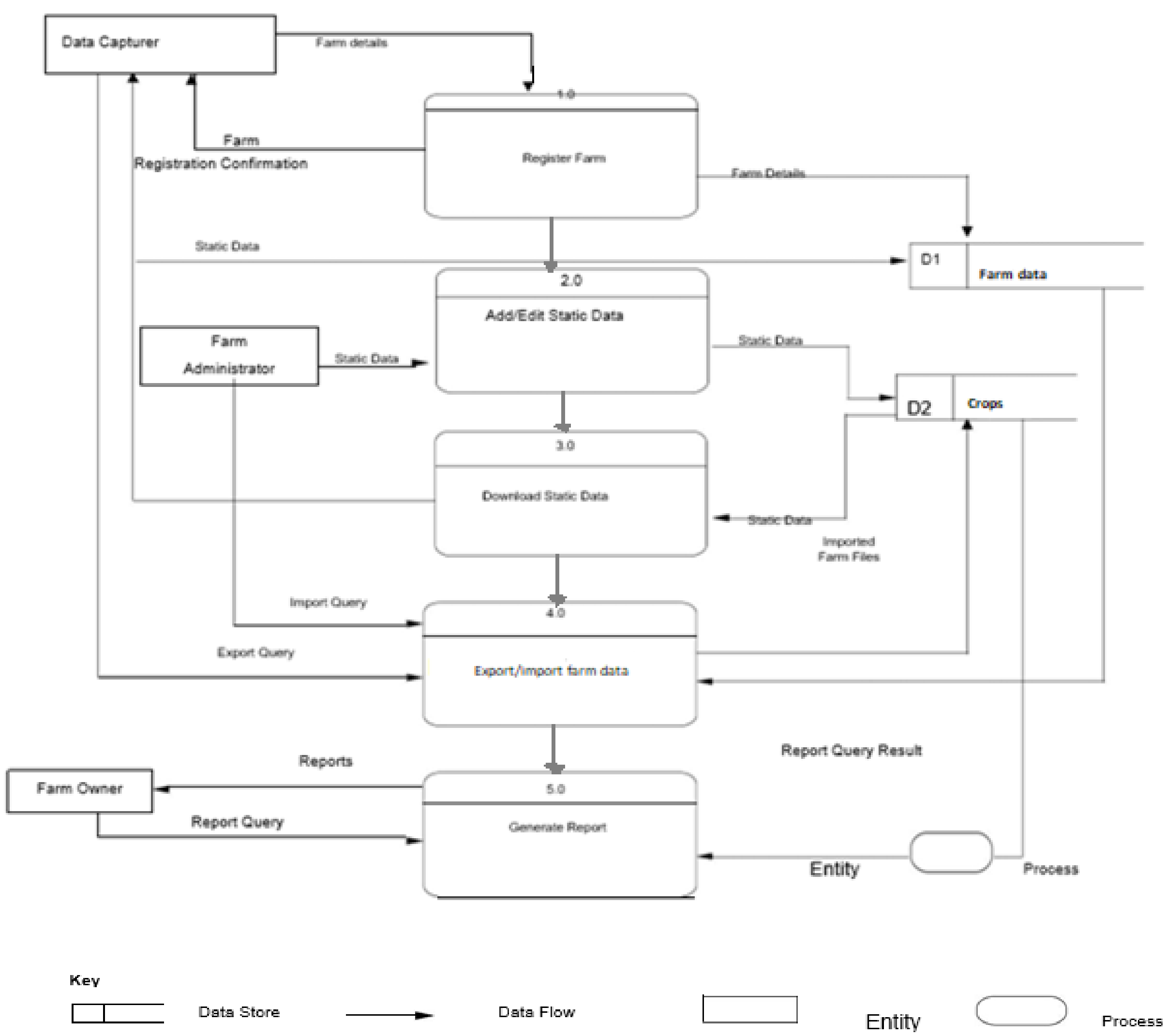


Fig 4.2 Big data system data flow diagram

4.2.3 Logic flow chart of big data system

It shows the graphical picture of the arrangements of the processes and transactions of the proposed system (Birrell, 1985). The logic flow charts provide the procedure manuals which are used when carrying out specific tasks. The logical flow chart documents and specifies the order to perform tasks. The following diagram shows the logic flow of big data system:

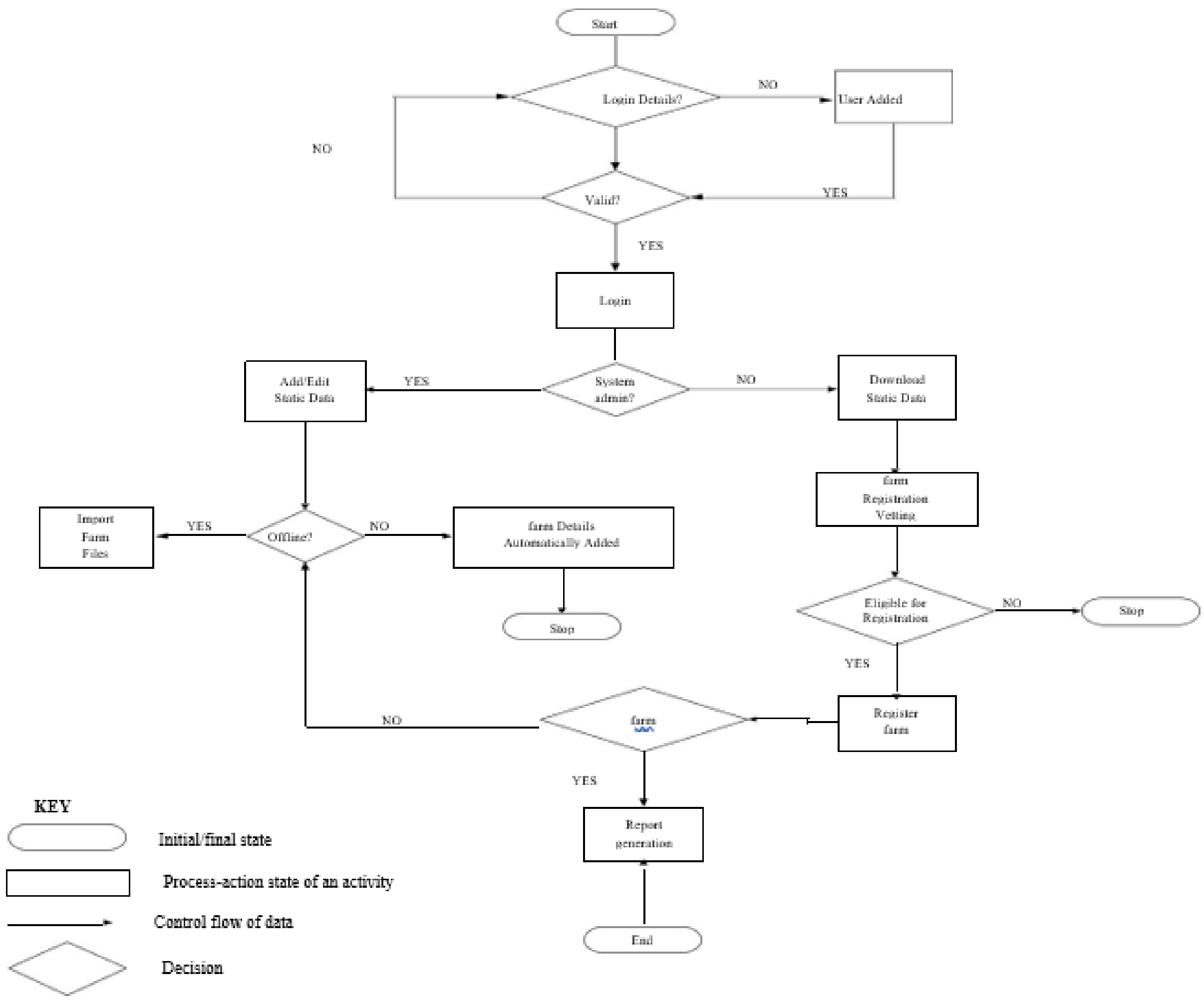


Fig 4.3 Big data system logical flow

4.3 Architectural design

Conger (1994) defined architectural design as the integration of software, network infrastructure and the hardware components. It deals with the physical layout of the hardware components and software installation on the hardware underlaying. The system environment should be reliable and it should not cause future problems.

4.3.1 Network architectural design

Network architectural design is the layout of the network, procedures, hardware, entailing the software and the transmission way (Conger, 1994). The big data system will be using the client and server architecture. The system user will login into the system which will be a web application in order to access the system. The following diagram shows the network architectural design:

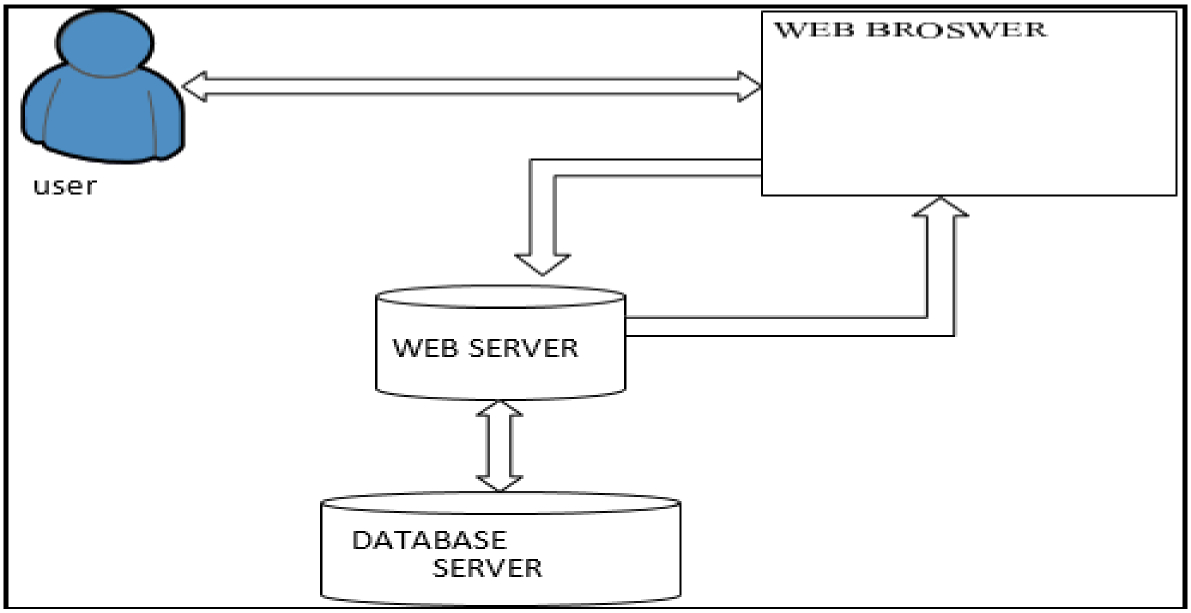


Fig 4.4 Network architecture diagram (source: McCabe (2010))

4.3.2 Client server approach

Client server approach makes a link between the client and the server (McCabe, 2010). It enables the system users to communicate with the database by using different mediums. The web services to be used will be contained in the web server. The query is send by the user to the database through GUI in this case it will be big data system. The server webserver will respond with a solution if it has it in the server.

4.4 Physical design

The raw inputs and the output processes of the system are related to the physical design (Davis and Alan ,1995). This shown by noting down the way in which the data enters the system as inputs, how it is validated and verified, the data processing way and the way in which the data is presented as outputs. During the physical design, requirements will be decided about the system such as the input, output, storage and processing requirements.

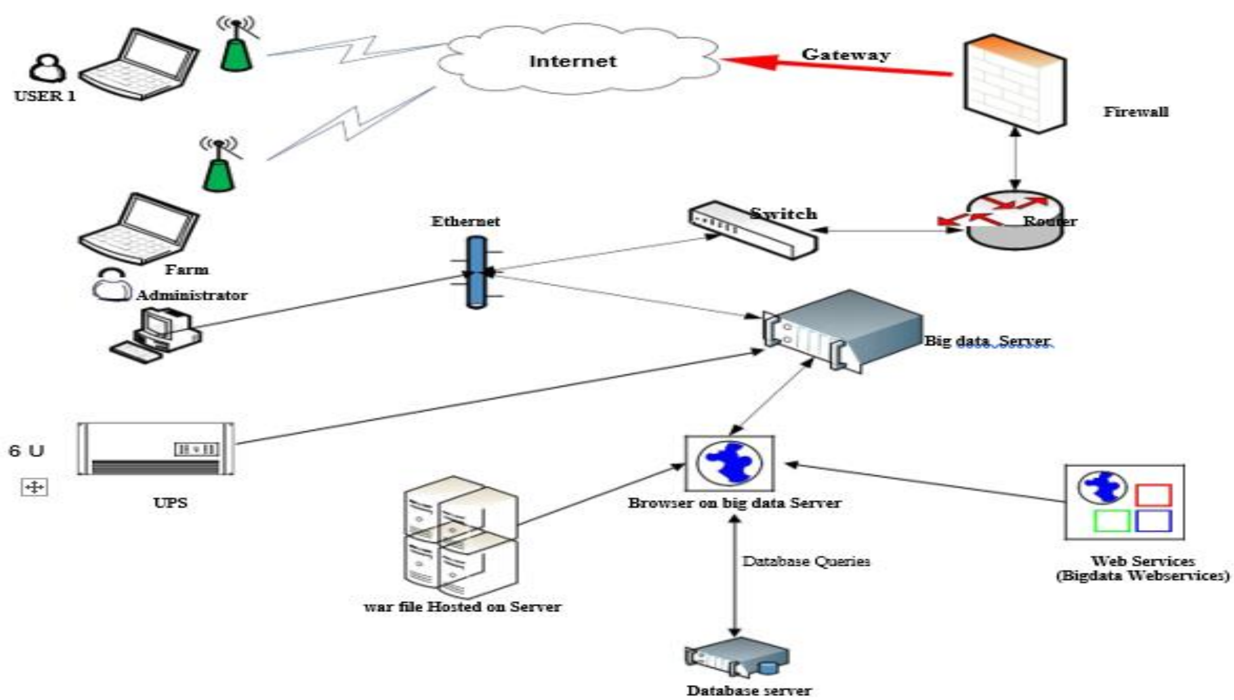
The system design's physical part can be divided into the following sub categories:

- Data Design- shows the data presentation and data storage in the system.
- User Interface design-shows how the users enter raw data into the system and how the system interprets back the information to the users.
- Process design-shows how the raw data passes through the system, the data validation points and the data transformation points within the system.

By the time system design phase is accomplished, the document describing the sub categories above will be produced for the next phase. The following diagram is the physical design of the proposed system

4.4.1 How software and hardware interact

By the time system design phase is accomplished, the document describing the sub categories above will be produced for the next phase. The following diagram is the physical design of the proposed system that shows how the software and the hardware interacts.



