CRIME ANALYSIS

REPORT

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR SIX MONTHS INDUSTRIAL TRAINING

at

NATIONAL INSTITUTE OF ELECTRONICS AND INFORMATION TECHNOLOGY
(from 8th Jan ,2018 to 8th July ,2018)

SUBMITTED BY

SIMRANPREET GULATI

B.TECH CSE

149/14

1408442

Department of Computer Science & Engineering
DAV Institute of Engineering & Technology
Jalandhar, India

CRIME ANALYSIS

REPORT

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR
SIX MONTHS INDUSTRIAL TRAINING

at

NATIONAL INSTITUTE OF ELECTRONICS AND INFORMATION TECHNOLOGY
(from 8th Jan ,2018 to 8th July ,2018)

SUBMITTED BY

SIMRANPREET GULATI

B.TECH CSE

149/14

1408442

Department of Computer Science & Engineering
DAV Institute of Engineering & Technology
Jalandhar, India

ABSTRACT

Big Data is a group of problems and technologies related to the availability of extremely large volumes of data that businesses want to connect and understand. The reason why the sector is hot now is that the data and tools have reached a critical mass. This occurred in parallel with years of education effort that has convinced organizations that they must do something with their data treasure.

Big Data is a technique which helps us to manipulate and analyze large amount of data (data in Petabytes / Zettabytes) and based upon that analysis we can undergo predictive analysis/on-time action et cetera. Various tools are required for Big Data analysis such as Hadoop, hive, Java, etc. These tools help in the manipulation of data and get out desired results. This data treasure has helped various companies to get out of the quagmire of financial losses, poor strategies and customer dissatisfaction. Hadoop is the main tool used which offers various features such as Map Reduce etc. Presently various companies are using this technique such as Google, Facebook, Private Hospitals etc.

Further, we will discuss about four dimensions of big data i.e. four V's; various sectors in which this technology is being used; scope of big data. Apart from that, we will also discuss various projects which can be undertaken under this technology depending upon the data availability. Then we will discuss about the project we made during the 6 months of internship i.e Crime Analysis where initially we will discuss about the brief introduction of the project , its objective , usage etc. Then we will discuss about the month wise work description and status of project during the month of march (mid) , april and may. It incorporates the extreme usage of softwares and tool like Ubuntu, VMware , apache hadoop, map- reduce , hive (hiveserver2, beeline , integrated with jdbc), apache pig ,apache hbase , and R language for statistical analysis. The work Screenshots also is been added to show the work status during each month. Work Includes collecting of Raw data, Inserting into Database system , working upon the database and applying queries to it. The final work is shown by making a Java front end which shows the major work done in the system. After that with the help of Hive a predictive analysis is also carried out which will predict the future result from the past result record of thirteen years record from 2001- 2013 which could be extended by uploading the data.

ACKNOWLEDGEMENT

I am highly grateful to the Dr. Manoj Kumar , Principal, DAV Institute of Engineering & Technology, Jalandhar, for providing this opportunity to carry out the six month industrial training at NATIONAL INSTITUTE OF ELECTRONICS & INFORMATION TECHNOLOGY.

The constant guidance and encouragement received from Ms. Harpreet Kaur Bajaj, HoD Department of Computer Science & Engineering, DAVIET Jalandhar has been of great help in carrying out the project work and is acknowledged with reverential thanks.

I would like to express my gratitude to Mr. Gursewak Garcha, Assistant Professor Department of Computer Science & Engineering DAVIET Jalandhar for his stimulating guidance, continuous encouragement and supervision throughout the course of present work.

I would like to express a deep sense of gratitude and thanks profusely to Ms. Sunita Goyal Director Incharge of the Institute . Without the wise counsel and able guidance, it would have been impossible to complete the report in this manner.

The help rendered by Mr. Manish Arora, HoD Training Division(NATIONAL INSTITUTE OF ELECTRONICS & INFORMATION TECHNOLOGY) for experimentation is greatly acknowledged.

I express gratitude to other faculty members of Computer Science & Engineering department of DAVIET for their intellectual support throughout the course of this work.

SIMRANPREET GULATI

LIST OF FIGURES

Fig No.	Figure Description	Page No.
1.1	NIELIT	5
1.2	Requirement Gathering	12
1.3	Analysis	13
1.4	Designing Stage	14
1.5	Coding Stage	15
1.6	Installation Stage	17
1.7	STLC	17
1.8	Architecture Of JDBC	27
1.9	HDFC Architecture	28
1.10	Mapreduce Architecture	29
1.11	HIVE Architecture	30
1.12	PIG Architecture	31
1.13	Steps Involved During Integration	34
1.14	Main Form	38
1.15	Upload Window	39

1.16	Analysis Window	40
1.17	Yearwise_CrimeType_Of_Specific_State	41
1.18	Yearwise_CrimeType_Of_Specific_State Result	42
1.19	Total_CrimeType_Of_State_In_Specific_Year	43
1.20	Total_CrimeType_Of_State_In_Specific_Year Result	44
1.21	CrimeType_State	45
1.22	CrimeType_State Result	46
1.23	CrimeType Increasing Or Decreasing	47
1.24	CrimeType Increasing Or Decreasing Result	48
1.25	CrimeType Increasing Or Decreasing Graph	49
1.26	Overall Increase Or Decrease	50
1.27	Overall Increase Or Decrease Result	51
1.28	Overall Increase Or Decrease Graph	52
1.29	Maximum and Minimum In Each State	53
1.30	Maximum and Minimum In Each State Result	54
1.31	Crime Prediction	55
1.32	Crime Prediction Result	56
1.33	Ratio of Crime	57
1.34	Ratio of Crime Result	58

1.35 Tables(1) 59

1.36 Tables(2) 59

1.37 Different States 60

1.38 Different States and Crimetype 60

1.39 Base Table 61

1.40 Max and Min Table 61

1.41 States and District Table 62

1.42 Sum within the range 62

1.43 Random 63

1.44 Summation 63

1.45 Ratio 64

1.46 Connection 64

TABLE OF CONTENTS

Contents	**Page no.**
Certificate	**1**
Abstract	**2**
Acknowledgement	**3**
List of Figures	**4**
Table of Contents	**7**
Chapter 1 Introduction	**9**
1.1 Introduction to Organization	9
1.2 Introduction to Project	10
1.3 Objectives	10
1.4 Problem Formulation	10
1.5 Identification of Need	11
1.6 Existing System	11
1.7 Proposed System	12
1.8 Unique Features of the System	12
Chapter 2. Requirement Analysis and System Specification	**13**
2.1 Feasibility study	13
2.2 Software Requirement Specification Document	15
2.3 SDLC model	16
2.4Introduction to STLC	21
Chapter 3. System Design	**22**
3.1 Design Approach	22
3.2 Detail Design	22
3.3 System Design	24
3.4 Methodology of system	26
Chapter 4. Implementation, Testing and Maintenance	**27**
4.1 Introduction to Languages, IDE's, Tools and Technologies used for Implementation	27
Chapter 5. Results and Discussions	**39**
5.1 Snapshots of system with brief detail of each	39
5.2 Back Ends Representation(Database to be used)	40

Chapter 6. Conclusion and Future Scope 66

References/Bibliography 67

Chapter 1 INTRODUCTION

1.1 Introduction To Organization

National Institute of Electronics & Information Technology (NIELIT), erstwhile known as the DOEACC (Department of Electronics and Accreditation of Computer Courses Society), is an autonomous scientific society under the administrative control of Ministry of Electronics and Information Technology of the government of India. The Society is registered under the Societies Registration Act, 1860. The NIELIT (National Institute of Electronic and Information Technology) Scheme, jointly developed by AICTE and DEITY, was launched in 1990 after detailed deliberations in a National Working Group. The objective of the scheme is to generate high quality man-power in the computer software and allied fields by utilizing the expertise and facilities available with various institutions in the non-formal sector. NIELIT is one of the National Examination Body in India, which accredits Computer courses in ‘O’, 'A' , 'B' & 'C' level approving institutes / organizations for conducting particular course, specializing in the non-formal sector of IT education and training. At present, NIELIT has thirty five (35) offices located at Agartala, Aizawl, Ajmer, Aurangabad, Calicut, Chandigarh and many more. It is also well networked throughout India with the presence of about 800 institutes.

रा.इ.सू.प्रौ.सं
NIELIT
राष्ट्रीय इलेक्ट्रॉनिकी एवं सूचना प्रौद्योगिकी संस्थान,
National Institute Of Electronics & Information Technology,
A Autonomous Scientific Society of Department of Electronics & IT (DeitY)
Government of India
सत्यमेव जयते
Shift

Fig 1.1 Nielit

The basket of activities of NIELIT is further augmented by the wide range of projects that it undertakes. NIELIT has demonstrated its capability and capacity to undertake R&D projects, consultancy services, turnkey projects in office automation, software development, website development etc. NIELIT is also the nodal implementing agency on behalf of DEITY for Data Digitization of the population of 15 assigned States and 2 Union Territories for the creation of National Population Register (NPR) project of Registrar General and Census Commissioner of India.

NIELIT is also successfully executing the Agriculture Census and Input Survey project under which tabulation of about 10 crore data records have to be done. NIELIT has planned a road map for adopting appropriate pedagogy for metamorphosing NIELIT into an Institute of National Importance. The Centre has a team of highly qualified, well trained, dedicated and experienced professionals. All of them undergo extensive training for developing and enhancing professional skill throughout their career.

1.2 Introduction To Project

The Crime Analysis System is developed to manage and review the past trends in crimes throughout India to ensure that steps can be taken on issues/crimes that are affecting the society the most. The software will be great relief to the crime department. The Crime Analysis System is developing to ensure the appropriate management, which works in batch mode. This application will allow the analysers to view the data more effectively and in compact form ,as the analysers want to see, since different categories and views are shown in this project.In this application with the help of previous data we can predict the future trends in a particular crime i.e. how much crime would be there in a particular year.

1.3 Objective

Crime analysis and prevention is a systematic approach for identifying and analyzing patterns and trends in crime. Our system can predict regions which have high probability for crime occurrence and can visualize crime prone areas. With the increasing advent of computerized systems, crime data analysts can help the Law enforcement officers to speed up the process of solving crimes. Using the concept of data mining we can extract previously unknown, useful information from an unstructured data. Here we have an approach between computer science and criminal justice to develop a data mining procedure that can help solve crimes faster. Instead of focusing on causes of crime occurrence like criminal background of offender, political enmity etc we are focusing mainly on crime factors of each day.

1.4 Problem Formulation

This application is used to deal with the problems of the society i.e. different crimes throughout India. It helps in the management so that steps can be taken to reduce crimes and make India free!!

The first step to resolve a problem is collecting data and analysing the data so that we can get to know that where do we need to focus.

1.5 Identification/Recognition of Need

The basic need of Crime Analysis is to detect the problem that occurs at various levels. Crime analysts study crime reports, different trends to identify emerging patterns, series, and trends as quickly as possible. They analyse these phenomena for all relevant factors, sometimes predict or forecast future occurrences, and issue bulletins, reports, and alerts to their agencies. They then work with their police agencies to develop effective strategies and tactics to address crime and disorder. Other duties of crime analysts may include preparing statistics, data queries, or maps on demand; analysing beat and shift configurations; preparing information for community or court presentations; answering questions from the public and the press; and providing data and information support for a police department's Comp Stat process which arises the need of Crime Analysis.

To see if a crime fits a certain known pattern or a new pattern is often tedious work of crime analysts, detectives or in small departments, police officers or deputies themselves. They must manually sift through piles of paperwork and evidence to predict, anticipate and hopefully prevent crime. The U.S. Department of Justice and the National Institute of Justice recently launched initiatives to support “predictive policing”, which is an empirical, data-driven approach. However this work to detect specific patterns of crime committed by an individual or group (crime series), remains a manual task.

1.6 Existing System

Day by day the crime rate is increasing considerably. Crime cannot be predicted since it is neither systematic nor random. Also the modern technologies and hi-tech methods help criminals in achieving their misdeeds, According to Crime Records Bureau crimes like burglary, arson etc have been decreased while crimes like murder, sex abuse, gang rape etc have been increased.In the existing system only we can see the details of a particular information about the crimes in a particular state, the existing system has more workload for the authorized person. The single person has to manage the whole data which is quite challenging. Also, it is difficult to analyse such a huge amount of data manually.

1.7 Proposed System

Crime analysis and prevention is a systematic approach for identifying and analyzing patterns and trends in crime. Our system can predict regions which have high probability for crime occurrence and can visualize crime prone areas. With the increasing advent of computerized systems, crime data analysts can help the Law enforcement officers to speed up the process of solving crimes. Using the concept of data mining we can extract previously unknown, useful information from an unstructured data.

Also the modern technologies and hi-tech methods help criminals in achieving their misdeeds. According to Crime Records Bureau crimes like burglary, arson etc have been decreased while crimes like murder, sex abuse, gang rape etc have been increased. Even though we cannot predict who all may be the victims of crime but can predict the place that has probability for its occurrence.

1.8 Unique Features Of The System

-To help the Law enforcement officers to speed up the process of solving crimes.

-To analyse the year wise, state wise and district wise crime patterns in India.

-To predict the regions which have high probability for crimes.

-To evaluate the ratio of crimes with respect to years.

-Graphical interpretation of data has helped understanding the scenario more precisely.

Chapter 2 REQUIREMENT ANALYSIS AND SYSTEM SPECIFICATION

2.1 Feasibility Study

Feasibility study is an evaluation of a proposal designed to determine the difficulty in carrying out a designated task. Generally, a feasibility study precedes technical development and project implementation. In other words, a feasibility study is an evaluation or analysis of the potential impact of a proposed project.

Feasibility study is used to measure how the development of a system should be beneficial to organizations. Feasibility study should be performed throughout the development of the system. Feasibility study involves making a preliminary determination of end user needs and to determine the feasibility, that system is satisfied goal of project or not. The goal of feasibility study is to evaluate alternative system and to provide most feasible and desirable system for development.

A feasibility study should provide enough information to decide:

-Whether the project can be done?

-Whether the final system will benefit its intended users?

-What are the alternatives among which a solution will be chosen?

-Is there a preferred alternative?

The key considerations in feasibility analysis are :

-Technical feasibility

-Operational feasibility

-Economic feasibility

Technical Feasibility

This is concerned with specifying equipment and software and hardware that will successfully satisfy the user requirement. The technical needs of the system may vary considerably, but might include

-The facility to produce output in a given time.

-Response time under certain conditions

-Ability to process a certain volume of transaction at a particular speed.

-Facility to communicate data to distant location.

In examining technical feasibility, configuration of the system is given more importance than the actual make of hardware. The configuration should give the complete picture about the system requirements. What speeds of input and output should be achieved at particular quility of printing.

Operational Feasibility

It is a measure of how well a proposed system solves the problems, and takes advantage of the opportunities identified during scope definition and how it satisfies the requirements identified in the requirements analysis phase of system development.

During operational feasibility, it is determined whether the system will operate in the way that user wants or not. Operational feasibility must determines how the proposed system will fit in with the current operations and what, if any, job reconstruction and training will be needed to implement the system. It is a measure of how well the solutions of the problem or specific alternative solutions will work in the organizations. It deals with the operations performed within the network.The essential questions that help in testing the operational feasibility of a system are following:

Does management support the project?

Are the users not happy with current business practices? Will it reduce the time (operation) considerably? If yes, then they will welcome the change and the new system.

Have the users been involved in the planning and development of the project? Early involvement reduces the probability of resistance towards the new system.

Will the proposed system really benefit the organization? Does the overall response increase? Will accessibility of information be lost? Will the system.

effect the customers in considerable way

What changes will be brought with the system?

What organization structures are disturbed?

Economic Feasibility

Economical feasibility determines the cost and benefits of the proposed system and compare with the budget. The system that is made must also be economical. Economic analysis commonly known as cost/benefit analysis. The cost of the project includes the cost of hardware, software development and implementation. If benefits are found more than costs then the decision is made to design and implement the network. The cost should be consistent according to the operations being performed within the network.

Economic analysis is the most frequently used technique for evaluating the effectiveness of a proposed system. More commonly known as cost/benefit analysis; the procedure is to determine the benefits and saving that are expected from a proposed system and compare them with cost. If benefits outweigh cost, a decision is taken to design and implement the system. Otherwise, further justification or alternative in the proposed system will have to be made if it is to have a chance of being approved. This is an ongoing effort that improves in accuracy at each phase of the system life cycle.

2.2 Software Requirement Specification Document

Functional Requirements

Upload Data

The data is uploaded and used to see the crime trends in every state of India.

Analysis

There are different categories for analysis of the crime data that includes state wise analysis of the crime, year wise analysis showing the increase or decrease in the crime in particular year with their graphical representation, prediction of the crime in particular year, ratio of crime with respect to particular year,etc.

Non Functional Requirements

Performance

Taking care of working of every module and making the application accessible on every platform

Usability

The application is easy to understand and use.

Data Requirements

The data used for analysis and prediction consists of twenty types of crimes occurring in all the states of India from year 2001 to year 2013.

2.3 SDLC Model

The Software Development Life Cycle is a process that ensures good software is built.Each phase in the life cycle has its own process and deliverables that feed into the next phase. There are typically 5 phases starting with the analysis and requirements gathering and ending with the implementation.

2.3.1 Requirement Gathering Stage

The requirements gathering process takes as its input the goals identified in the high-level requirements section of the project plan. Each goal will be refined into a set of one or more requirements. These requirements define the major functions of the intended application, define operational data areas and reference data areas, and define the initial data entities. Major functions include critical processes to be managed, as well as mission critical inputs, outputs and reports. A user class hierarchy is developed and associated with these major functions, data areas, and data entities. Each of these definitions is termed a Requirement. Requirements are identified by unique requirement identifiers and, at minimum, contain a requirement title and textual description.

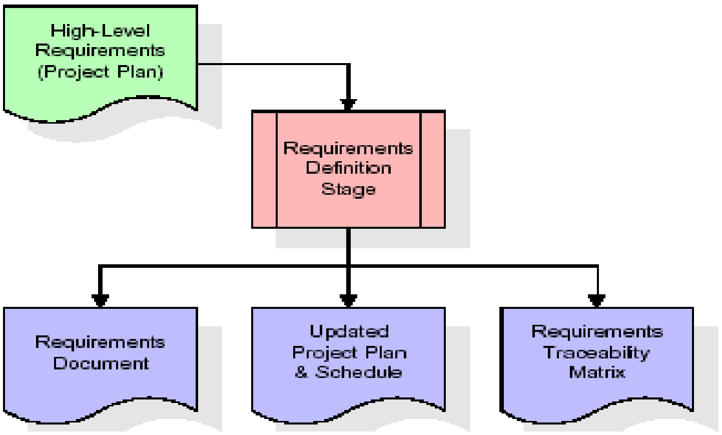

High-Level Requirements (Project Plan)
Requirements Definition Stage
Requirements Document
Updated Project Plan & Schedule
Requirements Traceability Matrix

Fig 1.2 Requirement gathering

These requirements are fully described in the primary deliverables for this stage: the Requirements Document and the Requirements Traceability Matrix (RTM). The requirements document contains complete descriptions of each requirement, including diagrams and references to external documents as necessary.The purpose of the RTM is to show that the product components developed during each stage of the software development lifecycle are formally connected to the components developed in prior stages.

2.3.2 Analysis Stage:

The planning stage establishes a bird's eye view of the intended software product, and uses this to establish the basic project structure, evaluate feasibility and risks associated with the project, and describe appropriate management and technical approaches.

