International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 03 - Issue 01 || January 2017 || PP. 28-41

BIG DATA IN SECURE CLOUD COMPUTING WITH SCALABLE DATA SHARING IN TERN MAGNIFYING SMART CITIES

Dr. H. Lilly Beaulah[1], Mrs. Latha P S[2], Mrs. Subaira A S[3]

Professor & Head, Department of Computer Science and Engineering,
Mahendra College of Engineering,Salem.
Assistant professor, Department of Computer Science and Engineering,
Mahendra College of Engineering,Salem.
Assistant professor, Department of Computer Science and Engineering,
Mahendra College of Engineering,Salem.

Abstract: Data has become an essential part of every Economy, Production, Organization, Business function and individual. The amount of data in world is growing day by day because of use of internet, Smartphone, social network, fine tuning of ubiquitous computing and many other technological advancements. Big Data is a term used to identify the datasets that whose size is beyond the ability of typical database software tools to store, manage and analyses. Generally size of the data is Petabyte and Exabyte. Most of the data is partly structured, unstructured or semi structured and it is heterogeneous in nature. Due to its specific nature, Big Data is stored in distributed file system architectures. Hadoop and HDFS by Apache are widely used for storing and managing Big Data.

In the same way, Cloud computing has changed the entire process that distributed computing used to present e.g. Grid computing, server client computing. Cloud computing security is an important aspect of quality of service from cloud service providers. Security concerns arise as soon as one begins to run applications beyond the designated firewall and move closer towards the public domain. In violation of security in any component in the cloud can be disaster for the organization (the customer) as well as for the provider.
Likewise using the cloud storage, users store their data on the cloud without the burden of data storage and maintenance and services and high-quality applications from a shared pool of configurable computing resources. Cryptography is probably the most important aspect of communications security and is becoming increasingly important as a basic building block for computer security, as data sharing is an important functionality in cloud storage.

As a result in this paper, we propose a new approach towards big data in secured cloud computing along with efficient system for scalable data sharing to develop and enhance smart cities.

1. INTRODUCTION:

It's a known fact that Big data is an evolving term that describes any voluminous amount of structured, semistructured and unstructured data that has the potential to be mined for information. Although big data doesn't equate to any specific volume of data, the term is often used to describe terabytes, petabytes and even exabytes of data captured over time. Big Data is a term used to identify the datasets that whose size is beyond the ability of typical database software tools to store, manage and analyses. Due to its specific nature, Big Data is stored in distributed file system architectures. Hadoop and HDFS by Apache are widely used for storing and managing Big Data.

In the same way, Cloud computing is emerged as the modern technology which developed in last few years, and considered as the next big thing, in the years to come. In recent years it has grown up from just being a concept to a major part of IT industry. Cloud computing widely accepted as the adoption of virtualization, SOA and utility computing, it generally works on three type of architecture namely SAAS, PAAS, and IAAS. Cloud computing security is an important aspect of quality of service from cloud service providers. Security concerns arise as soon as one begins to run applications beyond the designated firewall and move closer towards the public domain. In violation of security in any component in the cloud can be disaster for the organization (the customer) as well as for the provider.

And also Cloud storage is gaining popularity recently. In enterprise settings, we see the rise in demand for data outsourcing, which assists in the strategic management of corporate data. It is also used as a core technology behind many online services for personal applications. Considering data privacy, a traditional way to ensure it is to rely on the server to enforce the access control after authentication which means any unexpected privilege escalation will expose all data. In a shared-tenancy cloud computing environment, things become even

worse. Data from different clients can be hosted on separate virtual machines (VMs) but reside on a single physical machine. Data in a target VM could be stolen by instantiating another VM co-resident with the target one. Regarding availability of files, there are a series of cryptographic schemes which go as far as allowing a thirdparty auditor to check the availability of files on behalf of the data owner without leaking anything about the data, or without compromising the data owners anonymity. Likewise, cloud users probably will not hold the strong belief that the cloud server is doing a good job in terms of confidentiality. A cryptographic solution with proven security relied on number theoretic assumptions is more desirable, whenever the user is not perfectly happy with trusting the security of the VM or the honesty of the technical staff. These users are motivated to encrypt their data with their own keys before uploading them to the server. Data sharing is an important functionality in cloud storage. The challenging problem is how to effectively share encrypted data. Of course users can download the encrypted data from the storage, decrypt them, then send them to others for sharing, but it loses the value of cloud storage.

On whole, Here is an approach towards having big data to handle high volume of data in cloud computing where data security is assured with the help of a proposed security model and with the help of proposed aggregate alogorithm, a scalable data sharing is guaranteed. We look forward towards combination of three concepts which leads towards up gradation, advancement, enhancement and improvement in the technologies and magnifying smart cities.

2. BIG DATA ANALYSIS:

Big data is an evolving term that describes any voluminous amount of structured, semistructured and unstructured data that has the potential to be mined for information. Although big data doesn't equate to any specific volume of data, the term is often used to describe terabytes, petabytes and even exabytes of data captured over time.

Such voluminous data can come from myriad different sources, such as business sales records, the collected results of scientific experiments or real-time sensors used in the internet of things. Data may be raw or preprocessed using separate software tools before analytics are applied. Further, big data may involve multiple, simultaneous data sources, which may not otherwise be integrated. For example, a big data analytics project may attempt to gauge a product's success and future sales by correlating past sales data, return data and online buyer review data for that product.

Figure 1 : Big data

The 5Vs that define Big Data are Variety, Velocity and Volume, Variability and Veracity

1) Volume: There has been an exponential growth in the volume of data that is being dealt with. Data is not just in the form of text data, but also in the form of videos, music and large image files. Data is now stored in terms of Terabytes and even Petabytes in different enterprises. With the growth of the database, we need to re-evaluate the architecture and applications built to handle the data.

