1.1 INTRODUCTION

This chapter deals with the type of hardware that is used and the software to support the same. Selection and identification of suitable software is also taken into account. Connectivity between various technologies used is established. The designing of interfaces and identification of all functional requirements is carried out. Mainly this system is designed in two modules 1) Desktop application using open source softwares like java and data mining tool Weka for predicting air and water quality data. 2) Web based GIS application for finding its impact on human health and the representation of the same in the form of real time maps.

1.2 HARDWARE AND SOFTWARE SPECIFICATIONS

**Hardware** - Processor (two quad core processors) with 32GB RAM, 5 x146GB RAID hard drive and 4 x10/100mb NIC

**Software** -

1. Operating system : Windows 7
2. Database : postgresQL 9.2/PostGIS 9.2
3. Apache-Tomcat version 6.0
4. GeoServer version 2.1
5. Openlayer version 2.0
6. Java version 1.6
7. Data Mining Tool : Weka 3.7.5
<table>
<thead>
<tr>
<th>Software</th>
<th>Functionality</th>
<th>Use in Project</th>
<th>Why?</th>
</tr>
</thead>
<tbody>
<tr>
<td>PostgreSQL</td>
<td>An open source object-relational database built for Linux</td>
<td>Used to store map files, data files, and user Data.</td>
<td>PostgreSQL is a high-performance ORDBMS. For PostgreSQL databases, pgAdminIII is the management tool which is Open Source. All PostgreSQL features are supported by graphical interface supports and makes admin easy. PgAdminIII is used to write simple SQL queries and edit procedural code. It can be used as a remote data server on another O.S (a client to administer).</td>
</tr>
<tr>
<td>Java</td>
<td>Programming language</td>
<td>Underlies most other code within the Project</td>
<td>Java is secure, reliable and fast &amp; free to download.</td>
</tr>
<tr>
<td>WEKA</td>
<td>A powerful Data Mining tool which is open source.</td>
<td>Used to predict the future trend of environmental data.</td>
<td>It is a bundle of GUI based visualization tools for analyzing data and prediction also it very easily accessible.</td>
</tr>
<tr>
<td>PostGIS</td>
<td>Adds geographic object support to PostgreSQL, essentially allowing geo-referencing of database objects</td>
<td>Allowed database to key and search items based on geographic location</td>
<td>PostGIS is an open source RDBMS which is very powerful and has proved to be reliable. PostGIS is an extension of the O-R database PostgreSQL that supports geographic objects. PostGIS includes a large number of functions for spatial/topology analysis that extends PostgreSQL itself.</td>
</tr>
<tr>
<td>Geo Server</td>
<td>A platform for web publishing of spatial data which is open source and interactive maps Implements a JavaScript-based</td>
<td>The basic framework for publishing map data used by this project</td>
<td>GeoServer is a feature rich, standards compliant which allows users to share and edit geospatial data. It is designed for interoperability It is open source software server and developed in Java, using open standards; it publishes data from any spatial data source. To publish the data on the Web Geoserver needs to configure. To achieve this Postgres back-end, data styling, and integration with browser-based and desktop clients are required.</td>
</tr>
<tr>
<td><strong>JavaScript</strong></td>
<td>An object-oriented programming language with fast dynamic interpreters</td>
<td>An essential core software tool used in all Browsers.</td>
<td>Very easy to learn. It is a scripting, lightweight programming language, and it is a programmable code that can be embedded into HTML pages. This code can be performed by all web browsers. Old or modern.</td>
</tr>
<tr>
<td><strong>Openlayer</strong></td>
<td>Browser-based and desktop client.</td>
<td>Open Layers provides the base for a browser-based mapping client</td>
<td>Using Open Layers it is very easy to put a dynamic map in any web page. To display markers and map tiles loaded from any source can be displayed easily with the help of Open Layer. It is further extended to use geographic information of all types. It is free Open Source JavaScript.</td>
</tr>
<tr>
<td><strong>Apache-Tomcat</strong></td>
<td>An open source implementation of Java Server pages and Java Servlet.</td>
<td>Used for communication between server and client side of web applications</td>
<td>Apache is HTTP Server and generally known as Apache. In the preliminary growth of the World Wide Web Apache (web server software) played an important role. An open community of developers has designed &amp; developed it by under the auspices of the Apache Software Foundation.</td>
</tr>
<tr>
<td><strong>HTML</strong></td>
<td>Hyper Text Markup Language</td>
<td>HTML embeds objects &amp; images and it can be used to create interactive forms.</td>
<td>HTML objects form the architecture of all websites. And it is the markup language used for creating web pages and any information that could be showed on a web browser.</td>
</tr>
<tr>
<td><strong>CSS</strong></td>
<td>Cascading Style Sheets</td>
<td>Colors, fonts, positioning and styling of data has been controlled and taken care by CSS.</td>
<td>The exact content of the website can be searched by spider of the engine when it scans thoroughly. Externally, CSS code is getting stored and it streamlines the XHTML code for faster and smooth loading and working.</td>
</tr>
<tr>
<td><strong>NetBeans</strong></td>
<td>IDE</td>
<td>Used for powerful, multi-language development environment</td>
<td>Freeware Great GUI builder, Great Ant integration, Great and CONSISTENT IDE with Great speed, Great Debugger, Great Profiler</td>
</tr>
</tbody>
</table>
1.3 MAIN FRAMEWORK