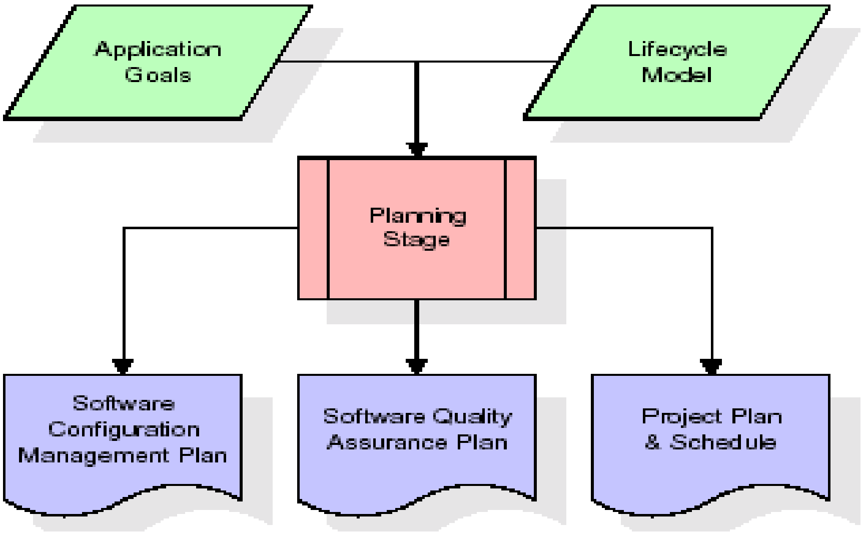

Application Goals
Lifecycle Model
Planning Stage
Software Configuration Management Plan
Software Quality Assurance Plan
Project Plan & Schedule

Fig 1.3 Analysis

The most critical section of the project plan is a listing of high-level product requirements, also referred to as goals. All of the software product requirements to be developed during the requirements definition stage flow from one or more of these goals. The minimum information for each goal consists of a title and textual description, although additional information and references to external documents may be included. The outputs of the project planning stage are the configuration management plan, the quality assurance plan, and the project plan and schedule, with a detailed listing of scheduled activities for the upcoming Requirements stage, and high level estimates of effort for the out stages.

2.3.3 Design Stage

Technical design requirements are prepared in this phase by lead development staff that can include architects and lead developers. The Business Requirements are used to define how the application will be written. Technical requirements will detail database tables to be added, new transactions to be defined, security processes and hardware and system requirements

.

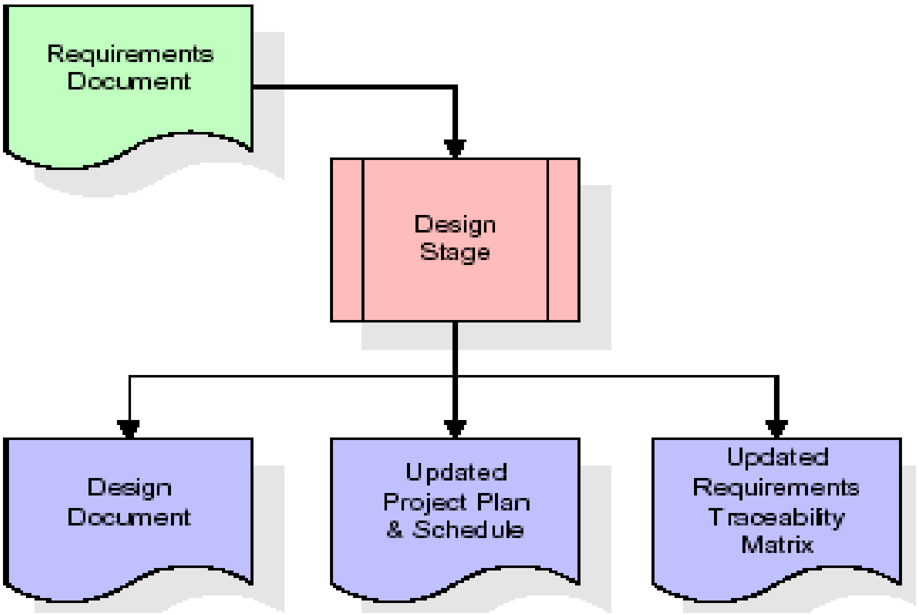

Requirements Document
Design Stage
Design Document
Updated Project Plan & Schedule
Updated Requirements Traceabilty Matrix

Fig 1.4 Designing Stage

2.3.4 Coding Stage

This phase is the actual coding and unit testing of the process by the development team. After each stage, the developer may demonstrate the work accomplished to the Business Analysts and tweaks and enhancements may be required. It’s important in this phase for developers to be open-minded and flexible if any changes are introduced. This is normally the longest phase of the SDLC. The finished product here is input to the Testing phase.

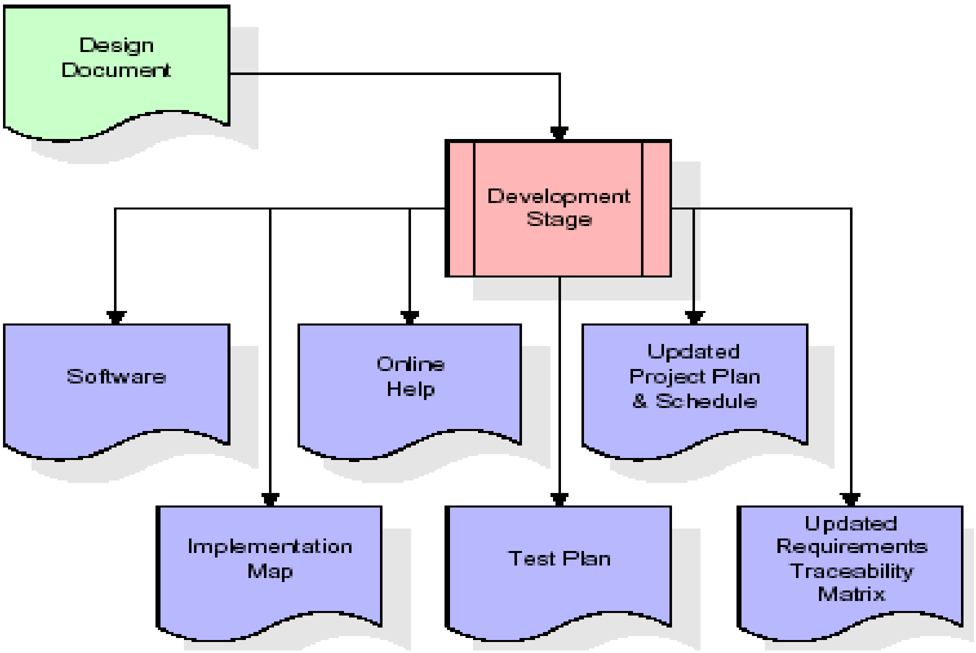


Fig 1.5 Coding Stage

2.3.5 Testing Stage

Once the application is migrated to a test environment, different types of testing will be performed including integration and system testing. User acceptance testing is the last part of testing and is performed by the end users to ensure the system meets their expectations. At this point, defects may be found and more work may be required in the analysis, design or coding. Once sign-off is obtained by all relevant parties, implementation and deployment can begin.

Different types Of Testing Performed in the organization are: -

Component Testing/Module Testing

A component is the lowest unit of any application, so Component testing; as the name suggest, is a technique of testing the lowest or the smallest unit of any application. An application can be thought of a combination and integration of many small individual modules. Before we test the entire system as a whole itis imperial that each and every component OR the smallest unit of the application is tested thoroughly.

In this case, the modules or the units are tested independently. Each module receives an input, does some processing and generates the output. The output is then validated against the expected feature.

Subsystem Testing

Sub systems are integrated to make up the entire system. The testing process is concerned with finding errors that result from unanticipated interactions between sub-systems and system components. It is also concerned with validating that the system meets its functional and non-functional requirements.

System Testing

System testing is defined as testing of a complete and fully integrated software product. This testing falls in black-box testing wherein knowledge of the inner design of the code is not a pre-requisite and is done by the testing team. System testing is performed in the context of a System Requirement Specification (SRS) and/or a Functional Requirement Specifications (FRS).

Field Testing

Experiment, research, or trial conducted under actual use conditions, instead of under controlled conditions in a laboratory. Also called as Field Experiment.

2.3.6 Installation & Acceptance Test:

During the installation and acceptance stage, the software artifacts, online help, and initial production data are loaded onto the production server. At this point, all test cases are run to verify the correctness and completeness of the software. Successful execution of the test suite is a prerequisite to acceptance of the software by the customer.

After customer personnel have verified that the initial production data load is correct and the test suite has been executed with satisfactory results, the customer formally accepts the delivery of the software.

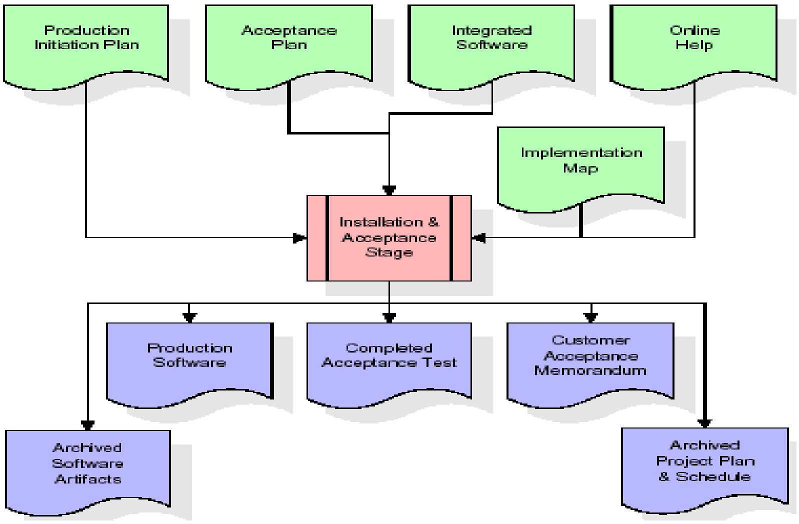

Production Initiation Plan
Acceptance Plan
Integrated Software
Online Help
Implementation Map
Installation & Acceptance Stage
Production Software
Completed Acceptance Test
Customer Acceptance Memorandum
Archived Software Artifacts
Archived Project Plan & Schedule

Fig 1.6 Installation Stage

2.3.7 Implementation/Deployment Stage

The size of the project will determine the complexity of the deployment. Training may be required for end users, operations and on-call IT staff. Roll-out of the system may be performed in stages starting with one branch then slowly adding all locations or it could be a full-blown implementation.

2.4 Introduction To Software Testing Life Cycle

Software Testing Life Cycle (STLC) is defined as a sequence of activities conducted to perform Software Testing. It consists of series of activities carried out methodologically to help certify your software product.

Diagram - Different stages in Software Test Life Cycle

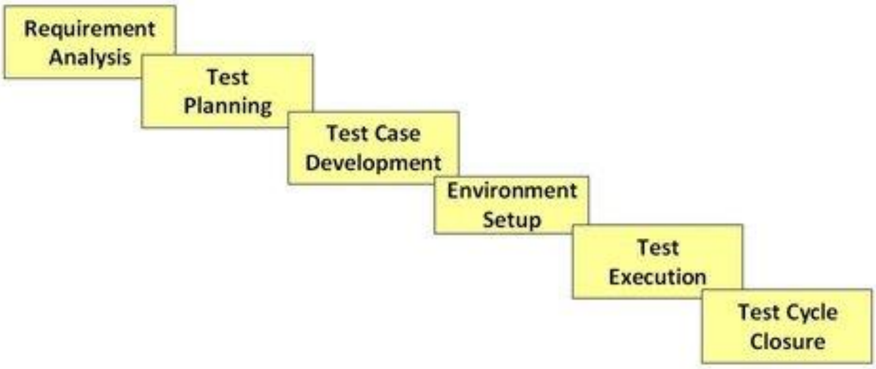


Fig 1.7 STLC

Chapter 3 SYSTEM DESIGN

3.1 Design Approach

The Object Oriented Approach has been used for the building the application.

Object-oriented analysis and design (OOAD) is a popular technical approach for analyzing and designing an application, system, or business by applying object-oriented programming, as well as using visual modelling throughout the development life cycles to foster better stakeholder communication and product quality. The purpose of any analysis activity in the software life-cycle is to create a model of the system's functional requirements that is independent of implementation constraints.

The main difference between object-oriented analysis and other forms of analysis is that by the object-oriented approach we organize requirements around objects, which integrate both behaviors (processes) and states (data) modeled after real world objects that the system interacts with.

The primary tasks in object-oriented analysis (OOA) are:

- Find the objects
- Organize the objects
- Describe how the objects interact
- Define the behaviour of the objects
- Define the internals of the objects

3.2 Detail Design

System design is the process of developing specifications for a candidate system that meet the criteria established in the system analysis. Major step in system design is the preparation of the input forms and the output reports in a form applicable to the user.

The main objective of the system design is to make the system user friendly. System design involves various stages as:

- Data Entry
- Data Correction

- Data Deletion
- Processing
- Sorting and Indexing
- Report Generation

Once analysis is completed, the analyst has a firm understanding of what is to be done. The next step is how the problem can be solved the design of a system uses the functional specification as basis and produces the details that state how a system will meet the requirements identified during systems analysis. The design process should take care of the following:

- Identification of reports and outputs the new system should produce.
- Sketch the form or display as expected to appear at the end of completion of the system. The may be done on paper or on a computer display, using one of the automated system design tools available.
- Description of data to be input calculated or stored.
- Individual data items and calculation procedures are written in detail.
- The procedures written should tell how to process the data and produce the output.

Some objectives guiding the design of the input focus on:

Effectiveness: This means that input forms and screens serve specific purpose.

Accuracy: Refers to design that assures proper completion.

Ease to use: Means that the forms and screens are straight forward and requires no time to understand.

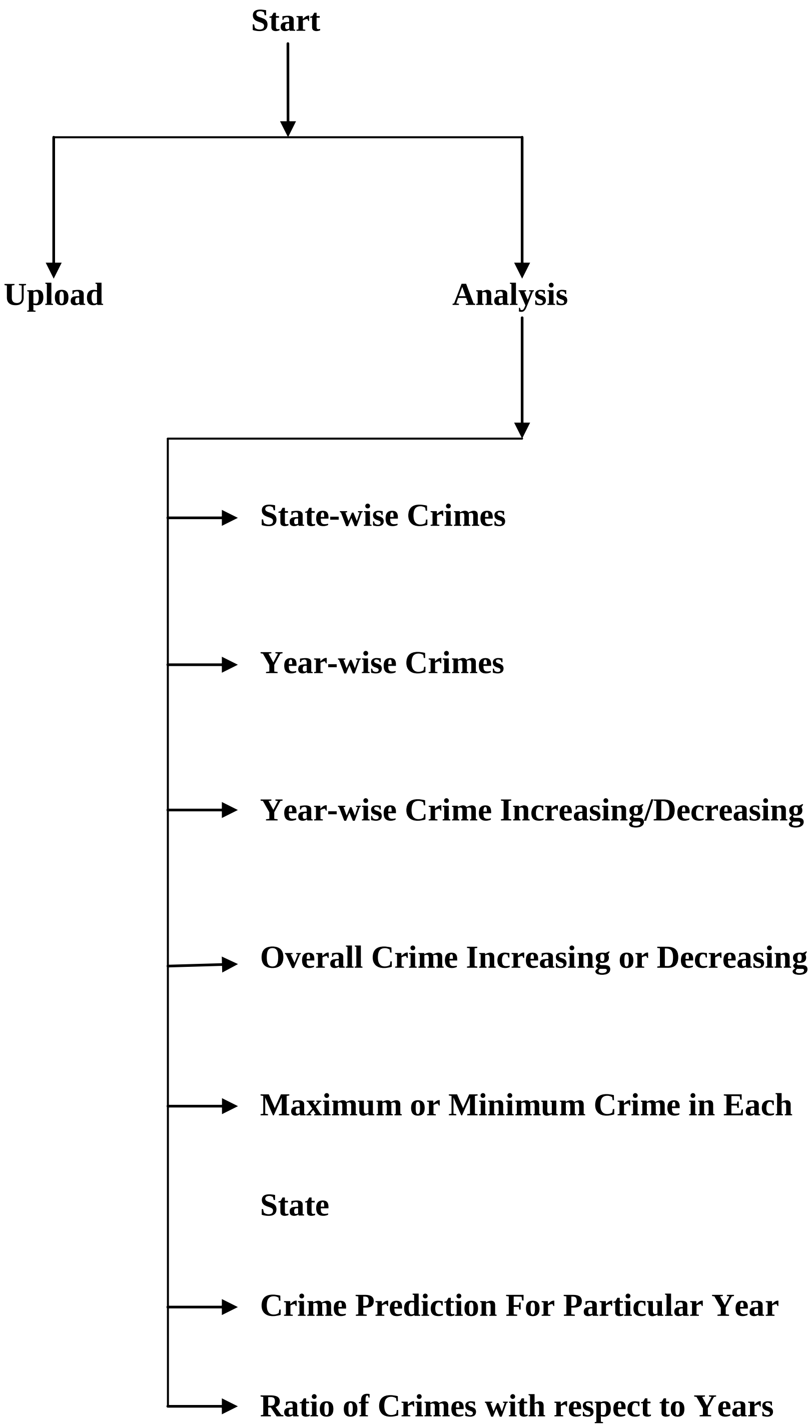
Consistency: Means that forms and screens should group data of similar nature together

Simplicity: refers to keeping the forms and screens simple and uncluttered.

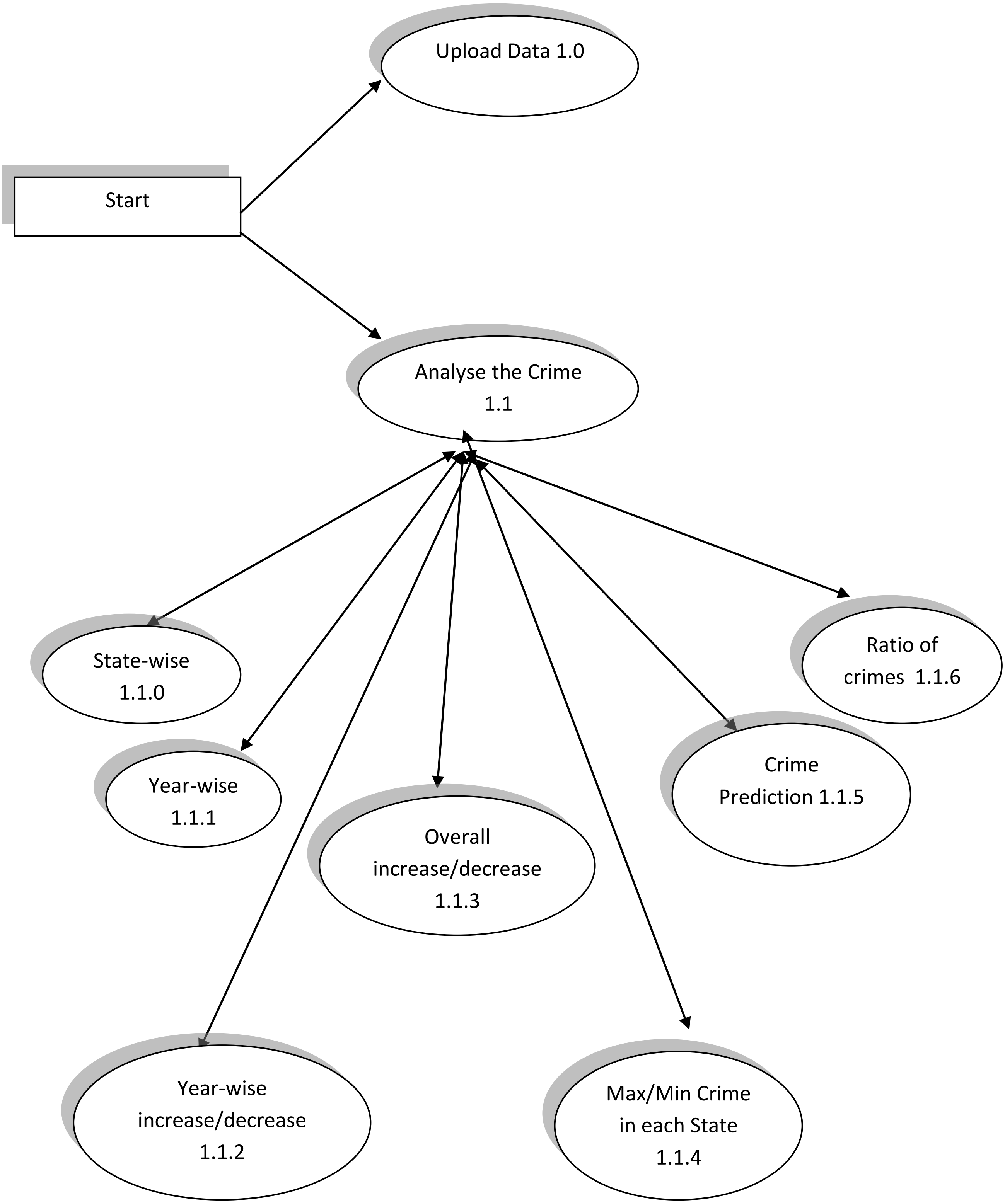
Attractiveness: Input forms should be of appealing design which should please the user

