Analyzing the Performance of Searchable Symmetric Encryption over Huawei Cloud

Matheus M. Silveira[1], **Danielle S. Silva**[1],
Luiz Z. Oliveira[2], **Rafael L. Gomes**[1]

[1] Universidade Estadual do Ceará (UECE), Fortaleza, Ceará, Brasil.

{matheus.monteiro,danielle.santos}@aluno.uece.br,rafa.lopes@uece.br

[2]Huawei Brasil, São Paulo, SP, Brasil.

luiz.zucas@huawei.com

Abstract. *Cloud computing environments store data from companies and end users, bringing reliability and scalability to this remote access. Additionally, encryption techniques have been applied to protect this data in the cloud, where Searchable Symmetric Encryption (SSE) enables a fast search of keywords in encrypted databases. However, the configuration of the cloud environment directly impacts the processing time of SSE and, consequently, affects Service Level Agreement (SLA) requirements. Within this context, this paper analyzes the impact of Huawei Cloud configuration over the SSE task, encompassing the possible existing options regarding computing and architectures available. The experiments performed using a real database over the Huawei Cloud suggest the most suitable configuration according to the requirements needed to meet.*

1. Introduction

Most of the existing business models are improved when executed over Cloud Computing environments since they benefit from the advantages of scalability, reliability, elasticity, measured services, accessibility, etc. These improvements support fulfilling requirements defined in Service Level Agreements (SLAs) [Gomes et al. 2020]. Among the most popular services offered by cloud providers is data storage, usually based on Relational Database Service (RDS) or Elastic Cloud Server (ECS).

However, this approach of online services may expose due to existing technology vulnerabilities and compromised involved parties, resulting in a scenario where these companies, government institutions, and end users are subject to intrusion attempts and possible data leakage situations [Gupta et al. 2022]. Several cases of data leakage have occurred around the world in the last few years. For example, Marriott hotels had the data of 500 million customers accessed by hackers [Yu et al. 2022]. Among other victims of the attacks are companies such as T-Mobile, Quora, Google, Orbitz, and Facebook, which faced significant breaches and incidents that affected more than 100 million users.

In this way, encryption techniques are applied to bring confidentiality and protection to the data in the cloud [Mosca et al. 2023]. These two issues are crucial due to the necessity of the companies to attend to the existing privacy laws [Gong et al. 2022], such as General Data Protection Regulation (GDPR) in Europe and Lei Geral de Proteção de Dados (LGPD) in Brazil. Thus, companies and government institutions, need to

comply with the points listed by privacy laws, since this may have an impact on business, when dealing with data from their customers and employees, at the time of making the data portable, when cooperating internationally, etc. , Data encryption techniques demand processing time, harming the possibility of frequent encryption and decryption of huge databases and increasing the time to perform searches and retrievals [Aparajit et al. 2022]. An alternative approach is the usage of Searchable Symmetric Encryption (SSE) [Li et al. 2019, Silveira et al. 2023], which is a technique to search for keywords in encrypted databases and retrieve this specific data, without the necessity to decrypt all the database. However, the existing studies in the literature do not perform an analysis of the performance of SSE solutions according to the configuration of the cloud environment.

Within this context, this paper analyzes the impact of Huawei Cloud configuration over the SSE solutions [Poh et al. 2017], encompassing the possible existing options regarding computing and architectures available. The experiments performed using a real database over the Huawei Cloud suggest the most suitable configuration according to the requirements needed to meet an SLA.

The remainder of this paper is organized as follows. Section 2 describes the SSE approach, while Section 3 present the experimental analysis performed and their results. Finally, section 4 concludes the paper and presents future works.

2. Searchable Symmetric Encryption (SSE) over Cloud Computing

An SSE solution consists of a request of a Trapdoor w made by the client (data owner) to the Cloud service provider (server) that will return a list of the index of the documents that contain w. A general SSE encryption scheme algorithm, that are [Li et al. 2019]: (A) Keygen(s), it is an algorithm that should run in the client side to generate a master key MK based on a security parameter; (B) Trapdoor(MK, w), which is an algorithm that should by run by the client by taking the master key MK and a keyword w as the input, and outputs the trapdoor T_w of word w; (C) BuildIndex(R, MK), that is an algorithm that should be run by the client by taking the master key MK and a record R as the input, and outputs the index IR for record R; and, (D) Search(T, I), it is an algorithm that should be run by the server by taking a trapdoor Tw and a document's index IR as the input, and outputs 1 if w ϵ R or 0 otherwise.

