A

Mini Project Report

On

**PRIVACY-AWARE PERSONAL DATA STORAGE(P-PDS)**

Submitted in partial fulfillment of the

Requirement for the award of degree of

**BACHELOR OF TECHNOLOGY**

**In**

**COMPUTER SCIENCE & ENGINEERING**

|  |  |
| --- | --- |
|  | ***by*** |
| **CH.SRIJA** | **(226C1A0528)** |
| **G.VARSHITHA** | | **(226C1A0537)** |
| **K.USHA** | **(226C1A0553)** |
|  |  |
| **K.SIVAMANI** | **(226C1A0549)** |
|  |  |

**G,MOUNIKA (226C1A0534)**

Under the Guidance of

**Mr.K.SAMSON**

**Associate Professor**

****

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**PRIYADARSHINI INSTITUTE OF SCIENCE & TECHNOLOGY FOR WOMEN**

SAI PRABHATH NAGAR,PEDDATHANDA,KHAMMAM-507003

(Approved by AICTE, Affiliated to JNTUH, Hyderabad-T.S.)

**2022– 2026**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**PRIYADARSHINI INSTITUTE OF SCIENCE & TECHNOLOGY FOR WOMEN**

SAI PRABHATH NAGAR,PEDDATHANDA,KHAMMAM-507003

(Approved by AICTE, Affiliated to JNTUH, Hyderabad-T.S.)



***CERTIFICATE***

This is to certify that the mini project report Entitled **“ PRIVACY – AWARE PERSONAL DATA STORAGE (P-PDS)”** is a bona fide record of work carried out by

|  |  |
| --- | --- |
| **CH.SRIJA** | **(226C1A0528)** |

**G.VARSHITHA (226C1A0537)**

**K.USHA (226C1A0553)**

**K.SIVAMANI (226C1A0549)**

**G.MOUNIKA (226C1A0534)**

We hereby accord our approval of it as a mini project report carried out and presented in a manner required for its acceptance in partial fulfillment for award of degree of

**BACHELOR OF TECHNOLOGY**

**In**

**COMPUTER SCIENCE AND ENGINEERING**

Of

**Jawaharlal Nehru Technological University, Hyderabad.**

During the academic year 2024-25

|  |  |
| --- | --- |
| **INTERNAL GUIDE** |  |
| Mr.K.SAMSON , M.Tech |  |
| Associate Professor |  |

**HEAD OF THE DEPARTMENT**

Mr.I.NARASIMHA RAO, M.Tech., (Ph.D.)

Associate Professor

**EXTERNAL EXAMINER**

**Dr. B.Gopal, M.Tech.,Ph.D**.

**Principal**

**DECLARATION**

We declare that the project entitled **“PRIVACY – AWARE PERSONAL DATA STORAGE(P-PDS)”** recorded in this reportdoes not form part of any other report on which a degree has been awarded earlier. We further declare that this project report is based on our work carried out at the

**“PRIYADARSHINI INSTITUTE OF SCIENCE & TECHNOLOGY FOR WOMEN”,** in

the B.Tech. course.

|  |  |
| --- | --- |
|  | **REPORTED BY** |

**CH.SRIJA 226C1A0528**

**G.VARSHITHA 226C1A0537**

**K.USHA 226C1A0553**

**K.SIVAMANI 226C1A0549**

**G.MOUNIKA 226C1A0534**

|  |  |
| --- | --- |
|  |  |

**ACKNOWLEDGEMENT**

We express our gratitude to **Dr . K .NAVEEN BABU** , M . Tech , Ph .D Chairman, Priyadarshini Institute of Science and Technology for Women, Khammam, for giving us an opportunity to do this project work.

We express our gratitude to **Dr .B. GOPAL** , M .Tech, Ph.D. Principal, Priyadarshini Institute of Science and Technology for Women, Khammam, for giving us an opportunity to do this project work.

We express our gratitude to **Mr. INR,** Head of the department, computer science and Engineering, Priyadarshini Institute of Science and Technology for Women, Khammam, for giving us an opportunity to do this project work.

We thankful to our Guide **Mr. K.SAMSON**, Associate Professor, computer science and Engineering, Who with his continuous efforts, unfailing interest, constant support and providing us the right infrastructure helped us in completing this project work.

We would like to thank all the faculty members of CSE Department and our friends for their valuable suggestions and support which directly or indirectly helped us mounding this project in to a comprehensive one.

We wish to express our gratitude to our parents, whose love and encouragement have a great support throughout my education.

**PROJECT ASSOCIATES**

**CONTENTS**

|  |  |  |
| --- | --- | --- |
|  | |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  | |  |
|  | |  |
|  |  |  |
|  |  |  |
|  |  |  |

**ABSTRACT**

Recently, Personal Data Storage (PDS) has inaugurated a substantial change to the way people can store and control their personal data, by moving from a service-centric to a user-centric model. PDS offers individuals the capability to keep their data in a unique logical repository, that can be connected and exploited by proper analytical tools, or shared with third parties under the control of end users. Up to now, most of the research on PDS has focused on how to enforce user privacy preferences and how to secure data when stored into the PDS. In contrast, in this paper we aim at designing a Privacy-aware Personal Data Storage (P-PDS), that is, a PDS able to automatically take privacy-aware decisions on third parties access requests in accordance with user preferences. The proposed P-PDS is based on preliminary results presented in , where it has been demonstrated that semi-supervised learning can be successfully exploited to make a PDS able to automatically decide whether an access request has to be authorized or not. In this

paper, we have deeply revised the learning process so as to have a more usable P PDS, in terms of reduced effort for the training phase, as well as a more conservative approach With respect to user privacy, when handling conflicting access requests. We run several experiments on a realistic dataset exploiting a group of 360 evaluators. The obtained results show the effectiveness of the proposed approach.

**INTRODUCTION**

Nowadays personal data we are digitally producing are scattered in different online systems managed by different providers (e.g., online social media, hospitals, banks, airlines, etc). In this way, on the one hand users are losing control on their data, whose protection is under the responsibility of the data provider, and, on the other, they cannot fully exploit their data, since each provider keeps a separate view of them. To overcome this scenario, Personal Data Storage (PDS) [2]– [4] has inaugurated a substantial change to the way people can store and control their personal data, by moving from a service-centric to a user-centric model. PDSs enable individuals to collect into a single logical vault personal information they are producing. Such data can then

be connected and exploited by proper analytical tools, as well as shared with third parties under the control of end users. This view is also enabled by recent developments in privacy legislation and, in particular, by the new EU General Data Protection Regulation (GDPR), whose art. 20 states the right to data portability, according to which the data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format, thus making possible data collection into a PDS. Up to now, most of the research on PDS has focused on how to enforce user privacy preferences and how to secure data when stored into the PDS (see Section 7 for more details). In contrast, the key issue of helping users to specify their privacy preferences on PDS data has not been so far deeply investigated. This is a fundamental issue since average PDS users are not skilled enough to understand how to translate their privacy requirements into a set of privacy preferences. As several studies have shown, average users might have difficulties in properly setting potentially complex privacy preferences [5]– [7]. For example, let us consider Facebooks privacy setting, where users need to configure the options manually according to their desire. In, authors survey user awareness, attitudes and privacy concerns on profile information and find that only a small number of users change the default privacy preferences on Facebook. Interestingly, in, authors find that even when users have changed their default privacy settings, the modified settings do not match the expectations (these are reached only for 39% of users). Moreover, another survey in has shown that Facebook users are not aware enough on protection tools that designed to protect their personal data. According to their study the majority (about 88%) of users had never read the Facebook privacy policy

**EXISTING SYSTEM**

* Oort is a user-centric cloud storage system that organizes data by users rather than applications, considering global queries which find and combine relevant data fields from relevant users. Moreover, it allows users to choose which applications can access their own data, and which types of data to be shared with which users. Sieve allows user to upload encrypted data to a single cloud storage. It utilizes key-homomorphic scheme to provide cryptographically enforced access control.
* Amber has proposed an architecture where users can choose applications to manipulate their data but it does not mention either how the global queries work or how the application providers interact with. In, authors developed a user-centric framework that share with third parity only the answers to a query instead of the raw data. Mortier et al. have proposed a trusted platform called Databox, which can manage personal data by a finegrained access control mechanism but do not focus on policy learning. Recently, proposed a Block chain-based Personal Data Store (BC-PDS) framework, which leverages on Block Chain to secure the storage of personal data. However, all the above proposals focus on access control enforcement whereas they do not consider user preference or policy learning.
* Privacy preference enforcement have been also investigated in different domains, such as for instance social networks where most of the platforms offer users a privacy setting page to manually set their privacy preferences. Research works have tried to alleviate the burden of this setting, by exploiting machine learning tools. For instance, have investigated the use of semi-supervised and unsupervised approaches to automatically extract privacy settings in social media. In, authors have considered location based data. They have compared the accuracy of manually set privacy preferences with the one of an automated mechanism based on machine learning. The results show that machine learning approaches provide better result than user-defined policies.

**Disadvantages**

* + In the existing work, the system doesn't have strong techniques to implement Privacy-aware Personal Data Storage (P-PDS).
  + The system doesn't have active learning which is to select from the training dataset the most representative instances to be labeled by users

**Proposed System**

* The system proposes a revised version of the ensemble learning algorithm proposed in , to enforce a more conservative approach With respect to user privacy. In particular, we reconsider how ensemble learning handles decisions for access requests for which classifiers return conflicting classes. In general, the final decision is taken selecting the class with the highest aggregated probabilities. However, this presents the limit of not considering user perspective, in that, it does not take into account which classifier is more relevant for the considered user.
* To cope with this issue, we propose an alternative strategy for aggregating the class labels returned by the classifiers. According to this approach, we assign a personalized weight to each single classifier used in ensemble learning. We also show how it is possible to learn these weights from the training dataset, thus without the need of further input from the P-PDS owner. Experiments show that this approach increases user satisfaction as well as the learning effectiveness.

**Advantages**

* PDS able to automatically take privacy-aware decisions on third parties access requests requires further investigation.
* The system proposes a revised version of the ensemble learning algorithm proposed in this system, to enforce a more conservative approach with respect to user privacy.

**SYSTEM**

**REQUIREMENTS**

**H/W System Configuration:-**

➢ Processor - Pentium –IV

➢ RAM - 4 GB (min)

➢ Hard Disk - 20 GB

➢ Key Board - Standard Windows Keyboard

➢ Mouse - Two or Three Button Mouse

➢ Monitor - SVGA

**Software Requirements:**

* Operating System - Windows XP
* Coding Language - Java/J2EE(JSP, Servlet)
* Front End - J2EE
* Back End - MySQL

**PRELIMINARY INVESTIGATION**

The first and foremost strategy for development of a project starts from the thought of designing a mail enabled platform for a small firm in which it is easy and convenient of sending and receiving messages, there is a search engine ,address book and also including some entertaining games. When it is approved by the organization and our project guide the first activity, ie. preliminary investigation begins. The activity has three parts:

* **Request Clarification**
* **Feasibility Study**
* **Request Approval**

**REQUEST CLARIFICATION**

After the approval of the request to the organization and project guide, with an investigation being considered, the project request must be examined to determine precisely what the system requires.

Here our project is basically meant for users within the company whose systems can be interconnected by the Local Area Network(LAN). In today’s busy schedule man need everything should be provided in a readymade manner. So taking into consideration of the vastly use of the net in day to day life, the corresponding development of the portal came into existence.

**FEASIBILITY ANALYSIS**

An important outcome of preliminary investigation is the determination that the system request is feasible. This is possible only if it is feasible within limited resource and time. The different feasibilities that have to be analyzed are

* **Operational Feasibility**
* **Economic Feasibility**
* **Technical Feasibility**

###### Operational Feasibility

Operational Feasibility deals with the study of prospects of the system to be developed. This system operationally eliminates all the tensions of the Admin and helps him in effectively tracking the project progress. This kind of automation will surely reduce the time and energy, which previously consumed in manual work. Based on the study, the system is proved to be operationally feasible.

**Economic Feasibility**

Economic Feasibility or Cost-benefit is an assessment of the economic justification for a computer based project. As hardware was installed from the beginning & for lots of purposes thus the cost on project of hardware is low. Since the system is a network based, any number of employees connected to the LAN within that organization can use this tool from at anytime. The Virtual Private Network is to be developed using the existing resources of the organization. So the project is economically feasible.

###### Technical Feasibility

According to Roger S. Pressman, Technical Feasibility is the assessment of the technical resources of the organization. The organization needs IBM compatible machines with a graphical web browser connected to the Internet and Intranet. The system is developed for platform Independent environment. Java Server Pages, JavaScript, HTML, SQL server and WebLogic Server are used to deve2lop the system. The technical feasibility has been carried out. The system is technically feasible for development and can be developed with the existing facility.

**4.3.3 REQUEST APPROVAL**

Not all request projects are desirable or feasible. Some organization receives so many project requests from client users that only few of them are pursued. However, those projects that are both feasible and desirable should be put into schedule. After a project request is approved, it cost, priority, completion time and personnel requirement is estimated and used to determine where to add it to any project list. Truly speaking, the approval of those above factors, development works can be launched.

**SYSTEM DESIGN AND DEVELOPMENT**

**INPUT DESIGN**

Input Design plays a vital role in the life cycle of software development, it requires very careful attention of developers. The input design is to feed data to the application as accurate as possible. So inputs are supposed to be designed effectively so that the errors occurring while feeding are minimized. According to Software Engineering Concepts, the input forms or screens are designed to provide to have a validation control over the input limit, range and other related validations.

This system has input screens in almost all the modules. Error messages are developed to alert the user whenever he commits some mistakes and guides him in the right way so that invalid entries are not made. Let us see deeply about this under module design.