3.3 System Design Using-

3.3.1 Control Flow Diagram

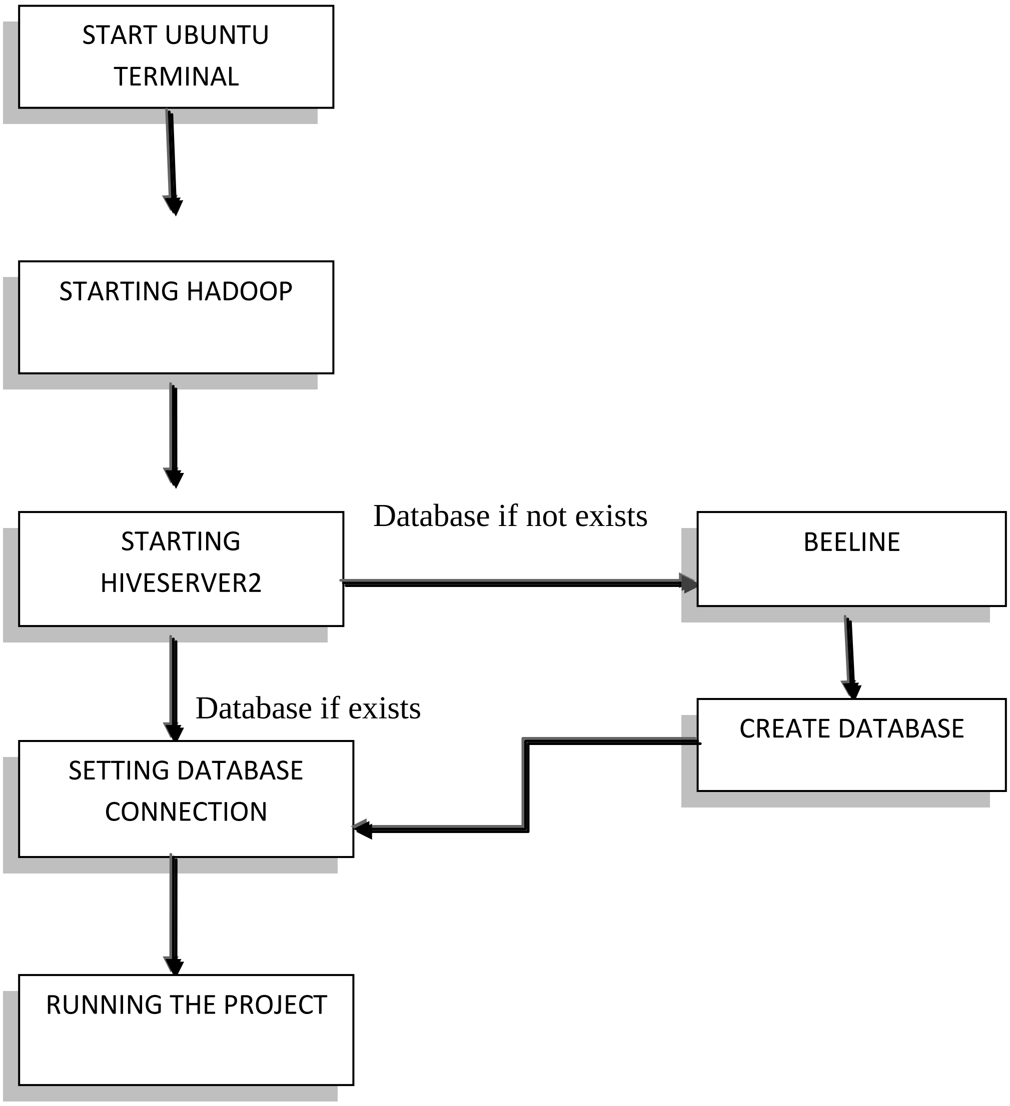

Start
Upload
Analysis
State-wise Crimes
Year-wise Crimes
Year-wise Crime Increasing/Decreasing
Overall Crime Increasing or Decreasing
Maximum or Minimum Crime in Each State
Crime Prediction For Particular Year
Ratio of Crimes with respect to Years

3.3.2 Data Flow Diagram



3.4 Methodology Of System

In order to implement the system followings steps are adopted:



Chapter 4 IMPLEMENTATION ,TESTING AND MAINTENANCE

4.1 Introduction to Languages, IDE’s, Tools and Technology Used

BigData

Every day, we create 2.5 quintillion bytes of data — so much that 90% of the data in the world today has been created in the last two years alone. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few. Such colossal amount of data that is being produced continuously is what can be coined as Big Data. Big Data decodes previously untouched data to derive new insight that gets integrated into business operations. However, as the amounts of data increases exponential, the current techniques are becoming obsolete. Dealing with Big Data requires comprehensive coding skills, domain knowledge and statistics. Despite being Herculean in nature, Big Data applications are almost ubiquitous- from marketing to scientific research to customer interests and so on. We can witness Big Data in action almost everywhere today. From Facebook which handles over 40 billion photos from its user base to CERN‘s Large Hadron Collider (LHC) which generates 15PB a year to Walmart which handles more than 1 billion customer transactions in an hour. Over a year ago, the World Bank organized the first WBG Big Data Innovation Challenge which brought forward several unique ideas applying Big Data such as big data to predict poverty and for climate smart agriculture and fore user focused Identification of Road Infrastructure Condition and safety and so on.

Big Data can be simply defined by explaining the 3V‘s – volume, velocity and variety which are the driving dimensions of Big Data quantification. Gartner analyst, Doug Laney introduced the famous 3 V‘s concept in his 2001 Met group publication 3D data management:

Volume: This essentially concerns the large quantities of data that is generated continuously. Initially storing such data was problematic because of high storage costs. However with decreasing storage costs, this problem has been kept somewhat at bay as of now. However this is only a temporary solution and better technology needs to be developed. Smartphones, E-Commerce and social networking websites are examples where massive amounts of data are being generated. This data can be easily distinguishes between structured data, unstructured data and semi-structured data.

Velocity: In what now seems like the pre-historic times, data was processed in batches. However this technique is IJARCCE ISSN (Online) 2278-1021 ISSN (Print) 2319 5940 International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 2, February 2016 only feasible when the incoming data rate is slower than the batch processing rate and the delay is much of a hindrance. At present times, the speed at which such colossal amounts of data are being generated is unbelievably high

Variety: Documents to databases to excel tables to pictures and videos and audios in hundreds of formats, data is now losing structure. Structure can no longer be imposed like before for the analysis of data. Data generated can be of any type- structures, semi-structured or unstructured. The conventional form of data is structured data. For example text. Unstructured data can be generated from social networking sites, sensors and satellites

It has become important to create a new platform to fulfill the demand of organizations due to the challenges faced by traditional data. By leveraging the talent and collaborative efforts of the people and the resources, innovation in terms of managing massive amount of data has become tedious job for organisations. This can be fulfilled by implementing big data and its tools which are capable to store, analyse and process large amount of data at a very fast pace as compared to traditional data processing systems (Picciano 2012). Big data has become a big game changer in today’s world.

To analyse This abundant amount of Data we used The tool Hadoop in our respected project. Hadoop is an open source, Java-based programming framework that supports the processing and storage of extremely large data sets in a distributed computing environment. It is part of the Apache project sponsored by the Apache Software Foundation. Hadoop makes it possible to run applications on systems with thousands of commodity hardware nodes, and to handle thousands of terabytes of data. Its distributed file system facilitates rapid data transfer rates among nodes and allows the system to continue operating in case of a node failure.

This approach lowers the risk of catastrophic system failure and unexpected data loss, even if a significant number of nodes become inoperative. Consequently, Hadoop quickly emerged as a foundation for big data processing tasks, such as scientific analytics, business and sales planning, and processing enormous volumes of Data

4.1.1 Tools Learned

-LINUX

Linux is a freely distributed implementation of a UNIX−like kernel, the low level core of an operating system. Because Linux takes the UNIX system as its inspiration, Linux and UNIX programs are very similar. In fact, almost all programs written for UNIX can be compiled and run under Linux. Also, many commercial applications sold for commercial versions of UNIX can run unchanged in binary form on Linux systems. Linux was developed by Linus Torvalds at the University of Helsinki, with the help of UNIX programmers from across the Internet. It began as a hobby inspired by Andy Tanenbaum's Minix, a small UNIX system, but has grown to become a complete UNIX system in its own right. The Linux kernel doesn't use code from AT&T or any other proprietary source.

Let's take a look at some popular distributions:

Red Hat: Red Hat is a billion dollar commercial Linux Company that puts a lot of effort in developing Linux. They have hundreds of Linux specialists and are known for their excellent support. They give their products (Red Hat Enterprise Linux and Fedora) away for free. While Red Hat Enterprise Linux (RHEL) is well tested before release and supported for up to seven years after release, Fedora is a distro with faster updates but without support.

Ubuntu: Canonical started sending out free compact discs with Ubuntu Linux in 2004 and quickly became popular for home users (many switching from Microsoft Windows). Canonical wants Ubuntu to be an easy to use graphical Linux desktop without need to ever see a command line.

Debian: There is no company behind Debian. Instead there are thousands of well organised developers that elect a Debian Project Leader every two years. Debian is seen as one of the most stable Linux distributions. It is also the basis of every release of Ubuntu. Debian comes in three versions: stable, testing and unstable

Other Distributions: like CentOS, Oracle Enterprise Linux and Scientific Linux are based on Red Hat Enterprise Linux and share many of the same principles, directories and system administration techniques. Linux Mint, Edubuntu and many other Ubuntu named distributions are based on Ubuntu and thus share a lot with Debian. There are hundreds of other Linux distributions.

-JAVA

Java is a general-purpose computer programming language that is concurrent, class-based, object-oriented, and specifically designed to have as few implementation dependencies as possible. It is intended to let application developers "write once, run anywhere" (WORA), meaning that compiled Java code can run on all platforms that support Java without the need for recompilation.

Java is:

Object Oriented: In Java, everything is an Object. Java can be easily extended since it is based on the Object model.

Platform Independent: Unlike many other programming languages including C and C++, when Java is compiled, it is not compiled into platform specific machine, rather into platform independent byte code. This byte code is distributed over the web and interpreted by the Virtual Machine (JVM) on whichever platform it is being run on.

Simple: Java is designed to be easy to learn. If you understand the basic concept of OOP Java, it would be easy to master.

Secure: With Java's secure feature it enables to develop virus-free, tamper-free systems. Authentication techniques are based on public-key encryption.

Architecture-neutral: Java compiler generates an architecture-neutral object file format, which makes the compiled code executable on many processors, with the presence of Java runtime system.

-JDBC

Java Database Connectivity (JDBC) is an application programming interface (API) for the programming language Java, which defines how a client may access a database.

JDBC stands for Java Database Connectivity, which is a standard Java API for database-independent connectivity between the Java programming language and a wide range of databases.

The JDBC library includes APIs for each of the tasks mentioned below that are commonly associated with database usage.

Making a connection to a database.

Creating SQL or MySQL statements.

Executing SQL or MySQL queries in the database.

Viewing & Modifying the resulting records.

JDBC provides the same capabilities as ODBC, allowing Java programs to contain database-independent code.

The JDBC API supports both two-tier and three-tier processing models for database access but in general, JDBC Architecture consists of two layers:

JDBC API: This provides the application-to-JDBC Manager connection.

JDBC Driver API: This supports the JDBC Manager-to-Driver Connection.

Following is the architectural diagram, which shows the location of the driver manager with respect to the JDBC drivers and the Java application:

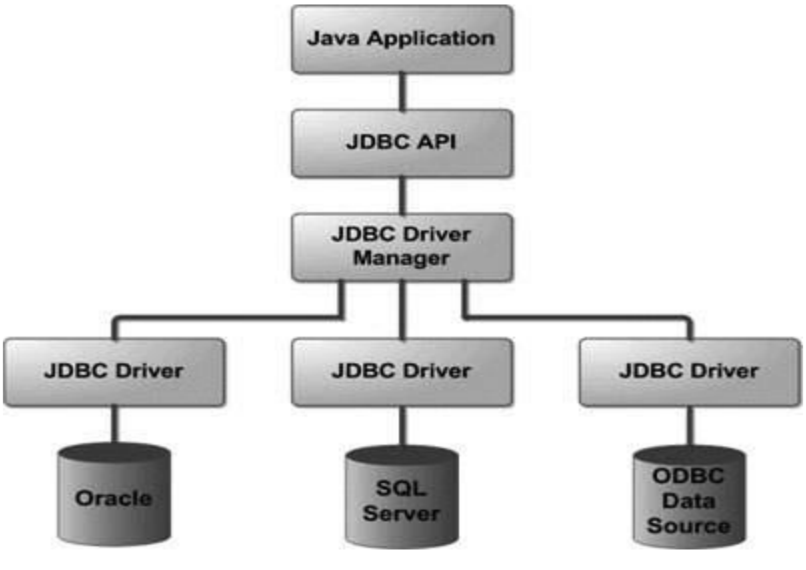


Fig 1.8 Architecture Of JDBC

-HADOOP

Hadoop is an open source, Java-based programming framework that supports the processing and storage of extremely large data sets in a distributed computing environment. It is part of the Apache project sponsored by the Apache Software Foundation.

The base Apache Hadoop framework is composed of the following modules:

Hadoop Common : contains libraries and utilities needed by other Hadoop modules;

Hadoop Distributed File System (HDFS) : a distributed file-system that stores data on commodity machines, providing very high aggregate bandwidth across the cluster;

Hadoop YARN : a resource-management platform responsible for managing computing resources in clusters and using them for scheduling of users' applications; and

Hadoop MapReduce: an implementation of the MapReduce programming model for large scale data processing.

Hadoop makes it possible to run applications on systems with thousands of commodity hardware nodes, and to handle thousands of terabytes of data. Its distributed file system facilitates rapid data transfer rates among nodes and allows the system to continue operating in case of a node failure.

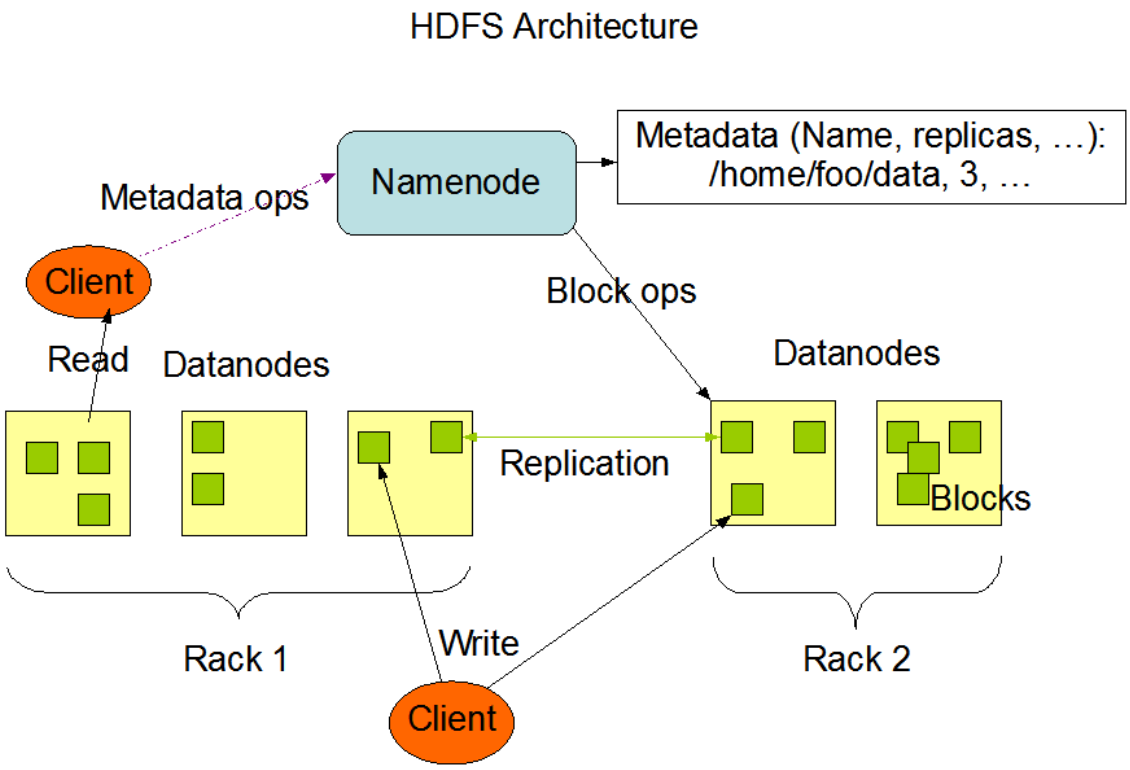


Fig 1.9 HDFS Architecture

-MAPREDUCE

Hadoop MapReduce is a software framework for easily writing applications which process vast amounts of data (multi-terabyte data-sets) in-parallel on large clusters (thousands of nodes) of commodity hardware in a reliable, fault-tolerant manner. A MapReduce job usually splits the input data-set into independent chunks which are processed by the map tasks in a completely parallel manner. The framework sorts the outputs of the maps, which are then input to the reduce tasks. Typically both the input and the output of the job are stored in a file-system. The framework takes care of scheduling tasks, monitoring them and re-executes the failed tasks. Typically the compute nodes and the storage nodes are the same, that is, the MapReduce framework and the Hadoop Distributed File System are running on the same set of nodes. This configuration allows the framework to effectively schedule tasks on the nodes where data is already present, resulting in very high aggregate bandwidth across the cluster. The MapReduce framework consists of a single master Job Tracker and one slave Task Tracker per cluster-node. The master is responsible for scheduling the jobs' component tasks on the slaves, monitoring them and re-executing the failed tasks. The slaves execute the tasks as directed by the master. Minimally, applications specify the input/output locations and supply map and reduce functions via implementations of appropriate interfaces and/or abstract-classes. These, and other job parameters, comprise the job configuration. The Hadoop job client then submits the job (jar/executable etc.) and configuration to the Job Trackerwhich then assumes the responsibility of distributing the software/configuration to the slaves, scheduling tasks and monitoring them, providing status and diagnostic information to the job-client.

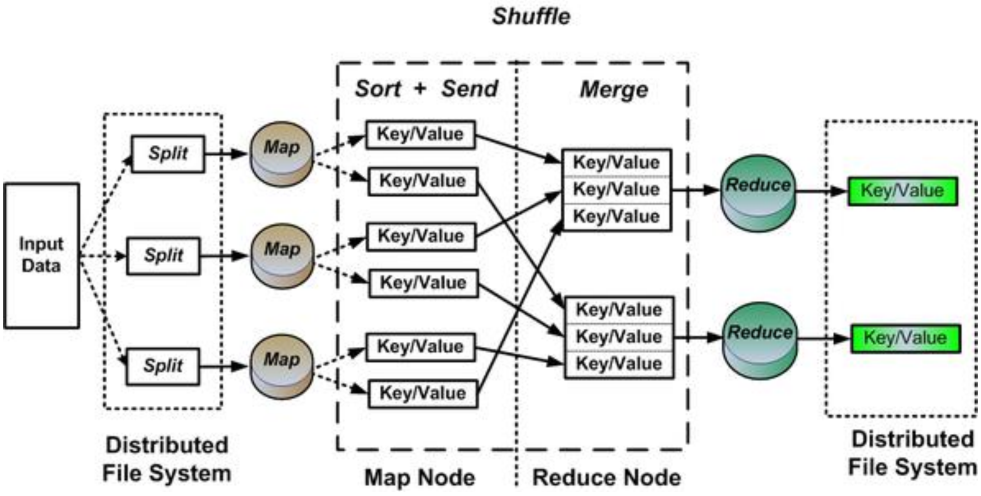


Fig 1.10 MapReduce Architecture

-HIVE

Hive is a data warehouse infrastructure tool to process structured data in Hadoop. It resides on top of Hadoop to summarize Big Data, and makes querying and analysing easy.Apache

Hive is a datawarehouse infrastructure built on top of Hadoop for providing data summarization, query, and analysis Hive gives an SQL-like interface to query data stored in various databases and file systems that integrate with Hadoop.Initially, Hive was developed by Facebook, later the Apache Software Foundation took it up and developed it further as an open source under the name Apache Hive. While Hive works on an SQL-dialect, there are a lot of differences in structure and working of Hive in comparison to relational databases.The storage and querying operations of Hive closely resemble with that of traditional databases.It is used by different companies. For example, Amazon uses it in Amazon Elastic MapReduce.

