



Vault: A Model for Investigative Journalism and Whistleblowers

Dr. V. M. Gayathri¹, Dr. P. Supraja², A. Helen Victoria³

¹SRM Institute of Science and Technology, India, vmg188@gmail.com

²SRM Institute of Science and Technology, India, p.supraja18@gmail.com

³SRM Institute of Science and Technology, India, helenvia@srmist.edu.in

ABSTRACT

Promoting the protection and security of journalists and combating exemption for people who attack them. On average, each 5 days a journalist is killed for transportation data to the general public. Attacks on media professionals are usually committed in non-conflict things by gangland teams, militia, security personnel, and even native police, creating native journalists among the foremost vulnerable. With many authoritarian regimes and increasing oppressive culture against journalism, especially investigative journalism and whistle-blowing which puts many lives on the line, it is the inherent duty of the state to protect these people from any harm that comes their way. But as evident, such measures have not been given significant importance. India's press freedom index ranks 138 among 180 countries, North Korea being at the last position. Death toll of journalists is on the rise each passing year. Hundreds of Activists and protesters are killed each year. This problem could be solved by designing a system to enhance criminal investigation, investigative journalism, whistleblowing on the basis of anonymity, security and decentralization, to protect and fight against authoritarianism and fascism.

Key words: Transaction Processing Management, Application Programming Interfaces, Operations Support Systems, Journalism.

1. INTRODUCTION

Newspapers, wire services, and independent journalists have traditionally performed the most inquiring journalism. With advertising's decline in financial gain, many ancient news services have struggled to finance investigative journalism, which is long and so overpriced. Print media investigations square measure slowly dole out by news outlets working along, even globally (as in the case of Panama Papers and Paradise Papers), or by outlets such as ProPublica, which have not traditionally worked as news publishers and have trust in the public and benefactors' help to fund their work.

A. Classification

Whistleblowers need a secure measure to go public with the knowledge being anonymous at a similar time, otherwise they will be subjected several however not restricted to the subsequent risks.

a. Employer Retaliation

For someone who's considering blowing the whistle on their leader, a primary worry is that same leader can retaliate against them. While their square measure laws in situ meant to shield whistleblowers, they're typically not enough.

b. Industry Blacklisting

So, what if one employee offers you the boot, there's an entire trade out there, right? Perhaps, however if there's referred to as a whistle-blower then even corporations United Nations agency doesn't have something to hide is also reluctant to hire you.

c. Legal Consequences

While speaking up regarding criminal activity may be a noble issue to try and do, it's conjointly potential for whistle-blowers to be fixed within the penalization that comes with the crime. Often, they'll ought to face charges of their own for being a part of the crime within the 1st place, might build them even a lot of reluctant to square up and say one thing (even although processing the whistle may lead to a lighter sentence).

d. Professional Violations

In several cases being a whistle-blower could violate a contractor knowledgeable obligation. In these instances, there could also be extra penalties like civil suits for breaching agreements and confidences. Whatever proof they need collected has to reach the suitable authorities even once their deaths.

B. Background

Surveillance, information storage capabilities, and digital attack techniques have become ever more sophisticated, cost-effective, and omnipresent, potentially putting journalists at the risk of digital attacks by rising state and non-state actors. With the universal accessibility of police work code and equipment, in a number of states across various territories, some interpret narrowly defined legislative actions as silencing digital opposition, punishing whistleblowers, and extending capricious police work across various digital channels. The on-the-line information was launched by the International Press Institute in late 2016, a project that seeks to regularly track on-line abuse of journalists as a response to their coverage. The project had recorded 1065 instances of online abuse inside the 2 countries (Turkey and Austria) during which the project collected information as of the Gregorian calendar month 2017.

