A Secure and efficient Database Management System based on Integrated Statistical Data Analysis modelling and Privacy Preserving Analytics

GAYATHRI.A[1] , THANGA REVATHI.S[2]
[1]Associate Professor, Department of Computer Science and Engineering,
Saveetha School of Engineering, SIMATS, Chennai, India
email: gayathribala.sse@saveetha.com
[2]Assistant Professor, Department of Computer Science and Engineering,
SRM University of Science and Technology, Chennai, India
email: thangarevathi84@gmail.com

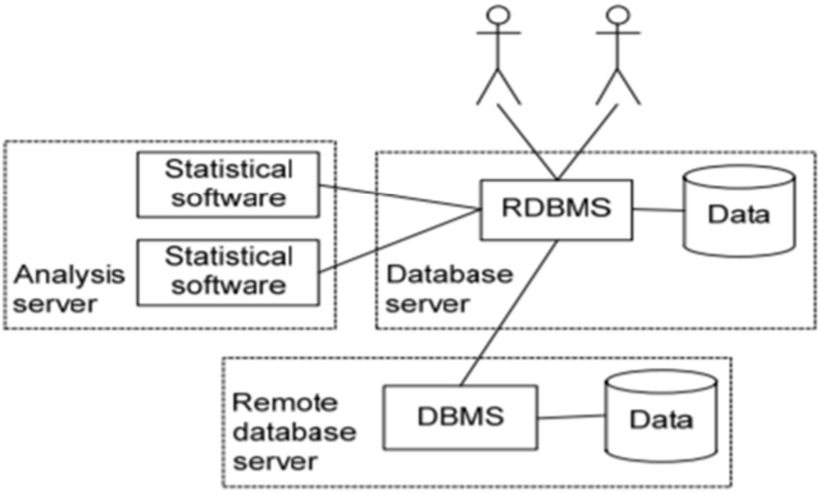
ABSTRACT: Access and incorporate the data in international research collaborations picked up in various countries. For different purposes e.g. Legislation, the owners of the data must monitor who has access to and how it analyses their results. Data review is carried out in the Statistical applications, usually named on top of a data Control method, for instance the Data Base Management System (DBMS). Access to data therefore is regulated by the DBMS, whereas, statistical analysis is normally reviewed by another Installation. In this paper proposes a novel architecture for improving health carrying out statistical analysis of the data contained in the DBMS. The proposed statistical software architecture is named from a DBMS. The architecture allows for the management of both data recovery and data recovery Statistical analysis of data from one system i.e. DBMS.

KEYWORDS: DBMS, SQL, RDMS, Data, SAQEL, Database server.

INTRODUCTION:

Data access is measured by the DBMS, Where as, statistical analysis is typically looked at by another Installing. [1][4] We propose a modern Health Improvement Architecture the statistical analysis of the data found in the DBMS is carried out. In the Spiel the proposed architecture of the statistical program is derived from a [2] DBMS. DBMS. DBMS: The architecture enables all data recovery to be handled. We implemented a primary prototype SAQEL, Statistical Analysis from SQL. It was tried on an investigation from epidemiological examination on cervical disease [3-6]. The fundamental exercise that we gained from the model is that it is anything but difficult to execute a shouting interface from a RDBMS, e.g., IBM DB2, to measurable programming, e.g., SAS. Our following stage is to actualize an easy to use interface to restore the investigation results, which are regularly a few tables and images analysts.

Building frameworks dependent on far off investigation workers [1-4, 7] needs huge assets.



A task subsists where actualizes an interface to invoke (call) investigation programs from a DBMS. Le Select [9] broadens an information base inquiry language SQL with capacity to call picture investigation programs. SQL is utilized in Le Select to give clear and effective approaches to perform distinctive picture investigation calculations over pictures in a dispersed shared condition. The aftereffect of an investigation is a solitary table. Interestingly, the focal point of our work is on factual examination, which brings about a few tables and charts.

LITERATURE SURVEY:

It gives an outline of past endeavors made in various zones of dispersed information base concerning the issues like Concurrency control, Load Balancing, Cache, Query Optimization, Checkpointing and Security. Simultaneousness control is needed to guarantee the consistency of the framework [11-15]. Burden adjusting builds the presentation of workers, brings about their ideal utilization and guarantees that no single worker is overpowered. Burden adjusting is particularly significant for occupied organizations, where it is hard to appraise the quantity of solicitations that will be sent to the worker. Inquiry Optimization utilizing Cache is a development idea in appropriated information base and security shields an information base from a unintended action.

I. CONCURRENCY CONTROL: Simultaneousness Control securities that the correct upshots for instantaneous activities are delivered whereas giving those upshots as speed as could reasonably be expected. PC frameworks, both programming and equipment, are comprised of components or parts. Each segment is anticipated to work effectively for example to conform to or meet definite uniformity rules. This clarifies why the locking calculation is as yet favored for business DBMS.