Features of Hive:

- It stores schema in a database and processed data into HDFS.
- It is designed for OLAP.
- It provides SQL type language for querying called HiveQL or HQL.
- It is familiar, fast, scalable, and extensible.

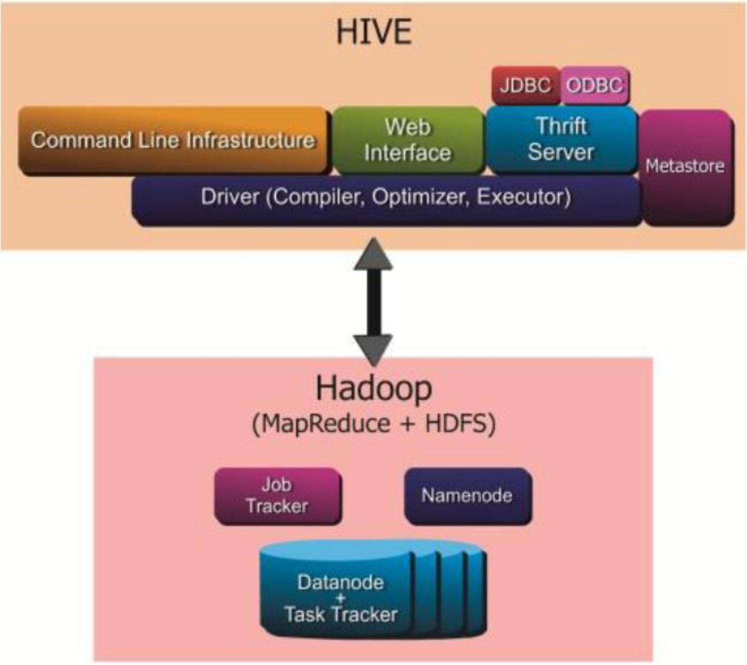


Fig 1.11 Hive Architecture

-PIG

Apache Pig is an abstraction over MapReduce. It is a tool/platform which is used to analyse larger sets of data representing them as data flows. Pig is generally used with Hadoop; we can perform all the data manipulation operations in Hadoop using Apache Pig. To write data analysis programs, Pig provides a high-level language known as Pig Latin. This language provides various operators using which programmers can develop their own functions for reading, writing, and processing data. To analyse data using Apache Pig, programmers need to write scripts using Pig Latin language. All these scripts are internally converted to Map and Reduce tasks. Apache Pig has a component known as Pig Engine that accepts the Pig Latin scripts as input and converts those scripts into MapReduce jobs.

Features of Pig:

- Using Pig Latin, programmers can perform MapReduce tasks easily without having to type complex codes in Java.
- Apache Pig uses multi-query approach, thereby reducing the length of codes. For example, an operation that would require you to type 200 lines of code (LoC) in Java can be easily done by typing as less as just 10 LoC in Apache Pig
- Pig Latin is SQL-like language and it is easy to learn Apache Pig when you are familiar with SQL.
- Apache Pig provides many built-in operators to support data operations like joins, filters, ordering, etc. In addition, it also provides nested data types like tuples, bags, and maps that are missing from MapReduce.

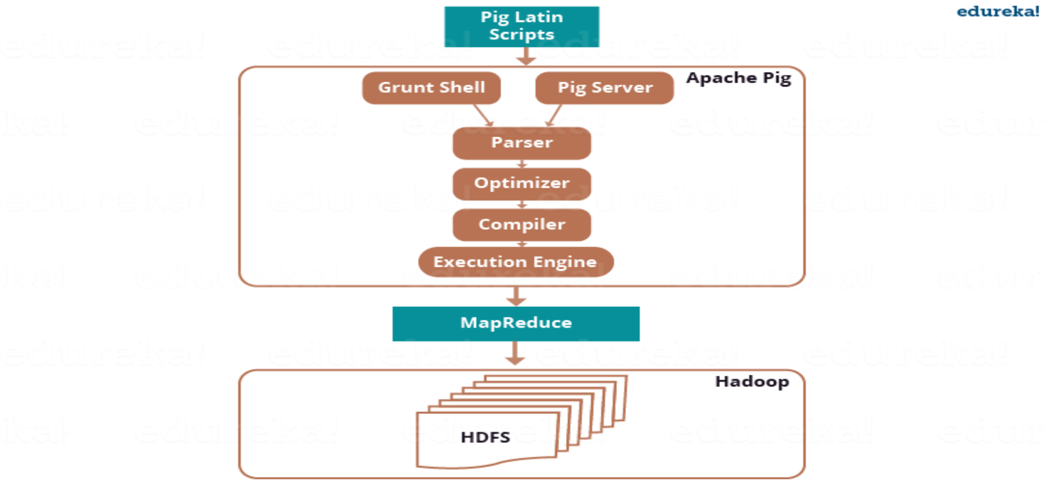


Fig 1.12 PIG Architecture

-JAQL

Jaql is currently used in IBM's Info SphereBig Insights and Cognos Consumer Insight products. Jaql's data model is inspired by JSON and can be used to represent datasets that vary from flat, relational tables to collections of semi structured documents. A Jaql script can start without any schema and evolve over time from a partial to a rigid schema. Reusability is provided through the use of higher-order functions and by packaging related functions into modules. Most Jaql scripts work at a high level of abstraction for concise specification of logical operations (e.g., join), but Jaql's notion of physical transparency also provides a lower level of abstraction if necessary. This allows users to pin down the evaluation plan of a script for greater control or even add new operators.

Features of JAQL are:

- Flexible data model
- Reusability
- Varying levels of abstraction
- Scalability

-HBASE

HBase is a distributed column-oriented database built on top of the Hadoop file system. It is an open-source project and is horizontally scalable. HBase is a data model that is similar to Google's big table designed to provide quick random access to huge amounts of structured data. It leverages the fault tolerance provided by the Hadoop File System (HDFS). It is a part of the Hadoop ecosystem that provides random real-time read/write access to data in the Hadoop File System. HBase is not a direct replacement for a classic SQL database, however Apache Phoenix project provides an SQL layer for HBase as well as JDBC driver that can be integrated with various analytics and business intelligence applications. The Apache Trafodion project provides a SQL query engine with ODBC and JDBC driversHBase is now serving several data-driven websites including Facebook's Messaging Platform. Unlike relational and traditional databases, HBase does not support SQL scripting; instead the equivalent is written in Java, employing similarity with a MapReduce application.[2]One can store the data in HDFS either directly or through HBase. Data consumer reads/accesses the data in HDFS randomly using HBase. HBase sits on top of the Hadoop File System and provides read and write access.

Features of HBase:

- HBase is linearly scalable.
- It has automatic failure support.
- It provides consistent read and writes.
- It integrates with Hadoop, both as a source and a destination.
- It has easy java API for client.
- It provides data replication across clusters.

4.1.2 Integrating Tools

For Big Data analysis where the data size is petabytes or zettabytes, analysing such big data and performing certain operations on the data could be cumbersome without using these tools and would involve large amount of time and man-hours for analysis. Moreover the results would be more prone to human error.

Using these tools to analyse the data would be accurate, will take less time and the system would be quite robust and free from any form of errors. The process is as follows:

Firstly data is collected from a survey or a firm and further analysis is performed depending on the customer's requirement.

There are certain operations on the data could be cumbersome without using these tools and would involve large amount of time and man-hours for analysis. Moreover the results would be more prone to human error.

Using these tools to analyse the data would be accurate, will take less time and the system would be quite robust and free from

Depending upon the data (Structured/Unstructured) appropriate database management system is used. My SQL is used for structured data and JQL is used for unstructured data.

Now JDBC is used to make the connectivity between the java programme and DBMS.

Using Hadoop, depending upon the requirement of the customer Map Reduce technique is used and the desired data set is retrieved.

GUI front end will be created using the AWT feature of JAVA.

Using Statistics tools we can show our analysis by bar graph, line graphs.

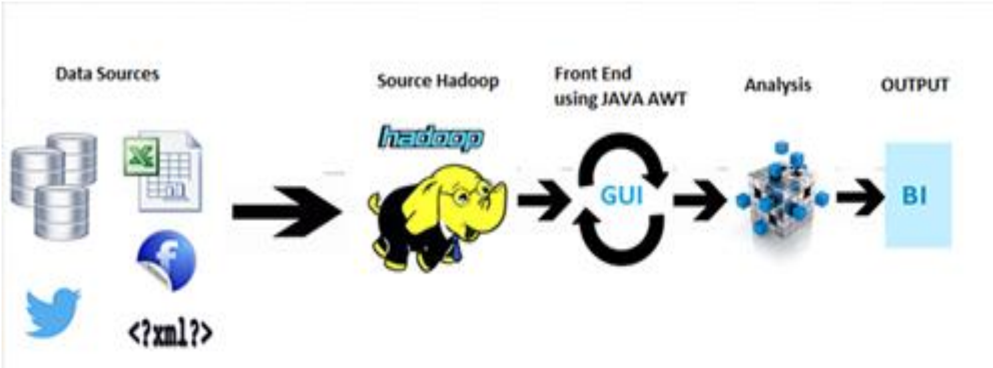


Fig 1.13 Steps Involved During Integration

5.1 Snapshots of the system

Main Form

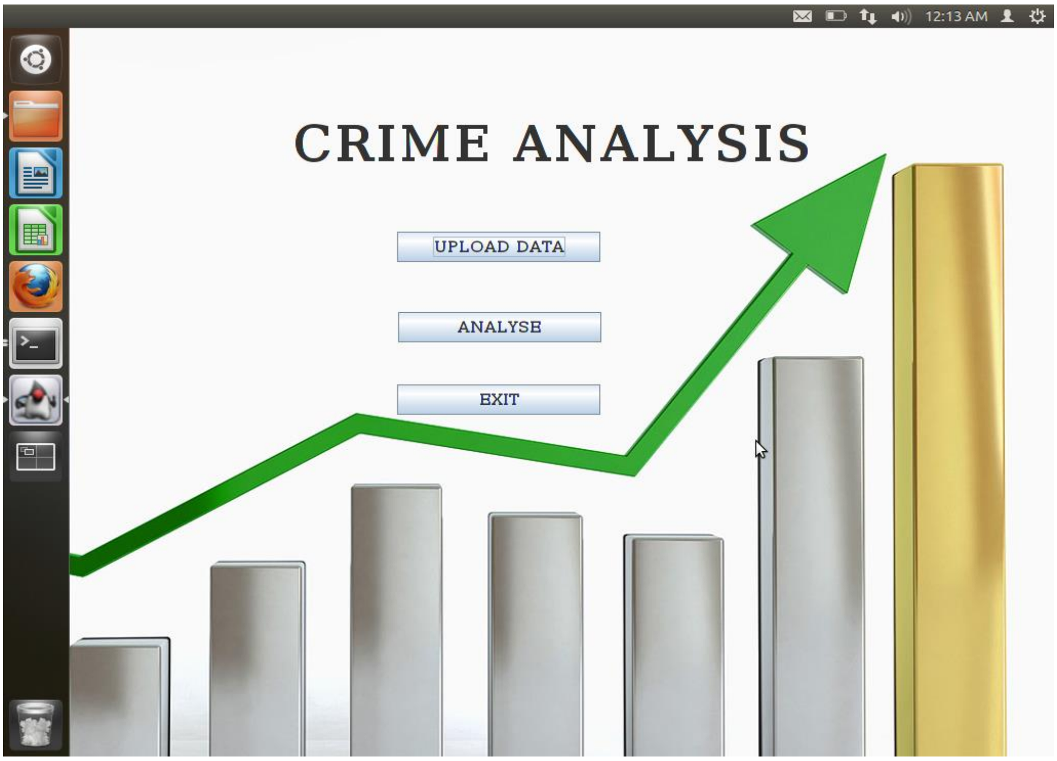


Fig 1.14 Main Form

The main form of the project describes the major two activities/sub-application that it performs i.e. **UPLOADING THE DATA** and **ANALYZATION.**

Uploading Window(for uploading data)

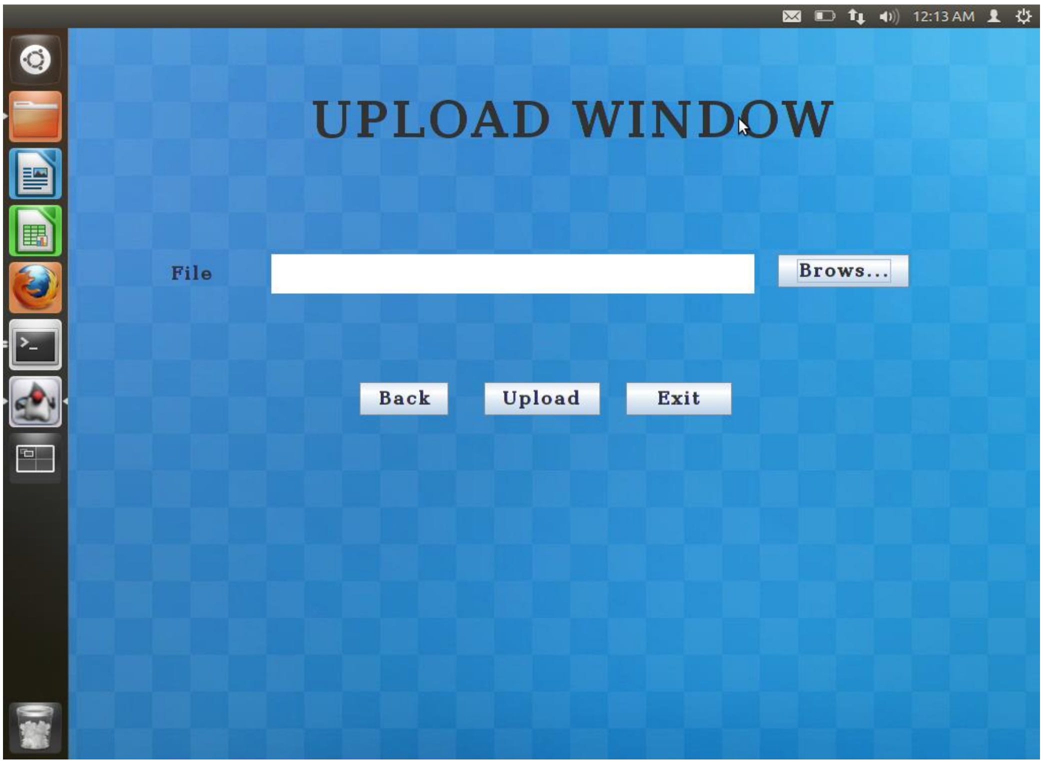


Fig 1.15 Upload Window

The upload window uploads the data in the structure when we browse a particular file in it.

Analysis Window(for analysation)

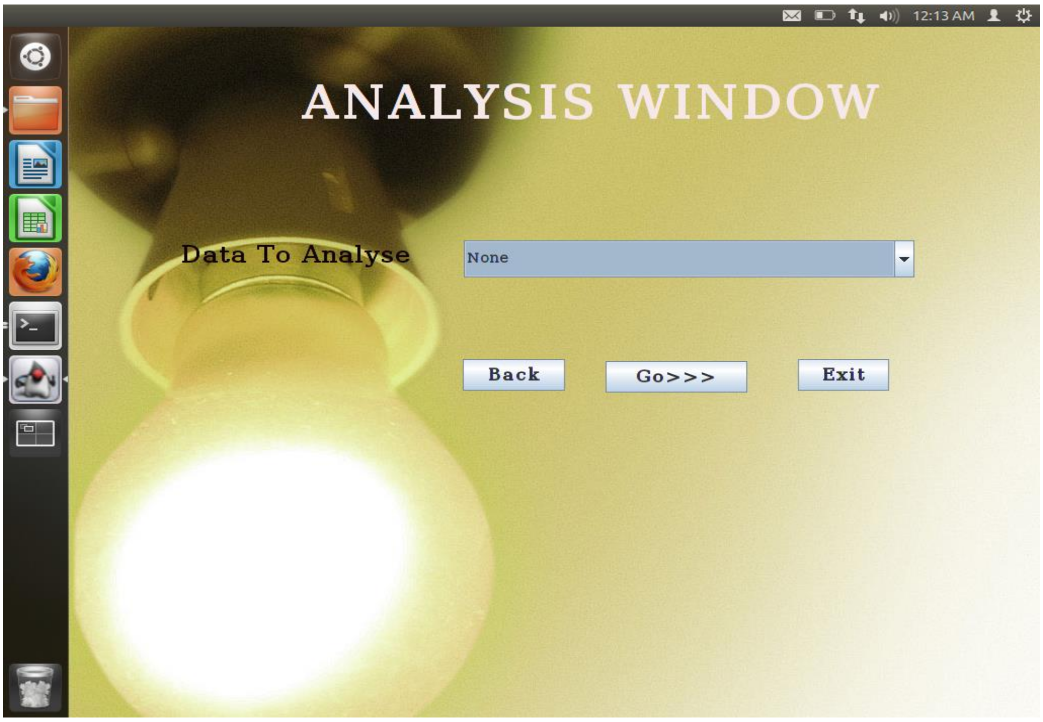
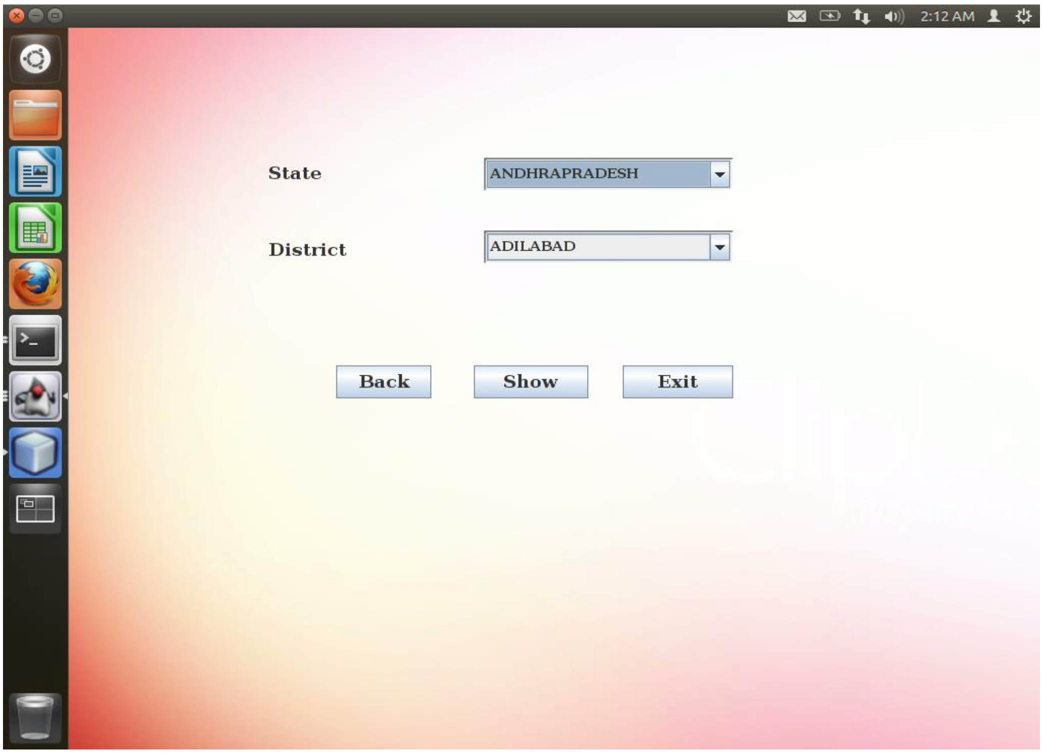


Fig 1.16 Analysis Window

The analysis window allows us to analyse the data in 8 different ways i.e. different perspectives of analysation for different views.