This Database is used to store the secondary data. Three different databases are created water quality data, Air quality data and third database created for storing health data. For details refer Annexure VI, database is attached there.

Data mining analysis tool can be used to predict the air and water pollution pattern and gives us the forecasted values and also help us to find its effect on human health, which will further enable local governing body to take preventive measures.

In this research the first part output i.e. SQL SERVER, JAVA AND GIS is in the form of all answers related to all water, air and health related queries, on interactive maps of Navi Mumbai. Then the database is connected to the data mining tool where that data will be preprocessed according to the system needs and then by applying neighborhood analysis we will get the
necessary information about in and around objects or places. Then by applying time series data and regression analysis trend prediction can be done.

1.3.2 **Geo Server:**

Geoserver is standards amenable and feature rich server (open source) which will help us to connect our information to the geospatial web.

1.3.3 **GIS:**

GIS extends spatial data capabilities to give the user results in the form of maps, graphs, 3-D analysis. Etc. GIS will help up giving reports in the form of graph will be easily understandable by the layman also. GIS can also do analyze (buffer analysis, neighborhood analysis) of the data.

1.4 **DESIGN AND FUNCTIONAL REQUIREMENTS**

After collecting the primary and secondary data the readings can be compared, analyzed and studied for future pollution pattern and its effect on us.

In this model the output of first module i.e. SQL SERVER, JAVA AND WEKA is in the form of all answers related to all water, air quality. Then the database is connected to the data mining tool where that data will be preprocessed according to the system needs and then by applying time series data and regression analysis trend prediction can be done. As output of the second module, health related queries are answered on the interactive maps of Navi Mumbai and the impact of air and water pollutants is shown in different colors.
Figure Error! No text of specified style in document.2: Visual Three Tier Architecture/ Main Framework
Figure Error! No text of specified style in document. 3 : Detailed Flow Chart
1.4.1 Implementation details of Module I

The customized package AWMS (Air & Water Monitoring System) is developed using open source data mining tool WEKA for predictive analysis and Java for designing the interface. As Java is reusable and platform independent, the front end is developed in Java. The detailed steps are given below for database creation, data analysis, interface design, and report generation.

1.4.1.1 Database Creation

Backend database is created in PostgreSQL 9.2.

Different tables for different season and different water bodies, different air monitoring stations in the study area are created.

The spatial and non-spatial components are available in database. The spatial component consist of coverage like shape, depth, area of the water body, the map of locality portraying the major localities of Navi Mumbai area.

