QUERY DRIVEN APPROACH TO ENTITY RESOLUTION

A Mini Project Report Submitted to

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, HYDERABAD

In Partial Fulfillment of the requirement For the Award of the Degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

Submitted by

1.U.POOJA   **(H.T.N0:17N01A05A9)**

2. POOJA ROY   **(H.T.N0:17N01A0588)**

3.V.SANJAY   **(H.T.N0:17N01A05B1)**

4. MD.SUMER ALI   **(H.T.N0:17N05A0569)**

Under the Supervision of

DR.MANISH TIWARI

Associate Professor

Department of Computer Science and Engineering

SREE CHAITANYA COLLEGE OF ENGINEERING

(Affiliated to JNTUH, HYDERABAD)

THIMMAPUR, KARIMNAGAR, TELANGANA-505 527
FEBRUARY-2021

SREE CHAITANYA COLLEGE OF ENGINEERING

(Affiliated to JNTUH, HYDERABAD
THIMMAPUR, KARIMNAGAR, TELANGANA-505 527

Department of Computer Science and Engineering

CERTIFICATE

This is to certify that **U.POOJA, POOJA ROY, V.SANJAY, MD.SUMER ALI,** bearing hallticket numbers **17N01A05A9, 17N01A0588, 17N01A05B1, 17N05A0569** are the students of **SREE CHAITANYA COLLEGE OF ENGINEERING**, KARIMNAGAR pursuing the Degree of B.Tech., in **COMPUTER SCIENCE AND ENGINEERING**, has carried out Mini Project work/ Internship titled **"QUERY DRIVEN APPROACH TO ENTITYRESOLUTION** " under our guidance during the period from **15/02/2021** to **20/02/2021** in partial fulfillment of the requirements for the award of the above mentioned Degree. The students are punctual, hardworking and shown keen interest to produce the project output and results.

Regards,

Mini Project Guide   **Head of the Department**

Dr.MANISH TIWARI   **Mr. KHAJA ZIAUDDIN**
Associate Professor   **Associate Professor**
Department of CSE   **Department of CSE**

EXTERNAL EXAMINER

Date : ....

MINI PROJECT / INTERNSHIP COMPLETION CERTIFICATE

This is to certify that **U.POOJA, POOJA ROY, V.SANJAY, MD.SUMER ALI,** bearing hall ticket numbers **17N01A05A9, 17N01A0588, 17N01A05B1, 17N05A0569**, are the students of **SREE CHAITANYA COLLEGE OF ENGINEERING**, KARIMNAGAR pursuing the Degree of B.Tech., in **COMPUTER SCIENCE AND ENGINEERING**, has carried out Mini Project work/ Internship titled **" QUERY DRIVEN APPROACH TO ENTITY RESOLUTION"** under our guidance during the period from **15/02/2021** to **20/02/2021** in partial fulfillment of the requirements for the award of the above mentioned Degree. The student are punctual, hardworking and shown keen interest to produce the project output and results.

Regards,

SREE CHAITANYA COLLEGE OF ENGINEERING

(Affiliated to JNTUH, HYDERABAD)
THIMMAPUR, KARIMNAGAR, TELANGANA-505 527

Department of Computer Science and Engineering

DECLARATION

I, U.POOJA, POOJA ROY, V.SANJAY, MD.SUMER ALI, is student of **Bachelor of Technology in Computer Science and Engineering,** during the academic year:2018-2019**,**hereby declare that the work presented in this Project Work entitled "**Query Driven Approach To Entity Resolution"** is the outcome of our own bona fide work and is correct to the best of our knowledge and this work has been undertaken taking care of Engineering Ethics and carried out under the supervision of **Dr.MANISH TIWARI, Associate Professor**.

It contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

1.U.POOJA   (H.T.N0:17N01A05A9)
2. POOJA ROY (H.T.N0:17N01A0588)
3.V.SANJAY   (H.T.N0:17N01A05B1)
4. MD.SUMER ALI (H.T.N0:17N05A0569)

Date:

Place:

SREE CHAITANYACOLLEGE OF ENGINEERING
(Affiliated to JNTUH, HYDERABAD)
THIMMAPUR, KARIMNAGAR, TELANGANA-505 527
Department of Computer Science and Engineering

ACKNOWLEDGEMENTS

The Satisfaction that accomplishes the successful completion of any task would be incomplete without the mention of the people who make it possible and whose constant guidance and encouragement crown all the efforts with success.

I would like to express my sincere gratitude and indebtedness to my project supervisor **Dr.MANISH TIWARI, Associate Professor**, Department of Computer Science and Engineering, Sree Chaitanya College of Engineering, LMD Colony, Karimnagar for his/her valuable suggestions and interest throughout the course of this project

I am also thankful to Head of the department **Mr.Khaja Ziauddin, Associate Professor & HOD**, Department of Computer Science and Engineering, Sree Chaitanya College of Engineering, LMD Colony, Karimnagar for providing excellent infrastructure and a nice atmosphere for completing this project successfully

We Sincerely extend out thanks to **Dr.G.Venkateswarlu, Principal,** Sree Chaitanya College of Engineering, LMD Colony, Karimnagar, for providing all the facilities required for completion of this project.

I convey my heartfelt thanks to the lab staff for allowing me to use the required equipment whenever needed.

Finally, I would like to take this opportunity to thank my family for their support through the work.

I sincerely acknowledge and thank all those who gave directly or indirectly their support in completion of this work.

> 1.U.POOJA   (H.T.N0:17N01A05A9)
>
> 2. POOJA ROY   (H.T.N0:17N01A0588)
>
> 3.V.SANJAY   (H.T.N0:17N01A05B1)
>
> 4. MD.SUMER ALI (H.T.N0:17N05A0569)

ABSTRACT

This paper explores "on-the-fly" data cleaning in the context of a user query. A novel Query-Driven Approach (QDA) is developed that performs a minimal number of cleaning steps that are only necessary to answer a given selection query correctly. The comprehensive empirical evaluation of the proposed approach demonstrates its significant advantage in terms of efficiency over traditional techniques for query-driven applications. The significance of data quality research is motivated by the observation that the of data-driven technologies such as decision support tools, data exploration, analysis, and scientific discovery tools is closely tied to the quality of data to which such techniques are applied. It is well recognized that the outcome of the analysis is only as good as the data on which the analysis is performed. That is why today organizations spend a substantial percentage of their budgets on cleaning tasks such as removing duplicates, correcting errors, and lying missing values, to improve data quality prior to pushing data through the analysis pipeline. Given the critical importance of the problem, many efforts, in both industry and academia, have explored systematic approaches to addressing the cleaning challenges.