YearWise_CrimeType_Of_Specific_State



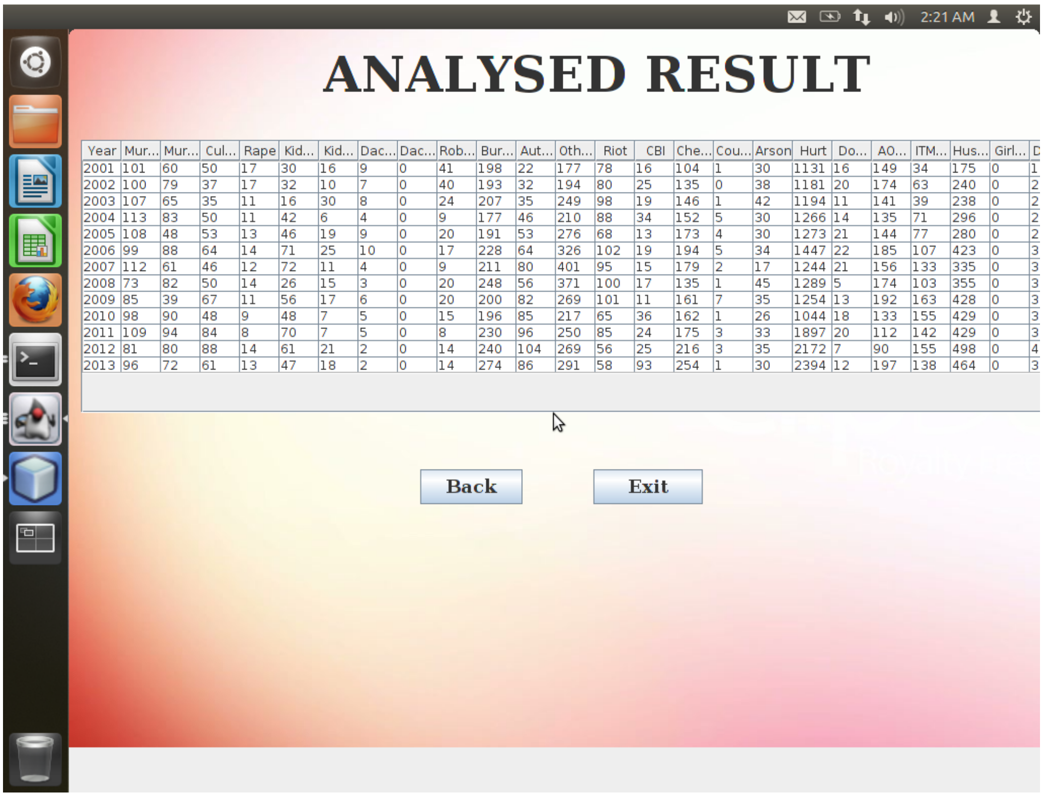
State: ANDHRAPRADESH

District: ADILABAD

Fig 1.17 Yearwise_Crimetype_Of_Specific_State

In this form, according to particular district of a particular state, all the counts of all the crimetypes will be analysed.

Result-



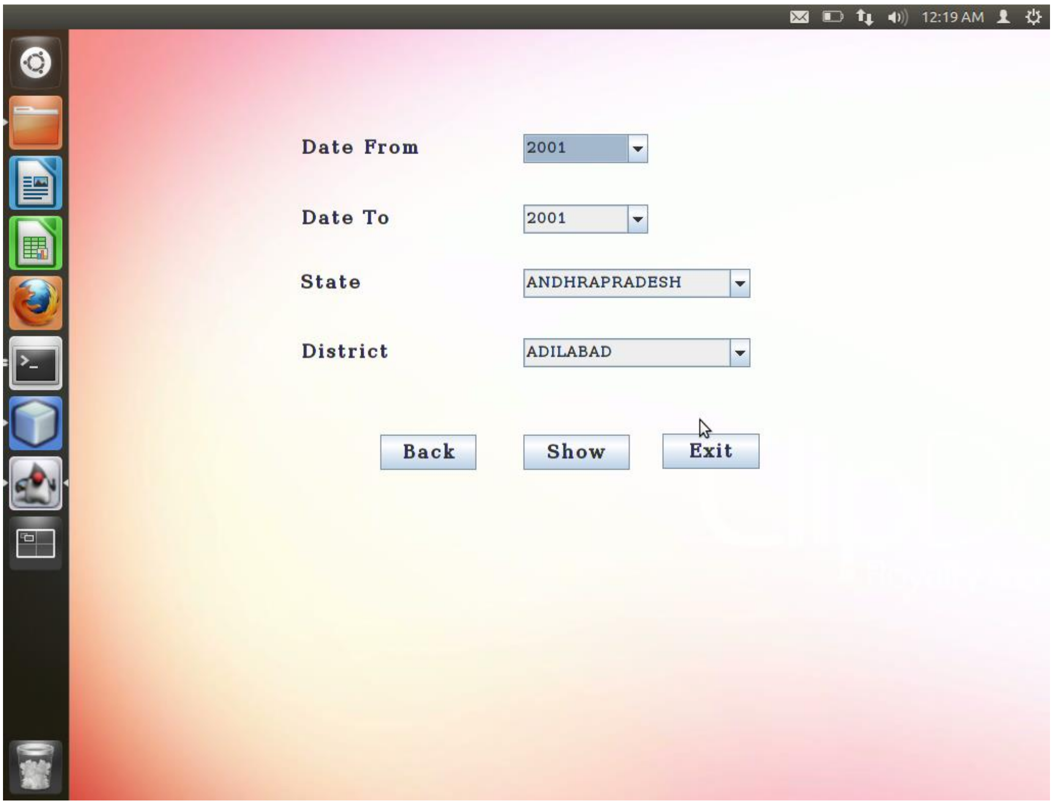
ANALYSED RESULT

Year	Mur...	Mur...	Cul...	Rape	Kid...	Kid...	Dac...	Dac...	Rob...	Bur...	Aut...	Oth...	Riot	CBI	Che...	Cou...	Arson	Hurt	Do...	AO...	ITM...	Hus...	Girl...	D
2001	101	60	50	17	30	16	9	0	41	198	22	177	78	16	104	1	30	1131	16	149	34	175	0	1
2002	100	79	37	17	32	10	7	0	40	193	32	194	80	25	135	0	38	1181	20	174	63	240	0	2
2003	107	65	35	11	16	30	8	0	24	207	35	249	98	19	146	1	42	1194	11	141	39	238	0	2
2004	113	83	50	11	42	6	4	0	9	177	46	210	88	34	152	5	30	1266	14	135	71	296	0	2
2005	108	48	53	13	46	19	9	0	20	191	53	276	68	13	173	4	30	1273	21	144	77	280	0	2
2006	99	88	64	14	71	25	10	0	17	228	64	326	102	19	194	5	34	1447	22	185	107	423	0	3
2007	112	61	46	12	72	11	4	0	9	211	80	401	95	15	179	2	17	1244	21	156	133	335	0	3
2008	73	82	50	14	26	15	3	0	20	248	56	371	100	17	135	1	45	1289	5	174	103	355	0	3
2009	85	39	67	11	56	17	6	0	20	200	82	269	101	11	161	7	35	1254	13	192	163	428	0	3
2010	98	90	48	9	48	7	5	0	15	196	85	217	65	36	162	1	26	1044	18	133	155	429	0	3
2011	109	94	84	8	70	7	5	0	8	230	96	250	85	24	175	3	33	1897	20	112	142	429	0	3
2012	81	80	88	14	61	21	2	0	14	240	104	269	56	25	216	3	35	2172	7	90	155	498	0	4
2013	96	72	61	13	47	18	2	0	14	274	86	291	58	93	254	1	30	2394	12	197	138	464	0	3

Fig 1.18 YEARWISE_CRIMETYPE_OF_SPECIFIC_STATE RESULT

The analysed result of the previous form showing data yearly of different crimetype.

Total_CrimeType_Of_State_In_Specific_Year

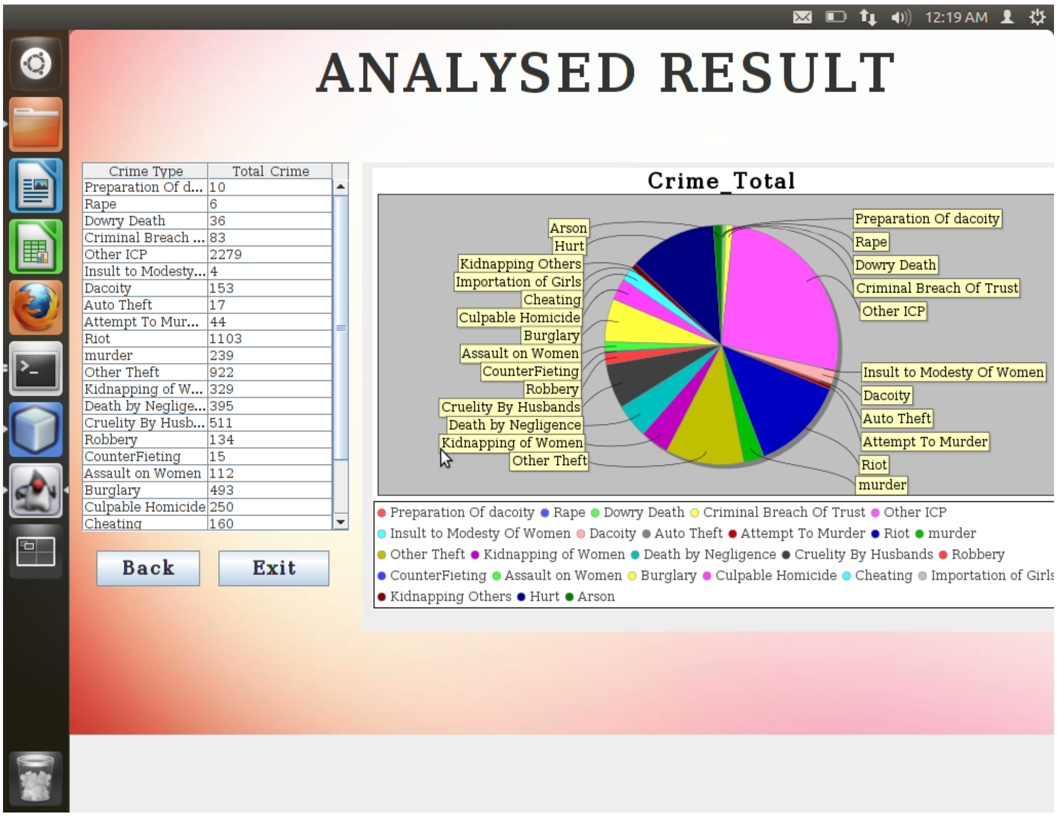


Date From	2001
Date To	2001
State	ANDHRAPRADESH
District	ADILABAD

Fig 1.19 Total_Crimetype_Of_State_In_Specific_Year

In this form, according to particular district of a particular state, total of the different crimetypes will be analysed within a range.

Result



ANALYSED RESULT

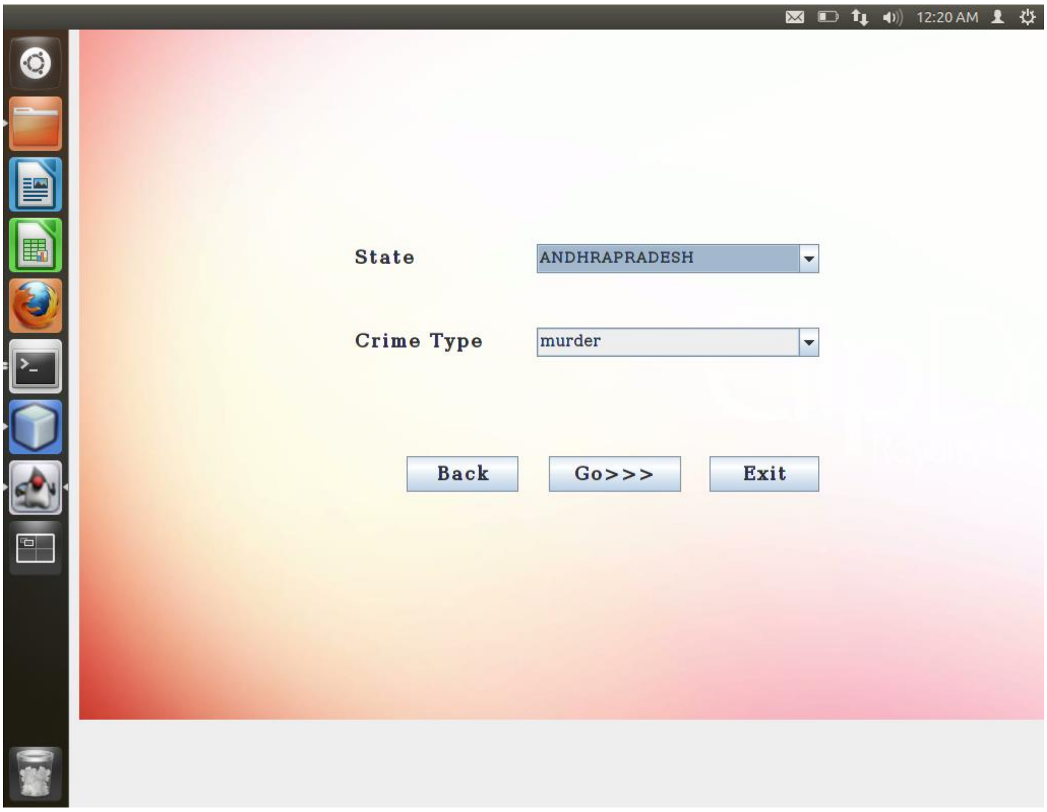
Crime Type	Total Crime
Preparation Of d...	10
Rape	6
Dowry Death	36
Criminal Breach ...	83
Other ICP	2279
Insult to Modesty...	4
Dacoity	153
Auto Theft	17
Attempt To Mur...	44
Riot	1103
murder	239
Other Theft	922
Kidnapping of W...	329
Death by Neglige...	395
Cruelity By Husb...	511
Robbery	134
CounterFieting	15
Assault on Women	112
Burglary	493
Culpable Homicide	250
Cheating	160

12:19 AM
Back
Exit
Crime_Total
Arson
Hurt
Kidnapping Others
Importation of Girls
Cheating
Culpable Homicide
Burglary
Assault on Women
CounterFieting
Robbery
Cruelity By Husbands
Death by Negligence
Kidnapping of Women
Other Theft
Preparation Of dacoity
Rape
Dowry Death
Criminal Breach Of Trust
Other ICP
Insult to Modesty Of Women
Dacoity
Auto Theft
Attempt To Murder
Riot
murder
● Preparation Of dacoity ● Rape ● Dowry Death ● Criminal Breach Of Trust ● Other ICP
● Insult to Modesty Of Women ● Dacoity ● Auto Theft ● Attempt To Murder ● Riot ● murder
● Other Theft ● Kidnapping of Women ● Death by Negligence ● Cruelity By Husbands ● Robbery
● CounterFieting ● Assault on Women ● Burglary ● Culpable Homicide ● Cheating ● Importation of Girls
● Kidnapping Others ● Hurt ● Arson

Fig 1.20 Total_CrimeType_Of_State_In_Specific_Year Result

The analysed result of the previous form showing data in a range of different crimetype.

CrimeType_State



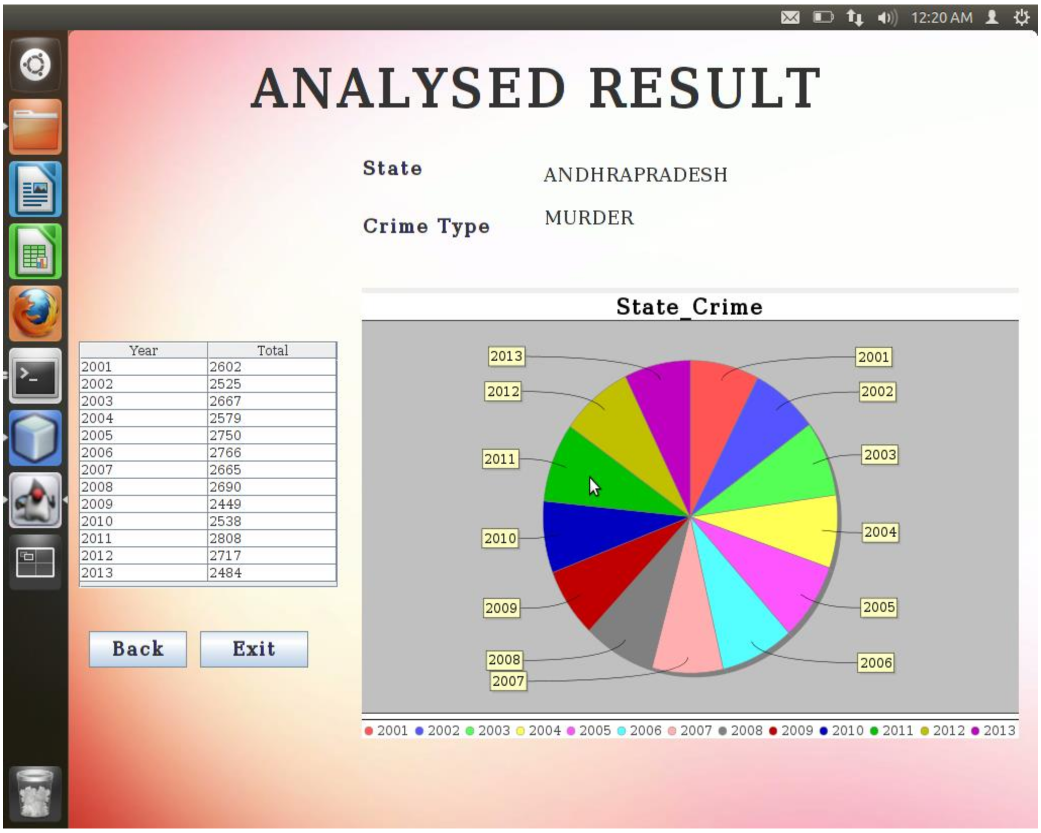
State: ANDHRAPRADESH

Crime Type: murder

Fig 1.21 CrimeType_State

In this form, according to a particular state and crimetype, total no. of crimes that has happened in different years will be analysed.