Input design is the process of converting the user created input into a computer-based format. The goal of the input design is to make the data entry logical and free from errors. The error is in the input are controlled by the input design. The application has been developed in user-friendly manner. The forms have been designed in such a way during the processing the cursor is placed in the position where must be entered. The user is also provided with in an option to select an appropriate input from various alternatives related to the field in certain cases.

OUTPUT DESIGN

The Output from the computer is required to mainly create an efficient method of communication within the company primarily among the project leader and his team members, in other words, the administrator and the clients. The output of VPN is the system which allows the project leader to manage his clients in terms of creating new clients and assigning new projects to them, maintaining a record of the project validity and providing folder level access to each client on the user side depending on the projects allotted to him. After completion of a project, a new project may be assigned to the client. User authentication procedures are maintained at the initial stages itself. A new user may be created by the administrator himself or a user can himself register as a new user but the task of assigning projects and validating a new user rests with the administrator only.

The application starts running when it is executed for the first time. The server has to be started and then the internet explorer in used as the browser. The project will run on the local area network so the server machine will serve as the administrator while the other connected system

#### **Client Server**

#### **Over view:**

With the varied topic in existence in the fields of computers, Client Server is one, which has

generated more heat than light, and also more hype than reality. This technology has acquired a

certain critical mass attention with its dedication conferences and magazines. Major computer

vendors such as IBM and DEC, have declared that Client Servers is their main future market. A

survey of DBMS magazine reveled that 76% of its readers were actively looking at the client server

solution. The growth in the client server development tools from $200 million in 1992 to more than

$1.2 billion in 1996.

Client server implementations are complex but the underlying concept is simple and powerful. A

client is an application running with local resources but able to request the database and relate the

services from separate remote server. The software mediating this client server interaction is often

referred to as MIDDLEWARE.

#### **What is a Client Server**

Two prominent systems in existence are client server and file server systems. It is essential to

distinguish between client servers and file server systems. Both provide shared network access to

data but the comparison dens there! The file server simply provides a remote disk drive that can be

accessed by LAN applications on a file by file basis. The client server offers full relational database

services such as SQL-Access, Record modifying, Insert, Delete with full relational integrity backup/

restore performance for high volume of transactions, etc. the client server middleware provides a

flexible interface between client and server, who does what, when and to whom.

#### **Why Client Server**

Client server has evolved to solve a problem that has been around since the earliest days of

computing: how best to distribute your computing, data generation and data storage resources in

order to obtain efficient, cost effective departmental an enterprise wide data processing. During

mainframe era choices were quite limited. A central machine housed both the CPU and DATA (cards,

tapes, drums and later disks). Access to these resources was initially confined to batched runs that

produced departmental reports at the appropriate intervals. A strong central information service

department ruled the corporation. The role of the rest of the corporation limited to requesting new or

more frequent reports and to provide hand written forms from which the central data banks were

created and updated. The earliest client server solutions therefore could best be characterized as

“SLAVE-MASTER”.

#### Time-sharing changed the picture. Remote terminal could view and even change the central data,

#### subject to access permissions. And, as the central data banks evolved in to sophisticated relational

#### database with non-programmer query languages, online users could formulate adhoc queries and

#### produce local reports with out adding to the MIS applications software backlog. However remote

#### Access was through dumb terminals, and the client server remained subordinate to the

#### Slave\Master.

#### **Front end or User Interface Design**

The entire user interface is planned to be developed in browser specific environment with a touch of Intranet-Based Architecture for achieving the Distributed Concept.

The browser specific components are designed by using the HTML standards, and the dynamism of the designed by concentrating on the constructs of the Java Server Pages.

#### **Communication or Database Connectivity Tier**

The Communication architecture is designed by concentrating on the Standards of Servlets and Enterprise Java Beans. The database connectivity is established by using the Java Data Base Connectivity.

The standards of three-tire architecture are given major concentration to keep the standards of higher cohesion and limited coupling for effectiveness of the operations.

#### **Features of The Language Used**

In my project, I have chosen *Java* language for developing the code.

#### **About Java**

Initially the language was called as “oak” but it was renamed as “Java” in 1995. The primary

motivation of this language was the need for a platform-independent (i.e., architecture neutral)

language that could be used to create software to be embedded in various consumer electronic

devices.

* Java is a programmer’s language.
* Java is cohesive and consistent.
* Except for those constraints imposed by the Internet environment, Java gives the programmer, full control.

Finally, Java is to Internet programming where C was to system programming.

#### **Importance of Java to the Internet**

Java has had a profound effect on the Internet. This is because; Java expands the Universe of

objects that can move about freely in Cyberspace. In a network, two categories of objects are

transmitted between the Server and the Personal computer. They are: Passive information and

Dynamic active programs. The Dynamic, Self-executing programs cause serious problems in the

areas of Security and probability. But, Java addresses those concerns and by doing so, has opened

the door to an exciting new form of program called the Applet.

#### Java can be used to create two types of programs

***Applications* and *Applets*:**  An application is a program that runs on our Computer under the

operating system of that computer. It is more or less like one creating using C or C++. Java’s

ability to create Applets makes it important. An Applet is an application designed to be transmitted

over the Internet and executed by a Java –compatible web browser. An applet is actually a tiny

Java program, dynamically downloaded across the network, just like an image. But the difference

is, it is an intelligent program, not just a media file. It can react to the user input and dynamically

change.

#### **Features Of Java**

**Security**

Every time you that you download a “normal” program, you are risking a viral infection. Prior to

Java, most users did not download executable programs frequently, and those who did scanned them

for viruses prior to execution. Most users still worried about the possibility of infecting their systems

with a virus. In addition, another type of malicious program exists that must be guarded against. This

type of program can gather private information, such as credit card numbers, bank account balances,

and passwords. Java answers both these concerns by providing a “firewall” between a network application and your computer.

When you use a Java-compatible Web browser, you can safely download Java applets without fear of virus infection or malicious intent.

**Portability**

For programs to be dynamically downloaded to all the various types of platforms connected to the

Internet, some means of generating portable executable code is needed .As you will see, the same

mechanism that helps ensure security also helps create portability. Indeed, Java’s solution to these

two problems is both elegant and efficient.

**The Byte code**

The key that allows the Java to solve the security and portability problems is that the output of Java

compiler is Byte code. Byte code is a highly optimized set of instructions designed to be executed

by the Java run-time system, which is called the Java Virtual Machine (JVM). That is, in its standard

form, the JVM is an interpreter for byte code.

Translating a Java program into byte code helps makes it much easier to run a program in a wide

variety of environments. The reason is, once the run-time package exists for a given system, any Java

program can run on it.

Although Java was designed for interpretation, there is technically nothing about Java that prevents

on-the-fly compilation of byte code into native code. Sun has just completed its Just In Time (JIT)

compiler for byte code. When the JIT compiler is a part of JVM, it compiles byte code into

executable code in real time, on a piece-by-piece, demand basis. It is not possible to compile an

entire Java program into executable code all at once, because Java performs various run-time

checks that can be done only at run time. The JIT compiles code, as it is needed, during execution.

**Java, Virtual Machine (JVM)**

Beyond the language, there is the Java virtual machine. The Java virtual machine is an important

element of the Java technology. The virtual machine can be embedded within a web browser or an

operating system. Once a piece of Java code is loaded onto a machine, it is verified. As part of the

loading process, a class loader is invoked and does byte code verification makes sure that the code

that’s has been generated by the compiler will not corrupt the machine that it’s loaded on. Byte

code verification takes place at the end of the compilation process to make sure that is all accurate

and correct. So byte code verification is integral to the compiling and executing of Java code.

**Overall Description**

# Java Source

## Java byte code

# JavaVM

Java

.Class

**Picture showing the development process of JAVA Program**

Java programming uses to produce byte codes and executes them. The first box indicates that the

Java source code is located in a. Java file that is processed with a Java compiler called javac. The

Java compiler produces a file called a. class file, which contains the byte code. The. Class file is

then loaded across the network or loaded locally on your machine into the execution environment

is the Java virtual machine, which interprets and executes the byte code.

**Java Architecture**

Java architecture provides a portable, robust, high performing environment for development. Java

provides portability by compiling the byte codes for the Java Virtual Machine, which is then

interpreted on each platform by the run-time environment. Java is a dynamic system, able to load

code when needed from a machine in the same room or across the planet.

**Compilation of code**

When you compile the code, the Java compiler creates machine code (called byte code) for a

hypothetical machine called Java Virtual Machine (JVM). The JVM is supposed to execute the

byte code. The JVM is created for overcoming the issue of portability. The code is written and

compiled for one machine and interpreted on all machines. This machine is called Java Virtual

Machine.

**Compiling and interpreting Java Source Code**

During run-time the Java interpreter tricks the byte code file into thinking that it is running on a Java

**Source**

**Code**

**………..**

**………..**

**………..**

**…………**

# PC Compiler

**Macintosh**

**Compiler**

**SPARC**

###### Compiler

**Java**

**Byte code**

**(Platform**

**Indepen**

**dent)**

**Java**

**Interpreter**

**(PC)**

**Java**

**Interpreter**

**(Macintosh)**

**Java**

**Interpreter**

**(Sparc)**

Virtual Machine. In reality this could be a Intel Pentium Windows 95 or Sun SARC station running

Solaris or Apple Macintosh running system and all could receive code from any computer through

Internet and run the Applets.

**Simple**

Java was designed to be easy for the Professional programmer to learn and to use effectively. If

you are an experienced C++ programmer, learning Java will be even easier. Because Java inherits

the C/C++ syntax and many of the object oriented features of C++. Most of the confusing concepts

from C++ are either left out of Java or implemented in a cleaner, more approachable manner.

**Object-Oriented**

Java was not designed to be source-code compatible with any other language. This allowed the

Java team the freedom to design with a blank slate. One outcome of this was a clean usable,

pragmatic approach to objects. The object model in Java is simple and easy to extend, while simple

types, such as integers, are kept as high-performance non-objects.

**Robust**

The multi-platform environment of the Web places extraordinary demands on a program, because

the program must execute reliably in a variety of systems. The ability to create robust programs

was given a high priority in the design of Java. Java is strictly typed language; it checks your code

at compile time and run time.

Java virtually eliminates the problems of memory management and deallocation, which is

completely automatic. In a well-written Java program, all run time errors can –and should –be

managed by your program.

**JAVASCRIPT**

JavaScript is a script-based programming language that was developed by Netscape Communication Corporation. JavaScript was originally called Live Script and renamed as JavaScript to indicate its relationship with Java. JavaScript supports the development of both client and server components of Web-based applications. On the client side, it can be used to write programs that are executed by a Web browser within the context of a Web page. On the server side, it can be used to write Web server programs that can process information submitted by a Web browser and then updates the browser’s display accordingly

Even though JavaScript supports both client and server Web programming, we prefer JavaScript at Client side programming since most of the browsers supports it. JavaScript is almost as easy to learn as HTML, and JavaScript statements can be included in HTML documents by enclosing the statements between a pair of scripting tags

<SCRIPTS>..</SCRIPT>.

<SCRIPT LANGUAGE = “JavaScript”>

JavaScript statements

</SCRIPT>

Here are a few things we can do with JavaScript :

* Validate the contents of a form and make calculations.
* Add scrolling or changing messages to the Browser’s status line.
* Animate images or rotate images that change when we move the mouse over them.
* Detect the browser in use and display different content for different browsers.
* Detect installed plug-ins and notify the user if a plug-in is required.

We can do much more with JavaScript, including creating entire application.

JavaScript Vs Java

JavaScript and Java are entirely different languages. A few of the most glaring differences are:

* Java applets are generally displayed in a box within the web document; JavaScript can affect any part of the Web document itself.
* While JavaScript is best suited to simple applications and adding interactive features to Web pages; Java can be used for incredibly complex applications.

There are many other differences but the important thing to remember is that JavaScript and Java are separate languages. They are both useful for different things; in fact they can be used together to combine their advantages.

**ADVANTAGES**

* JavaScript can be used for Sever-side and Client-side scripting.
* It is more flexible than VBScript.
* JavaScript is the default scripting languages at Client-side since all the browsers supports it.

**Hyper Text Markup Language**

Hypertext Markup Language (HTML), the languages of the World Wide Web (WWW), allows users to produces Web pages that include text, graphics and pointer to other Web pages (Hyperlinks).

HTML is not a programming language but it is an application of ISO Standard 8879, SGML (Standard Generalized Markup Language), but specialized to hypertext and adapted to the Web. The idea behind Hypertext is that instead of reading text in rigid linear structure, we can easily jump from one point to another point. We can navigate through the information based on our interest and preference. A markup language is simply a series of elements, each delimited with special characters that define how text or other items enclosed within the elements should be displayed. Hyperlinks are underlined or emphasized works that load to other documents or some portions of the same document.

HTML can be used to display any type of document on the host computer, which can be geographically at a different location. It is a versatile language and can be used on any platform or desktop.

HTML provides tags (special codes) to make the document look attractive. HTML tags are not case-sensitive. Using graphics, fonts, different sizes, color, etc., can enhance the presentation of the document. Anything that is not a tag is part of the document itself.

**Basic HTML Tags :**

**<!-- -->** Specifies comments

**<A>……….</A>** Creates hypertext links

**<B>……….</B>** Formats text as bold

**<BIG>……….</BIG>** Formats text in large font.

**<BODY>…</BODY>** Contains all tags and text in the HTML document

**<CENTER>...</CENTER>** Creates text

**<DD>…</DD>** Definition of a term

**<DL>...</DL>**  Creates definition list

**<FONT>…</FONT>** Formats text with a particular font

**<FORM>...</FORM>** Encloses a fill-out form

**<FRAME>...</FRAME>** Defines a particular frame in a set of frames

**<H#>…</H#>** Creates headings of different levels

**<HEAD>...</HEAD>** Contains tags that specify information about a document

**<HR>...</HR>** Creates a horizontal rule

**<HTML>…</HTML>** Contains all other HTML tags

**<META>...</META>** Provides meta-information about a document

**<SCRIPT>…</SCRIPT>** Contains client-side or server-side script

**<TABLE>…</TABLE>**  Creates a table

**<TD>…</TD>** Indicates table data in a table

**<TR>…</TR>** Designates a table row

**<TH>…</TH>** Creates a heading in a table

**ADVANTAGES**

* A HTML document is small and hence easy to send over the net. It is small because it does not include formatted information.
* HTML is platform independent.
* HTML tags are not case-sensitive.

