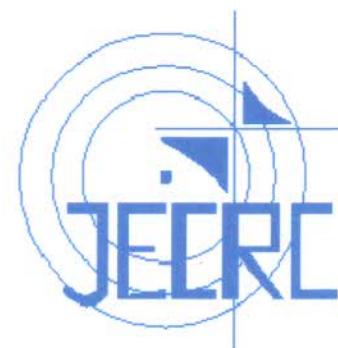


A
Project Report
On
“Back-door”
Submitted
In partial fulfillment of the requirement for the award of degree of
Bachelor of Technology
In
Information Technology



(Session 2017-2018)

Submitted to:

Mr. Sunil Kumar Jangir
HOD, IT

Mr. Naveen Kedia
Project Coordinator, IT

Mr. Sunil Kumar Jangir
Project Guide, IT

Submitted by:

Prashant Manghnani
[14EJCIT055]
VIII Semester

Department of Information Technology
Jaipur Engineering College & Research Centre, Jaipur
Rajasthan Technical University, Kota
May 2018

Table of Content

S.No.	TITLE	PAGE NO.
1	Certificate	i
2	Candidate's Declaration	ii
3	Acknowledgement	iii
4	Abstract	iv
5	List Of Chapter	v
6	List Of Tables	vii
7	List Of Figures	viii
8	Future Scope	ix
9	Bibliography	x
10	Conclusion	xi

RAJASTHAN TECHNICAL UNIVERSITY, KOTA



CERTIFICATE

*This is to certify that the work, which is being presented in the project entitled “Back-door” submitted by **Mr. Prashant Manghnani**, the student of fourth year (VIII Semester) B.Tech in Information Technology in partial fulfillment for the award of degree of Bachelor of Technology is a record of student's work carried out and found satisfactory for submission.*

Mr. Naveen Kedia
Project Coordinator
Assistant Professor, IT

Mr. Sunil Kumar Jangir
HOD, IT

Mr. Sunil Kumar Jangir
Project Guide
HOD, IT

CANDIDATE'S DECLARATION

I hereby declare that the work, which is being presented in the Project Stage II entitled "*Back-door*" in partial fulfillment for the award of Degree of "Bachelor of Technology" in Information Technology, and submitted to the **Department of Information Technology**, Jaipur Engineering College and Research Centre, Affiliated to Rajasthan Technical University is a record of my own work carried out under the Guidance of **Mr. Naveen Kedia** and **Mr. Sunil Kumar Jangir**, Assistant Professor and Professor of Department of Information Technology.



PRASHANT MANGHNANI
(14EJCIT055)

ACKNOWLEDGEMENT

I wish to express my deep sense of gratitude to my Project Coordinator **Mr. Naveen Kedia** and Project Guide **Mr. Sunil Kumar Jangir**, Jaipur Engineering College and Research Centre, Jaipur for guiding me from the inception till the completion of the project. I sincerely acknowledge them for giving their valuable guidance, support for literature survey, critical reviews and comments for our Project.

I would like to first of all express my thanks to **Mr. Arpit Agrawal**, Director of JECRC, for providing me such a great infrastructure and environment for my overall development.

I express sincere thanks to **Dr. V. K. Chandna**, Principal of JECRC, for his kind cooperation and extendible support towards the completion of our project.

Words are inadequate in offering my thanks to **Mr. Sunil Kumar Jangir**, HOD of Department of Information Technology, for consistent encouragement and support for shaping my project in the presentable form.

I also like to express my thanks to all supporting IT faculty members who have been a constant source of encouragement for successful completion of the project.

Also my warm thanks to **Jaipur Engineering College and Research Centre**, who provided me this opportunity to carry out this prestigious Project and enhance my learning in various technical fields.



PRASHANT MANGHNANI
(14EJCIT055)

ABSTRACT

Back-door works on two concepts: -

- Distributed storage
- Distributed Computing.

Distributed storage: For this we use HDFS (Hadoop distributed file system) in which there is a master node called namenode which controls all the other nodes of the network. Those other nodes are called datanodes. Now what happens, the datanodes will give some amount of their storage space to namenode thus creating an virtual HDD and the namenode which is only going to interact with the client, will show that it is having a storage space of nearly 1-2TB. Now what happens when a client uploads a data of 200 GB on this cluster the namenode breaks the data into blocks of equal sizes and then stores each block on one datanode and then 2-3 replications of that particular block on different datanodes.

Distributed Computing: Now when we have stored the data we need to perform some computation on that data. For this we use map reduce technique of hadoop. In this there is a jobtracker to which various tasktrackers are connected. Now when client tells the jobtracker to do some job on the data stored in hdfs, then the job tracker contacts with the namenode for the location of the datanodes and then assigns jobs to tasktrackers to perform jobs on the data stored.

LIST OF CHAPTER

S.No.	TITLE	PAGE NO.
1.	INTRODUCTION	
	1.1 Proposed system	1
	1.2 Purpose of the project	1
	1.3 Problem in existing system	2
	1.4 Solution to the problem	3
2.	REQUIREMENT ANALYSIS	
	2.1 System Requirement	4
	2.2 Need of the system	5
	2.3 Scope of the project	5
3.	FEASIBILITY STUDY	
	3.1 Operational Feasibility	6
	3.2 Economical Feasibility	6
	3.3 Technical Feasibility	7
4.	TECHNOLOGY USED	
	4.1 Characteristics of Big Data	8
5.	PROJECT ANALYSIS	
	5.1 Study of the system	10
	5.2 Input and output	10
	5.3 Process model used with justification	11
6.	UML DIAGRAMS	
	6.1 Use case diagrams	12
	6.2 Activity Diagram	14
	6.3 Deployment diagrams	16

7.	TESTING METHODOLOGY	
	7.1 Testing	18
	7.2 Validation and verification	18
	7.3 Test case design	21
8.	OUTPUTS	24
9.	Mapping with PO'S	29

LIST OF TABLES

S.No.	Table Name	PAGE NO.
1	Table1- Software Requirements	4
2	Table2-Login: Test Case 1	21
3	Table3- Login: Test Case 2	21
4	Table4- Sign up: Test Case 3	22
5	Table 5- Sign up: Test Case 4	22

LIST OF FIGURE

S.No.	Figure Name	PAGE NO.
1	Figure 1- Use case diagram	13
2	Figure 2-Activity diagram	15
3	Figure 3- Deploymentdiagram	17
4	Figure 4- Verification and validation model	20
5	Figure 5- Login page	24
6	Figure 6- Access page	25
7	Figure 7- Configuration page	26
8	Figure 8- Automatic Setup	27
9	Figure 9- Manual Setup	28

LIST OF FIGURE

S.No.	Figure Name	PAGE NO.
1	Figure 1- Use case diagram	13
2	Figure 2-Activity diagram	15
3	Figure 3- Deploymentdiagram	17
4	Figure 4- Verification and validation model	20
5	Figure 5- Login page	24
6	Figure 6- Access page	25
7	Figure 7- Configuration page	26
8	Figure 8- Automatic Setup	27
9	Figure 9- Manual Setup	28

FUTURE SCOPE

Today, Big Data is influencing IT industry few technologies have done before. The massive data generated from sensor-enabled machines, mobile devices, cloud computing, social media, satellites help different organizations improve their decision making and take their business to another level.

“Big data absolutely has the potential to change the way governments, organizations and academic institutions conduct business and make discoveries and its likely to change

- Backup Name node
- Advertisement portal

CHAPTER-1

INTRODUCTION

1.1 PROPOSED SYSTEM

Back-door works on two concepts:

- Distributed storage
- Distributed Computing.

Distributed storage: For this we use HDFS (Hadoop distributed file system) in which there is a master node called namenode which controls all the other nodes of the network. Those other nodes are called datanodes. Now what happens, the datanodes will give some amount of their storage space to namenode thus creating an virtual HDD and the namenode which is only going to interact with the client, will show that it is having a storage space of nearly 1-2TB. Now what happens when a client uploads a data of 200 GB on this cluster the namenode breaks the data into blocks of equal sizes and then stores each block on one datanode and then 2-3 replications of that particular block on different datanodes.