Result



ANALYSED RESULT

State ANDHRAPRADESH

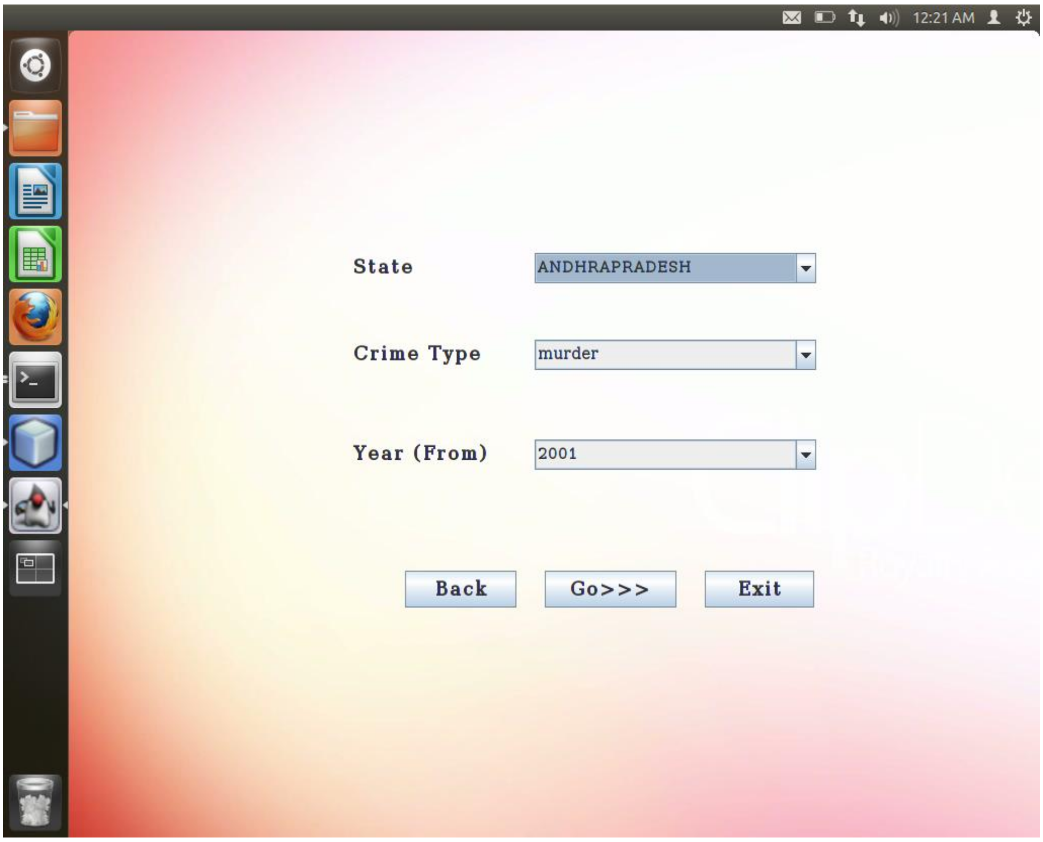
Crime Type MURDER

Year	Total
2001	2602
2002	2525
2003	2667
2004	2579
2005	2750
2006	2766
2007	2665
2008	2690
2009	2449
2010	2538
2011	2808
2012	2717
2013	2484

Fig 1.22 Crimetype_State Result

The analysed result of the previous form showing data yearly of total crime.

CrimeType Increasing Or Decreasing



State ANDHRAPRADESH

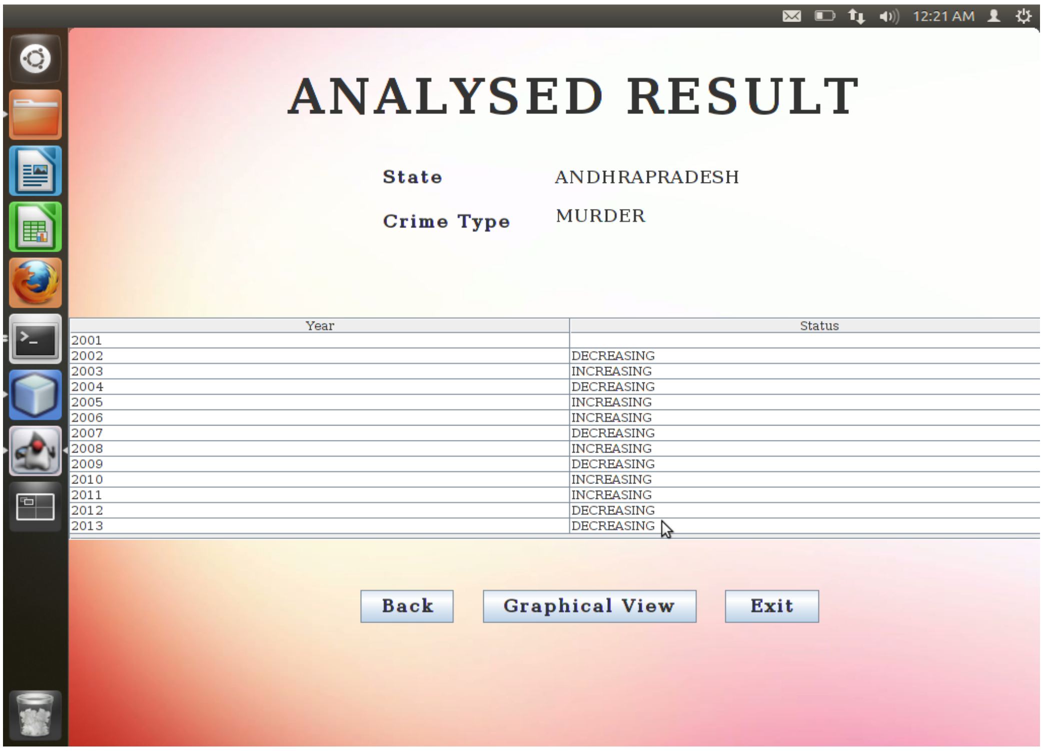
Crime Type murder

Year (From) 2001

Fig 1.23 Crimetype Increasing Or Decreasing

In this form, according to a particular state and crimetype along with the starting year(till the end), whether the total crime is increasing or decreasing is analysed with respect to previous year.

Result



ANALYSED RESULT

State ANDHRAPRADESH

Crime Type MURDER

Year	Status
2001	
2002	DECREASING
2003	INCREASING
2004	DECREASING
2005	INCREASING
2006	INCREASING
2007	DECREASING
2008	INCREASING
2009	DECREASING
2010	INCREASING
2011	INCREASING
2012	DECREASING
2013	DECREASING

Fig 1.24 Crimetype Increasing Or Decreasing Result

The analysed result of the previous form showing status yearly of total crime.

Graphical View

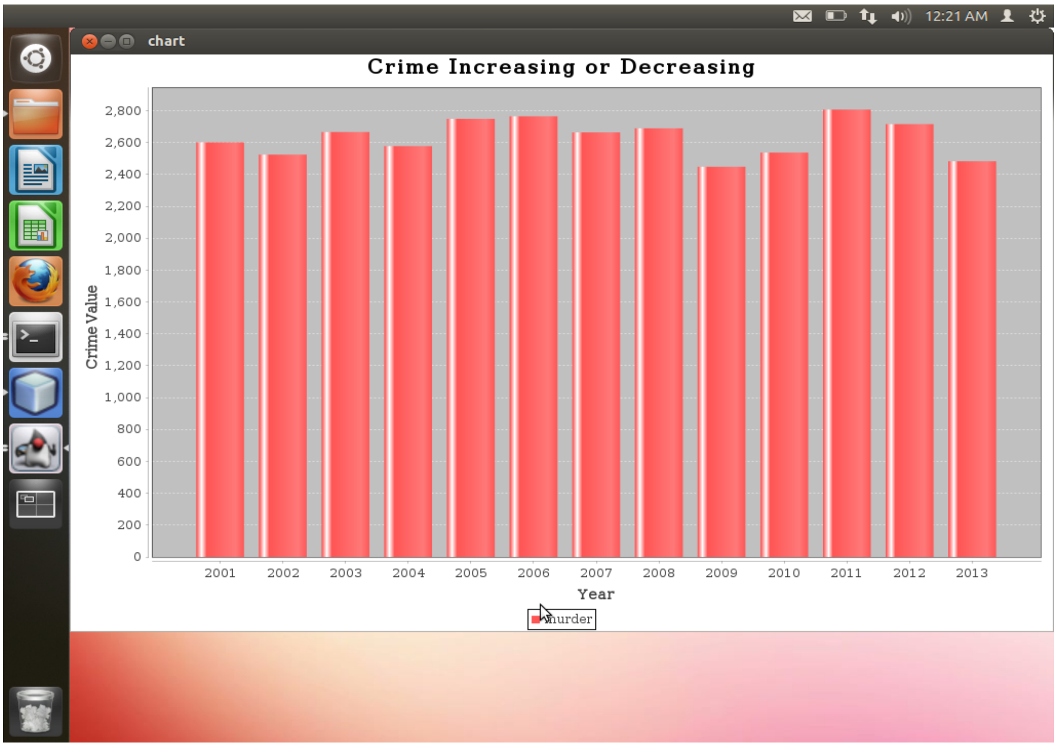
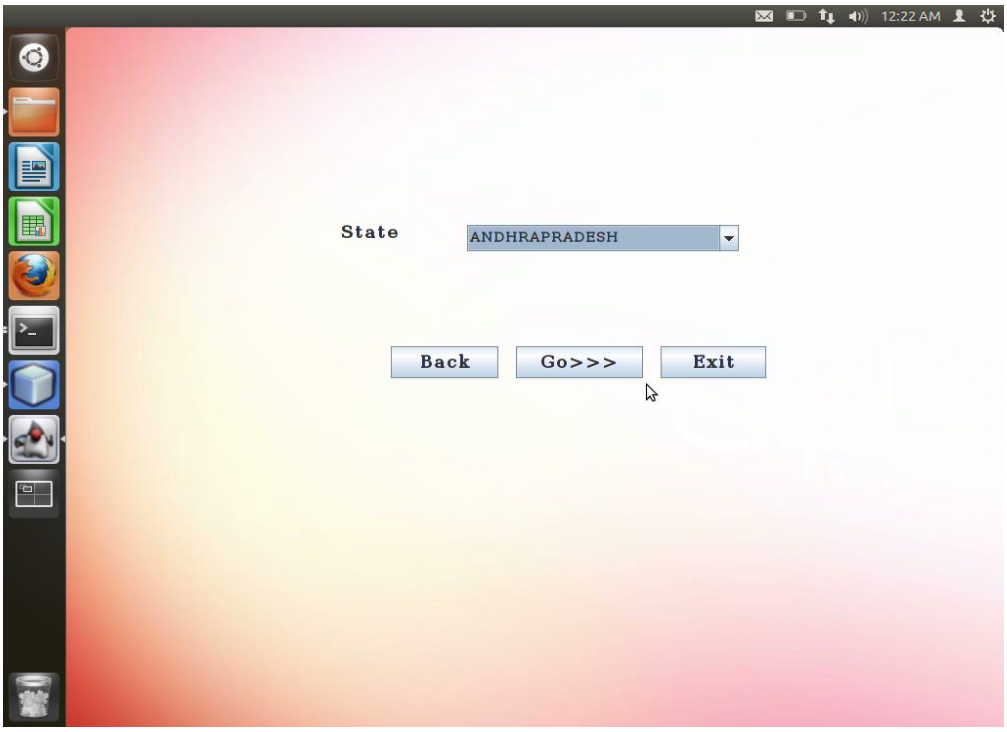


Fig 1.25 CrimeType Increasing Or Decreasing Graph

The graphical representation of the total crime in different years.

Overall Increase Or Decrease

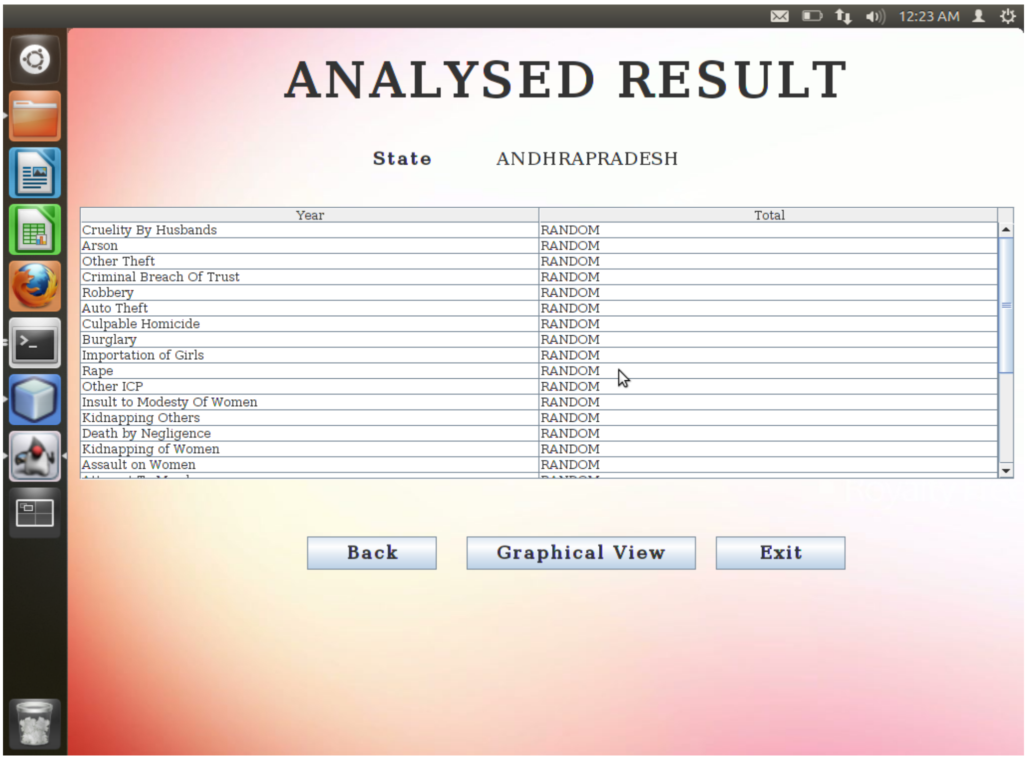


State ANDHRAPRADESH

Fig 1.26 Overall Increase Or Decrease

In this form, according to a particular state , whether the different crime is increasing or decreasing or random is analysed with respect to previous year.

Result



ANALYSED RESULT

State ANDHRAPRADESH

Year	Total
Cruelity By Husbands	RANDOM
Arson	RANDOM
Other Theft	RANDOM
Criminal Breach Of Trust	RANDOM
Robbery	RANDOM
Auto Theft	RANDOM
Culpable Homicide	RANDOM
Burglary	RANDOM
Importation of Girls	RANDOM
Rape	RANDOM
Other ICP	RANDOM
Insult to Modesty Of Women	RANDOM
Kidnapping Others	RANDOM
Death by Negligence	RANDOM
Kidnapping of Women	RANDOM
Assault on Women	RANDOM

Fig 1.27 Overall Increase Or Decrease Result

The analysed result of the previous form showing status yearly of different crime.

Graphical View

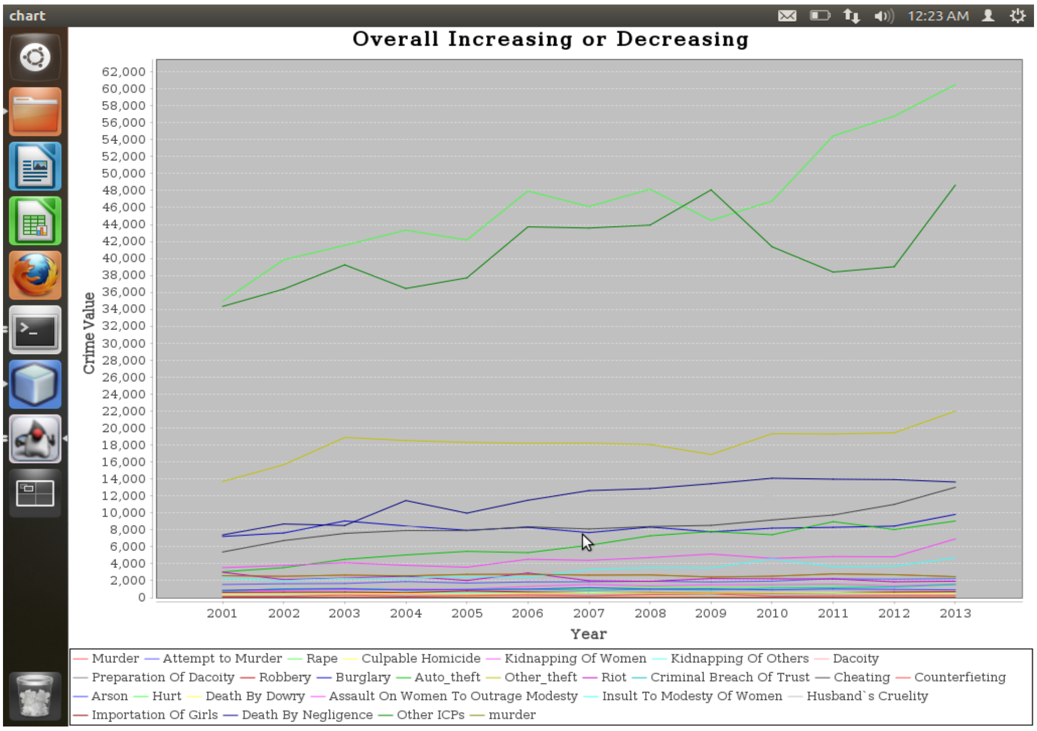
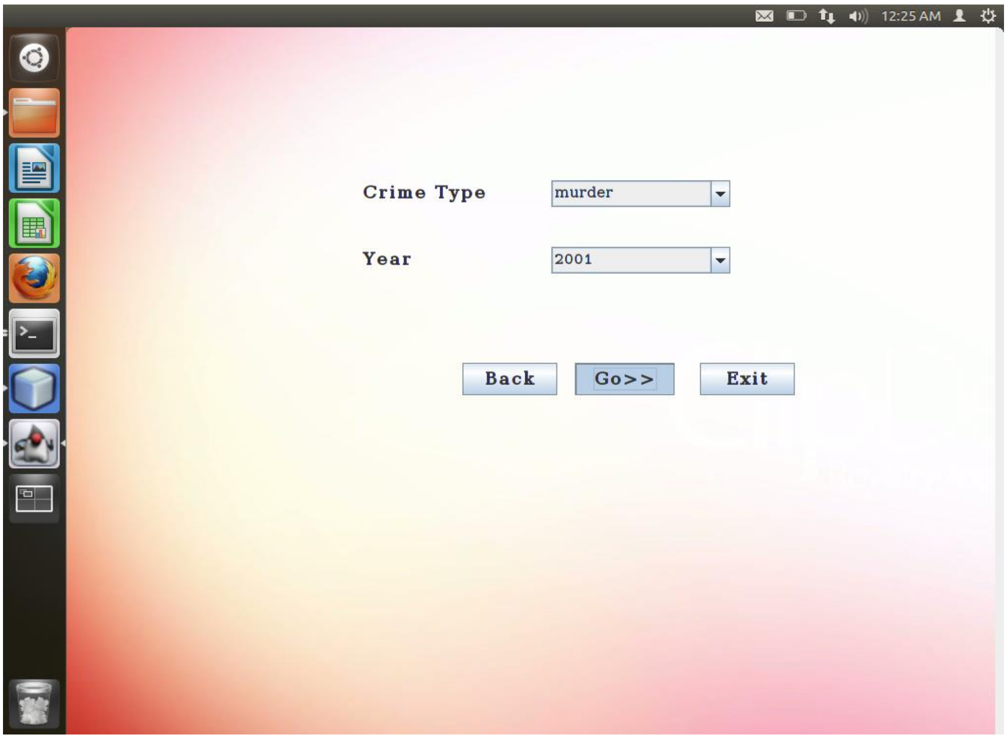


Fig 1.28 Overall Increase Or Decrease Graph

The graphical representation of the different crimes in different years.

Maximum and Minimum In Each State



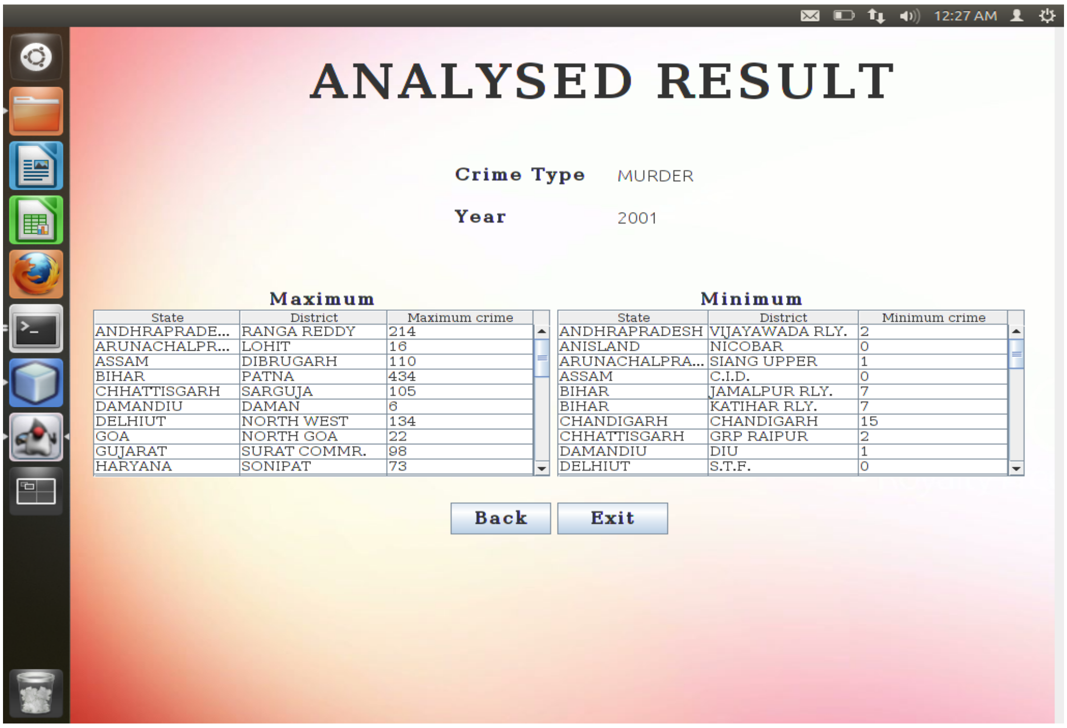
Crime Type murder

Year 2001

Fig 1.29 Maximum And Minimum In Each State

In this form, according to a particular crimetype and year , the maximum and minimum crime is shown along with the district.