TABLE OF CONTENTS

viii

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

This paper addresses the problem of **query-aware** data cleaning, wherein the needs of the query dictates which parts of the data should be cleaned. Query-aware cleaning is emerging as a new paradigm for data cleaning to support today's increasing demand for (near) real-time analytical applications. Modern enterprises have access to potentially limitless data sources, e.g., web data repositories, social media posts, clickstream data, etc. Analysts usually wish to integrate one or more such data sources (possibly with their own data) to perform joint analysis and decision making. As a result of merging data from different sources, a given real-world object may often have multiple representations, resulting in data quality challenges. In this paper, we focus on the Entity Resolution (ER) challenge.

1.1 Existing System:

- These rules can be discovered from existing high quality data such as master data or manually identified data. Inspired by the swoosh method, each cluster is then merged into a composite record via a merge function. Finally a traditional ER method, denoted by T-ER, can be applied to identify the new data set. Moreover, in order to identify more records, the current ER result can be used as the training data to discover new ER-rules. The training data can also be obtained by using techniques, such as relevant feedback, crowd sourcing and knowledge extraction from the web. Therefore, with the accumulated information, ER-rules for more entities can be discovered.

- Invalid rule. A rule r is invalid if there exist records that match LHS(r) but do not refer to RHS(r) . Invalid rules might be discovered when the information of entities is not comprehensive. For example, suppose the training data set involves the records. The rule r: (name ¼"wei wang")^(coa 2"zhang")) e1 can be generated. For o31, it matches LHS(r) but does not refer to e1. Therefore, r is an invalid rule.

1.2 Proposed System:

Entity resolution is a well-known problem and it has received signicant attention in the literature over the past few decades. A thorough overview of the existing work in this area can be found in surveys . We classify the ER techniques into two categories as follow:Generic ER. A typical ER cycle consists of several phases of data transformations that include: normalization, blocking,similarity computation, clustering, and merging ,which can be intermixed. In the normalization phase, the ER framework standardizes the data formats. The next phase is blocking which is a main traditional mechanism used for improving ER effciency The primary motivation of this paper is querie on online data.A key concept driving the QDA approach is that of vestigiality. A cleaning step (i.e., call to the resolve function fora pair of records) is called vestigial (redundant) if QDA can guarantee that it can still compute a correct answer without knowing the outcome of this resolve. We formalize the concept of vestigiality in the context of a large class of SQL selection queries and develop techniques to identify vestigial cleaning steps

1.3 System Configuration:

Hardware Requirements:

Hardware	-	Pentium IV
Speed	-	3.5 GHz
RAM	-	1GB
Hard Disk	-	20 GB
Key Board	-	Standard Windows Keyboard
Mouse	-	Two or Three Button Mouse
Monitor	-	SVGA

1.4 Software Requirements:

Operating System	: Windows
Technology	: Java and J2EE
Web Technologies	: Html, JavaScript, CSS
IDE	: My Eclipse
Web Server	: Tomcat
Database	: My SQL
Java Version	: J2SDK1

CHAPTER 2

SYSTEM STUDY

2.1 Feasibility Study

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

- ♦ Economical Feasibility
- ♦ Technical Feasibility
- ♦ Social Feasibility

2.2 Economical Feasibility

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

2.3 Technical Feasibility

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

2.4 Social Feasibility

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

2.5 Module Description:

Number of Modules:After careful analysis the system has been identified to have the following:

Modules:

1. User Module
2. Admin Module
3. Chart Module
4. Query Driven Approach Module

2.5.1 User Module

User module, the new user should register application form , before enter the particular site, after login , user should create the profile for that particular login user, user can search any author details ,they can view also related author details.

2.5.2 Admin Module:

Admin is a super user. they can view all the author details.admin can view the chart based on author details, admin can view the user details.A typical ER cycle consists of several phases of data transformations that include: normalization, blocking, similarity computation, clustering, and merging which can be intermixed.

2.5.3 Query Driven Approach Module:

Process of Query Driven Approach. Mentioned below is process of query driven data warehousing approach − When a query is issued to a client side, a

metadata dictionary translates the query into the queries, appropriate for the individual heterogeneous site involved.

2.5.4 Chart Module:

query driven approach project we will generate a chart based on author details , so user admin can easily find out which author can realise the most number of articals .

CHAPTER 3

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

3.1 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.

3.2 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 analysed are

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

3.2.1 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.

3.2.2 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.

3.2.3 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 develop the system. The technical feasibility has been carried out. The system is technically feasible for development and can be developed with the existing facility.

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.