2) Velocity: Data is streaming in at unprecedented speed and must be dealt with in a timely manner. RFID tags, sensors and smart metering are driving the need to deal with torrents of data in near-real time. Reacting quickly enough to deal with data velocity is a challenge for most organizations.

3) Variety: Today, data comes in all types of formats. Structured, Numeric data in traditional databases. Information created from line-of-business applications. Unstructured text documents, email, video, audio, stock ticker data and financial transactions. We need to find ways of governing, merging and managing these diverse forms of data.

4) Variability: In addition to the increasing velocities and varieties of data, data flows can be highly inconsistent with periodic peaks. Daily, seasonal and event-triggered peak data loads can be challenging to manage. Even more so with unstructured data involved.

5) Veracity: Today's data comes from multiple sources and it is still an undertaking to link, match, cleanse and transform data across systems. However, it is necessary to connect and correlate relationships, hierarchies and multiple data linkages or your data can quickly spiral out of control. A data environment can lie along the extremes on any one of the following parameters, or a combination of them, or even all of them together.

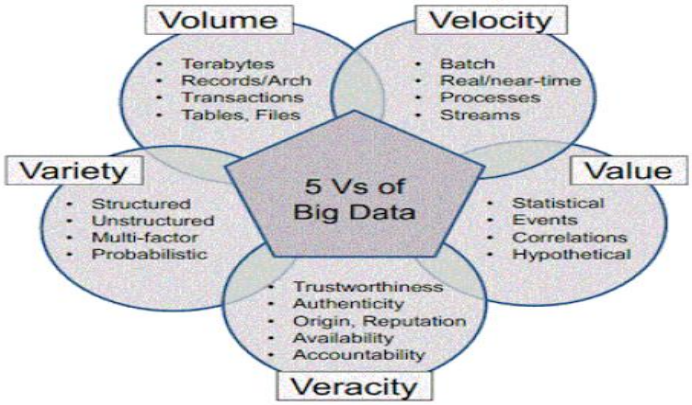

Figure 2 : Characteristics of Big Data

The data in it will be of three types.

- **Structured data** : Relational data.
- **Semi Structured data** : XML data.
- **Unstructured data** : Word, PDF, Text, Media Logs.

2.1 BENEFITS OF BIG DATA:

Big data is really critical to our life and its emerging as one of the most important technologies in modern world. Follow are just few benefits which are very much known to all of us:

- Using the information kept in the social network like Facebook, the marketing agencies are learning about the response for their campaigns, promotions, and other advertising mediums.
- Using the information in the social media like preferences and product perception of their consumers, product companies and retail organizations are planning their production.
- Using the data regarding the previous medical history of patients, hospitals are providing better and quick service.

2.2 BIG DATA TECHNOLOGIES

Big data technologies are important in providing more accurate analysis, which may lead to more concrete decision-making resulting in greater operational efficiencies, cost reductions, and reduced risks for the business.

To harness the power of big data, you would require an infrastructure that can manage and process huge volumes of structured and unstructured data in real-time and can protect data privacy and security.

There are various technologies in the market from different vendors including Amazon, IBM, Microsoft, etc., to handle big data. While looking into the technologies that handle big data, we examine the following two classes of technology:

2.3 OPERATIONAL BIG DATA

This include systems like MongoDB that provide operational capabilities for real-time, interactive workloads where data is primarily captured and stored.

NoSQL Big Data systems are designed to take advantage of new cloud computing architectures that have emerged over the past decade to allow massive computations to be run inexpensively and efficiently. This makes operational big data workloads much easier to manage, cheaper, and faster to implement.

Some NoSQL systems can provide insights into patterns and trends based on real-time data with minimal coding and without the need for data scientists and additional infrastructure.

2.4 ANALYTICAL BIG DATA

This includes systems like Massively Parallel Processing (MPP) database systems and MapReduce that provide analytical capabilities for retrospective and complex analysis that may touch most or all of the data.

MapReduce provides a new method of analyzing data that is complementary to the capabilities provided by SQL, and a system based on MapReduce that can be scaled up from single servers to thousands of high and low end machines.

These two classes of technology are complementary and frequently deployed together.

	Operational	**Analytical**
Latency	1 ms - 100 ms	1 min - 100 min
Concurrency	1000 - 100,000	1 - 10
Access Pattern	Writes and Reads	Reads
Queries	Selective	Unselective
Data Scope	Operational	Retrospective
End User	Customer	Data Scientist
Technology	NoSQL	MapReduce, MPP Database

Table 1: Operational vs. Analytical Systems

2.5 BIG DATA SECURITY AND PRIVACY CHALLENGES

1. Secure Computations in Distributed Programming Framework: Distributed programming framework utilize parallelism in computations and storage to process massive amounts of the data .A popular example is map reduce framework, which splits an input file into multiple chunks in the first phase of map reduce, a mapper for each chunk reads the data, perform some computation, and outputs a list of key/value pairs. In the next phase, a reducer combines the values belonging to each distinct key and outputs the result. There are two major attack prevention measures: securing the mappers and securing the data in the presence of an untrusted mapper.

2. Security Best Practices for Non-Relational Data Stores: Non-relational data stores popularized by NoSQL databases are still evolving with respect to security infrastructure.For instance, robust solutions to NoSQL injection are still not mature each NoSQL DBs were built to tackle different challenges posed by the analytics world and hence security was never part of the model at any point of its design stage.Developers using NoSQL databases usually embed security in the middleware .NoSQL databases do not provide any Support for Enforcing it explicitly in the database. However, clustering aspect of NoSQL databases poses additional challenges to the robustness of such security practices.

