

Data Sharing & Discussion Portal: A Literature Review

Bhagyashri Kshirsagar¹ Ravishankar Purne² Sanket Unche³ Rahul Adhao⁴

^{1,2,3}Student ⁴Assistant Professor

^{1,2,3,4}Department of Computer Engineering & Information Technology

^{1,2,3,4}College of Engineering, Pune, India

Abstract— Lack of communication between students and teachers has caused a great impact on the quality of knowledge the students gain in India. A Data Sharing and Discussion Forum equipped with all the necessary facilities can be a great tool to handle the situation very effectively. We try to study few technologies that can make such portal. We review technologies like NoSQL, MongoDB, Lucene, Elastic search, Postfix mail server, cloud computing and show a system can be based on them. The review indicates how these technologies form the best solution to the institutional problems.

Key words: Cloud Computing, Discussion, Portal, Review, Sharing

I. INTRODUCTION

Data Sharing and Discussion Portal defines as an application (more likely web-based), that provides capabilities for multiple users with different permission levels to manage Educational content and a comprehensive discussion portal for discussing various topics. It provides students the flexibility of time and place to reflect on the previous postings to the discussion thread and thus actively engages them in a meaningful and intellectual experience. It also provides a platform to share files among them to form a good educational network. The portal can be used for blogging also so the students can express them in a broader way. This insures a regular flow of information for increasing communication and ultimately to improve sharing of knowledge.

Sometimes, the data in an institute need not be on the internet. It is better kept inside the institute. In such situation, we can manage servers storing such information with ease as the number of users is limited. We can manage access to services on the server affordably as it is an intranet. It is mainly aimed at educational institutes. The students and faculty can interact with each other with very fewer efforts. Users can start discussion threads on any topics. We tried to gather most of the previous work done on this topic and designed a system based on it. The next section shows the study of related research articles and the third section describes the proposed model based on the study.

II. LITERATURE REVIEW

Data exchange standards of most colleges are out of date which led to the existence of information islands. The Information Systems cannot share data between each other to form a robust system. Cloud computing is an emerging model in which computing tasks' repository constitute a lot of computer resources so that all application can get

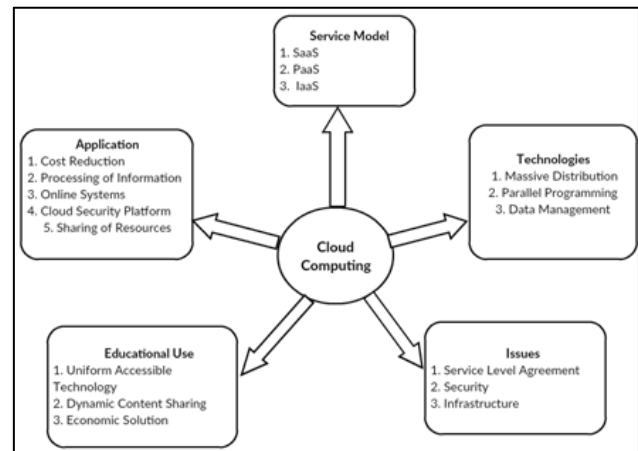


Fig. 1: Cloud Computing Overview

Computing power, storage space, and various software services. Key Technologies of Cloud Computing are massively distributed memory computing, Parallel programming model, Data management technology and Distributed resource management [1]. Cloud computing can be applied to reduce cost in Information processing, Online systems and Resource Sharing.

Fig.1 shows Application, Service models, Technologies, Educational use of cloud computing and Issues faced. Tuncay Ercan(2010) [2] focuses on the educational usage of cloud services and its benefits to higher educational institutes. SaaS supplies the hardware and software product. The interface can be provided by the web portal. University strives for cost-effective, broadcast feasible and equal access to technology for students and staffs. Availability and scalability are features of cloud that enables users to build customized solutions and enlarge their computing environment. As students' perspective is not limited to only contents, the contents can be changed dynamically. There are still some disadvantages of cloud-based solutions are constraints with application offering, service-level agreements, and security issues. But hopefully, these backdrops will be overcome.