Distributed Computing: Now when we have stored the data we need to perform some computation on that data. For this we use map reduce technique of hadoop. In this there is a jobtracker to which various tasktrackers are connected. Now when client tells the jobtracker to do some job on the data stored in hdfs, then the job tracker contacts with the namenode for the location of the datanodes and then assigns jobs to tasktrackers to perform jobs on the data stored.

1.2 PURPOSE OF THE PROJECT

Today, we're surrounded by data. People upload videos, take pictures on their cell phones, text friends, update their Facebook status, leave comments around the web, click on ads, and so forth. Machines, too, are generating and keeping more and more data.

The exponential growth of data first presented challenges to cutting-edge businesses such as Google, Yahoo, Amazon, and Microsoft. They needed to go through terabytes and petabytes of data to figure out which websites were popular, what books were in demand, and what kinds of

ads appealed to people. Existing tools were becoming inadequate to process such large data sets. Google was the first to publicize *MapReduce*—a system they had used to scale their data processing needs.

This system aroused a lot of interest because many other businesses were facing similar scaling challenges, and it wasn't feasible for everyone to reinvent their own proprietary tool. Doug Cutting saw an opportunity and led the charge to develop an open source version of this MapReduce system called Hadoop. Soon after, Yahoo and others rallied around to support this effort. Today, Hadoop is a core part of the computing infrastructure for many web companies, such as Yahoo, Facebook, LinkedIn, and Twitter. Many more traditional businesses, such as media and telecom, are beginning to adopt this system too.

Hadoop is an open source framework for writing and running distributed applications that process large amounts of data. Distributed computing is a wide and varied field, but the key distinctions of Hadoop are that it is

- **Accessible**—Hadoop runs on large clusters of commodity machines or on cloud computing services such as Amazon's Elastic Compute Cloud (EC2).
- **Robust**—Because it is intended to run on commodity hardware, Hadoop is architected with the assumption of frequent hardware malfunctions. It can gracefully handle most such failures.
- **Simple**—Hadoop allows users to quickly write efficient parallel code.

Hadoop's accessibility and simplicity give it an edge over writing and running large distributed programs. Even college students can quickly and cheaply create their own Hadoop cluster. On the other hand, its robustness and scalability make it suitable for even the most demanding jobs at Yahoo and Facebook.

1.3 PROBLEM IN EXISTING SYSTEM

A well thought out and executed Big Data and analytics strategy ultimately makes organizations smarter and more efficient. Today, Big Data is being leveraged in many industries from criminal justice to health care to real estate with powerful outcomes. The same common sense approach to Big Data should be employed by organizations desiring similar results.

As most organizations will agree (if we're honest with ourselves), it's simply not possible to carry out the conversations we once had with our customers. There's too much dialogue coming in from various sources. We need help. In Forrester's thought leadership paper, "Use Behavioral Marketing To Up The Ante In The Age Of The Customer," they note that building "the technical infrastructure to support dynamic, cross-channel conversations with customers" is absolutely necessary for organizational impact.

What is that inevitable future? The digitization of all customer-facing organizational systems from customer service to sales to marketing.

The iConsumer report makes an interesting and noteworthy case for why structural changes within organizations (related to Big Data) are necessary now as reversals are likely to come. Reversals, MGI notes, come gradually until they come suddenly, interrupting "life as we know it." They cite two interesting reversals. The first reversal was in the newspaper industry that moved from booming to near obsolete with the advent of online publishing. This happened within a decade. The second reversal was in the recording/music industry that moved from booming CD sales to obsolete (CD sales) with the advent of digital music. This also happened within a decade. Both reversals were gradual until they were sudden.

These are both great examples of the gradual takeover that Big Data management tools are having.

1.3 SOLUTION TO THE PROBLEM

The aim of this project is to provide a platform with the use of web UI(User Interface), Which provide configured cluster which is automated by the technology Ansible. It is created automatically on a single click.

Nextly, project is developed as an automation tool which creates cluster plus adds facilities of MapReduce by providing a set of data for analysis. This will hide complexities for a normal user who can solve BigData problems without getting into ambiguous configuration of clusters.

The "Back-Door Analyzer" Portal is a portal which provides a platform where a client can perform analysis of bigdata rather than worrying about the resources and servers it needs to have.

CHAPTER-2

REQUIREMENT ANALYSIS

2.1 SYSTEM REQUIREMENT

The process of deciding on the requirement of a software system, which determines the responsibilities of a system, is called requirement analysis. Requirement analysis is a software engineering task that bridges the gap between system level requirements engineering and software design. Requirement engineering activities result in the specification of software's operational characteristics, indicate the software's interface with other system elements and establish constraints that the software must meet. The following section presents the detailed requirement analysis of our project.

2.1.1 HARDWARE REQUIREMENT:

The system needs to have at least:

- 1) 6 GB RAM
- 2) 64-bit OS

2.1.2 SOFTWARE REQUIREMENT:

Software used	Description
Operating system	We have chosen Linux operating system for its best support and user-friendliness.
Java	For installation of hadoop package
Python	To implement the project we have chosen python language for its more interactive support.

2.2 NEED OF THE SYSTEM

1. To use existing computer network:

The system has a capability to use existing computer network for voice and text communication.

2. Reliable communication:

To provide more reliable communication through text chat between users.

3. Portable application:

Robust application is needed which portably used for voice and text communication for diverse computer network.

2.3 SCOPE OF THE PROJECT

The “BackDoor” Portal is a portal which provides a platform where a client can perform analysis of bigdata rather than worrying about the resources and servers it needs to have.

As the Data is increasing day by day, with it comes the responsibility of handling the data and properly analyzing the data so that useful information can be mined out and more researches can be carried out. These force corporates to find a solution for it and Apache Hadoop is the ultimate and reliable solution they can get.

Apache Hadoop provides not only HDFS (Hadoop distributed File System) but also processing clusters using MapReduce making it not only highly storable but also highly processing.

CHAPTER-3

FEASIBILITY STUDY

Feasibility is the measure of how beneficial the development of an information system would be to an organization. Important outcome of the Preliminary investigation is determination of whether the system is feasible or not.

There are three aspects of feasibility study:

- Operational Feasibility.
- Economical Feasibility.
- Technical Feasibility.

3.1 Operational Feasibility:

The developed system must be simple to use so that there should not be any confusion while operating it.

System helps in interaction with user while determining various requirements. It is user-friendly with appropriate error messages to help. The system offers validations & verifications.

The training is not required to operate the system in user's environment. One can easily work with the new system. In this way, the system is operationally feasible.

3.2 Economical Feasibility:

This is very important for considering the cost overheads while implementing the system. The cost overheads include Software cost, Hardware cost, operating cost etc.

The proposed system would provide right type of information, at right time, at right place & in right format. This will save time in decision making. The system is economically suited to the organization & the companies will have substantial benefits by the system. As the project is built on the platform which is regularly used by everyone, so there is no need to spend more amount of money on it.

Considering all the benefits, the value of the proposed system outweighs the cost of development of the system. Hence it is economical.

3.3 Technical Feasibility:

Technical Feasibility is usually raised during feasibility stage of investigation. Project will provide user friendly approach & so any person with basic computer knowledge will be able to handle the system easily. The system is technically feasible, as software and hardware are both easy available by performing proper installation. User friendly features added to improve the performance of the system.

CHAPTER-4

TECHNOLOGY USED

Big data is a term for data sets that are so large or complex that traditional data processing applications are inadequate to deal with them. Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, querying, updating and information privacy. The term "big data" often refers simply to the use of predictive analytics, user behavior analytics, or certain other advanced data analytics methods that extract value from data, and seldom to a particular size of data set. "There is little doubt that the quantities of data now available are indeed large, but that's not the most relevant characteristic of this new data ecosystem."

