MODELLING AND SECURING DATA TRANSMISSION FOR DATA-INTENSIVE TASKS IN HETEROGENEOUS CLOUD SYSTEMS

Oleksii Leunenko
Simon Kuznets Kharkiv National University of Economics

DOI 10.24917/9788368020861.25

Abstract

In this intense era of Big Data, the role of heterogeneous, or multi-cloud, technologies is exceedingly pivotal. This article explores the complexities of deploying simultaneous use of multiple leading cloud resources, while underlining the advancements these technologies offer in terms of flexibility, reliability, as well as performance optimization for both simple and complex tasks. Emphasis is placed on ensuring efficient data transmission particularly concerning the security of data-intensive tasks. The article also discusses modelling various stages of deploying heterogeneous cloud systems, with highlighted tools such as Terraform and the utilization of a cloud broker for effective implementation. The relevance of these insights is extended to the Internet of Things (IoT), which deals with vast, complex data that necessitates efficient data transfer for intensive processing tasks in heterogeneous cloud systems. The paper concludes with addressing the requirement for future research in this critical area for improving efficiency and performance in our rapidly digitizing era.

Keywords: heterogeneous cloud systems, multi-cloud systems, data-intensive tasks, modelling data transmission, big data, securing data transmission.

The use of heterogeneous cloud systems is of utmost importance for scientists and information technology professionals. Heterogeneous cloud or multi-cloud technologies involve the simultaneous use of several leading cloud providers, providing unparalleled flexibility and reliability. With our expertise in big data, we are confident in our ability to provide important levels of performance, speed, and security for both simple and critical tasks. Efficient processing and storage strategies are necessary to handle the large amount of data generated daily, especially in heterogeneous cloud systems. The security of data transferred between cloud systems is a particularly pressing issue that requires decisive action.

Rapid changes are occurring in the field of cloud computing, particularly in the use of heterogeneous cloud resources for processing big data. We can integrate and optimize these technologies to maximize performance and reliability.

Hosting your infrastructure across multiple clouds increases fault tolerance and enables smoother recovery from cloud provider failures [1]. It also allows for the utilization of cloud resources and services that are best suited for your current data set. Terraform is the solution to the complexity of deploying across multiple clouds due to differences in interfaces, tools, and workflows between providers.

With Terraform, managing multiple providers and handling cross-cloud dependencies is made easy.

Despite the heterogeneity of cloud resources and related security challenges, there is currently no widely accepted approach to cloud modeling for big data tasks. Ensuring data transmission security for data-intensive tasks in heterogeneous cloud systems is an urgent task. A model and methodology to effectively implement this will be a key area of future research in cloud technologies.

Related work. Sending data for processing intensive tasks in heterogeneous cloud systems is highly relevant for the Internet of Things (IoT). IoT devices collect vast amounts of data that require transmission, processing, and storage. To ensure real-time analysis and decision-making, it is crucial to consider factors such as transmission speed, low response time, resource efficiency, security, and the ability to process data at the edge of the network (edge computing and edge-cloud environments) [2].

It is important to note that not all IoT systems can handle large amounts of data on their own hardware or in a single cloud. IoT systems can greatly benefit from using a heterogeneous cloud system to optimize the data transfer process. By selecting the best cloud service for each specific type of data or task, the efficiency of IoT systems can be significantly increased. Sending data for tasks that require intensive processing in heterogeneous cloud systems is highly relevant for IoT. This enables proper processing and management of the vast data flows generated by IoT devices.

Loreti and Ciampolini (2015) [4], Mansouri et al. (2020) [5], Bicer et al. (2011) [6], Clemente-Castelló et al. (2018) [7], and Zhao et al. (2021) [8] have extensively researched and implemented hybrid cloud [9].

Heterogeneous cloud systems implementation. Steps to deploy heterogeneous cloud systems may include the following stages:

1. Analysis and planning: Before you start executing the deployment, it is important to understand the business requirements and technical needs. You will also need to assess your current IT infrastructure and analyze how a heterogeneous cloud system can fulfill your current and future needs.
2. Selecting cloud providers: Given your needs and budget, you need to determine which cloud service providers are best for your business. Consider various aspects such as pricing, technology, data center locations, service sets, security policies, support services, etc.
3. Create an architecture: In this stage, the structure of the cloud deployment is designed. This may include network design, security policy definition, data management strategy, etc.
4. Deployment: You can now start deploying cloud services using automated tools such as Terraform.

5. Monitoring and optimization: Once deployed, you should monitor the performance of the heterogeneous cloud system to implement best practices in cloud resource management and make necessary changes for optimization.
6. Service and resource management: Cloud management systems can promote productivity, control costs, and increase efficiency.

The process of implementing heterogeneous cloud systems can be complex and requires a deep understanding of cloud technology and your business requirements. For the best results, consider engaging experienced professionals or consultants.