2. RELATED WORK

[1] studied on the implementation challenges of Node.js. Node.js is easily efficient for serious files and network, loading applications because of its event oriented, asynchronous and non-blocking approaches. The analysis as a result of a literature survey shows that Node.js provided a task for Full Stack Developers to come true again. In the absence of Node.js, discovering several entirely different languages and environments had become more robust for a developer. It also encourages organizations and developers to build extremely burdensome, simpler applications and eliminates single-page Applications [SPA] server calls by victimization, as well as making additional applications user-friendly and simpler. The additional result shows that the technology has generated it straightforward to accomplish high-load graphic method and net GIS [Geographic Info System] even faster, and is being used faithfully. In the context of its use within the group, node.js have several problems similar to their adoption by developers and organizations over the latest programming technologies. No doubt Node.js has good benefits, it's also some group challenges. One such difficulty is that developers have the potential to abuse the commonly used vocabulary. One enthusiast has built a Node.js victimization backdoor software package on Raspberry Zero. This will generate backdoors within the target laptop and their network despite the computer being barred for countersign. [2] studied as technological needs develop there is an rise in the need for user awareness using the internet which has contributed in real time to the growth of Web services. The exam involves a point-by-point analysis of Node.js which contrasts with the normal web server-IIS. Node offers an exclusive, event-based server with asynchronous efficiency. Node.js is event-based,

allowing developers to use asynchronous interfaces to perform I / O operations using a single thread for their event loop. Traditional web servers (IIS) are somewhat old, and follow the pattern of multiple thread requests. This paper demonstrates how Node.js and traditional web server's structural decisions are affected by the way apps carry out the sudden spike in demand. Web administrations work over system efficiency which is considered an assessment metric parameter. We're both built in node.js and in. Com, Internet Information Services (IIS Server) Supported. Precise experiments are carried out incorporating various conditions to allow a connection between the Node and IIS exhibitions. The results of the exhibition evaluation offer some significant Node and IIS Server presentation details in a specific period. Results show that Node.js is lightweight and professional in design. A state of a web server that can function intensively can be operated with hub. Therefore, developing web-administration APIs is beneficial. [3] studied the effect on the scalability of a simple node framework of different device architectures and designs. The main goal was to analyze the impact of different components in system in terms of performance and scalability of REST services built in Node.js. The service created was deployed on a multitude of configurations, for example for single machine instance and nginx load balancer. The performance metrics that were considered were latency and throughput. Lower the latency, better the performance. Other than the dimensions of scalability, various other algorithms or methods for achieving a scalable system are Co-location, Caching and Divide and Conquer. Co-location defines the existence of data in a physically connected and accessible space. If the data is co-located then it is fast to access. Divide and Conquer is used to compute or process multiple requests.

[4] checked the PostgreSQL lists by using an Efficient Hybrid Join Algorithm, a DB2 principle known as half breed join, which uses the join-list and skip-progressive pre-get method to effectively recover data. The reasoning for this estimate is to figure out the outside table on the join row. By then, the record is joined to the outside on the internal join field. The intermediary addresses the internal tuple which is passed on in the record. The after-effect of the partial join is placed on the intermediary and the internal table is then successively reached to complete the product of join. Using the inner parallelism of strong state drives based on blaze memory in knowledge preparation clarifies that Streak memory based solid state drives (SSDs) can probably alter the storage establishment at a fundamental level via their tip top and low power.

The focal points include the use of documents to get to the data quicker. Adding a record inside and out to a portion will allow you to request the speedier of data. Nonetheless, the trade-off for each record you

have is that the data will be inserted at a slower rate. If you add it with a list locations, information is composed at two locations and also holds the sort on the file when you add information. [5] Ding Zhang [2017] examined the correspondence between web application and inserted frameworks utilizing Node.js. This paper exhibits that in order to set up correspondence between the Web applications and the installed frameworks, and to help the productivity implanted framework and PC program communication, Answer to creating a web server hypertext transfer protocol (HTTP) and using it to merge with the Node. JS stage transmission control protocol or user datagram (TCP / UDP) attachment system is advanced. This paper also puts forward the strategy to create the webserver on the Express Edge. The correspondence progress between have PCs and the implanted framework alongside the convention they follow are represented also. The conclusion drawn out from this paper is that the trending usage of remote operation centre on branch side host computer with a lighter weight local application. Most customers need to gain access to the control system anywhere, anytime, and even on any gadget without adding any extra applications. The web application interface operates with a backend server, which is a superior option for most clients and terms of use. In this way, lightweight, adaptable Web applications are actually being used instead of large applications in the work field, and are being replaced step by step. The proposed way to deal with the HTTP Web application in this paper, permits the client to get to the program to get simpler strategy to see, change the inserted gadget, and also provides means to embed cluster management system

