INDEX

1. CERTIFICATE

This is to certify that the Seminar & Progress Report/ Dissertation entitled Data Repository for Human Development Index done by Mr. Jatin Verma, Roll No. 05816401514 is an authentic work carried out by him at under my guidance. The matter embodied in this Seminar & Progress Report/ Dissertation has not been submitted earlier for the award of any degree or diploma to the best of my knowledge and belief.

Date:

Name of the Supervisor: Prof C. S. Rai

2. ACKNOWLEDGEMENT

I am using this opportunity to express my gratitude to everyone who supported me throughout the course of this project. I am thankful for their aspiring guidance, invaluably constructive criticism and friendly advice during the project work. I am sincerely grateful to them for sharing their truthful and illuminating views on a number of issues related to the project.

I express my warm thanks to Prof C. S. Rai for his support and guidance.

I would also like to thank all the people who provided me with the facilities being required and conductive conditions for my project.

Thank you,

Jatin Verma

3. ABSTRACT

Data Repository for Human Development Index has been designed and developed as part of data application project. The system allows a provider to select catagories in which it can provide data and then a consumer can subscribe to whatever category it wants. The system generates and sends a notification to consumer whenever a provider updates data that consumer has subscribed to. A provider can provide data in multiple catagories and a subscriber can also subscribe to data in multiple categories. It is the maintainer's job to make sure a provider updates data of the same category it has registered itself with. The system also provides the consumer with the facility of viewing the data in graphical, tabular or pie chart form.

4. INTRODUCTION

4.1 Aim

To provide easy and fast access to large sized data like literacy rates, neonatal birth and mortality rate(data generated from nationwide census) etc. Fast and easy access to such data is paramount to study and analyze a country's human development indices. This type of data is used by world organizations like United Nations and World Health Organization for deciding a country's human development index which is calculated on the basis of a country's life expectancy for health, expected years of schooling, mean of years of schooling for education and Gross National Income per capita for standard of living by United Nations Development Program. Every year UNDP ranks countries based on the HDI report released in their annual report. HDI is one of the best tools to keep track of the level of development of a country, as it combines all major social and economic indicators that are responsible for economic development.
Access to this data is paramount and is made easily possible through Data Repository for Human Development Indicators.

4.2 Objective

Today when tons and tons of data is generated every day the storage and generation of that data is not really much of an issue anymore, technological advancements have now made it possible to store multitudes of data on tiny chips, but the instant access to meaningful data is. The objective of this project is to bring together surveying agencies and data consumers together in order to provide fast, easy and convenient access to real time data generated through censuses conducted on a large geographical area and provide tools to analyze that data and view that data in graphical form for producing a comparison based report.

4.3 Scope

The scopes of the project are listed below:

- A consumer or provider both can register themselves with different types of data they want to access or provide.
- The system provides tools(D3 js) to view that data in tabular or graphical form for better reading and conclusion drawing.
- The maintainer keeps check on the activity of consumers and providers and also keeps the data organized.

4.4 Motivation

Large amount of data is generated everyday, the real value of data is derived when it's used. Furthermore, the speed and efficiency in accessing and obtaining that data is even more important as it significantly impacts the value of the data and one of the ways to use it crucially is Data Repository System. The generation or storage of data is not much of concern anymore but the access to that data, much specifically access to some particular required data is, Data Repository System brings together the data generating agencies and data consumers and helps the data consumers churn out meaningful conclusion from that data.

5. PROJECT ANALYSIS

5.1 Project Description

Data Repository for Human Development Indicators is a web based application which provides platform for data sharing, a user can register itself and gain access to data generated by a producer(typically an agency conducting surveys or censuses over a large geographical area). This project is an integral feature of effective and efficient research. In this project access control protocols are also maintained i.e. users are unaware of the kind of data subscribed to by the other users.
This system broadly provides the following facilities to it's users:

- A user can register itself as a data provider or a data consumer or even both and then can provide/access the kind of data it has registered itself to.
- A consumer can view the data in graphical or tabular forms using the visual tools provided by the system.

Users are broadly classified into:

1. Provider
2. Maintainer
3. Consumer

A user can access the system as a data provider or as a data consumer at a time, after logging into the system the user will be prompted to select the kind of data he/she will be providing or accessing, for example if WHO(World Health Organization) logs into the system as a data consumer then after logging in they will be prompted to select the kind of data they are looking for and the region from where they require data, let's suppose they require data from India on infant mortality rates from all the states occurred in the last 10 years, then they will choose region as India, indicator

as health and subgroup as the mortality rate. Now the WHO is a subscriber to the data provided by the Indian government and as soon as some new data is published by the provider (which in this example is Indian government) the system will send a notification to the consumer i.e. WHO and they can then directly access that data stored over a distributed database, the system itself doesn't provides any database facility but only the means to share it over a single platform from multiple different sources and then to analyze it better conclusion drawing.