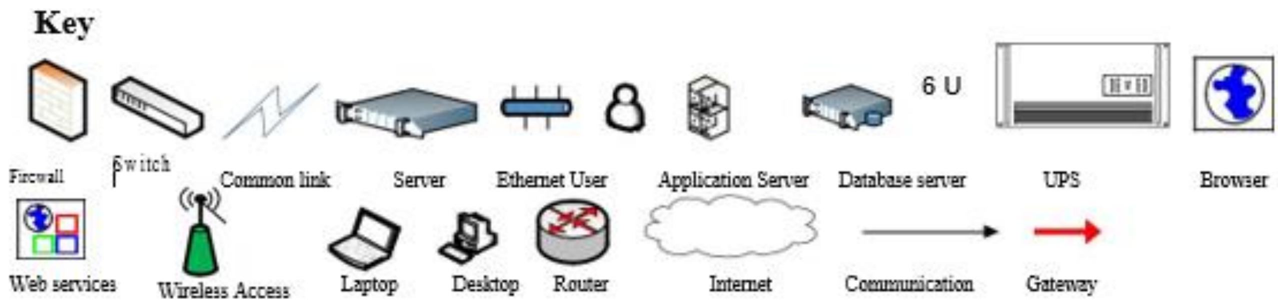


Fig 4.5 Physical design of the proposed system

4.5 Database design

Fournier (1991) explained the database as the central repository for data. A database is a collection of interrelated data designed to meet the varied information needs of an organization. A tool used to store information, or data. Information is something that we all use on a daily basis for a variety of reasons. Well-designed database should address all the business functions to be modelled. The following are the three database design techniques:

- Conceptual database design
- Logical database design
- Physical database design

4.5.1 Conceptual database design

Conceptual database design is the process of constructing a model of information used in an enterprise, independent of all physical considerations. This step involves:

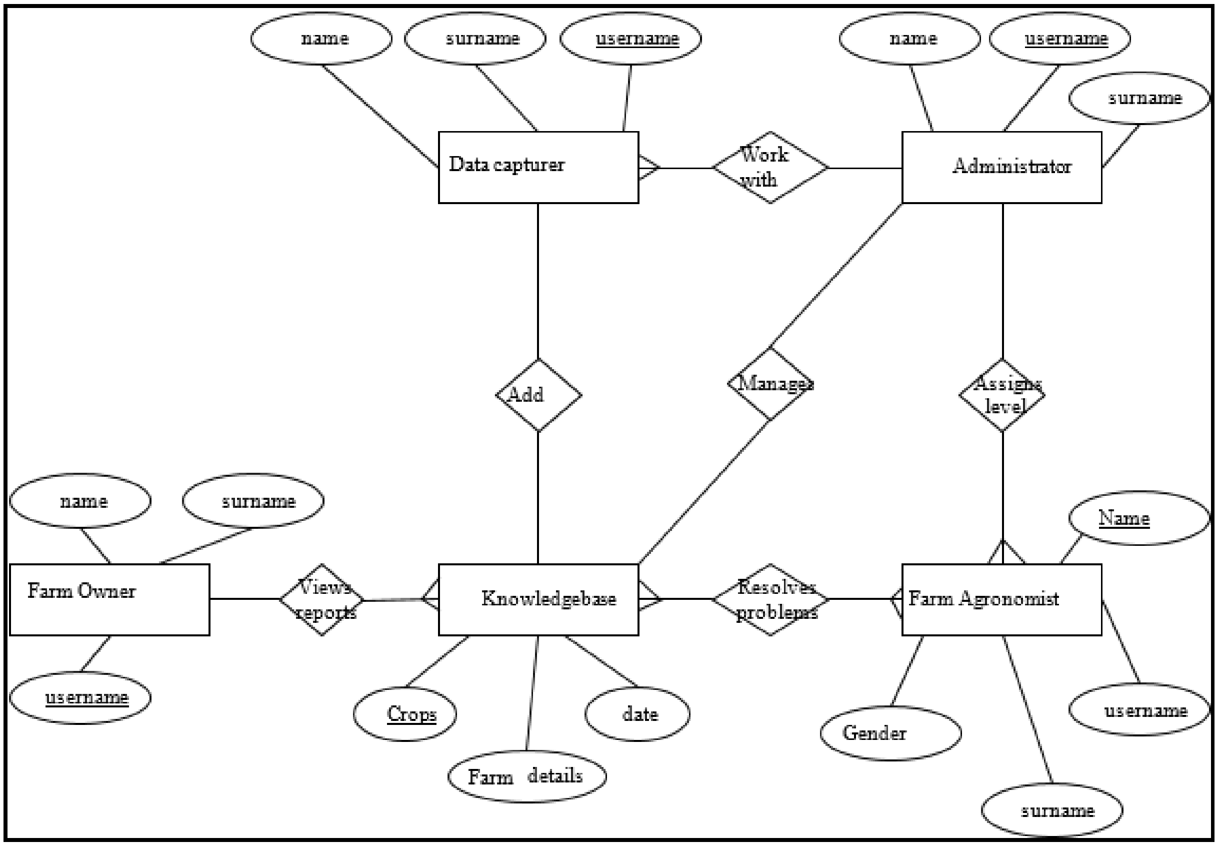
- The ER Model Construction
- Redundancy checking on the Model
- User transaction validation to ensure that all the scenarios are supported.

4.5.1.1 Entity Relation Modelling

It is a pictorial depiction of the real-world problem in terms of the entities (which have attributes) and the relations between the entities McGraw (2010)

- Entity: It is a class of distinct identifiable objects or concepts
- Relations: associations among entities.

- Attributes: are the properties or characteristics of entities



Key

Fig 4.6 Entity Relation Diagram

4.5.2 Logical database design

It is the process of creating a model of data used in an enterprise based on certain data model (for example relational), but separate of the DBMS and other physical considerations. The following steps are involved:

- Table generation from Entity relation model
- Table normalisation

Table Extraction from ER Model

During the generation of the tables from ER Model the relationships cardinality is grouped into the following groups:

One to one -comes from a single entity. Primary key is modelled from a table.

One to many – The primary key is referenced from another table to be a foreign key in the second table.

Many to many -This relationship occurs when the table contains entities form different tables that are referred as foreign keys.

Normalization of Tables:

(Pierce,1992) Normalization is a process of eliminating redundancy and other anomalies in the system. In most cases in the enterprise world, normalization up to Third Normal form would suffice. In certain cases or some transactions it is desirable that certain table be demoralized for efficiency in querying the database tables. In those cases, tables can be in demoralized form.

4.5.2.1 Database Tables

Table 4.1 User Table

Table Name	Description	Field	Data Type	Foreign Keys
User	user details	First name	Varchar(40)	
		Surname	Varchar(40)	
		username	Varchar(40)	
		password	Varchar(40)	
		Role	Varchar(40)	
		Gender	Varchar(40)	

Table 4.2 User_role Table

Table Name	Description	Field	Data Type	Foreign Keys
user_role	A user's roles	user_id	Varchar(40)pk	user_id in table user
		role	Varchar(40)	role in the table role
		username	Varchar(40)	user_id in table user

Table 4.3 Crops Table

Table Name	Description	Field	Data Type	Foreign Keys
Crops	Crops to grow	Crop name	Varchar(40)pk	
		Favourable conditions	Varchar(40)	

Table 4.4 Farm Table

Table Name	Description	Field	Data Type	Foreign Keys
Farm	Farm details	Name	Varchar(40)pk	
		Address	Varchar(40)	
		Owner	Varchar(40)	
		Farming Type	Varchar(40)	
		Location	Varchar(40)	
		Region	Varchar(40)	
		Soil type	Varchar(40)	
		Soil PH	Varchar(40)	

Table 4.5 Crop Rotation Table

Table Name	Description	Field	Data Type	Foreign Keys
Crop rotation	Crop rotation	Crop name	Varchar(40)pk	Fk
		Type of crop	Varchar(40)	
		Date planted	Datetime	
		Date harvested	Datetime	
		Comment	Varchar(40)	

Physical Database Design

The storage media on the database is described by the physical configuration of the description process of implementing of the secondary storage of a database (Maier et al, 2000). It consists of describing the file organisation, the relations and the design indexes used to link the data and the constraints of the entities and the measures of security.

The architecture of the database consists of three abstraction levels which are internal, external and conceptual as a standard by the ANSI-SPARC.

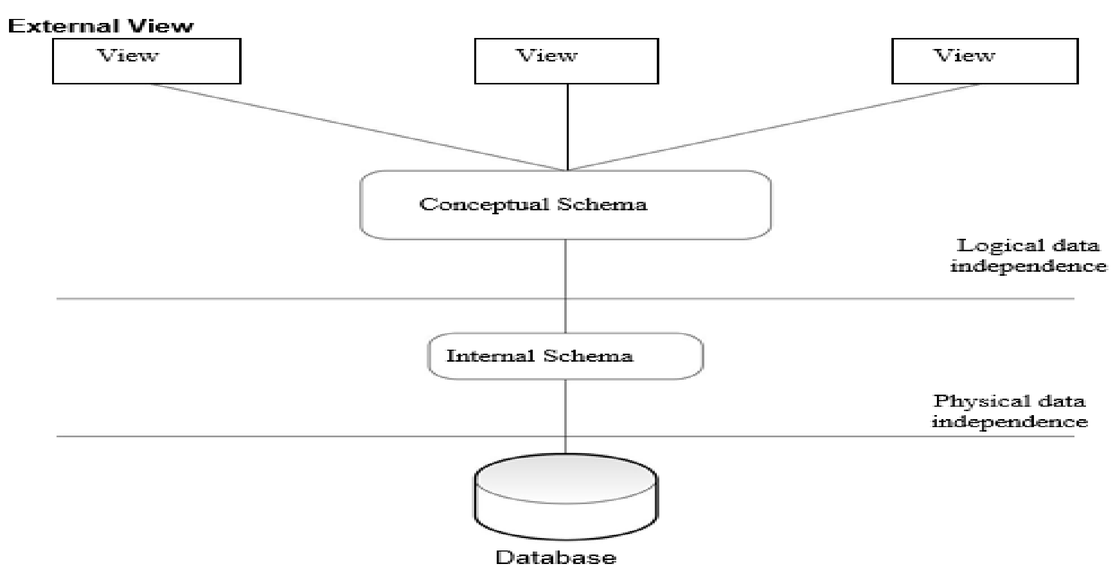


Fig 4.7 ANSI-SPARC Database design

4.5.3.1 External Level

It shows the external view of the database by the users and it shows the database parts that is crucial to a certain user.

4.5.3.2 Conceptual Level

Conceptual level shows the actual data that is stored in a database and the existing relationships of the data. (Gerald, 2005) described the tables as the data storage and tables shows specific attributes features (format, data type)

4.5.3.3 Internal Level

Internal level depicts the greatest abstraction level and database physical representation on the machine. Describes the data storage in the database. SQL is used to manipulated data in the database. DBMS operates below the internal level that enable users to create, define and the database maintenance.

4.6 Program design

The phase of program design focuses on producing the working software modules as stated in the designing phase (Gerald, 2005). The individual modules are being coded by using the program specifications highlighted in the design phase. This phase of designing program is fully shown in three diagrams listed below:

- Package diagram
- Sequence diagram
- Class diagram

4.6.1 Package Diagram

The dependencies amongst the packages that construct a model is depicted by the package diagram in the UML (Gerald, 2005). There are two forms of the dependencies explained within packages in the UML dependency:

Package import: is a relationship of a package and importing namespace, showing that addition of the names members after importing namespace its own package.

Package merge: is the direct relationship of the two packages that shows the combination of two packages.

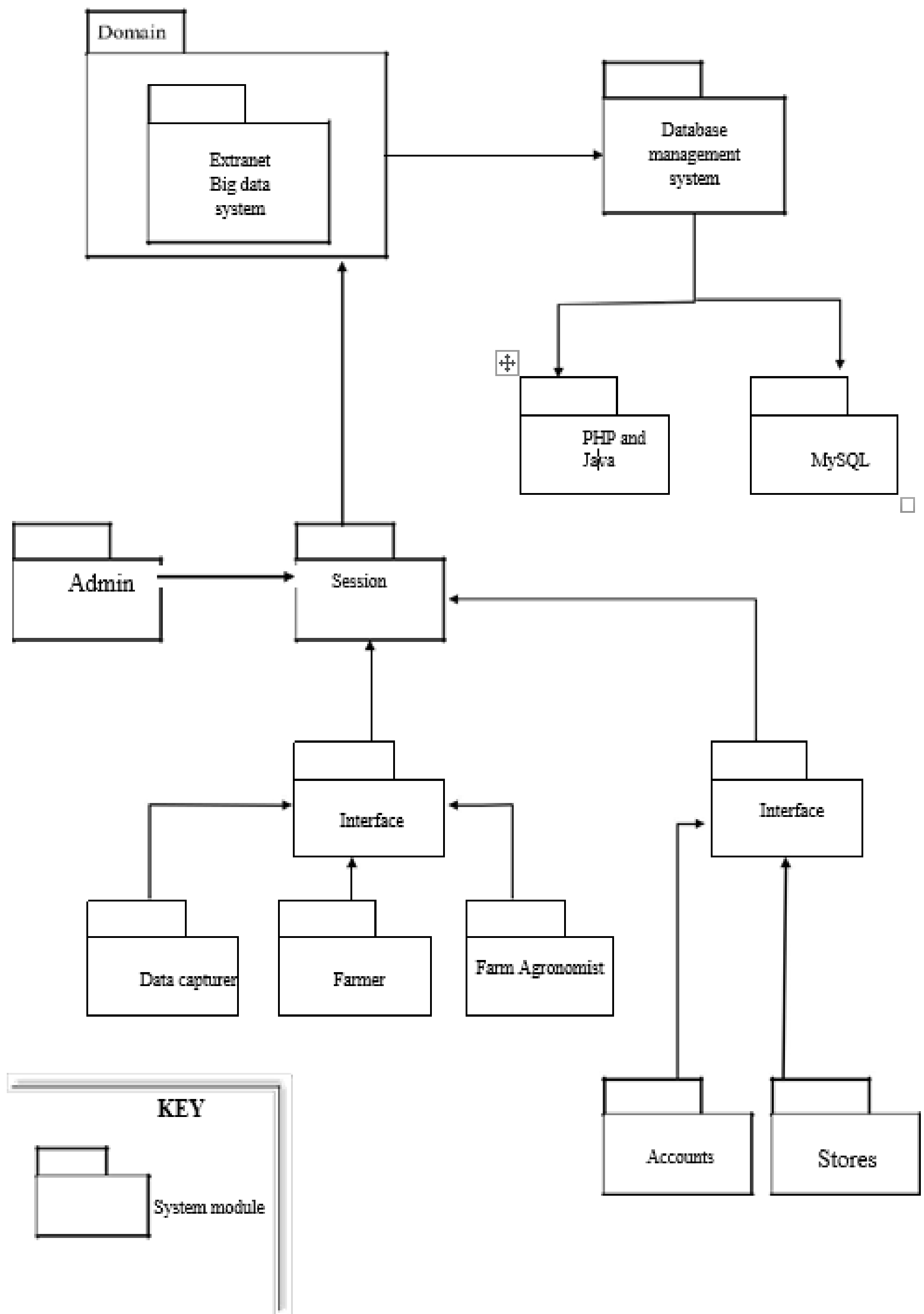


Fig 4. 8 Package diagram

4.6.2 Class Diagram of big data system

(Gerald, 2005) defined the class diagram as a figure of the static structure that shows system structure by depicting classes of the system, the class attributes and the classes relationships. System static models shows the structural relationship that exist among the data pieces that is manipulated by the system. Inheritance and the association are the two types of relationship that within classes. The super class relationship is generalized by the inheritance by using the arrow that is empty to point the super class from the subclass.

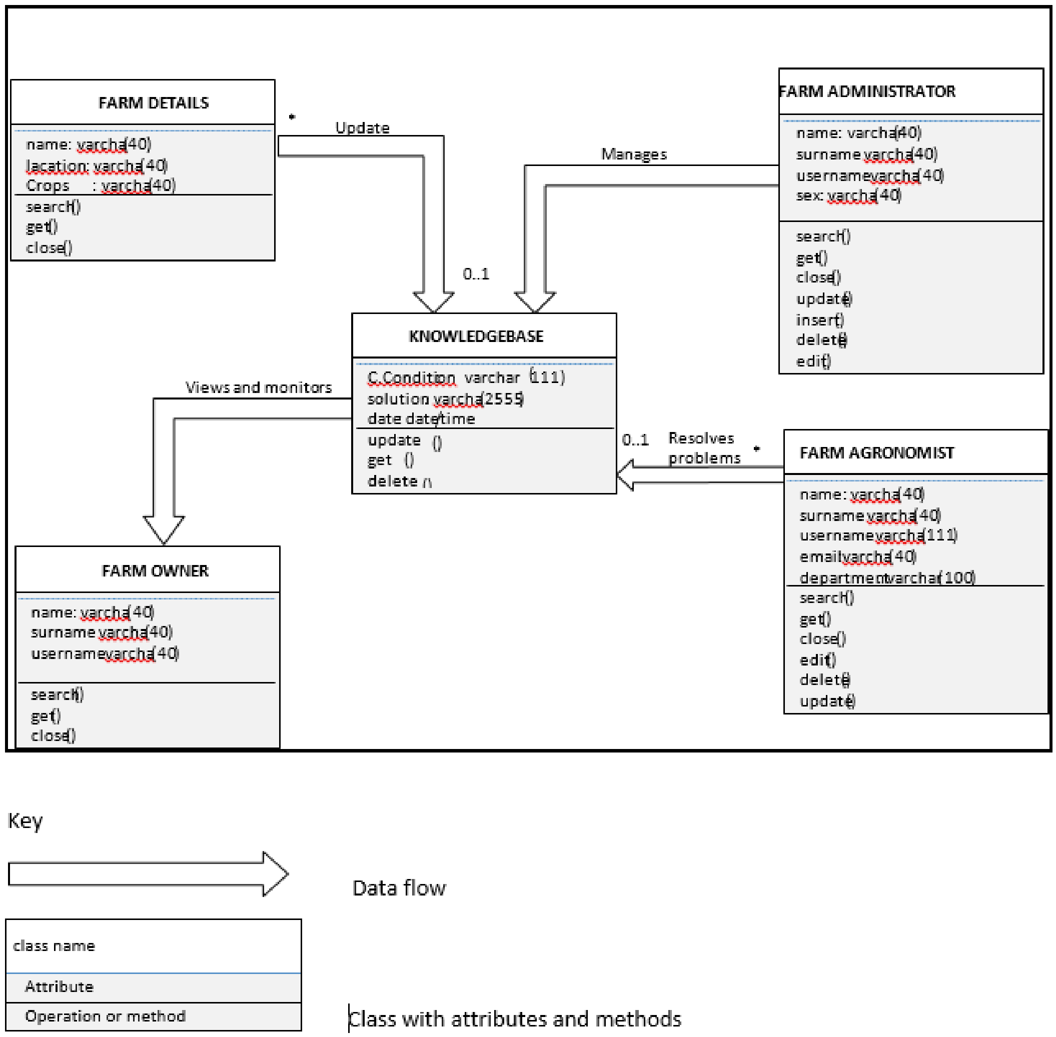


Fig 4.9 Big data system class diagram

4.6.3 Sequence Diagram

The operation of processes with one another is shown by the sequence diagram and its show the process interaction order (Hoglund and McGraw, 2004). The objects interaction is shown arranged in the sequence of time. It describes the classes and the objects tangled in the situation and the arrangement of exchanged messages amongst the objects required to carry situation functionality. The following diagram is the sequence diagram of the proposed system:

