Keyword search technique enables end users to selectively get back files of their interest and this technique has been widely applied when the scenarios is that of users doing plaintext search. But since data encryption restricts the ability of the user to do keyword search and also asks for protection of keyword privacy, this plaintext search technique is not suitable for encrypted data, where any such search fails when data is encrypted.

Ahuja R [4] proposed a novel technique to exploit the cloud for the purpose of qualitative improvement in services being rendered. In this context, a service level agreement is executed between the service provider and the user, incorporating stipulations regarding terms & conditions relevant to the transactions. Agreement envisages the provisioning of services to the users at the appointed time and charges commensurate with the market scenario. The focus of the model is to provide fair deal to the users and consumers, enhancing the quality of service as well as generation of optimal revenue.

Adrea Zanella et al [6] presented an urban IOT system. This article gives a survey of providing technologies, protocols and architectures for an urban IOT. They have also presented how cloud environment is useful for urban IOT.

Ye Yuan, Guoren Wang, Lei Chen, and Haixun Wang [77] developed the solution for searching through uncertain graphs using the keyword. To answer the query they have used the concept of filtering and verification framework. The result produced is a sub tree to a search word query.

Zheng Yan et al [79] developed a new technique for secure data storage, data access control and revocation. They developed a new scheme to de-duplicate encrypted data stored in cloud and proxy re-encryption.
Zhiyuan Yin, F. Richard Yu [80] addresses the security issues by Mutual Trust Based Access Control Model arise in cloud environment. Cloud computing can provide users scalable web services with virtualized, which faced with serious security challenges. The significant feature of the cloud computing which ensures is the access control. They proposed an efficient model based on mutual trust based access control.

Anton Akusok et al [7] suggested learning machines Toolbox (ELM) for Big Data. They proposed a solution to linear algebraic performance. They also summarize the advantages in algorithmic performance.

Cloud-based pervasive system must be secured holistically and discussed in such a way this might be accomplished. B. Lagesse [8], in his article 'Challenges in Securing the Interface between the cloud and Pervasive Systems’, offers the pervasive system utilizing cloud computing resources and the issues that must be addressed in the system. In this system, the user's mobile device cannot always have network access to leverage resources from the cloud, so it must make intelligent decisions so that the data is being stored locally and what processes should run locally. The user becomes vulnerable to attacks while interfacing with the pervasive system.

Dorit S Hochbaum and Philip Baumann [21] proposed a method of sparse computation. Instead of computing complete similarity matrix their method computes only relevant similarities. They implemented their approach for devised supervised normalized cut (SWC).

Elisa Bertino [22] proposed the role based access model in data management system. He also presented the challenges, approaches and concepts in database security. The limitation was that a new device has to be issued if the user changes the subscription.

In his article 'Designing Cloud services adhering to Government privacy Laws', Reich C [20] introduced the Cloud Data Security (CloudDataSec) technique that focus to propose cloud services adhering to government privacy laws and thereby introducing these privacy and security challenges. They mainly focus on the six layer security model and the
three level assurances for SMEs in cloud computing. They have also proposed security management as a service (SMaaS) method to provide needed privacy and security operations.

**Gabriele Cavallaro** and **Morris Riedel** [25] talks about an efficient classification method to support vector machines in a big data environment. Some satellite sensors like Hymap and AVIRIS is used in their method.

**Nguyen, Myeong-Seon Gil** and **Yang-Sae Moon** [29] suggested to extent Hadoop focusing on YARN (yet another resource negotiator) that could provide multitenant functions. First, they thoroughly analyzed reasons that made it difficult for Hadoop 2.x to provide multitenancy and clearly defined three problems. There are four requirements derived which are necessary to support multitenancy and resolve these problems. This work also presents an overall multitenant YARN framework to satisfy the requirements and proposes a detailed component structure and design scheme based on the actual Hadoop source code.

The concept of cloud evaluation model and network resistance ability is proposed by **Huimei Wang** and **Ming Xian** [32]. The proposed model is three-fold: (1) they propose a series of network attack resistance metrics that measure the attack impact and attack cost during an attack. (2) They present the weight of evaluation metrics using 1-dimensional normal distribution clouds, which cut down the evaluation results' impressible. (3) They illustrate the algorithm of the attack resistance ability evaluation model. They successfully demonstrate that our evaluation method using cloud model is brief and effective.

**Jevin D West et al** [35] proposed an efficient solution for improving the scholarly navigation. They introduce a new method called hierarchical clustering of an article level citation network. They have solved the problem of delivering scholarly articles, if one knows a little of an article.