II. LOAD BALANCING: In circulated information bases, alongside simultaneousness control, other primary contemplations are: decreasing the quantity of obstructed exchanges, sparing framework assets and improving the productivity of the framework as far as speed. While looking after consistency, the quantity of hindered exchanges can increment bringing about diminished framework execution. Along these lines, load adjusting must be contemplated in order to diminish the recurrence of the hindered and streamline the presentation of each site in the appropriated information base.

III. CHECKPOINTING PROCESS: A checkpoint is a preview of the condition of a cycle saved money on the steady stockpiling which can be reloaded into memory to decrease the measure of lost work in recuperation. [JCA2004] tended to the need of applying diverse checkpointing plans to various subsystems inside a solitary objective framework. The proposed calculation has a few preferences. It is anything but difficult to actualize and just subsystems utilizing free checkpointing plans must be adjusted.

IV. CACHING: One of the significant way to improve the presentation of web administration in circulated information base is to utilize storing systems. By reserving web reports at intermediary workers or workers near end clients, client solicitations can be satisfied by getting the mentioned record from a close by web store rather lessening the appropriate response season of the underlying worker.

V. QUERY OPTIMIZATION: Ceaseless question and inquiry improvement over information streams have become a problem area in data set exploration. Arrangement of ceaseless and methodical information delivered in budgetary data observing, network checking, security, web applications fabricating, media communications information the executives, sensor networks are called information streams. It has the qualities of extraordinary greatness changing, much of the time restricting the ideal opportunity for question. Thusly, it is hard to control the request in which it streams out and it is difficult to spare all the information. Along these lines, its inquiry and capacity must be changed over from a customary mode to another strategy to be specific nonstop question and dynamic preparing chronicled information. network data transfer capacity utilization just as worker load.

VI. SECURITY: A protected framework is one that can be trusted to hold privileged insights, and the watchword here is "trust." Individuals, governments and associations, for example, banks , medical clinics and other business elements can possibly pass their data to the PC organization in the event that they can be completely guaranteed of privacy.

The weight for security bunch is the equivalent for all the undertakings. Actually, various undertakings may require diverse level of the different kinds of security administrations. An errand may require more validation administrations than classification administration and the other inquiry enhancement.

COMPARISON OF ALGORITHMS ON TOP 5 OPEN SOURCE NoSQL DATABASE:

A.) MongoDB: MongoDB data files aren't encrypted and all data is automatically stored as plain text. Accordingly, Hackers can directly access the data, and also the data may be read immediately. MongoDB sometimes doesn't support authentication while it is running in split mode.

The secret to this authentication is the password that has been named 'pre shared code' Hashed with MD5 algorithm before saving in file page. Hence, when hackers get a key access File, they will see the pre-shared secret hash value, rather than plain text, where the data becomes unusable for them. Hackers may however break the pre-shared the hashed meaning of secrecy.

For relational databases i.e. Oracle SQL Server and MySQL, hackers may launch database attacks server, or bypass web server authentication using Injection SQL. Similarly, hackers will invade MongoDB uses JavaScript. Assault samples Started on MongoDB, seen in.

B.) Cassandra: The data files of Cassandra are stowed without encryption, and there is no automatic data encryption at the servers, it helps hackers to access the data that can be read straight away. Cassandra can handle Query Language in Cassandra (CQL) so that DBA can handle easily and clearly data inside database. This CQL does have a similar, Syntax as SQL's, so it is thought it could be attacked as with the SQL injection.

C.) CouchDB: CouchDB has no automatic encryption of the data as other NoSQL's, data archives are at risk retrieved directly, and read. In Contact Words among server and client, or between CouchDB servers, authentication is achieved by means of a system known as CRUD. Beside the communication, there is also an embedded SSL encryption which is a Compounding of CouchDB itself.