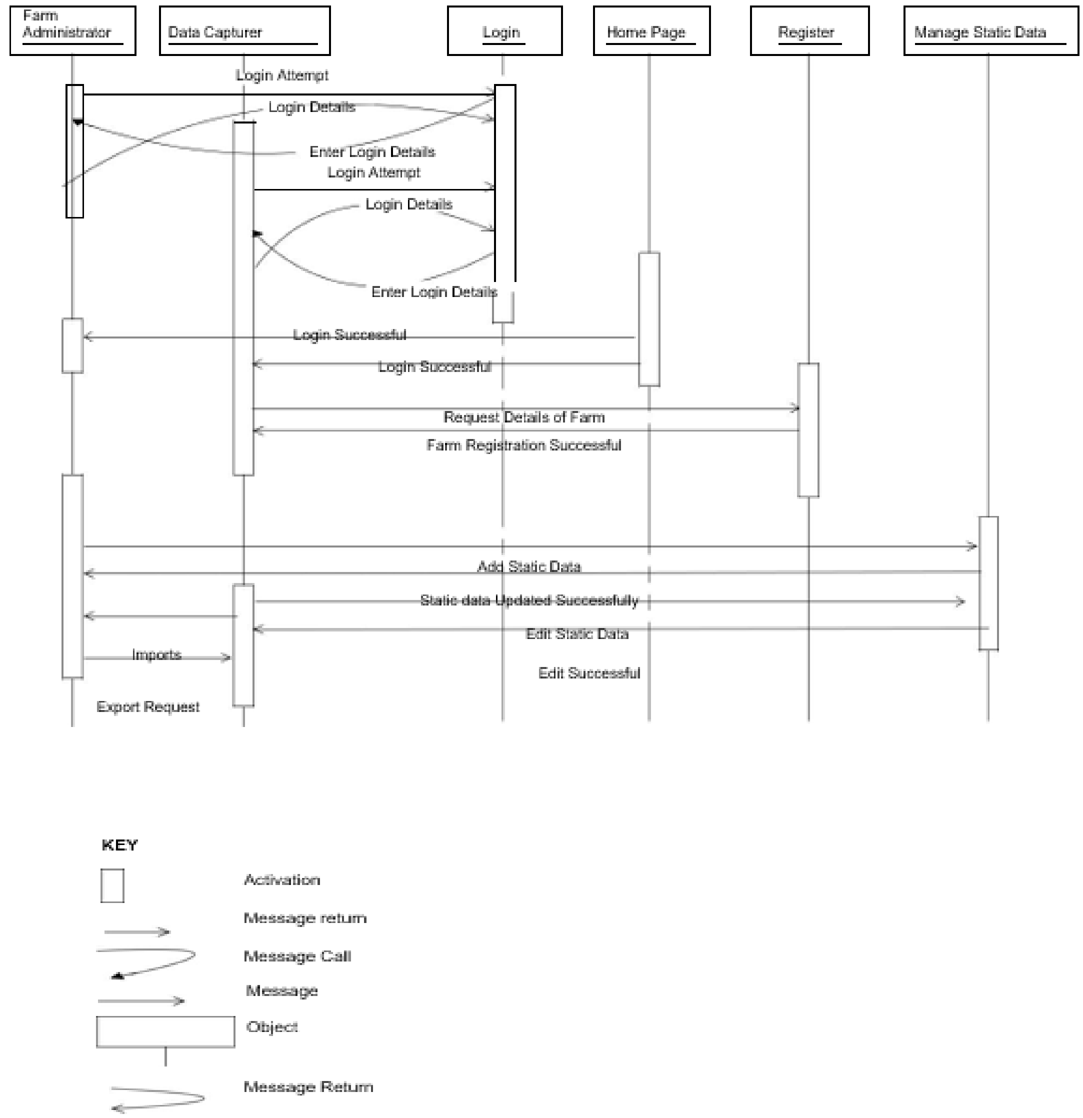


Fig 4.10 Big data system sequence diagram

4.7 Interface design

It's the engineering of the forms, activities and menus that enables users to interact with the system easily (Valacich, 2009). Defines how the users input and output the data required by the system through the use of the button clicks. The graphic user interface will be designed to enable the users to complete their tasks easily and efficiently without or minimum supervision

4.7.1 Menu design

It is the engineering of the forms or activities that the system redirects the user after successful login (Whitten et al, 1994). The user will see the main and the sub menus

4.7.1.1 Main menu

When the user credentials are correct the system will redirect the user to the main menu below:

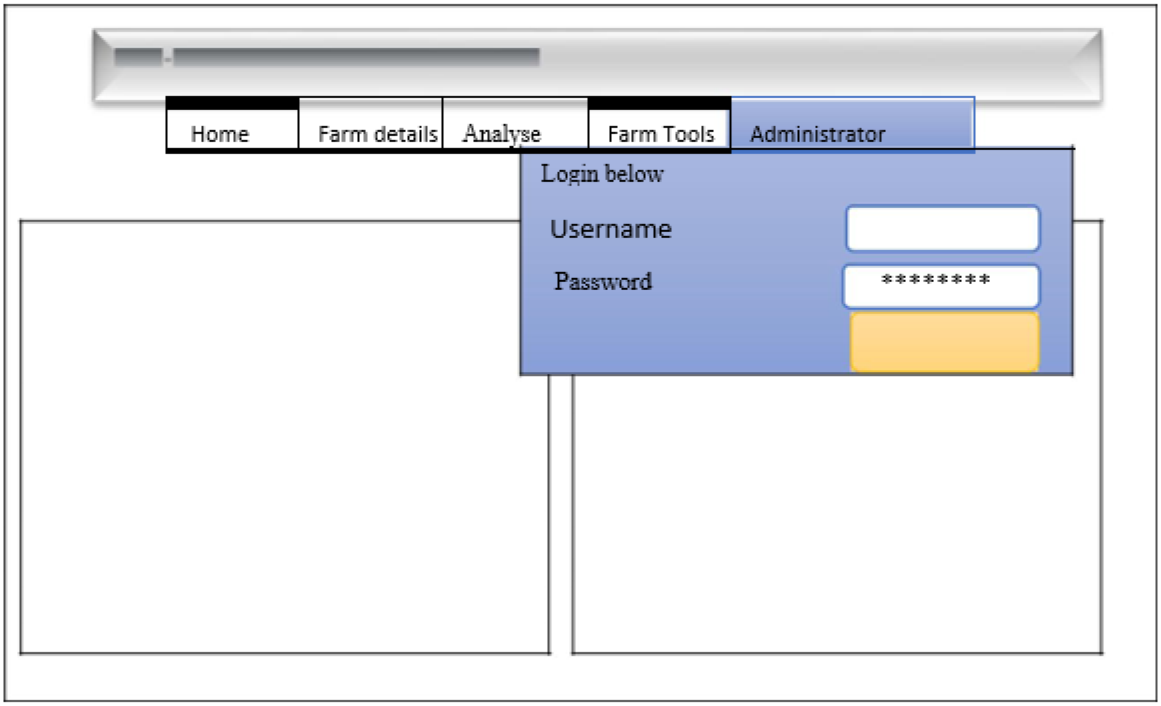


Fig 4.11 Main Menu Form

4.7.1.2 Sub menus

If the user login successfully, he/she will be able to capture daily activities, register, import static data. If the farm administrator logs into the system he/she will be able to edit, add, view and delete users, static data and to analyse. If the farm agronomist logs into system he/she will be able to analyse data and generate the reports.

If the farm administrator logs into the system he/she will be able to edit, add, view and delete users, static data and to analyse

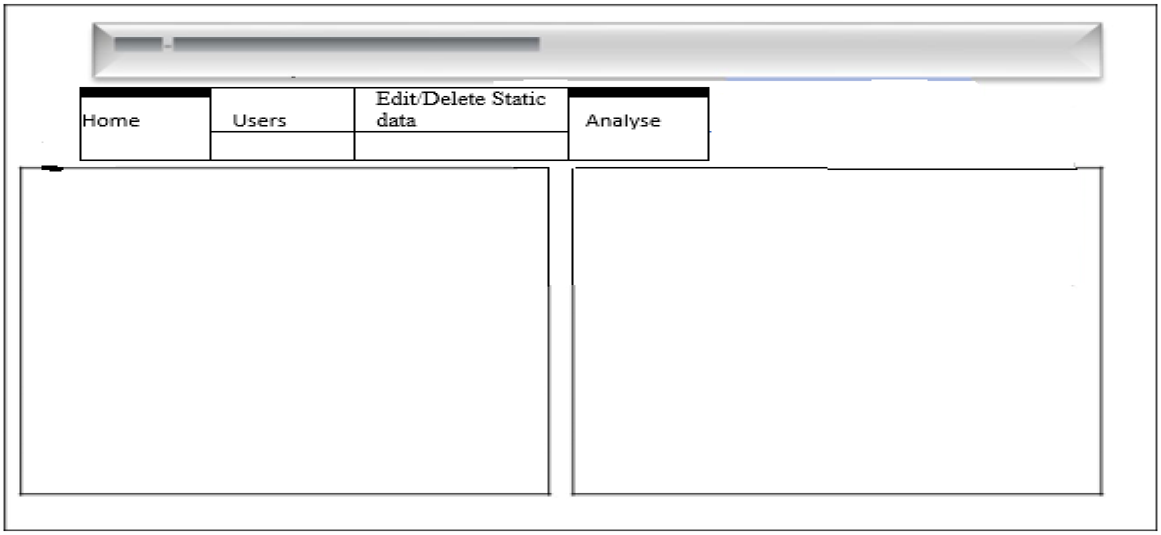


Fig 4.12 Farm administrator Menu Form

If the farm agronomist logs into system he/she will be able to analyse data and generate the reports

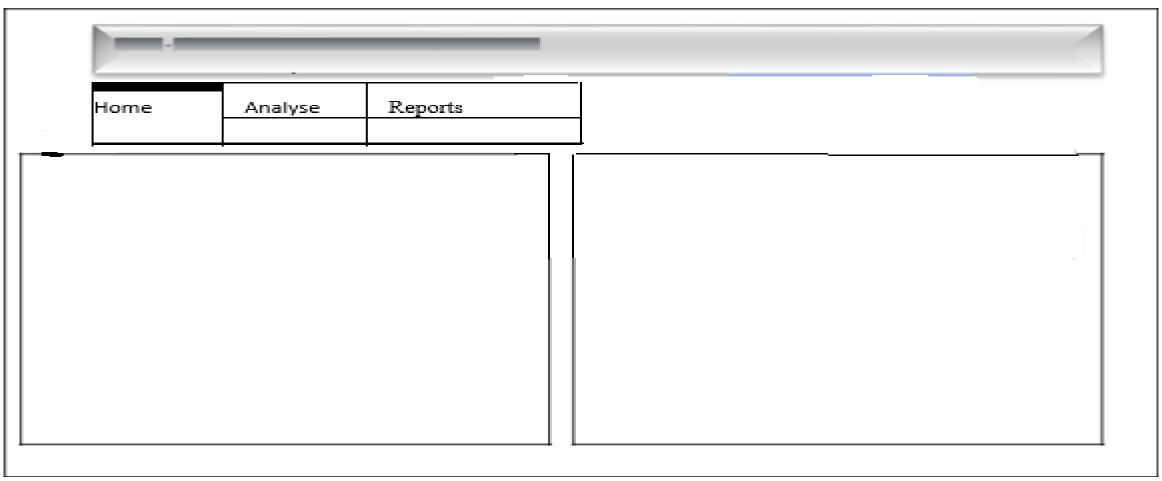


Fig 4.18 Farm Agronomist Menu Form

4.7.2 Input design

Is the engineering of the graphical user interface forms that enables the users to enter information into the system (Pierce, 1992). The system will validate the inputted data and accepts it into the system through the use of input forms.

The user login interface of the big data system.

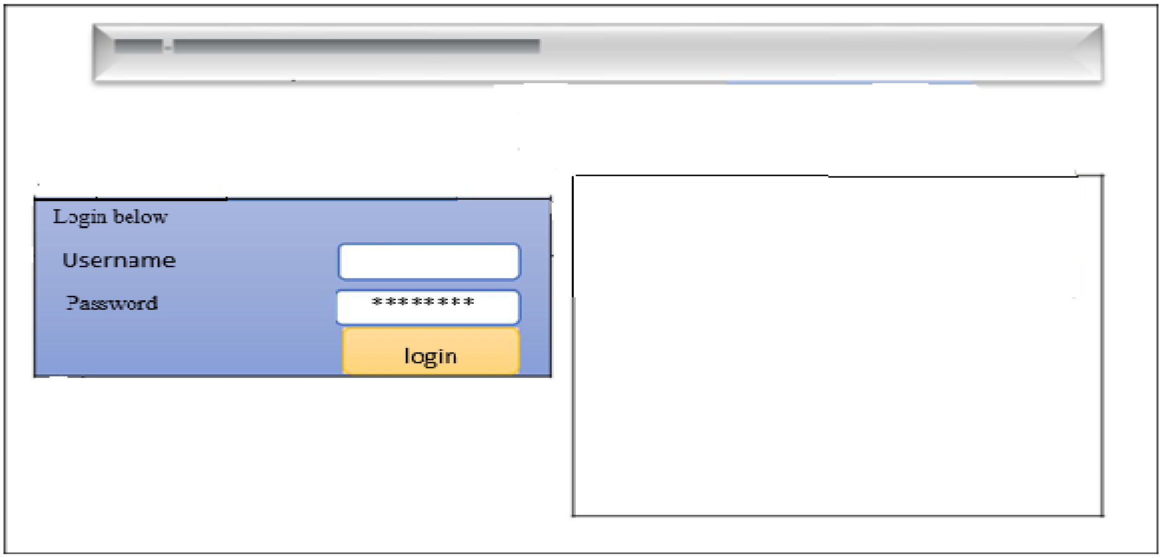


Fig 4.13 Login Form

All the individual accounts are created by the system administrator. The account creation form is shown below:

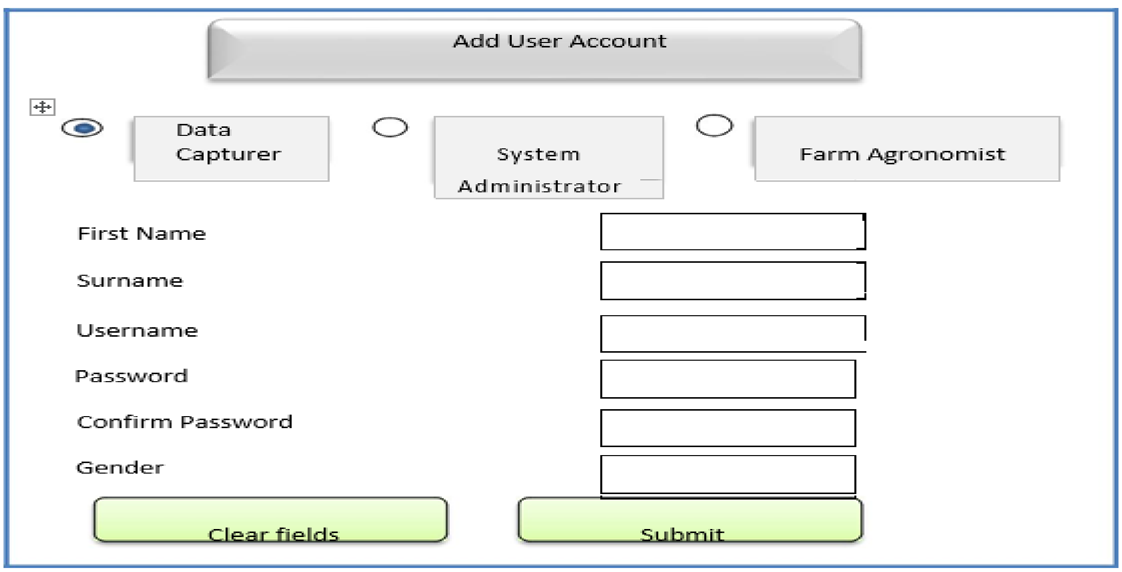


Fig 4.14 Account creation Form

The Farm agronomist enters the crop details to update the knowledge base about crops favourable conditions.