The generated encrypted data, along with the associated indexes of all the keywords, are in general kept in safety by the server. Thus, using the private-key encrypted data searching method, the server will only be accessed by using the given access key to it. To search in the encrypted database generated by using the build function, the algorithm will re-calculate the same keys used to previously encrypt the data. That means, in the search function we will have the same encryption functions used to build the database to make it possible re-generating the original cipher for the given keyword. After acquiring this AES cipher, the function will look for matches in all the required tables, saving the matches' indexes. Finally, a list with all the matched identifiers will be returned to the client.

The search process will be requested by the client when they provide a query to a required database containing keyword information. After this, the server will receive this query and run the SSE search algorithm to respond with a list of all of the matched iden-

tifiers in the required database that contain that keyword. It is the safer way to guarantee that clients will receive what they look for and still will not compromise the security of the stored sensitive data.

3. Experiments

This section presents the experiments performed to evaluate the performance of the SSE over Huawei Cloud. These experiments considered a real case study to enable a suitable evaluation of the tool and its impact. Subsection 3.1 presents the testbed configuration of the experiments, whereas Subsection 3.2 discusses the results.

3.1. Cloud Configuration

A number of experiments were made by varying the size of the databases used and the number of keyword queries made in each one. As cloud environment, we used several configuration of Elastic Cloud Server (ECS)[1] in Huawei Cloud[2] and, for the sake of comparison, two physical machines. It is worth mentioning that all the Huawei Cloud ECSs deployed were configured to 4 Virtual CPUs (vCPUs) and 16Gb of memory (the exception was Disk-intensive, which has a minimum of 32Gb). Therefore, the following processing cases were considered:

- Huawei Cloud - General Computing Plus (GCP - C3 version): It provides a balance of computing, memory, and network resources and a baseline level of vCPU performance with the ability to burst above the baseline.
- Huawei Cloud - Memory Optimized (MO - M3 version): It has a large memory size and provides high memory performance, being designed for memory-intensive applications that process large volumes of data.
- Huawei Cloud - High-Performance Computing (HPC - H1 version): It is suitable for applications that require a large number of parallel computing resources and high-performance infrastructure services to meet computing and storage requirements and ensure high rendering efficiency.
- Huawei Cloud - Disk-Intensive (DI - D3 version): It is delivered with local disks for high storage bandwidth and IOPS, i.e., it uses local disks to provide high sequential read/write performance and low latency, while providing powerful and stable computing capabilities, ensuring efficient data processing.
- Huawei Cloud - Kunpeng (KPG - KC1 version): It uses the Kunpeng 920 processors to cost-effectively provide a baseline level of vCPU performance with the ability to burst above the baseline, meeting the requirements of migrating infrastructure services to the cloud.
- Physical Machine (PM): Day-use machine for general tasks, composed of CPU Intel Core i5-12400, 8GB DDR4 2666MHz of memory and SSD Disk of 256GB.

Regarding the database, it was created using the Python Faker Package[3] to generate the data by accessing properties named after the type of data in the generator initialized. We varied the size of the database from 100 to 4000 rows (the number of columns was fixed to 20) and the search requests from 100 to 1000.

[1] support.huaweicloud.com/intl/en-us/productdesc-ecs/ecs_01_0073.html

[2] huaweicloud.com

[3] pypi.org/project/Faker/

Thus, by controlling the database size we can configure several different case studies and analyse the performance of the SSE over the Cloud. During the database population all the possible keywords found in it are also saved, that will be used to perform client requests in the cloud. During the experiments the searching time is considered as evaluation metric, i.e., the time to request, search, retrieve and return the data to the end user.

3.2. Results and Discussion

This subsection presents the experimental results we conducted to evaluate the performance of SSE over Huawei Cloud. Figure 1 present the results of all cases evaluated. In general, the behavior of the search time is linear for most requests performed by the client, where the differences in searching times become more significant in high-demand scenarios (the size of the databases is more than 2000).

