A Cloud Resources Portfolio Optimization Business Model - From Theory to Practice

Valentin Haag[1], Maximilian Kiessler[1], Benedikt Pittl[1], and Erich Schikuta[1]

Faculty of Computer Science, University of Vienna, Austria
valentin.haag94@gmail.com, m.kiessler@hotmail.com,
benedikt.pittl@univie.ac.at, erich.schikuta@univie.ac.at

Abstract. Cloud resources have become increasingly important, with many businesses using cloud solutions to supplement or outright replace their existing IT infrastructure. However, as there is a plethora of providers with varying products, services, and markets, it has become increasingly more challenging to keep track of the best solutions for each application. Cloud service intermediaries aim to alleviate this problem by offering services that help users meet their requirements. The business model of such intermediaries is missing in current works and introduced in this paper. This paper aims to lay the groundwork for developing a cloud portfolio management platform and its business model, defined via a business model canvas. Furthermore, a prototype of a platform is developed offering a cloud portfolio optimization service, using two algorithms developed in previous research to create suitable and well-utilized allocations for a customer's applications. The developed business model forms the baseline for realizing cloud intermediaries on cloud markets.

Keywords: Cloud Economics · Portfolio Optimization · Business Model

1 Introduction

Over the past years, the cloud resource market has been one of the fastest-growing IT segments. The biggest providers, Amazon Web Services (AWS), Microsoft Azure, and Google Cloud, have seen annual sales increases of over 20% for several years. Just AWS itself reported a revenue of 90 Billion US dollars for the year 2023, and the entire cloud market achieved revenue of 545.8 Billion dollars worldwide with 19% growth compared to 2022 [1]. This large and expanding market has resulted not only in a growing cloud service provider (from here on, often referred to as CSPs) market but also in several methods of delivery of cloud resources to the customer.

One big challenge facing both industry and academia is finding a cost-effective solution when buying cloud capacities. Pittls' study [16] concludes that almost all observed resource requests were oversized and offered significant cost reduction potential, which lies not only in reducing the amount of resources bought but also in their composition. Depending on the planning period of the operations to be performed, a different mix of procurement from different market spaces is

optimal. The author suggests tackling this problem via a cloud resource trading intermediary.

Offering a service that aids businesses in managing their cloud portfolio and proposing efficient allocations to run their applications, even across various providers, is of great interest in the market. While most larger CSPs offer some functionalities and services that claim to help prevent over-provisioning, like AWS Lambda and Fargate, it is ultimately not in the provider's best interest to reduce the costs for the customer. For the same reason, offering cross-platform support should not be expected of them either.

One way to better leverage the opportunities of the cloud market and make the market more accessible is the use of cloud intermediaries between the CSPs and the customer. Given that, this paper introduces a business model for a intermediary termed *Cloud Portfolio Manager*. The presented work has a twofold focus, on SaaS (business model) and IaaS (mangement of infrastructure resources), and addresses the following research questions:

- *Under what kind of business model could such an intermediary operate?* To answer this question, a detailed proposal and description of a viable business model for a cloud resource intermediary will be presented. Intermediaries in a similar fashion have been put forward, but as far as the authors' best knowledge, no conclusive business model exists on how these intermediaries could operate.
- *How could such a platform be implemented?* The other goal of this work is the design and implementation of a cloud resource intermediary.

Thus, the paper is structured as follows: The next section section 2 gives the reader a literature survey on the target research area. Section 3 focuses on cloud portfolio optimization theory, and two respective algorithms, a genetic and a greedy one, are described and their performance evaluated - this chapter is mainly based on previous work. In section section 4, we propose our proposal of a business model for a cloud portfolio management platform, consisting of the nine building blocks defined by the business model canvas framework of Osterwalder [13]. Based on the business model description, we present the implementation of our respective prototype in section section 5. Finally, section 6 summarises the paper's findings, including what future work to expand upon the topic.

While we followed the basic Research Design principals [5], a solid evaluation of our artefact is outside of the scope of the paper, but can be found here [8].