**Weiwei Zhu, Haojin Cao** [36] proposed a secure mobile based user data service mechanism to access data in cloud. This can be used at a minimized security management overhead. The core idea of this is that not only data management is outsourced but security also. This method has many advantages over traditional methods.
Jiang Yunxia et al [37] proposed a method to avoid idleness of IT resources in enterprise or institution using openstack. They have designed the private enterprise cloud platform for resource virtualization mode and the utilization and distribution of the virtual machine. They also provide the user interface design scheme.

Due to rapid advancements, VM is used by clouds. In order to safeguard data, it is required to identify the kernel of the VM, the software stack and hardware platform. Kai Zhang, Ying Song [42] suggested a trusted connection based virtual machine architecture and implements a system prototype. It also discovers isolation to untrusted VMs. There are three main parts in our system: trusted chain, attestation and isolation.

Kun Liu et al [44] propose an efficient algorithm for balancing the overload racks. This method prepares prior balance list and next balance list. The racks are balanced using these lists.

Kwang Mong Sim [45] proposed a new agent paradigm to build software tools and test beds for managing Cloud resources. The Agent based cloud computing is to design and develop the software agents for strengthening cloud service recovery, service negotiation and service composition.

Mashayekhy and Mahyar Movahed Nejad [47] show that there are significant optimization opportunities within the MapReduce framework in terms of reducing energy consumption. Also proposes two energy aware MapReduce scheduling algorithms, EMRSA-I and EMRSA-II that schedule the individual tasks of a MapReduce job for energy efficiency while meeting the application deadline. Both proposed algorithms provide very fast solutions making them suitable for execution in real-time settings.

The main idea of the article which is proposed by Li Dongming, Li Yanis [48] to increase the efficiency of storage and load capacity of the platform is by providing the seeds based on hadoop in cloud environment. This paper is designed and implemented the seeds based on Hadoop trace data processing model, through the processing of the crawl data of seed, such as ETL process, parallel computing, distributed storage.
A Hadoop based efficient economic system is proposed by Lin Weiwei, Chen Liang [50] to improve condition and solve problems of poor data storage, efficiency, data security and reliability, low resource utilization and high input costs. A Hadoop uses efficient resource scheduling and probabilistic redundant scheduling to efficiently organize in the available resources existing within an enterprise.

Muhammad Adnan [54] proposes the idea of addressing the big data issues by cloud computing based on Hadoop. Knowledge erudite in big data is being developed by sensors, personal computers, mobiles and the process retrieving that knowledge embedded is quite challenging millions of dollar spent to do the same. This article encourages the IT people to make use of Hadoop infrastructure to create cloud system for processing their data reliably end effectively.

The main idea of this article proposed by Qiang Duan et al [59] is to combine networking and cloud computing that that allows management, control and optimization of both networking and computing resources in a Cloud environment which leads networking to play a significant role in cloud computing calls for a whole version.

Shuhui Jiang et al [67] developed a personalized travel sequence recommendation from community contributed photos from travelogues. Their approach not only gives the user a travel interest but also give a travel sequence. The famous routes are ranked based on similarity between user package and route package.

Adam J Elbirt et al [2] by using FPGA based kit developed AES block cipher algorithm. They suggested that for encryption algorithms field programmable gate arrays were highly advantageous. But the limitation was that the number of rounds increases when the size increases. The main focus is on the implementations which provide high throughput and security.

Ahmed S, Raja M [3] has focused on the information security issues when migrating to a cloud environment. The confidence of end user can be won partially by the guarantee of service provider and they use mainly the cryptographic techniques. The
Intrusion Detection Systems (IDS) should be included in the models which is important to support the forensic study whenever required.

**Albeshri & Caelli W** [5] introduced architecture which is based upon a concept of mutual trust and the specification of definable profiles in vector matrix form. Based on a concept of mutuality, within that cloud computing environment this architecture aims to achieve better, more generic and flexible authentication, authorization and control.

**Chang Lung Tsai & Uei –Chin Lin** [12] developed an efficient method for the detection of intrusion to strength the security in cloud computing systems dynamically. Through multi layers and multi stages deployment many intrusion detectors are dispatched on the network system to detect the intruders.

**W.A. Pauley** [71] discussed about the emerging properties of cloud computing that provides IT capabilities needed by business. The purpose of this study is to develop an instrument for evaluating a cloud provider's transparency of security, privacy and service level competencies via its self-service web portals and web publications and then to empirically evaluate cloud service providers to measure how transparent by using the instrument.