Data transfer for data-intensive tasks in heterogeneous cloud systems refers to the process of transferring large amounts of data between different cloud systems or within a single cloud system. Heterogeneous cloud systems typically include different cloud service platforms from different providers, which can complicate the data transfer process. This process can include many tasks, including:

- Data migration: moving large data sets from one location to another, such as on-premises storage to cloud storage;
- Inter-service communication: sharing data between different services or applications in the cloud;
- Fast transfer: transferring large amounts of data at high speeds for real-time or low latency tasks;
- Security: ensuring the security of data transmission through encryption, access control, and other measures;
- Data flow management: Ensuring reliable, consistent, and efficient data transfer in complex network environments.

The importance of cloud technologies that support efficient data transfer increases with the growth of data that needs to be processed and stored.

Let's build a graphical model (Fig. 1) of data transfer for data-intensive tasks in heterogeneous cloud systems.

The process of data transfer for data-intensive tasks in heterogeneous cloud systems may include the following steps:

1. Receiving and forming tasks: At this stage, you define the type of task to be performed and the large amount of data to be processed. Data-intensive tasks can include analyzing large databases, machine learning, modeling, IoT, and more.
2. Create infrastructure as code (IaC) with Terraform: In this phase, you use Terraform to build the infrastructure needed to perform the tasks. This can include running virtual machines, setting up the network, configuring storage, etc. Infrastructure as code allows you to automate this process by running it on different services from different cloud providers.

3. Using a cloud broker: A cloud broker helps automate the distribution of tasks between different cloud providers and services. It decides where to send each task based on factors such as cost, performance, availability, and other requirements.
4. Using the selected cloud services: This step involves using the selected cloud services that were chosen in the previous step to run the computation.
5. Performing data-intensive tasks: The selected cloud services perform tasks by processing large data sets. The cloud broker monitors the execution to ensure that the task is successfully completed.

Depending on the requirements, additional steps may be included, such as optimizing the data transfer process, monitoring, and operating services, ensuring data security and confidentiality, etc.

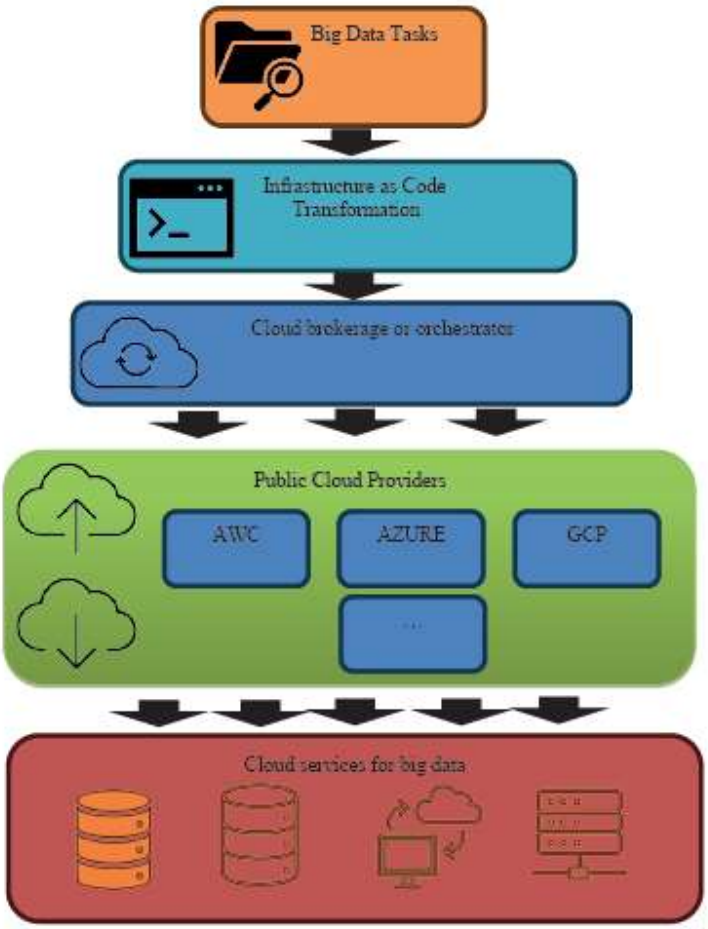

Figure 1. Overview of the process for using multi-cloud systems for data-intensive tasks

Data protection is of utmost importance for any organization that processes data. In today's world, where data distortion, loss, or unauthorized access can lead to serious consequences, data protection is an absolute necessity. Here are some reasons why it is so important:

- Confidentiality: Much of the data that organizations process, and store is sensitive. This could be personal customer data, sensitive company data, or information that is subject to data protection legislation such as the GDPR;
- Reputation: Data leaks or security breaches can severely damage a company's reputation, which can lead to the loss of customers or revenue;
- Financial losses: Unauthorized access to financial or other sensitive information can lead to financial losses;
- Legal requirements: Many countries require a certain level of data protection, and failure to comply with these requirements can result in heavy fines and penalties;
- Security: In the event of a cybercrime, unauthorized access to systems can lead to further security breaches, such as the installation of malware. Modern organizations must pay due attention to data protection in heterogeneous cloud systems. It is not only advisable but also necessary for ensuring the security of sensitive information.