4.1 Characteristics of Big Data

- **Volume:** - Big data implies enormous volumes of data. It used to be employees created data. Now that data is generated by machines, networks and human interaction on systems like social media the volume of data to be analyzed is massive. Yet, Inderpal states that the volume of data is not as much the problem as other V's like veracity.
- **Variety:** - Variety refers to the many sources and types of data both structured and unstructured. We used to store data from sources like spreadsheets and databases. Now data comes in the form of emails, photos, videos, monitoring devices, PDFs, audio, etc. This variety of unstructured data creates problems for storage, mining and analyzing data. Jeff Veis, VP Solutions at HP Autonomy presented how HP is helping organizations deal with big challenges including data variety.
- **Velocity:** - Big Data Velocity deals with the pace at which data flows in from sources like business processes, machines, networks and human interaction with things like social media sites, mobile devices, etc. The flow of data is massive and continuous. This real-time data can help researchers and businesses make valuable decisions that provide

strategic competitive advantages and ROI if you are able to handle the velocity. Inderpal suggest that sampling data can help deal with issues like volume and velocity.

- **Veracity:** - Big Data Veracity refers to the biases, noise and abnormality in data. Is the data that is being stored, and mined meaningful to the problem being analyzed. Inderpal feel veracity in data analysis is the biggest challenge when compares to things like volume and velocity. In scoping out your big data strategy you need to have your team and partners work to help keep your data clean and processes to keep ‘dirty data’ from accumulating in your systems.
- **Validity:** - Like big data veracity is the issue of validity meaning is the data correct and accurate for the intended use. Clearly valid data is key to making the right decisions. Phil Francisco, VP of Product Management from IBM spoke about IBM’s big data strategy and tools they offer to help with data veracity and validity.
- **Volatility:** - Big data volatility refers to how long is data valid and how long should it be stored. In this world of real time data you need to determine at what point is data no longer relevant to the current analysis.

CHAPTER-5

PROJECT ANALYSIS

5.1 STUDY OF THE SYSTEM

Hadoop framework allows the user to quickly write and test distributed systems. It is efficient, and it automatically distributes the data and work across the machines and in turn, utilizes the underlying parallelism of the CPU cores.

Servers can be added or removed from the cluster dynamically and hadoop continues to operate without interruption. Hadoop is compatible on all platforms since it is Java based.

In this project we are providing a platform for the big data problems. For the easy usage of clients we have done automation by hiding the complexity getting into the understanding of configuration of clusters and mapReduce.

The first phase of project deals with logging in a web portal with specific username and password. The later part concerns about automatic configuration of following:

- Namenode
- Datanodes
- Jobtracker
- Tasktrackers

At last, we provide a set of sample data to analyze it using available java programs.

Else user can develop his/her own program in form of .jar file to use it for analysis, but this part is done manually by running some commands.

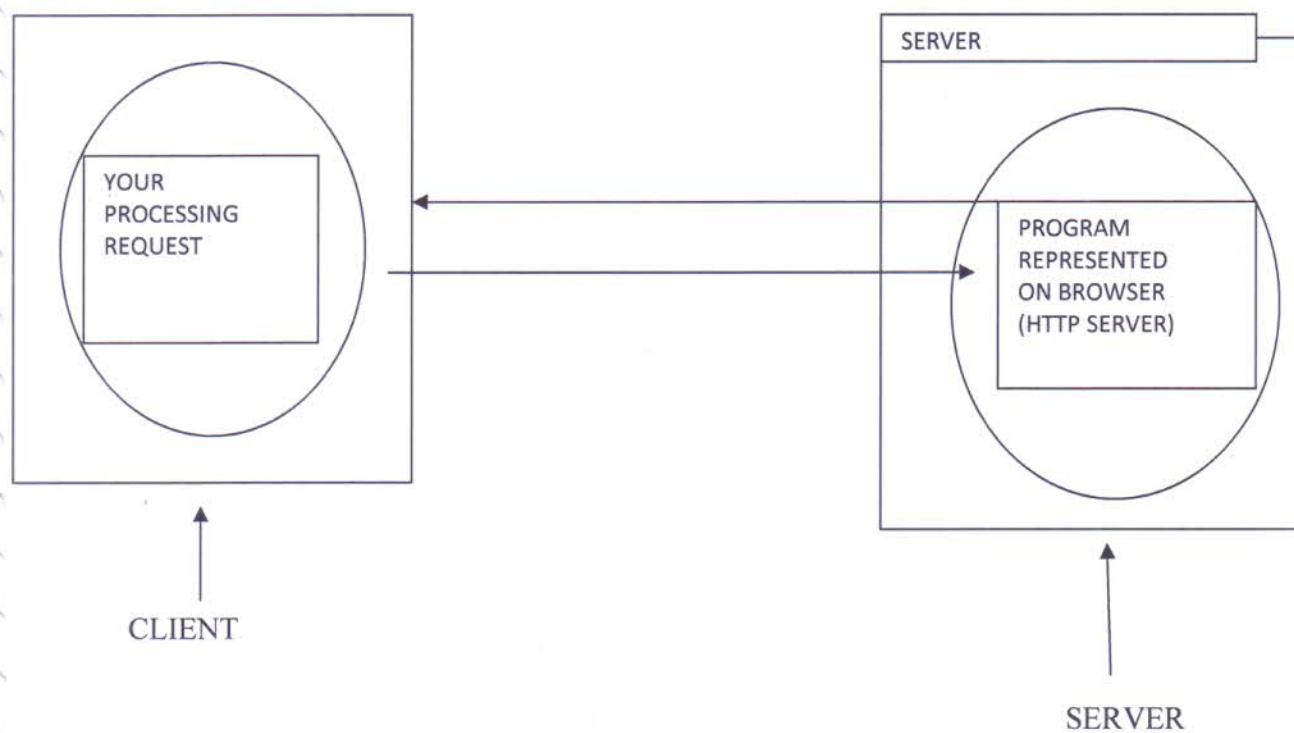
5.2 INPUT AND OUTPUT:

The user has to provide to which server it has to connect. In this, the server name to be provided is local host.

The user has to provide the username and the password for proceeding with the server for project access.

5.3 PROCESS MODEL USED WITH JUSTIFICATION

The model used here is a SPIRAL MODEL. This Model demands a direct consideration of technical risk at all stages of the project and if properly applied it reduces risk before they become problematic, hence it becomes easier to handle a project when using this kind of model where in the end user can evaluate the program at the end of each stage and suggest modification if required. In this way the Risk Management of Project is carried out efficiently through Spiral model.



This is your program

This program by URL

CHAPTER-6

UML DIAGRAMS

6.1 USE CASE DIAGRAMS

Use-cases model the system from the end-user's point of view. Created during requirements elicitation, use-cases should achieve the following objectives:

- To define the functional and operational requirements of the system (product) by defining a scenario of usage that is agreed upon by the end-user and the software engineering team.
- To provide a clear and unambiguous description of how the end-user and the system interact with one another.
- To provide a basis for validation testing.

During OOA, use-cases serve as the basis for the first element of the analysis model. Using UML notation, a diagrammatic representation of a use-case, called a use-case diagram, can be created. Like many elements of the analysis model, the use-case diagram can be represented at many levels of abstraction. The use-case diagram contains actors and use-cases. Actors are entities that interact with the system. They can be human users or other machines or systems that have defined interfaces to the software.