**Java Database Connectivity**

**What Is JDBC**?

JDBC is a Java API for executing SQL statements. (As a point of interest, JDBC is a trademarked name and is not an acronym; nevertheless, JDBC is often thought of as standing for Java Database Connectivity. It consists of a set of classes and interfaces written in the Java programming language. JDBC provides a standard API for tool/database developers and makes it possible to write database applications using a pure Java API.

Using JDBC, it is easy to send SQL statements to virtually any relational database. One can write a single program using the JDBC API, and the program will be able to send SQL statements to the appropriate database. The combinations of Java and JDBC lets a programmer write it once and run it anywhere.

What Does JDBC Do?

Simply put, JDBC makes it possible to do three things:

* Establish a connection with a database
* Send SQL statements
* Process the results.

**JDBC versus ODBC and other APIs**

At this point, Microsoft's ODBC (Open Database Connectivity) API is that probably the most widely used programming interface for accessing relational databases. It offers the ability to connect to almost all databases on almost all platforms.

So why not just use ODBC from Java? The answer is that you can use ODBC from Java, but this is best done with the help of JDBC in the form of the JDBC-ODBC Bridge, which we will cover shortly. The question now becomes "Why do you need JDBC?" There are several answers to this question:

1. ODBC is not appropriate for direct use from Java because it uses a C interface. Calls from Java to native C code have a number of drawbacks in the security, implementation, robustness, and automatic portability of applications.
2. A literal translation of the ODBC C API into a Java API would not be desirable. For example, Java has no pointers, and ODBC makes copious use of them, including the notoriously error-prone generic pointer "void \*". You can think of JDBC as ODBC translated into an object-oriented interface that is natural for Java programmers.
3. ODBC is hard to learn. It mixes simple and advanced features together, and it has complex options even for simple queries. JDBC, on the other hand, was designed to keep simple things simple while allowing more advanced capabilities where required.
4. A Java API like JDBC is needed in order to enable a "pure Java" solution. When ODBC is used, the ODBC driver manager and drivers must be manually installed on every client machine. When the JDBC driver is written completely in Java, however, JDBC code is automatically installable, portable, and secure on all Java platforms from network computers to mainframes.

**Two-tier and Three-tier Models**

The JDBC API supports both two-tier and three-tier models for database access.

In the two-tier model, a Java applet or application talks directly to the database. This requires a JDBC driver that can communicate with the particular database management system being accessed. A user's SQL statements are delivered to the database, and the results of those statements are sent back to the user. The database may be located on another machine to which the user is connected via a network. This is referred to as a client/server configuration, with the user's machine as the client, and the machine housing the database as the server. The network can be an Intranet, which, for example, connects employees within a corporation, or it can be the Internet.

In the three-tier model, commands are sent to a "middle tier" of services, which then send SQL statements to the database. The database processes the SQL statements and sends the results back to the middle tier, which then sends them to the user. MIS directors find the three-tier model very attractive because the middle tier makes it possible to maintain control over access and the kinds of updates that can be made to corporate data. Another advantage is that when there is a middle tier, the user can employ an easy-to-use higher-level API which is translated by the middle tier into the appropriate low-level calls. Finally, in many cases the three-tier architecture can provide performance advantages.

**Java applet or**

**Html browser**

**Application**

**Server (Java)**

**JDBC**

## DBMS

**Client machine (GUI)**

**HTTP, RMI, or CORBA calls**

**Server machine (business Logic)**

**DBMS-proprietary protocol**

**Database server**

**JAVA**

**Application**

### JDBC

### DBMS

**Client machine**

**DBMS-proprietary protocol**

**Database server**

Until now the middle tier has typically been written in languages such as C or C++, which offer fast performance. However, with the introduction of optimizing compilers that translate Java byte code into efficient machine-specific code, it is becoming practical to implement the middle tier in Java. This is a big plus, making it possible to take advantage of Java's robustness, multithreading, and security features. JDBC is important to allow database access from a Java middle tier.

**JDBC Driver Types**

The JDBC drivers that we are aware of at this time fit into one of four categories:

* JDBC-ODBC bridge plus ODBC driver
* Native-API partly-Java driver
* JDBC-Net pure Java driver
* Native-protocol pure Java driver

**JDBC-ODBC Bridge**

If possible, use a Pure Java JDBC driver instead of the Bridge and an ODBC driver. This completely eliminates the client configuration required by ODBC. It also eliminates the potential that the Java VM could be corrupted by an error in the native code brought in by the Bridge (that is, the Bridge native library, the ODBC driver manager library, the ODBC driver library, and the database client library).

**What Is the JDBC- ODBC Bridge?**

The JDBC-ODBC Bridge is a JDBC driver, which implements JDBC operations by translating them into ODBC operations. To ODBC it appears as a normal application program. The Bridge implements JDBC for any database for which an ODBC driver is available. The Bridge is implemented as the

SUN.JDBC.ODBC Java package and contains a native library used to access ODBC. The Bridge is a joint development of Inter solv and Java Soft.

**Java Server Pages (JSP)**

Java server Pages is a simple, yet powerful technology for creating and maintaining dynamic-content web pages. Based on the Java programming language, Java Server Pages offers proven portability, open standards, and a mature re-usable component model .The Java Server Pages architecture enables the separation of content generation from content presentation. This separation not eases maintenance headaches, it also allows web team members to focus on their areas of expertise. Now, web page designer can concentrate on layout, and web application designers on programming, with minimal concern about impacting each other’s work.

**Features of JSP**

**Portability:**

Java Server Pages files can be run on any web server or web-enabled application server that provides support for them. Dubbed the JSP engine, this support involves recognition, translation, and management of the Java Server Page lifecycle and its interaction components.

**Components**

It was mentioned earlier that the Java Server Pages architecture can include reusable Java components. The architecture also allows for the embedding of a scripting language directly into the Java Server Pages file. The components current supported include Java Beans, and Servlets.

**Processing**

A Java Server Pages file is essentially an HTML document with JSP scripting or tags. The Java Server Pages file has a JSP extension to the server as a Java Server Pages file. Before the page is served, the Java Server Pages syntax is parsed and processed into a Servlet on the server side. The Servlet that is generated outputs real content in straight HTML for responding to the client.

**Access Models:**

A Java Server Pages file may be accessed in at least two different ways. A client’s request comes directly into a Java Server Page. In this scenario, suppose the page accesses reusable Java Bean components that perform particular well-defined computations like accessing a database. The result of the Beans computations, called result sets is stored within the Bean as properties. The page uses such Beans to generate dynamic content and present it back to the client.

In both of the above cases, the page could also contain any valid Java code. Java Server Pages architecture encourages separation of content from presentation.

**Steps in the execution of a JSP Application*:***

1. The client sends a request to the web server for a JSP file by giving the name of the JSP file within the form tag of a HTML page.
2. This request is transferred to the Java Web Server. At the server side Java Web Server receives the request and if it is a request for a JSP file server gives this request to the JSP engine.
3. JSP engine is program which can understands the tags of the JSP and then it converts those tags into a Servlet program and it is stored at the server side. This Servlet is loaded in the memory and then it is executed and the result is given back to the Java Web Server and then it is transferred back to the result is given back to the Java Web Server and then it is transferred back to the client.

**JDBC connectivity**

The JDBC provides database-independent connectivity between the J2EE platform and a wide range of tabular data sources. JDBC technology allows an Application Component Provider to:

* Perform connection and authentication to a database server
* Manager transaction
* Move SQL statements to a database engine for preprocessing and execution
* Execute stored procedure

# CHAPTER 3 Tomcat 6.0 web server

Tomcat is an open source web server developed by Apache Group. Apache Tomcat is the servlet container that is used in the official Reference Implementation for the Java Servlet and Java Server Pages technologies. The Java Servlet and Java Server Pages specifications are developed by Sun under the Java Community Process. Web Servers like Apache Tomcat support only web components while an application server supports web components as well as business components (BEAs Web logic, is one of the popular application server).To develop a web application with JSP/servlet install any web server like J Run, Tomcat etc to run your application.



**Bibliography:**

**References for the Project Development were taken from the following Books and Web Sites**.

**Oracle**

PL/SQL Programming by Scott Urman

SQL complete reference by Livion

**JAVA Technologies**

JAVA Complete Reference

Java Script Programming by Yehuda Shiran

Mastering JAVA Security

JAVA2 Networking by Pistoria

JAVA Security by Scotl oaks

Head First EJB Sierra Bates

J2EE Professional by Shadab siddiqui

JAVA server pages by Larne Pekowsley

JAVA Server pages by Nick Todd

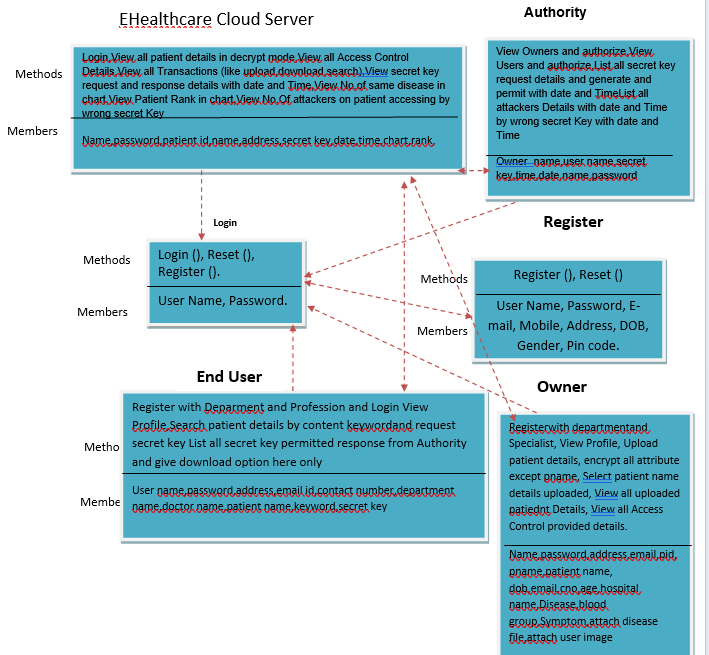
**HTML**

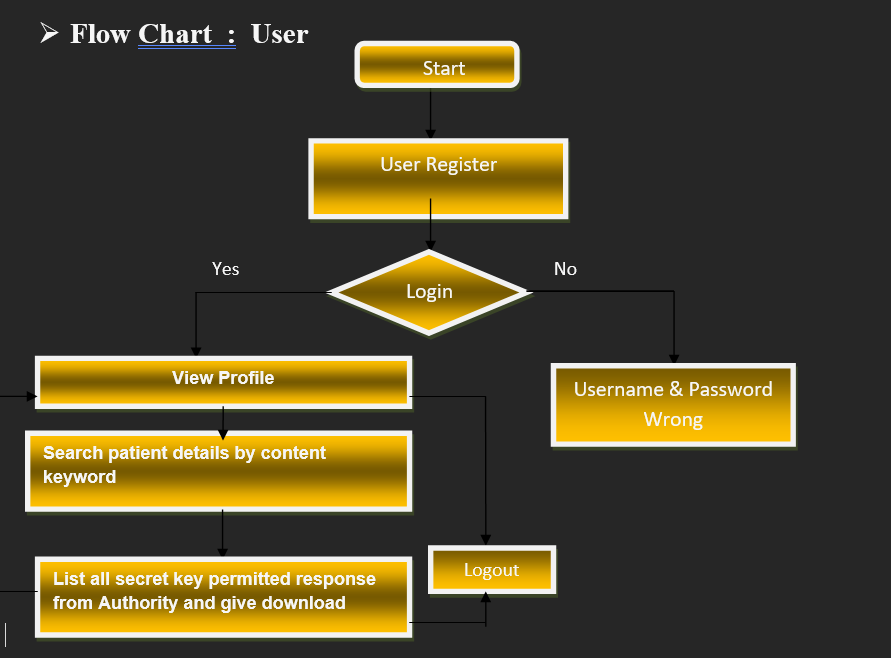
HTML Black Book by Holzner

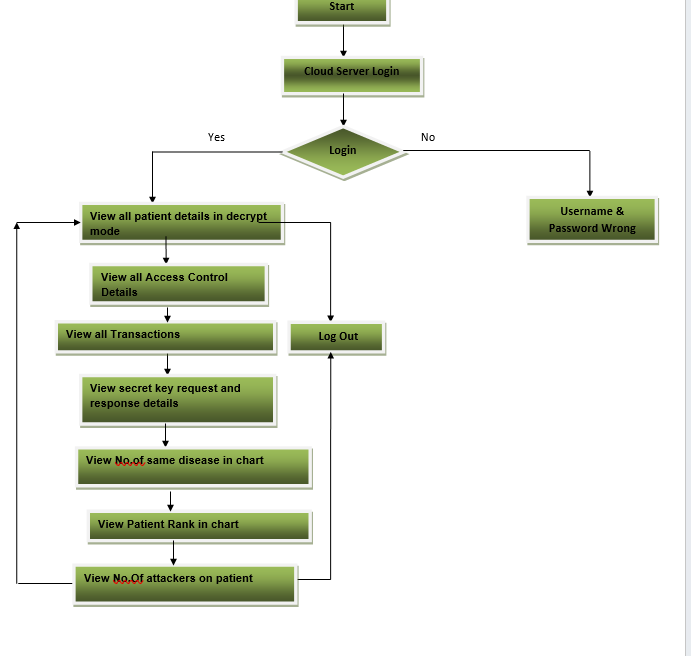
**JDBC**

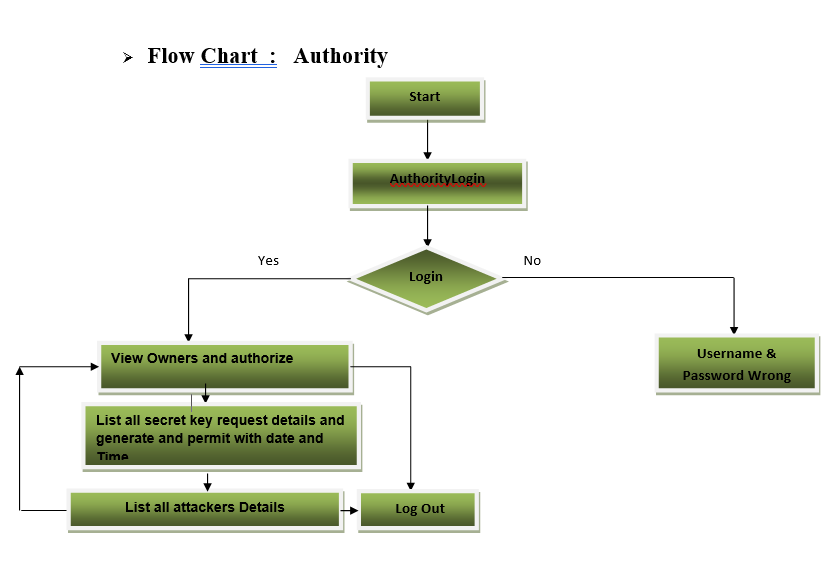
Java Database Programming with JDBC by Patel moss.