[6] focused on the unwavering nature of PostgreSQL and its developed OSS RDBMS, which in the ongoing past is being widely used in large business frameworks. It provided an equal framework of execution and the Dynamic Shared Memory and the Dynamic Context Workers. In any case, it doesn't ensure about the common information being mapped to a similar location space in the procedures and about information with pointers not ready to be shared legitimately without local duplicates. This paper proposes another mutual memory structure to be specific Shared Memory Context that gives a typical shared memory mapped to all procedures to a similar location in the endeavor framework. It additionally empowers a proficient equal preparing in PostgreSQL as it maintains a strategic distance from pointless between process shared information duplicates.

The end result of this paper is the introduction of a new shared memory framework called SMC with a strong PostgreSQL interface. For a viable parallelization in PostgreSQL, SMC gives a proficient between process information sharing instrument which is basic and furthermore permits

ease creation of equal handling codes. The execution of SMC are currently utilizing it for the information sharing instrument and for the equal handling broadening PostgreSQL of another columnar information store. Presently work is being done to additionally look at the presentation and suitability of SMC, and furthermore its ease of use in different applications.

[7] researched the file sharing and synchronization process and equipment using a computer-implemented synchronization process provided for file management and sharing. The system comprises in one personification of the present invention, the production of a server platform which provides a data store for maintaining and preserving Copies of the initial files exchanged by a large number of sharing clients in maximum resolution. On the server is given a synchronization technique that is designed to send real-time updates to a multiple number of sharing clients when at least one of the sharing clients updates or changes one of those files. There is also a web interface which allows the user to access files in the data store using a web browser. File sharing and data management is the biggest trouble we face today. With different types of files on different devices such as computers, phones, watches, or personal digital assistants, the data management in an effective manner from device-to-device is a challenge for users.

[8] contemplated the circumstance of analytical reporting in China which is at a dubious state. Weight is from different sides, both from the gathering state and publicists that have diminished this sort of news-casting to a disheartening condition. Then again, insightful reporting likewise demonstrates out to be a significant apparatus in the monetary advancement of certain papers, and is a basic piece of different hierarchical structure just as giving an expert belief system to columnists. Weight on news associations keep on becoming much further, and they will undoubtedly react. A few, TV news outlet and furthermore numerous papers, have pretty much abandoned insightful news coverage. Many endeavor to hold the training, however receive an extremely cautious and mindful technique. In the vast majority of the cases the market circumstance of the paper and the self-personality of the writers, which whenever put their lives on chance, imply that they hold a solid promise to this field of analytical news coverage. In this, the Internet has come to extraordinary guide for the writers, which gives a decent source to stories, a ground on which it is conceivable to distribute material and make it open information that couldn't show up in the customary media, and a method for guaranteeing that exciting stories increase a more extensive crowd. Then again, some news outlets have created systems, even those papers that value keeping up their promise to analytical news coverage, to limit the negative political and financial results of their movement.