The non-spatial component consists of air and water parameters (like, DO, COD, Iron, MPN etc.) name, address, sample number, sample date, physical properties of water like color, odour, etc. The non-spatial and spatial data is being correctly associated to produce the essential knowledge.

The data entry of various parameters of water and air, lake names, air stations, zones and date with month and year.

1.4.1.2 Selection of appropriate data mining tool:

To select a suitable platform, a through comparison of data mining tools was done.

A detailed study of all available data mining tools was done. For doing comparative analysis the research started with more than 25 data mining tools, which were narrowed down to 10 with best support for visual analysis. Then as suggested by [Weka] an apt methodology was applied to classify number of computational, functional, usability, and support criteria necessary for this study.
The requirement of the project was to have a system which can be operated on a wide variety of open source platforms. This reduced the list to three: Weka (Weka, University of Waikato), Orange (Nguyen Thai Nghe, et.al.) and Yale (Mohd Fauzi bin et.al.).

A wider range of algorithms are supported by Weka and Yale than Orange, and have superior data preparation tools. Then based on Weka’s support for huge data sets, I have decided to use Weka.

**WEKA (Waikato Environment for Knowledge Analysis, version WEKA 3.7.5)**

Data mining software – Weka is designed in New Zealand by the University of Waikato which applies data mining algorithms. It is developed in JAVA language. An open source system – WEKA is released under General Public License. WEKA is a modern resource for developing machine learning (ML) method and its functions to real world data mining problems. It is a bundle of machine learning algorithms used for doing different data mining jobs.

Directly on a dataset the data mining algorithms are applied. ARFF file format generally used by WEKA. To specify different things, it consists of special tags. [Weka, University of Waikato, New Zealand]

WEKA consist of various standard data mining jobs like preprocessing of data, regression analysis, classification, visualization, clustering and feature selection. Weka is Java based open source data mining tool and it has compatibility with all operating system.
1.4.1.3 Integration of all technologies used

Java and PostGreSQL Database Connectivity Code:

PostgreSQL driver is required for connection between java and PostgreSQL database. The driver has to be placed in the folder where the program resides.

Class.forName("org.postgresql.Driver");

Connection connection = null;

ConnectionDriverManager.getConnection("jdbc:postgresql://127.0.0.1:5432/postgres","postgres","swati");
Java and Weka Connectivity:

To make import files of Weka into java code, weka.jar file has to be placed in jdk folder.

PostgreSQL Database and Weka Connectivity:

To make connection between PostgreSQL database and Weka, the PostgreSQL driver has to be placed in the home directory. Also the Databaseutilis.props file of PostgreSQL has to be placed in the home directory.  jdbc:postgresql://localhost:5432/postgres

To make connection between PostgreSQL database and Weka, the PostgreSQL driver has to be placed in the home directory. Also the Databaseutilis.props file of PostgreSQL has to be placed in the home directory. Jdbc: postgresql://localhost:5432/postgres.

(a) (b) (c)  Weka and PostgreSQL connectivity
These screen shots shows how database is connected to weka and how .ARFF or .CSV file is selected and linked to the database and how database tables are browsed in Weka Environment.

### 1.4.1.4 Data Mining Algorithms Applied

An ARFF (Attribute-Relation File Format) file is an ASCII text file that illustrates a list of occurrences distributing the field set. These files are created to store the data of database.

Data is then converted into ARF format required by WEKA. The current trend and pattern of the data is studied using WEKA.

Based on this data prediction/forecasting for the future is done. Being time series data, algorithms such as linear regression, non linear regression algorithm SMOreg, bagging and rep tree is used for forecasting. The results obtained using SMOreg were found much closer to the actual figures and hence the same was used for final prediction.

### 1.4.1.5 Forms Designing, Coding and Analysis using Java

Interface is designed using Java. The forms are designed and coded according to the user’s requirement using Java.