Flow Chart1: User
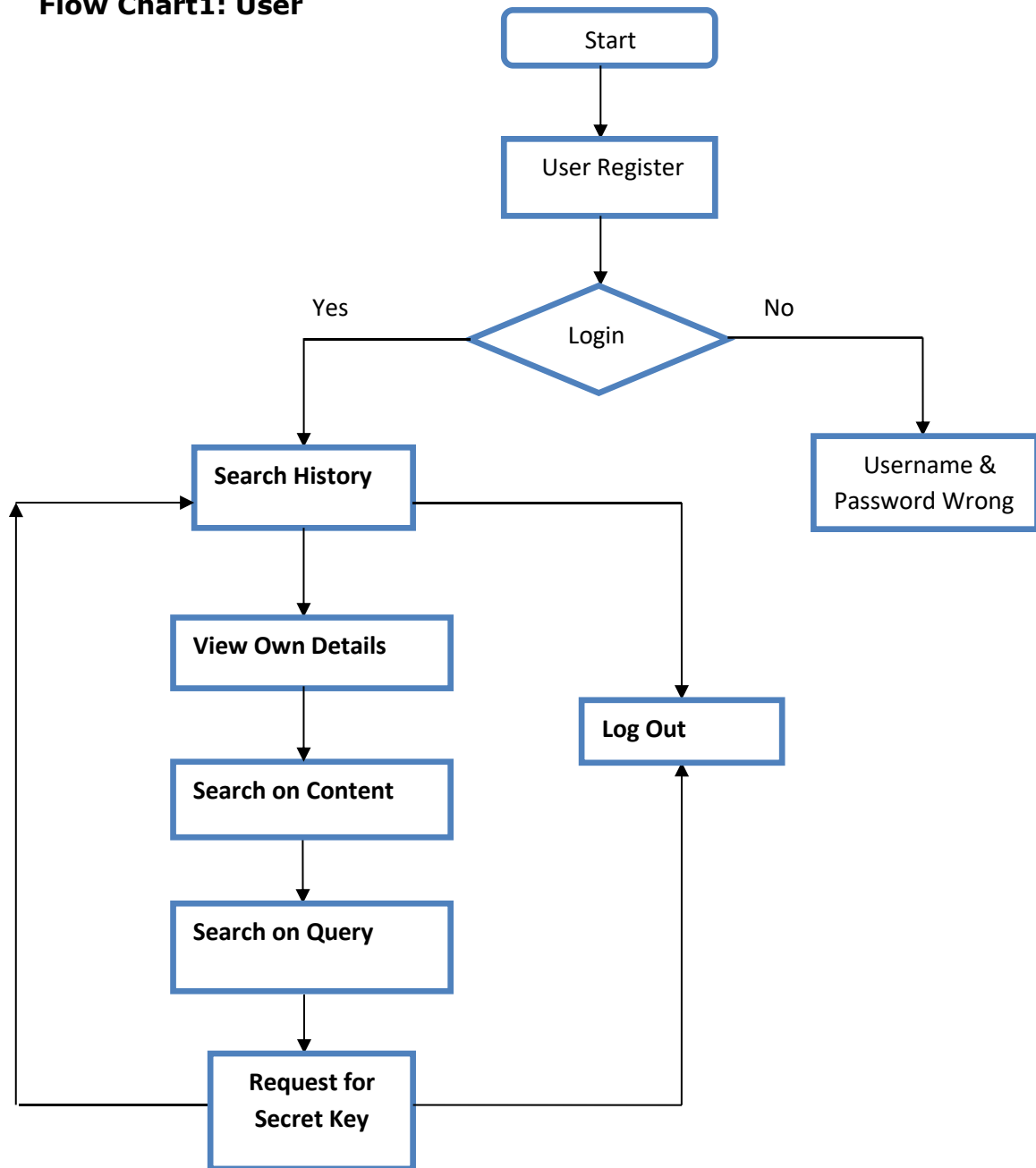
Fig 3.1 user flow chart

Flow Chart2: Author
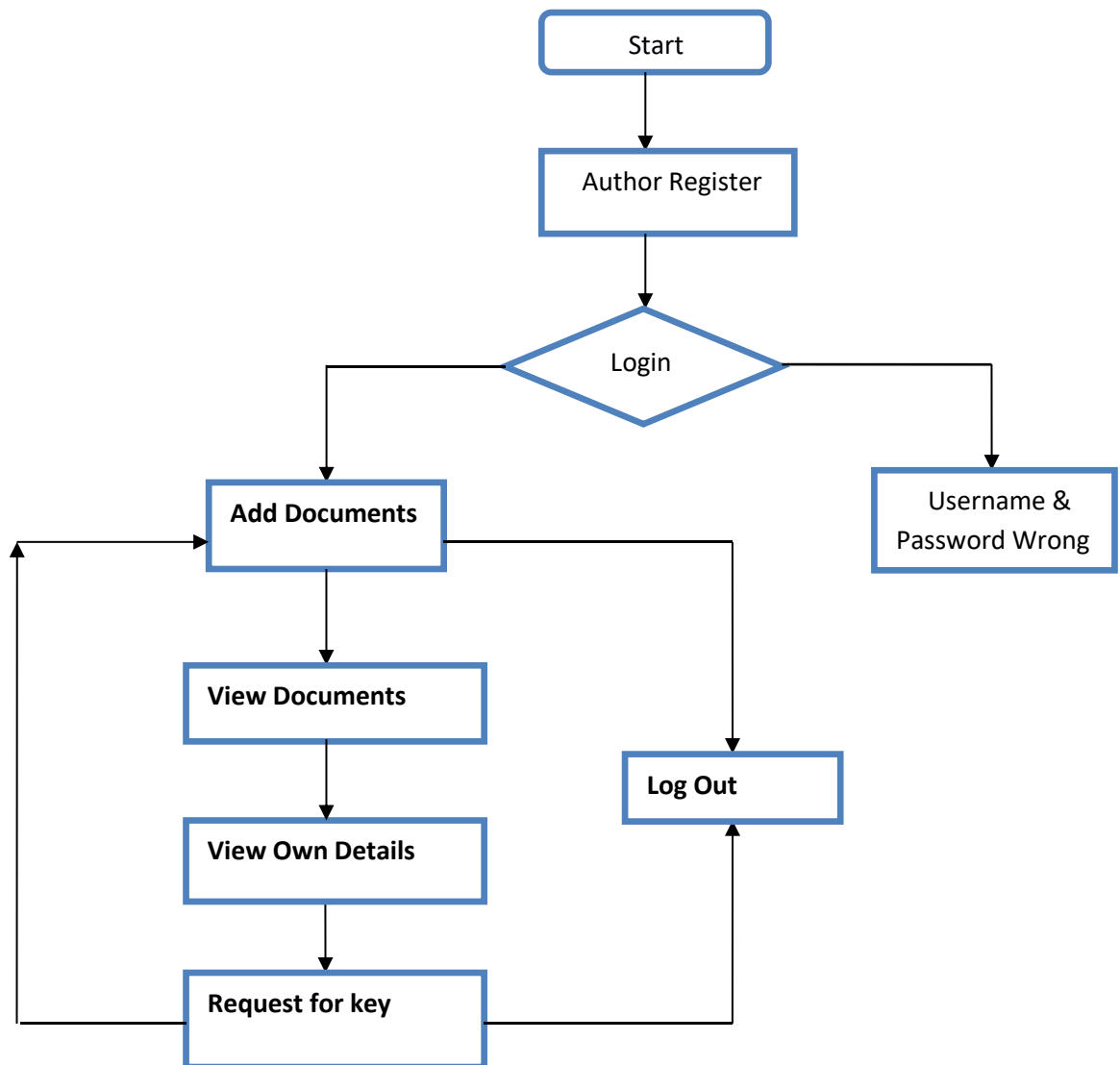
Fig 3.2 Author Flow Chart