Data protection is an important aspect when using cloud-based systems, especially for data-intensive tasks. Here are a few ways you can organize this at each stage:

1. Receiving and generating tasks: This includes protecting incoming data that may be sensitive or confidential. Apply encryption to data during transmission and storage and use secure data transfer methods.
2. Create infrastructure with Terraform: Sensitive data such as passwords or access keys should not be stored in Terraform code. Instead, use secure means of storing such data, such as secret stores or environment variables. Using Terraform also comes with the prerequisite of creating backups and using version control for your IaC code.
3. Using a cloud broker: Special attention should be paid to securing communication with the cloud broker with encryption and strong authentication methods.
4. Use of selected cloud services: The cloud services you use should support security best practices, such as encryption of data on disks and in transit, role-based authorization, security logs, and compliance with security standards.
5. Perform data-intensive tasks: Encrypt results before they are transmitted and store them in a secure location. Initiate security audits to identify any potential issues.

In addition, it is important that all employees who work with data and systems have appropriate security training. They should understand how security mechanisms work and the threats they protect against. Ensuring continuous security monitoring and conducting regular security reviews also play a key role in data protection.

Conclusions

In conclusion, the paradigm shift towards heterogeneous cloud systems presents an innovative future for big data processing. The utilization of several leading cloud providers not only provides flexibility and reliability, but also opportunities for optimization across a variety of tasks. Despite the complex challenges particularly concerning security during data transmission for data-intensive tasks, solutions such as Terraform and utilizing a cloud broker have paved the way for smooth implementation. Alongside, emphasis on appropriate security strategies, mindful planning, and consistent monitoring are crucial to make the most out of these technologies. Furthermore, with the Internet of Things generating increased volumes of data requiring complex processing and real-time analysis, optimized data transfer for data-intensive tasks in heterogeneous cloud systems turns incredibly essential. While there is exhaustive research and development still required in this realm, its integration with IoT stands to bring about substantial improvements in efficiency and overall performance. As we continue to expand in the digital era, the exploration and enhancement of heterogeneous cloud systems remain of great relevance and potential.

References

1. *Multi-Cloud Deployment. Use case | Terraform | HashiCorp Developmen*t, electronic resource: https://developer.hashicorp.com/terraform/intro/use-cases#multi-cloud-deployment (date of access: 10.02.2024).
2. M. Zheng, X. Du, Z. Lu, Q. Duan (2024). *A Balanced and Reliable Data Replica Placement Scheme Based on Reinforcement Learning in Edge – Cloud Environments*, https://doi.org/10.1016/j.future.2024.02.004
3. D.C. Marinescu (2022). *Cloud Computing. Theory and Practice*, third edition, Morgan Kaufmann, https://doi.org/10.1016/C2020-0-02233-4
4. D. Loreti, A. Ciampolini (2015). *A Hybrid Cloud Infrastructure for Big Data Applications*, IEEE 17th International Conference on High Performance Computing and Communications, https://ieeexplore.ieee.org/abstract/document/7336418
5. Y. Mansouri, V. Prokhorenko, M.A. Babar (2020). *An Automated Implementation of Hybrid Cloud for Performance Evaluation of Distributed Databases*, J. Netw. Comput. Appl., https://www.sciencedirect.com/science/article/abs/pii/S1084804520302149

6. T. Bicer, D. Chiu, G. Agrawal (2011). *A Framework for Data-Intensive Computing with Cloud Bursting*, IEEE International Conference on Cluster Computing, pp. 169–177, https://ieeexplore.ieee.org/abstract/document/6061052
7. F.J. Clemente-Castelló, B. Nicolae, R. Mayo, J.C. Fernández (2018). *Performance model of mapreduce iterative applications for Hybrid Cloud Bursting*, IEEE Trans. Parallel Distrib. Syst., vol. 29, no. 8, pp. 1794–1807, https://ieeexplore.ieee.org/abstract/document/8283575
8. *A Deep Reinforcement Learning Approach to Resource Management in Hybrid Clouds Harnessing Renewable Energy and Task Scheduling*, 2021 IEEE 14th International Conference on Cloud Computing, CLOUD, pp. 240–249, https://ieeexplore.ieee.org/abstract/document/9582195
9. U. Faheem, D. Shagun, X. Xiaoyu, M. Ali Babar, *Evaluation of Distributed Data Processing Frameworks in Hybrid Clouds*, https://www.sciencedirect.com/science/article/pii/S1084804524000146#b38