3. Secure Data Storage and Transaction Logs: Data and transaction logs are stored in multi-tiered storage media manually moving data between tiers gives the it manager direct control over exactly what data is moved and when. However as the size of data set has been and continues to be, growing exponentially, scalability and availability have necessitated auto-tiring for big data storage management. Auto-tiering solutions do not keep track of where the data is stored, which poses new challenges to secure data storage.

4. End Point Input Validation/Filtering: Many big data use cases in Enterprise settings require data collection from many sources, such as end point devices for example, a security information and event management system (SIEM) may collect event logs from millions of hardware devices and software application in an enterprise network. A key challenge in the data collection process is input validation: how can we trust the data? How can we validate that a source of input data is not malicious and how can we filter malicious input from our collection? Input validation and filtering is a daunting challenge posed by untrusted input sources, especially with the Bring Your Own Device (BYOD) model.

5. Real –Time Security/Compliance Monitoring: Real time security monitoring has always been a challenge, given the number of alerts generated by (Security) devices.These alerts (correlated or not) lead to many false positive, which are mostly ignored or simply‖ clicked away‖ , as humans cannot cope with the shear amount. This problem might even increase with the bid data given the volume and velocity of data streams however, big data technologies might also provide an opportunity, in the sense that these technologies do allow for fast processing and analytics of different types of data.

6. Scalable and Composable Privacy-Preserving Data Mining and Analytics: Big data can be seen as a troubling manifestation of big brother by potentially enabling invasions of privacy, invasive marketing, decreased civil freedoms, and increase state and corporate control. A recent analysis of how companies are leveraging data analytics for marketing purpose identified an example of how a retailer was able to identify that

teenager was pregnant before her father knew. Similarly anonym zing data for analytics is not enough to maintain user privacy. For example AOL released anonymized search logs for academic purposes ,but users were easily identified by their searchers .Netflix faced a similar problem when users of their anonymized data set were identified by correlating their Netflix movie scores with IMDB scores. Therefore, it is important to establish guidelines‟ and recommendations for preventing inadvertent privacy disclosures.

7. Cryptographically Enforced Access Control And Secure Communication: To ensure that the most sensitive private data is end to end secure and only accessible to the authorized entities, data has to be encrypted based on access control policies. Specific research in this area such as attribute-based encryption (ABE) has to be made richer, more efficient, and scalable. To ensure authentication, agreement and fairness among the distributed entities, a cryptographically secure communication framework has to be implemented.

8. Granular Access Control: The security Property that matters from the perspective of access control is secrecy-preventing access to data by people that should not have access .The problem with course-grained access mechanisms is that data that could otherwise be shared is often swept into a more restrictive category to guarantee sound security granular access control gives data managers a scalpel instead of a sword to share data as much as possible without compromising secrecy.

9. Granular Audits: With real time security monitoring, we try to be notified at the moment an attack takes place. In reality, this will not always be the case (e.g., new attacks, missed true positives). In order to get to the bottom of the missed attack, we need audit information. This is not only relevant because we want to understand what happened and what went wrong ,but also because compliance, regulation and forensics reasons .in that regard ,auditing is not something new, but the scope and granularity might be different. For example, we have to deal with more data objects, which probably are (but not necessarily) distributed.

10. Data Provenance: Provenance metadata will grow in complexity due to large provenance graphs generated from provenance- enabled programming environments in big data applications. Analysis of such large provenance graphs to detect metadata dependencies for security/confidentiality applications is computationally intensive.

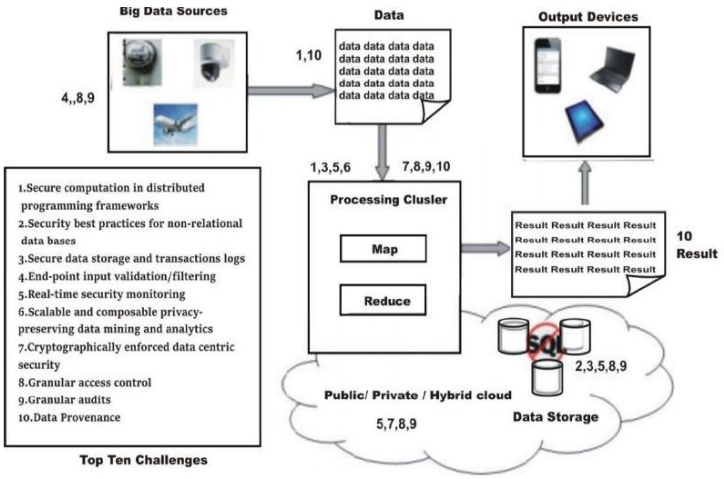

Figure 3: Top ten challenges in Big data

2.6 Traditional Approach

In this approach, an enterprise will have a computer to store and process big data. Here data will be stored in an RDBMS like Oracle Database, MS SQL Server or DB2 and sophisticated softwares can be written to interact with the database, process the required data and present it to the users for analysis purpose.

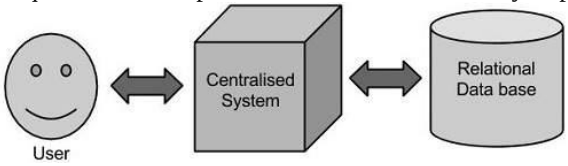

Figure 4: Traditional Approach towards big data.

2.6.1 Limitation

This approach works well where we have less volume of data that can be accommodated by standard database servers, or up to the limit of the processor which is processing the data. But when it comes to dealing with huge amounts of data, it is really a tedious task to process such data through a traditional database server.

2.7 Google's Solution

Google solved this problem using an algorithm called MapReduce. This algorithm divides the task into small parts and assigns those parts to many computers connected over the network, and collects the results to form the final result dataset.