2 State of the Art

The cloud market currently consists of many cloud service providers, each featuring various ways and markets for customers to access their products. There is no single space from which all suppliers can be accessed which slows down the

development of multi-cloud environments [14]. The market is an oligopoly, dominated by the most prominent three players, Amazon Cloud Services (AWS), Microsoft Azure, and Google Cloud [12]. In the course of this work, we focus on AWS, the largest cloud service provider today, which offers three different resource markets with varying pricing models: *On-demand marketplace*, *Saving plans marketplace* and *Spot marketplace*.

The definition of a business model has not only been one of the earliest focuses of research, but also one of the most hotly debated. Almost every paper concerning this topic has defined its own take on this task, and to this day, there is no general agreement on a universally accepted definition. We follow the definition of [13], where a business model describes the rationale of how an organization creates, delivers, and captures value. Similar to the definition of the term, there is also a wide array of frameworks describing the elements a business model comprises. The paper will focus on the Business Model Canvas from Osterwalder and Pigneur [13], one of the most established and widely used frameworks today. This framework proposes to describe any business model with nine building blocks: (i) The *Customer Segment* block describes the groups a company tries to offer its products to. (ii) The *Value Proposition* describes the reason why a customer chooses to work with one business over others (iii) *Channels* describe the way a company reaches its customer segment to make its proposition of value. (iv) The different kinds of relationships companies can have with their customers are described in the block *Customer Relationships*. (v) The block *Revenue Streams* deals with the income a company receives from its customers. (vi) The *Key Resources* block describes the resources that allow the company to operate its business and earn revenues. (vii) The block *Key Activities* defines the activities a company must perform to be successful. They are necessary to create value and earn revenue, and they can differ widely depending on the company. (viii) *Key partnerships* are the suppliers and partners a business model includes to make it work. (ix) The last block *Cost Structure* of the Canvas deals with all operating costs of the business model.

Actual cloud broker implementations on the market are still sparse, while the high number of cloud service offerings makes it hard to find the right service as a customer [6]. For example, spot.io[1] offers a range of tools for customers to analyze, manage and optimize their cloud portfolios. While they offer services like Eco and Elasticgroup, with a similar value proposition of cost reduction for a customers cloud portfolio, their underlying optimization algorithms differs from ours. Additionally, Eco also relies on trading and reselling resources among their customers and does not factor in spot instances for its calculations. Densify[2] offers a cloud management and optimization service with a similar value proposition to our platform but differences regarding revenue stream and pricing model. Seeing as their pricing model is based on the number of instances

[1] `https://spot.io/`
[2] `https://www.densify.com/`

managed, Densify does not seem incentivized to reduce the number of instances used in a portfolio.

While a plethora of research has been conducted in the field of cloud intermediaries, the number of implementations on the market is limited. This situation, in conjunction with the fact that the services provided can vary wildly, has resulted in a lack of research concerning the business models of cloud brokers so far. Nevertheless, the work of Filiopoulou at al. [7] gives an overview of benefits, common pricing models, and an evaluation of cloud brokers. It is concluded that brokers assist companies in developing themselves and creating a more competitive environment for providers while earning revenues themselves. Other works such as [15] focus on the introduction of a cloud ecosystem as a generic framework where intermediates are foreseen, but no business model is described.

3 Portfolio Optimization Model

In the paper at hand we refer cloud portfolio optimization alogrithms which were introduced by Kiessler et al. [11]. This section summarizes the most important findings of our previous work.