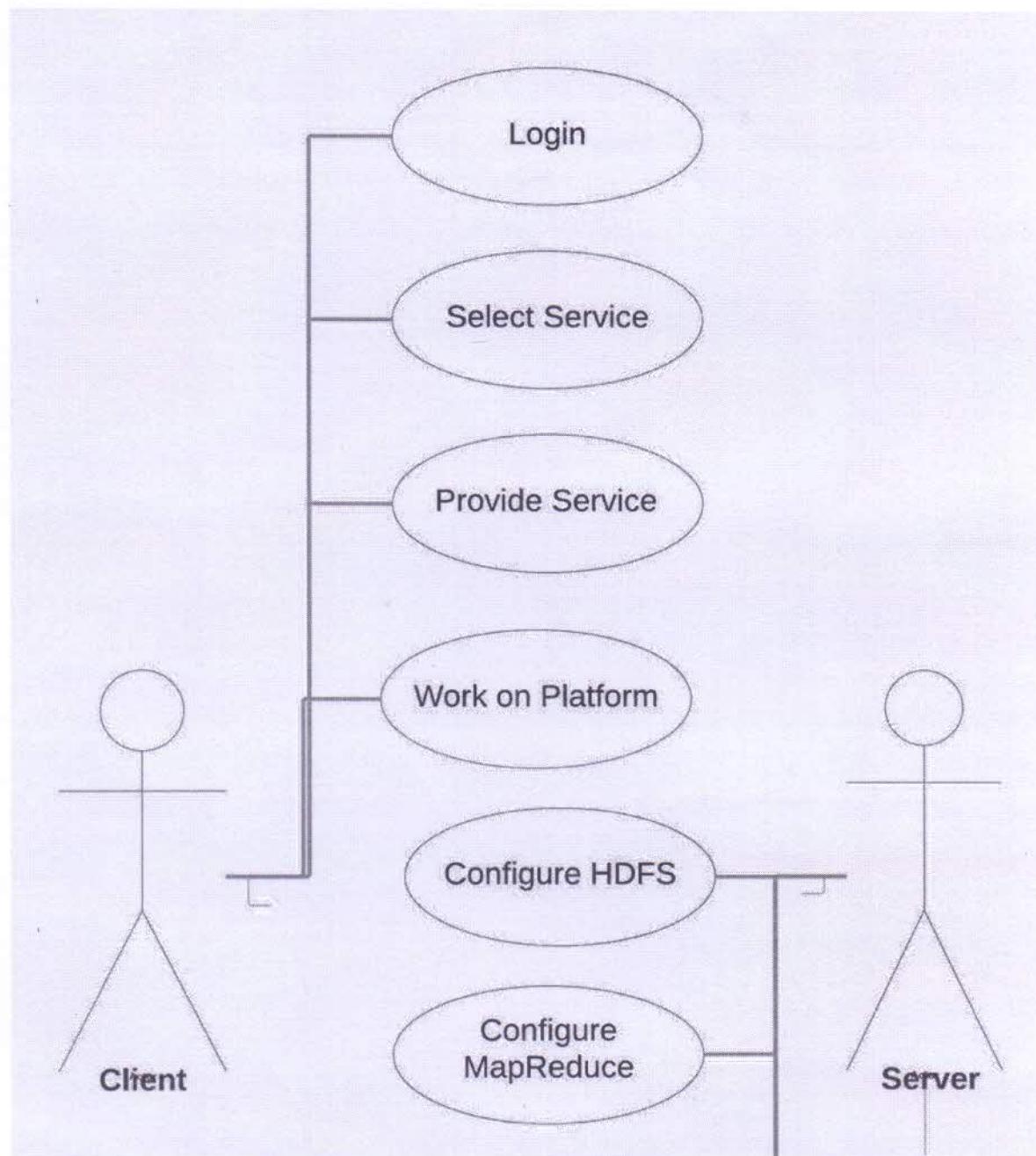


Fig.2 Use case diagram

6.2 ACTIVITY DIAGRAM

Activity diagram is another important diagram in UML to describe the dynamic aspects of the system.

Activity diagram is basically a flowchart to represent the flow from one activity to another activity. The activity can be described as an operation of the system.

The control flow is drawn from one operation to another. This flow can be sequential, branched, or concurrent. Activity diagrams deal with all type of flow control by using different elements such as fork, join, etc

6.3 DEPLOYMENTDIAGRAM:

Deployment diagram is a structure diagram which shows architecture of the system as deployment (distribution) of software artifacts to deployment targets.

Artifacts represent concrete elements in the physical world that are the result of a development process. Examples of artifacts are executable files, libraries, archives, database schemas, configuration files, etc.

Deployment target is usually represented by a node which is either hardware device or some software execution environment. Nodes could be connected through communication paths to create networked systems of arbitrary complexity.

Note, that components were directly deployed to nodes in UML 1.x deployment diagrams. In UML 2.x artifacts are deployed to nodes, and artifacts could manifest (implement) components.

Components are deployed to nodes indirectly through artifacts.

Deployment diagrams could describe architecture at specification level (also called type level) or at instance level (similar to class diagrams and object diagrams).

Specification level deployment diagram shows some overview of deployment of artifacts to deployment targets, without referencing specific instances of artifacts or nodes.

Instance level deployment diagram shows deployment of instances of artifacts to specific instances of deployment targets. It could be used for example to show differences in deployments to development, staging or production environments with the names/ids of specific build or deployment servers or devices.

Some common types of deployment diagrams are:

- Implementation (manifestation) of components by artifacts,
- Specification level deployment diagram,
- Instance level deployment diagram,
- Network architecture of the system.

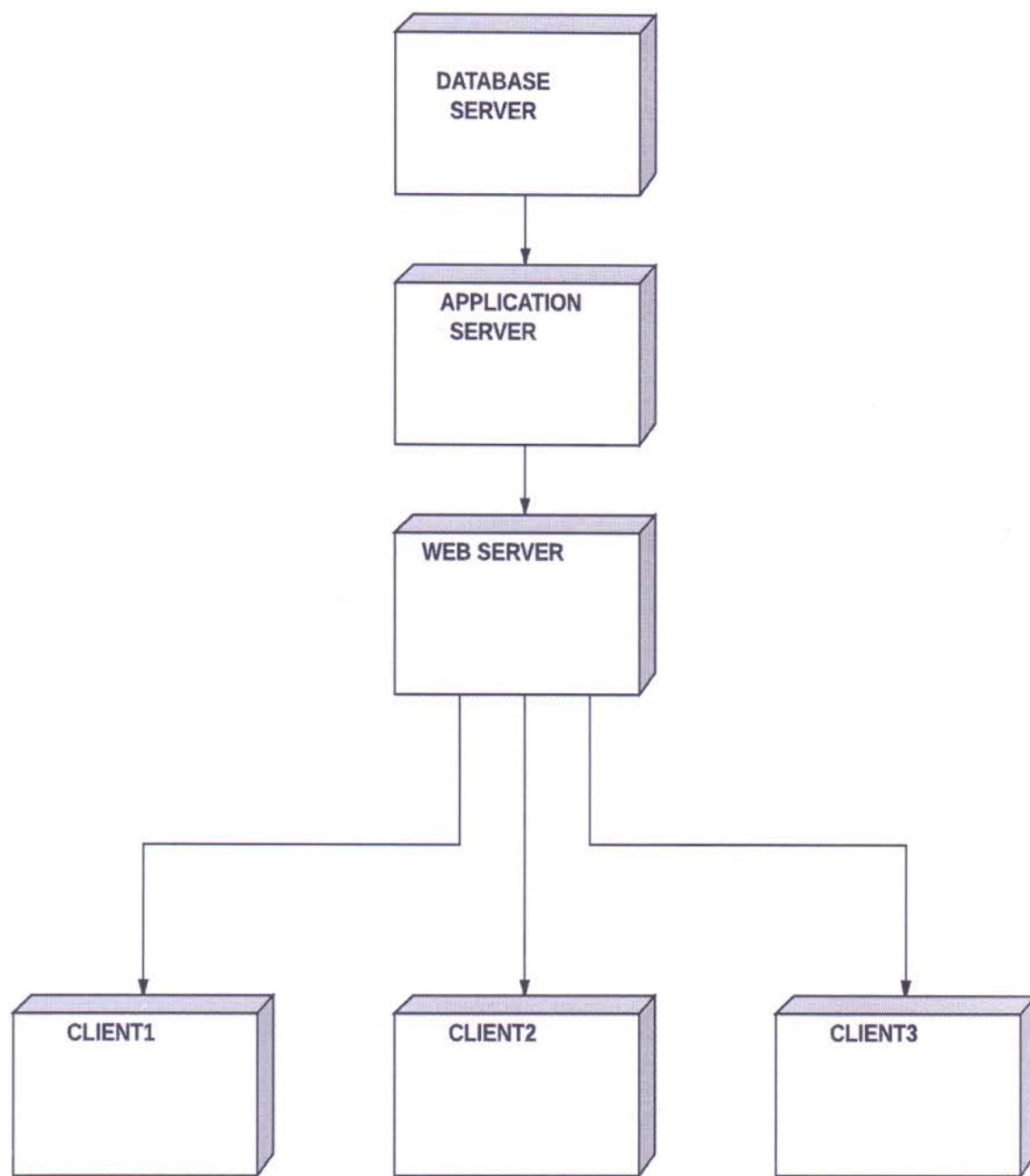


Fig.3 Deployment Diagram

CHAPTER-7

TESTING METHODOLOGY

7.1 TESTING:

Testing is the set of activities that can be planned in advance and conducted systematically. Numbers of testing strategies are proposed. All provide software developer with a template for testing and all have following characteristics.