It is possible to note that the critical issue regarding SSE is the processing capacity of the cloud environment since all the operations are performed over the memory with sporadic reads from disk. This fact can be observed in the HPC case (with Intel Xeon Scalable Processor 4.2GHz), which reached the best results (almost half of the time when compared to the Disk-intensive case with Intel Xeon Processor E5 2.6GHz). Another interesting point is the advantage of Memory Optimized over GCP (around 12%), which occurred due to its design for memory-intensive solutions because both types have the same CPU specification (Intel Xeon Processor 3GHz).

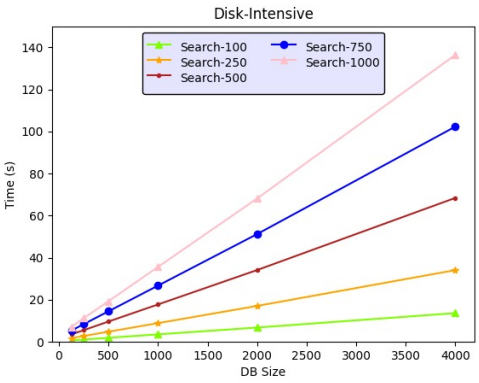
The results indicate that in the cases of a higher number of search requests with big databases, it is necessary to apply HPC or MO types, enabling better processing of the search requests. On the other hand, it is possible to note that a small number of requests, regardless of the size of the database, have a suitable searching time, enabling all the types in the cloud to be applied. Another important issue is the comparison between the physical machine and cloud virtual machines, where GPC and KPG have a similar behavior of PM, highlighting the applicability of cloud environments for SSE since it has a similar performance, but it includes the advantages of scalability, reliability, etc. Additionally, it is interesting the similar processing capacity of Huawei Kunpeng processor (in KPG) and Intel Xeon Scalable Processor (in GPC), as stated in Figures 1(e) and 1(c). Based on the experiments performed, it was possible to identify the most suitable cloud configuration according to the requirements needed to meet an SLA.

4. Conclusion and Final Discussion

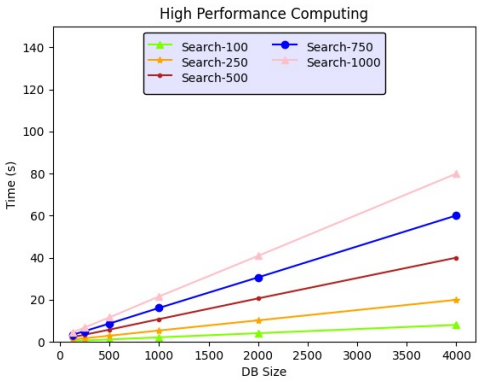
Usually, companies and organizations aim to protect their data in the cloud through encryption approaches, such as SSE, but it is necessary to understand the computational cost of these approaches in the processing of the services in the cloud. Therefore, this paper presented a study about the performance of SSE technique according to the configuration of Huawei Cloud environment, where it was possible to identify the most suitable cloud configuration according to the requirements needed to meet an SLA.

Acknowledgments

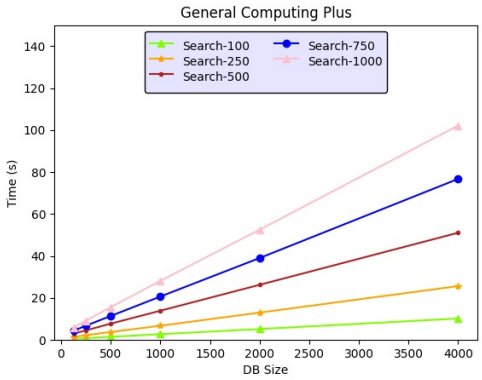
The authors would like to thank National Council for Scientific and Technological Development (CNPq) of Brazil (N^o 303877/2021-9) and Huawei Brazil for the financial support.