[9] contemplated the circumstance of analytical reporting in China which is at a dubious state. Weight is from different sides, both from the gathering state and publicists that have diminished this sort of news-casting to a disheartening condition. Then again, insightful reporting likewise demonstrates out to be a significant apparatus in the monetary advancement of certain papers, and is a basic piece of different hierarchical structure just as giving an expert belief system to columnists. Weight on news associations keep on becoming much further, and they will undoubtedly react. A few, TV news outlet and furthermore numerous papers, have pretty much abandoned insightful news coverage. Many endeavor to hold the training, however receive an extremely cautious and mindful technique. In the vast majority of the cases the market circumstance of the paper and the self-personality of the writers, which whenever put their lives on chance, imply that they hold a solid promise to this field of analytical news coverage. In this, the Internet has come to extraordinary guide for the writers, which gives a decent source to stories, a ground on which it is conceivable to distribute material and make it open information that couldn't show up in the customary media, and a method for guaranteeing that exciting stories increase a more extensive crowd. Then again, some news outlets have created systems, even those papers that value keeping up their promise to analytical news coverage, to limit the negative political and financial results of their movement.

[10] studied about E-mail synchronization which involves data type synchronization, such as a flag, content, unique identifier (UID), and header. The client devices implementation of an optimized retrieval scheme to receive a multiple number of messages when a trigger is identified. This retrieval scheme involves retrieving various types of data (e.g., UIDs, flag(s), and text) before retrieving certain types of data for one group of messages to other groups of messages. Retrieval schemes are automatically adjusted and determined so that it provides faster retrieval of such data based on the user input for particular messages. It can also be optimized when, in the system there are multiple mailboxes are present. Users can retrieve e-mail messages on their device from various mailboxes. For example, some users have, with different account providers, one or more e-mail accounts, such as an e-mail service provider, an employer, and the like. Under the omnipresent Internet Message Access Protocol (IMAP) which is used for retrieving e-mail messages, it may comprise a Unique Identifier (UID), a content part (text and/or attachment), one or more flags, and a header. Existing retrieval schemes do not account for user experience and also does not account for a likely immediate need of a user. This is one of the disadvantages of the existing systems. Existing methods are also slow from the viewpoint of a user.

For example, once a user starts a synchronization process on a mailbox, it may take a notable amount of time (e.g., tens of second to minutes) after which the user is shown with a screen of useful data about messages in one of the user's mailboxes.

[11] studied the disclosure related to identifying resource rules for email access, identifying email resources associated with client devices, determining whether the email access resources fulfil the resource access rules, and modifying the email access resources as well as the ability to access the email resources based on the resource access rules if the resource rules are not satisfied by the email resources. Controlling email access is crucial to ensure that email access resources may only be accessed in authorized manners and only by authorized parties. To date, enterprises have utilized access credentials to have control on which, if any, parties can access email resources. This method may ensure that, for controlling email access system, only authorized parties may access email resources. However, one disadvantage of this method is that it fails to control how the access for proper recipients of those email resources. One more disadvantage is that it fails to modify the email resources to obey with enterprise security standards. As enterprises look for ways to protect sensitive information, systems and methods for controlling email access are important to ensure that email access resources may only be accessed in authorized manners by authorized parties. [12] explains the procedure to improve the reliability of the node. [13] proposed cryptographic technique for secure communication.

3. PROPOSED WORK

Designing a system to enhance criminal investigation, investigative journalism, whistle blowing on the basis of anonymity, security and decentralization to protect and fight against authoritarianism and fascism. Some of the key features of the system are:

1. A time-based dead-man switch to check the activity of the author.
2. Encrypted storage to encrypt and store the files of the author.
3. Automated mailer system to mail to news and media outlets.
4. Disassociation of identity of the author and the sensitive information.
5. Rollout of encryption keys on a regular basis.

Figure 1 describes the actual flow of this work. It represents the modules and their specific state transfer into respective modules with certain criteria. The interaction with the system firstly starts with login page having Auth0 server also known as

password-less login directing the user to certain privileges of the system.

Some of the advantages of the proposed system are:

1. Guarantees the delivery of the sensitive information, should anything happen to the author.
2. No mediator is involved in transfer of information.
3. Anonymity is preserved.
4. Files are encrypted.
5. Authenticity can be verified by the news and media outlets.
6. Does not put the author at any risk.