- Testing begins at component level & works “outward” towards the integration of the entire computer based system.
- Different testing techniques are appropriate at different points in time.
- Testing is conducted by the developer of the software & independent test group.
- Testing & debugging are different activities, but debugging must be accommodated in any testing strategy.

7.2 VALIDATION AND VERIFICATION

7.2.1 What is Verification?

The standard definition of Verification goes like this: "Are we building the product RIGHT?" i.e. Verification is a process that makes it sure that the software product is developed the right way. The software should confirm to its predefined specifications, as the product development goes through different stages, an analysis is done to ensure that all required specifications are met. Methods and techniques used in the Verification and Validation shall be designed carefully, the planning of which starts right from the beginning of the development process. The Verification part of ‘Verification and Validation Model’ comes before Validation, which incorporates Software inspections, reviews, audits, walkthroughs, buddy checks etc. in each phase of verification (every phase of Verification is a phase of the Testing Life Cycle)

During the Verification, the work product (the ready part of the Software being developed and various documentations) is reviewed/examined personally by one or more persons in order to find

and point out the defects in it. This process helps in prevention of potential bugs, which may cause in failure of the project.

The activities involved in Verification process are:

- Requirement Specification verification
- Functional design verification
- Internal/system design verification and code verification

Each activity makes it sure that the product is developed right way and every requirement; every specification, design code etc. is verified!

7.2.2 What is Validation?

Validation is a process of finding out if the product being built is right? I.e. whatever the software product is being developed; it should do what the user expects it to do. The software product should functionally do what it is supposed to, it should satisfy all the functional requirements set by the user. Validation is done during or at the end of the development process in order to determine whether the product satisfies specified requirements.

Validation and Verification processes go hand in hand, but visibly Validation process starts after Verification process ends (after coding of the product ends). Each Verification activity (such as Requirement Specification Verification, Functional design Verification etc.) has its corresponding Validation activity (such as Functional Validation/Testing, Code Validation/Testing, System/Integration Validation etc.).

All types of testing methods are basically carried out during the Validation process. Test plan, test suits and test cases are developed, which are used during the various phases of Validation process.

The phases involved in Validation process are:

- Code Validation/Testing
- Integration Validation/Integration Testing
- Functional Validation/Functional Testing
- System/User Acceptance Testing/Validation.

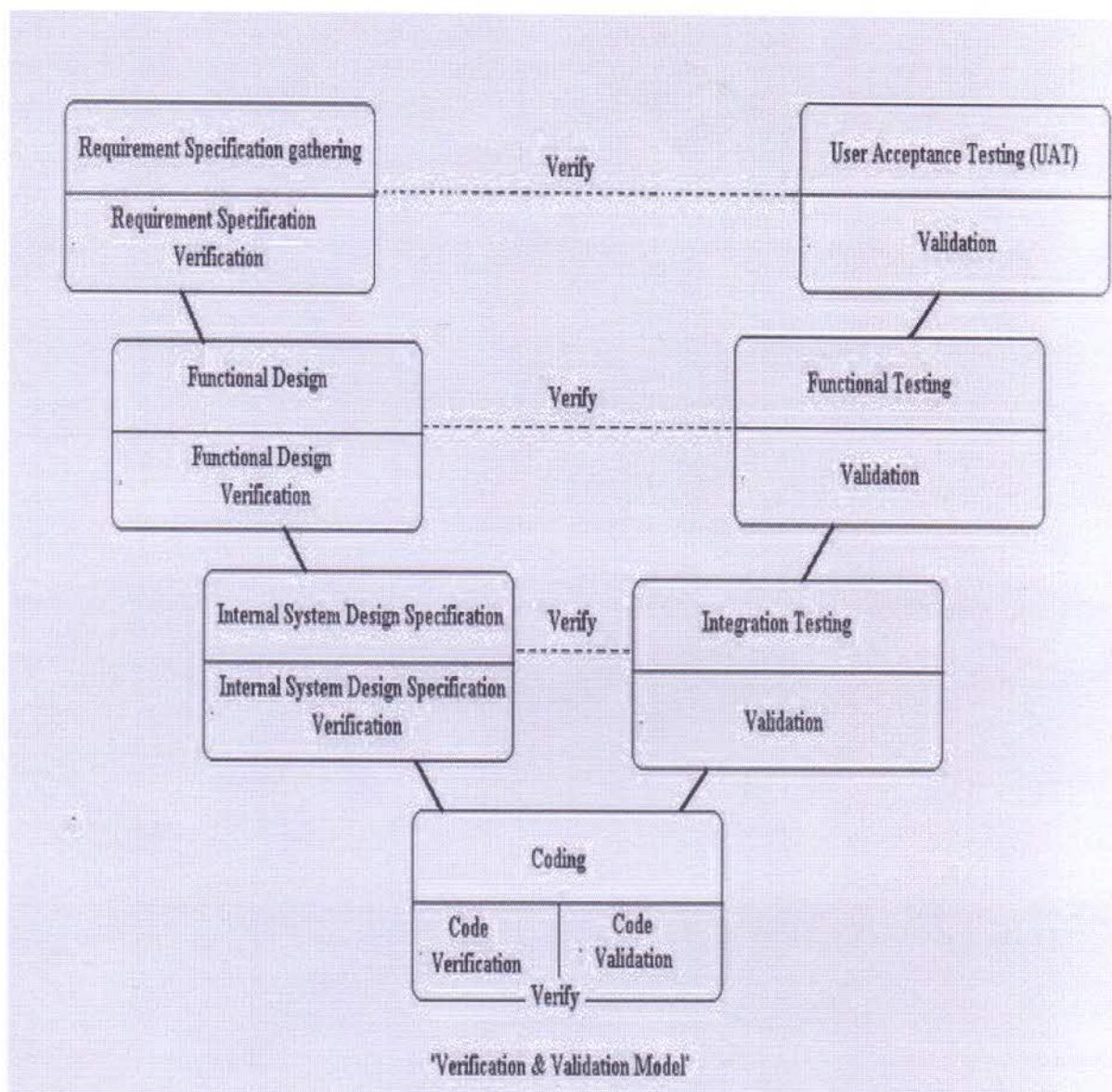


Figure10- Verification & Validation Model

7.3 TEST CASE DESIGN:

Test case specification has to be done separately for each unit. Test case specification gives, for each unit to be tested, all test cases, inputs to be used in the test cases, conditions being tested by the test case, and outputs expected for those test cases.

Test case specification is a major activity in the testing process. Careful selection of test cases that satisfy the criterion and approach specified is essential for proper testing. Test case specification document gives plan for testing that evaluates quality of test case.

With the specification of test cases, the next step in the testing process is to execute them. The steps to be performed to execute the test cases are specified in a test procedure specification which gives procedure for setting test environment and describes the methods and formats for reporting the results of testing. Test log, test summary report, and error report are some common outputs of test case execution.