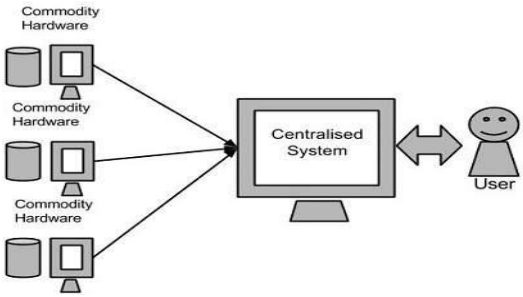

Figure 5: A new approach towards big data [map reduce]

Above diagram shows various commodity hardwares which could be single CPU machines or servers with higher capacity.

2.7.1 Hadoop

Doug Cutting, Mike Cafarella and team took the solution provided by Google and started an Open Source Project called HADOOP in 2005 and Doug named it after his son's toy elephant. Now Apache Hadoop is a registered trademark of the Apache Software Foundation.

Hadoop runs applications using the MapReduce algorithm, where the data is processed in parallel on different CPU nodes. In short, Hadoop framework is capable enough to develop applications capable of running on clusters of computers and they could perform complete statistical analysis for a huge amounts of data.

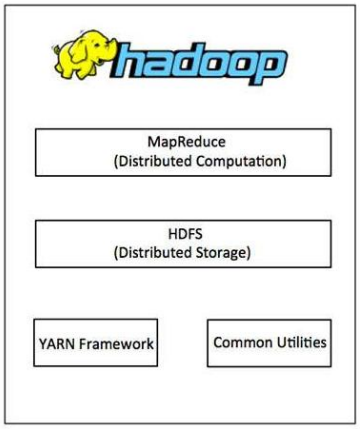

Figure 6: Hadoop frame work

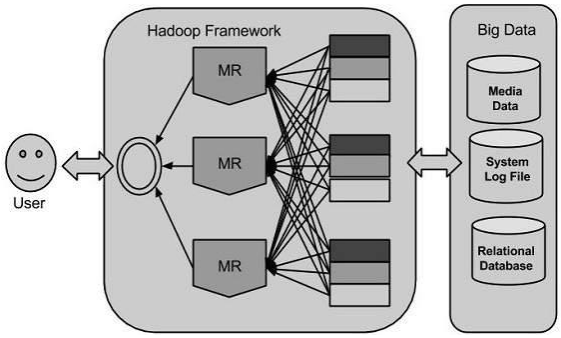

Figure 7: Hadoop frame work for big data

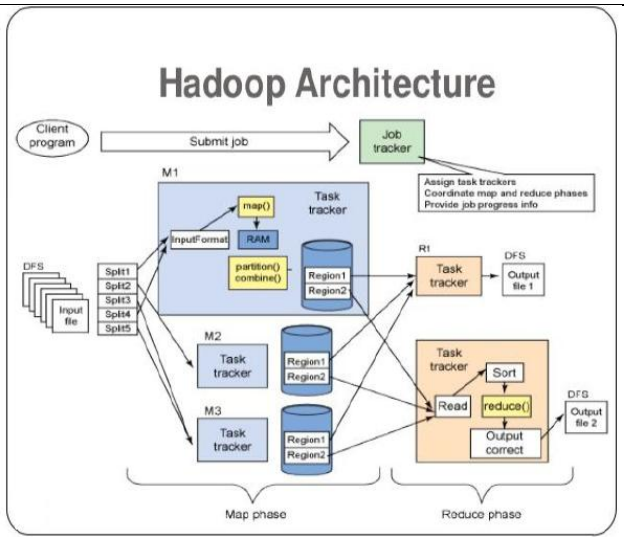

Figure 8: Hadoop architecture for big data

3. SECURE CLOUD COMPUTING MODEL:

Cloud computing is emerged as the modern technology which developed in last few years, and considered as the next big thing, in the years to come. Cloud computing technology requires new security issues and need to face new challenges as well. In recent years it has grown up from just being a concept to a major part of IT industry. Cloud computing widely accepted as the adoption of virtualization, SOA and utility computing, it generally works on three type of architecture namely SAAS, PAAS, and IAAS.

There are different issue and challenges with each cloud computing technology. Security concerns and challenges are addressed in and reviewed in terms of standards such as PCI-DSS, ITIL, and ISO- 27001/27002. Cloud computing has three main aspects: SaaS (software as a service), PaaS (platform as a service) and IaaS (infrastructure as a service). As shown in Figure, SaaS provider hosts and manages a given application in their data centre and makes it available to multiple users over the Web. Oracles CRM on Demand, Salesforce.com are some of the well known SaaS examples. PaaS is an application development and deployment platform which delivered over the web to developers. PaaS facilitates development and deployment of applications without the cost and complexity of buying and managing the underlying infrastructure. All of the facilities required to support the complete life cycle of building and delivering web applications and services entirely available through Internet. This platform includes a database, middleware , development tools and infrastructure software. Well-known PaaS service providers include Google App Engine, Engine Yard. IaaS is the delivery of hardware and software as a service.

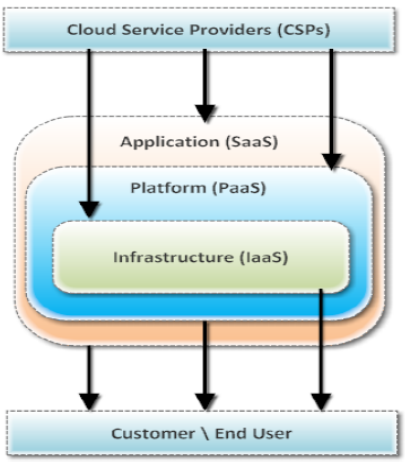

Figure 9: Cloud Computing Environment

3.1 THREATS TO CLOUD COMPUTING

In this section, we discuss threats relevant to the security architecture of Cloud services. We discuss here some potential threats relevant to Cloud and their remedies based on our experience of implementing the cloud.