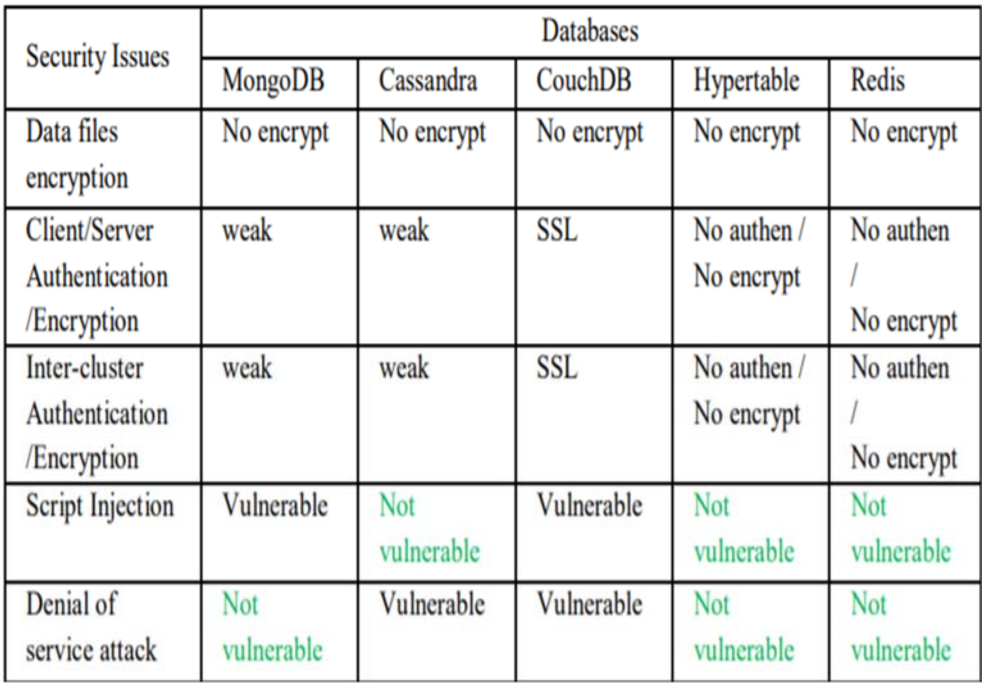
D.) Hypertable:

As with other, NoSQL's, there is no encryption at Hypertable the correspondence between the consumer and the server or between its servers is carried out for its data files without authentication, and encryption of data. Amazingly, Hypertable is free of injection vulnerabilities this has a data processing HQL, close to that of the Relational Database commands SQL.

E.) Redis: Redis does not sustenance data encryption in the same way as MongoDB and Cassandra, so the hackers or everyone with an access to the Redis server will indeed to retrieve all database data.

Not at all data encryption is carried out in the communiqué between the Redis client and the Redis server; In both supplementary servers, either on the same or different servers. Cluster, because Redis was designed primarily to function quickly, so there are not many components for its security. It is because of that reason that hackers have access to the network of Redis servers will be capable of detecting data while it's being conveyed.

Table 1. The Security Comparison of the Top 5 Open Source NoSQL Databases



Security Issues	Databases				
	MongoDB	Cassandra	CouchDB	Hypertable	Redis
Data files encryption	No encrypt	No encrypt	No encrypt	No encrypt	No encrypt
Client/Server Authentication /Encryption	weak	weak	SSL	No authen / No encrypt	No authen / No encrypt
Inter-cluster Authentication /Encryption	weak	weak	SSL	No authen / No encrypt	No authen / No encrypt
Script Injection	Vulnerable	Not vulnerable	Vulnerable	Not vulnerable	Not vulnerable
Denial of service attack	Not vulnerable	Vulnerable	Vulnerable	Not vulnerable	Not vulnerable

METHODOLOGY:

The technique for enhancing security through the use of database management is deponds on Encryption, a web-based solution. Server Protection, Adverse Registry, Authentication and Access Control, Timeliness and Control Safety in Real world Database Systems, SQL Injection Testing Schemes.

Encryption: This is the method of converting normal text info using encryption algorithms (called ciphers) that make it indecipherable to everyone but those with distinct features, Knowledge, which is usually referred to as the secret. Current database structures use plain text there are also risks of data manipulation and database failure. The data were used to stop these attacks. Stored in encrypted form.

Web-based Database Security: a range of methods are proposed to develop database security an unauthorized intrusion database. The transfer of data from the server to the client should be carried out using the Stable Socket Layer in a safe way. The Host Identity of the End Machine Regulated.

Deleterious Database: Negative data is applied to the original data in the database to avoid misuse of data from malevolent users and afford secure data recovery for all legitimate consumers.

Authentication and Access Control: Authentication is used to correctly verify the identity of the user. User and Access Management controls the behaviour or activities of the user. Access Control is given, Similar rights for specific authenticated users

Real-time Server Systems Timeliness and Security: Trade-off has to be made between the two. Security and transaction priority. Various approaches are suggested to ensure health and security have a small risk of exceeding time limits in real world database systems.

SQL Injection Testing Schemes: SQL Injection is a code inoculation technique that takes advantage of SQL Injections. A protection flaw that exists in the application's database layer.

CONCLUSION:

In this paper introduced a new framework for the execution of statistical analyses on interconnected data in a safe manner. Proposed method to make use of good protection of the privacy of RDBMS. This will make it possible to build stable assimilated infrastructures at a fair cost. A Secure and efficient Database Management System based on Integrated Statistical Data Analysis modelling and Privacy Preserving Analytics, first SAQeL prototype efficaciously exhibits the interface between DB2 and SAS.

