

Database-driven Spatial Data Auditing using QGIS against LDAP Authentication and Authorization

Santosh Gaikwad^{1*}, Maulik Bhagat², Arjan Odedra³, Rahul Kanani⁴, Aditya Saraswat⁵

^{1,2,3,4,5}Nascent Info technologies Pvt. Ltd., AF-01, Shapath-4, S.G. Highway, Ahmedabad-380015, India

*Corresponding author: santosh.gaikwad@nascentinfo.com

DOI: <https://doi.org/10.26438/ijcse/v10i12.2833> | Available online at: www.ijcseonline.org

Received: 28/Oct/2022, Accepted: 01/Dec/2022, Published: 31/Dec/2022

Abstract—Efficient spatial database management system is now considered to be a primary component of any GIS architecture or Spatial Data Infrastructure (SDI) which acts as a central repository that disseminates data across web and desktop clients. In such scenarios, a common requirement for production databases is the ability to track history and authenticity of the data generated. Using the database and the trigger system as regular practice, it's possible to add history tracking, however, there exists a challenge of using desktop GIS tools for editing and validation of the normalized data residing in spatial databases, intacting data integrity with privileged and authenticated users. Considering the facts and importance of auditing, jurisdiction based user access management and limitation of editing spatial data in a web browser, the present study is an attempt to develop an Open Source solution based on authorization and authentication against LDAP, QGIS, PostgreSQL/PostGIS, and Multicorn FDW software.

Keywords— LDAP, Web- Editing, QGIS, Open Source

I. INTRODUCTION

In today's world, data proves to be an integral part for decision making processes. These decisions prove to be more effective and tend to provide better insights into the business governance, when backed up with the verified datasets. Approximately, 80% of this available data has spatial components in it, which can be visualized and analysed through the Geographic Information Systems (GIS) for effective and strategic decision-making practices. Spatial database being an integral part of Spatial Data Infrastructure (SDI) or any GIS architecture, acts as a central repository that disseminates and shares data across web and desktop clients. Besides that, centralized database storage leads to the advantages such as data security, data sharing, disaster recovery, reduced maintenance cost. Thus, it is really important to maintain the datasets and the changes that have been done on a centralized system. Data normalization is an important step while designing the structure in order to reduce data redundancy and improve data integrity.

PROBLEM STATEMENT: The GIS projects, worldwide, often require a large amount of spatial data collection and production which are time consuming and quite difficult to manage. Thus, maintaining and sharing spatial data is essential to avoid unnecessary costs from duplicate production which utilizes unnecessary space and creates duplication of data. Such a data shared environment is necessary to maintain the quality of data, where multiple users work on a specific task and try to discover and edit errors within the data. In such scenarios, there is a common requirement for production databases which is the ability to track history and authenticity of the

data generated. It is also helpful to know the details pertaining to , how has the data changed between two dates, who made the changes, and where did they occur? The traditional approach was to make edits locally and then update it on the production database. But with the advancement of robust technologies such as web, smart phones and cloud computing, there are ways available for online editing and validation but it lacks powerful topological error correction tools in comparison with desktop GIS clients.

MOTIVATION: Considering the importance of auditing, jurisdiction based user access of data, management of data and limitation of editing spatial data in a web browser, the present study focuses on developing a solution based on an open source technology. The solution provides a two step workflow for editing and validation of spatial and non-spatial data that is stored in a centralized PostgreSQL database with PostGIS capabilities. This is done using desktop through QGIS with authorization and authentication against LDAP.

The centralized database storage leads to the advantages such as data security, data sharing, disaster recovery, reduced maintenance cost etc. However, there exists a challenge of using desktop GIS tools for editing and validation of the normalized data that is residing in a spatial database keeping data integrity intact with privileged and authenticated users from LDAP. This is being addressed in the current study by providing a viable solution to maintain such datasets.