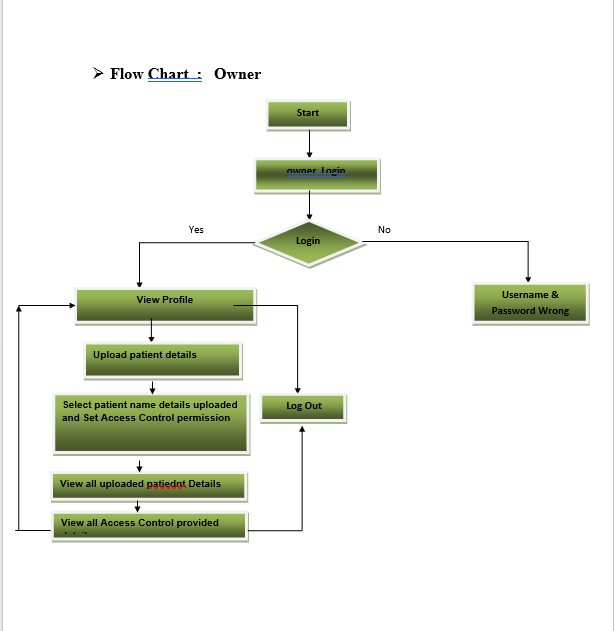
Software Engineering by Roger Pressman

**FLOW CHART**

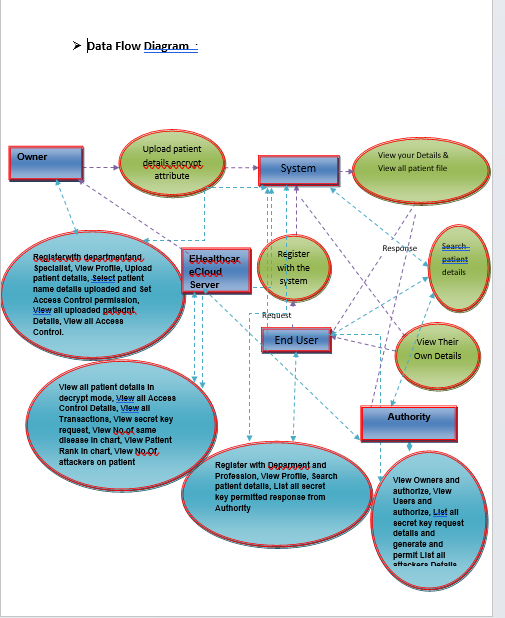


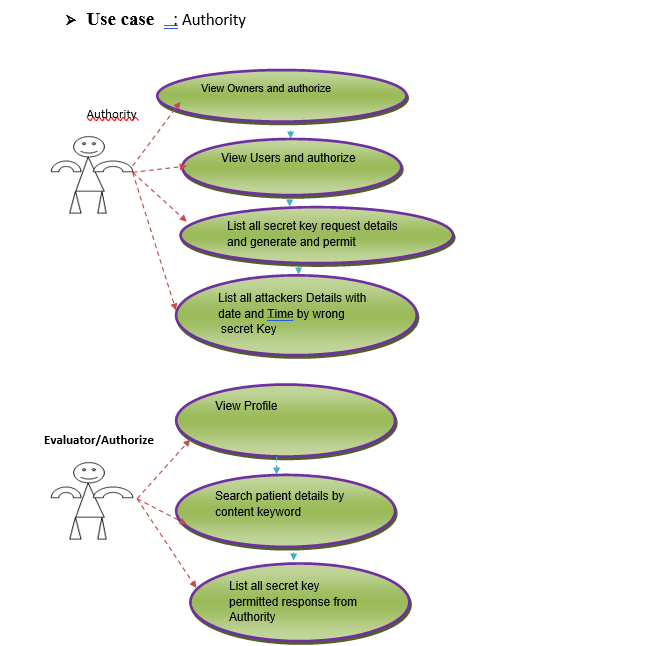
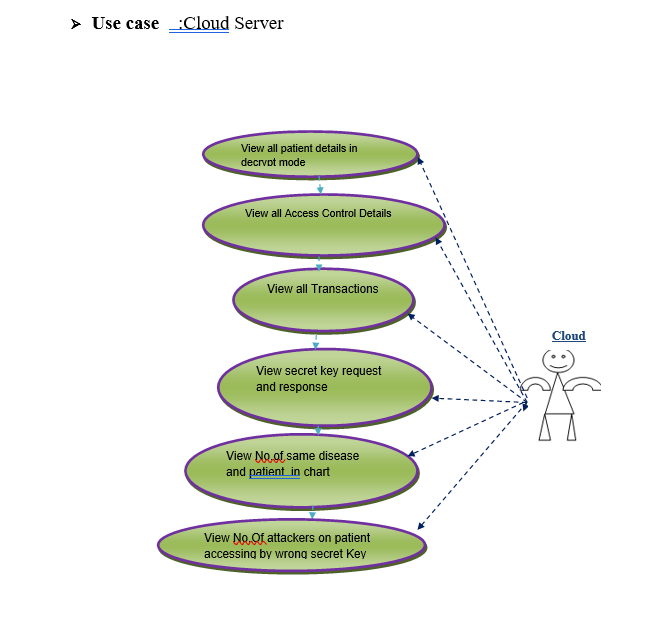


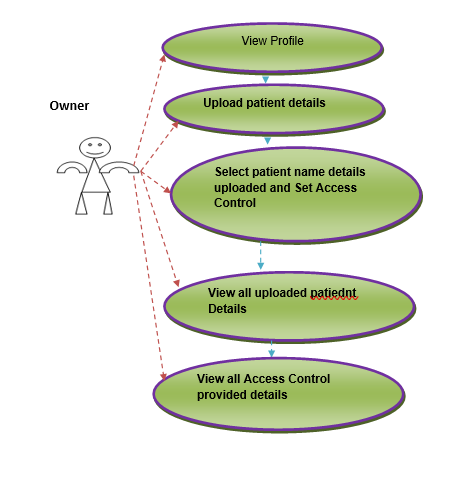


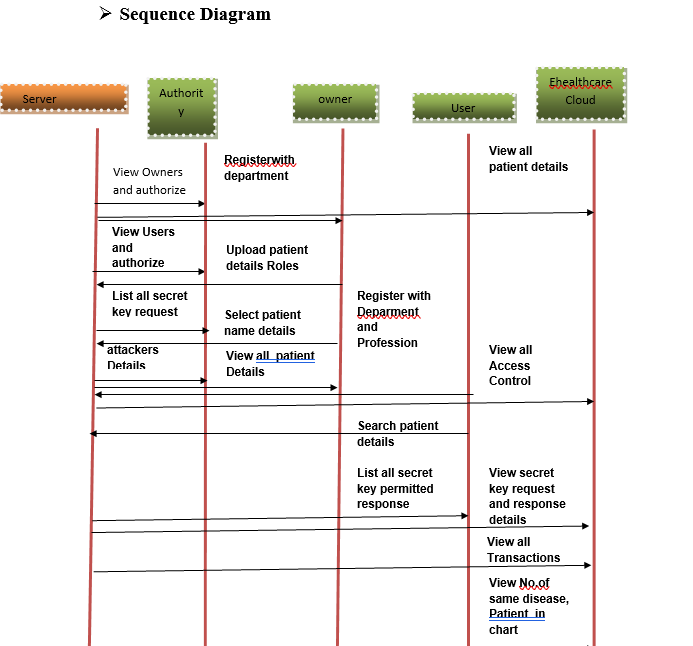


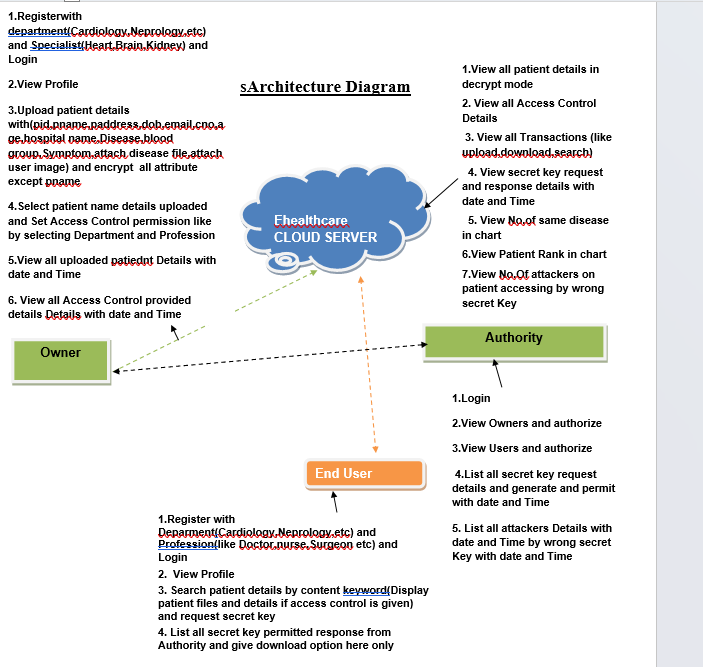
**DATA FLOW DIAGRAM**











**Implementation**

* **Data Owner**

In this module, the data provider uploads their encrypted data in the Cloud server. For the security purpose the data owner encrypts the data file and then store in the server. The Data owner can have capable of manipulating the encrypted data file and performs the following operations Register and Login, Attackers, Upload File, View Files, Verify data(Verifiability), View and Delete Files, View All Transactions.

**Cloud Service Provider**

The Cloud server manages which is to provide data storage service for the Data Owners. Data owners encrypt their data files and store them in the Server for sharing with data consumers. To access the shared data files, data consumers download encrypted data files of their interest from the Server and then Server will decrypt them. The server will generate the aggregate key if the end user requests for file authorization to access and performs the following operations such as Login, View and Authorize Users, View and Authorize Owners, View Files, View All Search Transactions, View All File Transactions, View All Top Searched, View Attackers, Search Requests, View Time Delay, View Throughput.

* **User**

In this module, the user can only access the data file with the secret key. The user can search the file for a specified keyword. The data which matches for a particular keyword will be indexed in the cloud server and then response to the end user and performing the following operations Register and Login, My Profile, View Files, Search Files, Search Ratio, Top K Search, Req Search Control.

### . SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the

Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

### TESTING METHODOLOGIES

The following are the Testing Methodologies:

* **Unit Testing.**
* **Integration Testing.**
* **User Acceptance Testing.**
* **Output Testing.**
* **Validation Testing.**

**Unit Testing**

Unit testing focuses verification effort on the smallest unit of Software design that is the module. Unit testing exercises specific paths in a module’s control structure to ensure complete coverage and maximum error detection. This test focuses on each module individually, ensuring that it functions properly as a unit. Hence, the naming is Unit Testing.

During this testing, each module is tested individually and the module interfaces are verified for the consistency with design specification. All important processing path are tested for the expected results. All error handling paths are also tested.

**Integration Testing:**

Integration testing addresses the issues associated with the dual problems of verification and program construction. After the software has been integrated a set of high order tests are conducted. The main objective in this testing process is to take unit tested modules and builds a program structure that has been dictated by design.

**The following are the types of Integration Testing:**

**1.Top Down Integration**

This method is an incremental approach to the construction of program structure. Modules are integrated by moving downward through the control hierarchy, beginning with the main program module. The module subordinates to the main program module are incorporated into the structure in either a depth first or breadth first manner.

In this method, the software is tested from main module and individual stubs are replaced when the test proceeds downwards.

**2. Bottom-up Integration**

This method begins the construction and testing with the modules at the lowest level in the program structure. Since the modules are integrated from the bottom up, processing required for modules subordinate to a given level is always available and the need for stubs is eliminated. The bottom up integration strategy may be implemented with the following steps:

* The low-level modules are combined into clusters into clusters that perform a specific Software sub-function.
* A driver (i.e.) the control program for testing is written to coordinate test case input and output.
* The cluster is tested.
* Drivers are removed and clusters are combined moving upward in the program structure

**OTHER TESTING METHODOLOGIES**

**User Acceptance Testing**

User Acceptance of a system is the key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with the prospective system users at the time of developing and making changes wherever required. The system developed provides a friendly user interface that can easily be understood even by a person who is new to the system.

