

# Locating Points of Interest for GIS Maps Using Apache Solr

Vivek P. Talekar, Sajidha S.A.

**Abstract** — The geographical information system (GIS) is an approach to encapsulate, load, handle or control analysis, manage, and present all kinds of geographical Big Data. This map making contains recording the images, saving the data, inspecting, and flourish information identified with positions on Earth's surface. This will contain unique picture mosaics with separated vector and customer provided ascribe information to make single, information-rich pictures for GIS Mapping ventures. A geomap is a map of a country, continent, or region map, with colors and values assigned to specific regions. The Points of Interest (POIs) are businesses and landmarks important in the Find, Guide, and Display functionality of vehicle and pedestrian navigation, Internet mapping, and Enterprise mapping solutions. POI will represent the activity at a specification location. The division of the Point of Interest into its subclasses is driven by specific point of interest behavior in POI relationships and POI attribution. As the data required for map making process, is in very huge volume. Finding of particular information about the point of interest is very hectic process. Initially this process has been taken huge time. Fetching the data which comes in fallout report should be faster. The main objective of the paper is to fetching required information from terabyte of data in millisecond using apache solr technology. This apache solr is having different functionality, indexing method, which helps out fetch or searing faster.

**Index Terms** — GIS, map, Position, GeoMap, Point of Interest (POI), Apache Solr, Sharding the data, Indexing.

## 1. INTRODUCTION

*“Suppose you are having a young boy, and you asked to him to travel into a small house, and in short time period collect more, you can do this easily as you know everything about that house; but he can't. Let the boy carry with him some map or paper which describing the house, where he travelled, which will be a key to his search...”*

This illustration is solid counsel and sets the stage well to think on different applications and distinctive research open doors for land data frameworks (GIS) in business maps. GIS are quickly expanding being utilized for multipurpose in light of the fact that they are capable devices that can be utilized to bring the great data that is kept in the information that depicts area (e.g., scope and longitude, addresses, postal codes, provinces,). Land data is a choice instrument which enables a client to bring aggregate maps.

The main feature of the system is that it distinguishes this from other object information. This provides the greater help for the user to get actual geographic locations. As we know that most of maps contain lots of useful information like hotel, school, colleges, offices, restaurants and malls etc. Map reading and map drawing are important skills to learn in GIS.

## 2. PROBLEM STATEMENT

As mostly using traditional way, to search the data or fetch the data, which is important for business, will take more time for performing the operations. This is because of traditional way like text search, or word search may take much more time. This happened because of that used old method. For overcoming that

problem, we introduced faster search engine using Apache Solr technology. Solr is used for to extract useful information from database.

This will take very little time for fetching the data from even terabyte of data. This technique is having more advantages because of its Indexed searching. To make the searching and fetching faster than the traditional way, this method gives good results.

## 3. GEOGRAPHICAL INFORMATION SYSTEM (MAPPING)

It should be noted that the GIS is much more useful tool for creation of any kind of map. Or this GIS can be used to show the graphical presentation. For example, different kind of sheets packages now contains GIS usability that permits users displaying the new kind of map displays. Although this kind of ability is useful for making graphical presentation and similar kind of displays, this represent just some kind of abilities that can woks fully on GIS possessing.

The geocoding process is to describe the method of connecting attribute information with the actual coordinates of the location on a map. For which, if somebody required to place the locations of all stores on a map from their way, they could geocode all stores address with coordinates on the map to define POI that is used to locate all stores nearby each store. The geocoding is process of creating the data fields in the attribute datasets for the longitude (Coordinate X) and latitude (Coordinate Y) of each point of interest. The emerging link generates a GIS database which is a great merging of the two different data sets, the attribute data and the map data. Because of this significant dataset, it will provide more accurate data for number of applications. This dataset is having

huge amount of data which contains the all information about the POI or Node.

Another feature of a GIS application is the display of geographic information. In other words, maps can be represented with the help of GIS. Adding to that, more than one different set of data can be represented.