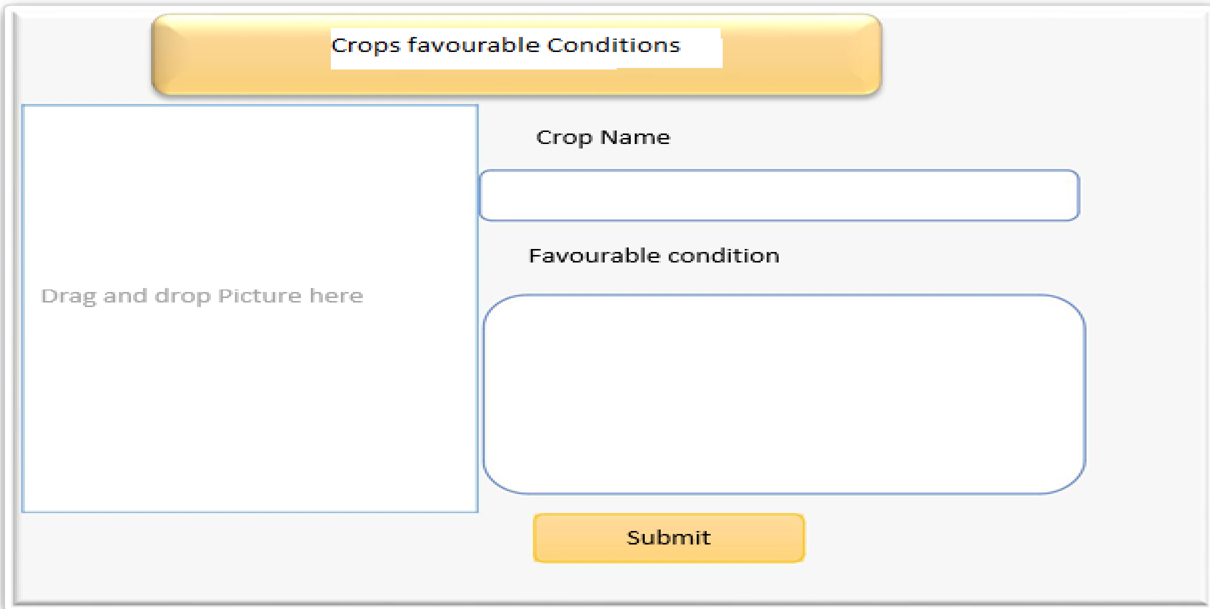


Fig 4.15 Crops favourable conditions form

The data capture will be able to upload farm data form the excel sheets.

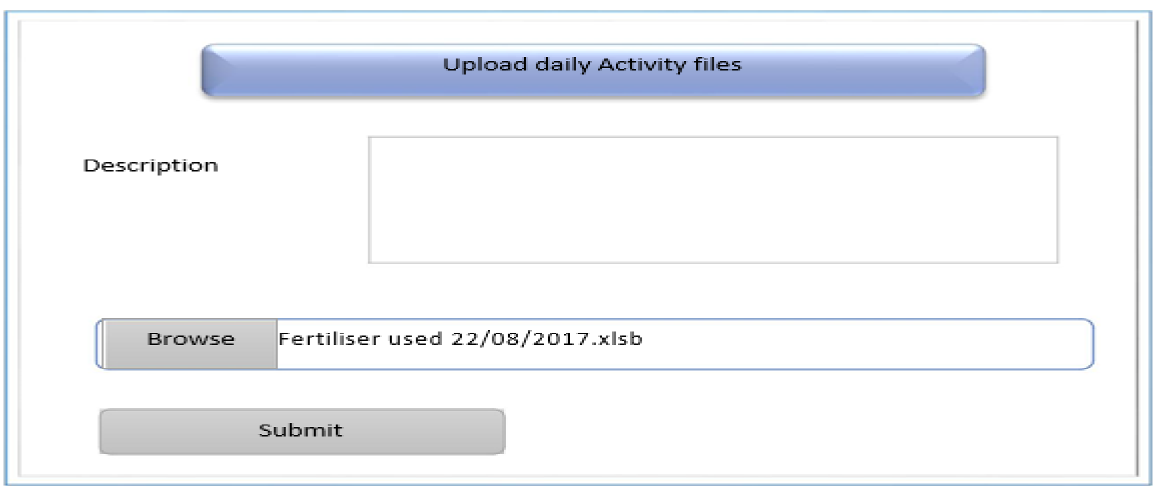


Fig 4.16 Upload daily activities form

The crops grown are logged for the crop rotation use.

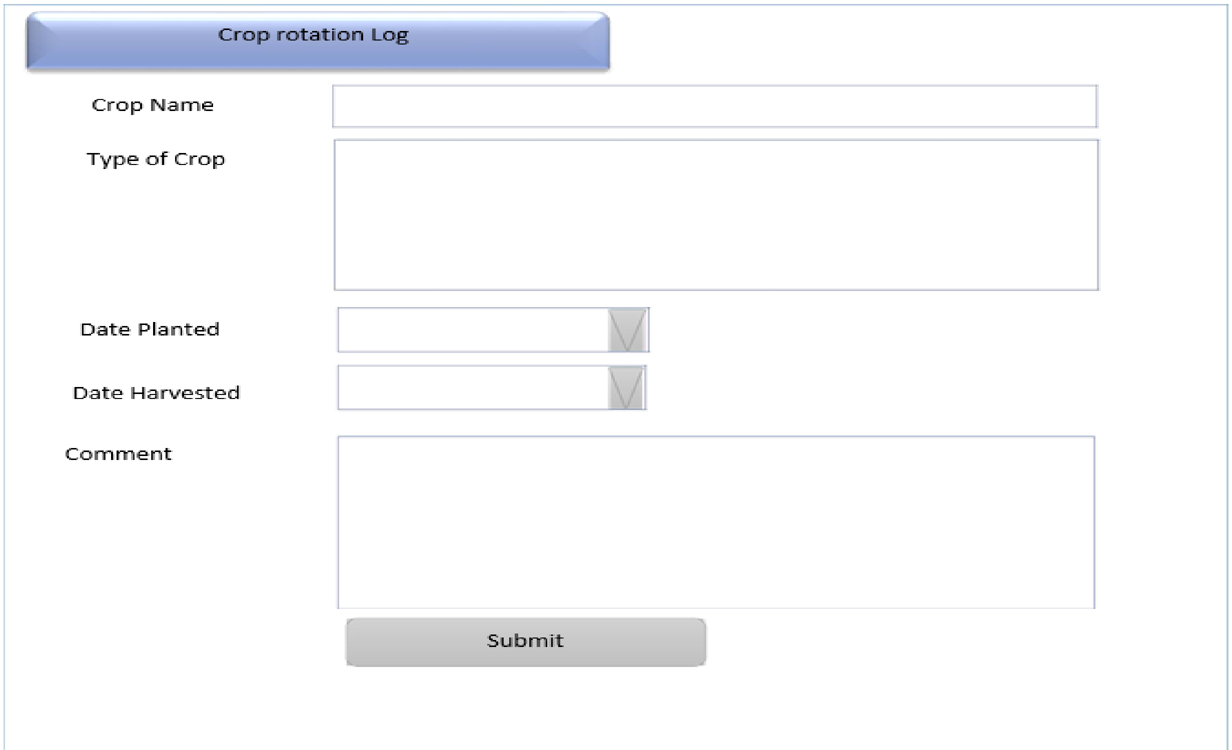
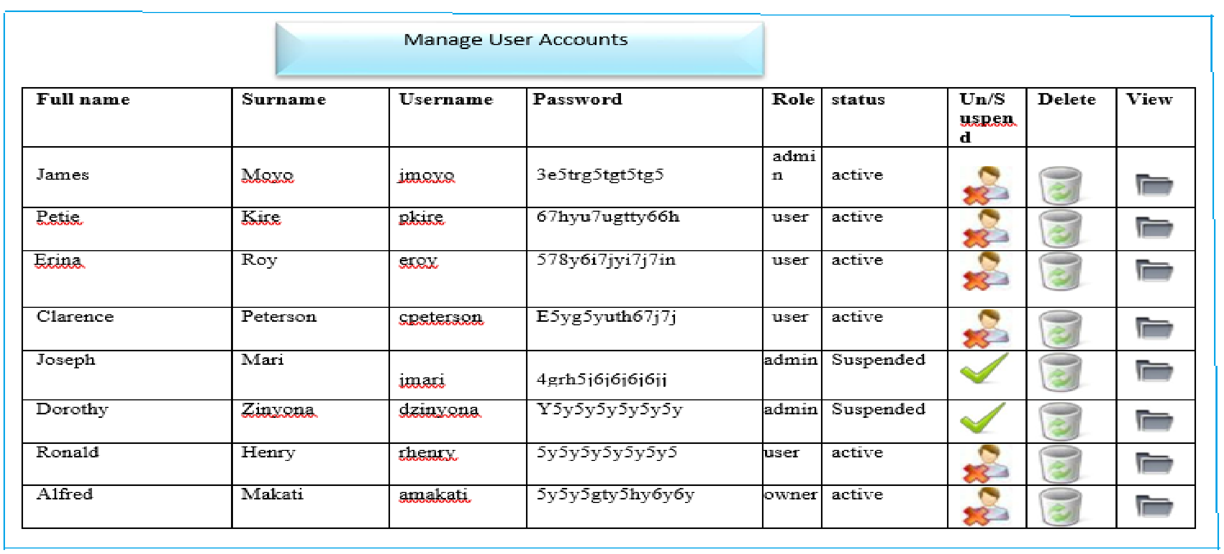


Fig 4.17 Crop rotation log

4.7.3 Output design

The farm administrator can add, edit and delete the user accounts.



Manage User Accounts

Full name	Surname	Username	Password	Role	status	Un/Suspend	Delete	View
James	Moyo	jmoyo	3e5trg5tgt5tg5	admin	active			
Petie	Kire	pkire	67hyu7ugtty66h	user	active			
Erina	Roy	eroy	578y6i7jyi7j7in	user	active			
Clarence	Peterson	cpeterson	E5yg5yuth67j7j	user	active			
Joseph	Mari	jmari	4grh5j6j6j6jj	admin	Suspended			
Dorothy	Zinyona	dzinyona	Y5y5y5y5y5y5y	admin	Suspended			
Ronald	Henry	rhenry	5y5y5y5y5y5y5	user	active			
Alfred	Makati	amakati	5y5y5gty5hy6y6y	owner	active			

Fig 4.18 Manage user accounts

The farm agronomist generates the report about the best plants to grow in the coming season after analysing the relationships of the farm conditions.

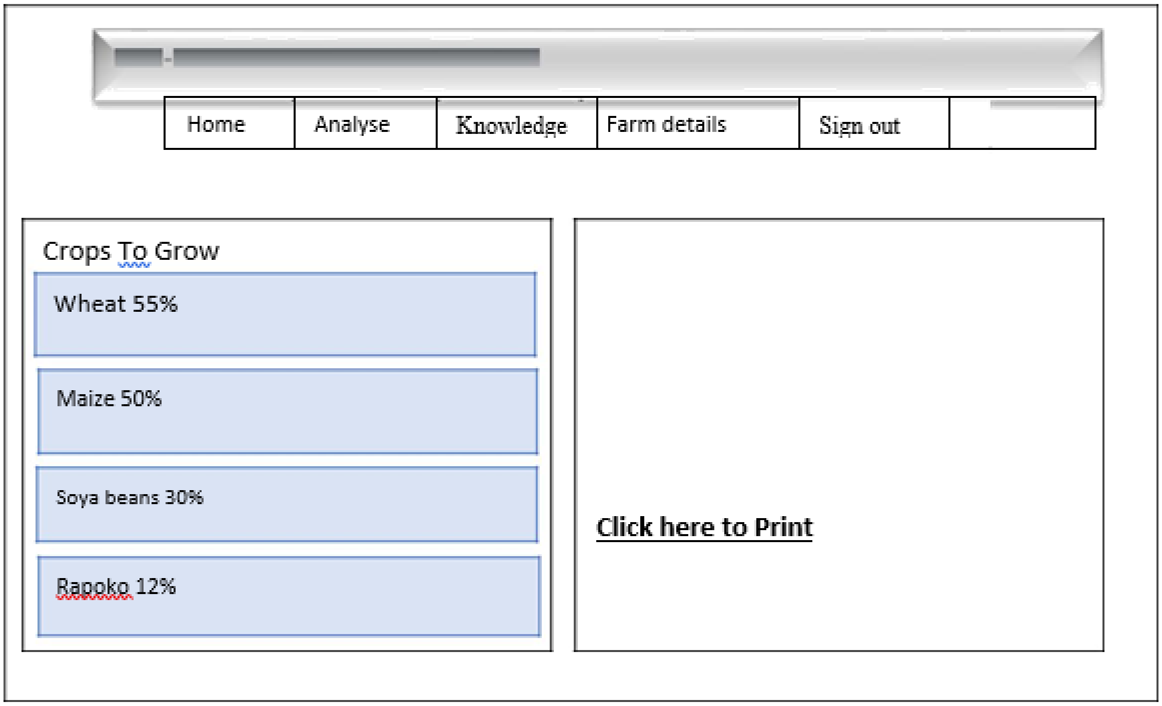


Fig 4.19 crops to grow

4.8 Pseudo Code

Sindre and Opdahl (2000) defined pseudo code as the way of presenting instructions as a program. Pseudocode is written before the system actual code to make sure that all needed modules are included and its enable easy statement translation into any programming language. The following pseudo codes were used:

Database connection

Check if there are connections

If no connection, exits

Establish connection

Else

Ignore

End if

Login

Enter username and password

If username and password are correct

Open the main menu.

Else

Invalid login

End if

Add Crop Favourable Conditions

Enter crop details

Validate data entered

If input is not valid

Return error

Else

Save entered data using current connection

Return confirmation message

End if

Updating database data

Find key Field

Search if record exist

If record not found then

Log error

Else

Save new record

End if

Database backup

Specify backup location

Create backup file

Save file specified location

Retrieve back up file

Locate backup file

Retrieve backup file

End

4.9 Security design

McGraw(2004) explained security design as the structures that enables the system to be completely protected from the threats and attacks. The big data system is very secure to avoid possible threats, the following diagram shows the big data system security design from the database down to the browser where the system will be accessed.

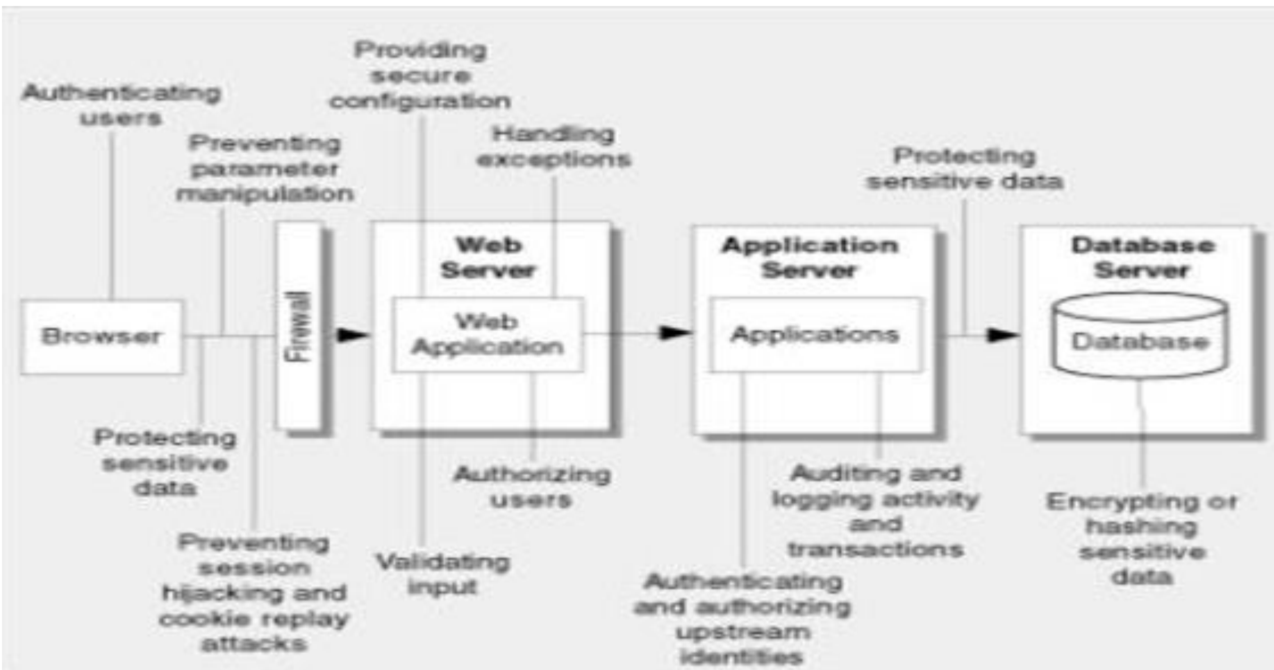


Fig 4.20 Security design diagram

4.9.1 Physical security

It is the actions structured to avoid unwanted access to the system tools and the resources and to secure the property from physical damage. The following measures were used:

- The intrusion detectors were installed in server room.
- Security cameras were installed to help to monitor the people’s actions.
- Biometric sensors were installed at the doors to allow only authorised people to enter.