**Output Testing**

After performing the validation testing, the next step is output testing of the proposed system, since no system could be useful if it does not produce the required output in the specified format. Asking the users about the format required by them tests the outputs generated or displayed by the system under consideration. Hence the output format is considered in 2 ways – one is on screen and another in printed format.

**Validation Checking**

Validation checks are performed on the following fields.

**Text Field:**

The text field can contain only the number of characters lesser than or equal to its size. The text fields are alphanumeric in some tables and alphabetic in other tables. Incorrect entry always flashes and error message.

**Numeric Field:**

The numeric field can contain only numbers from 0 to 9. An entry of any character flashes an error messages. The individual modules are checked for accuracy and what it has to perform. Each module is subjected to test run along with sample data. The individually tested modules are integrated into a single system. Testing involves executing the real data information is used in the program the existence of any program defect is inferred from the output. The testing should be planned so that all the requirements are individually tested.

A successful test is one that gives out the defects for the inappropriate data and produces and output revealing the errors in the system.

**Preparation of Test Data**

Taking various kinds of test data does the above testing. Preparation of test data plays a vital role in the system testing. After preparing the test data the system under study is tested using that test data. While testing the system by using test data errors are again uncovered and corrected by using above testing steps and corrections are also noted for future use.

**Using Live Test Data:**

Live test data are those that are actually extracted from organization files. After a system is partially constructed, programmers or analysts often ask users to key in a set of data from their normal activities. Then, the systems person uses this data as a way to partially test the system. In other instances, programmers or analysts extract a set of live data from the files and have them entered themselves.

It is difficult to obtain live data in sufficient amounts to conduct extensive testing. And,

although it is realistic data that will show how the system will perform for the typical processing requirement, assuming that the live data entered are in fact typical, such data generally will not test all combinations or formats that can enter the system. This bias toward typical values then does not provide a true systems test and in fact ignores the cases most likely to cause system failure.

**Using Artificial Test Data:**

Artificial test data are created solely for test purposes, since they can be generated to test all combinations of formats and values. In other words, the artificial data, which can quickly be prepared by a data generating utility program in the information systems department, make possible the testing of all login and control paths through the program.

The most effective test programs use artificial test data generated by persons other than those who wrote the programs. Often, an independent team of testers formulates a testing plan, using the systems specifications.

The package “Virtual Private Network” has satisfied all the requirements specified as per software requirement specification and was accepted.

**USER TRAINING**

Whenever a new system is developed, user training is required to educate them about the working of the system so that it can be put to efficient use by those for whom the system has been primarily designed. For this purpose the normal working of the project was demonstrated to the prospective users. Its working is easily understandable and since the expected users are people who have good knowledge of computers, the use of this system is very easy.

**MAINTAINENCE**

This covers a wide range of activities including correcting code and design errors. To reduce the need for maintenance in the long run, we have more accurately defined the user’s requirements during the process of system development. Depending on the requirements, this system has been developed to satisfy the needs to the largest possible extent. With development in technology, it may be possible to add many more features based on the requirements in future. The coding and designing is simple and easy to understand which will make maintenance easier.

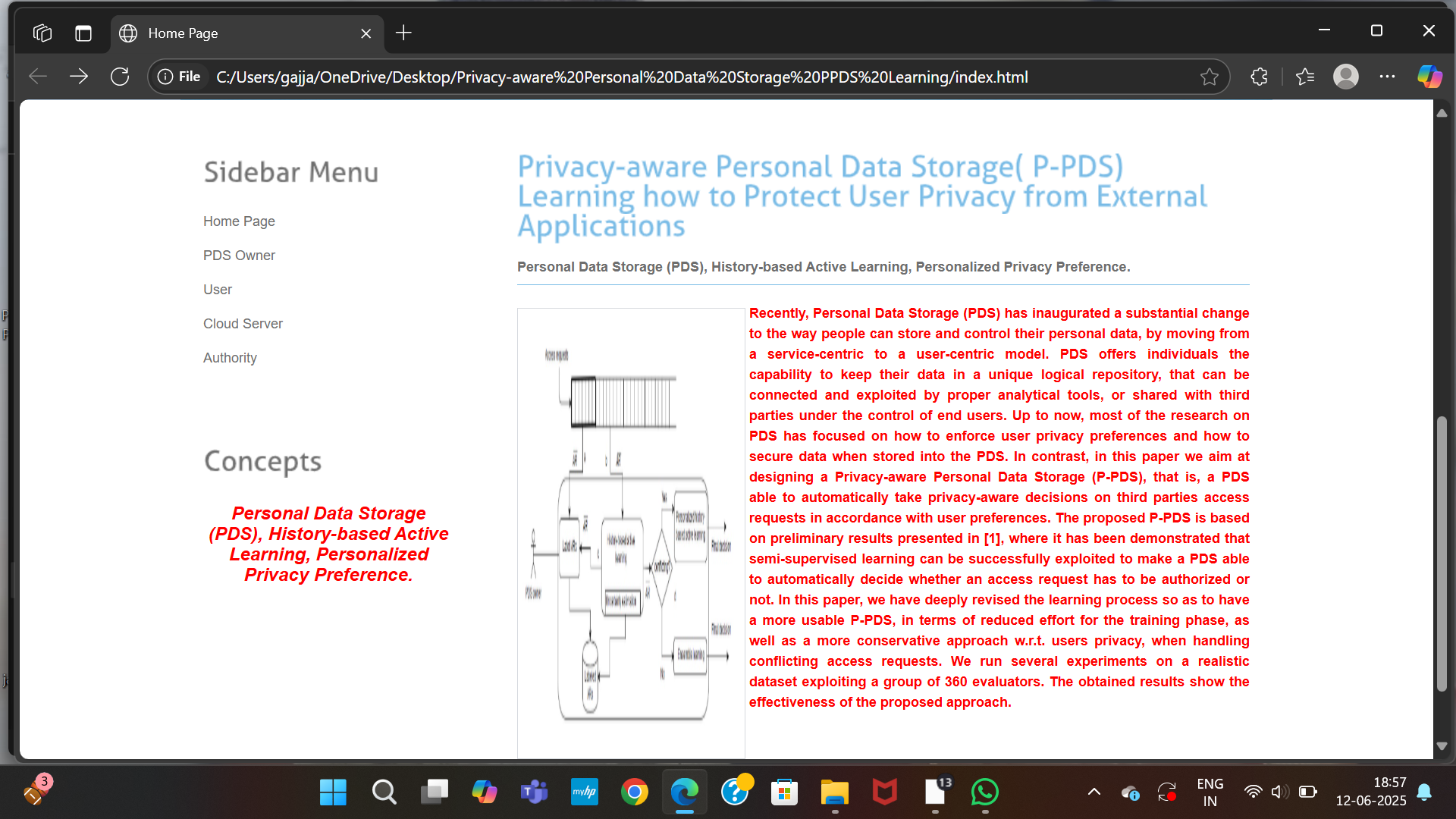
**TESTING STRATEGY :**

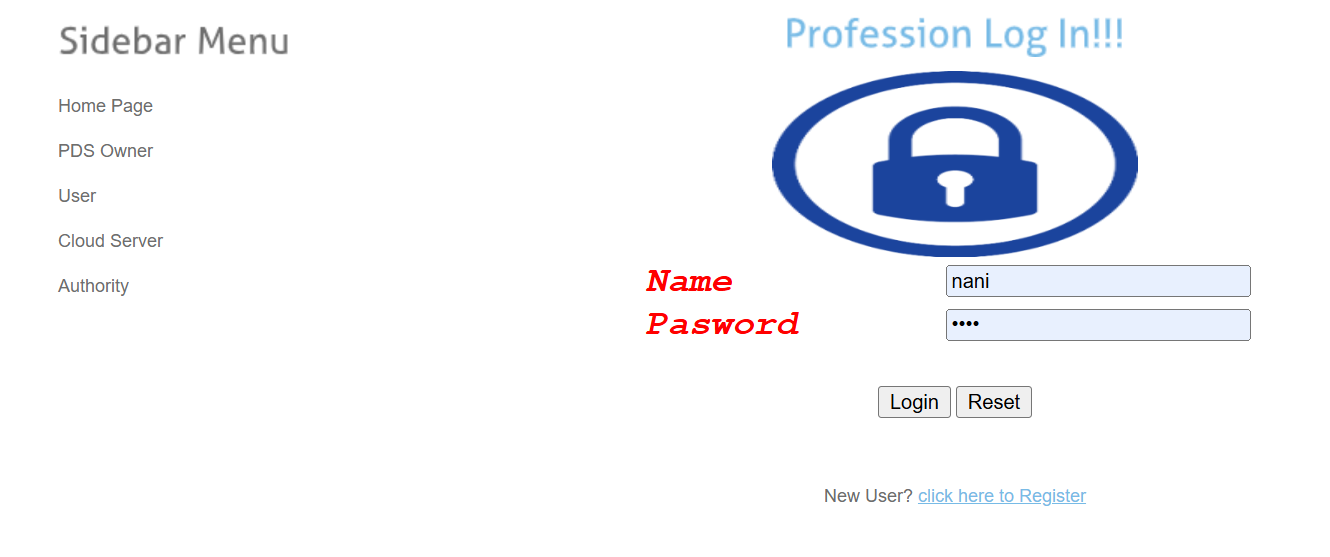
A strategy for system testing integrates system test cases and design techniques into a well planned series of steps that results in the successful construction of software. The testing strategy must co-operate test planning, test case design, test execution, and the resultant data collection and evaluation .A strategy for software testing must accommodate low-level tests that are necessary to verify that a small source code segment has been correctly implemented as well as high level tests that validate major system functions against user requirements.

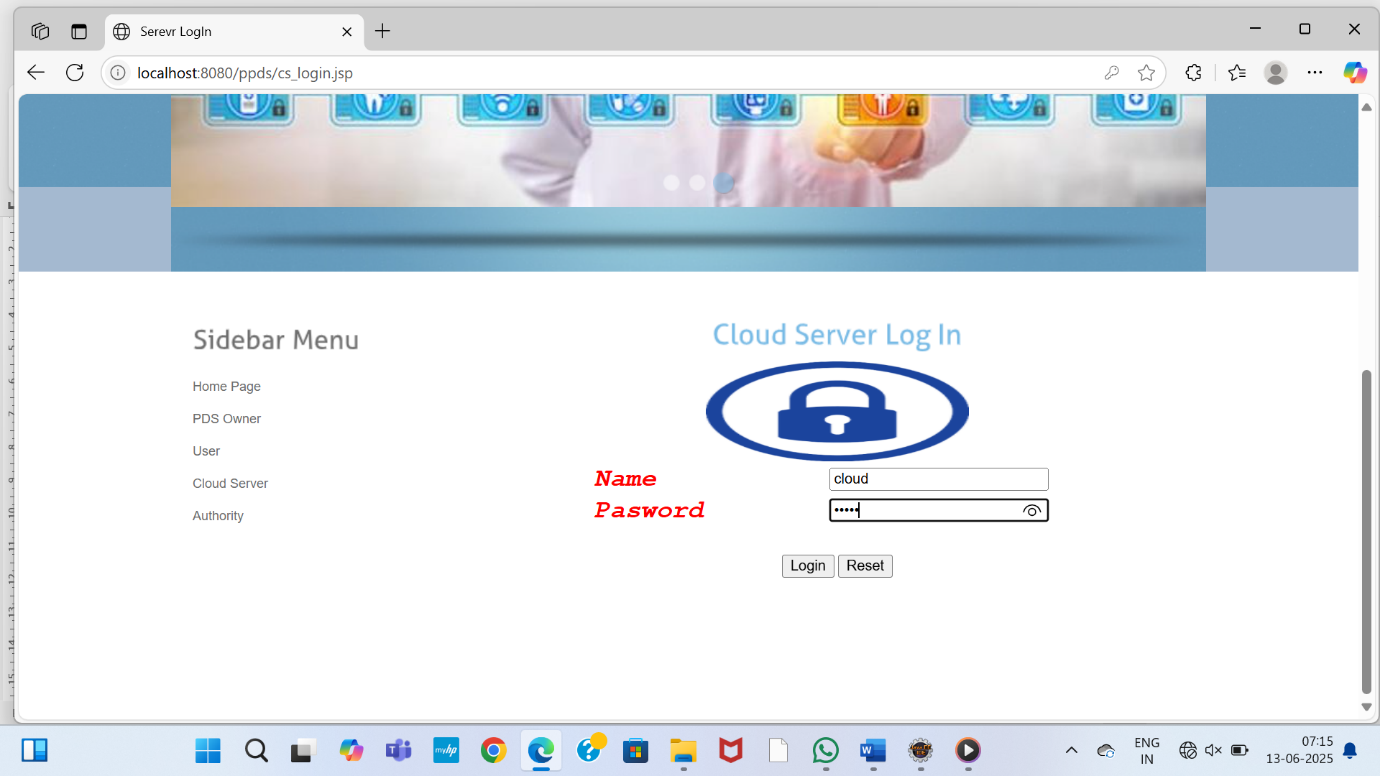
Software testing is a critical element of software quality assurance and represents the ultimate review of specification design and coding. Testing represents an interesting anomaly for the software. Thus, a series of testing are performed for the proposed system before the system is ready for user acceptance testing.