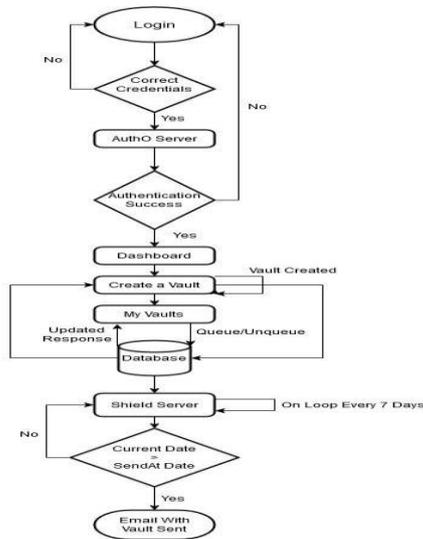


Figure 1: Flow Diagram of proposed work

4. RESULTS AND DISCUSSION

This work consists of two main sections; Front-end – the Graphical User Interface [GUI] of the system, Back-end – consists of the implementation logic of the system and the database which is used to store information of the users. It is a compiler and interpreter which compiles the templates into HTML code, CSS Scripts. Node.js – is an open-source, cross-platform, runtime framework for JavaScript that executes JavaScript outside of a browser. PostgreSQL – also known as Postgres, is an open-source, free relational database management system that emphasizes extensibility and compliance with SQL.

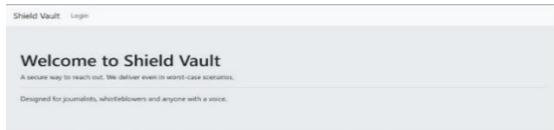


Figure 2: Home page

Figure 2 represents the home page of our website. The front-end has been created in Pug.js which is the template engine for Node.js. The browser reads HTML and CSS, then shows compressed images and HTML-based text to the client, and CSS asks them to do so. Pug is the middleman and it enables us to insert data and then to construct HTML. Pug framework is highly flexible which supports the generation of HTML content dynamically for almost any sort of data.

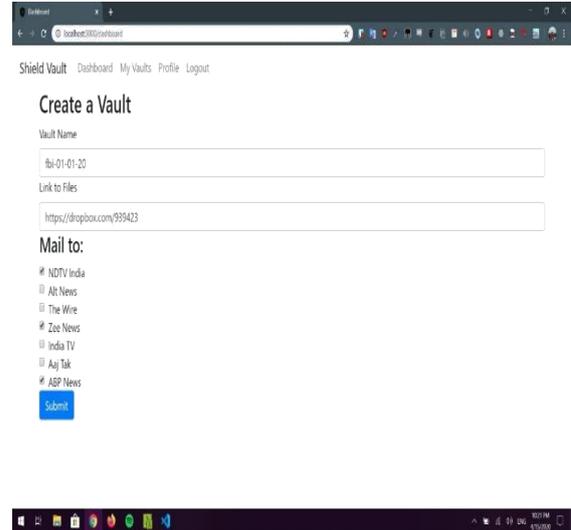


Figure 3: Vault Creation

Figure 3 represents web page where the user can create a vault to store his gathered information or collected confidential information. The user can choose which media outlets he/she wants to send the confidential information and the vault is stored in the database with a respective timestamp.

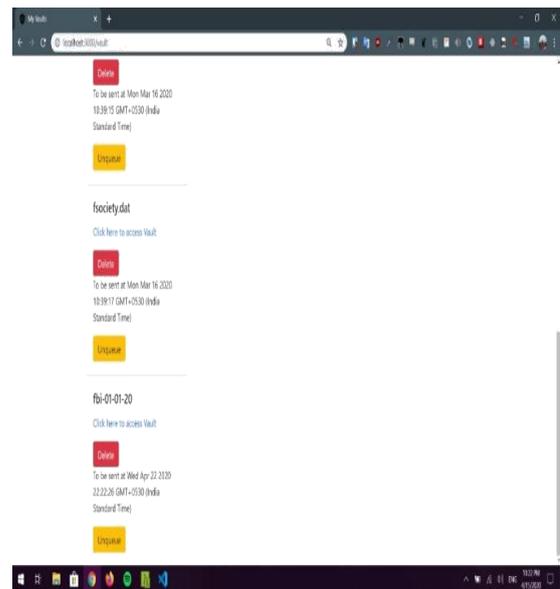


Figure 4: Vault Information

Figure 4 represents a web page with the vaults containing confidential information created by the