**Huaizhi Yan** [13] reported a study saying private face recognition is a method to solve cloud computing security issues. This method has three parts: user part to provide face images; cloud initialization part has a face subspace and template database; cloud private matching identification part contains the core algorithm of the method, comparing two encrypted numbers under double-encrypted conditions. They developed a credible, efficient, low-complex method to ensure cloud computing security. The experimental result shows that the method can ensure cloud neither know user's real face data nor the face private matching identification result thus making user's face data secure.

**Somani U, Lakhani K** [68] implemented in his research the digital signature with RSA algorithm. In this Research Paper, they have tried to assess Cloud Storage Methodology and Data Security in cloud by the Implementation of digital signature with RSA algorithm.
Cong Wang, Qian Wang [15], in his article “Ensuring data storage security in Cloud Computing” developed an efficient method for identifying the misbehaving server in the cloud computing environment. They introduced the concept of opposing to its predecessors for data file storage security. Data dynamic operations are also provided.

Dawei Sun et al [19] proposed a concept called service level agreements and virtualization techniques. Virtualization technique provides security in the virtual machine. Automated monitoring technique is also introduced.

Gul and Rehman [26], suggested a model that contains of SaaS, PaaS and IaaS. The data is maintained by a third party service provider. Losing of control over data brings security threats, privacy and availability. The protection problem in cloud is significant. IT auditing mechanisms play an important factor in Cloud security policies.

H.C Williams [27] suggested a new encryption algorithm based on RSA public key encryption. He modified the RSA public key algorithm. But this method uses very large prime numbers. The limitation was some mathematical errors were observed.

Hsiao-Ying Lin [30] proposed an effective and flexible method for providing confidentiality in the data. Their method combines the concepts of proxy re-encryption and decentralized erasure code. They also suggested the method to forward the data from one storage server to another server without downloading the data. Their basic concept is based on encoding the information.

Hua Li et al [31] presented a dual core architecture that has done the encryption and decryption at the same time. Round keys for encryption and decryption was provided by 32 bit data path. The main disadvantage was that it requires additional hardware resources.

Popovic et al [57] tells about the growth of cloud computing in IT industry. It also talks about importance of cloud computing in order to regain their fast access to best-of-breed business applications or drastically boosts their infrastructure resources, all at negligible cost once they faced business decline. This paper discusses security issues, requirements and challenges that cloud service providers (CSP) face during cloud engineering.
The first public-key-based searchable encryption scheme was suggested by Hyun Sook Rhea et al [33]. In their construction, anyone with the public key can write to the data stored on the server. The difference here was that only authorized users with the private key can search.

Dimensional security in Cloud Computing paper has been written Sagar Tirodik et al [61] to focus on the problem of data leakage and proposes a framework works in two phases. First phase which is known as Data classification is done by client before storing the data. This paper also categorized the cloud computing on the basis of CIA (Confidentiality, Integrity, and Availability).

The vulnerability in Amazon’s AWS is discussed by Jun Feng and Yu Chen [38] in their article. The three ongoing issues in cloud storage are repudiation, fairness and roll back attacks. A novel fair multi party non-repudiation protocol was suggested by them which are capable of preventing roll-back attacks.

To provide stronger security services more computing space is required and so user designated spaces become less. Large number of logs also makes it difficult to analyze. So to ensure effective usage, trade-offs to be avoided. Jun-Ho Lee, Min-Woo Park [39] in their article. They have suggested a new scheme of log management and detraction system.

Taher Elgamal [70] implemented Diffi-Hellman key distribution scheme based on discrete logarithms. The security depends on computing discrete logarithms over finite fields.

K. Bhatele et al [40] suggested the dual RSA algorithm. They also analyzed the security of the algorithm. Their technique can be used in situation where storage requirement for the keys has to be reduced. Blind signatures and authentication are the main application of dual RSA.
Trust and security have prevented businesses from fully accepting cloud platforms. To protect clouds, the data centers have to uphold privacy and preserve data integrity. **Kai Hwang, Deyi Li** [41] suggested a trust overlay network to be used over multiple data centers. Coloring and watermarks helps protecting data. These safeguard form multi-way authentications and tighten access control for sensitive data.