The basic conception of cloud computing is not limited to only business but it can be used in education also. So here the discussion is not only about possibilities offered by cloud in education but also how it is perceived by students. In the survey conducted [3], the author found that most working people are using cloud services mostly. Pensioners are in the second place and students are behind them. The main reason for using the cloud is its ability to use file format on various devices. Authors conducted an additional study on students of the Technical University of Czestochowa [3] and they found out that largest percentage of people used the cloud for less than a year and most commonly used cloud application is Dropbox followed by facebook, twitter, skype.

Many students also agreed that the use of cloud may reduce the use of printers, papers. When asked if they will use a cloud solution offered by University, vast majority students

responded positively. A wide range of benefits and positive impact on the level and quality of education, should be a cause for the implementation of this solution by universities.

Jaroslav Pakorny [4] discusses NoSql in context of cloud computing, architecture, a basic feature of these databases and their horizontal scalability. Fig. 2 shows types of NoSQL databases. The author writes about cloud computing, transactional processing, scalability. Databases which do not implement ACID property fully are eventually consistent. He also introduces CAP theorem, Consistency, Availability, and Partial Tolerance. This theorem is also known as Brewer's Theorem. Usually, RDBMS prefers Cover A and P. In scalability RDBMS does not provide a cost-effective way. While NoSQL can be scaled linearly with a number of servers used. Horizontal data distribution enables division of computation task. NoSQL restricts overhead occurring in RDBMS. Data model consists of a key-value pair. Querying in NoSQL can be done by specific API. Nosql databases are eventually consistent. So, for now, NoSQL is still far from advanced technology and they are improving day by day.

Cloud Educational Resource Datacenter (CERD) [6] based on DaaS model which can provide economically profitable remote access for educational facilities. Authors try to solve the problem of acquiring, upgrading software due to lack of qualified staff by introducing CERD. CERD is described as {Scheduled Templates, Software, Requests} where users are treated as a request. The accessible Software is stored on a computer server which then serves to different user requests. The users' requests are handled by round-robin scheduling algorithm. Thus CERD eliminated the need of a proprietary license for each system per software and significantly reduces cost.

Jose, Abraham (2017) [7] had tried to find out advantages of NoSQL database as the conventional approaches in RDBMS does not support flexible scalability needed for recent systems and tried to show how data spoken to and queried in MongoDB. Big databases are the need of time but supporting systems have not emerged yet. Big Data represents a huge volume of data. NoSQL proposed for storing unstructured data. A study was made to integrate MySQL and MongoDB using middleware system another research proposed data analytics framework for knowledge discovery. NoSQL database has a different type such as Key-value, Document, Column family, Graph. MongoDB is document based one. NoSQL has several merits such as elastic scaling of the volume of data, Big data can be managed, No special DBA is required and economically viable option.

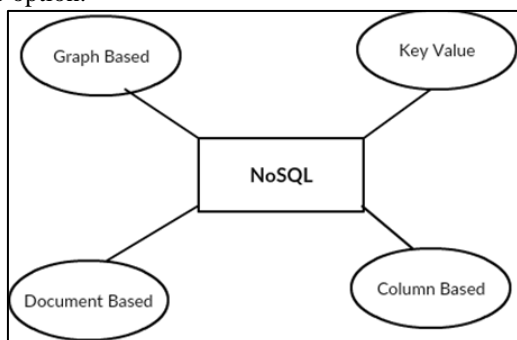


Fig. 2: Types of NoSQL

Antonios, Konstantinos, Andronikou and Anagnostopoulos [8] emphasized on analyzing key design characteristics of NoSQL data store. Further authors tried to analyze the relationship between NoSQL and cloud infrastructure. "Shared Nothing" horizontal scaling nature of NoSQL enables a large number of read/write operations per second. Instead of ACID properties, NoSQL provides BASE semantics. (Table Comparison). NoSQL data model contains Key-value, Column family, Document store, Graph store models. MongoDB is of type Document store. Paper also discusses data placement which is an optimal distribution of data. NoSQL leverages on elastic infrastructure thus enables itself to add more remote VM, storage space as their need. Paper also compares various NoSQL data model with respect to data partitioning techniques and replication techniques for Nosql. Due to increasing requirement of data and digitization in every field NoSQL databases are set to dominate.