Flow Chart 3: Admin
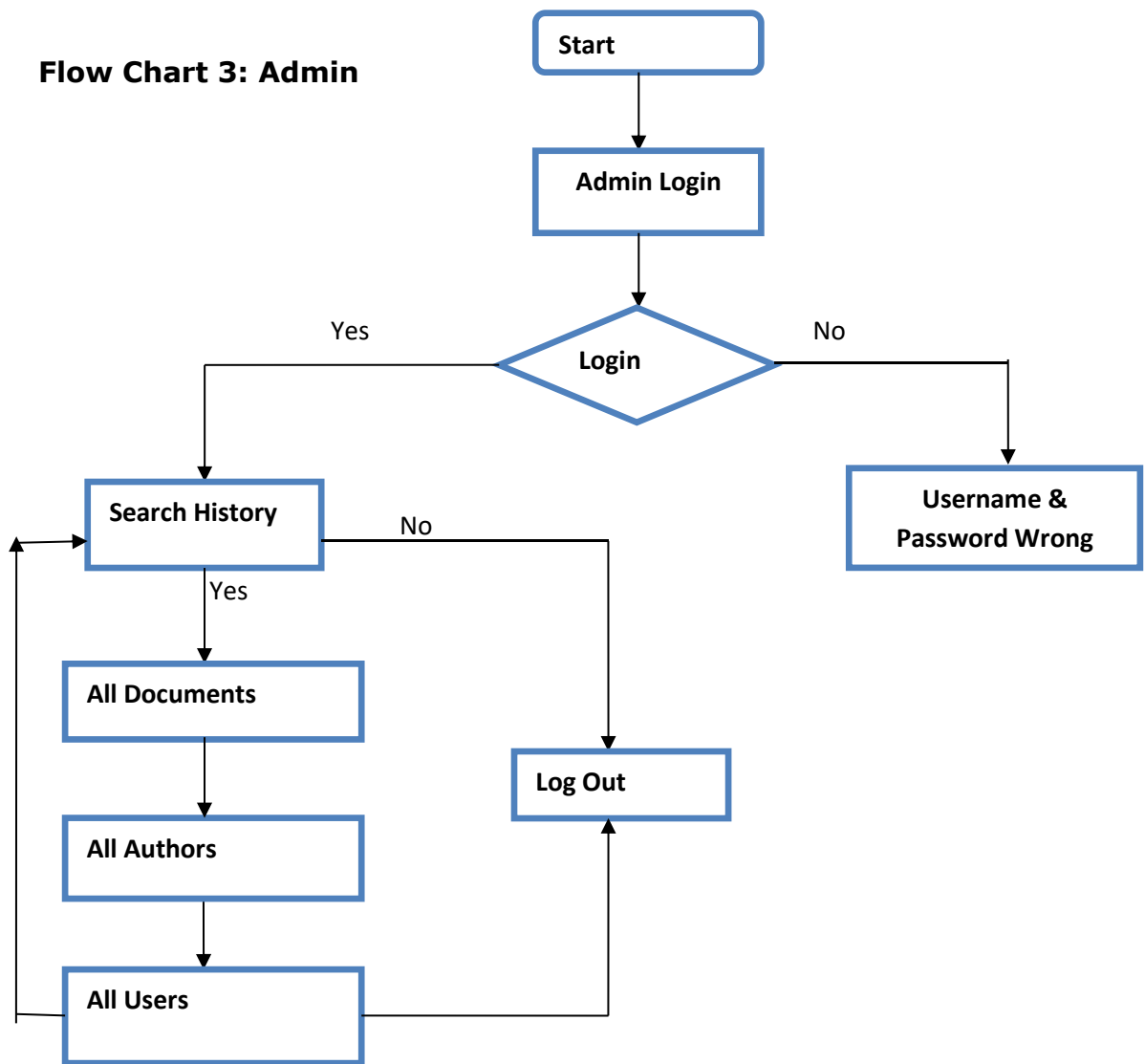
Fig 3.3 Admin Flow Chart

Sequence Diagram:

Admin

Web Server | User | Author

Register with the system

Search History

View their own details

All Documents

Search on Query & Content

List of all Users

Add / View Documents

List of all Authors

Request for Secret Key

View their own details
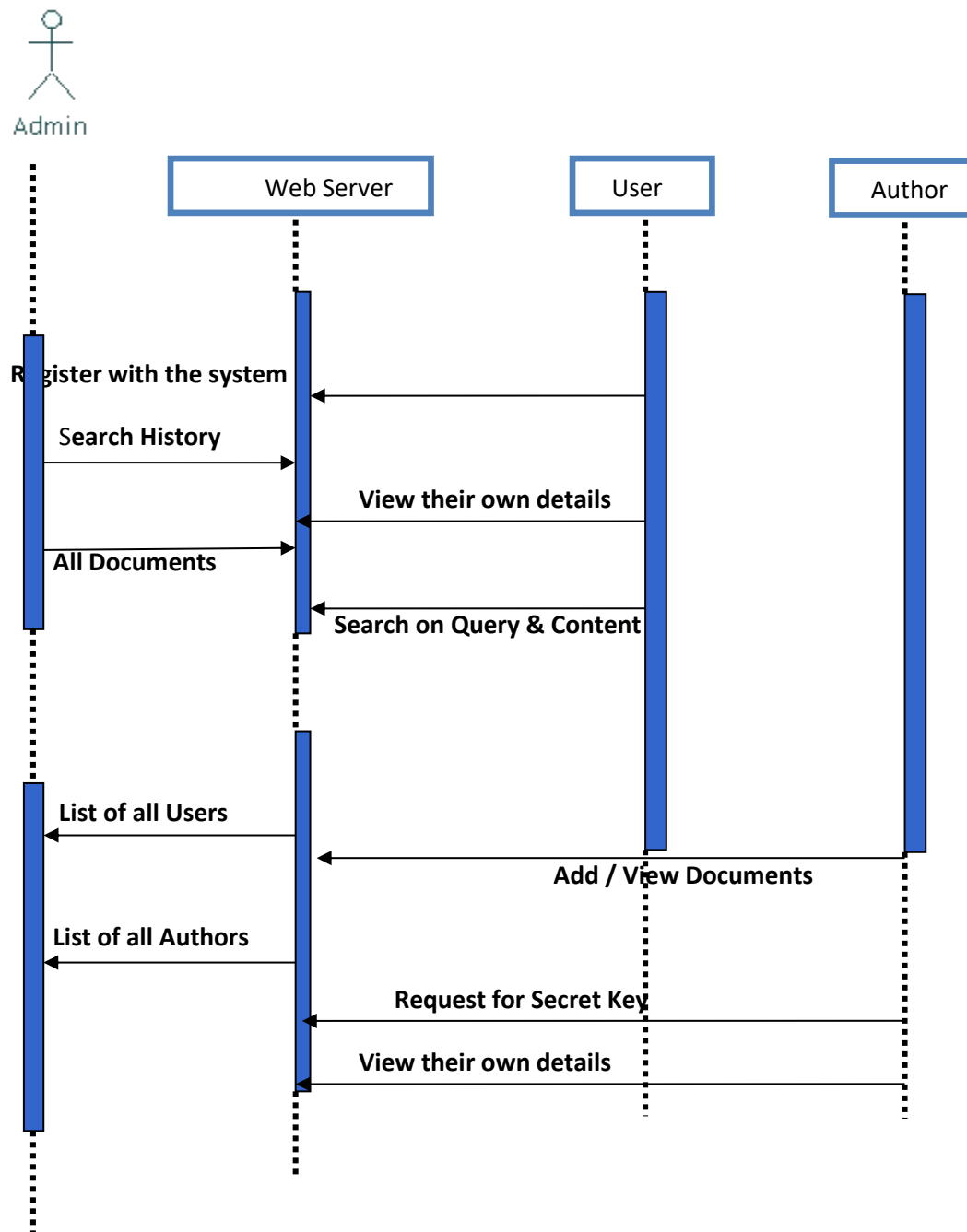
Fig 3.4 Sequence Diagram