Result



ANALYSED RESULT

Crime Type MURDER

Year 2001

Maximum

State	District	Maximum crime
ANDHRAPRADE...	RANGA REDDY	214
ARUNACHALPR...	LOHIT	16
ASSAM	DIBRUGARH	110
BIHAR	PATNA	434
CHHATTISGARH	SARGUJA	105
DAMANDIU	DAMAN	6
DELHIUT	NORTH WEST	134
GOA	NORTH GOA	22
GUJARAT	SURAT COMMR.	98
HARYANA	SONIPAT	73

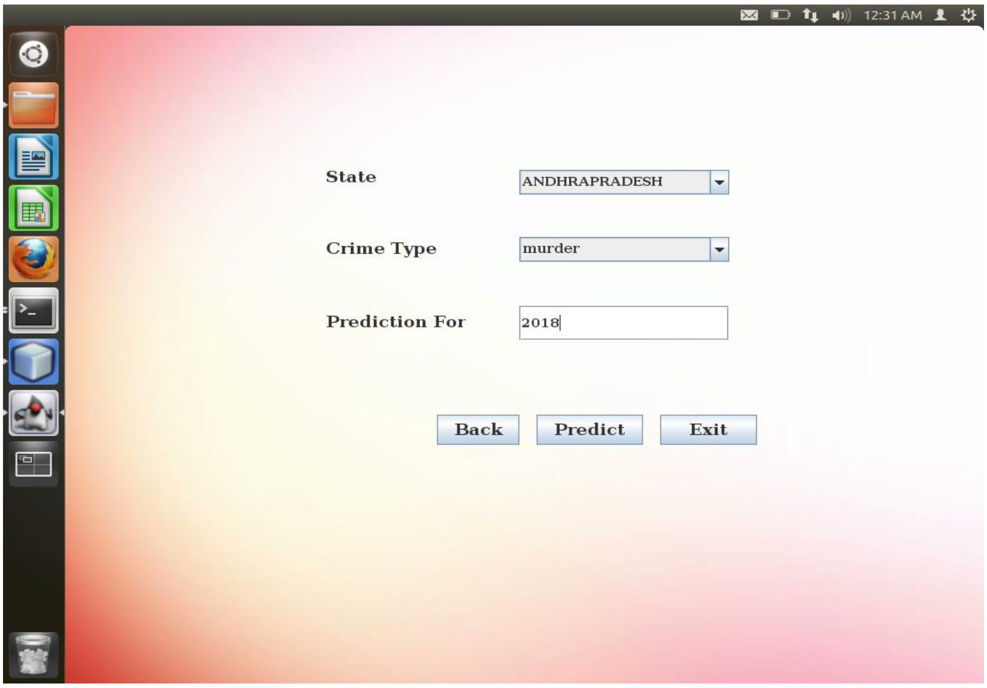
Minimum

State	District	Minimum crime
ANDHRAPRADESH	VIJAYAWADA RLY.	2
ANISLAND	NICOBAR	0
ARUNACHALPRA...	SIANG UPPER	1
ASSAM	C.I.D.	0
BIHAR	JAMALPUR RLY.	7
BIHAR	KATIHAR RLY.	7
CHANDIGARH	CHANDIGARH	15
CHHATTISGARH	GRP RAIPUR	2
DAMANDIU	DIU	1
DELHIUT	S.T.F.	0

Fig 1.30 Maximum And Minimum In Each State Result

The analysed result of the previous form showing maximum and minimum of a crime with their district in which it happened.

Crime Prediction



State: ANDHRAPRADESH

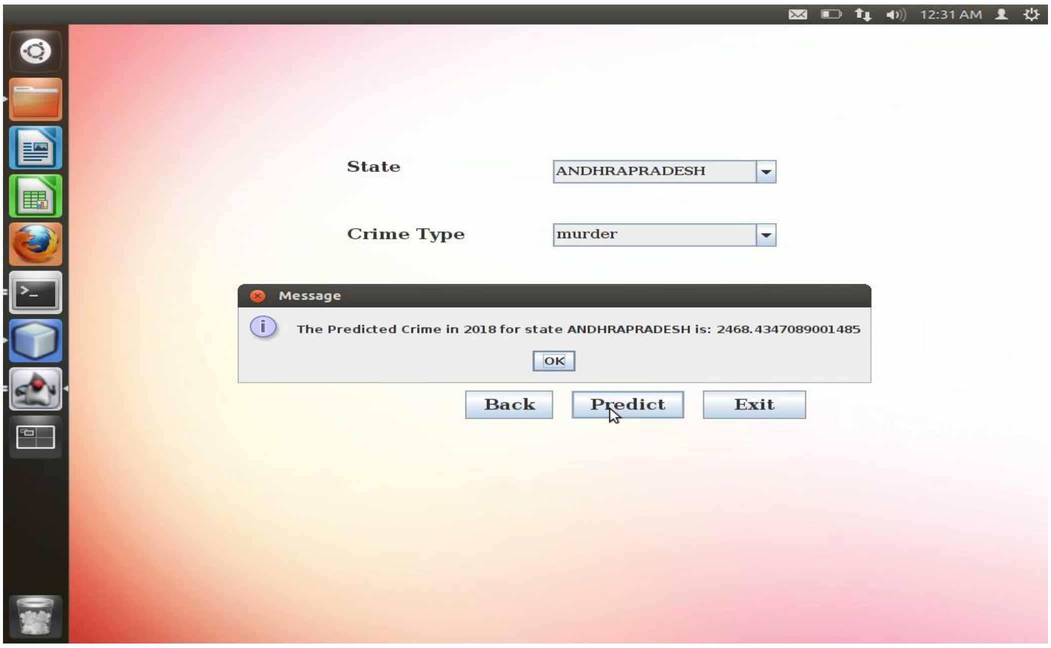
Crime Type: murder

Prediction For: 2018

Fig 1.31 Crime Prediction

In this form, according to a particular crimetype and state , the count of the crimetype is predicted as asked.

Result



State ANDHRAPRADESH

Crime Type murder

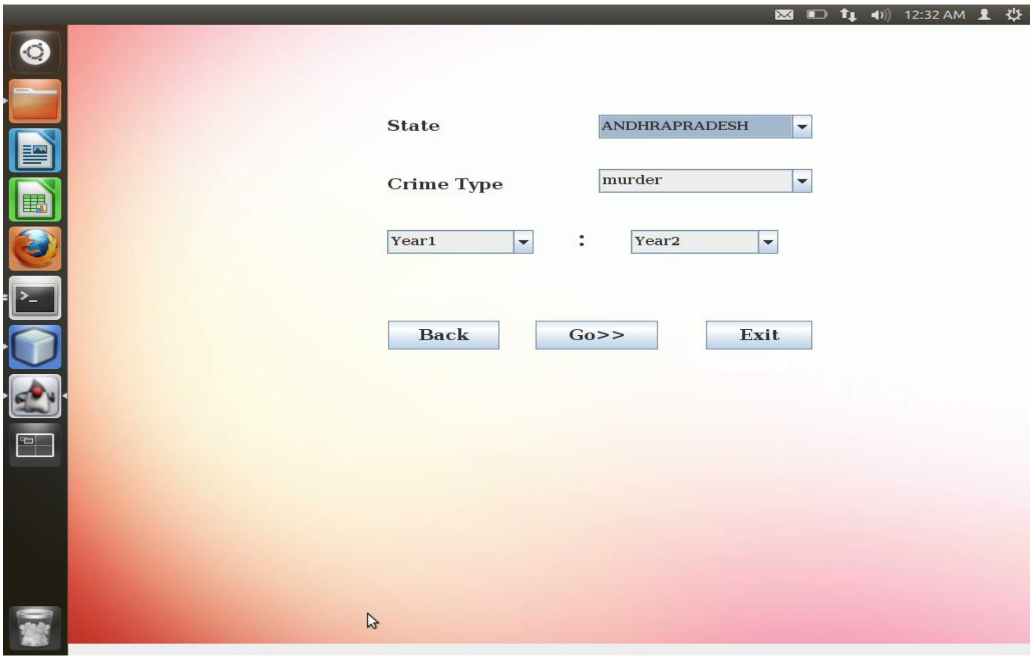
Message

The Predicted Crime in 2018 for state ANDHRAPRADESH is: 2468.4347089001485

Fig 1.32 Crime Prediction Result

The analysed result of the previous form showing prediction of a crime in a particular year.

Ratio of Crime



State ANDHRAPRADESH

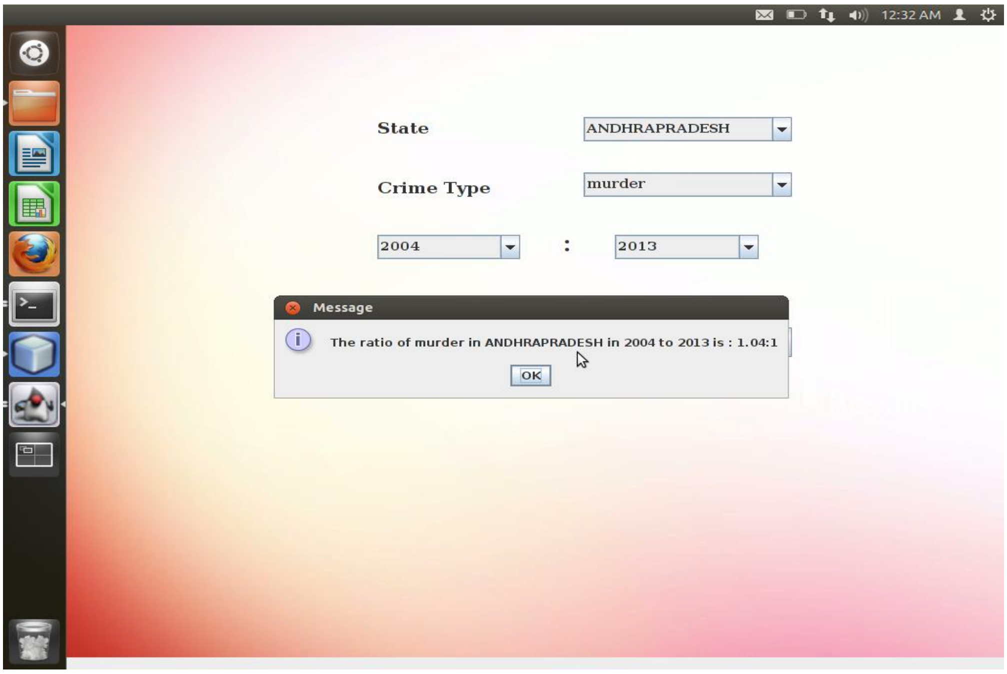
Crime Type murder

Year1 : Year2

Fig 1.33 Ratio Of Crime

In this form, according to a particular crimetype and state , the ratio of the count of crime of two years is calculated.

Result



State ANDHRAPRADESH

Crime Type murder

2004 : 2013

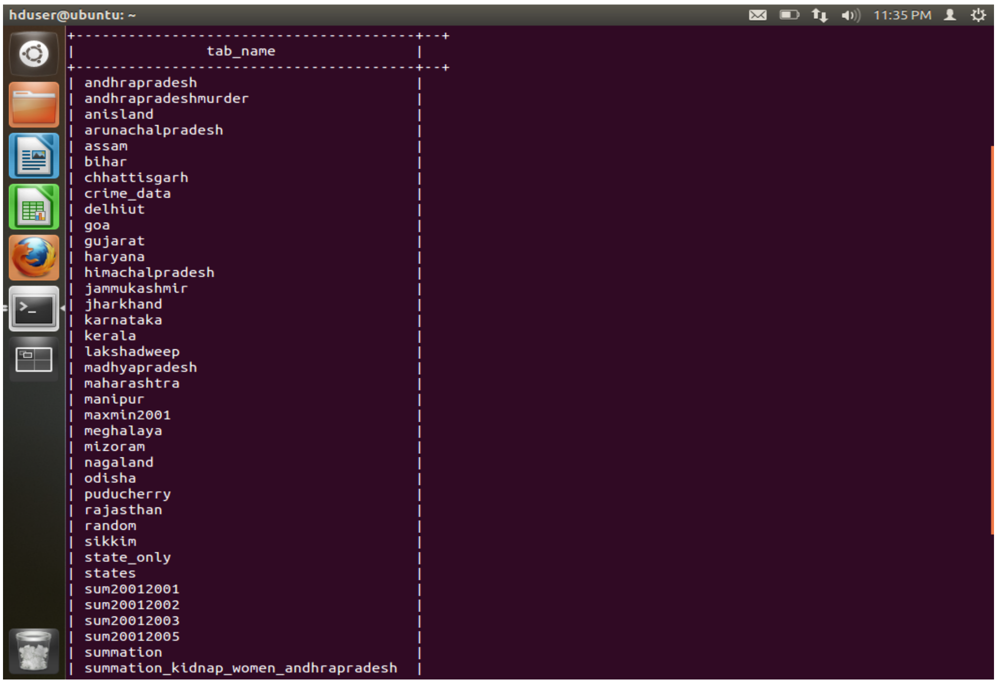
The ratio of murder in ANDHRAPRADESH in 2004 to 2013 is : 1.04:1

Fig 1.34 Ratio Of Crime Result

The analysed result of the previous form showing ratio of a crime in a state of two years.

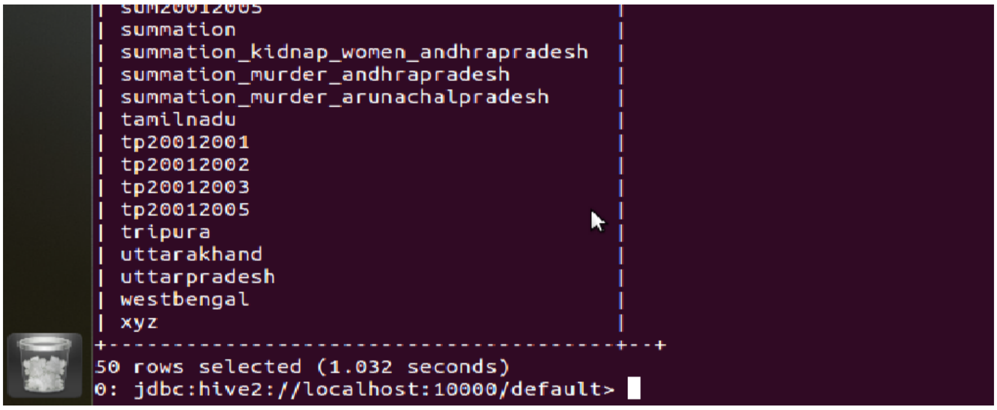
5.2Backeneds Representation Snapshots

Tables used in the Database(crime)



```
+----------------------------------------+--+
|                tab_name                |
+----------------------------------------+--+
| andhrapradesh                          |
| andhrapradeshmurder                    |
| anisland                               |
| arunachalpradesh                       |
| assam                                  |
| bihar                                  |
| chhattisgarh                           |
| crime_data                             |
| delhiut                                |
| goa                                    |
| gujarat                                |
| haryana                                |
| himachalpradesh                        |
| jammukashmir                           |
| jharkhand                              |
| karnataka                              |
| kerala                                 |
| lakshadweep                            |
| madhyapradesh                          |
| maharashtra                            |
| manipur                                |
| maxmin2001                             |
| meghalaya                              |
| mizoram                                |
| nagaland                               |
| odisha                                 |
| puducherry                             |
| rajasthan                              |
| random                                 |
| sikkim                                 |
| state_only                             |
| states                                 |
| sum20012001                            |
| sum20012002                            |
| sum20012003                            |
| sum20012005                            |
| summation                              |
| summation_kidnap_women_andhrapradesh   |
```

Fig 1.35 Tables(1)



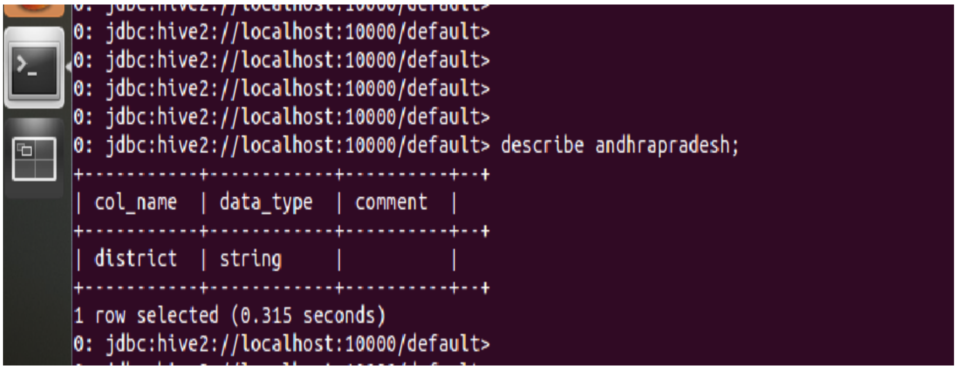
```
| sum20012005                            |
| summation                              |
| summation_kidnap_women_andhrapradesh   |
| summation_murder_andhrapradesh         |
| summation_murder_arunachalpradesh      |
| tamilnadu                              |
| tp20012001                             |
| tp20012002                             |
| tp20012003                             |
| tp20012005                             |
| tripura                                |
| uttarakhand                            |
| uttarpradesh                           |
| westbengal                             |
| xyz                                    |
+----------------------------------------+--+
50 rows selected (1.032 seconds)
0: jdbc:hive2://localhost:10000/default>
```

Fig 1.36 Tables(2)

The tables created statically and dynamically are as shown

Description of different tables

1. Different States

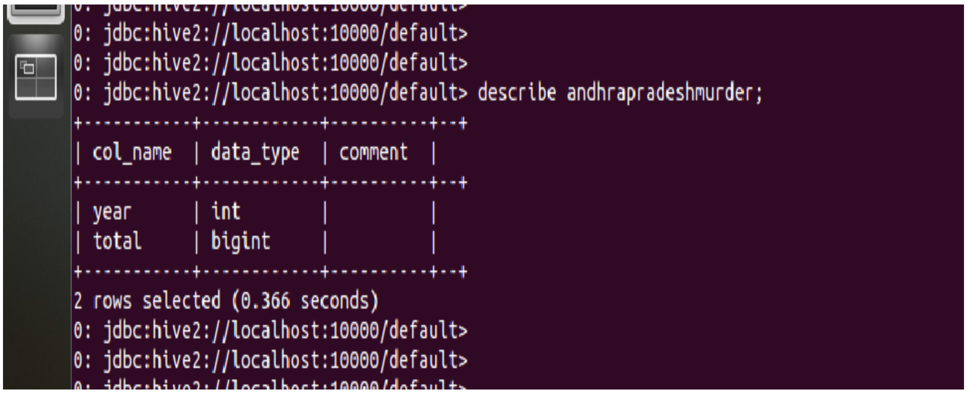


```
0: jdbc:hive2://localhost:10000/default>
0: jdbc:hive2://localhost:10000/default>
0: jdbc:hive2://localhost:10000/default>
0: jdbc:hive2://localhost:10000/default>
0: jdbc:hive2://localhost:10000/default>
0: jdbc:hive2://localhost:10000/default> describe andhrapradesh;
+-----------+-----------+----------+--+
| col_name  | data_type | comment  |
+-----------+-----------+----------+--+
| district  | string    |          |
+-----------+-----------+----------+--+
1 row selected (0.315 seconds)
0: jdbc:hive2://localhost:10000/default>
0: jdbc:hive2://localhost:10000/default>
```

Fig 1.37 Different States

Similarly, there are 28 tables of different states.