RDBMS is known for its Simplicity, efficiency, robustness, and compatibility. More recently need of user are increasing this RDBMS lack of fulfilling these needs, New generations software has a huge volume of data which is unstructured and scalable. Thus RDBMS would not able to handle it optimistically. NoSQL can help with its distinguished features such as schema-less, scalability, data replication, and partitioning. RDBMS misses these functions. Leonardo Roacha [9] presented a framework in this paper which can perform impulsively and evidently data and model migration from RDBMS to NoSQL. Abstract layer present in framework help to access data in NoSql model without affecting existing queries in the database.

Rakan Alsowail and Ian Mackie (2017) [10] presented an approach of using a type system to solve the problem of accidental misuse of shared files by trusted authorized users. The violation of file policies is misuse. They designed language of commands to manipulate files and specify their policies. A type system is implemented to enforce these policies. It intercepts the commands issued on the file and determines whether or not the command is safe to be executed. The policy includes an approach to limit the number of times a file can be read or copied. If cannot be copied, then it can be read once. They define three security types of files: UC, LCn, and NC which stand for Unrestricted copy, Linear copy, and No copy respectively. Executing commands that violate files policies causes Run time error.

Masnizah Mohd (2010) [11] shared the experience of using Lucene. The experience is based on the use of Lucene by 28 second-year students from Information Science program. Indexing and query processes are the two major components of a search engine. Creating data structures or the indexes is aimed at the Indexing process while the query process uses the structures and queries to generate a list of documents. The ranking of documents is also an important part of query processing. The sub-processes in Indexing are Text acquisition, Text transformation, and Index creation. User interaction, Ranking, and Evaluation come under Query process.

D.Saravanan (2016) [12] proposed an efficient framework for image retrieval from the given video. When a user inputs an image query, the framework processes it and compares with relevant images from database and then the result is computed. It involves two important techniques for

the efficient retrieval. One is the RGB feature which removes redundant frames in a video. The other is Segment Based Indexing Technique which indexes the individual frames. The hierarchical clustering mechanism is used for clustering video frames.

Advanced Direct Connect (ADC) [13] is a peer to peer file sharing and chat protocol, using the same network technology, concepts, and terminology as the Direct Connect (DC) protocol. The ADC protocol is a text-based protocol, where commands and their information are sent in clear text, except during password negotiation. The client-server aspect of the protocol stipulates that the client speak first when a connection has been made. There are no port defaults, for hubs or clients. Hub addresses are in the following form: adc://example.com:411, where 411 is the port. Searches use a token, as well, to identify each result of a search. Downloads are transported using TCP. Searches can be transported using TCP or UDP. The protocol allows for extensions such as compression with bzip2 or encryption with TLS.

While comparing Solr and Elasticsearch Ugur Kilic and Isil Karabey (2016) [14] mention that Elasticsearch is schema-less, Solr is not. Solr is used in content searches, data analysis and inquiries for projects such as Instagram, SourceForge, eBay. Solr operates such like REST type API server using Lucene libraries. Elasticsearch is less complex, less detailed compared to Solr search engine it is real-time and distributed. As the data stored at databases grows, we encounter speed/performance problems in query operations carried out over this data. Elasticsearch has many full-text search capacities such as a multi-language option, a strong query language and autocomplete. It is a document-oriented search engine. It does not request definitions such as index, type, field type before the indexing process as it is schema-free while Solr requires all the fields to be known in advance.