For the security of the user’s data, the Company has to provide some SLAs. Organizations use cloud computing as a service architecture critically like to examine the security and confidentiality issues for their business critical insensitive applications. The providing of security is not impossible as they provide SaaS, PaaS and IaaS. Each service suffers from their own security issues and hence different levels of security have to be specified and the customer has to understand these. A standardized procedure needs to be made for SLAs. Security issues faced by Cloud have been summarized by **Kandukuri B.R, Paturi V.R.** [43]

**Lei Xu et al** [46] reviewed the security issues in data mining and the various approaches to protect the sensitive information. In their work, they thoroughly studied the privacy preserving data mining (PPDM) concept and gave solution to protect the sensitive information in Big Data.

‘SecCloud’, a novel auditing scheme is proposed by **Lifei Wei & Haojin Zhu** [49] to consider data security and its functions based on the probabilistic sampling and designated verifier technique. It decides on the optimal sampling size thereby minimizing costs.

The heavy load on the thin client of cloud storage sharing arises due to mutual authentication scheme with inherent public key operation. **Lishan Kang, Xuejie Zhang** [51] proposes an identity based authentication, based on non-interactive, identity-based. It exercises mutual authentication by dividing the sharing users into the same domain that relies on sharing global master key. This doubles the user-side performance in both computational and communicating cost.
The Shannon theory approach to cryptography was extended by Martin. E Hellman. [52]. The tradeoff between global and local uncertainty was suggested. The method was matching cipher to a language was also developed. But this method cannot be applicable directly in designing cryptography systems.

Mao-Yin Wang et al [53] developed single and multi pole AES architecture for security. Group of AES processors are the building blocks for their architecture. Interrupt handling load of the host is reduced in this architecture

Nguyen Q, Sood A [55] suggested the method to effectively utilize the resources like storage, memory and CPU time. They explained the challenges and issues in intrusion tolerance to applications and basic services offered by the cloud. Then they have proposed the method called Cloud-based Self-Cleansing Intrusion Tolerant.

Cong Wang [14] gave the solution for searching the keyword in the cipher text. Before outsourcing the inverted index is created, this contains the score of each file for each keyword. They also developed one-to-many order preserving mapping technique to secure the score. In this system to score the files for each keyword, the system will have to go through all the files in the collection. So, creating the posting list takes more time. And also if the data owner has not properly chosen the range size R for mapping the keyword we may get many duplicates. From the duplicates others will be able to get partial information.

Further, to enhance search efficiency, C. Wang, N. Cao, J. Li, K. Ren, and W. Lou [9], in his article “Secure Ranked Keyword Search over Encrypted Cloud Data suggested a per-keyword-based approach, where a single encrypted hash table index is constructed for the full file group of files. Here, each entry consists of the trapdoor of a keyword and an encrypted set of related file identifiers. Searchable encryption has also been considered in the public-key setting.

Boneh et al [16] has presented the public-key-based searchable encryption method. He has presented an equivalent state to that of Song et al. In their structure, any person with the public key can write to the data stored on the cloud server. But here the users
with private key can access or search the data stored in the cloud. The user then downloads the files which are sent by the cloud server.

Song et al [17] introduced the method of searching the data in the encrypted environment. A special two layered encryption scheme is proposed. Each word of the data file is encrypted using the symmetric key method. Thus, a searching cost is linear to the full file group length.

Dan Boneh et al [18] proposed the security scheme for mail server. They have introduced two methods to search keyword in the mail. The first one is identity based encryption and the second one is the public key encryption with search word search.

Goh [24] introduced a Bloom filter-based per-file index, reducing the cost for every search request relative to the number of data files in the collection. A similar per-file index scheme was also developed by Chang and Mitzenmacher [75]

To enhance the search efficiency, Curtmola et al [60] introduced a method called multi keyword search. This method is based on ranking the files. Their basic concept is inner product similarity and co-ordinate matching.

The heavy load on the thin client of cloud storage sharing arises due to mutual authentication scheme with inherent public key operation. Carlo Blundo et al [11] propose an identity based authentication based on non-interactive, identity-based. It exercises mutual authentication by dividing the sharing users into the same domain that relies on sharing global master key. This doubles the user-side performance in both computational and communicating cost.

Hung-Yu Chien [28] proposed time bound hierarchical key management method. He used tamper resistant device for key management scheme. But the implementation cost was very high in this scheme.
Jason H. Li et al [34] suggested an efficient clustering and key management scheme for WSM. The scalability was solved by their method but their method was not suitable for bit cluster size.

A. Juels and B.S. Kaliski Jr [1] proposed Proof of Retrievability scheme, to make sure the data integrity and privacy of the files that the user retrieve. In this scheme the prover or auditor has no knowledge about the files they retrieve. They used the concept of cryptographic proof of knowledge to manage large files.

Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li [10] proposed a suitable method for auditing cloud based data. Their auditing protocol supports dynamic data operations. They have also provided an efficient method of batch auditing for both several cloud and several owners. Initially they have designed the framework for storage system auditing. Then they extended with dynamic auditing.

E. Ateniese et al [23] developed Provable Data Possession model for every block the cloud customer calculate the tag. Then he uploaded the tag and data file to the cloud server. By generating the random challenge on a certain set of file segments, the cloud customer can verify the availability of the data file in the cloud storage. The file blocks are retrieved for verification.

Pa. Varalakshmi, Hamsavardhini Deventhiran [56] proposed Encryption Algorithm to check the integrity of the data. The system consists of three entities, viz., a trusted third party called cloud broker, clients and cloud service provider. In this system, the broker acts as interface between cloud data storage server and clients. Request handling and integrity checking are performed by this entity. For each segment the tag generator creates the tag. Hash method is used to find out the tag.

Q. Wang et al [58] suggested a new method for data storage integrity. They have introduced the third party auditor for verification. Dynamic data modification, deletion and insertion are also discussed. Merkle hash tree algorithm is used for authenticating the block tag. The proposed is included with multiuser setting which uses bilinear aggregate signature.
Sailendra S L, and Arun Kumar Sangaiah [62] introduced the concept called accountability in cloud information. The data used by the user is tracked. It provides the end-to-end distributed accountability. Two phases of auditing called push mode and pull mode is done for accountability. The logs are periodically transmitted to user in push mode where as in pull mode the user gets the log when needed.

Sanchika Gupta et al [63] developed a method based on MD5 function. In this method, first the file is read and compressed. This compressed content is input to the MD5 function which then generates a message digest. This Message digest is encrypted and appended with the original file content which is there within the pre defined tag. For verification, the file has to be read back. File is compressed and after which MD5 is used to generate hash value. This hash value is encrypted and compared with the appended file content in original file.

The method proposed by Wenjun Luo, Guojing Bai [73] for data integrity is based on the RSA algorithm which makes use of big prime numbers. Public key and private keys are used for encryption and decryption. For each of the block authenticators are generated. The main three algorithms are SigGen, GenProof, VerifyProof. SigGen: calculates proof metadata. This function develops the authenticator for every block of the file. GenProof: Run by cloud server to generate a proof of data storage correctness. VerifyProof: Run by the TPA to verify proof from cloud server.

The method suggested by Saravana Kumar R, Ashutosh Saxena [65] is based on XOR operation. First the data file is divided into various blocks. Bits are identified for each of the block from various locations. Another random byte sequence is generated which is then XORed with the previous values. These values are then used for data integrity verification.

Y.Zhu, H.Hu, G.Ahn and M.Yu [76] developed a scheme termed as Cooperative provable data possession for integrity verification in multi-cloud storage”. In this scheme, the file is first divided into blocks and after that it is encoded with error correcting codes. Then check blocks also called as sentinels are embedded for each block. Encryption is done to
ensure that the check blocks are indistinguishable from other file blocks. The verifier challenges the prover by specifying the positions of a collection of sentinels.

**Wajid et al** [72] provides the basic concept of Hadoop MapReduce Framework and various issues in data skew problems. They have also reviewed the various methods of the data skew and the different solutions available to minimize these problems.

**Kin Liu et al** [78] introduced a solution for load balancing in Hadoop framework. They focused on overload machines and developed an improved algorithm for balancing the over load racks.

**Xingwei Wang et al** [74] developed an efficient solution for resource allocation and management in cloud computing environment. They introduced a method called Microeconomic method for resource allocation and management in cloud computing. They also developed a method for the purpose of trading fraud prevention.

**Suppawong Tuarob et al** [69] proposed a new technique to identify and extract algorithm representation in a Big Data environment. To discover algorithm representations, hybrid machine learning approaches were developed. They also discussed the techniques to extract textual metadata for each algorithms.

**Sara Landset et al** [64] did a survey on selecting the appropriate machine learning tools for big data. They have analyzed the advantages and disadvantages of various tools available for big data. They also compared different mechanisms which implement the tools like MapReduce, Spark, Flink and Storms.

**Mukesh Borana et al** [66] suggested the use of Hadoop in health care industry. They proposed a framework which has improved the performance of MapReduce work loads and also maintained the fairness.

One of the important objectives of our research is file retrieval procedure and the proposed method is discussed in detail in our next chapter.