II. RELATED WORK

Recent studies have indicated that most of the desktop clients provide data editing facilities based upon the shapefile format. Whilst, few of the software such as COTS provides database driven editing and validation facilities. In addition to the same, [1] proposed a strategy that offers multi-user parallel editing for spatial data stored inside Oracle database based on ArcSDE version mechanism. Two different modes, online edit pattern and offline pattern, were raised with respect to various environments. Later, [2] conferred that web based data editing can also be carried out by individuals which lack prior GIS experience. However, this unique functionality comes with its own set of challenges, wherein the dataset is exposed to the non-accessible users imposing the threat of malicious entry or data corruption. In the virtue of the same, [3], puts forward the online data editing method that designs a Windowing Query for large geometric features along-with Simplify Query that controls the vertex scale of geometric features. It follows an indexing structure based on DBLG-tree for geometric networks that helps in editing and updating of spatial features with minimal errors. [4] also inferred a framework for updating and maintaining multi-user geospatial databases for an electric utility data management. The study mainly emphasizes data maintenance protocols using native software functionalities. The architecture adopted, helps in multi-user access, share and edit spatial data subject to DBMS permissions. [5] also discussed the importance of the change detection and the corresponding data update, wherein the data update should be “data model aware”. The study emphasized the importance of the rule based and service oriented architecture mechanisms for update processes. For the multi-user spatial editing facility, [6] demonstrated the network data models to detect the topological errors within the spatial dataset. The requirement of these network models is that multiple users should be able to create and update the features simultaneously. A versioned database allows multiple concurrent users to make edits on it without affecting the integrity of the database. Similarly, [7], developed the web application for managing remote spatial data in crisis situations with the use of Web Feature Service - Transactional (WFS-T) that enables client application to alter the state of feature through insert, update, and delete operations through standard Geography Markup Language (GML) using GeoServer and OpenLayers. [8] proposed the method of spatial data version updating based on V-tree and controlling the errors raised out of updating by multi-user, multi-affair and offline update.[9] further demonstrated that spatial database management system (SDBMS) plays very important role in efficient data maintenance. Data integrity, quality and security is a critical bottleneck of maintaining and creating such databases. In the virtue of the same, [10] described the new approach for spatial data consolidation, editing and map revision. This approach was based on a unique spatial identifier given for every spatial feature in the database in order to Add, Delete and Update tables. Nevertheless,

these tables presented an efficient method to store and distribute revisions and updates of the national spatial database to various organizations.

III. METHODOLOGY

In order to implement the secured database editing and validating functionality for the GIS desktop clients, the present study demonstrates the solution for its product called “CityLayers” for managing the large database management of the municipal corporations. The methodology is mainly divided into the following parts:

- Automated generation of denormalized data layer and its replica from the existing normalized database to be used in QGIS for editing and validation.
- Syncing LDAP users and user access rights with PostgreSQL using Foreign Data Wrapper (FDW).
- Defining policies for Editing and Validation workflows.
- Series of Triggers and Procedures for maintaining multiple actions inside the editing and validation workflow.

The normalized structure restricts the data editing in QGIS as it accepts the denormalized data structure. To overcome this issue, automated generation of denormalized layers and its replica has been accomplished using triggers. The purpose of generating two denormalized layers is to provide editing on one layer where edits get validated while the other layer acts as reference layer.

Users and their respective access rights, as per jurisdiction boundaries and the data pertaining to it are defined at LDAP level. There is a challenge to provide the user and their respective access right information from LDAP to the database for accessing the data from it. Hence, there is a need to sync LDAP users with PostgreSQL databases. Here comes the role of Foreign Data Wrapper (FDW) which enables PostgreSQL server to proxy the information from other data sources. We have used a Multi-corn Foreign Data Wrapper (FDW) to access any data source in a form of foreign table inside PostgreSQL database. The LDAP-PostgreSQL sync mechanism is shown in Fig.1.

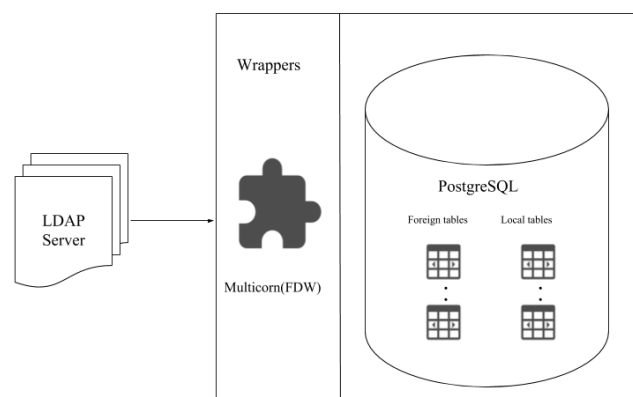


Fig.1 A schematic illustration of LDAP-PostgreSQL sync mechanism.