7.3.1 LOGIN PAGE:

TestCase	TC-1
TestDescription	Logintextarea is not selected
TestData	Textfield is left blank
ExpectedResult	Message should display "Incorrect selection"
ActualResult	Correct output is displayed
Status	Pass

TestCase	TC-2
TestDescription	Enter invalid password
TestData	Incorrect password
ExpectedResult	Message should display "Wrong password"
ActualResult	Correct output is displayed
Status	Pass

7.3.2SIGN UP:

TestCase	TC-3
TestDescription	Enterincorrectpasswordinretypepasswordtextfield
TestData	Incorrectpassword
ExpectedResult	Messageshoulddisplay“passworddoesnotmatch”
ActualResult	Correctoutputisdisplayed
Status	Pass

TestCase	TC-4
TestDescription	Entervalidusernameandpassword
TestData	Username:root,password:redhat, retypepassword:root
ExpectedResult	Messagewilldisplay“usercreated”
ActualResult	Correctoutputisdisplayed
Status	Pass

7.3.3 Recovery Testing:

Recovery testing is a system test that enforces the software to fail in a variety of ways and verifies that recovery is properly performed. If the recovery is automatic, re-initialization, check pointing mechanism and data recovery and restart are each evaluated for correctness. If recovery requires human intervention, the mean time to repair is evaluated to determine whether it is within acceptable limits.

7.3.4 Security Testing:

Security testing attempts to verify that protection mechanisms built into a system will protect it from improper penetration.

7.3.5 Stress Testing:

Stress tests are designed to handle programs with abnormal situations. Stress testing executes a system in a manner that demands resources in abnormal quantity, frequency or volume.

CHAPTER-8

OUTPUTS

1. Login Page

The below screenshot shows the login page of web portal.

Where username= root and password= redhat.

On entering the wrong password , a message with “incorrect details” will be shown.

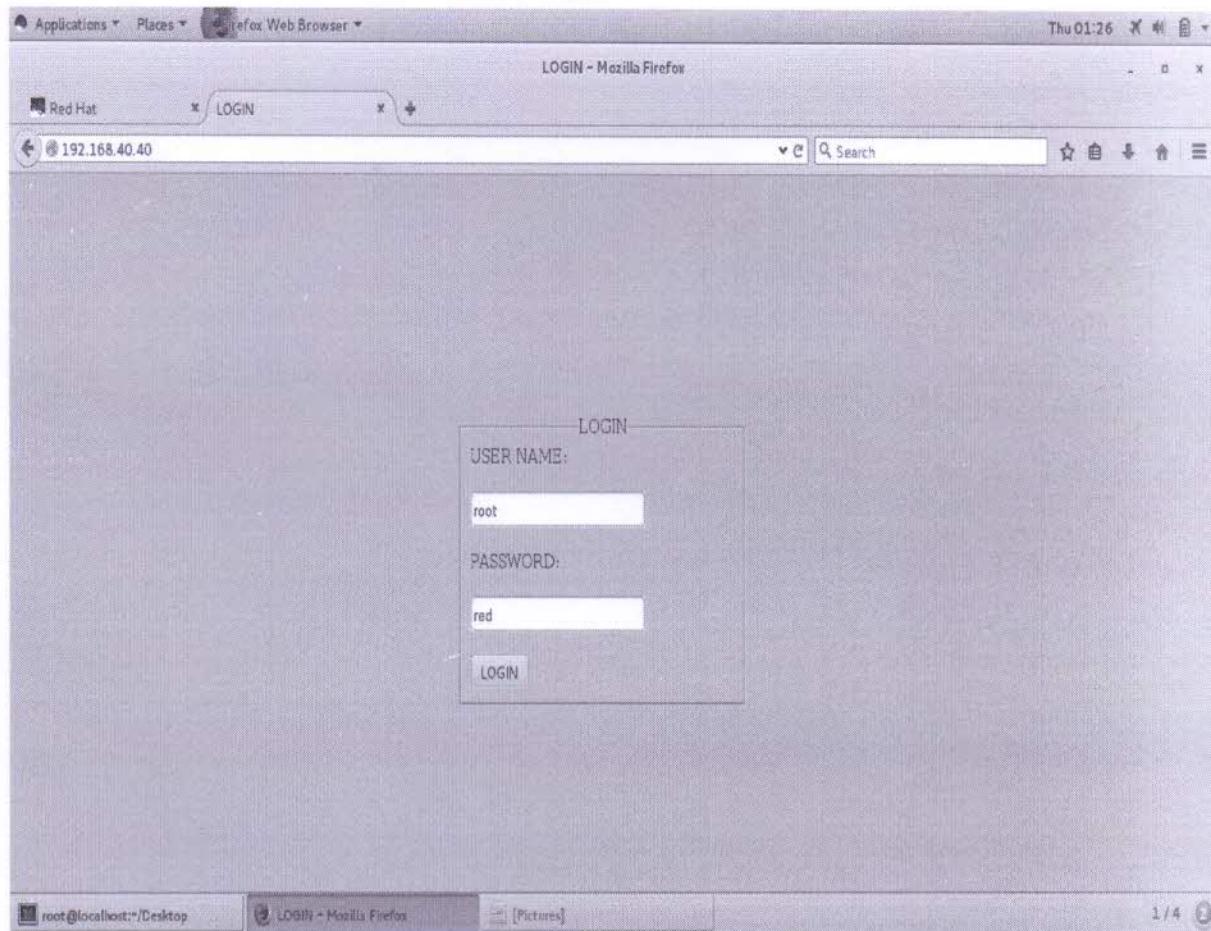


Fig.1 Login page

2. Access page

The screenshot below shows the immediate page after entering the correct username and password.

The user will be redirected to a page showing “ACCESS GRANTED”.

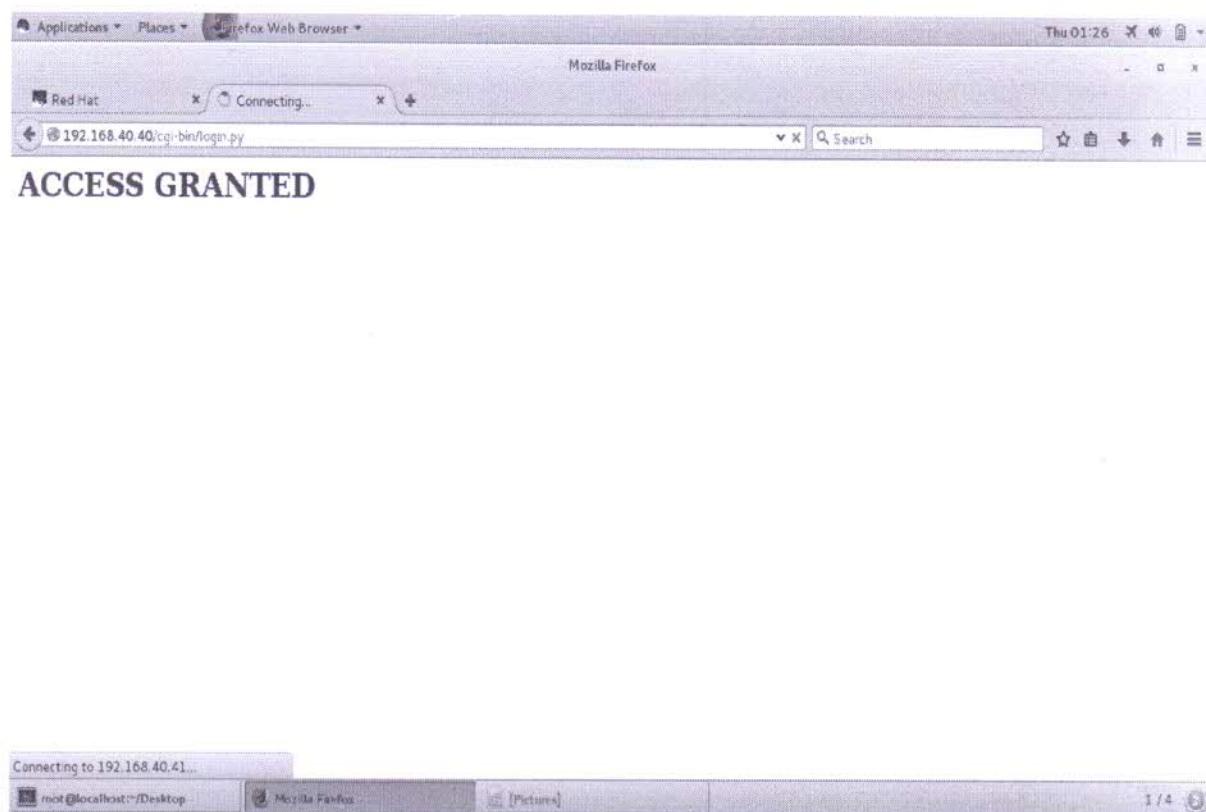


Fig.2 Access page

3. Configuration page

This page will show the setup option for configuration i.e. either manual or automatic.