The system has been designed using HTML and CSS using JAVASCRIPT for scripting and is implemented using node js framework, the system also uses express js framework for handling the communication i.e. sending and receiving of notifications between the provider and consumers. D3 js framework has also been for providing the graphical tools for data analysis.

At a time when multitudes of data are generated everyday, storage or generation of that data is not really much of a concern anymore but fast, easy, systematic and convenient access to that data is paramount, Data Repository for Human Development Indicators does just that by providing a unified platform for data sharing.

6. KEY NOTES

6.1 Node JS

Node.js is an open source, cross-platform runtime environment for developing server-side and networking applications. Node.js applications are written in JavaScript, and can be run within the Node.js runtime on OS X, Microsoft Windows, and Linux. Node.js is a platform built on Chrome's JavaScript runtime for easily building fast and scalable network applications. Node.js uses an event-driven, non-blocking I/O model that makes it lightweight and efficient, perfect for data-intensive real-time applications that run across distributed devices.
Following are some key features of node js:

- Node.js is an open source server framework
- Node.js is free
- Node.js runs on various platforms (Windows, Linux, Unix, Mac OS X, etc.)
- Node.js uses JavaScript on the server

This project also uses the node js framework.

6.2 D3 JS

D3 allows you to bind arbitrary data to a Document Object Model (DOM), and then apply data-driven transformations to the document. For example, you can use D3 to generate an HTML table from an array of numbers. Or, use the same data to create an interactive SVG bar chart with smooth transitions and interaction.
D3.js (or just D3 for Data-Driven Documents) is a JavaScript library for producing dynamic, interactive data visualizations in web browsers. It makes use of the widely implemented SVG, HTML5, and CSS standards. It is the

successor to the earlier Protovis framework.[In contrast to many other libraries, D3.js allows great control over the final visual result. D3 js has been heavily used in this project to provide tools for analyzing the data in graphical and tabular format.

6.3 Express JS

ExpressJS is a web application framework that provides you with a simple API to build websites, web apps and back ends. Express provides a minimal interface to build our applications. It provides us the tools that are required to build our app. It is flexible as there are numerous modules available on npm, which can be directly plugged into Express. Express Js has been used in this project for performing back end operations like sending notifications to a subscriber whenever a provider uploads relevant data.

6.4 Javascript

Javascript is a dynamic computer programming language. It is lightweight and most commonly used as a part of web pages, whose implementations allow client-side script to interact with the user and make dynamic pages, it is a programming language that adds interactivity to your website (for example: games, responses when buttons are pressed or data entered in forms, dynamic styling, animation). It is an interpreted programming language with object-oriented capabilities.
The JavaScript client-side mechanism provides many advantages over traditional CGI server-side scripts. For example, you might use JavaScript to check if the user has entered a valid e-mail address in a form field.
Javascript has been used in this project as the scripting language.

6.5 HTML

HTML stands for Hypertext Markup Language, and it is the most widely used language to write Web Pages. Every webpage we look at is written in HTML. It can be thought of as the skeleton that gives every webpage structure. It is the standard markup language for creating web pages and web applications. Hypertext refers to the way in which Web pages (HTML documents) are linked together. Thus, the link available on a webpage is called Hypertext.
As its name suggests, HTML is a Markup Language which means you use HTML to simply "mark-up" a text document with tags that tell a Web browser how to structure it to display. HTML has also been used to write web pages in this project.

6.6 CSS

Cascading Style Sheets, fondly referred to as CSS, is a simple design language intended to simplify the process of making web pages presentable. CSS handles the look and feel part of a web page. Using CSS, you can control the color of the text, the style of fonts, the spacing between paragraphs, how columns are sized and laid out, what background images or colors are used, layout designs, variations in display for different devices and screen sizes as well as a variety of other effects. CSS has been used in designing of web pages in this project. CSS stands for Cascading Style Sheets. CSS describes how HTML elements are to be displayed on screen, paper, or in other media.CSS saves a lot of work. It can control the layout of multiple web pages all at once. External style sheets are stored in CSS files

7. REQUIREMENT SPECIFICATION

To run this project on various platform hardware and software requirements are detailed below:

7.1 Hardware Description

Memory: 30 GB
RAM: 512 Mb
Processor: Dual core

7.2 Software Description

Technologies: HTML, CSS, JavaScript, Node Js, D3 Js, Express Js
Language: Javascript, HTML
IDE: Sublime text, Notepad++, Eclipse

8. METHODOLOGY

The project was made in following four phases:

Phase 1 - Developed Data Provider Interface to input the data in the system.