There is a mandate to have a user to be created at the database end to perform editing and validation actions along with enabling row level security. Besides, there also exists a challenge to track the changes happening for user’s privileges at the LDAP side and accordingly make changes at database end. To address this issue, there is a need to create an intermediate table which can be mapped with a foreign table, which only has a readable format. The changes happening at the LDAP side for users are being tracked through a cron job generation which keeps foreign and intermediate tables in sync with each other. The tables have row level security policies that restrict, on a per-user basis, regarding the rows which can be returned by normal queries or inserted, updated, or deleted by data modification commands in addition to the SQL-standard privilege system available through GRANT operation. This feature is known as Row-Level Security (RLS) which has been implemented through use of triggers that run due to the effect of cron jobs on intermediate tables. The policies can be created based on user role as well as column level attributes. Row level security added another level of security for PostgreSQL users. Hence, users with privileges to that table can access all rows for updating or querying based on AOI. Based on the user access rights and policies defined at database end, the user with editor rights will be able to edit inside a QGIS environment by making a secure connection with the PostgreSQL database and similarly the users with validation rights will validate the same, the resultant of which is reflected inside the database in the back-end (Fig.2). The auditing of data that is being edited or validated is recorded inside a separate log table.

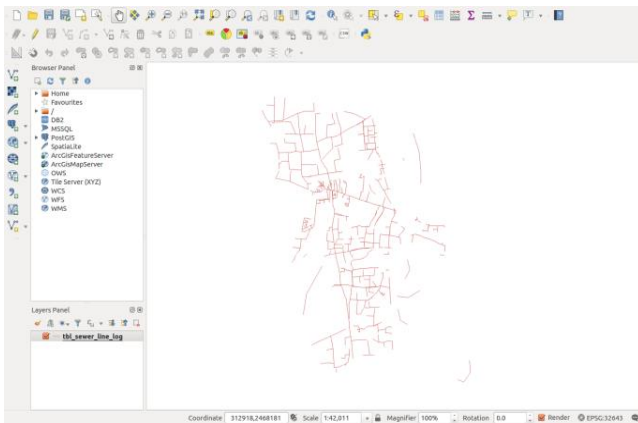


Fig.2 An illustration of the Data access as per editing rights

Besides, the softwares used in the study are listed in the Table. (1) for achieving the aforementioned objectives.

Table 1

Sr. No	Softwares Used
1.	PostgreSQL+PostGIS(RDBMS)
2.	Apache Directory Studio
3.	LDAP directory server
4.	QGIS, an open source desktop-GIS software.
5.	Multi-corn(FDW)
6.	Operating System (OS): Ubuntu Linux

IV. RESULTS & FINDINGS

The present study discusses the advantages of database driven spatial auditing using QGIS against LDAP authentication and authorization along with its workflow. QGIS, is an open source, cross-platform geographic system application that offers viewing, editing and analysis of geospatial datasets. In the present study, connection with the database has been done against the LDAP authentication service that is configured for various users. Upon setting up the connection with the PostgreSQL database in QGIS, either the users with the types editor or validator can access the layers. There are four types of policies that have been applied while implementing row level security which are as follows: **Select, Insert, Update and Delete**. The data is accessed in QGIS by users which is made available through a select policy applied for the respective users. The connection with the PostGIS database can be done as shown in Fig.3.

The editing of spatial data includes majorly three types of operations such as Insert, Update and Delete besides merging and splitting. When the editor inserts a record or feature, a new row with geometry and attribute data gets added into the denormalized table. The editor can add his remarks for entering the new feature into the dataset as shown in Fig.4.

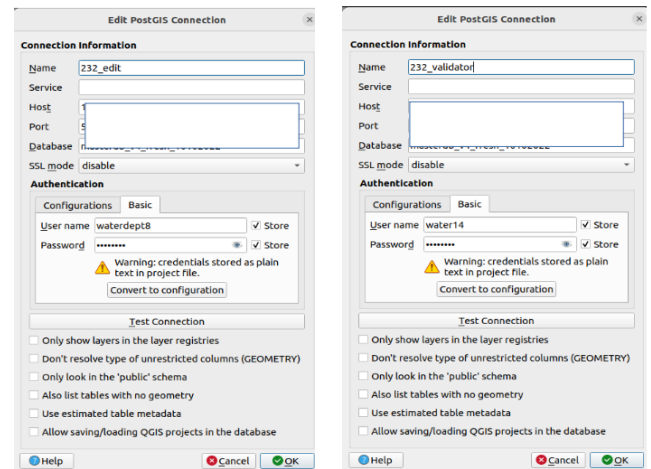


Fig.3 An ideal view of the editor and validator connection in QGIS.