user. The vaults are in the state of unqueued and once its state transfers to queued, it is ready to perform the designated operation.

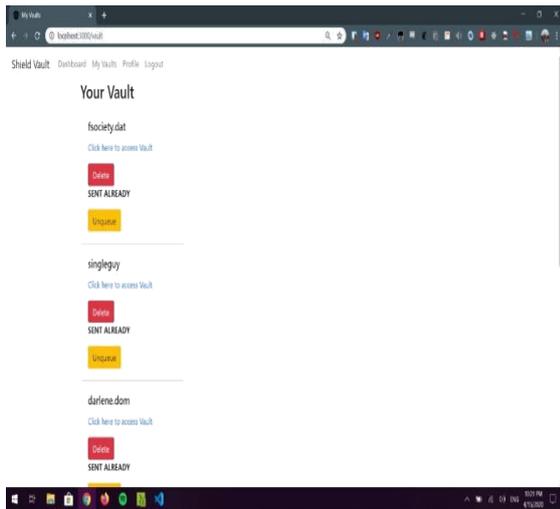


Figure 5: Vault Sent

Figure 5 represents a web page having all the vaults containing confidential information created by the user which have been sent to the specified media outlets.

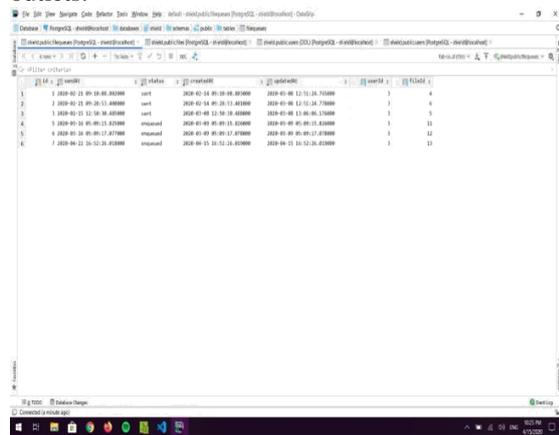


Figure 6: Updated Database

Figure 6 represents database having all the vaults created by the users. These vaults have respective links which directs it to the information and to which media outlets it is to be sent.

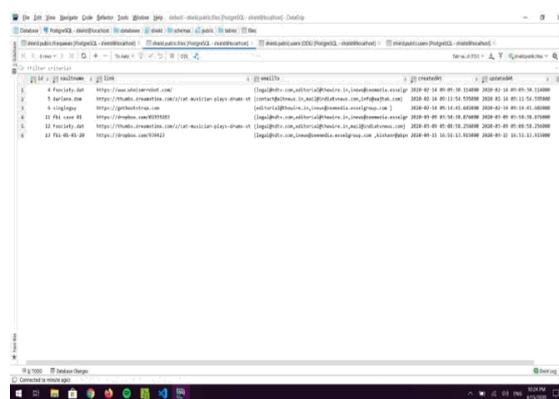


Figure 7: Vault Status

Figure 7 represents a database having the representational status of the vaults created by the users. The status of the vaults whether they are “sent”, “queued” and “unqueued”.

5. CONCLUSION

The system offers services for journalists and whistle-blowers making certain that the rightful information is transported among people. The system allows the users to store any confidential information and put it on a queue of time. If the user doesn't dequeue it, the information will automatically spread among all the local news channels. The main focus of the system is to preserve the authenticity of the information and prevents from spreading false information thus preventing chaos and hatred among people. The system provides security to the users by using OAuth technology which uses password-less login so that there is no way that the system can be hacked by another person as the passwords are not being stored anywhere.