The main goal of cloud portfolio management is usually achieving the lowest costs for running a specific set of applications over time. Some research like Jangjaimon and Tzeng [10] and Sharma et al. [19] tried to deal with this problem by creating a checkpointing mechanism or by focusing on preemptible servers in combination with concepts taken from financial modeling, to meet the requirements of applications when using spot instances. Meanwhile, Pittl et al. [16] took a more comprehensive approach to cloud portfolio management, which resulted in the findings that a more heterogeneous portfolio tends to be more cost-efficient. Cloud portfolio is defined as a set of *cloud instances*, which are used to run a set of *applications* on them. The main goal of our optimization problem, hereby, is finding a cost-efficient allocation of these applications and instances [11]. The problem is essentially a multi-dimensional packing problem, it is NP-hard, very complex to solve, and finding an optimal solution is usually not computationally feasible [4]. Therefore, we developed two optimization heuristics to find good approximations of an optimal solution. First, we developed a greedy algorithm called *Efficient Resource Inference for Cloud Hosting* (ERICH). It integrates the approach of the widely known bin packing algorithm first fit decreasing (FFD), combining it with the proposed portfolio management strategy by Hwang and Pedram [9]. Our second algorithm is a genetic algorithm named *Genetic Optimization of Resource Groupings* (GEORG) which outperformed ERICH in almost all evaluated secenarios [11].

4 Cloud Portfolio Manager Business Model

Based on Osterwalders Canvas [13] the proposed Cloud Portfolio Managers building blocks are shown in figure 1.

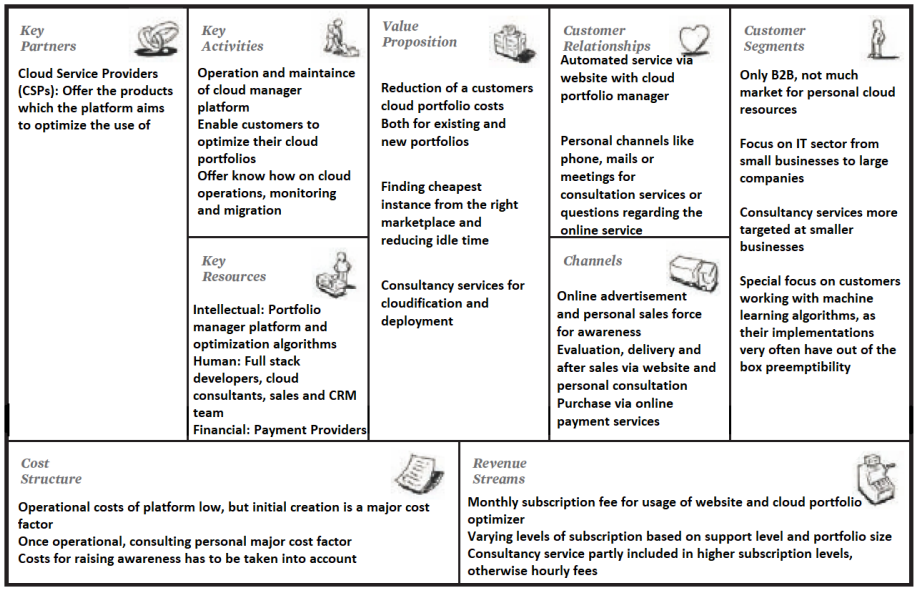


Fig. 1: Business Model Canvas [13] of Cloud Portfolio Manager

Customer Segments. For our customer segment, the sole focus will be on the B2B area, as consumers, so far, have little to no reason to purchase cloud resources for personal use. Within the B2B sector, a big emphasis will be put on the IT sector, ranging from small businesses to large companies, seeing as large parts of this sector move more and more of their IT infrastructure into the cloud such as Netflix, which moved its entire IT infrastructure to AWS [2]. Surprisingly, when it comes to smaller businesses start-up companies looking for cloud solutions prefer the reputation of a cloud provider over other aspects such as price, security, and reliability [17]. Hence, establishing a reputation should be a primary business goal. Besides this, the envisioned Portfolio Manager can increase its attractiveness to small businesses and start-ups by offering cloud consulting and optimization services.

One customer segment our business will specifically focus on within the IT sector are those companies, businesses, and possibly research facilities working with machine learning algorithms. These are uniquely well-suited to be run on cloud resources, as their implementations often offer out-of-the-box preemptibility like the previously mentioned PyTorch and Google TensorFlow [3,20].