2. Different States and Crimetype

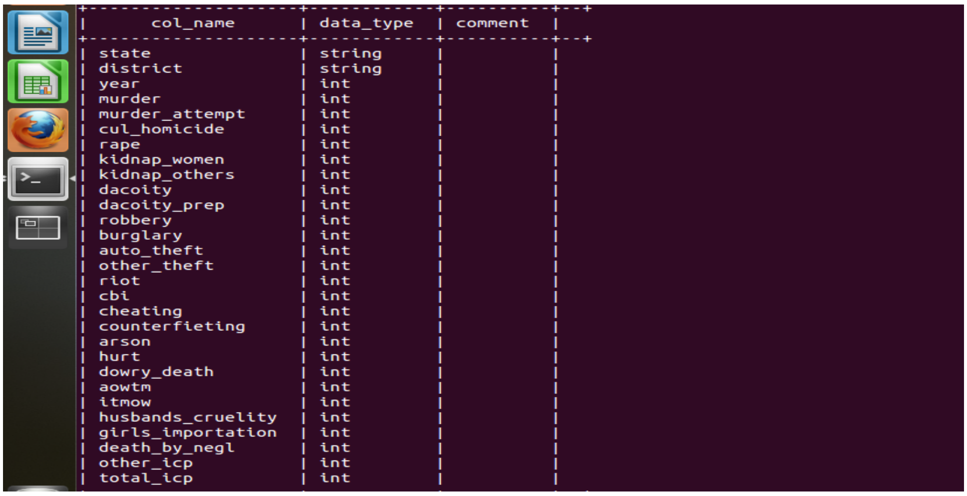


```
0: jdbc:hive2://localhost:10000/default>
0: jdbc:hive2://localhost:10000/default>
0: jdbc:hive2://localhost:10000/default>
0: jdbc:hive2://localhost:10000/default> describe andhrapradeshmurder;
+-----------+-----------+----------+--+
| col_name  | data_type | comment  |
+-----------+-----------+----------+--+
| year      | int       |          |
| total     | bigint    |          |
+-----------+-----------+----------+--+
2 rows selected (0.366 seconds)
0: jdbc:hive2://localhost:10000/default>
0: jdbc:hive2://localhost:10000/default>
0: jdbc:hive2://localhost:10000/default>
```

Fig 1.38 Different States And Crimetype

Table for each crimetype of each state.

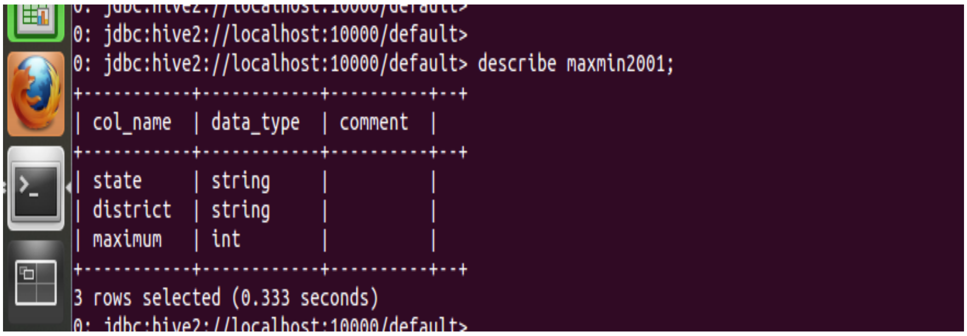
1. Base Table



```
+---------------------+-------------+-----------+--+
|       col_name      | data_type   | comment   |
+---------------------+-------------+-----------+--+
| state               | string      |           |
| district            | string      |           |
| year                | int         |           |
| murder              | int         |           |
| murder_attempt      | int         |           |
| cul_homicide        | int         |           |
| rape                | int         |           |
| kidnap_women        | int         |           |
| kidnap_others       | int         |           |
| dacoity             | int         |           |
| dacoity_prep        | int         |           |
| robbery             | int         |           |
| burglary            | int         |           |
| auto_theft          | int         |           |
| other_theft         | int         |           |
| riot                | int         |           |
| cbi                 | int         |           |
| cheating            | int         |           |
| counterfieting      | int         |           |
| arson               | int         |           |
| hurt                | int         |           |
| dowry_death         | int         |           |
| aowtm               | int         |           |
| itmow               | int         |           |
| husbands_cruelity   | int         |           |
| girls_importation   | int         |           |
| death_by_negl       | int         |           |
| other_icp           | int         |           |
| total_icp           | int         |           |
```

Fig 1.39 Base Table

2. Maximum and Minimum Table

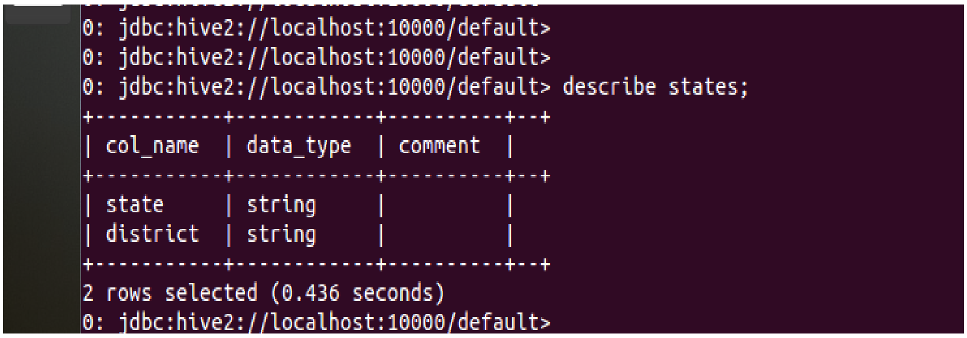


```
0: jdbc:hive2://localhost:10000/default>
0: jdbc:hive2://localhost:10000/default>
0: jdbc:hive2://localhost:10000/default> describe maxmin2001;
+-----------+-----------+----------+--+
| col_name  | data_type | comment  |
+-----------+-----------+----------+--+
| state     | string    |          |
| district  | string    |          |
| maximum   | int       |          |
+-----------+-----------+----------+--+
3 rows selected (0.333 seconds)
0: jdbc:hive2://localhost:10000/default>
```

Fig 1.40 Max And Min Table

Similarly, there are tables for each year's maximum and minimum.

3. States and District

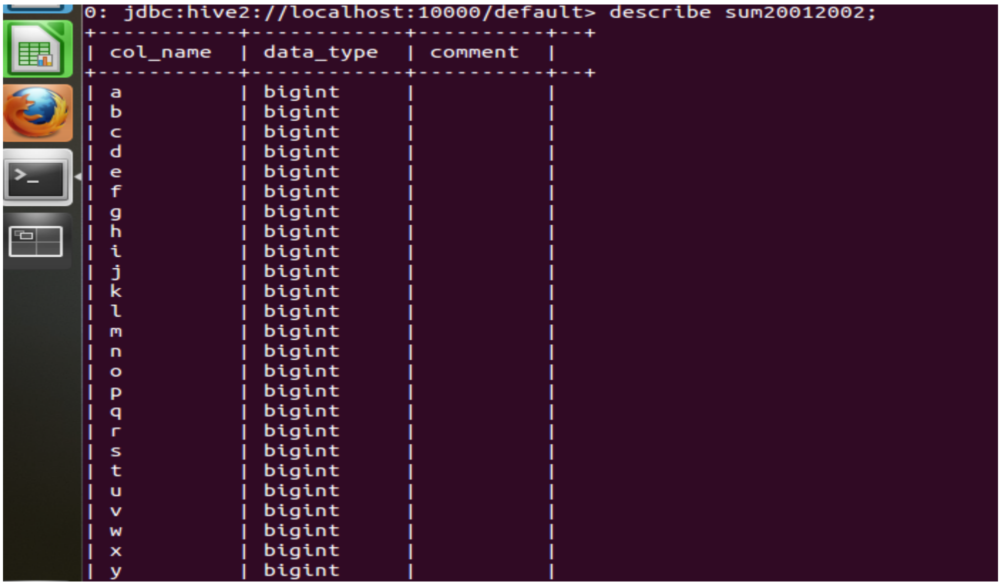


```
0: jdbc:hive2://localhost:10000/default>
0: jdbc:hive2://localhost:10000/default>
0: jdbc:hive2://localhost:10000/default> describe states;
+-----------+-----------+----------+--+
| col_name  | data_type | comment  |
+-----------+-----------+----------+--+
| state     | string    |          |
| district  | string    |          |
+-----------+-----------+----------+--+
2 rows selected (0.436 seconds)
0: jdbc:hive2://localhost:10000/default>
```

Fig 1.41 States And District Table

It shows the different states and their different districts.

4. Sum within the Range

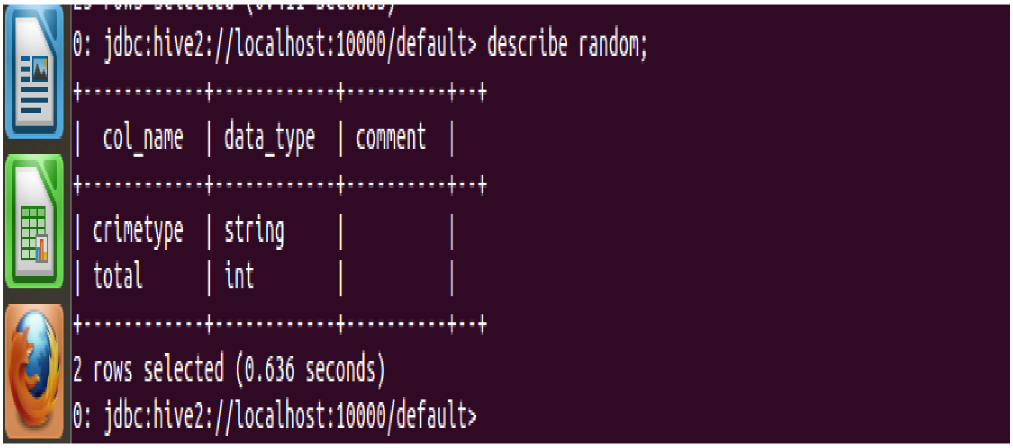


```
0: jdbc:hive2://localhost:10000/default> describe sum20012002;
+-----------+------------+----------+--+
| col_name  | data_type  | comment  |
+-----------+------------+----------+--+
| a         | bigint     |          |
| b         | bigint     |          |
| c         | bigint     |          |
| d         | bigint     |          |
| e         | bigint     |          |
| f         | bigint     |          |
| g         | bigint     |          |
| h         | bigint     |          |
| i         | bigint     |          |
| j         | bigint     |          |
| k         | bigint     |          |
| l         | bigint     |          |
| m         | bigint     |          |
| n         | bigint     |          |
| o         | bigint     |          |
| p         | bigint     |          |
| q         | bigint     |          |
| r         | bigint     |          |
| s         | bigint     |          |
| t         | bigint     |          |
| u         | bigint     |          |
| v         | bigint     |          |
| w         | bigint     |          |
| x         | bigint     |          |
| y         | bigint     |          |
```

Fig 1.42 Sum Within The Range Table

Similarly, sum of different ranges is stored in different tables.

5. Random

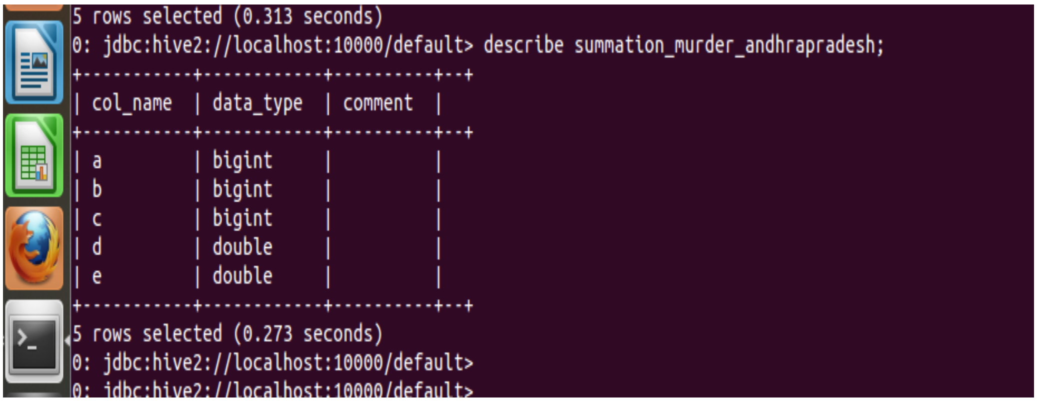


```
0: jdbc:hive2://localhost:10000/default> describe random;
+-----------+-----------+----------+--+
|  col_name  | data_type  | comment  |
+-----------+-----------+----------+--+
| crimetype  | string     |          |
| total      | int        |          |
+-----------+-----------+----------+--+
2 rows selected (0.636 seconds)
0: jdbc:hive2://localhost:10000/default>
```

Fig 1.43 Random Table

It contains the different crimetypes and its total count.

6. Summation

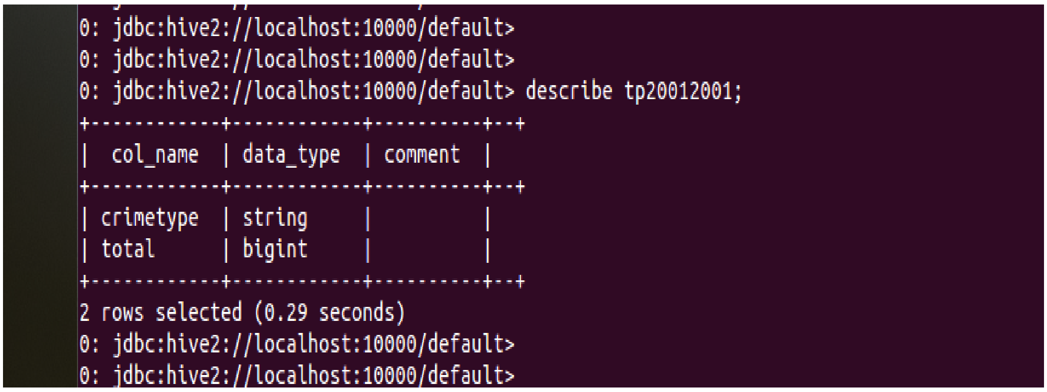


```
5 rows selected (0.313 seconds)
0: jdbc:hive2://localhost:10000/default> describe summation_murder_andhrapradesh;
+-----------+-----------+----------+--+
| col_name  | data_type  | comment  |
+-----------+-----------+----------+--+
| a         | bigint     |          |
| b         | bigint     |          |
| c         | bigint     |          |
| d         | double     |          |
| e         | double     |          |
+-----------+-----------+----------+--+
5 rows selected (0.273 seconds)
0: jdbc:hive2://localhost:10000/default>
0: jdbc:hive2://localhost:10000/default>
```

Fig 1.44 Summation Table

It contains the summation of different crimetypes and different states.

7. Ratio

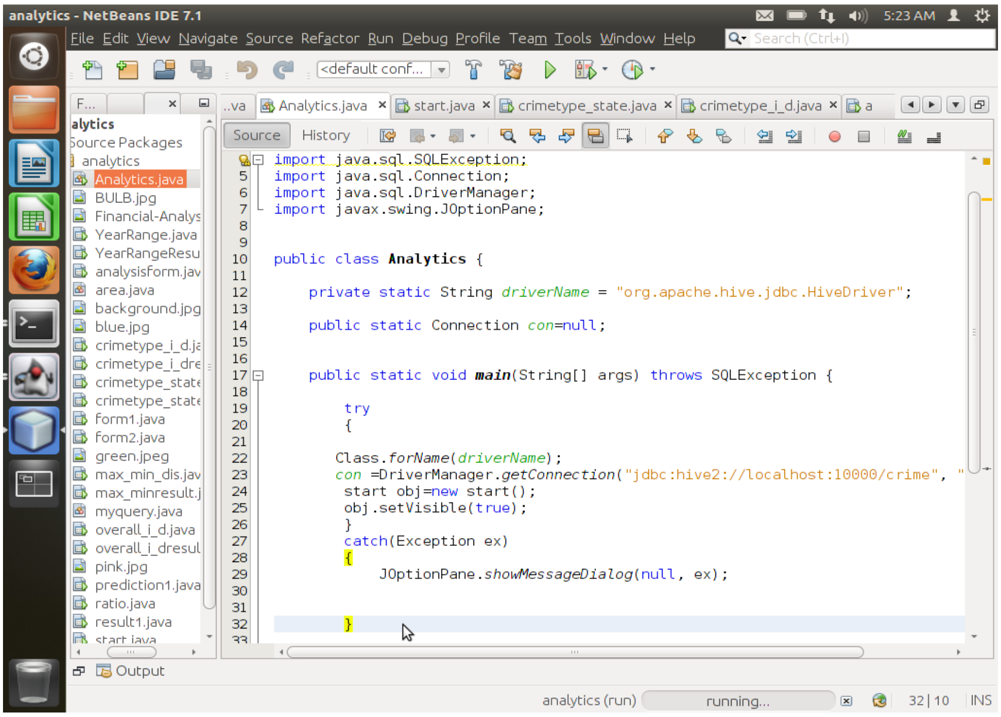


```
0: jdbc:hive2://localhost:10000/default>
0: jdbc:hive2://localhost:10000/default>
0: jdbc:hive2://localhost:10000/default> describe tp20012001;
+-----------+-----------+---------+--+
|  col_name  | data_type  | comment  |
+-----------+-----------+---------+--+
| crimetype  | string     |          |
| total      | bigint     |          |
+-----------+-----------+---------+--+
2 rows selected (0.29 seconds)
0: jdbc:hive2://localhost:10000/default>
0: jdbc:hive2://localhost:10000/default>
```

Fig 1.45 Ratio Table

Different ratios are stored in different tables.

Conection with java



```
import java.sql.SQLException;
import java.sql.Connection;
import java.sql.DriverManager;
import javax.swing.JOptionPane;


public class Analytics {

    private static String driverName = "org.apache.hive.jdbc.HiveDriver";

    public static Connection con=null;


    public static void main(String[] args) throws SQLException {

        try
        {

       Class.forName(driverName);
       con =DriverManager.getConnection("jdbc:hive2://localhost:10000/crime", "
        start obj=new start();
        obj.setVisible(true);
        }
        catch(Exception ex)
        {
            JOptionPane.showMessageDialog(null, ex);


        }
```

Fig 1.46 Connection

Chapter 6 Conclusion and Future Scope

When the data involved in analysis is in terabytes or even huge huge then using Big Data for data analysis is an optimal approach,there is a need of enhanced hardware for running programs involving huge data and their processing.

Scope of involving different statistical techniques for data prediction.

The GUI should be improved to allow users for a better interface and more user-friendly.Acess can be controlled when used in real time application.

The choice of using different frameworks but new frameworks as being added every year should be adopted as a practice to increase the horizons of knowledge and achieve results faster.

BIBLIOGRAPHY AND REFRENCES

1. Herbert Schildt ,"osborne java2",the complete reference, tata mcgraw-hill publishing company limited,2007.
2. Bipin C.Desai, an introduction to big data, Galgotia publications pvt ltd., second edition,1990.
3. Thomas A. Powell, hadoop and map-reduce complete reference, tata mcgraw-hill, publishing company limited, edition 2004.
4. E.P. Wigner, "cookbook mapreduce and hive", international journal of computer studies. vol-53,p.475.sep-2000

WEBSITES REFERRED:
http://www.w3schools.com
http://www.javabegineer.com
http://www.bigdatauniversity.com
http://www.hotrnworkshive.com
http://www.onlinehadoop.com
http://www.hbaselive.com