Phase 2 - Login of Data provider entity to login into the Data Provider system.

Phase 3 - Development of data maintainer's side.

Phase 4 - Developed Data Consumer Interface to register the request and to analyze the data.

Phase 5- Testing

In phase1 data provider interface was developed, functionalities like selection of category, region, unit, subgroup and indicators were implemented for data input by the provider.
In phase2 login facility for the data provider was developed.
In phase3 data maintainer's side was developed, now a data maintainer can keep the system sorted, whether or not the posted data matches it's category can be checked by the maintainer.
In phase4 data consumers interface was developed, functionalities like subscribing to different indicators, regions, subgroups, units etc were implemented.
Finally in phase4 the system was locally hosted and tested whether or not working properly.

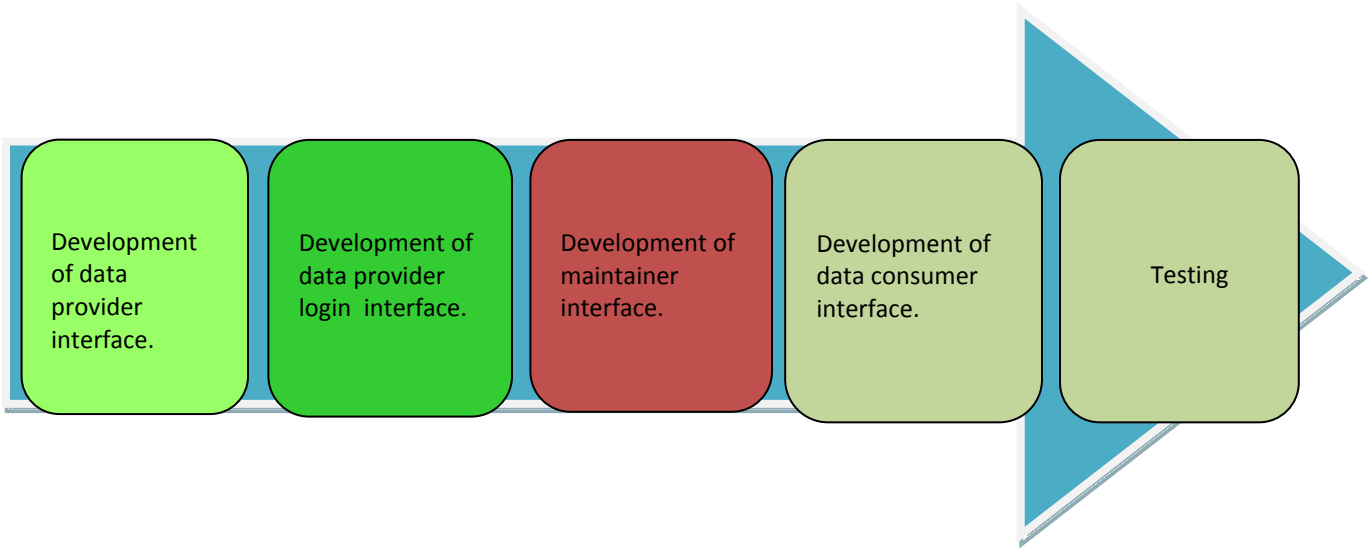

Figure0: Different phases of project development.

9. USE CASE DIAGRAM

A use case diagram is used to model functionalities and requirements of a system. A use case diagram shows the interaction of various users with the system, it has actors and the actions they can perform with the system. An actor is any person or external system that interacts with the system to achieve a user goal. The Use Case Diagram is a visualization of a use case. The Use Case Diagram for the actions that the Users (provider, maintainer and consumer) can perform in the system. Usually referred to as behavior diagrams they are used to describe a set of actions that some system or systems should or can perform in collaboration with one or more external users of the system. Each use case should provide some observable and valuable result to the actors or other stakeholders of the.
The use case diagrams shown here are for the two most common users of the system namely provider and data consumers.

Figure1 on the following page shows the use case diagram for a data provider with provider as an actor and it's various functionalities listed in the diagram.
Figure2 on the page after that shows the use case diagram for a data consumer and lists the functionalities provided by the system to a data consumer. The whole interface is designed simplistic in nature and self explanatory for a better user experience.