A Survey on Security and Privacy in Cloud Computing for Big Data Process

R.Kalaivani[1] Dr. A.Viswanathan[2]
[1,2]Department of Computer Science & Engineering
[1,2]K.S.R College of Engineering, Tamilnadu, India

Abstract— Cloud Computing promises a scalable communications for meting out big data applications. It is web based computing, used to share resources to a computers. The phrase big data has expand hastily in the framework of data mining and business brains. The need of sheltered big data storage service is more enviable than ever to date. We must highlight that big data does not just involve large volumes of data also the stipulation for scalability, to guarantee a response in a conventional elapsed time. This model is cloud computing, and among its main features we has to hassle its flexibility in the use of computing resources and spaces, less management effort, and bendy cost. We edge on those systems for wide analytics based on the Map Reduce scheme, its open-source implementation. In this paper, we first explain the security model of cloud computing, and then analyze the feasibility, threats and security in cloud computing in terms of wide existing methods to control them along with their pros and cons.

Key words: Cloud, Big Data, Map Reduce

I. INTRODUCTION

In recent years, Cloud Computing and big data receives mammoth attention internationally due to assorted business-driven promises and potential such as lower upfront IT cost, a faster time to market and opportunities for creating value-add business. As the newest computing hypothesis, cloud is branded by delivering hardware and software resources as virtualized services by which users are free from the load of acquiring the low-level system supervision details. Cloud computing promises a scalable infrastructure for giving out big data applications such as the analysis of vast amount of data.

Data is recorded everyday resulting in a large volume of information; this incoming information arrives at a high rate and its dispensation involves real-time desires implying a high velocity; we may find a wide variety of structured, semi-structured, and unstructured data; and data have to be cleaned before the integration into the system in order to maintain veracity. This four property is one of the most widespread definitions of what is known as the big data problem, which has become a hot topic of interest within university and corporations.

- On-demand self services: A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed routinely without requiring human interaction with each service's provider.
- Broad network access: Capabilities are available over the network and accessed through benchmark mechanisms that endorse use by diverse thin or thick client platforms.
- Resource pooling: The cloud enables your workforce to enter and use data within the business management software hosted in the cloud at the same time, from any location, and at any time. This is an attractive feature for numerous business offices and field service or sales teams that are usually exterior the office.
- Rapid elasticity: If anything, the cloud is flexible and scalable to suit your immediate business needs. You can quickly and easily add or remove users, software features, and other resources.
- Measured service: Going back to the affordable nature of the cloud, you only pay for what you use. You and your cloud provider can measure storage levels, processing, bandwidth, and the number of user accounts and you are billed appropriately. The amount of resources that you may use can be monitored and controlled from both your side and your cloud provider's side which provides transparency. These are the five personalities in cloud computing.

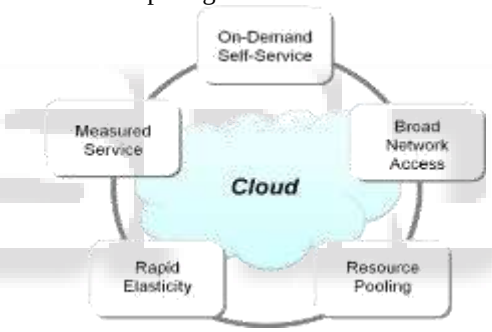

Fig. 1: personalities in cloud computing.

The main cloud computing applications are given below,

1) Infrastructure as a service and platform as a service.
2) Private cloud and hybrid cloud.
3) Test and development.
4) Big data analytics.
5) File storage.
6) Disaster recovery.
7) Back up.

II. RELATED WORK

Wanchun Dou et al [1] proposed the cross clouds are formed with the private cloud data resources and public cloud service components. Cross cloud service composition provides a tangible loom capable for large scale big data processing. Private cloud refuses to unveil all details of their service transaction records. History record based service optimization method (Hiresome-II) is a privacy aware cross cloud service composition method. QoS history records are used to guess the cross cloud service composition plan. K-means algorithm is used as a data filtering tool to pick representative history records. It reduces the time

complexity of cross cloud service composition plan for big data.