4.9.2 Network Security

Mishra (2002) explained network security as protection of directories and file access in a computer network against misuse and hacking. The following will be used to secure network:

- Firewall will be put in place to block unwanted access.
- Use of usernames and passwords to secure the networks and machines.

- The Avast antivirus will be installed on computers for internet security.

4.9.3 Operational security

Mishra (2002) defined the operational security as the ability to protect the critical information against the public. The big data system will be operationally secured by using the following measures:

- The users will be given different access levels as per their daily tasks.
- The system will prompt the users to change password after two weeks.
- The outgoing and retiring workers account will be deleted after they leave.
- More sensitive information will be encrypted every day.

4.10 Conclusion

The researcher was able to come up with brilliant structure and the system operation during the design stage. The researcher was able to draft the physical architecture of the system, the way in which the big data system will be integrated. The system's inputs and he processes were well designed so that the outputs of the system will be presented well to meet the system goals. The system data flows, the relationships between the entities and the system database were designed very well. The solid authentication was well designed and the authorisation strategy was put in place. The researcher managed to solve the fundamental aspects of building and designing the secure application to prepare for the system implementation.

Chapter 5: Implementation Phase

5.1 Introduction

Jones (1998) defined the implementation stage is considered as the final stage of the development life cycle of a system. At this stage there are three sub stages which are system testing to ensure that the product performs as required and to check errors on the system, system installation will be done to enable the users to use the software and the post implementation to check out the need if there is a need for upgrade or to make some system changes. At this stage the researcher will be taking into consideration about the software performance, baselines, quality, libraries and system debugging. Th user training will be conducted and the user document will be delivered. As soon as the system get into its steady-state production, it is revised to make sure that the systems in the project plan were implemented successfully.

5.2 Coding

At the coding phase the programmer turns the specifications of the program into the computer instructions, which are called programs (Howard and LaBlanc, 2003). These programs organise the movement of data and they control the system's entire process. The programs are modular naturally. This enables quick development, system maintenance and the changes in the future, if they are required. The programmers gather the design documents and the tools for development (compilers, debuggers and editors etcetera) and begin writing software. The coding phase is the longest phase of product development life cycle. The well-written and structured software reduces the system testing and maintenance effort.

5.3 Testing

Viega (2001) described testing as the evaluation process of the system, to see whether all the requirements are being met by the system and to correct errors. The system is executed in order to notice the gaps, missing requirements and errors. The program testing was performed step by step manually and the functional tests were jotted down in detail. The testing processes below shows how the big data system for agricultural system was tested to ensure it meets the requirements specified the users:

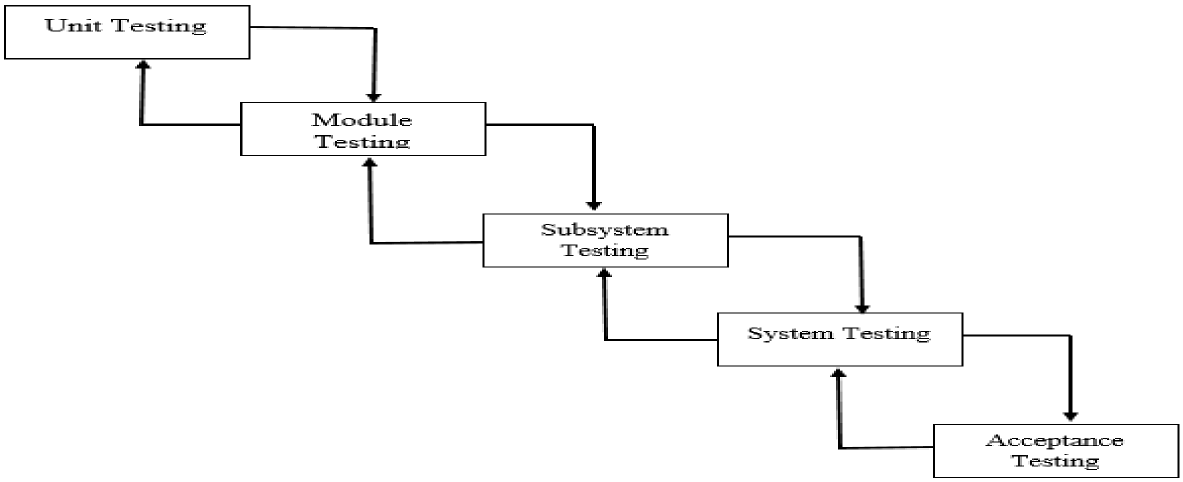


Fig 5.1 Testing stages(Viega,2001)

5.3.1 Unit Testing

The researcher employed unit testing to isolate each and every part and to show that the individual blocks of codes are correct in terms of functionality and requirements before the system setup was handled over the system to the testing team. According to (William, 1998) unit testing is conducted when the whole system is done and the system will be divided into separate blocks of code that performs the main functions of the system. The unit testing was performed for the programs like the logs of the system components, it was done and all the errors were corrected.

5.3.2 Module Testing

The blocks of codes that had been tested at unit test are then integrated together to build a class which is called module. These are then tested to see how they interact with each other as a module. The small sub systems joined by the several units are sometimes referred as a module. The big data system forms were referred as the modules and they were being tested to whether they interact as required by the user specifications. The module testing was conducted successfully and all the errors were corrected.

5.3.3 Sub-System Testing

According to Conger (1994) the sub-system testing is performed when the individual modules are combined so that they formed the sub-system. The system functionality will be tested, there will be a better view of the system since the modules were being joined. The big data system modules were divided into the following sub-systems which are, the system database was tested and all the errors were corrected and documented, web services created in Php were being tested as the sub-system and all the errors were fixed.

5.3.4 System testing

According to Beatty (2013) the system testing is conducted when sub-systems are joined together to form one whole system. The whole system is tested rigorously and evaluated to see if it is effective and efficient. The system testing is done to verify if the system meets the technical specifications and functionality and the system will be tested in an environment that is close to the production environment in which the system will be operating. The system testing enables the testing team to validate and verify the business requirements and the system performance. The big data system was tested using the following popular methods:

- Black box testing
- White box testing

5.3.4.1 Black box testing

The black box testing is done without having knowledge about the system interior structures (Tait and Vessey, 1998). It involves checking of the results of the system rather than the source code. The black box testing enables the system testing specialist to interact with the user interface of the system by entering the data as inputs and examining the outputs without having the knowledge of how the data processes (Taggart, 1999). The black box testing was employed to the big data system and the errors were fixed and documented. The diagram below shows the how the black box testing was conducted:

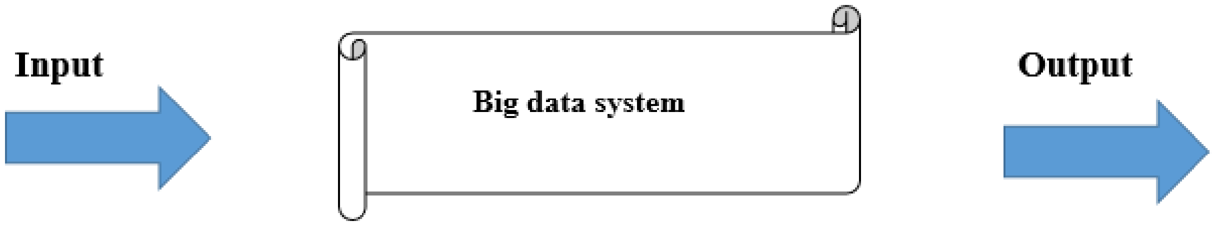


Fig 5.2 Black box testing (source: Taggart, 1999)

5.3.4.1.1 Merits of Black box testing

- Fast and easy way of testing for big systems with lot of pages.
- It's a secure way of testing since no one will be allowed to access the system code.
- Lot of people can test without any programming language knowledge

5.3.4.1.1 Demerits of Black box testing

- Blind coverage, since the code will be not accessed.
- Insufficient testing, usually this method of testing will cause the system to require too much maintenance.

5.3.4.2 White box testing

It is the detailed investigation of the inner logic and the code structure. The white box testing involves the analysis of the lines of the source codes and the system internal processes Taggart (1999). This type of testing requires the expert who possess the programming language knowledge. The Big data was tested using the white box testing by examining the structure of the source code and from the web services and the stored procedures structure. The following diagrams shows how the white box testing was performed:

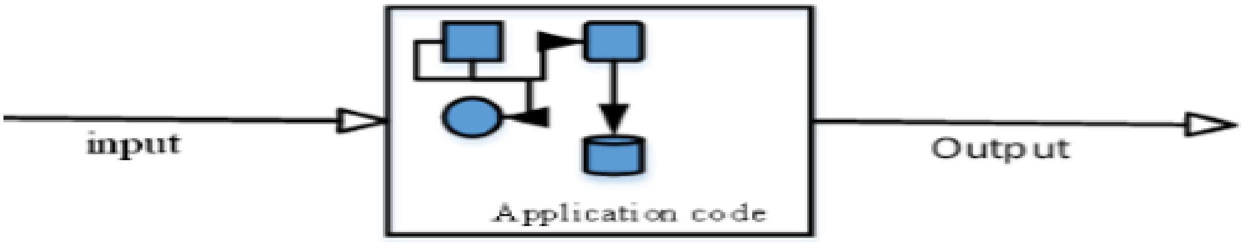


Fig 5.3 White box testing (source: Taggart (1999)

5.3.4.2.1 Merits of Black box testing

- Lot of bugs are noted and fixed since the code will be well examined.

- System maintenance will be easy since more bugs will be noted and fixed.
- Secure way of testing since only specialist will be allowed to examine the code

5.3.4.2.2 Merits of Black box testing

- White box testing is very expensive since specialist will be hired
- Time consuming since each line of code will be inspected

5.3.5 Acceptance testing

Acceptance testing is one of the crucial type of testing, the system is handled over to the users, stakeholders and any quality assurance teams to see whether the system met the intended specifications and the requirements specified by the users. The users test the system under the production environment and they check the interface gaps, spelling mistakes and the bugs. The big data system was tested using the acceptance test and all the bugs were corrected

5.3.6 Test cases

Test 1: Null Fields Validation.

The system was validated, every input form was validated. In order to have clean data the inputs should be validated. If a user saves a null field the following error messages appear

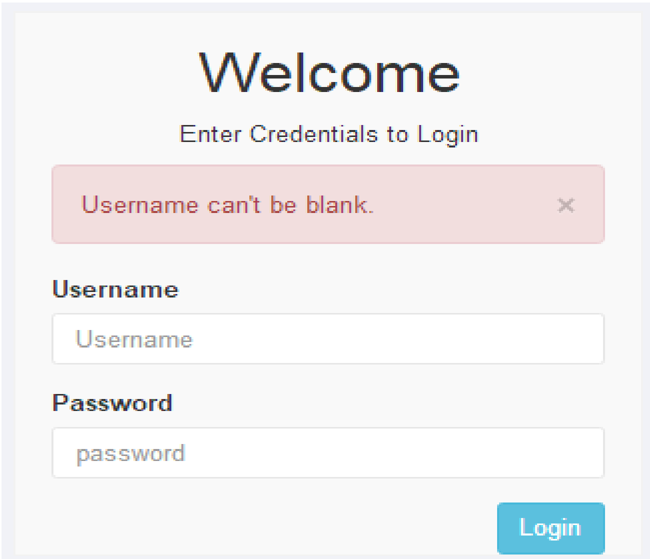


Fig 5.4 Null Fields Validation

Test 1: Login Validation

If the user enters wrong details the system will notify the users as shown in the diagram below:

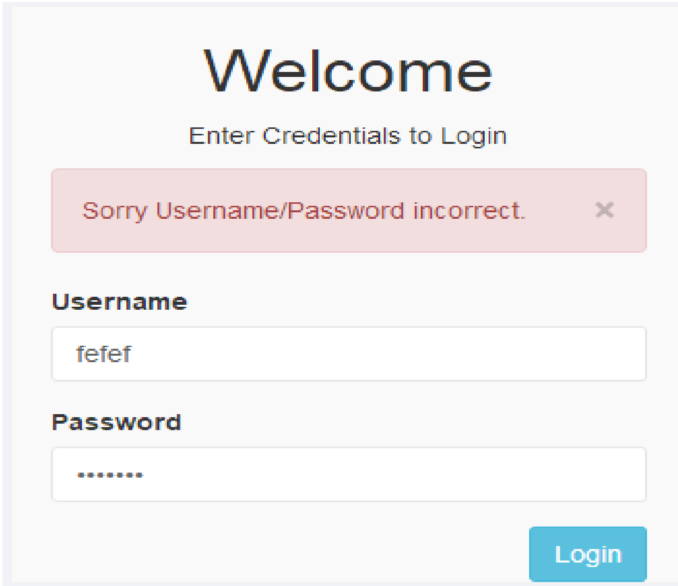


Fig 5.5 Login validation

Test case 2: Adding User

The farm administrator has the rights to add new users and assign them levels according to their daily tasks and job descriptions. The system gives back the confirmation message to confirm that user had been added successfully. The following diagram shows how the new user is being added.

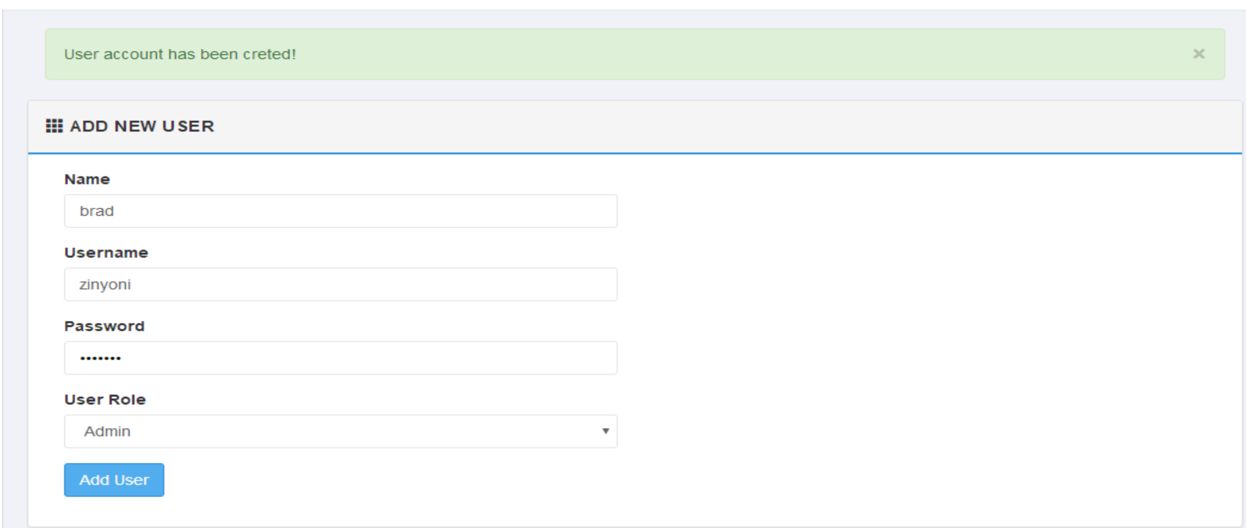


Fig 5.6 Adding user

Test case 3: Null CSV file upload

The system is validated to reject null csv file upload, the user has to select the file to upload.

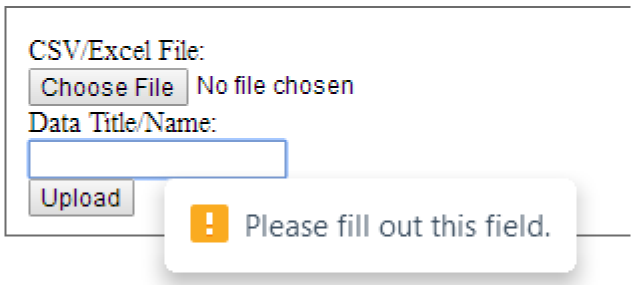


Fig 5.7 Null CSV file upload Validation

Test case 4: Crop condition Display

The crops favourable conditions are retrieved from the database before the data is being analysed, if there is error the crops will not be displayed.