Value Proposition. The primary value proposition, a cost reduction for their cloud portfolio, is unlike many other business models, targeting all customer segments. It can be applied to both existing portfolios and first-time cloud deployments. A portfolio can be set up in two ways: Either, directly via interacting with the Portfolio Manager via its website or via API. The API allows customers

to integrate the Portfolio Manager with their own systems and automate the process of creating portfolios and creating allocations.

The previously mentioned cost reduction is achieved by a combination of choosing the cheapest instances from the right marketplace and reducing their idle time through continuous monitoring of the needs of the applications. The resulting benefit for the customer entails the direct cost reduction itself and offers easier access to the complex world of cloud computing, which could be especially useful for small and medium-sized businesses.

Channels. Listed here, we will address the channels used throughout the five phases of customer interaction through which we aim to reach our customers:

(i) *Awareness:* For the first phase, raising the customer's awareness of our service, a mixture of online advertisement and direct contact with potential customers through an in-house sales force seems applicable. Furthermore, targeted online advertisement, for example, via Google ads[3], word-of-mouth between different businesses as well as reach potential customers is trade fairs could be considered an option. (ii) *Evaluation:* Our website is the primary channel used for evaluation and the center of the operation. It provides example calculations, which showcase the potential cost reductions offered by the service and gives an overview of the pricing structure. (iii) *Purchase:* Regarding purchasing our products, many online payment services such as PayPal, Amazon Payments, and Credit cards are available and can be provided with relative ease. (iv) *Delivery:* The delivery of the optimization service to the customer can also be achieved through the platform's direct channel, either by direct customer interaction on the website or via API call. (v) *After sales:* On the website the user can overview his portfolios, optimizations, subscriptions, and general account information. A FAQ page as well as personal customer support can help to clear issues.

Customer Relationships. The website will be used for the majority of interactions, such as setting up an account, creating and managing a portfolio as well as for communicating with consumers. Another important aspect of this building block is the interaction between the consultancy force and the customers. Here, direct interaction with the customer is preferable, as service experience is valued even higher than the actual service quality in B2B services [18].

Revenue Streams. Regarding revenue streams, a plethora of options are available at first glance. However, most of them are not readily applicable to our platform for one reason or another, leaving us with one very widely used revenue stream as our primary source of revenue.

The first option we want to discuss is advertisement. While it is the main revenue stream of many large online platforms such as YouTube and Facebook, the Cloud Portfolio Manager focuses on a too small audience.

[3] `https://ads.google.com/intl/de_at/home/`

Next, we have considered the option of transaction fees. This could be implemented on a usage-based model, where the customer would pay a certain amount for each optimization based on the portfolio size.The usage-based model does not lend itself well to a product that is meant for continuous optimization. This would either lead to a need to frequently pay for a new allocation or prevent customers from getting the full benefit of an approach that is meant to adapt to changing demands in their portfolio.

Another alternative would be a system based on a brokerage fee. In this case, a part of the cost reduction achieved by the Cloud Portfolio Manager would be taken as our revenue. The significant flaw with this idea, though, is that our system does not aim to directly access and manage the customer's cloud instances. This would result in customers needing to accurately and honestly report their current cloud expenses and their achieved cost reductions, which lends itself to be abused way too easily.

Finally, we propose that subscription fees are the revenue stream best suited to our business model. They tackle several disadvantages mentioned in the previously discussed systems, such as fitting well with a continuously running service, unlike pay-per-use transaction fees and a low entry barrier compared to a one-time charge. The subscription fee, due in a monthly interval, could either be based upon a system with different levels of subscriptions, offering support to differing sizes of cloud portfolios and varying levels of customer support, or directly scaling with the size of the optimized cloud portfolios.

As for revenue streams concerning the consultancy side of the business model, a classic hourly fee would most suit customers needing only a more minor assistance contingent. Another variant would be offering package deals with a fixed price, such as offering to help set up the first cloud portfolio for a customer. Finally, higher-level subscription models for the cloud portfolio optimizer platform could include a certain amount of consultancy services for free.