#### 4. POINT OF INTEREST APPLICATION

Because of accurate POI, it is very easy to reach at desired location. This will help to user of map, to finding the more accurate location. Accuracy of POI is important concern in the POI building. This will also be beneficial for a new user for finding the exact location. Map gives exact shortest distance in between POI. Accurate POI will give the all detail information about that location. Blindness affects nearly 45 million people worldwide. Because of rapid population broadening, this number is expected to more than double by the year 2020. Navigation in indoor conditions is highly demanding.

#### 5. RELATED WORK

GIS is a more effective decision support tool because it enables their clients to oversee property information, as well as to catch, oversee, and fuse space information in their analysis. Because of these abilities and also other industrial facts, the GIS business had huge growth in the past few years. Despite the fact that a government administration still speaks to the biggest section of the GIS-client group, quite a bit of this current development can be credited to across the board dissemination of GIS into the business group. [1]

The search innovation for the source, they are using The k-shortest path method [2] to calculate the minimum or shortest path query. Here, the variable k remains for the quantity of transit point that are contained in the inquiry Result. transit point focuses are only the Point of Interest (POI). For finding the minimum distance in between the two nodes can be find by using this method. [2]

Points of Interest (POIs) and Regions of Interest (ROIs) are two kind image features rapidly used in many computer graphics applications. Detection of the features has received further studies. [3]

Statistics information like addition, average, and distribution of points of interests (PoIs), e.g., hotel, restaurant on guide administrations, for example, Google maps and Foursquare give significant data to applications, for example, showcasing basic leadership like decision making. For example, the learning of the POI rating dissemination empowers us to assess a specific Pol's relative administration quality positioning.

In addition, a restaurant start-up can surmise nourishment inclinations of individuals in a geographic zone by looking at the notoriety of restaurant PoIs serving different cuisines within the area of interest [4]. In the meantime, it can likewise assess its market measure in light of POI total insights, for example, the quantity of foursquare clients checked in PoIs inside the region. Also, a lodging start-up can use hotel PoIs properties, for example, appraisals and surveys to comprehend its market and rivals.

In the technique, a record is produced progressively utilizing an arrangement of competitor travel focuses and an arrangement of applicant goal focuses. The framework assesses shorted way rapidly utilizing the Index. Any client can change beginning stage. A similar file is utilized for a similar arrangement of hopeful travel focuses and an arrangement of applicant goal focuses.

Social networking like Facebook, YouTube or LinkedIn contains numerous data like preferences, aversions, telephone number, and Email id. The author utilized the solr innovation to construct a web index utilizing apache solr to remove helpful data from Facebook. Apache solr is utilized to get more imperative data. [6].

The creator utilizes the Apache Hadoop and Apache Storm to parallelize the procedure, with the goal that it can be quick than some time recently. The current ways to deal with recover of information advancement of various substance based ordering strategies that permit to proficiently get the information in sight and sound dataset, for example, pictures.

In any case, for that move, record capable information ought to be removed first from accessible dataset. Indexing is necessary move can be very time to speed the process up. Apache Hadoop and Apache Storm are used for the making system more parallelized. [7]

The author uses the apache solr for carrying out their different activities, tagging tools which allow activities to Bulk/group tag or untag big size data sets of objects in temporary work sessions, for experiencing the real time. Their actions before making the changes visible to end-users. Apache solr is used for completely functional annotation tagging environment over full-text index Apache Solr, [8]

#### 6. APACHE SOLR TECHNOLOGY

Apache Solr is open source Search Server / web application, an open source enterprise search server which will provide huge-scale scalability, deployed on top of application server like Tomcat. Apache Solr can accomplish quick hunt reactions in light of the fact that, rather than seeking the content straightforwardly, it looks an index instead.

Solr is powered by Lucene which able to powerful matching abilities like phrases, joins, an open source enterprise search server which will provide large-scale scalability, also it can be deployed on top of application server like Tomcat. Solr is controlled by Lucene which empowers intense coordinating capacities like expressions, special cases, joins, bunching and significantly more crosswise over different information sorts.