Kaitai Liang et al [2] described the system shares the big data with privacy and security shield between senders and recipients. Privacy-preserving ciphertext multi-sharing machinery is used to attain unsigned data sharing with service clients. Data sharing is achieved with cipher text attack control mechanism. Cross cloud environment based data sharing is not supported. This research work aims to solve the above problems. To preserve anonymity, some well-known encryption mechanisms are anticipated in the fiction, such as anonymous. By employing these primitives, the foundation and the target of data can be cosseted secretly.

Xuyun Zhang et al [3] designed the system manages the big data sharing with proximity aware local recording anonymization mechanism. Scalable two-phase clustering approach and proximity-aware agglomerative clustering algorithm are used to share big data with privacy. Privacy preserved big data mining operations are not supported. A practical and widely-adopted technique for data privacy preservation is to anonymize data via sweeping statement to satisfy a given privacy model.

Joonsang Baek et al [4] described big data collection from electricity services is shared with security and privacy in clouds. Identity-based encryption, signature and proxy re-encryption methods are incorporated in secure cloud computing based framework for big data process. Cross cloud based service composition is not supported. The main idea of our framework is to build a hierarchical constitution of cloud computing centers to provide dissimilar types of computing services for information management and big data analysis.

Xiaoyong Li et al [5] proposed that the system manages the trust verification between the user, broker and service resources in multiple cloud environment. Service operator-aware trust scheme(SOTS) evaluates the trust using multi attribute based model. Resource security level assessment is not optimized. This adaptive loom can overcome the margins of traditional trust schemes, whereby the trusted operators are slanted physically or intuitively.

Rajiv Ranjan et al [6] described the cloud resource provisioning methods are used to build big data applications. Iterative ordinal optimization (IOO) method is used to assemble big data applications with high scalability in clouds. Privacy and security features are not supported. Virtualized clouds introduce concert variability in resources. Elasticity has how become the rudimentary feature of cloud computing as it enables the ability to energetically add or remove virtual machine instances when workload changes.

Yanfeng Zhang et al [7] proposed the result of data mining become fusty and archaic over time. Incremental processing is a promising loom to uplifting mining results. Iterative map reduce models are used to execute mining on big data atmosphere. Incremental processing extension to Map Reduce framework is used for mining big data. Private cloud based data sharing is not personalized with the mining model. Fine-grain incremental processing using MRBG-store, General-purpose iterative computation with modest extension to Map Reduce API, Incremental processing for iterative computation these are the three novel features. The basic idea for incremental computation for Map is straightforward. We simply appeal to the Map function for the inserted or deleted. K means is a commonly used clustering algorithm that partitions points into k clusters.

Luis M.Vaquero et al [8] dercribed the virtual machines shares and enthusiastically loads the big data values based on the user desires. Big data provisioning service incorporates hierarchical and peer-to-peer data sharing techniques to speed-up data loading into the VMs used for data processing. Data transfer scheduling is not supported. The sequence of task needed to prepare a big data for parallel analysis on a set of newly deployed VMs is as follows: partitioning, Data distribution, Application configuration, Load data in memory. There is four approach are implemented in data distribution centralised approach, semi-centralised approach, hierarchical approach, P2P approach.

Weikuan Yu et al [9] proposed the virtual shuffling strategy is used to enable proficient data movement and reduce I/O for map reduce shuffling. Map Reduce uses shuffling period to internationally exchange the intermediary data generated by the mapping phase. Multi cloud data movement is not supported. Virtual shuffling is realized through a mixture of three techniques including a three-level segment table, near-demand merging, and dynamic and balanced merging sub trees. MapReduce is popularized by Google as a very simple but influential program model that offers parallelized computation, fault-tolerance and distributing data processing.

Qi Zhang et al [10] designed task level scheduling schemes are used to allocate resources for map reduce. Phase and Resources information-aware scheduler for Mapreduce clusters (PRISM) framework is used to allocate resources. Security and privacy factors are not considered. Map Reduce can significantly reduce the running time of data-intensive jobs.