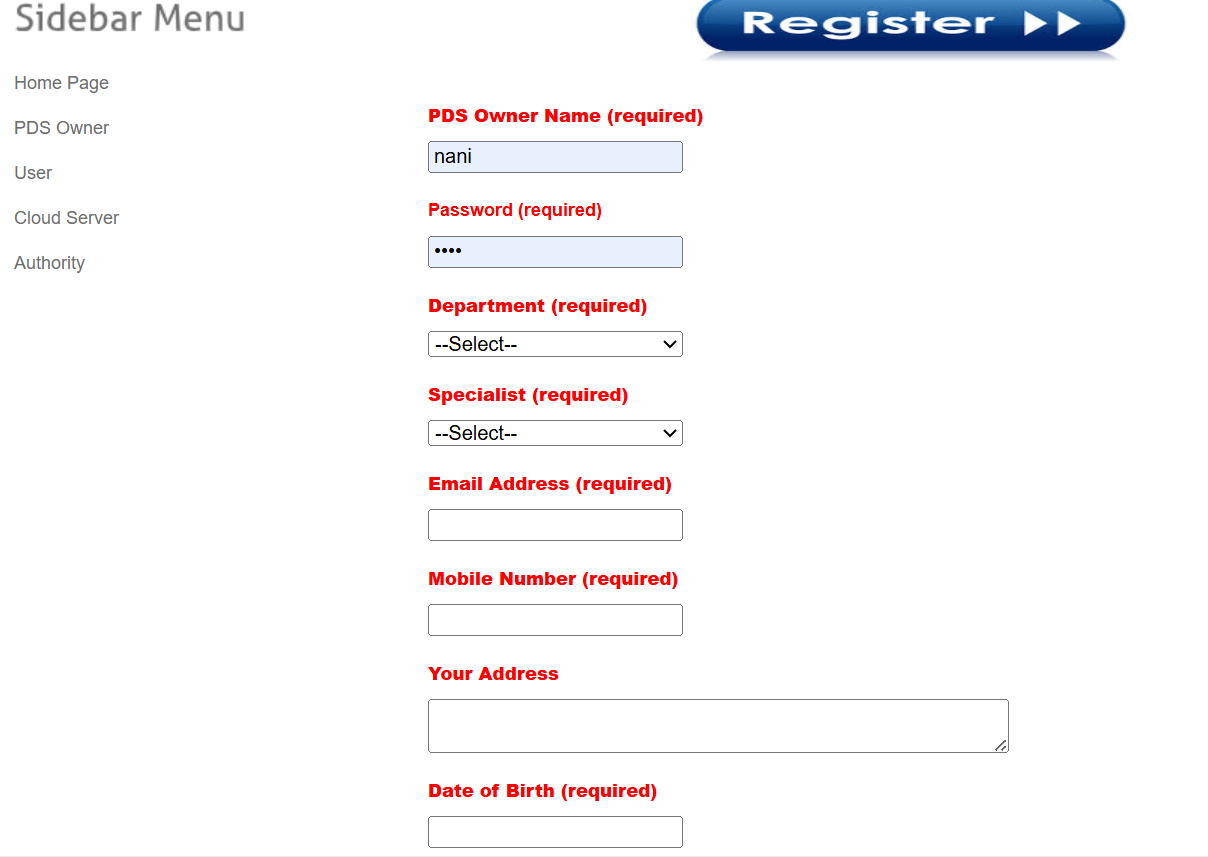
**SYSTEM TESTING:**

Software once validated must be combined with other system elements (e.g. Hardware, people, database). System testing verifies that all the elements are proper and that overall system function performance is achieved. It also tests to find discrepancies between the system and its original objective, current specifications and system documentation.