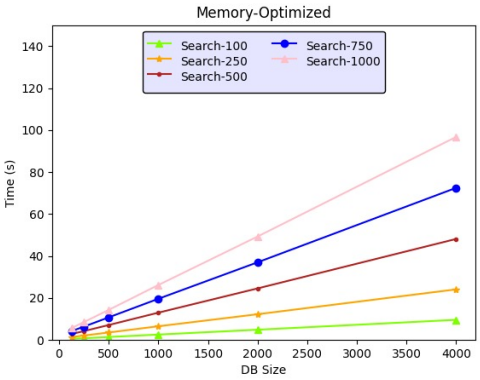
(a) Disk-Intensive (DI)



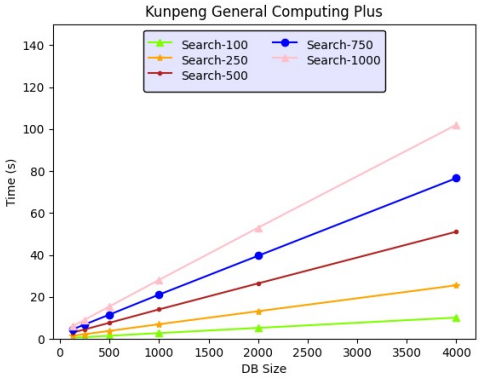
(b) High Processing Computing (HPC)



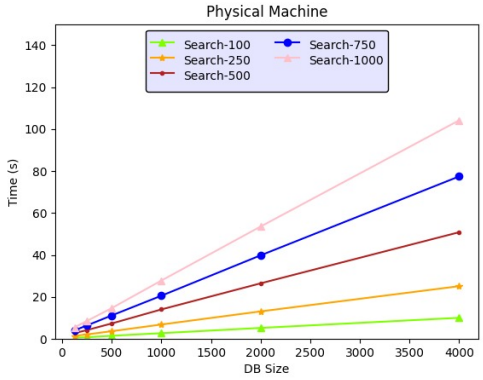
(c) General Computing Plus (GPC)



(d) Memory Optimized (MO)



(e) Kunpeng (KPG)



(f) Physical Machine (PM)

Figura 1. Results of SSE over Huawei Cloud

Referências

Aparajit, S., Shah, R., Chopdekar, R., and Patil, R. (2022). Data protection: The cloud security perspective. In *2022 3rd International Conference for Emerging Technology (INCET)*, pages 1–5.

Gomes, R. L., Bittencourt, L. F., and Madeira, E. R. M. (2020). Reliability-aware network

slicing in elastic demand scenarios. *IEEE Communications Magazine*, 58(10):29–34.

Gong, X., Chen, Y., Wang, Q., Wang, M., and Li, S. (2022). Private data inference attacks against cloud: Model, technologies, and research directions. *IEEE Communications Magazine*, 60(9):46–52.

Gupta, I., Singh, A. K., Lee, C.-N., and Buyya, R. (2022). Secure data storage and sharing techniques for data protection in cloud environments: A systematic review, analysis, and future directions. *IEEE Access*, 10:71247–71277.

Li, J., Huang, Y., Wei, Y., Lv, S., Liu, Z., Dong, C., and Lou, W. (2019). Searchable symmetric encryption with forward search privacy. *IEEE Transactions on Dependable and Secure Computing*, 18(1):460–474.

Mosca, E. E. P., Ribeiro, S., Urbano, A., Silva, D. S., and Gomes, R. L. (2023). Evaluation of security techniques in heterogeneous iot devices. In *Proceedings of the 11th Latin-American Symposium on Dependable Computing*, LADC '22, page 91–94, New York, NY, USA. Association for Computing Machinery.

Poh, G. S., Chin, J.-J., Yau, W.-C., Choo, K.-K. R., and Mohamad, M. S. (2017). Searchable symmetric encryption: designs and challenges. *ACM Computing Surveys (CSUR)*, 50(3):1–37.

Silveira, M. M., Silva, D. S., Rodriguez, S. J. R., and Gomes, R. L. (2023). Searchable symmetric encryption for private data protection in cloud environments. In *Proceedings of the 11th Latin-American Symposium on Dependable Computing*, LADC '22, page 95–98, New York, NY, USA. Association for Computing Machinery.

Yu, J., Moon, H., Chua, B.-L., and Han, H. (2022). Hotel data privacy: strategies to reduce customers' emotional violations, privacy concerns, and switching intention. *Journal of Travel & Tourism Marketing*, 39(2):213–225.