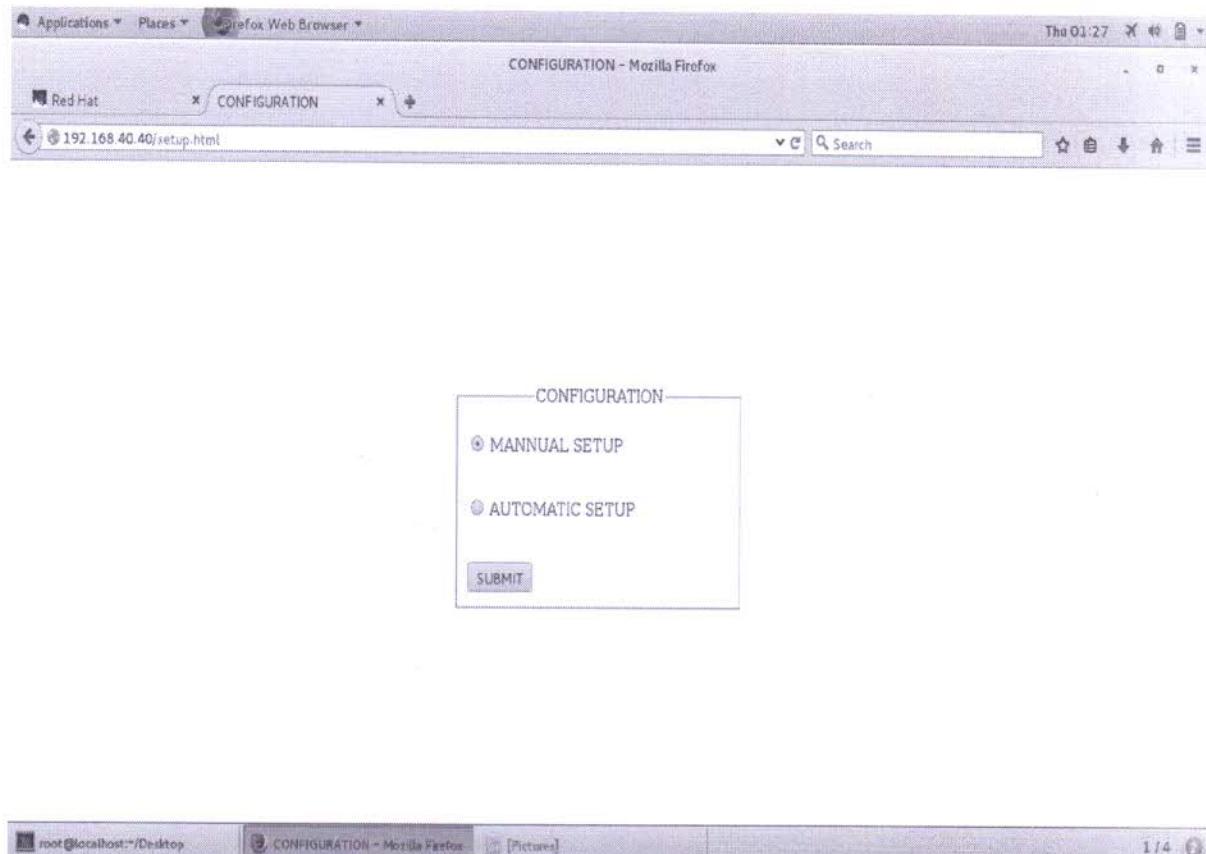


Fig.3 Configuration page

4. Automatic setup page

The resultant page will show the available systems along with their IP's and cores and an option to select the system as name node and job tracker.

Remaining systems were automatically configured as data nodes and task trackers.

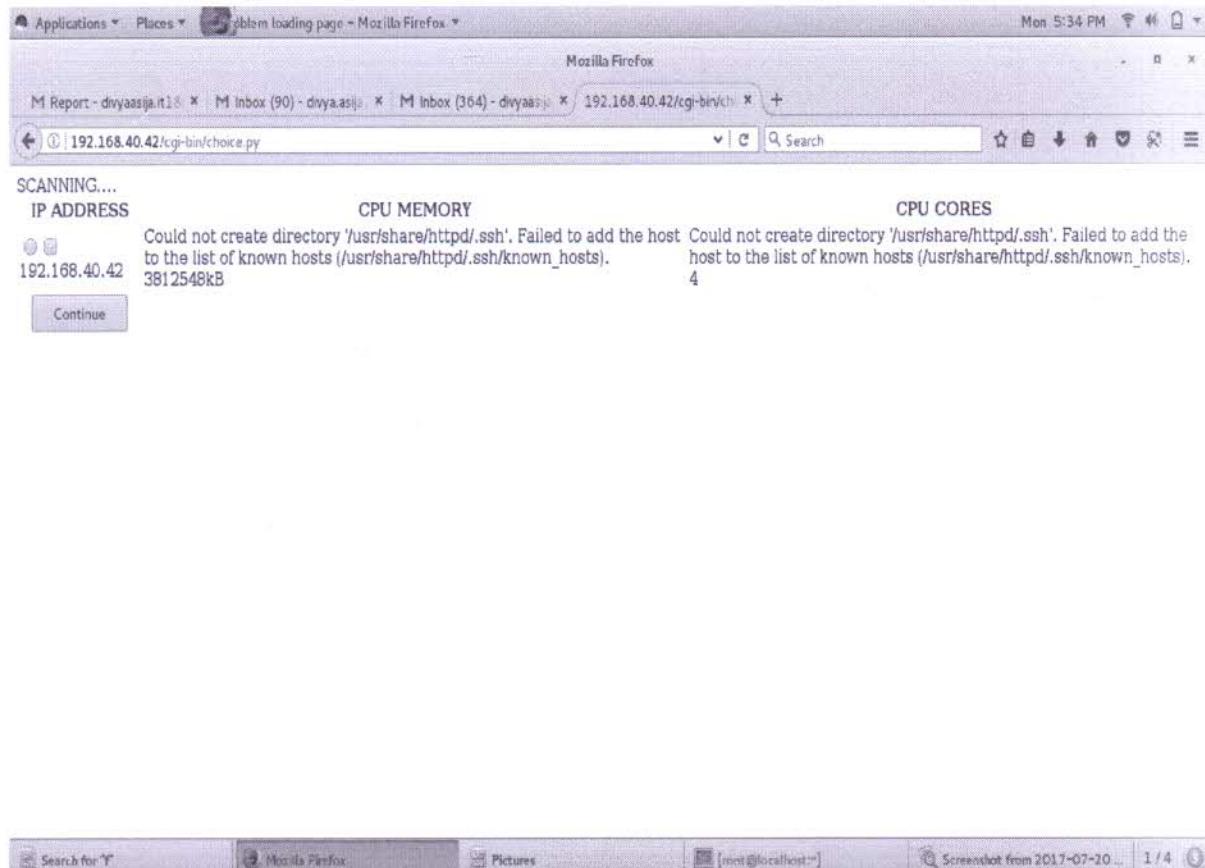


Fig.4 Automatic setup

5. Manual setup page

On selecting manual setup option in configuration file, system with maximum internal memory is configured as name node and job tracker whereas remaining systems were configured as data nodes and task tracker.