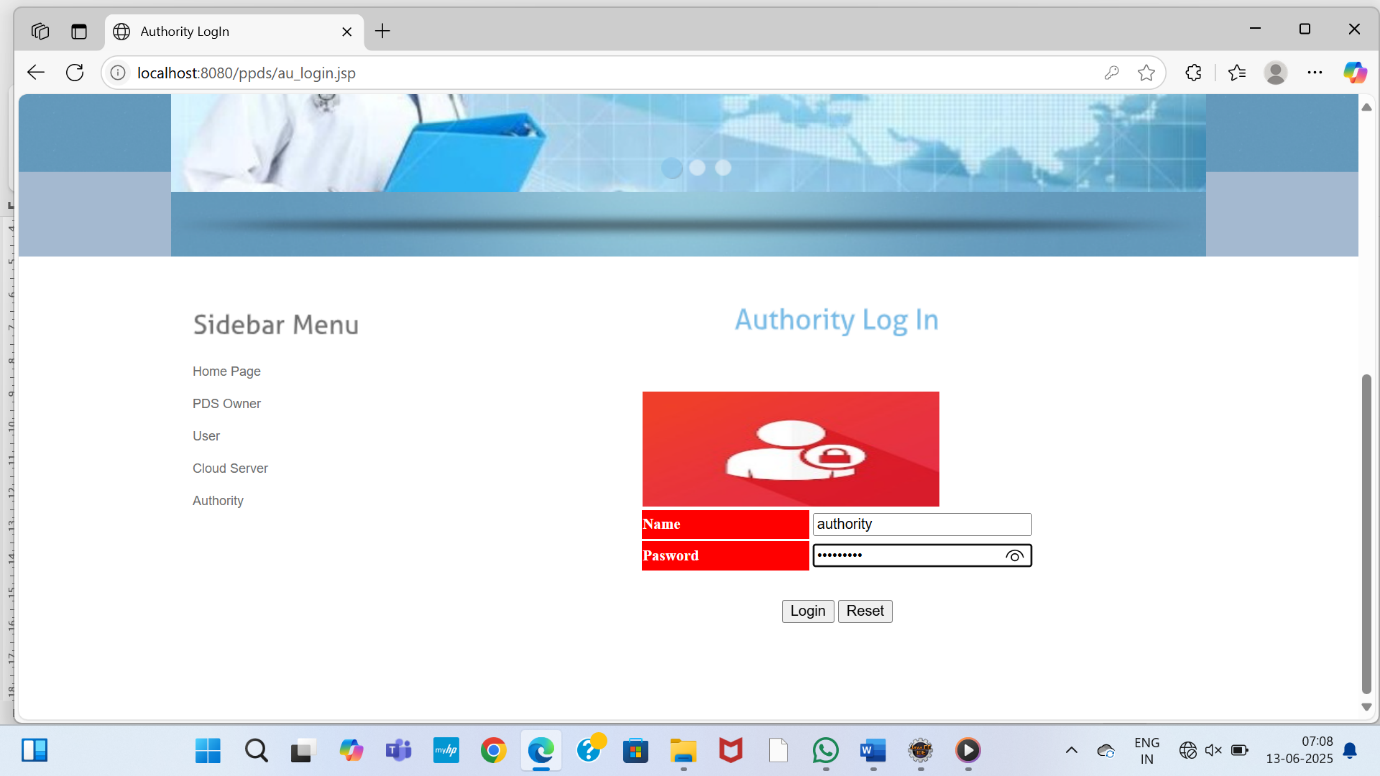


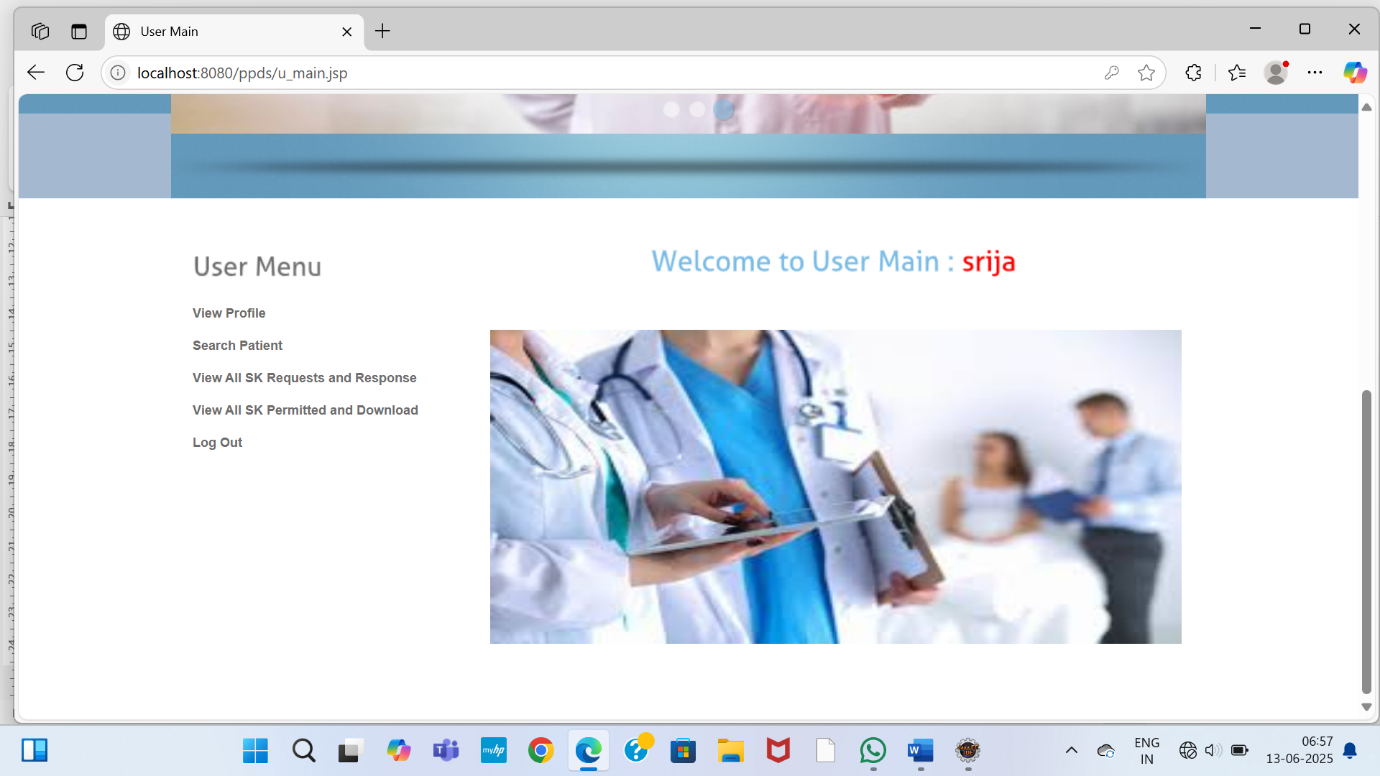


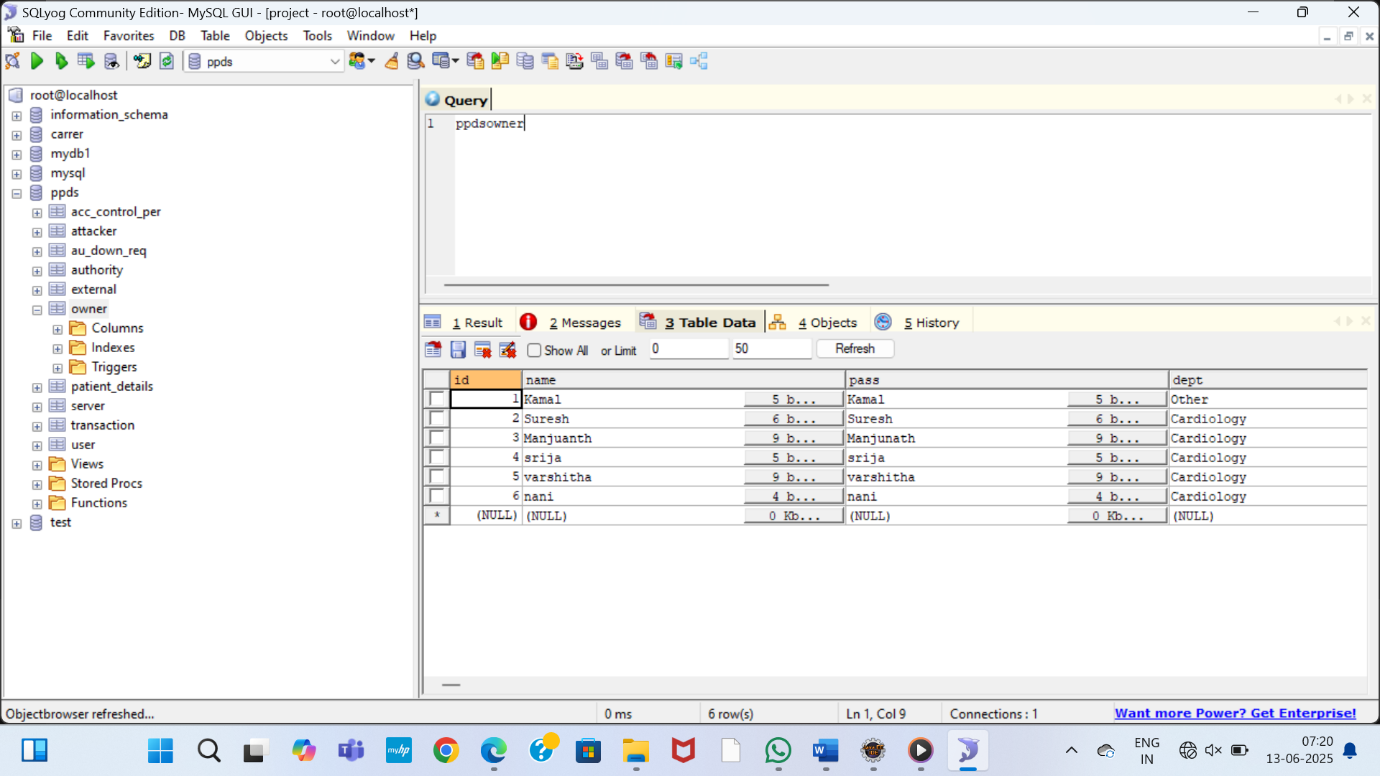
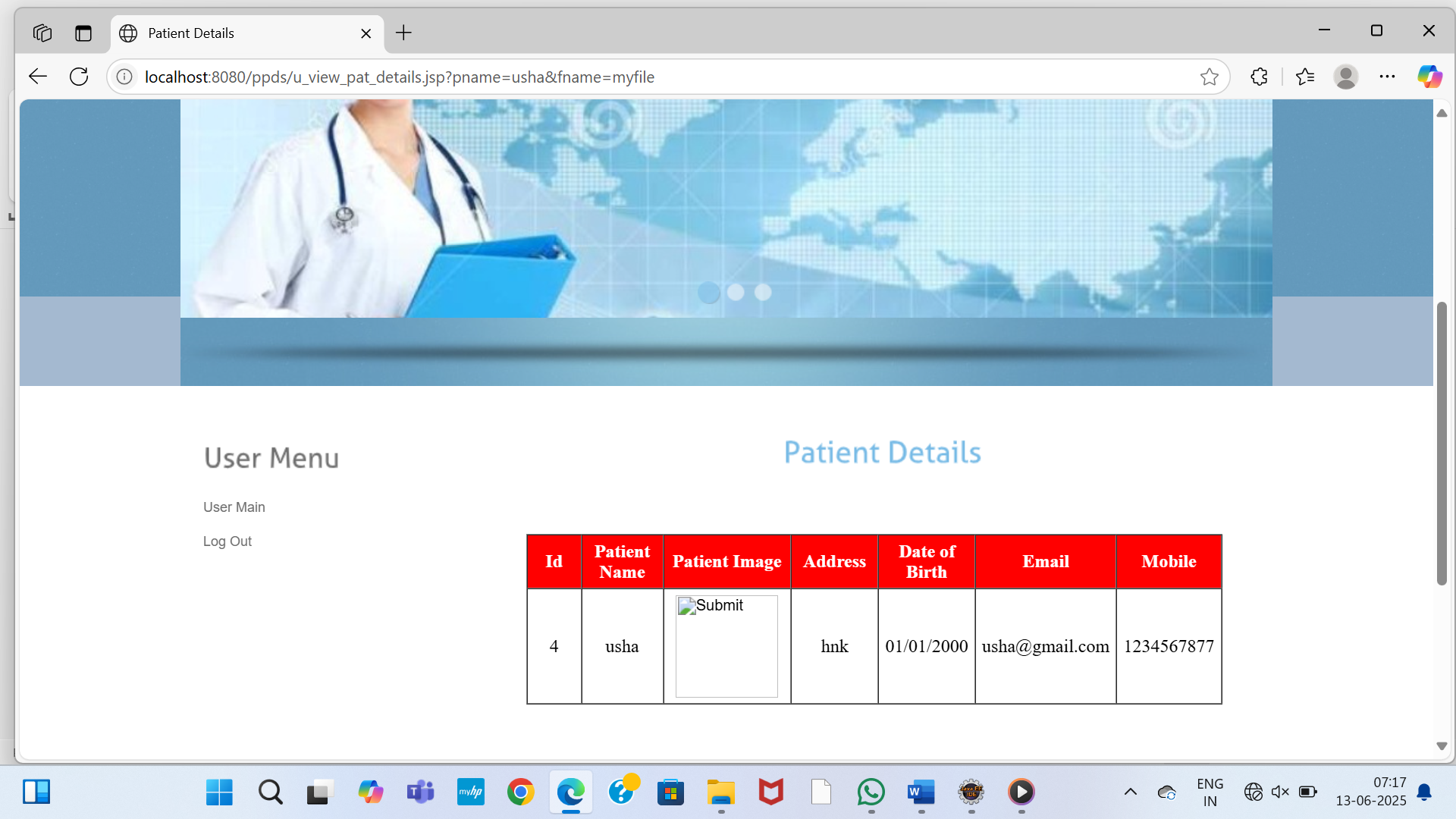
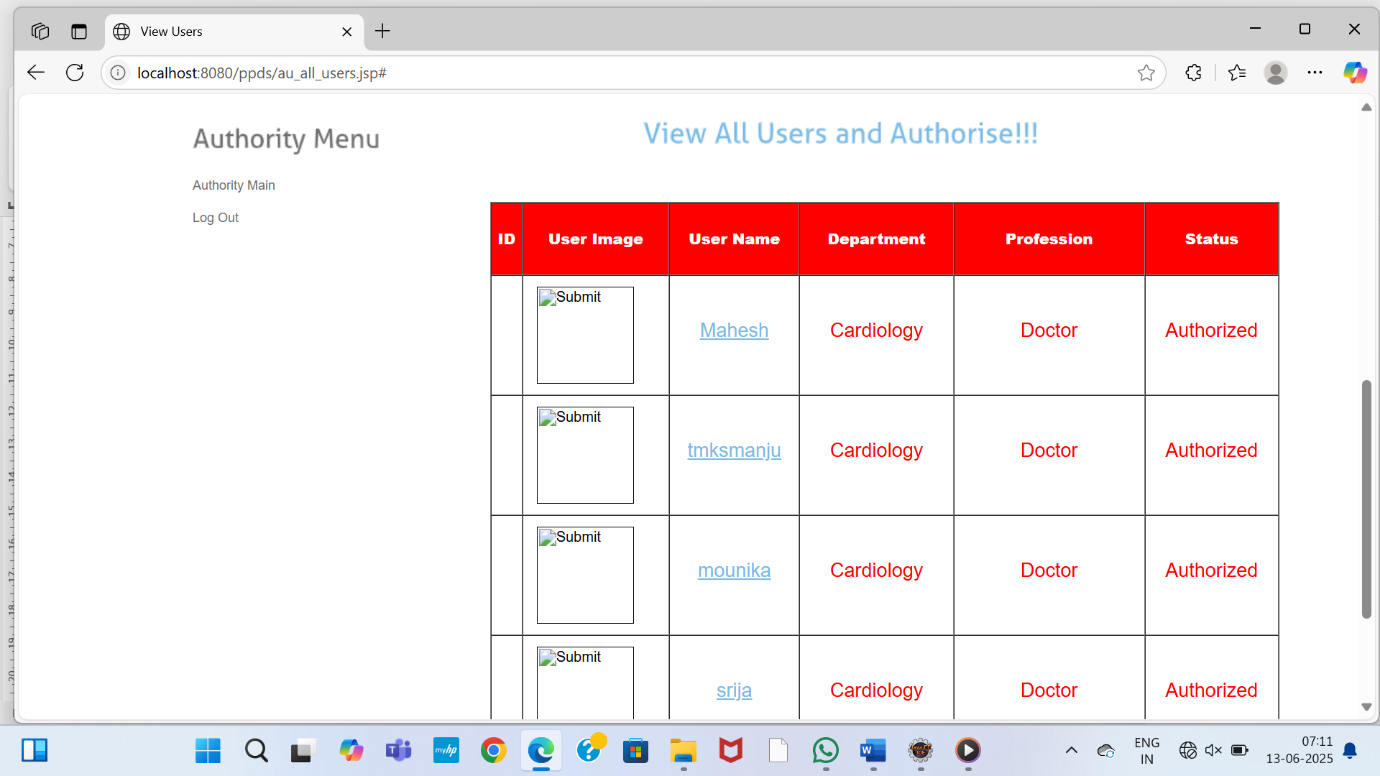


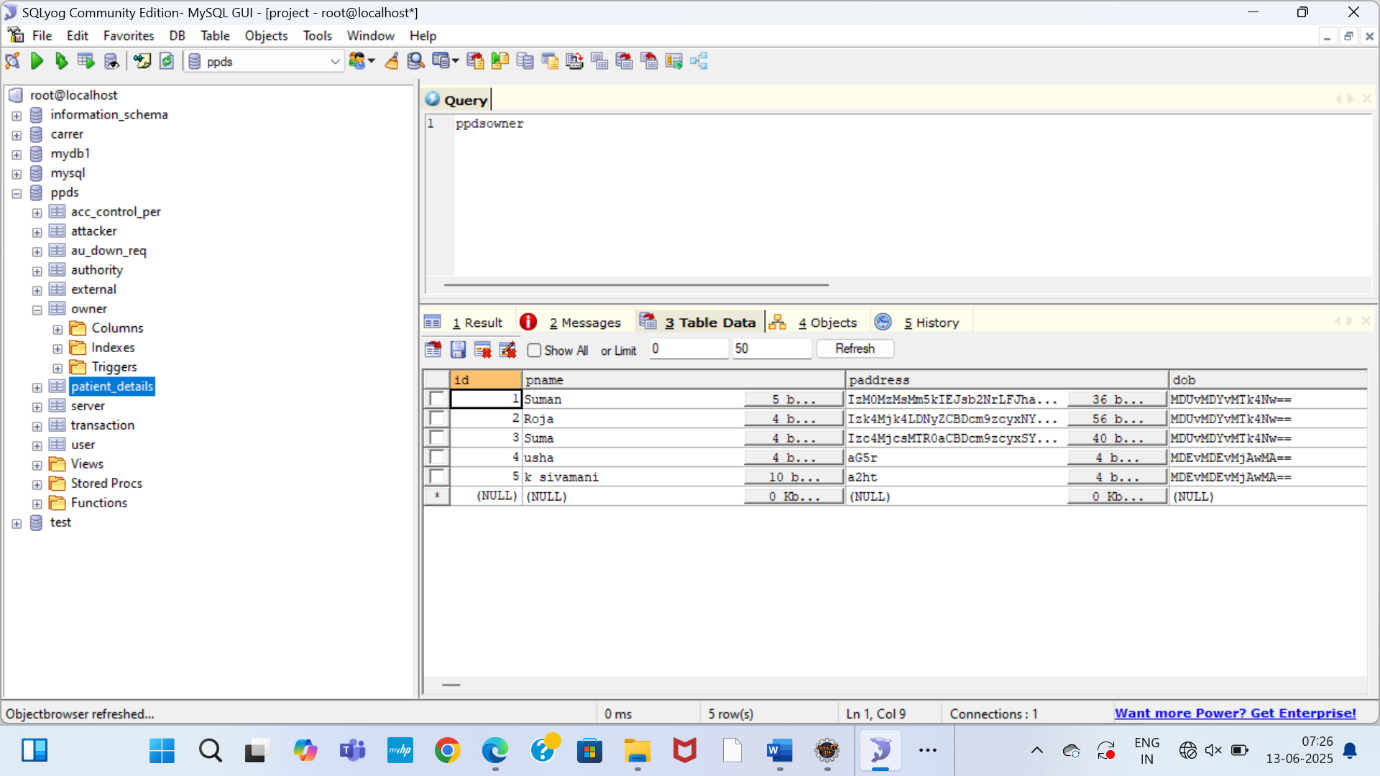


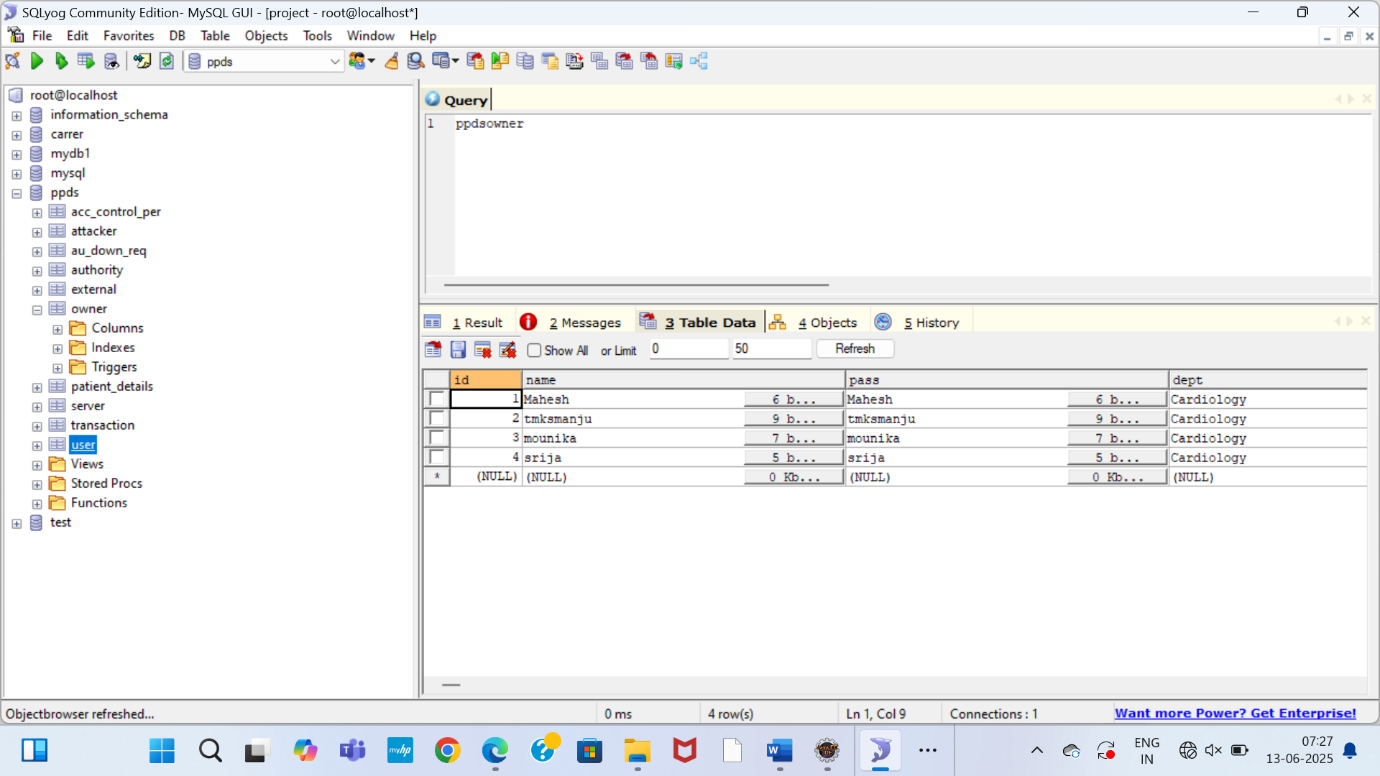












**SYSTEM TESTING**

### TESTING METHODOLOGIES

The following are the Testing Methodologies:

* **Unit Testing.**
* **Integration Testing.**
* **User Acceptance Testing.**
* **Output Testing.**
* **Validation Testing.**

**Unit Testing**

Unit testing focuses verification effort on the smallest unit of Software design that is the module. Unit testing exercises specific paths in a module’s control structure to

ensure complete coverage and maximum error detection. This test focuses on each module individually, ensuring that it functions properly as a unit. Hence, the naming is Unit Testing.