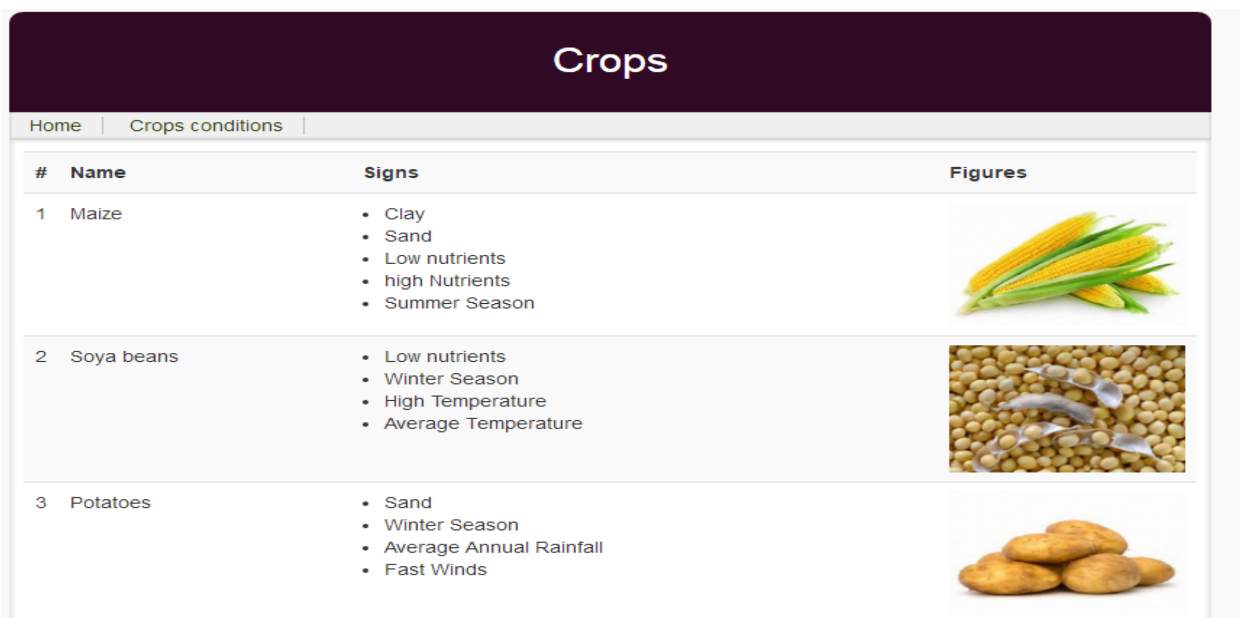


Fig 5.8 Crops display

Test Case 5: Crop Analysis

The crops will be analysed by comparing the conditions at a certain area by the favourable conditions of a certain crop, if the analyser returns 0% it means there is no match.

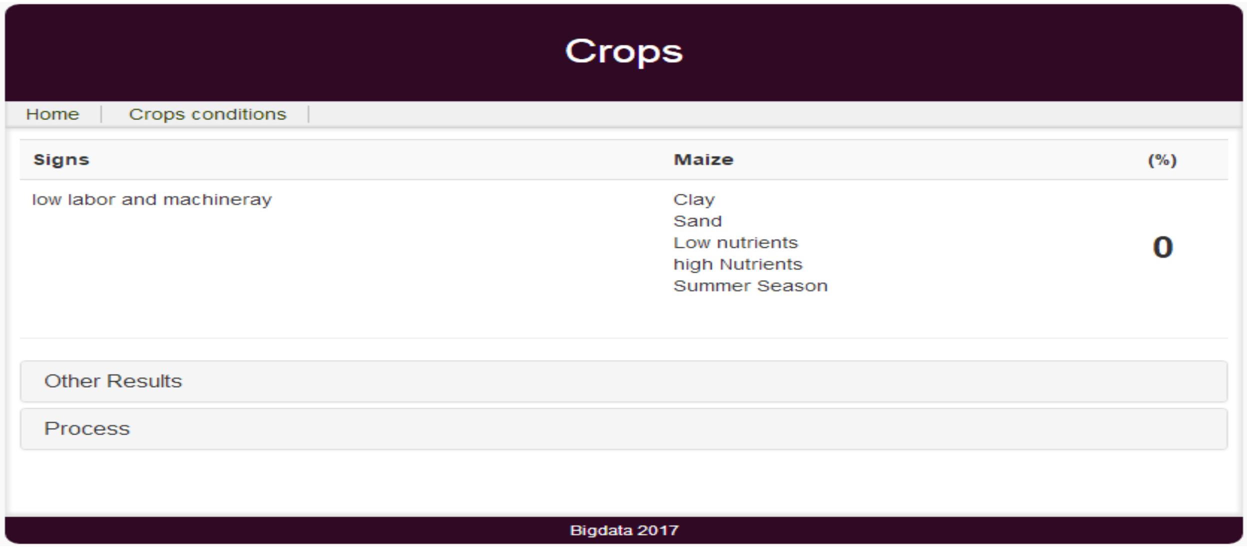


Fig 5.9 Crop analysis

Test case 6: Access level Testing

The users with different access level has different views.

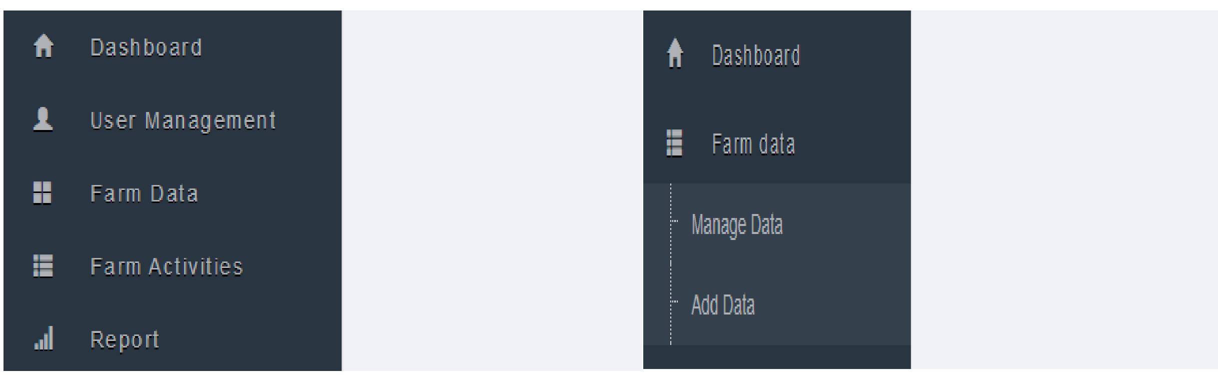


Fig 5.10 User Access Level Testing

5.4 Installation

The big data system was developed for any farm or agricultural company, the system is used whilst the manual system is being used for the trial period. The FileZilla and the apache tomcat server was used to install the big data system web portal and the web services. The web services allow the farm workers to access the system from anywhere. After the installation was successfully done, the farm data capturers started to capture the historical information about the farm and the details are synchronised to the big data database.

5.4.1 Steps in software installation process

The software installation process also encompasses the installation of the environment software required to support the developed system. In this case of big data system software (Birrell and Ould, 1999). The installation consists of the war file building and full production system deployment.

The following steps shows how the software is installed:

- Install Linux Operation system(Ubuntu)
- Open terminal to install php 5 and apache2
- Install the php myadmin and MySQL on the server for database management.
- Build the war file and install the tomcat server
- Restore the system database in linux by typing mysql -username and password ,database name </FolderWithSqlFile/databasename.sql
- Type the URL http: //serverip:8080/bigdatasystem
- Run the system and set the system properties, set the host IP address to http://bigdatasystem:8080 .

5.4.2 User training

Bently et al (1994) defined the user training as the process of familiarizing the system to the users highlighting the steps to use the system. For the big data system for agriculture software, the users that were trained include the farm administrator, the farm agronomist, data capturers and the farm owner. **Refer to Appendix A** for User Manual.

5.4.2.1 Training plan

The user training was conducted for the farm staff members and other system developers to enable them to know the way to troubleshoot if the system is down and to help the users facing problems during work time as well as system maintenance and further development. The following users were trained:

Table 5.1 Big data system Administrators training

Venue	Farm
Users	Farm administrator and system developers
Training scope	Administrator modules, system functionality, database backup, troubleshooting.
Requirements	Projector and a single computer
Trainer	System developer (Bradwin Zinyoni)

Table 5.2 Big data system Data capturers training

Venue	Farm
Users	Data capturers
Training scope	Data input modules, system functionality (limited view)
Requirements	Projector and a single computer
Trainer	System developer (Bradwin Zinyoni)

Table 5.3 Big data system Farm owner and agronomist training

Venue	Farm
Users	Farm agronomist, farm owner
Training scope	system functionality, report generation, exporting files.
Requirements	Projector and a single computer
Trainer	System developer (Bradwin Zinyoni) and the farm developers

5.4.3 System Conversion

Davis and Alan (1995) defined the system conversion as the process whereby the new system (Big data) will be deployed into production environment. After the training was conducted, the system operational environment was setup to install the new system. There are various ways that can be used to deploy the system which are pilot conversion, direct conversion, parallel conversion and phased conversation.

5.4.3.1 Direct Conversion

Direct conversion is the type of system conversion whereby the new system completely replaces the current system. The old system stops and the new system starts running (Davis and Alan, 1995). This type of conversion is very risky and it need intensive testing and training so that the users will be able to use the system without having problems. The farm will be using the complete big data system completely and the old system will stop operating. However, the direct conversion involves data loss of the farm and this method is very expensive. This method was not used by the researcher. The following diagram shows the direct conversion.

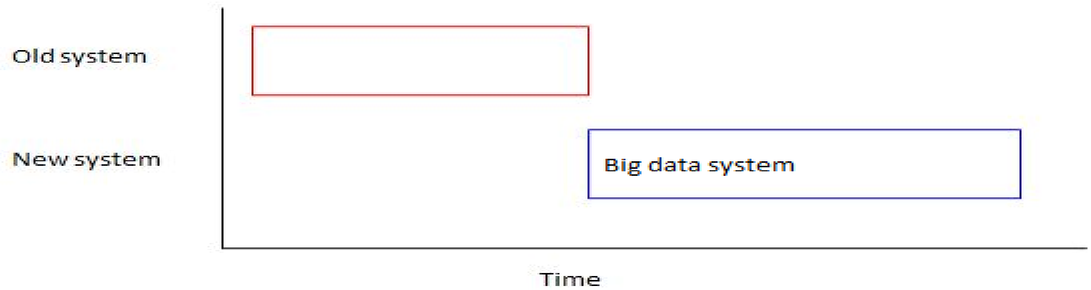


Fig 5.10 Direct Conversion

5.4.3.2 Pilot conversion

The type of system conversion, the new system operates with the data from one of the old system for the part of the system (Hoffer et al, 2006). This is done to enable comparison of the results with the old system. It is less risky than the direct conversion. This method builds confidence and the bugs are easily fixed without affecting the daily operations, however the researcher did not use method to deploy the system. The diagram below shows the pilot conversion:

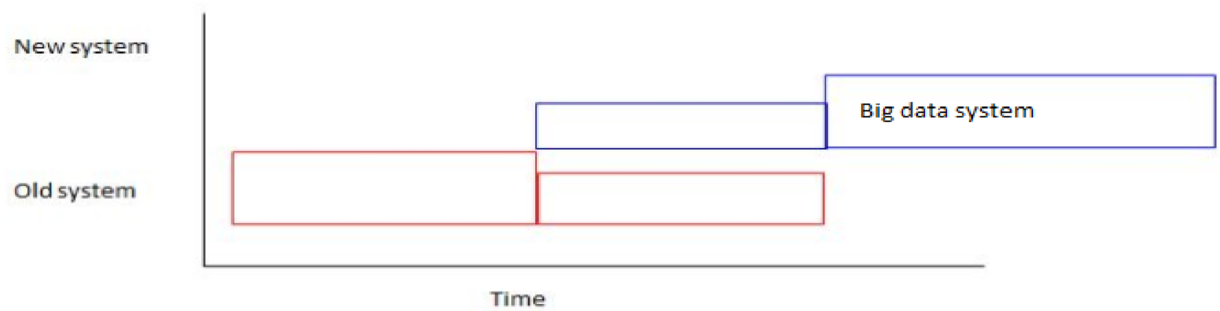


Fig 5.11 Pilot Conversion

5.4.3.3 Phased conversion

This type of system conversion, the new system is deployed in phases rather than full transformation once (Hoffer et al, 2006). The phases are installed one-step at a time until the full system is deployed. The conversion can be done as per department. The big data system was not deployed using this method. The following diagram shows the phased conversion:

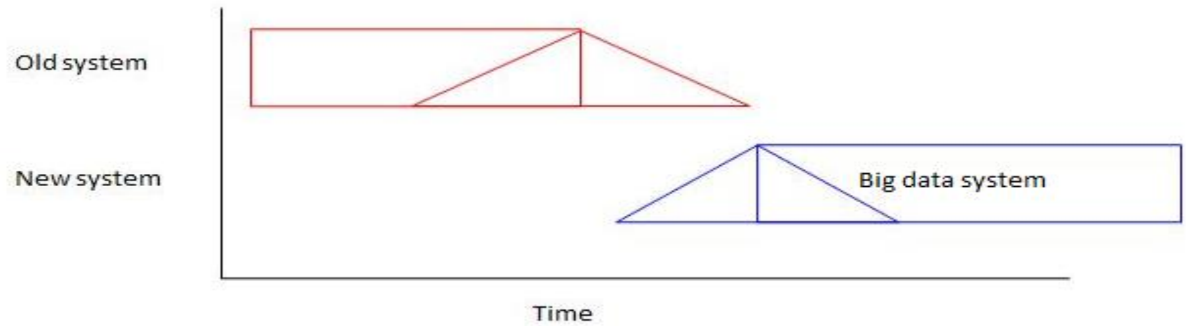


Fig 5.12Phased Conversion

5.4.3.4 Parallel conversion

Parallel conversion method enables the new system and the old system to run concurrently for a certain period. The data is processed by the both systems. This method of conversion is the best method because of the following reasons:

- The results from both systems can be compared.
- The early failure of the new system does not affect the daily operations.
- Workers can learn to use the new system step by step whilst using the old system.

The big data system was deployed using the parallel conversion method. The conversion process was done as shows by the diagram below:

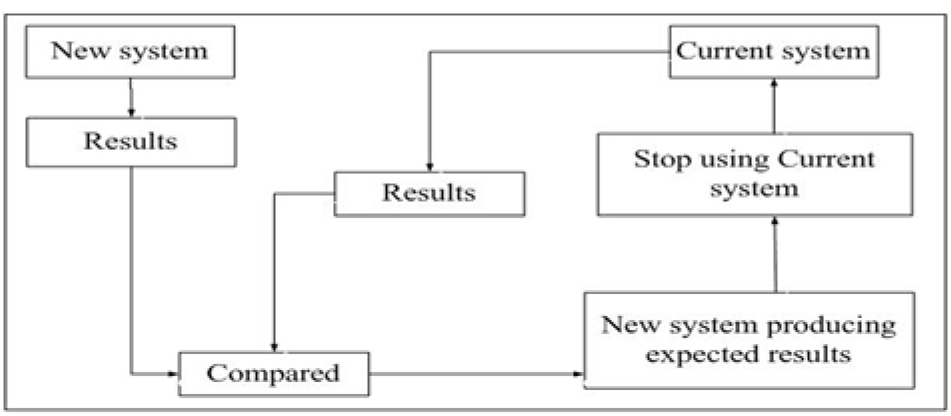


Fig 5.13 Parallel Conversion

5.4.3.5 Recommended conversion method

The big data system was deployed using the parallel conversion method to be able to compare the results of both systems and to avoid loss of data.