Xue-meng Li and Yong-yi Wang (2015) [15] described how to manage the index in an Elasticsearch based search engine. Elasticsearch starts its job from analyzer which is based on Lucene. The process of analysis has three main components. The text flows from character filter then tokenizer and finally through the token filter. These tokens are used for creating a dictionary and posting list. Dictionary contains every token and has a pointer pointing to posting list. This structure is used for faster retrieval of data. For compressing the index, Delta rule and Prefix-Suffix rule are

used. Authors put forward an indexing technique which uses category and sub-category while building the index.

Yong Zhang and Jian-lin Li (2009)[16] analyzed Lucene system architecture, indexing, and searching mechanism. They also studied sorting of indexes and presented a new retrieval sorting algorithm and a way to adjust indexing performance. Lucene indexing mechanism is not configured to deal a large number of documents by default. There are three parameters given in the IndexWriter for adjusting the performance: mergeFactor controls the number of documents to be stored before writing them to disc and how often to merge index segments together. maxMergeDocs decides the number of documents per segment. minMergeDocs Controls minimum number of Documents to be buffered before they are merged to a segment. Retrieval sorting algorithm of Lucence is based on Vector Space Model. It considers document and keyword similarity but neglects the importance of page while the pure Page Rank algorithm considers the importance of page but neglects the query. Yong Zhang and Jian-lin Li present a new algorithm which is a combination of above two algorithms.

Liangqiu Meng (2015) [17] presented a computer-aided system for student management. The system developed is very modular as it is built in four layers considering the hierarchy. Web display layer is where the client application is deployed on a web platform for end users. Logic layer performs the task of handling the functionality of the application. Data can be accessed through Data access layer and Database layer is deployed for storing the student information. Various levels of authorization can be assigned to different types of users for the smooth working of application and giving privileges to users.

SMTP (Simple Message Transfer Protocol) used for delivery of email which includes control headers along with the body of messages. Control header enables the users to exchange email among each other within and from outside area network. [18] Email is used for communication between customers, employees, managers for sharing the information. The email system offers a fast, reliable and easy solution for such communication. SMTP (simple mail transfer protocol) is a transportation protocol used to transfer e-mail messages over the Internet. SMTP also used to send and receive the mail from client to server.

Sr. No.	Paper Name	Security	Protocol Used	Theory/ Algorithm
1.	Postfix Mail Server Configuration In RHEL6 [18]	Control headers along with the body of messages.	SMTP protocol used for sending and retrieving mail.	-
2.	Secure E-Mail Protocols Providing Perfect Forward Secrecy [19]	Using a public key cryptosystem, such as RSA and ElGamal	A. The receiver requires a portable device to remember a used secret random integer. To overcome the requirement. B. The mail server takes over the job instead of the receiver. The second protocol is more flexible and useful to the e-mail system.	RSA algorithm, Diffie-Hellman
3.	Implementation Of Secure Email Server In Cloud Environment [20]	A. Configuring a secure email server on Linux platform (Ubuntu OS) to use Local-	SMTP authentication by configuring SMTP_AUTH framework as well as encryption mechanism by TLS (Transport Layer Security).	-

		based Email Transfer Agent such as Postfix. B. SASL(Simple Authentication and Security Layer) C. Two authentication modules Dovecot SASL and Cyrus SASL.		
4.	Secure And Privacy Enhanced Email System As A Cloud Service [22]	S/MIME cryptographic format by maintaining the privacy of user's identities.	A. S/MIME, PGP B. After designing and developing the protocol, they verified it using security verification tool Scyther. They found that their protocol provides protection against the replay attack.	-
5.	A Performance Model For Domino Mail Server [23]	-	IMAP, POP3	Queue theory
6.	Evaluation Of Linux SMTP Server Security Aspects-A Case Study [24]	Security provided by A. Amavis B. SpamAssassin C. ClamAV D.Integration with Postfix(VMI)	SMTP	Anti-spamming techniques.
7.	Design And Implementation Of A Web-Based Internet/Intranet Mail Server Management System [25]	-	SNMP(Simple Network Management Protocol)	-
8.	Active E-Mail System SMTP Protocol Monitoring Algorithm [5]	DNS system and TCP protocol for sending and receiving by a valid user.	SMTP	Monitoring algorithm
9.	Identification Of Spam Email Based On Information From Email Header [21]	Email header	-	Header-message-based spam detector