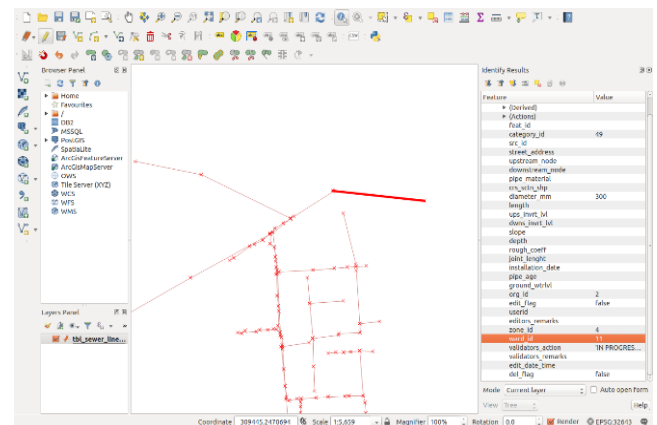


Fig.4 An ideal view of the Editor's feature INSERT action.

Upon addition, a flag gets generated which is a notification for the validator to understand the entry of a new feature along with a timestamp Fig.4. Besides this, the log table also adds the edit entry. The log table maintains the information about username, its action, table name, respective remarks, timestamp etc. The validator, thereafter, approves or rejects the change by adding the keywords 'ACCEPTED' and 'REJECTED' respectively along with remarks. Upon approval, the change that has been made in denormalized data gets added into reference data which triggers further changes to make those changes accordingly in the normalized data. Upon rejection, the feature that has been added into denormalized data gets deleted. In case of update the editor can either edit attribute data only or spatial data only or both at the same time. There is the provision of storing old records and new records in the log table. Upon approval by the validator, the edited entry remains unaffected in de-normalized data while it gets reflected in the reference table which in turn gets reflected in normalized data structure. Upon rejection, the old record gets restored in denormalized data without making any kind of changes into reference and normalized data. For delete scenarios, the feature/s that has been deleted by the editor get marked by delete flag. Upon approval by the validator, the features get deleted from denormalized data and reference tables and finally from normalized data structure. Upon rejection, the record remains as it is in the denormalized data without making any kind of changes into reference and normalized data. There are special cases of editing such as merging and splitting related to geometry. In case of merging of the features, the single merged record gets added into the table after deletion of subsequent features which got merged. While in case of splitting, the feature that gets split keeps its record into the table but other split parts also make entry into the table. There was a challenge of handling one-to-many relationship attribute data which is handled through the feature called relations in QGIS. This feature allows the user to connect feature layers with respective attributes tables with common relation keys between them. After making a connection, users can view features and their underlying data from other attributes tables using QGIS attribute viewer or information tool. In case of one-to-many relationships between attribute data, the relationship is maintained between the main attribute table and its associated relation tables where the primary key is for the main attribute table. The denormalized layer until now, used for editing is brought into QGIS along with its associated relation tables for editing one-to-many attribute relationship data using QGIS relation feature. Similar to the earlier approach, the users are given privileges on associated relationship tables using row level security policies mechanism to carry out editing or validation operations.

The advantage of this system is that concurrent users can make edits at the same time on the same feature class with feature-locking facility. If any edit is done on a feature by a particular user, that feature will be locked for other concurrent users for any further edit till the time the

previous edit is validated by validator as shown in Fig.5. It prevents multiple editing by users on a single feature. Hence using this editing/validation workflow, spatial data edits/validation can be handled at organization level with data integrity and relative user rights.