5.5 Maintenance

According to the system development life cycle the maintenance is the last stage of system development which is positive affected by everything happened in all the stages (Swanson, (1999). All the errors and mistakes made during the analysis and design phase can impact the maintenance. The system maintenance starts when the system is deployed into the production environment and continues till the system is changed. System maintenance is necessary to reduce errors in the system during system operation time and to tune the system into different working environments. Swanson (1999) suggested the following types of maintenance:

- Corrective maintenance
- Adaptive maintenance
- Perfective maintenance

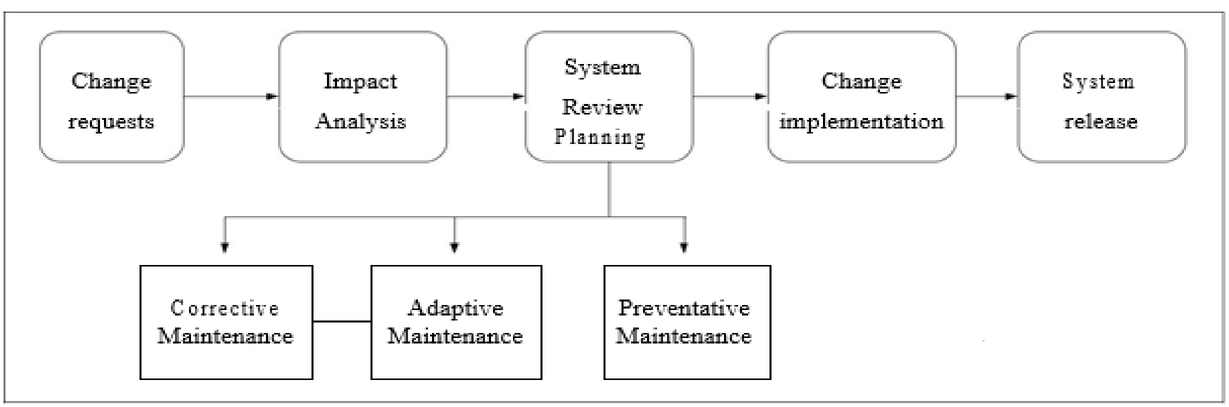


Fig 5.14 System maintenance

5.5.1 Corrective Maintenance

This is the most common type of maintenance which consists of repairing the defects in the design, coding or system implementation (Fournier, 1991) This maintenance method happens soon after the system deployment. Some of the defects happens during daily operations, they need to be resolved to continue with normal business activities. Corrective maintenance is done to fix the defects without adding new functionality

5.5.2 Adaptive maintenance

This type of maintenance consists of making the changes of system to evolve its functionality to business changes or migrating from a single operating environment to multi-user operating environment (West, 1999) This method of maintenance is usually less urgent than the corrective maintenance because the technical and business shift occurs over long period.

5.5.3 Perfective maintenance

Perfective maintenance consists of making enhancements to advance processing performance, the usability of interface or to add new interface (Fournier, 1991). The main reasons of this type of maintenance is to improve the system efficiency, response time and reliability. As the system data and operations increases the system performance start to degrade in efficiency, perfective maintenance might require to resolve all those challenges. The perfective maintenance will be scheduled after every two months by the IT department to improve the performance of big data system.

5.5.4 Recommended maintenance strategies

The researcher recommends the perfective maintenance to be carried out regularly to increase system performance.

5.6 Constraints

Big data system building was not that easy, the researcher faced lot of difficulties during the development process, some of the limitation faced are listed below:

- Financial problems- development budget was not able to cover everything required to complete the project. The software that were required were bit expensive, so the development team had to look for the alternatives so they could develop the system.
- Time factor- the system had to be developed in a very short time which made the researcher to develop the system under pressure.
- Unavailability of resources- The resources that were needed to develop the big data system were not readily available due to the service providers reasons, so the researcher had to find the alternatives.

5.7 Recommendations for future/further development

- The system database has to be backed up continuously as this help to recover data if the system fails.

- The system maintenance must be done after six months to prevent possible system fails and the maintenance must be done by inhouse developers.
- The researcher recommends that the developers should gather the requires for the mobile app development.
- The inhouse developers should use Apache Hadoop when they update the system in order to have better analysis.

5.7 Conclusion

All the phases of the system development were completed successfully. The big data system was successfully implemented and the users were well trained on how to use the system. The system was well tested, all the user requirements were met. The user documentation was given to the users and developers for continuous training and further development. The researcher hast listed some recommendations to be considered for further development of the big data framework for agricultural sector. All the processes were completed very well and the big data system were deployed to the production environment successfully.

References

Books

1. Bennatan, E. M., (1995), On Time, Within Budget Software Project Management Practices and Techniques, Wiley.

2. Birrell, N. D. and Ould, M.A., A Practical Handbook for Software Development, Cambridge University Press, Cambridge, UK, 1985.

3. Cavusoglu H, Mishra B and Raghunathan S, The Effect of Internet Security Breach, (2002) Univ.of Texas at Dallas.

4. Clifton, D. S. and Fyffe, Project Feasibility Analysis: A Guide to Profitable New Ventures, (1977), John Wiley & Sons, New York.

5. Conger, S Software Engineering (1994) Belmont California: Wadsworth, Inc

6. Davis C, Alan M, 201 Principles of Software Development, (1995), New York, NY: McGraw-Hill, Inc.

7. Fournier, R., Practical Guide to Structured System Development and Maintenance, Yourdon, Englewood Cliffs, NJ, 1991.

8. Gerald, V. (2005), Database Management System (3rd ed.), New York: McGraw Hill

9. Hoffer, J., George, J., & Valacich, J. (2006), Modern systems analysis and design 6th. Prentice Hall: U.S.A.

10. Hoglund G and McGraw G, Exploiting Software, (2004), Addison-Wesley.

11. Howard M and LaBlanc D, Writing Secure Code, (2003), Microsoft Press.

12. Jeffrey L Whitten and Lonnie D Bentley, (2001), System Analysis and Design Methods (7th Edition), Prentice Hall of India

13. Jones, H Handbook of team design (1998). New York McGraw-Hill

16. Kotonya, G. and Sommerville, I, (1998), Requirements Engineering: Processes and Techniques, UK: John Wiley and Sons.

17. Levin, Mark Sh., Composite Systems Decisions (2006), Springer, New York,

18. Maier, Mark W., and Rechtin, Eberhardt, The Art of Systems Architecting, Second Edition, (2000), CRC Press, Boca Raton.

19. McGraw G, "Software Security," IEEE Security & Privacy, (2004), Prentice Hall

20. Pierce, S Software System engineering: a first course (1992) Wilsonville, Oregon: Franklin, Beedle & Associates, Inc

21. Rajaraman V, (2004), Analysis and Design of Information Systems, Prentice Hall of India

22. Reutter, J., Maintenance is a management problem and a programmer's opportunity, AFIPS Conf. Proc. 1981 Natl. Comput.

23. Saltzer, J.H, et al., End-to-End arguments in Systems Design in: ACM Transactions in Computer Systems, (1984), Springer, New York,

24. Schach, S Classical and object oriented software engineering with UML & C++ 4th Edition (1999) New York McGraw- Hill.

25. Sindre G and Opdahl A.L, Technology of Object-Oriented Languages and Systems (2000), IEEE CS Press.

26. Steward, D V, (1987), Software engineering with systems analysis and design, Belmont California: Wadsworth, Inc

27. Swanson, E., (2004), The dimensions of maintenance, San Francisco.

28. Taggart W.M, and Silbey V, (1999), Information Systems: People and Computers in Organizations, Allyn and Bacon, Inc., Boston.

29. Tait, P. and Vessey, I. (1998), The Effect of User Involvement on System Success: A Contingency Approach. MIS Quarterly.

30. Ulrich, Karl T. and Eppinger, Steven D., Product Design and Development, Second Edition, (2000), Irwin McGraw-Hill, Boston,

31. Valacich, J.S., George, J. F., and Hoffer, J.A. (2009) Essentials of System Analysis and Design 4th Ed., Prentice Hall, Upper Saddle River, NJ.

32. Viega J and McGraw G, Building Secure Software, (2001), McGraw-Hill, Inc.

33. West Churchman C, the Design of Inquiring Systems: Basic Concepts of Systems and Organization (1999), Basic Books, New York.

33. Whitten, J. L., Bentley, L. D., and Barlow, V. M. (1994) Systems Analysis and Design Methods 3rd Ed. Richard D. Irwin, Inc., Burr Ridge, IL.

34. Whitten, J. L., Bentley, L. D., and Dittman, K. C. (2004) Systems Analysis and Design Methods 6th Ed., McGraw Hill Irwin, Boston.

35. Whitten, J. L., Bentley, L. D., and Ho, T.I.M. (1986) Systems Analysis and Design, Times Mirror/Mosby College Publishing, St. Louis.

36. Whitten, J.L., and Bentley, L.D. (2008) Introduction to Systems Analysis and Design 1st Ed. McGraw-Hill, Boston.

37. Wiegers, Karl E. (2003), Software Requirements (2nd ed.), Microsoft Press

38. William S. Davis, David C, (1998), The Information System Consultant's Handbook: Systems Analysis and Design by. Yen CRC Press, CRC Press LLC

Journals

1. Agricultural information systems http://dx.doi.org/10.1016/j.agsy Accessed 2/07/2017
2. Available at:http://softwarefeasibilitystudy.blogspot.com/2009/07/feasibility-study-software-engineering.html Accessed 25/06/17
3. Bogaardt, J.(2009)Agricultural Systems, Volume 153, 2017, 69 – 80 Accessed 13/08/2017
4. Dacom collaborate on precision farming solution to maximize yields and reduce costs. https://www.nec-enterprise.com/news/Latest-press-releases/NECand-Dacom-collaborate-on-precision-farming-solution-to-maximize-yields-andreduce-costs-791 Accessed: 7/05/ 2017.

5. Looking at Big Data One Plant at a Time. Farm Industry News. http://farmindustrynews.com/blog/looking-big-data-one-plant-time Accessed: 7/05/2017
6. Mohammad, R. (2006), Dilemma between the structured and object-oriented approaches to systems analysis and design. The Journal of computer information systems: 32-42. Accessed 15/08/2017
7. Poppe, K.J., Wolfert, J., Verdouw, C.N., Renwick, A., 2015. A European perspective on the economics of Big Data. Farm Policy Journal 12, 11 –19.
8. Sjaak Wolfert, Lan Ge, Cor Verdouw(2007) Big data for smart farming, Volume 140,2016,49-50 Accessed 29/05/2017
9. Software testing http://www.tutorialspoint.com/software_testing/index.htm http://www.tutorialspoint.com/software_testing/testing_iso_standards.htm Accessed 2/09/2017

Appendices

Appendix A: User Manual

The user manual was designed to familiarize the users with the necessary knowledge of the Big data analytics system software without any problems.

Getting Started

1. Go to **http://localhost:8080/bigagric/**

2. Enter your **Username and Password**

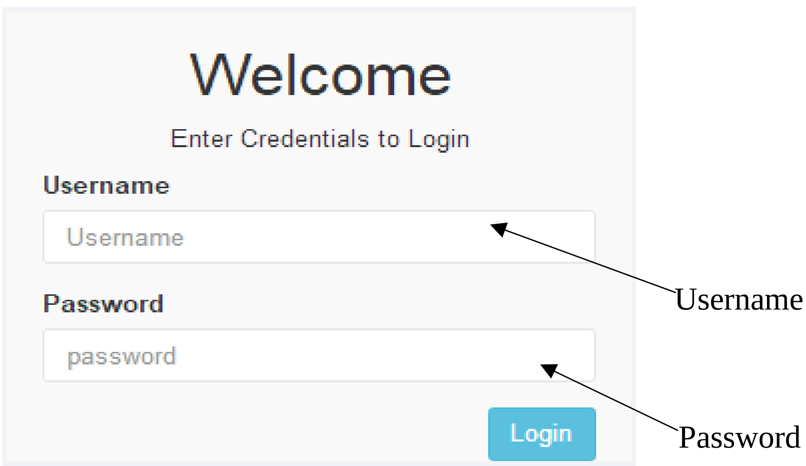


Fig A1 Login Page

3. Click Login Button
4. The following screen should appear if Login details are correct:

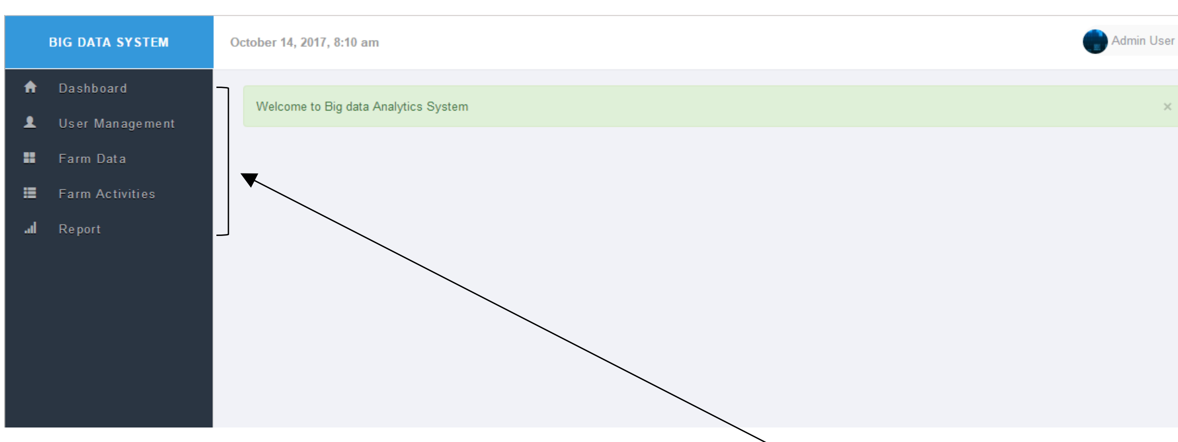


Fig A2 Main Menu

The Admin user can add new user

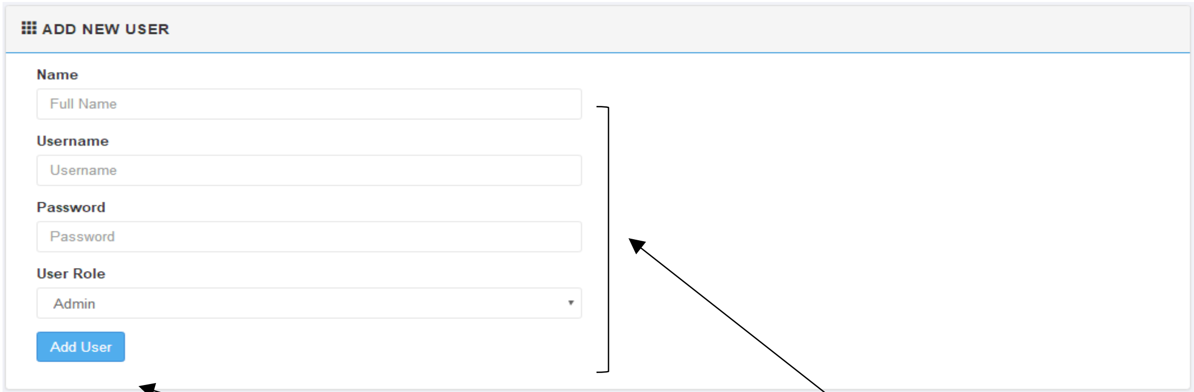
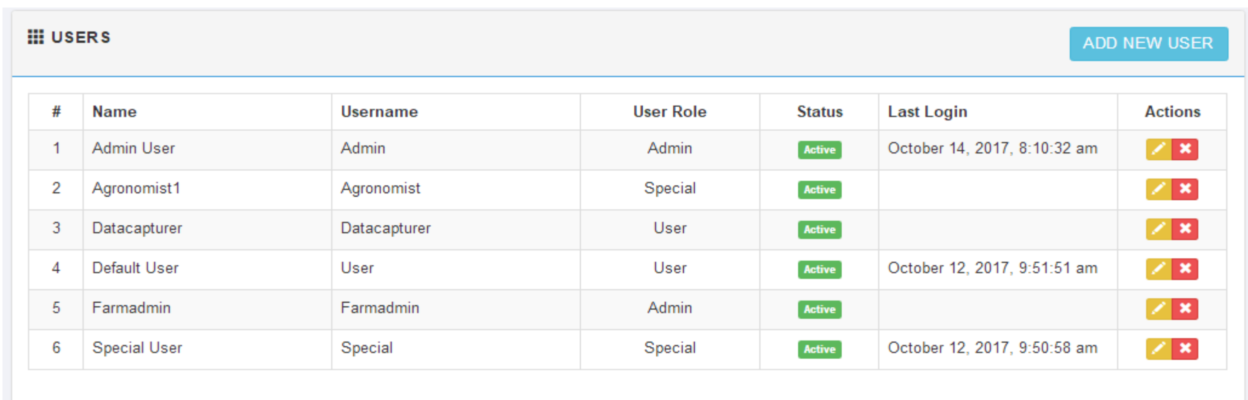