For initial data analysis Java language’s an open-source framework called JFreeChart is used. Data is graphically represented (pie chart, bar chart, stacked bar chart etc.) using JFreeChart. It helps to create variety of interactive charts like

- Merged Charts
- X-Y charts (line, spline and scatter).
- Pie charts
- Gantt charts
- Bar charts

Import files used for use of JFreeChart are as follows:

```java
import org.jfree.chart.*;
```

**Code snippet:**

```java
JFreeChart chart = ChartFactory.createBarChart("Bar Chart of Nerul Sec-20","Year", "Range", dataset, PlotOrientation.VERTICAL, true,true, false);
```
DefaultPieDataset dataset = new DefaultPieDataset();

JFreeChart chartn5 = ChartFactory.createPieChart("Pie Chart of Nerul Lake ", // chart title
dataset, // data
true, // include legend
true,
false);

At last zone-wise and parameter wise reports are generated using forms in Java. Based on the standards (Refer Appendix III & IV for Air & Water) rule based system is designed.

1.4.2 Implementation details of Module II:
Using PostGIS software shape files for each water body and each air monitoring centre in the study area are created. Connectivity between PostGreSQL and PostGIS is established. By using different PostGIS tools and modules spatial analysis is done. Different spatial queries can be answered such as which area is more polluted? Why? What is the concentration of one particular parameter in given area? How many people are affected by air borne diseases?

**Steps to create database, Interface Designing and Analysis for Module II**

For finding out an air pollution impact on human health, data of both Turbhe Industrial and Vashi Residential zone is used for comparison.

Boundaries for both the above zones are identified on Google Earth and KML (Keyhole Mark-up Language) files are created.

The location of the UHP’s for both the zones are identified on Google Earth.

Data from both the zones is pre-processed using ArcGis also health data of same area is pre-processed and shape files are created.

Using GeoServer these shape files are stored in PostGis database (an extension of PostGreSQL)

The web based interface is designed using HTML, Open Layers & JavaScript.

WMS technology is used to present static map data on web browser.

The detailed steps are given below.

**1.4.2.1 Spatial Data Pre-processing**

1) Digitize the polygon in Google Earth as per the boundary of a zone marked

2) Save the digitized polygon as Google earth KML/KMZ format

3) Use Spatial data import for importing the Google Earth KML to ArcGIS Shape file format.

4) Edit the Imported shape file in order to check and correct the data imported

5) Use Spatial data import function of GeoServer to Import the data into GeoServer

6) The data will be imported into GeoServer

7) Perform the similar steps to import all the respective approx 20 Layers into Geoserver.

PostGreSql Data Storage Structure.
1.4.2.2 Loading a shape file

Postgres supports only shape file to get uploaded into Database. Each layer/shape file goes into Postgres database as a table, and all its attributes are represented as its columns in the respective table. The name of the table is defined as the layer name and the name of columns are also represented as the attribute name. It also creates a special reference to hold the projection references.

**Example**: Create a Demogis database and upload a shape file to it.

Let the name of the shape file is “Zone1.shp”. It has attribute say: “id, polygid, area_sqkm, perimtr_km, xcentroid, ycentroid, geom”. So the table structure would be:

1.4.2.3 Database Preparation

PostgreSQL 9.2 with its spatial extension PostGIS is widely used as database in backend for GIS application. It’s an Open Source tool which has functionality of storing attribute data along with their spatial references.

After Installing Postgres9.2 Database server, open “PgAdminIII” tool to create new Database.

**New Database Configuration**: Name: DemoGISUsername: postgres
Roles and privileges also can be provided into the database.

Then the new Extensions of the database to make it capable of storing spatial references in the database are provided.

1. Postgis.
2. Postgis_topology.
After creating database, open postGIS2.0 Shape file Import/Export tool.

The connection parameter to connect the tool with your Database (DemoGIS) is provided.