CHAPTER 4

SYSTEM DESIGN AND DEVELOPMENT

4.1 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.

Validations are required for each data entered. Whenever a user enters an erroneous data, error message is displayed, and the user can move on to the subsequent pages after completing all the entries in the current page.

4.2 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 systems can act as the clients. The developed system is highly user friendly and can be easily understood by anyone using it even for the first time.

4.3 Data Flow
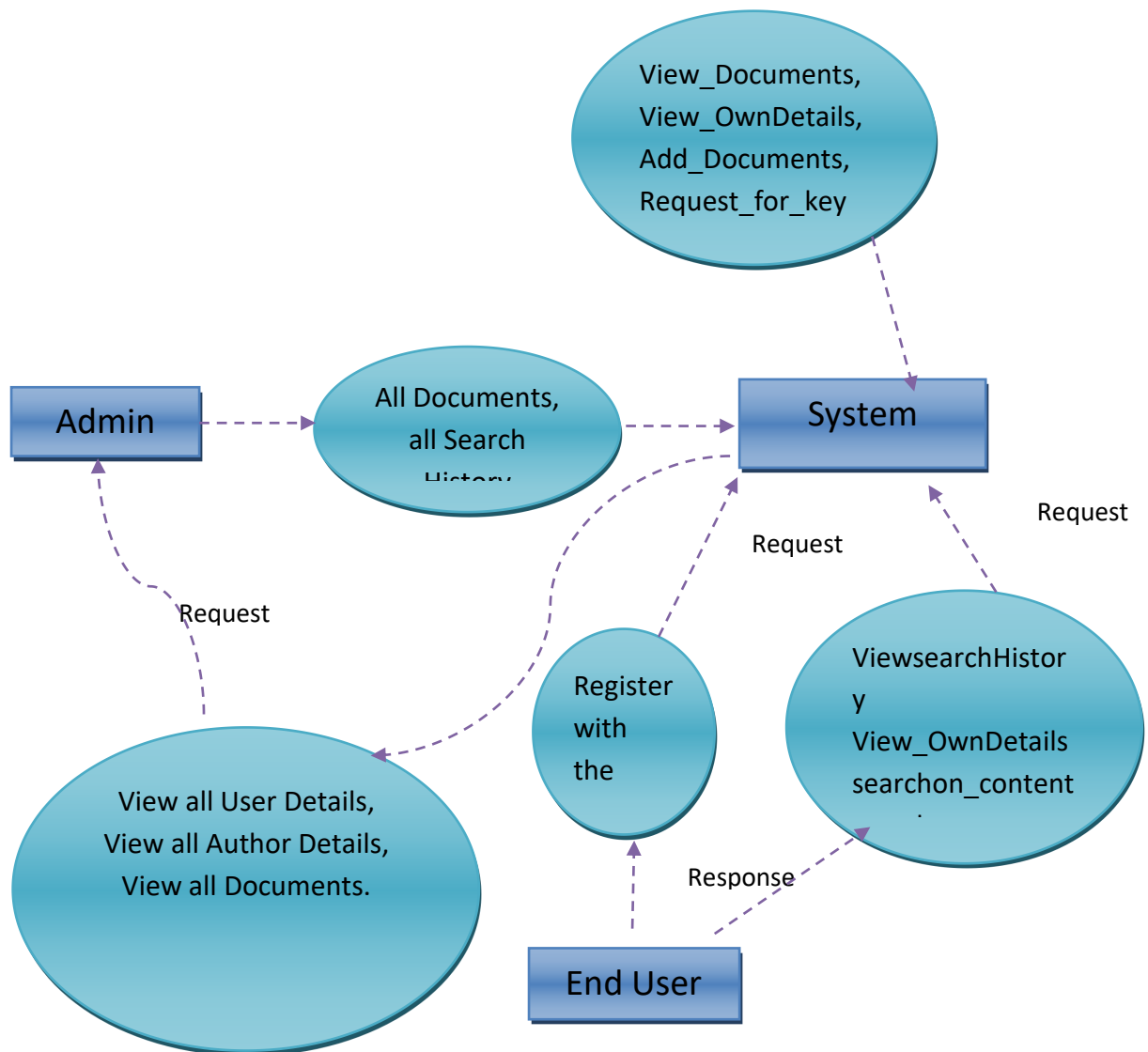
Fig 4.1 Data Flow

CHAPTER 5

SYSTEM REQUIREMENT

This Chapter describes about the requirements. It specifies the hardware and software requirements that are required in order to run the application properly. The Software Requirement Specification (SRS) is explained in detail, which includes overview of this dissertation as well as the functional and non-functional requirement of this dissertation.

SRS

Functional	Admin login by using valid user name & password, admin can view list of all user & author, view searched history details, provide secret key to author and user, author should login by using user name & password, author can upload document & also view document, requesting secret key, user login by using authorized user name & password, user can view user details, search documents based on content & query, request secret key.
Non- Functional	Admin never monitors the user activities
External interface	LAN , WAN
Performance	Admin login, user login, author login, generate secret key, view all user & author, upload document, view search history, view details, request secret key, search on query & content.
Attributes	Data Sharing platform, Document Type, Date, Location, content based document, Admin, author, users, ANTIHUBS AND OUTLIERS

Table: 5.1 Summaries of SRS

5.1 Functional Requirements

Functional Requirement defines a function of a software system and how the system must behave when presented with specific inputs or conditions.

These may include calculations, data manipulation and processing and other specific functionality.