1. Changes to business model
2. Abusive use of Cloud computing:
3. Insecure interfaces and API
4. Malicious insiders
5. Identity theft
6. Risk profiling
7. Service hijacking
8. Data loss and leakage
9. Shared technology issues/multi-tenancy nature

3.2 ATTACKS ON CLOUD COMPUTING

By exploiting vulnerabilities in Cloud, an adversary can launch the following attacks.

1. Zombie attack
2. Service injection attack
3. Attacks on virtualization
4. *VM Escape*
5. *Rootkit in Hypervisor*
6. Man-in-theMiddle attack
7. Metadata spoofing attack
8. Phishing attack
9. Backdoor channel attack

3.3 PROPOSDED SECURITY MODEL

In this subsection we describe security model for cloud computing against threats mentioned in previous section, which focus on scalability and security. The model is shown in Figure and it consists following security units.

Figure : Security Model for Cloud Computing User can be certificated by the 3rd party certificate authority, then can be issued token for service by End User Service Portal. After joining service portal, user can purchase and use cloud services which are provided by single service provider. End User Service Portal which is composed access

control, security policy, Key management, service configuration, auditing management, and virtual environments provides secure access control using Virtual Private Network (VPN) and cloud service managing and configuration.

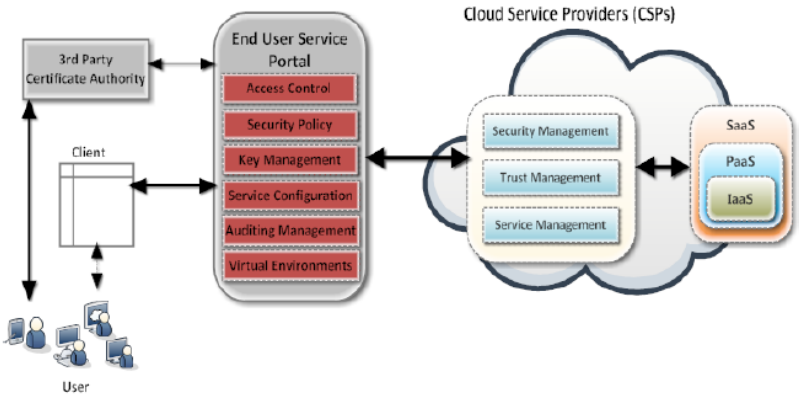

Figure 10: Security Model for Cloud Computing

User can be certificated by the 3rd party certificate authority, then can be issued token for service by End User Service Portal. After joining service portal, user can purchase and use cloud services which are provided by single service provider. End User Service Portal which is composed access control, security policy, Key management, service configuration, auditing management, and virtual environments provides secure access control using Virtual Private Network (VPN) and cloud service managing and configuration.

3.4 FRAMEWORK FOR SECURE CLOUD COMPUTING

The framework for secure cloud computing as shown in figure is based on the security model that will describe the details of each component and apply the needed security technologies for implementation between components in the Cloud Computing. Access control process for providing flexible service on each component is as follows:

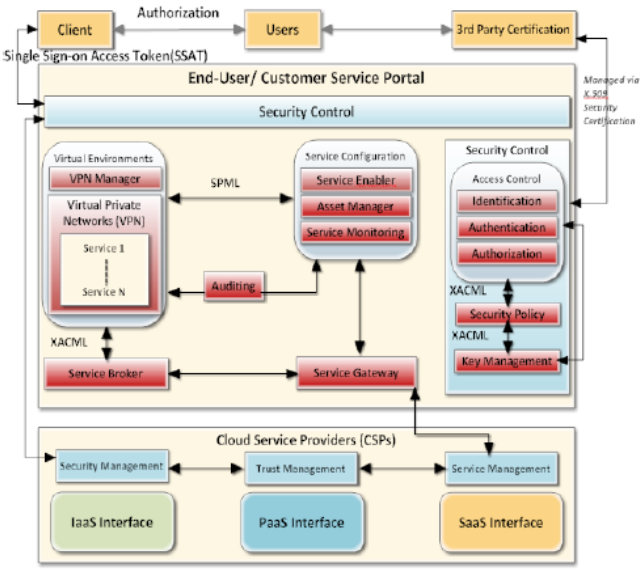

Figure 11: Framework for Secure Cloud Computing

1. **Client**: users could access the client side (i.e.: web browser or host installed application) via diverse devices like PDA, laptop, or mobile phone with Multifactors authentication provided by End-User Service Portal. The client side is the portal where users get their personal cloud. Multi-factors authentication based on certification issued by 3rd party Certification Authority.
2. **End-User Service Portal:** When clearance is granted, a Single Sign-on Access Token (SSAT) could be issued using certification of user. Then the access control component share the user information related with security policy and verification with other components in end-user service portal and cloud service providers by using XACML and KIMP User could use services without limitation of service providers.
3. **Single Sign-on (SSO) :** Currently, Users are having multiple accounts in various Service Providers with different usernames accompanied by different password. Therefore the vast majority of network users tend to use the same password wherever possible, posing inherent security risks. The inconvenience of multiple authentications not only causes users to lose productivity, but also imposes more administrative overhead. Enterprises today are seriously considering the use of Single Sign On (SSO) technology to address the password explosion because they promise to cut down multiple network and application passwords to one. To overcome this problem, it is suggested that, to streamline security management and to implement strong authentication within the cloud, organizations should implement Single Sign- On for cloud users. This enables user to access multiple applications and services in the cloud computing environment through a single login, thus enabling strong authentication at the user level.
4. **Service Configuration:** the service enabler makes provision for personalized cloud service using user's profile. This user's profile is provided to the service management in cloud service provider for the integrationand interoperation of service provisioning requests from user. The SPML can be used to share user's profile. The asset manager requests user's personalized resources with {user's profile} SPML to cloud service provider and configure service via VPN connection.
5. **Service Gateway, Service Broker:** a service gateway manages network resources and VPN on the information lifecycle of service broker.
6. **Security Control:** the security control component provides significant protection for access control, security policy and key management against security threats. Access Control Module is responsible for supporting providers' access control needs. Based on the requirements, various access control models can be used. Role Based Access Control (RBAC) has been widely accepted as the most promising access control model because of its simplicity, flexibility in capturing dynamic requirements, and support for the principle of least