Fig A3 Adding New user Click Add user button Fill in the fields

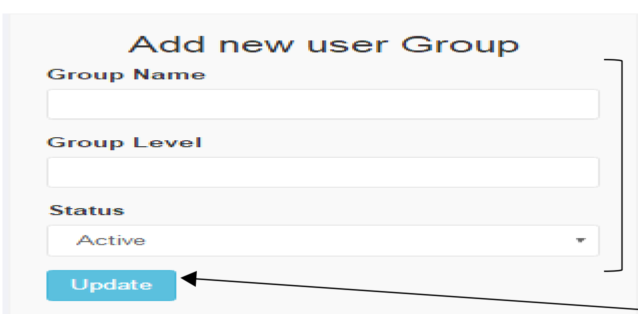
The admin user can manage users



#	Name	Username	User Role	Status	Last Login	Actions
1	Admin User	Admin	Admin	Active	October 14, 2017, 8:10:32 am	
2	Agronomist1	Agronomist	Special	Active		
3	Datacapturer	Datacapturer	User	Active		
4	Default User	User	User	Active	October 12, 2017, 9:51:51 am	
5	Farmadmin	Farmadmin	Admin	Active		
6	Special User	Special	Special	Active	October 12, 2017, 9:50:58 am	

Fig A4 Manage users

Admin can add new group

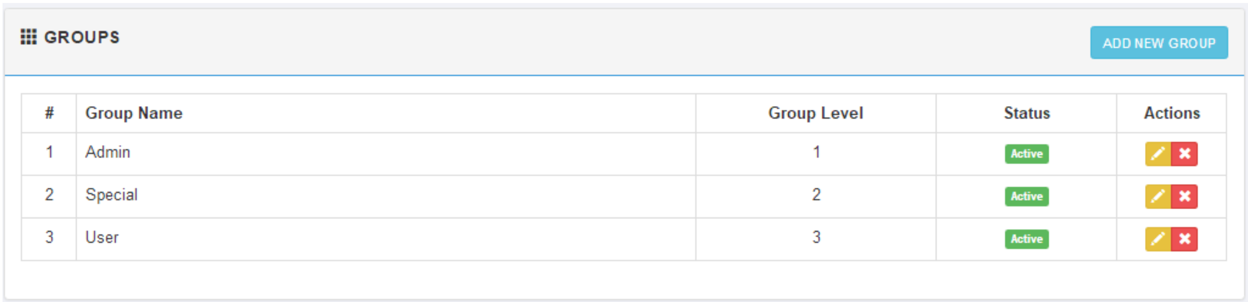


Fill in the fields

Click update button

Fig A5 Adding New user group

Admin can manage user group levels



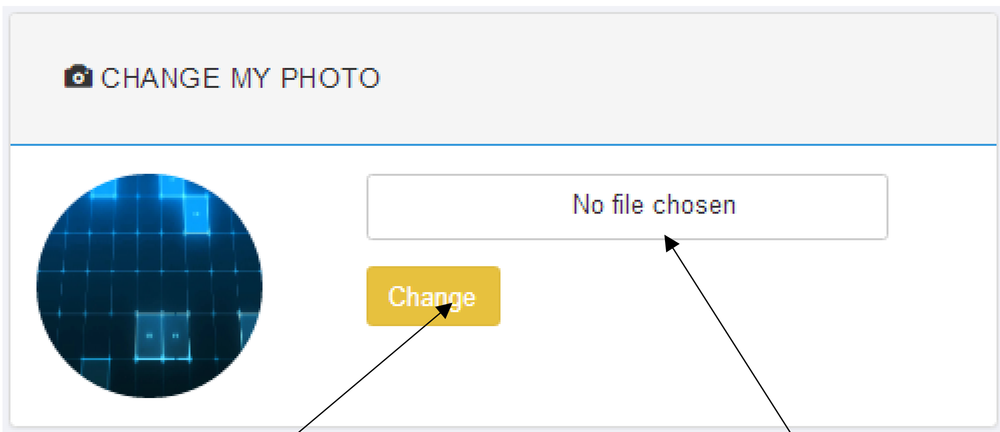
GROUPS

ADD NEW GROUP

#	Group Name	Group Level	Status	Actions
1	Admin	1	Active	
2	Special	2	Active	
3	User	3	Active	

Fig A6 Manage User Group Levels

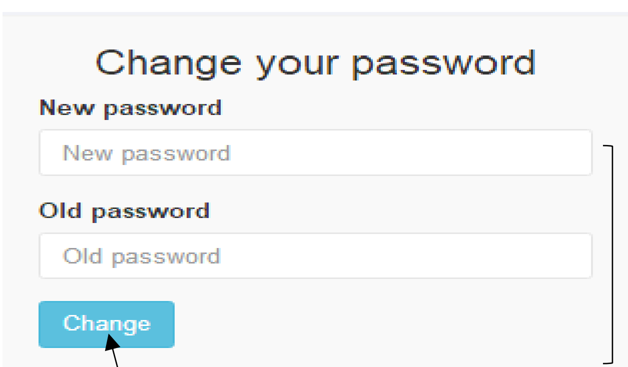
User can change profile picture by choosing the new picture in his/her computer



Click button Change to confirm   Click here to choose photo

Fig A7 Change profile picture

User can change password

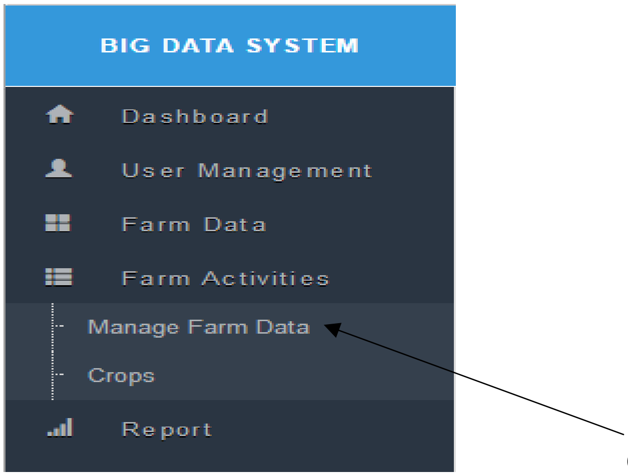


Input the new password

Click button to confirm change

Fig A8 Change password

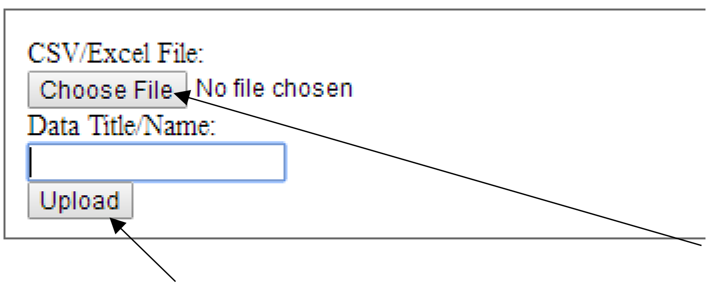
Click manage farm data to add, upload, add and delete data



Click here to add, upload, add and delete data

Fig A9 Manage Farm data

Upload data from excel sheet

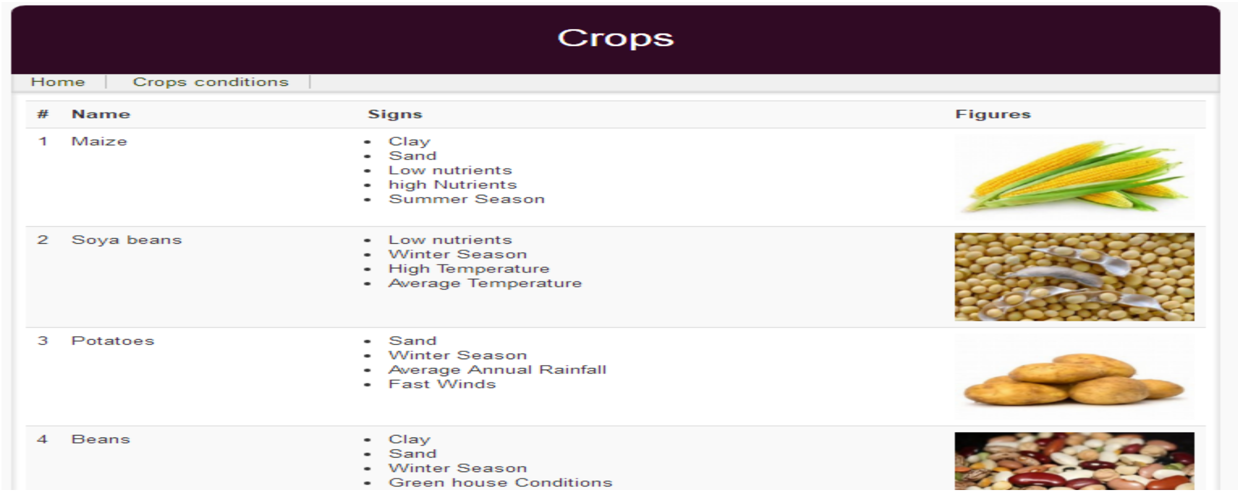


Select excel file from computer

Click button to upload

Fig A10 Upload data from excel sheet

Viewing the crop favourable conditions in the data base by clicking crop conditions



Crops

Home | Crops conditions

#	Name	Signs	Figures
1	Maize	- Clay - Sand - Low nutrients - high Nutrients - Summer Season	
2	Soya beans	- Low nutrients - Winter Season - High Temperature - Average Temperature	
3	Potatoes	- Sand - Winter Season - Average Annual Rainfall - Fast Winds	
4	Beans	- Clay - Sand - Winter Season - Green house Conditions	

Fig A11 Crop conditions

Crop analyser enable the agronomist to access the crops before they grow them by comparing the favourable conditions and the farm condition.

Crops

Home | Crops conditions

#	Signs
1	Clay
2	Sand
3	Low Rain fall
4	High Rainfall
5	Low nutrients
6	high Nutrients
7	Summer Season
8	Winter Season
9	Green house Conditions
10	High Temperature
11	low temperature

Fig A12 Crop analyser

The crops analyser report will show the matching percentage by comping the favourable conditions and the farm conditions

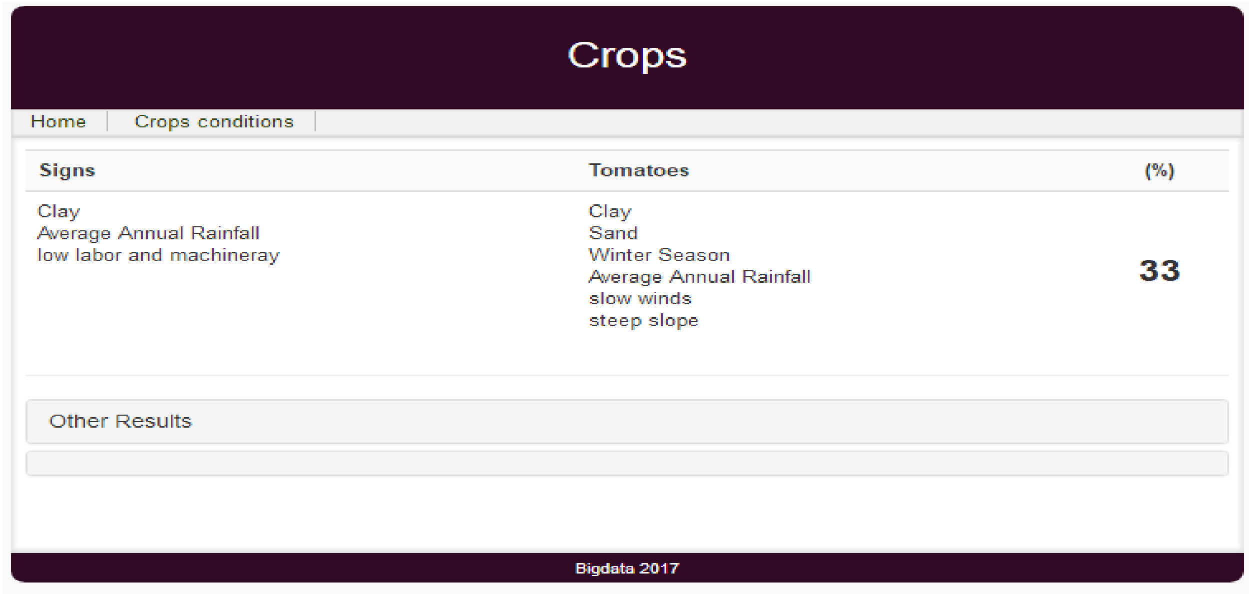


Fig A13 Crop analyser report

After harvesting the crops, the data capturer enters the data into the data base so the agronomist and the arm owner can analyse

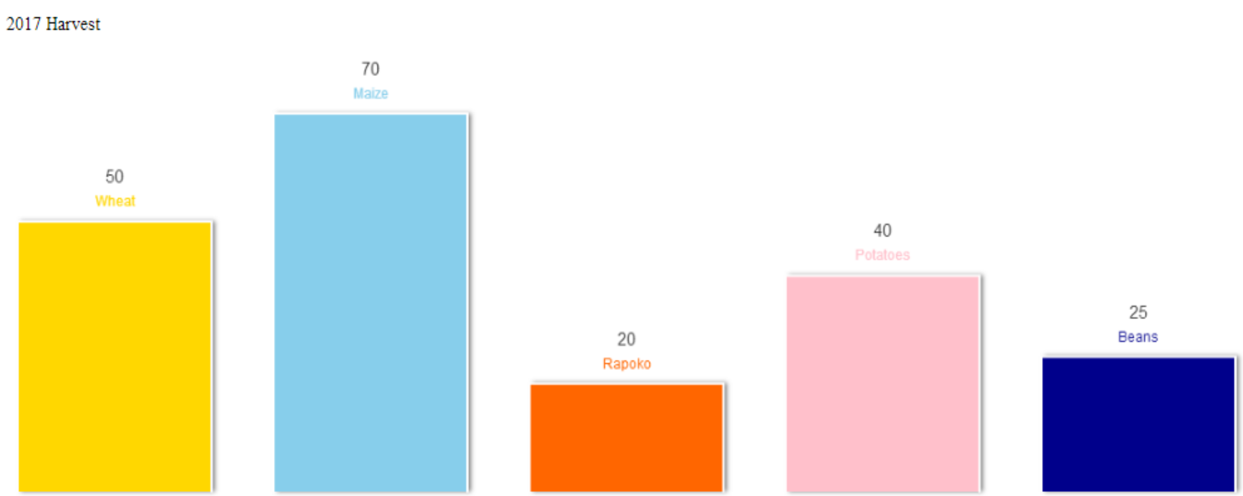


Fig A14 Harvest report

Appendix B: Observation Score Sheet

Location:……………………………………………………………………

Process Being Observed:…………………………………………………

Date: ……………   Time: ……………

Objective of observation:

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

Brief description of current system Process:

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

Areas of strength:

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

Areas that need development:

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

Signed: ………………………………   Date: ………………………………

Signed: ………………………………   Date: ……………………………..

Appendix C: Snippet of Code
Code to Connect to the Database

```
<?php
require_once(LIB_PATH_INC.DS."config.php");


class MySqli_DB {


   private $con;
   public $query_id;


   function __construct() {
     $this->db_connect();
   }



public function db_connect()
{
  $this->con = mysqli_connect(DB_HOST,DB_USER,DB_PASS);
  if(!$this->con)
         {
           die(" Database connection failed:". mysqli_connect_error());
         } else {
           $select_db = $this->con->select_db(DB_NAME);
             if(!$select_db)
             {
```

```
                die("Failed to Select Database". mysqli_connect_error());
            }
        }
}



public function db_disconnect()
{
  if(isset($this->con))
  {
    mysqli_close($this->con);
    unset($this->con);
  }
}
```

Code for Harvest Analyser

```
<?php
mysql_select_db('bigagric',mysql_connect('localhost','root',''))or die(mysql_error());
?>
```

Code for Crop Analyzer

```
/*      Get the name of favourable_conditions from Database        */
            $sql = "SELECT name FROM `favourable_conditions` WHERE id IN (".implode(', ', $_POST['favourable_conditions']).")";
            $res = mysql_query($sql) or die(header('Location: '.$_SERVER['HTTP_REFERER']));
```

```
if(mysql_num_rows($res) > 0){
	while($row = mysql_fetch_array($res)){
		/*	Store the name of favourable_conditions to the Array
	*/
		$signs[] = $row['name'];
	}

	/*	Formatting the name of favourable_conditions from Array	*/
	$asigns = '<ul class="list-unstyled">';
	foreach($signs as $val){
		$asigns .= '<li>'.$val.'</li>';
	}
	$asigns .= '</ul>';

	/*	Get the name of Disease from Database		*/
	$sql = "SELECT * FROM `crops`";
	$res = mysql_query($sql) or die(mysql_error());

	if(mysql_num_rows($res) > 0){
		while($row = mysql_fetch_array($res)){
			/*	Get the name of favourable_conditions of Disease
from database */
			$sql0 = "SELECT name FROM `favourable_conditions`
WHERE id IN (".implode(', ', explode(' ', $row['signs'])).")";
			$res0 = mysql_query($sql0) or die(mysql_error());
```