REFERENCES:

[1] Barry, S., Marc, C.: remote access systems for microdata statistical review. Statistics and Computation 13 (2003) 381-389. See http:/www.lisproject.org as well.

[2] Borchsenius, L.: new innovations in the Danish Micro Data Access Method. Monographs on official statistics (2005) 13-20.

[3] Hibbert, M., Gibbs, P., O'Brien, T., Colman, P., Merriel, R., Rafael, N., Georgeff, M., Molecular Medicine Informatics Model (MMIM). Stud Health Techno Data 126 (2007) 77-86. See also: http:/www.biogrid.org.au.

[4] Hjelm, C.G.: MONA-Microdata ON-Access line at Statistics Sweden. Meta analyses on official figures (2005) 21-28. See http:/www.scb.se as well.

[5] SAS, http:/www.sas.com.

[6] DB2 Technology, http:/www.ibm.com/software/data/db2/.

[7] Sparks, R., Carter, C., Lehman, J.B., O'Keefe, C.M., Duncan, J., Keighley, T., McAullay, D.: Remote access approaches for excavation data processing and mathematical modelling: Privacy-Preserving Analytics. Computational methods Systems Biomed 91 (2008) 208-222.

[8] Haas, L.M., Lin, E.T., Roth, M.A.: Data fusion through database federation. It's IBM Syst. P. 41 › 2002 578-596.

[9] Luc, B., Fran, Oise, F., Fabio, P., Patrick, V.: Processing Queries for Costly Functions and Large Objects in Distributed Mediator Systems. The sessions of the 17th International Computer Engineering Conference. Computer Society of IEEE (2001) 91-98.

[10] Tisell, C., Orsborn, K.: A system for multibody analysis based on object-relational database technology. Advances in Engineering Software 31 (2000) 971-984.

[11] Ruslan, F., Magnus, S., Jan-Eric, L.: Federated Databases as a Basis for Infrastructure Supporting Epidemiological Research. Proceedings of the 2009 20th International Workshop on Database and Expert Systems Application. IEEE Computer Society (2009) 313-317.

[12] Stata: Data Analysis and Statistical Software, http://www.stata.com.

[13] The R Project for Statistical Computing, http://www.rproject.org.

[14] Open Database Connectivity Overview, http://support.microsoft.com/kb/110093

[15] Gayathri, A., Christy, S."Image de-noising using optimized self similar patch based filter",International Journal of Innovative Technology and Exploring Engineering, 2019, 8(12), pp. 1570-1578.

[16] Rama, A., Gayathri, A., Christy, S."Fine tuning data mining algorithm for an efficient classification of E-coli",International Journal of Innovative Technology and Exploring Engineering, 2019, 9(1), pp. 109-113

[17] Kumar Reddy, A.J., Gayathri, A., Mahalakshmi, D. "Automatic spam detection on twitter based on content and online social interaction",Test Engineering and Management, 2020, 82(1-2), pp. 10603-10606.

[18] Reddy, K.N., Gayathri, A., Devi, T., "A primary warning methodology of train following interval supported government agency" Test Engineering and Management, 2020, 82(1-2), pp. 10499-10505.

[19] Manjusha, V., Gayathri, A., Logu, K., "Design of efficient multi-server password authenticated key management protocol for cloud computing environments", Test Engineering and Management, 2020, 82(1-2), pp. 10493-10498.

[20] Reddy, M.H., Gayathri, A., Deepa, N. "An automatic method to prevent cybercrime incidents using artificial intelligence approach" ,Test Engineering and Management, 2020, 82, pp. 10488-10492.

AUTHORS PROFILE

Dr.A.Gayathri, received the B.E degree in Electronics and Communication Engineering from Periyar Maniammai College of Technology for Women (Bharathidasan University, India) in 2001 and the M.Tech (CSE) degree in Computer Science and Engineering specialization from Bharath University, Chennai, India in 2005. She completed the Doctorate in the Department of Information and Communication Engineering at Anna University. She is currently working as Associate Professor in Saveetha School of Engineering (Department of CSE), SIMATS, Chennai, and Tamil Nadu. She is the member of CSI, IAENG and ACM.

Dr.S. Thanga Revathi, working as Assistant Professor in SRM University of Science and Technology, Chennai. She has completed her Bachelor of Engineering degree in Computer Science and Engineering from Anna University, Chennai with distinction. She has completed her Master of Engineering in Computer Science and Engineering from Bharath University, Chennai. She completed her PhD in Anna University, Chennai and indulged in research work in the field of Data Security. She has a overall experience of 13 years in the field of Teaching. She has published papers in referred International and National journals. She has also presented papers in various National and International Conferences