Key Resources. As for the differing categories of key resources, the following can be said: When it comes to physical key resources, there is little to be mentioned here. Financial key resources may also not play a huge role in starting off. Of course, financial resources such as cash or credit will be needed to set up the business, but due to its nature, these will be of a small volume.

Finally, when it comes to human key resources, the following groups can be expected to be part of those: Especially for development and improvements to the platform, further full-stack developers could be needed. Furthermore, a small team of cloud consultants would be responsible for providing customers. Besides that, a group of employees helping with sales and CRM-related topics should be employed as well.

Key Activities. The most important key activity of the business is the operation and maintenance of the Cloud Portfolio Manager platform. In this capacity, the platform offers the customer an automated service. Besides a simple interface to

directly manipulate a portfolio, the main feature for management is the possibility to upload load profiles based upon which an optimized portfolio of instances is calculated and displayed to the customer.

In addition to the service provided by the platform, the other main activity is problem-solving for the customer by offering our consultancy service.

Key Partnerships. Within this building block, the most prevalent partnerships are the various CSPs for which the platform offers portfolio optimization. While actively managing the customer's portfolio is not part of the business plan so far, it is crucial to the platform's functionality to access the instances and their respective pricing offered by the various providers. If the business grows beyond a small scale, it would be possible for certain activities, such as customer support or cloud consulting, to be outsourced to external partners, which would turn these into key partnerships as well.

Cost Structure. The cost structure of the business model is intended to lean towards being value-driven, focusing on creating value for the customer. As the business operates online with a web platform at its center, scaling should be achievable relatively easily.

5 Portfolio Manager Prototype

This section will present the prototype of our Cloud Portfolio Manager. It will present an overview of important pages and showcase the various functionalities of the application. The main motivation is to gather product and pricing information automatically from cloud providers so that matching providers can used for pending cloud workloads.

After authentication, the *instance page* becomes accessible as shown in figure 2 contains information about available instances from the various providers. For our prototype, we chose a range of instances from the four biggest CSPs: AWS, Google Cloud, Microsoft Azure, and Alibaba. The list on this page gives an overview of each instance's main attributes: provider, name, market space, capacity, and price.

The *Apps and portfolios* page is depicted in Figure 3. The left side lists the user's applications and details like mean resource demand, demand variance, preemtibility, and starting and finishing time. On the right side of the page is a list of the user's portfolio, including details like which providers should be considered for any possible allocation, a minimum quality of service, the number of apps in the portfolio, and a list of which applications exactly the portfolio consists of.

To create a new application or portfolio, two green buttons depict a plus sign on each side of the page. These open the respective application and portfolio forms, as seen in Figure 4a and Figure 4b. To create an application, the user