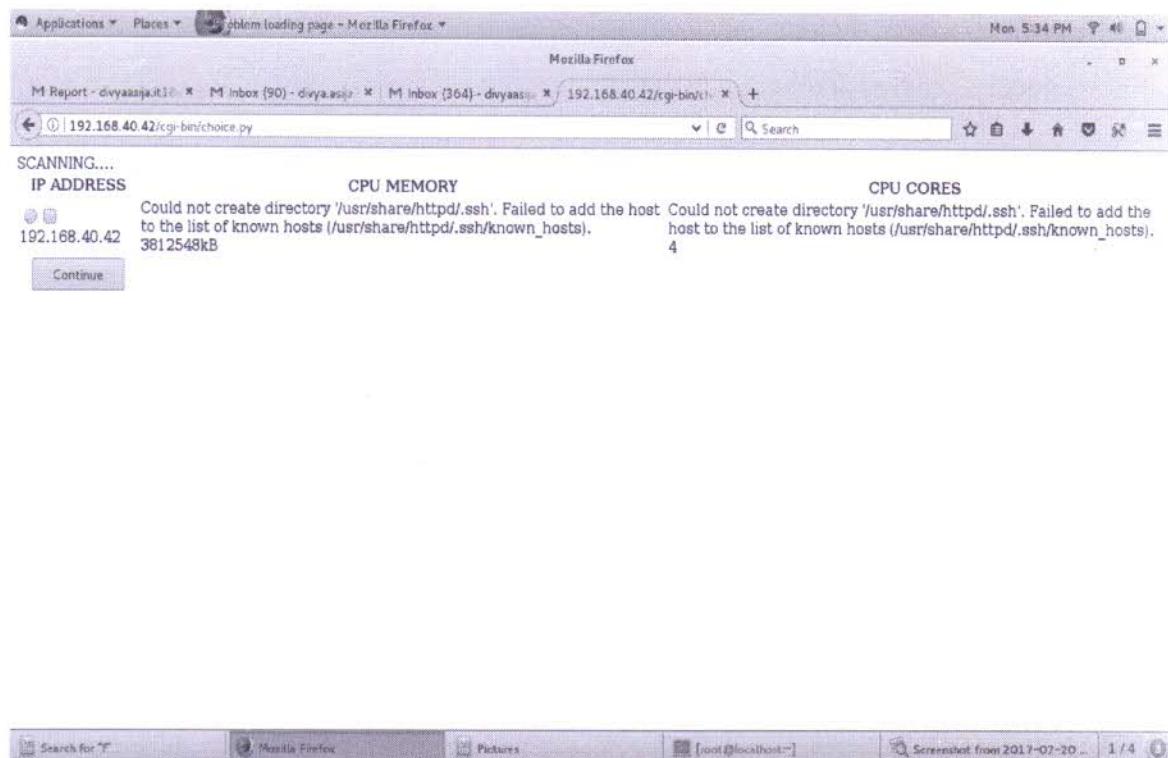


Fig.5 Manual setup

PROGRAM OUTCOMES

1. **Engineering Knowledge:** Apply the knowledge of distributed computing and distributed storage and an engineering specialization to the solution of complex storage problems in IT.
2. **Problem analysis:** Identify, formulate, research literature, and analyze complex storage problems reaching substantiated conclusions using first scripting and linux platform.
3. **Design/development of solutions:** Design solutions for complex operational problems and design system components using IT.
4. **Conduct investigations of complex problems:** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions using IT.
5. **Modern tool usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations in IT.
6. **The engineer and society:** Apply reasoning informed by the contextual knowledge to assess storage, computation and implementation of responsibilities relevant to the professional engineering practice using IT.
7. **Environment and sustainability:** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development in IT.
8. **Ethics:** Apply computation and processing on various systems together in order to access data from multiple locations.
9. **Individual and team work:** Function effectively as an individual and team to manage multiple system as a node to form cluster.
10. **Communication:** Communicate effectively on complex bigdata problems and activities with the community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
11. **Project Management and finance:** Demonstrate the need to use hadoop to solve bigdata problem in order to access data effectively.
12. **Life -long Learning:** A portal available to configure bigdata in certain number of clicks to get work done automatically.

PROJECT STAGE-II

Course Outcomes (COs):

On completion of the course:

CO1: Graduates will be able to understand the concepts of real world complex problems with analysing social impact for sustainable development in IT.

CO2: Graduates will be able to apply design, development and testing methodologies.

CO3: Graduates will be able to create cost effective solutions in multidisciplinary environments.

CO4: Graduates will be able to demonstrate their work with writing effective reports and design documentation via presentation tools.

SEM	SUBJECT	CODE	L/T/P	CO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12
7 th	Project Stage-II	7ITPR	P	Graduates will be able to understand the concepts of real world complex problems with analysing social impact for sustainable development in IT.	M	M	M	M	H	M	H	L	M	M	L	H
				Graduates will be able to apply design, development and testing methodologies.	M	L	H	H	H	H	M	M	M	H	M	H
				Graduates will be able to create cost effective solutions in multidisciplinary environments.	M	M	M	M	H	H	H	H	M	M	M	H
				Graduates will be able to demonstrate their work with writing effective reports and design documentation via presentation tools.	L	L	L	L	L	L	L	L	F	I	I	H

BIBLIOGRAPHY

RESEARCH ARTICLE:

- Details about research paper

WEB SITE FOR REFERENCES:

- [www.coursera.org,Introductiontobigdata](http://www.coursera.org/Introductiontobigdata)
- BigData:TheManagementRevolution.Andrew McAfee
- https://www.tutorialspoint.com/uml/uml_standard_diagrams.htm

CONCLUSION

By the above description we can understand the need of Big Data in future, so Hadoop can be the best of maintenance and efficient implementation of large data. This technology has bright future scope because day by day need of data would increase and security issues also the major point. In now-a-days many Multinational organizations are prefer Hadoop over RDBMS. So major companies like Facebook, Amazon, yahoo, LinkedIn etc. are adapting Hadoop and in future there can be many names in the list.

Hence Hadoop Technology is the best appropriate approach for handling the large data in smart way and its future is bright...



IT IDEATHON EXHIBITION 2017

CERTIFICATE OF PARTICIPATION

AWARDED TO

PRASHANT MANGHANI

FOR HIS/HER VALUEABLE PARTICIPATION IN
"IT IDEATHON EXHIBITION 2017" ORGANISED BY
INFORMATION TECHNOLOGY DEPARTMENT ON 13 Dec 2017, JECRC

V. Renu

Dr. VINAY KUMAR CHANDNA
PRINCIPAL

SUNIL KUMAR JANGIR
HOD, IT Department

NAVEEN KEDIA
P.Coordinator, IT Department



JAIPUR ENGINEERING COLLEGE AND RESEARCH CENTRE, JAIPUR
DEPARTMENT OF INFORMATION TECHNOLOGY
International Conference On Information Technology & Digital Applications 2018
April 6-7, 2018



CERTIFICATE

This is to certify that Prof./Dr./Ms. PRASHANT MANGHANI
of JECRC, JAIPUR has participated/
presented a paper/ delivered Invited Talk/Chaired Session entitled DATA AND
WEB MINING in the International Conference On Information
Technology & Digital Applications (ICITDA-2018) organized on April 6-7, 2018 by Department of
Information Technology, JECRC, Jaipur.

Mr. SUNIL KUMAR JANGIR
CONVENER, ICITDA-2018

Prof.(Dr.) V.K. CHANDNA
GENERAL CHAIR, ICITDA-2018

Mr. NAVEEN KUMAR KEDIA
CONVENER, ICITDA-2018

Prof.(Dr.) S.K. SINGH
ORGANIZING SECRETARY, ICITDA-2018