privilege and efficient privilege management. Furthermore, RBAC is policy neutral, can capture a wide variety of policy requirements, and is best suited for policy integration needs discussed earlier. RBAC can also be used for usage control purpose which generalizes access control to integrate obligations and conditions into authorizations. Obligations are defined as requirements that the subjects have to fulfill for access requests. Conditions are environmental requirements independent from subject and object that have to be satisfied for the access request. Due to the highly dynamic nature of the cloud, obligations and conditions are crucial decision factors for richer and finer controls on usage of resources provided by the cloud.

7. **Security Management:** The security management component provides the security and privacy specification and enforcement functionality. The authentication and identity management module is responsible for authenticating users and services based on credentials and characteristics.

8. **Trust Management:** In the cloud, there is a challenging need of integrating requirementsdriven trust negotiation techniques with fine- grained access control mechanisms. Due to the cloud's nature that is service oriented, the trust level should also be integrated with the service. The idea is that the more services a cloud service provider provides the higher trust level needs to be established. Another problem is that we need to establish bidirection trust in the cloud. That is, the users should have some level of trust on the providers to choose their services from, and the providers also need to have some level of trust on the users to release their services to. One possible approach is to develop a trust management approach that includes a generic set of trust negotiation parameters, is integrated with service, and is bi-directional. As the service composition dynamics in the cloud are very complex, trust as well as access control frameworks should include delegation primitives. Existing work related to access control delegation, including role-based delegation, has been focused on issues related to delegation of privileges among subjects and various levels of controls with regard to privilege propagation and revocation. Efficient cryptographic mechanisms for trust delegation involve complex trust chain verification and revocation issues raising significant key management issues with regard to its efficiency.

9. **Service Monitoring:** an automated service monitoring systems to guarantee a high level of service performance and availability. Security framework proposed provides secure connection and convenient to the user for accessing to the cloud service. We consider cloud orchestration environments and Single Sign- On Token to provide seamless experience to user. Furthermore, we provide possible technologies for cloud collaboration.

4. SECURE DATA SHARING IN CLOUD STORAGE:

Cloud storage is gaining popularity recently. In enterprise settings, we see the rise in demand for data outsourcing, which assists in the strategic management of corporate data. It is also used as a core technology behind many online services for personal applications. Considering data privacy, a traditional way to ensure it is to rely on the server to enforce the access control after authentication which means any unexpected privilege escalation will expose all data.

In a shared-tenancy cloud computing environment, things become even worse. Data from different clients can be hosted on separate virtual machines (VMs) but reside on a single physical machine. Data in a target VM could be stolen by instantiating another VM co-resident with the target one . Regarding availability of files, there are a series of cryptographic schemes which go as far as allowing a thirdparty auditor to check the availability of files on behalf of the data owner without leaking anything about the data, or without compromising the data owners anonymity. Likewise, cloud users probably will not hold the strong belief that the cloud server is doing a good job in terms of confidentiality. A cryptographic solution with proven security relied on numbertheoretic assumptions is more desirable, whenever the user is not perfectly happy with trusting the security of the VM or the honesty of the technical staff. These users are motivated to encrypt their data with their own keys before uploading them to the server. Data sharing is an important functionality in cloud storage. The challenging problem is how to effectively

share encrypted data. Of course users can download the encrypted data from the storage, decrypt them, then send them to others for sharing, but it loses the value of cloud storage. Here is an efficient approach to secure the data for scalable data sharing in cloud.

4.1 PROPOSED SYSTEM

The best solution for the above problem is that Alice encrypts files with distinct public-keys, but only sends Bob a

single (constant-size) decryption key. Since the decryption key should be sent via a secure channel and kept secret, small key size is always desirable. For example, we cannot expect large storage for decryption keys in the resource-constraint devices like smart phones, smart cards or wireless sensor nodes.

Especially, these secret keys are usually stored in the tamperproof memory, which is relatively expensive. The present research efforts mainly focus on minimizing the communication requirements (such as bandwidth, rounds of
communication) like aggregate signature. However, not much has been done about the key itself.

4.1.1 ADVANTAGES

- It is more secure.
- Decryption key should be sent via a secure channel and kept secret.
- It is an efficient public-key encryption scheme which supports flexible delegation.