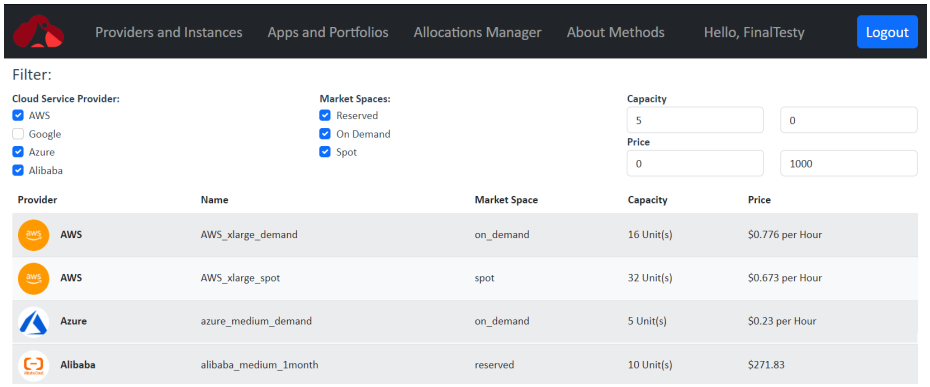


Fig. 2: Instances page

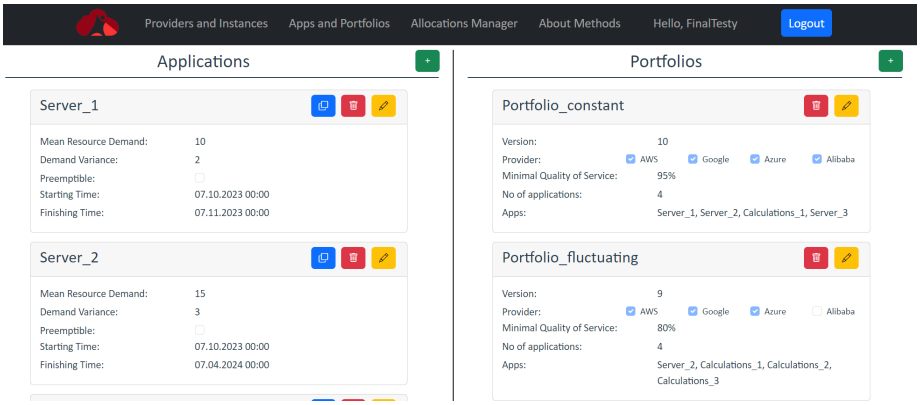
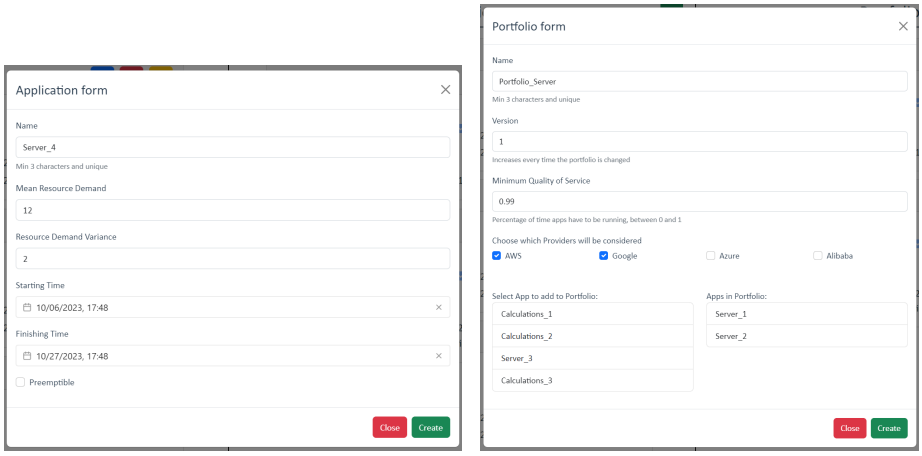


Fig. 3: Apps and portfolios page overview

has to fill out the application form, including a unique name, mean resource demand, demand variance, a checkbox for preemtibility, and finally, the starting and finishing time chosen via a date-time picker. Should the user wish to create a portfolio, the portfolio form requires a unique name and a minimum quality of service, which gives a percentage of time the apps in the portfolio are required to run. e portfolio form also requires the user to choose at least one CSP to be considered for allocations and which apps should make up the portfolio. To ensure suitable inputs for both forms, they also feature various checks, giving instant feedback to invalid inputs, such as an application's finishing time before its starting time. The user can update each application and portfolio by clicking the yellow button, which displays a pen icon for every application and portfolio.

The *Allocations Page* as shown in figure 5 is the primary feature of the Cloud Portfolio Manager: creating allocations for cloud portfolios. The user has a drop-down menu at the top of the page, which lists their created portfolios. Choosing a portfolio shows its details on the side, and all already existing allocations for



(a) Application Form (b) Portfolio Form

Fig. 4: Screenshots of application and portfolio forms

this portfolio are below. Every allocation has an overview stating which algorithm was used, which portfolio version it was made for, its total costs, and the mean overall utilization achieved with this allocation: an excerpt is shown in figure 6b. As allocations can take a while to be calculated, especially in the case of using the GEORG algorithm, there is also a field stating if the allocation is already completed. Below the general stats are two fields, which can be extended by clicking on them. The first contains a complete list of all instances used for this allocation and some statistics like capacity, price, and the beginning and end of the instance's run time. The second field contains more detailed statistics about the allocation, such as separate statistics for reserved, on-demand, and spot instances. To create a new allocation, the "New Allocation" button opens a form where the user can choose which algorithm should be used such as shown in figure 6a.