Loss of Anonymity can result in many issues, loss of life being one of them. Many journalists have been killed over the years as their anonymity has been lost to the Organizations in power. This system provides anonymity by not storing the User Details on the website, thus preventing even the possibility of the leak of user information to anyone.

Fake news is the cause of many unrests in any civil society. Fake news can result in innocent people getting killed and struggles between different sections of society and even with the government. Thus, the spread of fake news is being prevented by the system as the news is being sent to the news agencies where they can verify the facts with the help of their sources. Since the system is designed to send the files at a 7-day interval, it gives the user enough time to decide whether the information gathered is enough to send or more info need to be collected before sending it to news agencies. The user can easily **queue and dequeue** the files to deal with this issue. Thus, this work solves the issues of security, anonymity, the spread of fake news and issues in Investigative Journalism and Whistleblowing. Future Enhancements could include: A native file uploading can be included so that the user does not have to look for other storage providers. An application for the website could be developed for increasing its accessibility and its audience reach. Mechanisms could be added so that the user can customize the time limit of sending the messages. The uploaded files could be encrypted which would result in better security. In the worst-case scenario, even if the website is hacked, if the files are encrypted, the hacker would not be able to access the files. The site could be enhanced so that it could create its own multiple proxies to preserve information.

REFERENCES

- [1] Hezbullah Shah and Tariq Rahim, “**Node.js Challenges in Implementations**”, Vol 17, No 2-E: Global Journal of Computer Science and Technology, May 2019.
- [2] L. P. Chitra and R. Satapathy, “**Performance comparison and evaluation of Node.js and traditional web server(IIS)**”, 2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET), Chennai, pp. 1-4, 2017
- [3] Denis Rucaj et al, “**Evaluate scalable and high-performance Node.js Application Designs**”,2019,https://www.academia.edu/40939578/Scalable_and_High_Performance_No_dejs_App.
- [4] S.Sultana and S. Dixit, “**Indexes in PostgreSQL**,” 2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, pp. 512-515, 2017.
- [5] Ding Zhang, “**The communication system between web application host computers and embedded systems based on Node.JS**”, IEEE,2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, 2018.
- [6] Y.Ujibashi, M. Nakamura, T. Tabaru, T. Hashida, M. Kawaba and L. Harada, “**Design of a Shared Memory mechanism for efficient parallel processing in PostgreSQL**,” 2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA), Corfu, pp. 1-6, 2015.
- [7] Jack Strong and Gibu Thomas, “**Method and apparatus for a file sharing and synchronization system**”,<https://patents.google.com/patent/US20060101064>, 2019.
- [8] J.Tong& C. Sparks, “**Investigative Journalism in China Today**”, Journalism Studies, Pages 337-352, May 2009.
- [9] G.L.Ciampaglia et al ,“**ComputationalFact Checking from Knowledge Networks**”,PlosOne,<https://doi.org/10.1371/journal.pone.0128193>, June 2015.
- [10] Neil V. JHAVERI et al, “**Prioritized email synchronization with server**”, <https://patents.google.com/patent/US20160359786>, 2019.
- [11] Alan Dabbieri, “**Systems and Methods for Controlling Email Access**”, <https://patents.google.com/patent/US20140165153>, 2014.
- [12] Makmur A et. al, “**The use of Smart Contracts in Utility Business**”, International Journal of Advanced Trends in Computer Scienceand Engineering, Vol. 9, No.3, June 2020.
- [13] Yasser Lamalem, Khalid Housni, “**An Improved Algorithm for Network Reliability based on Finding all Node Cutsets**”, International Journal of Advanced Trends in Computer Science and Engineering, Vol. 9, No.3, June 2020. <https://doi.org/10.29121/ijesrt.v9.i6.2020>