During this testing, each module is tested individually and the module interfaces are verified for the consistency with design specification. All important processing path are tested for the expected results. All error handling paths are also tested.

**Integration Testing**

Integration testing addresses the issues associated with the dual problems of verification and program construction. After the software has been integrated a set of high order tests are conducted. The main objective in this testing process is to take unit tested modules and builds a program structure that has been dictated by design.

**The following are the types of Integration Testing:**

1. **Top Down Integration**

This method is an incremental approach to the construction of program structure. Modules are integrated by moving downward through the control hierarchy, beginning with the main program module. The module subordinates to the main program module are incorporated into the structure in either a depth first or breadth first manner.

In this method, the software is tested from main module and individual stubs are replaced when the test proceeds downwards.

**2. Bottom-up Integration**

This method begins the construction and testing with the modules at the lowest level in the program structure. Since the modules are integrated from the bottom up, processing required for modules subordinate to a given level is always available and the need for stubs is eliminated. The bottom up integration strategy may be implemented with the following steps:

* The low-level modules are combined into clusters into clusters that perform a specific Software sub-function.
* A driver (i.e.) the control program for testing is written to coordinate test case input and output.
* The cluster is tested.
* Drivers are removed and clusters are combined moving upward in the program structure

The bottom up approaches tests each module individually and then each module is module is integrated with a main module and tested for functionality.

**7.1.3 User Acceptance Testing**

User Acceptance of a system is the key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with the prospective system users at the time of developing and making changes wherever required. The system developed provides a friendly user interface that can easily be understood even by a person who is new to the system.

**7.1.4 Output Testing**

After performing the validation testing, the next step is output testing of the proposed system, since no system could be useful if it does not produce the required output in the specified format. Asking the users about the format required by them tests the outputs generated or displayed by the system under consideration. Hence the output format is considered in 2 ways – one is on screen and another in printed format.

**7.1.5 Validation Checking**

Validation checks are performed on the following fields.

**Text Field:**

The text field can contain only the number of characters lesser than or equal to its size. The text fields are alphanumeric in some tables and alphabetic in other tables. Incorrect entry always flashes and error message.

**Numeric Field:**

The numeric field can contain only numbers from 0 to 9. An entry of any character flashes an error messages. The individual modules are checked for accuracy and what it has to perform. Each module is subjected to test run along with sample data. The individually tested modules are integrated into a single system. Testing involves executing the real data information is used in the program the existence of any program defect is inferred from the output. The testing should be planned so that all the requirements are individually tested.

A successful test is one that gives out the defects for the inappropriate data and produces and output revealing the errors in the system.

**Preparation of Test Data**

Taking various kinds of test data does the above testing. Preparation of test data plays a vital role in the system testing. After preparing the test data the system under study is tested using that test data. While testing the system by using test data errors are again uncovered and corrected by using above testing steps and corrections are also noted for future use.

**Using Live Test Data:**

Live test data are those that are actually extracted from organization files. After a system is partially constructed, programmers or analysts often ask users to key in a set of data from their normal activities. Then, the systems person uses this data as a way to partially test the system. In other instances, programmers or analysts extract a set of live data from the files and have them entered themselves.

It is difficult to obtain live data in sufficient amounts to conduct extensive testing. And, although it is realistic data that will show how the system will perform for the typical processing requirement, assuming that the live data entered are in fact typical, such data generally will not test all combinations or formats that can enter the system. This bias toward typical values then does not provide a true systems test and in fact ignores the cases most likely to cause system failure.

**Using Artificial Test Data:**

Artificial test data are created solely for test purposes, since they can be generated to test all combinations of formats and values. In other words, the artificial data, which can quickly be prepared by a data generating utility program in the information systems department, make possible the testing of all login and control paths through the program.

The most effective test programs use artificial test data generated by persons other than those who wrote the programs. Often, an independent team of testers formulates a testing plan, using the systems specifications.

The package “Virtual Private Network” has satisfied all the requirements specified as per software requirement specification and was accepted.

**7.2 USER TRAINING**

Whenever a new system is developed, user training is required to educate them about the working of the system so that it can be put to efficient use by those for whom the system has been primarily designed. For this purpose the normal working of the project was demonstrated to the prospective users. Its working is easily understandable and since the expected users are people who have good knowledge of computers, the use of this system is very easy.

**7.3** **MAINTAINENCE**

This covers a wide range of activities including correcting code and design errors. To reduce the need for maintenance in the long run, we have more accurately defined user’s requirements during the process of system development. Depending on the requirements, this system has been developed to satisfy the needs to the largest possible extent. With development in technology, it may be possible to add many more features based on the requirements in future. The coding and designing is simple and easy to understand which will make maintenance easier

**CONCLUSION**

This paper proposes a Privacy-aware Personal Data Storage, able to automatically take

privacy-aware decisions on third parties access requests in accordance with user

preferences. The system relies on active learning complemented with strategies to

strengthen user privacy protection. As discussed in the paper, we run several

experiments on a realistic dataset exploiting a group of 360 evaluators. The obtained

results show the effectiveness of the proposed approach. We plan to extend this work

along several directions. First, we are interested to investigate how P-PDS could scale

in the IoT scenario, where access requests decision might depend also on contexts, not

only on user preferences. Also, we would like to integrate P-PDS with cloud computing

services (e.g., storage and computing) so as to design a more powerful P-PDS by, at the

same time, protecting users privacy.

**REFERENCES**

[1] B. C. Singh, B. Carminati, and E. Ferrari, “Learning privacy habits of pds owners,” in Distributed Computing Systems (ICDCS), 2017 IEEE 37th International Conference on. IEEE, 2017, pp. 151–161.

[2] Y.-A. de Montjoye, E. Shmueli, S. S. Wang, and A. S. Pentland, “openpds: Protecting the privacy of metadata through safeanswers,” PloS one, vol. 9, no. 7, p. e98790, 2014.

[3] B. M. Sweatt et al., “A privacy-preserving personal sensor data ecosystem,” Ph.D. dissertation, Massachusetts Institute of Technology, 2014.

[4] B. C. Singh, B. Carminati, and E. Ferrari, “A risk-benefit driven architecture for personal data release,” in Information Reuse and Integration (IRI), 2016 IEEE 17th International Conference on. IEEE, 2016, pp. 40–49.

[5] M. Madejski, M. Johnson, and S. M. Bellovin, “A study of privacy settings errors in an online social network,” in Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on. IEEE, 2012, pp. 340–345.

[6] L. N. Zlatolas, T. Welzer, M. Heriˇcko, and M. H¨ olbl, “Privacy antecedents for sns self-disclosure: The case of facebook,” Computers in Human Behavior, vol. 45, pp. 158–167, 2015.

[7] D. A. Albertini, B. Carminati, and E. Ferrari, “Privacy settings recommender for online social network,” in Collaboration and Internet Computing (CIC), 2016 IEEE 2nd International Conference on. IEEE, 2016, pp. 514–521.

[8] A. Acquisti and R. Gross, “Imagined communities: Awareness, information sharing, and privacy on the facebook,” in International workshop on privacy enhancing technologies. Springer, 2006, pp. 36– 58.

[9] R. Gross and A. Acquisti, “Information revelation and privacy in online social networks,” in Proceedings of the 2005 ACM workshop on Privacy in the electronic society. ACM, 2005, pp. 71–80.

[10] Y. Liu, K. P. Gummadi, B. Krishnamurthy, and A. Mislove, “Analyzing facebook privacy settings: user expectations vs. reality,” in Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference. ACM, 2011, pp. 61–70.

[11] P. Nyoni and M. Velempini, “Privacy and user awareness on facebook,” South African Journal of Science, vol. 114, no. 5-6, pp. 27–31, 2018.

[12] R. Polikar, “Ensemble learning,” in Ensemble machine learning. Springer, 2012, pp. 1–34.

[13] T. G. Dietterich et al., “Ensemble methods in machine learning,” Multiple classifier systems, vol. 1857, pp. 1–15, 2000.

[14] B. Settles, “Active learning literature survey,” University of Wisconsin, Madison, vol. 52, no. 55-66, p. 11, 2010.

[15] X. Zhu, “Semi-supervised learning literature survey,” Computer Science, University of Wisconsin-Madison, vol. 2, no. 3, p. 4, 2006.

[16] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, “Supervised machine learning: A review of classification techniques,” Emerging artificial intelligence applications in computer engineering, vol. 160, pp. 3–24, 2007.

[17] S. Borman, “The expectation maximization algorithm-a short tutorial,” Submitted for publication, pp. 1–9, 2004.

[18] D. Cohn, L. Atlas, and R. Ladner, “Improving generalization with active learning,” Machine learning, vol. 15, no. 2, pp. 201–221, 1994.

[19] M.-R. Bouguelia, Y. Bela¨ıd, and A. Bela¨ıd, “A stream-based semisupervised active learning approach for document classification,” in Document Analysis and Recognition (ICDAR), 2013 12th International Conference on. IEEE, 2013, pp. 611–615.

[20] A. K. McCallumzy and K. Nigamy, “Employing em and poolbased active learning for text classification,” in Proc. International Conference on Machine Learning (ICML). Citeseer, 1998, pp. 359– 367.

[21] I. Muslea, S. Minton, and C. A. Knoblock, “Active+ semisupervised learning= robust multi-view learning,” in ICML, vol. 2, 2002, pp. 435–442.

[22] D. Martin, M. Paolucci, S. McIlraith, M. Burstein, D. McDermott, D. McGuinness, B. Parsia, T. Payne, M. Sabou, M. Solanki et al., “Bringing semantics to web services: The owl-s approach,” in International Workshop on Semantic Web Services and Web Process Composition. Springer, 2004, pp. 26–42.

[23] D. Martin, M. Burstein, J. Hobbs, O. Lassila, D. McDermott, S. McIlraith, S. Narayanan, M. Paolucci, B. Parsia, T. Payne et al., “Owl-s: Semantic markup for web services,” W3C member submission, vol. 22, pp. 2007–04, 2004.

[24] P. Jaccard, “ ´ Etude comparative de la distribution florale dans une portion des alpes et des jura,” Bulletin del la Soci´et´e Vaudoise des Sciences Naturelles, vol. 37, pp. 547–579, 1901.

[25] S. M. Mohammad and G. Hirst, “Distributional measures of semantic distance: A survey,” arXiv preprint arXiv:1203.1858, 2012.

[26] I. Zliobaite, A. Bifet, B. Pfahringer, and G. Holmes, “Active learning with drifting streaming data,” IEEE transactions on neural networks and learning systems, vol. 25, no. 1, pp. 27–39, 2014.

[27] T. Chajed, J. Gjengset, M. F. Kaashoek, J. Mickens, R. Morris, and N. Zeldovich, “Oort: User-centric cloud storage with global queries,” 2016.

[28] F. Wang, J. Mickens, N. Zeldovich, and V. Vaikuntanathan, “Sieve: Cryptographically enforced access control for user data in untrusted clouds.” in NSDI, 2016, pp. 611–626.

[29] T. Chajed, J. Gjengset, J. Van Den Hooff, M. F. Kaashoek, J. Mickens, R. Morris, and N. Zeldovich, “Amber: Decoupling user data from web applications.” in HotOS, vol. 15, 2015, pp. 1–6.

[30] R. Mortier, J. Zhao, J. Crowcroft, L. Wang, Q. Li, H. Haddadi, Y. Amar, A. Crabtree, J. Colley, T. Lodge et al., “Personal data management with the databox: What’s inside the box?” in Proceedings of the 2016 ACM Workshop on Cloud-Assisted Networking. ACM, 2016, pp. 49–54.

[31] Z. Yan, G. Gan, and K. Riad, “Bc-pds: Protecting privacy and selfsovereignty through blockchains for openpds,” in Service-Oriented System Engineering (SOSE), 2017 IEEE Symposium on. IEEE, 2017, pp. 138–144.

[32] L. Fang and K. LeFevre, “Privacy wizards for social networking sites,” in Proceedings of the 19th international conference on World wide web. ACM, 2010, pp. 351–360.

[33] G. Danezis, “Inferring privacy policies for social networking services,” in Proceedings of the 2nd ACM workshop on Security and artificial intelligence. ACM, 2009, pp. 5–10.

[34] N. Sadeh, J. Hong, L. Cranor, I. Fette, P. Kelley, M. Prabaker, and J. Rao, “Understanding and capturing peoples privacy policies in a mobile social networking application,” Personal and Ubiquitous Computing, vol. 13, no. 6, pp. 401–412, 2009.

[35] I. Bilogrevic, K. Huguenin, B. Agir, M. Jadliwala, M. Gazaki, and J.-P. Hubaux, “A machine-learning based approach to privacyaware information-sharing in mobile social networks,” Pervasive and Mobile Computing, vol. 25, pp. 125–142, 2016.

[36] T. Nakamura, S. Kiyomoto, W. B. Tesfay, and J. Serna, “Easing the burden of setting privacy preferences: A machine learning approach,” in International Conference on Information Systems Security and Privacy. Springer, 2016, pp. 44–63.