validation_id	validation_date	pipe_age	ground_water	org_id	edit_flag	user_id	dton_remark	zone_id	ward_id	validator_action	last_year	edit_data_time
16113	NULL	0	0	2	f			4	6	IN PROGRESS		
16114	NULL	0	0	2	f			4	6	IN PROGRESS		
16115	NULL	0	0	2	f			4	6	IN PROGRESS		
16116	NULL	0	0	2	f			4	6	IN PROGRESS		
16117	NULL	0	0	2	f			4	6	IN PROGRESS		
16118	NULL	0	0	2	f			4	11	IN PROGRESS		
16119	NULL	0	0	2	f			4	4	IN PROGRESS		
16120	NULL	0	0	2	f			4	11	IN PROGRESS		
16121	NULL	0	0	2	f			3	9	IN PROGRESS		
16122	NULL	0	0	2	f			1	7	IN PROGRESS		
16123	NULL	NULL	NULL	2	f			4	11	IN PROGRESS		
16124	NULL	NULL	NULL	2	f			4	11	IN PROGRESS		
16125	NULL	NULL	NULL	2	f			4	11	IN PROGRESS		
16126	NULL	0	0	2	L	sewer_edit...		4	11	IN PROGRESS		2018-04-13 15:59:34
16127	NULL	NULL	NULL	2	L	sewer_edit...		4	11	REJECTED		2018-05-17 15:54:36
16128	NULL	NULL	NULL	2	f			4	11	IN PROGRESS		
16129	NULL	NULL	NULL	2	L	sewer_edit...		4	11	IN PROGRESS		2018-04-11 11:24:00
16130	NULL	0	0	2	L	sewer_edit...		4	11	IN PROGRESS		2018-05-17 16:14:03
16131	NULL	NULL	NULL	2	L	sewer_edit...		4	11	IN PROGRESS		2018-04-16 10:42:49
16132	NULL	NULL	NULL	2	f			4	11	IN PROGRESS		
16133	NULL	NULL	NULL	2	f			4	11	IN PROGRESS		
16134	NULL	NULL	NULL	2	L	sewer_edit...		4	11	ACCEPTED		2018-08-03 18:27:35
16135	NULL	0	0	2	L	sewer_edit...		4	11	IN PROGRESS		2018-05-17 13:41:53
16136	NULL	NULL	NULL	2	f			4	11	IN PROGRESS		
16137	NULL	0	0	2	f			4	11	IN PROGRESS		

Fig.5. An ideal view of the Validator's action approving/rejecting the inserted changes.

V. DISCUSSION & CONCLUSIONS

The given solution is the part of CityLayers product which has been successfully implemented at Vadodara Municipal Corporation (VMC), Ahmedabad Municipal Corporation (AMC), and Bhavnagar Municipal Corporation (BMC) in Gujarat, India. In a municipal corporation, a centralized GIS repository has been hosted for the data needed for all departments. There is a department hierarchical structure of users in the corporation which demands for implementation of workflow systems even in the case of spatial auditing to track the changes. In such organizations, the user management is mostly handled through LDAP. The traditional way of making edits and approval was tedious and time consuming imposing difficulties in maintaining data audit. Online editing has become a new go to solution recently which lacks powerful editing tools and there are constraints of editing large data which affects the performance and transmission of the vector data for editing. The solution was developed considering the workflow needed for the organizations to ease the data auditing as per the privileges. The editing and validation of spatial data in a desktop GIS environment proves to be more effective and user friendly than online editing due to the availability of advanced editing tools and also in terms of performance criteria. The users are acquainted with shape file format like structure and that is the reason to adopt the denormalized layer approach as a solution. For one-to-many attribute relation tables, there is a need to have an understanding about the table relationship. Advantage of using this is that now multilevel hierarchical data structure can be easily maintained. The disadvantage is that notifications are not available at the QGIS side that have been generated from the database side. With the use of row level security policies, multi-corn extension for

foreign data wrapper for LDAP and thorough use of stored procedures and trigger functions made the solution robust and secure. The present study also implemented the proposed solution on various flavors of Linux such as Ubuntu, Suse, RHEL and even on Windows Server. Thus, the solution that has been provided and implemented is a cross platform solution. Currently, QGIS has been used as a desktop client and in future there is a plan for integrating the solutions with other open source desktop clients. The database driven spatial auditing using QGIS against LDAP authentication and authorization is the part of CityLayers product which provides the complete solution for the municipal GIS requirements of a corporation. The out-of-the-box solution provides an upper hand in handling database driven spatial auditing by integrating with LDAP which centrally manages user and their rights. Hence, the present study probes to be the first hand estimates of establishing a secured web based editing and validating functionality, considering Digital India policy. Besides, an effort has been made in the open source geospatial domain to provide a simplified OGC compliant solution for spatial data edits and validation along with auditing.

ACKNOWLEDGEMENT

This work would not have been possible without the support of the Geosolutions team at Nascent Infotechnologies Pvt Ltd, for providing their support and respective knowledge in implementing this solution. Besides, the authors would also like to thank various Municipal Corporations for providing the data and to get it implemented and utilise it as a part of that spatial decision support system for smart cities initiative.