Table 1: Comparison between Approaches to Implement Email System

The table I shows the comparison between different approaches to implement email system, protocols, security and algorithms used.

independently and has separate storage. All types of mails are supported by the system. The discussion forum pages and files uploaded by the user are searchable through the Elasticsearch API.

III. PROPOSED SYSTEM

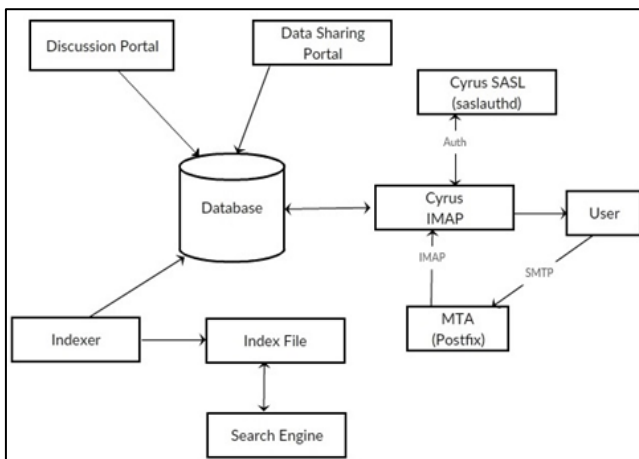


Fig. 3: Proposed System

The proposed system is made by studying the above discussed research papers as shown in Fig.3. The main elements of the system are Elasticsearch, MongoDB and PostFix mail server. MongoDB and Elasticsearch are configured to work together while the mail server uses the users' email-id from the users table. The system is made to index text as well as video data. The mail server works

Elasticsearch was chosen over Solr as it supports NoSQL which might be useful for further applications of this system. Postfix was chosen as the mail server considering the security aspect of the system. MongoDB is chosen for the best support to the unstructured database. The swiftness of the search engine, security of the mail server and flexibility of the database will make the system robust.

IV. CONCLUSION

Cloud computing is a great technology to use the available resources in an efficient manner. NoSQL allows scaling of databases which is used to store unstructured data in large quantity. Combination of cloud computing with NoSQL databases can give immense strength to the system. With presence of Lucene based search engine (Elasticsearch), the system becomes more relevant solution to current day problems. In addition, a secure mail server removes all the communication gaps among students and teachers. The system binds sharing of information and communication together without letting any third party interfere with the data on intranetwork. The whole system is implemented with open source technologies.