 In this system following are the functional requirements: -

- The Admin has to login by using valid user name and password.
- After login successful he can do some operations such as list of all documents, list of author, list of all users, list of searched history and logout.
- The Admin can view the all user details. And also admin will generate secret key for users and authors.
- The admin can view list all documents.
- If the admin click on the list all documents button, then the admin will get list of all document details.
- The Admin can view list of all author. Here all registered authors are stored with the all details.
- The admin can view all searched history. If the admin clicks on search history button, then the admin will get all searched history details.
- The Author has to login by using valid user name and password. After login successful he can do some operations such as view my details, upload documents, request for secret key, and view my documents and logout.
- The author can upload number of documents. Before uploading any documents, the author should request secret key to admin, then the admin will provide a secret key.
- After getting a secret key, the author can upload documents.
- The author can view all uploaded documents. If the author clicks on the view document button, then author will get all document details.
- There are n numbers of users present. User should register before doing any operations. After registration successful he has to login by using authorized user name and password.

- After logged in he will do some operations such as view my details, search on queries, search on content, request for secret key, view my search history and logout.

- The user can search a documents based on query. Before searching any document, the user should request secret key.

- The admin will provide secret key, after getting a secret key user can search a documents in three ways such as query using storm name &warnings, query using storm name & category, query using document type, location and date.

- And also the user can search document based on content.

- The attributes are Data Sharing platform, Document Type, Date, Location, content based document, Admin, author, users, Antihubs And Outliers.

5.2 Non – Functional Requirements

Non – Functional requirements, as the name suggests, are those requirements that are not directly concerned with the specific functions delivered by the system. They may relate to emergent system properties such as reliability response time and store occupancy.

Alternatively, they may define constraints on the system such as the capability of the Input Output devices and the data representations used in system interfaces.

Many non-functional requirements relate to the system as whole rather than to individual system features. This means they are often critical than the individual functional requirements.

The following non-functional requirements are worthy of attention

The Key Non-Functional Requirements Are:

➢ Security: The system should allow a secured communication between server, Admin, author and users.

➢ Energy Efficiency: The Energy consumed by the Users to receive the File information from the server and admin.

➢ Reliability: The system should be reliable and must not degrade the performance of the existing system and should not lead to the hanging of the system.

5.3 Architechture

Admin

Accepting all user & author Information

View user & author details

Authorize the Admin

Process all user queries

Store and retrievals

WEB Database

Admin Web

1. List of All Documents
2. List of All Authors
3. List of All Users
4. List All Users Search

Registering the User

1. View Own Details
2. Upload Documents 3.View All Documents 4.Request for Secret Key

1. View Own Details
2. Search On Query 3.Search on Content 4.Request for Secret Key
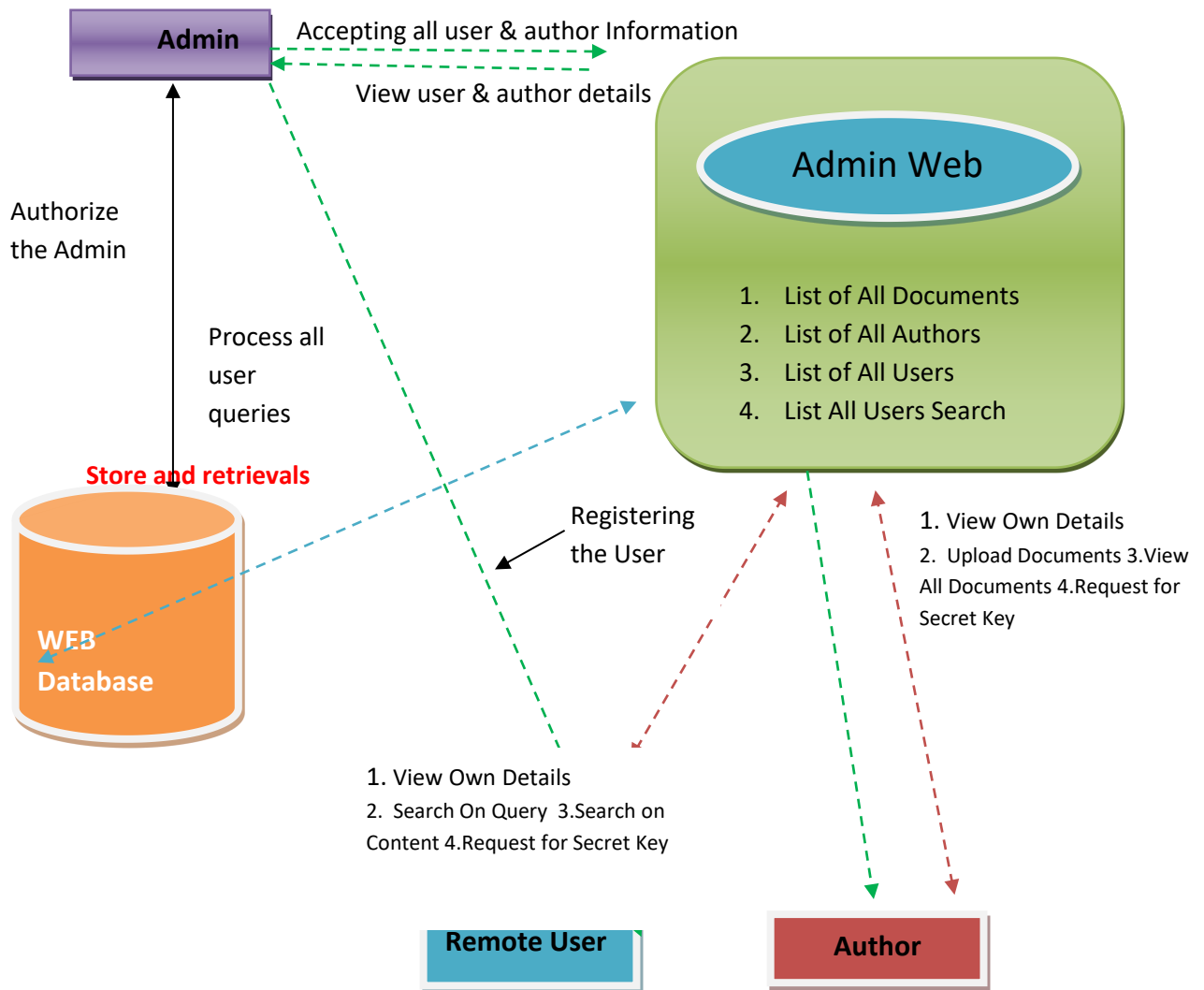
Remote User

Author

Fig 5.1 Architecture

CHAPTER 6

J2EE SOFTWARE ENVIRONMENTS

6.1 Client Server

Overview:

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.