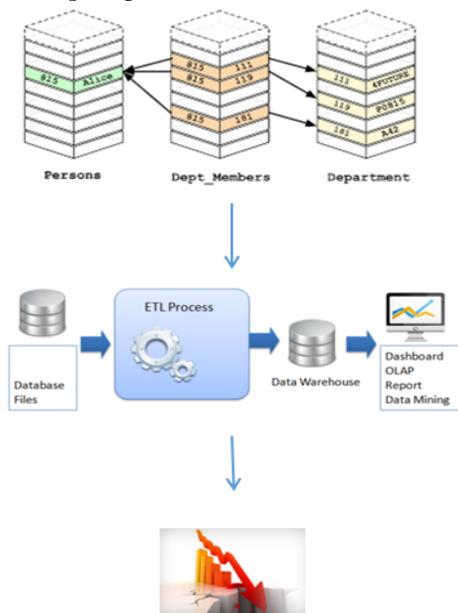
The apache solr is having support to Query language for structured and textual search. The main top Hadoop users from Cloudera, Hortonworks and MapReduce all having bond with Solr as the search-fetch engine as their Big data Projects. Solr is also work as end point in various big data processing projects.

Solr will work with HTTP Rest APIs with different xml and json formats and solr will integrate with any programming language which supporting these standards. For more comfortable of java, Python, Ruby, C# client

**7. REGULAR APPROACH AND NEW APPROACH.**

Traditional Approach to getting the report is more time consuming. For report generation, it will take almost two to three days.

Traditional way to fetch the data required lot of time for the process. This will take more time for processing the data and report generation.



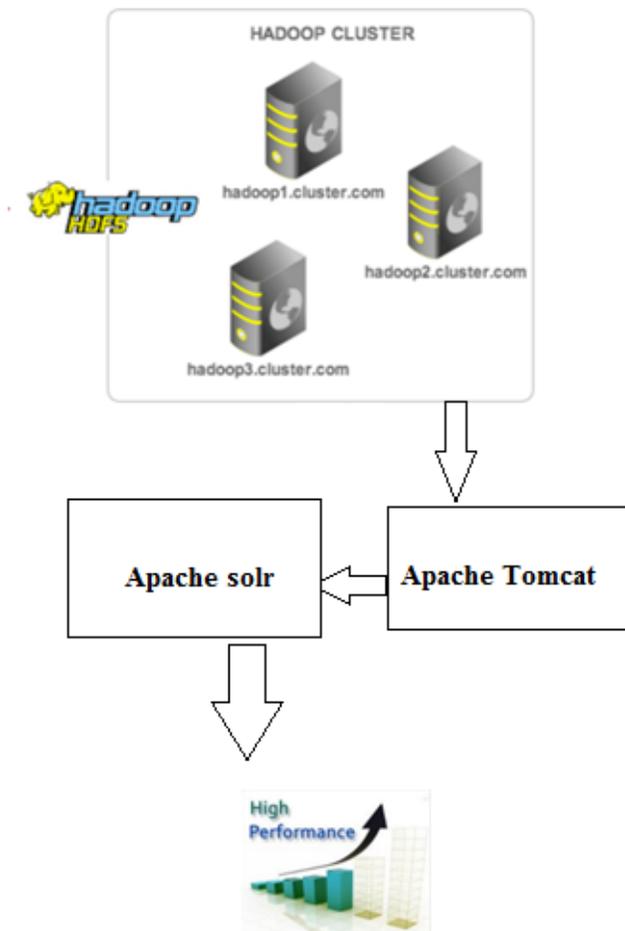
**Fig 1 Traditional approach of report generation**

For this project on first day, the production engineer will start the task of PPP pipelining process. This PPP pipeline is the main task of this project. After completion of PPP pipeline process, the actual report

generation will start. This contains different kind intermediate steps, which perform the different tasks for allotted to them.

In above method, the data is loaded taken from supplier or number of suppliers. The data may be in different format. This data loaded into pipeline for further process. After that, database get ready with all given data.

In the Extract-Transmit-Load process, all data will process. First data will be extracted from PPP pipeline process.