4.1.2 SYTEM ARCHITECTURE DESIGN

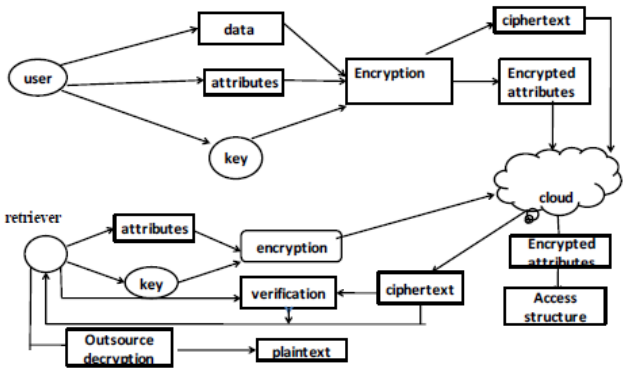

Figure 12: System Architecture Design

4.1.3 PROPOSED DATA SHARING SYSTEM IN CLOUD

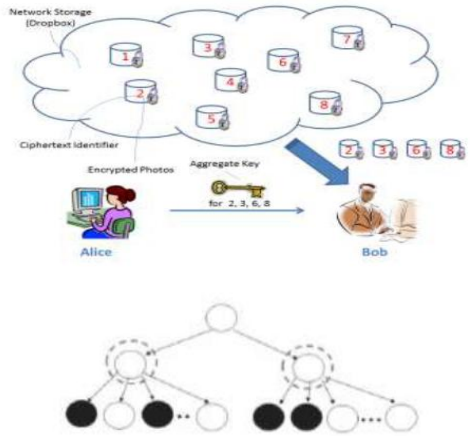

Figure 13: Proposed data sharing system &Key Assignment

5. TECHNIQUES:

5.1 STRING MATCHING ALGORITHM

We show that how to securely, efficiently and flexibly share data with others in cloud storage, Cloud-validation based Flexible Distributed, Migration, ciphertext with aggregate key encryption for data stored in cloud. This scheme provides secure data storage and retrieval. Along with the security the access policy is also hidden for hiding the user's identity. This scheme is so powerful since we use aggregate encryption and string matching algorithms in a single scheme. The scheme detects any change made to the original file and if found clear the error's. The algorithm used here are very simple so that large number of data can be stored in cloud without any problems. The security, authentication, confidentiality are comparable to the centralized approaches. A set of constant size cipher texts such that efficient delegation of decryption rights for any set of cipher texts is possible the best.

5.2 AGGREGATE KEY ENCRYPTION ALGORITHM

A. Framework

The basis or outline of the key-aggregate encryption scheme consists of five polynomial-time algorithms, which are

elucidated below: Setup ensures that the owner of the data can construct the public system stricture or parameter. KeyGen, as the name suggests generates a public/master secret (not to be confused with the delegated key explained later) key pair. By using this public and master-secret key cipher text class index he can convert plain text into cipher text via use of Encrypt.

Using Extract, the master-secret can be utilized to generate an aggregate decryption key for a set of cipher text classes. These generated keys can be safely transported to the appointees by use of secure mechanisms with proper security measures adhered to. If and only if the cipher text's class index is enclosed in the single key, then every user with an aggregate key can decrypt the given cipher text provided through the use of Decrypt.

B. Algorithm

1. Setup(Security level parameter, number of cipher text classes): Setup ensures that the owner of the data can construct the public system stricture or parameter he create account on cloud. After entering the input, the total of cipher text classes n and a security level parameter 1, the public system parameter is given as output, which usually skipped from the input of other algorithms for the purpose of conciseness. 2. KeyGen: it is for generation of public or master key secret pair.

3. Encrypt(public key,index,message):run any person who want to convert plaintext into cipher text using public and

master-secret key

4. Extract(master key, Set): Give input as master secret key and S indices of different cipertext class it produce output aggregate key. This is done by executing extract by the data owner himself. The output is displayed as the aggregate key represented by Ks, when the input is entered in the form the set S of indices relating to the various classes and mastersecret key msk.

5. Decrypt (Ks,S,i,C): When an appointee receives an aggregate key Ks as exhibited by the previous step, it can execute Decrypt. The decrypted original message m is displayed on entering Ks, S, i, and C, if and only if I belongs to the set S.

6. CONCLUSION:

Big data: A broad term that deals with handling of very, very large amount of data which is not possible with traditional systems. This includes collection, storage, search, analysis, visualization of the data. **Cloud computing**: A new paradigm of how we gain access to computing resources. Unlike the traditional captive model, cloud computing lets us "pay as we go" for the computing resources which can expand/contract on real-time based on need and resources can be managed "as a service". **Combine the two, and we have a "made for each other".** While it is possible for big-data to exist without cloud computing, the elastic nature of cloud makes it ideal for big-data projects.

Big Data simply means that huge amount of structured, semi-structured and unstructured data that has the potential to be processed for information. Now a days massive amount of data generated due to technological growth, digitalization and by a variety of sources, including business application transactions, web pages, videos, images, e mails, social media, and soon. So to process these data the big data concept is introduced. **But Cloud computing** means obtain or provide IT resources (Infrastructure, Platform, Software, Database, Storage so on) as service. It provides so many features like: On-demand self-service, Resource pooling, Rapid elasticity, Flexible scaling and High availability etc.

In this paper, big data is considered to handle high volume of data ,in a secured cloud computing where proposed security model assures the data security and also can have secured data sharing with the help of proposed aggregate algorithm. Hence, combination of the above three concepts leads to development, improvement, advancement and up gradation in the technologies towards enhancing the smart cities.