III. Problem Identification

The problem identified in hiresome-II is there is no big data processing and also not integrated with the system. There is no security and privacy for big data. For big data process there is limited scalability. Mining operations are not integrated. For proximity-aware the problem is privacy preserved big data mining operations are not supported. A secure cloud computing based framework for big data information management, cross cloud based service composition is not supported in this paper. In service operator-aware trust there is no resource security level assessment. In recent advances in autonomic provisioning of big data application on cloud the problem is there is no privacy and security. In i2 mapreduce there is no adaptation of private cloud data sharing with mining model. In deploying large-scale datasets data transfer scheduling is not supported. In virtual shuffling there are no multi cloud data movements. In PRISM there is no security and privacy factor.

IV. Conclusion

In this survey there is many optimization method based on security and privacy in cloud computing with big data process. In cloud environment, the privacy preservation for data analysis, share and mining is a challenging research

issue due to increasingly larger volumes of datasets, thereby requiring intensive investigation. We have to overcome the time complexity and also protect cloud privacy. Simulation and analytical results have demonstrated the validity of our method.

REFERENCES

[1] Wanchun Dou, Xuyun Zhang, Jianxun Liu and Jinjun Chen, "HireSome-II: Towards Privacy-Aware Cross-Cloud Service Composition for Big Data Applications", IEEE Transactions On Parallel And Distributed Systems, Vol. 26, No. 2, February 2015.
[2] Kaitai Liang, Willy Susilo and Joseph K. Liu,"Privacy-Preserving Ciphertext Multi-Sharing Control for Big Data Storage", IEEE Transactions On Information Forensics And Security, Vol. 10, No. 8, August 2015
[3] Xuyun Zhang, Wanchun Dou, Jian Pei, Surya Nepal, Chi Yang, Chang Liu and Jinjun Chen, "Proximity-Aware Local-Recoding Anonymization with MapReduce for Scalable Big Data Privacy Preservation in Cloud", IEEE Transactions On Computers, Vol. 64, No. 8, August 2015.
[4] Joonsang Baek, Quang Hieu Vu, Joseph K. Liu, Xinyi Huang and Yang Xiang, "A Secure Cloud Computing Based Framework for Big Data Information Management of Smart Grid", IEEE Transactions On Cloud Computing, Vol. 3, No. 2, April/June 2015.
[5] Xiaoyong Li, Huadong Ma, Feng Zhou and Xiaolin Gui , "Service Operator-Aware Trust Scheme for Resource Matchmaking across Multiple Clouds", IEEE Transactions On Parallel And Distributed Systems, Vol. 26, No. 5, May 2015.
[6] Rajiv Ranjan, Lizhe Wang, Albert Y. Zomaya, Dimitrios Georgakopoulos, Xian-He Sun and Guojun Wang, "Recent Advances in Autonomic Provisioning of Big Data Applications on Clouds", IEEE Transactions On Cloud Computing, Vol. 3, No. 2, April/June 2015
[7] Yanfeng Zhang, Shimin Chen, Qiang Wang and Ge Yu,"i2MapReduce: Incremental MapReduce for Mining Evolving Big Data", IEEE Transactions On Knowledge And Data Engineering, Vol. 27, No. 7, July 2015
[8] Luis M. Vaquero,Antonio Celorio, Felix Cuadrado and Ruben Cuevas, "Deploying Large-Scale Datasets on-Demand in the Cloud: Treats and Tricks on Data Distribution", IEEE Transactions On Cloud Computing, Vol. 3, No. 2, April/June 2015.
[9] Weikuan Yu, Yandong Wang, Xinyu Que and Cong Xu, "Virtual Shuffling for Efficient Data Movement in MapReduce",IEEE Transactions On Computers, Vol. 64, No. 2, February 2015.
[10] Qi Zhang,Mohamed Faten Zhani, Yuke Yang, Raouf Boutaba and Bernard Wong "PRISM: Fine Grained Resource-Aware Scheduling for MapReduce", IEEE Transactions On Cloud Computing, Vol. 3, No. 2, April/June 2015