REFERENCES

- [1] Chen Lin, "A Novel College Network Resource Management Method using Cloud Computing", International Conference on Applied Physics and Industrial Engineering, Physics Procedia 24 2293 – 2297, 2012.
- [2] Tuncay Ercan, "Effective use of cloud computing in educational institutions", World Conference on Educational Sciences, 2010.
- [3] Tomasz Lis, Bajdor Paula, "The use of Cloud Computing by Students from Technical University – the Current State and Perspectives", International Conference on Communication, Management and Information Technology, 2015.
- [4] Jaroslav Pokorny, "NoSQL Databases: a step to database scalability in Web environment", International journal of Web Information Systems, 2013, Vol. 9 Issue: 1, pp.69-82.
- [5] R. Sureswaran, Hussein Al Bazar, O. Abouabdalla, Ahmad M. Manasrah, Homam El-Taj, "Active E-mail System SMTP Protocol monitoring algorithm", 2nd IEEE International Conference on Broadband Network & Multimedia Technology, 2009.
- [6] Shukhman, I. Bolodurina, P. Polezhaev, L. Legashev, "Cloud Educational Resource Datacenter Simulator", XIIth International Symposium Intelligent Systems Moscow, Russia, 2016.
- [7] Benymol Jose, Sajimon Abraham, "Exploring the Merits of NoSQL: A Study Based on MongoDB", International Conference on Networks & Advances in Computational Technologies (NetACT), Trivandrum, 20-22 July 2017.
- [8] Antonios Makris, Konstantinos Tserpes, Vassiliki Andronikou, Dimosthenis Anagnostopoulos, "A classification of NoSQL data stores based on key design characteristics", Cloud Futures: From From Distributed to Complete Computing, Spain, 18-20 October 2016.
- [9] Leonardo Rocha, Fernando Vale, Elder Cirilo, Dárlinton Barbosa, and Fernando Mourão, "A Framework for Migrating Relational Datasets to NoSQL", International Conference On Computational Science, 2015
- [10] Rakan Alsowail and Ian Mackie, "Controlling File Access with Types", Electronic Notes in Theoretical Computer Science 332, March 2017.
- [11] Masnizah Mohd, "Development of Search Engines using Lucene: An Experience", Procedia Social and Behavioral Sciences 18 282–286, 2011.
- [12] D.Saravanan, "Segment Based Indexing Technique for Video Data File", Procedia Computer Science 87 12-17, 2017.
- [13] Amol Bhagat, Radhika Chaudhari, Kiran Dongre, "Content-based file sharing in peer-to-peer networks using threshold", Procedia Computer Science 79 53-60, 2016 .
- [14] kılıç, uğut & Karabey Aksakalli, Isil, "Comparison of Solr and Elasticsearch Among Popular Full Text Search Engines and Their Security Analysis", 10.13140/RG.2.2.24563.32803.
- [15] Xue-meng Li and Yong-yi Wang, "Design and Implementation of an Indexing Method Based on Fields for Elasticsearch", Fifth International Conference on Instrumentation and Measurement, Computer, communication, and control, 2015.
- [16] Yong Zhang and Jian-lin Li, "Research and Improvement of Search Engine Based on Lucene", International Conference on Intelligent Human-Machine Systems and Cybernetics, 2009.
- [17] Liangqiu Meng, "College Student Management System Design Using Computer Aided System", International Conference on Intelligent Transportation, Big Data & Smart City, pp.212-215, 2015.
- [18] Mr. Rathin Mapa, Nirmal Kole, "Postfix Mail Server Configuration in Rhel6".
- [19] Hung-Min Sun, Bin-Tsan Hsieh, and Hsin-Jia Hwang, "Secure E-mail Protocols Providing Perfect Forward Secrecy", IEEE Communications Letters, Vol. 9, No. 1, January, 2005.
- [20] Nadim K. M. Madi, Solmaz Salehian, arzaneh Masoumiyan, zizol Abdullah, "Implementation of Secure Email Server in Cloud Environment", International Conference on Computer and Communication Engineering, Kuala Lumpur, Malaysia, 2012.
- [21] Shukor Bin Abd Razak, Ahmad Fahrulrazie Bin Mohamad, "Identification of Spam Email Based on Information from Email Header", 13th International Conference on Intelligent Systems Design and Applications (ISDA), 2013.
- [22] Amna Joyia, Abdul Ghafoor, Maryam Sajjad, Muhammad Qaisar Choudhary, "Secure and Privacy Enhanced Email System as a Cloud Service", Eighth International Conference on Digital Information Management, 2013.
- [23] Yi Liang, Lei Wang, Ruihua Di, "A Performance Model for Domino Mail Server", International Conference on Computer Science and Software Engineering, 2008.
- [24] Salam Khanji, Raja Jabir, Liza Ahmad, Omar Alfandi, "Evaluation of Linux SMTP server security aspects-A case Study", 7th International Conference on Information and Communication Systems (ICICS), 2016.
- [25] Jae-Young Kim, James Won-Ki, "Design and Implementation of a Web-based Internet/intranet Mail Server Management System", IEEE International Conference on Communications, 1999.