

SIT226 Cloud Automation Technologies

Pass Task 8.2C

Kubernetes and Public Cloud

Background

Throughout the trimester we have learned how to deploy and manage applications in Kubernetes, however we have also explored the relationship between Kubernetes, different cloud models, and application architectures. In this task we explore the relationship between Kubernetes and cloud computing.

Get Prepared

This question considers how Kubernetes can be used for deploying applications to public, private, or hybrid cloud models. Historically, deploying an application to the cloud with features that would ensure availability, scalability, fault tolerance, and so on, would require developers to adopt the specific services offered by a single cloud provider. These services are still available, review some of these services to get an understanding of cloud development without Kubernetes (do not consider container services which were introduced much later).

Take the time to review the concepts of private cloud, public cloud, and hybrid cloud, and consider how applications would be deployed in these models both with and without Kubernetes. From the understanding of Kubernetes your studies this trimester, you should already have a reasonable understanding of how Kubernetes could support each of these models.

Complete the Task

Page Limit: 1 page of text formatted reasonably, e.g., 2cm margins, 11 or 12 point font, appropriate headings/spacing, etc.

Prepare a document according to the following requirements:

1. Explain the concept of elasticity regarding cloud computing and why it is beneficial for our applications (approximately $\frac{1}{3}$ page).
2. Discuss the problem of vendor lock-in and explain why using Kubernetes is a good approach to avoid this problem (approximately $\frac{1}{3}$ page).
3. Explain why it is beneficial to use Kubernetes on private infrastructure and how this would facilitate expanding to a hybrid cloud (approximately $\frac{1}{3}$ page).

Submit Your Task

Prepare your submission using the word processor of your choice and submit a PDF to OnTrack.

Taking it Further (Optional)

Kubernetes significantly simplifies deploying applications to public cloud but it isn't without its problems. The following topics are important considerations for any applications deployed in different situations:

- Extending a private cloud to adopt public cloud services, i.e., using a hybrid cloud, represents a significant advantage for organisations with existing infrastructure. A significant problem faced in this situation however is partitioning, where network connectivity is not maintained across all the entire cloud footprint. Consider how you might address this in a realistic situation – what kind of components would you deploy in the private infrastructure vs the public infrastructure? How would the application detect and adapt to the cloud being partitioned? Equally important, if the application can survive partitioning, how would the application detect and handle the various partitions being reconnected?
- Message queues, such as RabbitMQ, provide an important mechanism through which many applications are able to achieve asynchronous handling of requests, scaling of applications, fault tolerance, and so on. What is a message queue, how does it work/how is it used by applications, and how does it support these features?
- Another key application element that is often used in advanced architectures is to overlay or encapsulate access to other microservices using proxy servers, e.g., Envoy proxy, adding another layer of abstraction. What are the advantages of adding abstraction when invoking another service/microservice? How is the Envoy proxy used by service meshes, e.g., Istio, and what functionality is achieved by this?
- One of the issues raised above is what elements of the application would be deployed to private or public cloud. One of the factors that might be relevant here is the demand for resources from a particular component – a high demand could imply the public cloud is better able to satisfy those requirements or could equally mean the cost is too high. Consider alternative ways to capture the likely resource usage for a particular component. There are many different approaches, including manual estimation, application profiling tools, system accounting tools, public cloud calculators (estimates) and accounting systems (actual usage), and so on. You could look into any of these aspects to get some insight.

Citations and Referencing

When completing any work it is necessary to acknowledge the work of others that you have relied upon. For written assessment, we achieve this through the use of citations and references. Failing to correctly identify the work of others is known as plagiarism and is considered an issue of Academic Integrity.

If your submission to this task has involved the work of others, you must include citations and references where appropriate. Deakin provides a web site that explains how to use citations and references, and includes explanations of various referencing styles:

<https://www.deakin.edu.au/students/studying/study-support/referencing>

You may select any style for your citations/references, however you must be consistent in applying that style in this task (you can use other styles in other tasks if you wish).

Note that any bibliography/list of references is not included in page limits.