The typical client either a PC or a Work Station connected through a network to a more powerful PC, Workstation, Midrange or Main Frames server usually capable of handling request from more than one client. However, with some configuration server may also act as client. A server may need to access other server in order to process the original client request.

The key client server idea is that client as user is essentially insulated from the physical location and formats of the data needs for their application. With the proper middleware, a client input from or report can transparently access and manipulate both local database on the client machine and remote databases on one or more servers. An added bonus is the client server opens the door to multi-vendor database access indulging heterogeneous table joins.

6.2 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.

6.3 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.

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

6.4about 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.

6.5 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.

6.6 Features Of Java

6.6.1 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.

6.6.2 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.

6.6.3 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.

6.7 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.
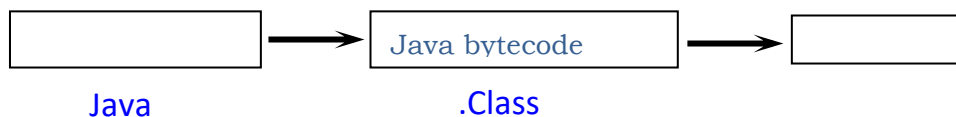
Overall **Description**

Java   Java bytecode   .Class

Fig 6.1 JAVA Development Process

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.
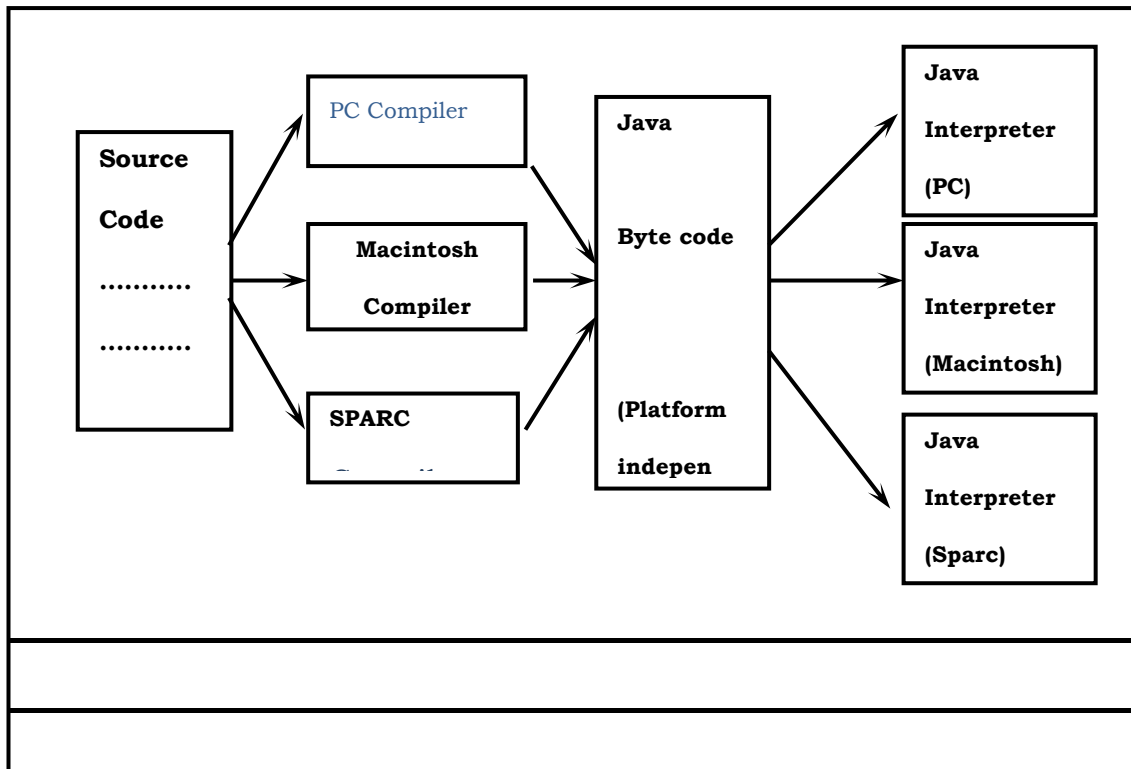
Fig 6. 2 Compiling and interpreting Java Source Code

During run-time the Java interpreter tricks the byte code file into thinking that it is running on a Java 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. In Java there are a small number of clearly defined ways to accomplish a given task.

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.

6.7.1 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.

6.7.2 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.

6.8 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.
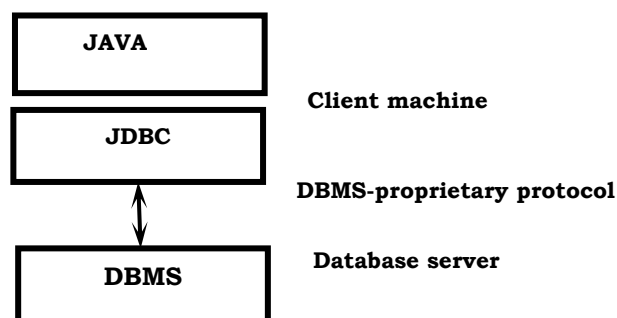
2. A literal translation of the ODBC C API into a Java API would not be desirable. For example, Java has no pointers, and ODBC makes copious use of them, including the notoriously error-prone generic pointer "void *". You can think of JDBC as ODBC translated into an object-oriented interface that is 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,

Fig 6.3 Two-Tier And Three-Tier Models

| JAVA |
| Client machine |

| JDBC |
| DBMS-proprietary protocol |

| DBMS |
| Database server |

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
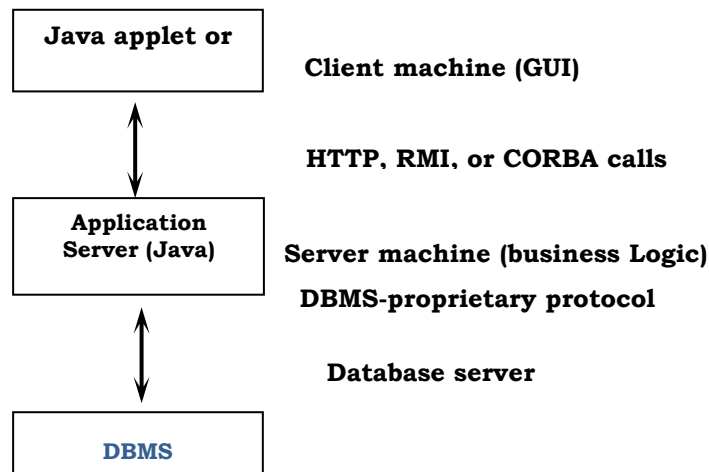

Fig 6.4 client machine

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.