6 Conclusion and Future Work

The main contribution of the paper is the development of a business model for a Cloud Portfolio Manager which was implemented including two optimization algorithms. The work is seen as a corner-stone towards adopting Cloud intermediaries in industry. As the cloud computing market has been on a meteoric rise over the past years and is still expanding, the topic of cloud portfolio management will likely keep or expand its relevance in the coming years. Future work will see the prototype developed into a fully functional public platform operating with our business model or a modification of it. Therefore, feedback from practitioners is required and the business model itself needs to be tightly evaluated along the Research Design framework. Furthermore, the platform's offered services can be extended by various monitoring functionalities and direct control of portfolios via the platform. .

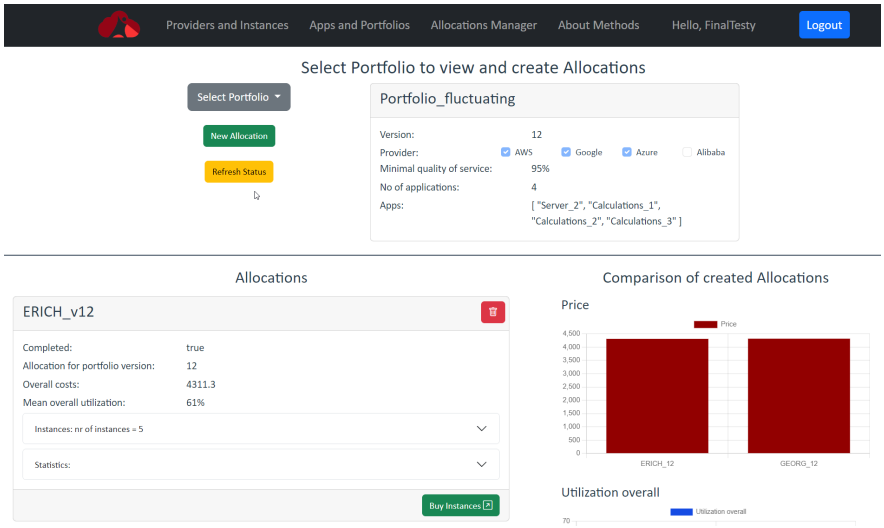
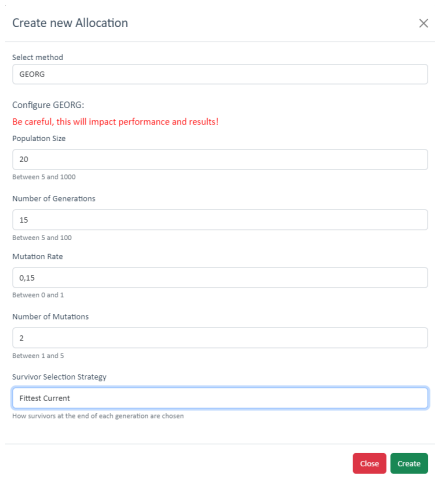