After connection successfully established, then the pre-prepared Layers shape files are provided to import to the database.
After successfully import, the Layers table in the DemoGIS database was created.
1.4.2.4 Geoserver Data Preparation

Start Geoserver.

The username and password is provided which is used during the installation of Geoserver to enter to the admin page.
A workspace name and uri is provided on the Workspace link available on the left hand side and then saved it.

Then go to link Stores → New Store → Vector Data Sources → PostGIS – PostGIS Database.
PostGIS configuration details are provided here.

Then save it.

It will then ask for publishing the layer.
After clicking on Publish link as shown in below Fig 1, it will go to Edit Layer page where the Layer’s Coordinate Reference Systems data and calculate the “Native bound“ and “Lat/Log bound” is filled as shown in Fig 2.
After going to Publish Tab, the WMS Feature Settings are filled.

Then save it. The Layer is ready to display on Client machine.
1.4.2.5 Deployment of application on Application Server

Functions of an application server:

- Business logic is exposed to client applications by a server by means of various protocols together with HTTP.
- A range of applications are delivered to another device and also allows all the peers from the network to execute the same application.
• Web applications are served by application server. As application server contains web server internally, its purpose is to serve web based applications and enterprise based applications. Example: Sun Java Application server, Geronimo, weblogic server, Apache Tomcat.

Apache Tomcat

The software model which gives an atmosphere where all functions execute is termed as an application server. To expand web servers and to support for dynamic content, web application servers are developed. First the application server software embedded in to the web server software. Then it dynamically captures any user’s requests for active content. The web server provides web pages and graphic files. The application server also creates dynamic content by combining data with templates, running programs, and by retrieving databases.

Open source software - Apache Tomcat is implemented in Java Servlet and Java Server Pages technologies, the specifications of which are developed under the process of Java Community. In an open and participatory environment Apache Tomcat is developed. Further, it is released under the Apache License version 2. Apache Tomcat can be collaborated with the best developers from the world. You all are invited to participate in this open development project.

Steps for Deployment on Apache Tomcat:

• Directory hierarchy which is unpacked needs to be copied into a subdirectory in directory $CATALINA_HOME/webapps/. Based on the subdirectory name chosen by us, Tomcat assigns a context path to the application

• As it is the quickest and easiest development approach, this technique is used to build build.xml file. Then Tomcat is restarted after installing or updating the application.

• Copy an archive file of web application to directory $CATALINA_HOME/ webapps/. On starting the web application, Tomcat automatically expands archive file into its unpacked form. Then it executes the application.

• When the applications are provided by a third party vendor or even by the internal development staff, into an existing Tomcat installation, such types of approaches can be used.
Web applications are deploy and un-deploy by Tomcat 4 "Manager" tab. On executing Tomcat server, default context path /manager deployed by Tomcat 4 allows us to deploy & un-deploy web applications or we can deploy another war file.

1.5 CONCLUSION

A modular approach has been followed in software design. A variety of tasks carried out by the main module of the customized package are spatial and non-spatial database organization, modifying spatial features, creating network, finding nearest neighbors, finding objects within the said proximity and asking of related information. A special care has been taken while developing the said modules like not to refer jargons like points, line and polygon coverage. Any Government employees will communicate to the system without knowing the underlying GIS terminologies also.

Various technologies like Java and Weka etc. are used to implement the prediction model. All the available technologies such as PostGIS, PostGreSQL, Open Layers, HTML, and Java Script are studied and are tested and verified to develop a secure, economic, user friendly model. Different servers like Geoserver and Tomcat Apache are studied and also different front ends like Open layers, HTML, Java Script are studied to present the data in the best possible way. After pre-processing of spatial data using ArcGIS, shape files are created and those are linked with PostGIS and then it is published on Geoserver. Through the application server Tomcat Apache the spatial data is presented on the browser using the frontends listed above. WMS technology is used to fetch the data from the PostGreSQL database and its spatial extension PostGIS. Execution of all the technologies is done in line with the planning done in the previous chapter.