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).

6.9 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 Intersolv and JavaSoft.

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.

6.10 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 JavaWebServer. At the server side JavaWebServer 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 theJavaWebServer and then it is transferred back to the result is givenback to the JavaWebServer 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 transactions
- Move SQL statements to a database engine for preprocessing and execution
- Execute stored procedures
- Inspect and modify the results from Select statements.

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 Weblogic, is one of the popular application server).
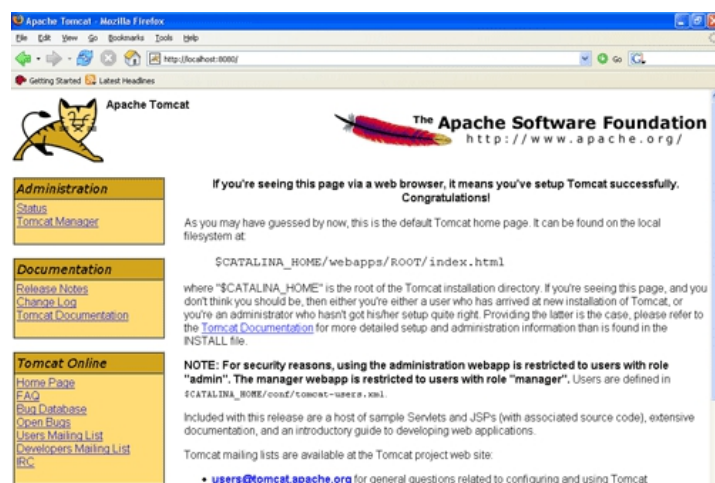

Fig 6.5 Tom Cat Server

CHAPTER 7

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.

7.1types Of Tests

7.1.1 Unit Testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

7.1.2 Integration Testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

7.1.3 FUNCTIONAL TEST

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input   : identified classes of valid input must be accepted.

Invalid Input   : identified classes of invalid input must be rejected.

Functions   : identified functions must be exercised.

Output   : identified classes of application outputs must be exercised.

Systems/Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

7.1.4 System Test

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

7.2 White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

7.3 Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot "see" into it. The test provides inputs and responds to outputs without considering how the software works.

7.4 Unit Testing:

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

Test Strategy And Approach

Field testing will be performed manually and functional tests will be written in detail.

Test Objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

Features To Be Tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

7.5 Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

7.6 System Testing

Testing Methodologies

The following are the Testing Methodologies:

- o Unit Testing.
- o Integration Testing.
- o User Acceptance Testing.
- o Output Testing.
- o 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 Following Are The Types Of Integration Testing:
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.

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.

7.7other 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.

Unit Testing:

In unit testing different are modules are tested against the specifications produced during the design for the modules. Unit testing is essential for verification of the code produced during the coding phase, and hence the goals to test the internal logic of the modules. Using the detailed design description as a guide, important Conrail paths are tested to uncover errors within the boundary of the modules. This testing is carried out during the programming stage itself. In this type of testing step, each module was found to be working satisfactorily as regards to the expected output from the module.

In Due Course, latest technology advancements will be taken into consideration. As part of technical build-up many components of the networking system will be generic in nature so that future projects can either use or interact with this.The future holds a lot to offer to the development and refinement of this project.

CHAPTER 8

SYSTEM TESTING

8.1 Testing Methodologies

The following are the Testing Methodologies:

- o Unit Testing.
- o Integration Testing.
- o User Acceptance Testing.
- o Output Testing.
- o 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.

8.2 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.

8.2.1 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.

8.2.2 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.

8.3preparation 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.

8.3.1 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.

8.4maintainence

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.

8.5testing 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.

8.5.1 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.

8.5.2 Unit Testing:

In unit testing different are modules are tested against the specifications produced during the design for the modules. Unit testing is essential for verification of

the code produced during the coding phase, and hence the goals to test the internal logic of the modules. Using the detailed design description as a guide, important Conrail paths are tested to uncover errors within the boundary of the modules. This testing is carried out during the programming stage itself. In this type of testing step, each module was found to be working satisfactorily as regards to the expected output from the module.

CONCLUSION

In this paper, we have studied the Query-Driven Entity Resolution problem in which data is cleaned \on-the-y" in the context of a query. We have developed a query-driven entity resolution framework which efficiently issues the minimal number of cleaning steps solely needed to accurately answer the given selection query. We formalized the problem of query-driven ER and showed empirically how certain cleaning steps can be avoided based on the nature of the query. This research opens several interesting directions for future investigation. While selection queries (as studied in this paper) are an important class of queries on their own, developing QDA techniques for other types of queries (e.g.,joins) is an interesting direction for future work. Another direction is developing solutions for efficient maintenance of a database state for subsequent querying.

Future Enhancement:

In future, we will future develop our algorithm in the following aspects:

This research opens several interesting directions for future investigation. While selection queries (as studied in this paper) are an important class of queries on their own, developing QDA techniques for other types of queries (e.g., joins) is an interesting direction for future work. Another direction is developing solutions for efficient maintenance of a database state for subsequent querying.

REFERENCES

Related Articles

- Face recognition with radial basis function (RBF) neural networks
- Meng Joo Er; Shiqian Wu; Juwei Lu; Hock Lye Toh
- Competitive learning with floating-gate circuits
- D. Hsu; M. Figueroa; C. Diorio
- A state-of-the-art survey on software merging
- T. Mens
- A modified Chi2 algorithm for discretization
- F.E.H. Tay; Lixiang Shen
- Constructive feedforward ART clustering networks. I
- Baraldi; E. Alpaydin
- Cluster number selection for a small set of samples using the Bayesian Ying-Yang model
- Ping Guo; C.L.P. Chen; M.R. Lyu
- Constructive feedforward ART clustering networks. II
- Baraldi; E. Alpaydin
- The multisynapse neural network and its application to fuzzy clustering
- Chih-Hsiu Wei; Chin-Shyurng Fahn
- Learning sensory maps with real-world stimuli in real time using a biophysically realistic learning rule
- M.A. Sanchez-Montanes; P. Konig; P.F.M.J. Verschure
- A study of concurrency control in real-time, active database systems
- Datta; S.H. Son