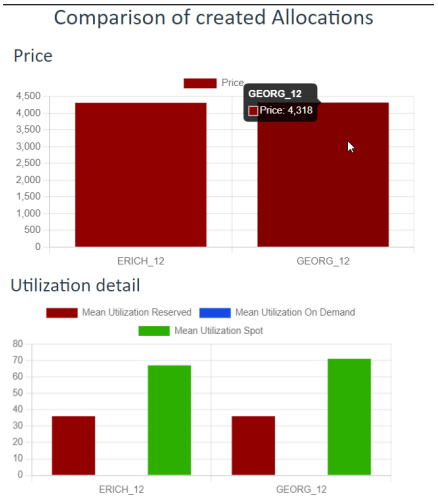
Fig. 5: Allocations page overview

References

1. Statista online services: Annual revenue of amazon web services (aws) from 2013 to 2022. `https://www.statista.com/statistics/233725/development-of-amazon-web-services-revenue/`, Accessed: 2024-08-07
2. Amazon web services: Netflix in aws (2017), `https://aws.amazon.com/solutions/case-studies/netflix-kinesis-data-streams/`, Accessed: 2024-08-07
3. Pytorch foundation: Pytorch Checkpoints. `https://pytorch.org/tutorials/recipes/recipes/saving_and_loading_a_general_checkpoint.html` (2023), Accessed: 12.05.2023
4. Chekuri, C., Khanna, S.: On multidimensional packing problems. SIAM journal on computing **33**(4), 837–851 (2004)
5. Creswell, J.: Research design (2012)
6. Elhabbash, A., Samreen, F., Hadley, J., Elkhatib, Y.: Cloud brokerage: A systematic survey. ACM Comput. Surv. **51**(6) (2019)
7. Filiopoulou, E., Mitropoulou, P., Michalakelis, C., Nikolaidou, M.: The rise of cloud brokerage: Business model, profit making and cost savings. In: Bañares, J.Á., Tserpes, K., Altmann, J. (eds.) Economics of Grids, Clouds, Systems, and Services. pp. 19–32. Springer International Publishing, Cham (2017)
8. Haag, V.: A cloud portfolio manager - a novel business model. Master's thesis, University of Vienna (2024)
9. Hwang, I., Pedram, M.: Portfolio theory-based resource assignment in a cloud computing system. In: 2012 IEEE Fifth International Conference on Cloud Computing. pp. 582–589. IEEE (2012)
10. Jangjaimon, I., Tzeng, N.F.: Effective cost reduction for elastic clouds under spot instance pricing through adaptive checkpointing. IEEE Transactions on Computers **64**(2), 396–409 (2013)
11. Kiessler, M., Haag, V., Pittl, B., Schikuta, E.: Optimization heuristics for cost-efficient long-term cloud portfolio allocations. In: International Conference on Information Integration and Web. pp. 309–323. Springer (2022)



(a) Form for creating a new GEORG allocation (excerpt of paramters)



(b) Graphs for comparison between a GEORG and ERICH allocation for an example portfolio

Fig. 6: Details on allocation page

12. Law, M.: Technology biggest cloud providers 2023. `https://technologymagazine.com/top10/top-10-biggest-cloud-providers-in-the-world-in-2023` (2023), Accessed: 16.06.2023
13. Osterwalder, A., Pigneur, Y.: Business model generation: a handbook for visionaries, game changers, and challengers, vol. 1. John Wiley & Sons (2010)
14. Petcu, D.: Multi-cloud: expectations and current approaches. In: Proceedings of the 2013 international workshop on Multi-cloud applications and federated clouds. pp. 1–6 (2013)
15. Pittl, B., Mach, W., Schikuta, E.: A classification of autonomous bilateral cloud sla negotiation strategies. In: Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services. pp. 379–388 (2016)
16. Pittl, B., Mach, W., Schikuta, E.: Cost-evaluation of cloud portfolios: An empirical case study. In: Proceedings of the 9th International Conference on Cloud Computing and Services Science (CLOSER). pp. 132–143. SciTePress (2019)
17. Repschlaeger, J., Erek, K., Zarnekow, R.: Cloud computing adoption: an empirical study of customer preferences among start-up companies. Electronic markets **23**(2), 115–148 (2013)
18. Roy, S., S., S., Bhatia, S.: Service quality versus service experience: An empirical examination of the consequential effects in b2b services. Industrial Marketing Management **82**, 52–69 (2019)
19. Sharma, P., Irwin, D., Shenoy, P.: Portfolio-driven resource management for transient cloud servers. Proceedings of the ACM on Measurement and Analysis of Computing Systems **1**(1), 1–23 (2017)
20. TensorFlow: Tensorflow Checkpoints. `https://www.tensorflow.org/guide/checkpoint` (2023), Accessed: 12.05.2023