**Fig 2 Our Approach using Apache Solr**

Transmission of data from PPP pipeline into oracle database will take too much time. As the data is transferred from local machine to oracle machine. This will take more time for data processing.

Loading of the data from local machine to oracle machine will take more time for uploading the data. After that report generation will start. For time reduce purpose we introduce a new method using the apache

solr technology and apache tomcat. Apache solr will work on indexing method. Because of this indexing search, required time is very less for report generation as compared to traditional way.

In new approach, the input from supplier added into PPP pipeline process. The data is loaded and it will start executing all the process. This will take the sometime for report generation.

After the report generation, the data is loaded into apache solr with the help of scripting language like python. Because of apache solr, time taken for report generation is very less. This will work on indexing method instead of other traditional way.

As this is using the indexing method, fetching of data is very simple and in faster manner.

Apache Tomcat is needed here for server. This apache tomcat will act as server in your machine. The apache tomcat will be used for deploying the java servelets and JSP pages.

## 8. APACHE SOLR ARCHITECTURE

Apache Solr contains different number of shards. These shards are nothing but data divided into small parts.

For this row data is stored in HDFS. Row data is nothing different than the data on which we have to process.

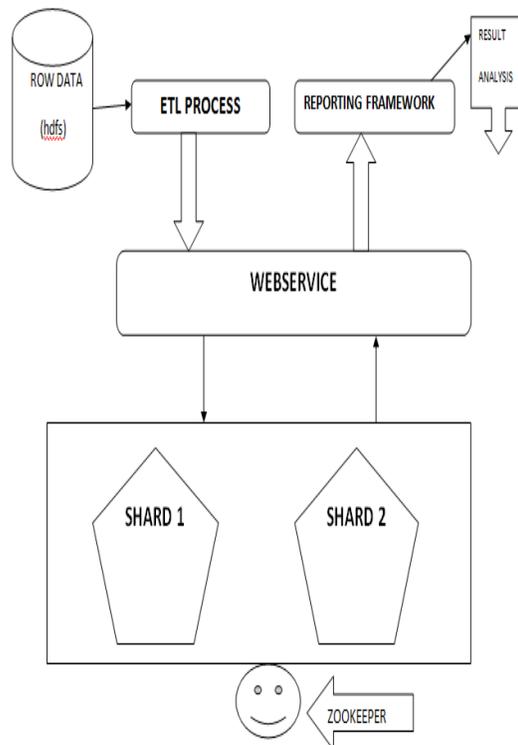
This kind of data is in very large volume. Mostly it contains the terabyte of data. This data cannot handle by traditional approaches. Apart from this section, the important task is fetch the required data from the terabyte of data.

Fetching the data from large volume is very important task. The old approach is very time consuming and also a hectic process.

The SOLR method will take very less time to fetch the data as it uses the indexed searching. Indexed searching is one of the fastest search available now.

The row data is undergo the ETL (extract-transform-load) process. Extracting the required data from terabyte of row data is essential work. Web service comes into action after the ETL process done.

Then data is merged into different kind of shards. These shards are able to keep the data into smaller version. As solr is using index method, that index is divided into chunks, known as shards. It is just a logical partition of data.



**Fig 3 Solr architecture**

In traditional way, dividing the index into shard is done by manually. And also, it is not supporting to distributed indexing. Load balancing is also major factor.

Using apache solr these problems are solved explicitly. Apache solr is supporting to distributed indexing and failover. Zookeeper plays an important role while load balancing and failure of process.

The important factor is that, it is not master or slave process. Each and every shard, will contains the replica of its data. This will be helpful while load balancing or failure happened.

The zookeeper automatically selects the leader for their process. If the leader get fails, one of the replica get selected as new leader of process.

## 9. EXPERIMENTAL RESULTS

The apache Solr gives the results as expected in lesser time. As Solr is working on indexed searching, will gives the faster result. With the help of apache solr and apache Tomcat as an server, this makes system faster than other traditional ways of searching techniques. The finding of the solr is faster, indexing plays important role in this finding.

Before using this apache solr method, the required result would take more time for fetching the data from