7. REFERENCES

[1]. On technical security issues in cloud computing , Meiko Jensen etal, 2009
[2]. Cloud computing security issues and challenges, Balachandran reddy et al, 2009
[3]. Cloud Computing security issues and challenges Kresimir Popovic, et al, 2010
[4]. Dikaiakos, M.D., Katsaros, D., Mehra, P., et al.: Cloud Computing: Distributed Internet Computing for IT and Scientific Research 13, 10–13 (2009)
[5]. Amazon Web Services. Amazon Virtual private Cloud, http://aws.amazon.com/vpc/

[6]. C. Băsescu, A. Carpen-Amarie, C.Leordeanu, A. Costan, and G. Antoniu, "Managing Data Access on Clouds: A Generic Framework for Enforcing Security Policies", In proceeding of IEEE International Conference on Advanced Information Networking and Applications (AINA), 2011
[7]. R. Sravan Kumar and A. Saxena, "Data integrity proofs in cloud storage", Third International Conference on Communication Systems and Networks (COMSNETS), 2011.
[8]. Z. Wang, "Security and Privacy Issues Within Cloud Computing" IEEE Int. conference on computational and Finformation sciences, Chengdu, China, Oct. 2011.
[9]. James B.D. Joshi, Elisa Bertino, Usman Latif, Arif Ghafoor, "A Generalized Temporal Role-Based Access Control Model",IEEE Computer Society, 2005.
[10]. Moonam Ko, Gail-joon Ahn, Mohamed Shehab, "Privacy enhanced User-Centric Identity Management", IEEE International Conference on Communications,2009.
[11]. Big Data: science in the petabyte era , Nature 455 (7209):1, 2008
[12]. Douglas and Laney, ―The importance of ‗Big Data: A definition ,2008
[13]. Agrawal, Amr El Abbadi et al., Big data and cloud computing: current state and future opportunities , Proceedings of the 14th International Conference on Extending Database Technology, ACM, Sweden, March 21-24, 2011
[14]. http://dashburst.com/infographic/Big-data-volume-variety-velocity/
[15]. http://www.wired.com/insights/2013/05/the-missing-vs-in-Big-data-viability-and-value/
[16]. Lu, Huang, Ting-tin Hu, and Hai-shan Chen. "Research on Hadoop Cloud Computing Model and its Applications , Hangzhou, China: 2012,pp. 59 – 63, 21-24 Oct. 2012.
[17]. Wie, Jiang, Ravi V.T and Agrawal G., "A Map-Reduce System with an Alternate API for Multi-core Environments , Melbourne, VIC: 2010,pp. 84-93, 17-20 May. 2010.International Journal of Network Security & Its Applications (IJNSA), Vol.6, No.3, May 2014
[18]. K, Chitharanjan, and Kala Karun A. "A review on hadoop — HDFS infrastructure extensions , JeJu Island: 2013, pp. 132-137, 11-12 Apr.2013.
[19]. F.C.P, Muhtaroglu, Demir S, Obali M, and Girgin C. "Business model canvas perspective on big data applications", IEEE International Conference, Silicon Valley, CA, pp. 32 – 37, Oct 6-9, 2013.
[20]. Zhao, Yaxiong , and Jie Wu. "Dache: A data aware caching for big-data applications using the Map Reduce framework." INFOCOM, 2013 Proceedings IEEE, Turin, pp. 35 – 39, Apr 14-19, 2013.
[21]. S. S. M. Chow, Y. J. He, L. C. K. Hui, and S.-M. Yiu, "SPICE - Simple Privacy-Preserving Identity-Management for Cloud Environment," in Applied Cryptography and Network Security – ACNS 2012, ser. LNCS, vol. 7341. Springer, 2012, pp. 526–543.
[22]. L. Hardesty, "Secure computers aren't so secure," MIT press, 2009, http://www.physorg.com/news176107396.html.
[23]. C. Wang, S. S. M. Chow, Q. Wang, K. Ren, and W. Lou, "Privacy- Preserving Public Auditing for Secure Cloud Storage," IEEE Trans. Computers, vol. 62, no. 2, pp. 362–375, 2013.
[24]. B. Wang, S. S. M. Chow, M. Li, and H. Li, "Storing Shared Data on the Cloud via Security-Mediator," in International Conference on Distributed Computing Systems - ICDCS 2013. IEEE, 2013.
[25]. S. S. M. Chow, C.-K. Chu, X. Huang, J. Zhou, and R. H. Deng, "Dynamic Secure Cloud Storage with Provenance," in Cryptography and Security: From Theory to Applications - Essays Dedicated to Jean-Jacques Quisquater on the Occasion of His 65th Birthday, ser. LNCS, vol. 6805. Springer, 2012, pp. 442–464.
[26]. D. Boneh, C. Gentry, B. Lynn, and H. Shacham, "Aggregate and Verifiably Encrypted Signatures from Bilinear Maps," in Proceedings of Advances in Cryptology - EUROCRYPT '03, ser. LNCS, vol. 2656. Springer, 2003, pp. 416–432.
[27]. M. J. Atallah, M. Blanton, N. Fazio, and K. B. Frikken, "Dynamic and Efficient Key Management for Access Hierarchies," ACM Transactions on Information and System Security (TISSEC), vol. 12, no. 3, 2009.
[28]. J. Benaloh, M. Chase, E. Horvitz, and K. Lauter, "Patient Controlled Encryption: Ensuring Privacy of Electronic Medical Records," in Proceedings of ACM Workshop on Cloud Computing Security (CCSW '09). ACM, 2009, pp. 103–114.
[29]. F. Guo, Y. Mu, Z. Chen, and L. Xu, "Multi-Identity Single- Key Decryption without Random Oracles," in Proceedings of Information Security and Cryptology (Inscrypt '07), ser. LNCS, vol. 4990. Springer, 2007, pp. 384–398.
[30]. V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute- Based Encryption for Fine-Grained Access Control of Encrypted data," in Proceedings of the 13th ACM Conference on Computer and Communications Security (CCS '06). ACM, 2006, pp. 89–98.

[31]. S. G. Akl and P. D. Taylor, "Cryptographic Solution to a Problem of Access Control in a Hierarchy," ACM Transactions on Computer Systems (TOCS), vol. 1, no. 3, pp. 239–248, 1983.