REFERENCES

- [1]. He. B., G. Zhu. "Multi-user parallel edit strategy based on version management" *Geospatial Information*, Vol. 5 No.4, pp.1-4, 2007.
- [2]. F. Hardisty. "Web Editing: Opportunities and Challenges". *Cloud and Server GIS, GEOG*, Vol. 865, 2005.
- [3]. M.Liao., X.Q. "A Web-GIS Online Vector Data Editing Method based on Multi-scale Representation Data Structure" *Technical gazette*, Vol. 25, pp.171, 2018
- [4]. M.A.Owoola. "A Framework for maintaining a Multiuser Geodatabase: An Empirical Example". pp. 1-5
- [5]. P.A.Woodsford. "Spatial Database Update and Validation-Current Advances in Theory and Technology". *ISPRS*, 2007.
- [6]. P. Bakalo., E. Hoel., S. Menon., V.J.Tsostras. "Versioning of network models in a multiuser environment". In *International Symposium on Spatial and Temporal Databases*, pp. 6-24, 2009.
- [7]. T. Reznik., Z. Hynek. "Data management in crisis situations through WFS-T Client". In *Cartography and Geoinformatics for early warning and emergency management: towards better solutions (Joint symposium of ICA working group on CEWaCM and JBGIS Gi4DM)*, pp. 386-395, 2009.
- [8]. W. Xi, H. Y. "Based Version-tree Spatial Data Update". *Urban Geotechnical Investigation & Surveying*, 2011.
- [9]. Lin, X., Zhang, Y., Liu, Y., Gao, Y. "Spatial data integrity ensuring mechanism in SDBMS". In *Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS 05*, Vol. 1, pp. 4-, 2009
- [10]. Felusab, Y., Srebroa, H., Tala, Y. "Gis versioning management—the approach of the survey of israel". *International Archives of Photogrammetry and Remote Sensing*, Vol. 38, 2010.

AUTHOR'S PROFILE

Santosh Gaikwad, is the Director, Geosolutions at Nascent Infotechnologies Pvt. Ltd., Gujarat, Ahmedabad, India for CityLayers. He holds a Master's degree in Agriculture and Advanced Diploma in Bioinformatics. He has more than 20 years of experience in Spatial Informatics and has worked on geospatial aspects in various domains such as wetlands, biodiversity, agriculture, marine, education, insurance, etc. With a Chevening scholarship from UNEP-WCMC, Cambridge, UK, under his belt, he is also a charter member of OSGeo foundation and closely associated with the OSGeo-India chapter, has worked with international organizations such as British Antarctic Survey, UK; Pacific Biodiversity Institute, USA; The Salim Ali Centre for Ornithology and Natural History (SACON); and Professional Assistance for Development Actions (PRADAN), India.



Maulik Bhagat is a Technocrat and an Entrepreneur in multiple domains. He is a keen planner and strategist with abilities of conceptualizing, framing and implementing projects in multiple domains. He holds a master's degree in BE (ICT) with more than 16 years of Entrepreneurship experience. With the intention of touching the lives of common citizens, he has lead the development of an indigenous GIS platform "CityLayers" that is useful to the Smart Cities for managing the daily operations in a very cost effective way. This initiative is also in tune with various initiatives of the GOI towards like "Atma Nirbhar Bharat", "Vocal For Local", "Make In India". These initiatives have helped in drastic reduction of the Total Cost of Ownership for GIS solutions, thus can help in the increased adoption of the GIS technology in the country.



Arjan Odedra, is a Senior Team Lead at Nascent Infotechnologies Pvt. Ltd., Gujarat, Ahmedabad, India. With over 4 years of experience in the area of Open Source Web-GIS Software Development, Front End Development and PostgreSQL+ PostGIS database development, his contribution to CityLayers has been invaluable. He holds a Master's degree in Computer Applications (MCA) from Gujarat Technological University (GTU).



Rahul Kanani is a Senior Team Lead at Nascent Infotechnologies Pvt. Ltd. He has been working with open source geospatial technologies and development of applications for over 4 years. His expertise lies in OpenLayers, GeoServer, PostgreSQL/PostGIS, to name a few.



He has past experience working with principal scientists and organizations like MPCST, MPRDC, MPGRP. He holds a Master's degree in Computer Applications.

Aditya Saraswat is a GIS developer at Nascent Infotechnologies Pvt. Ltd. He has an experience of 1.5 years in working with open source geospatial technology and providing GIS and remote sensing based solutions. He majorly works on PostgreSQL and PostGIS database development. He has done his post graduation and holds a M.tech Degree in Geomatics from CEPT University.

