SIT226 Cloud Automation Technologies

Pass Task 4.2C

Cost Comparisons

Background

Previously we considered the difference between capital and operational expenditure as one dimension of how using cloud computing can impact infrastructure costs for business. Differences in costings however are not limited to accounting however, with using locally hosted infrastructure requiring significant initial expenditure while a public cloud approach typically incurs costs in line with business activity. In this week we consider what costs would be incurred for three alternative approaches: in-house servers, cloud-based virtual machines, or cloud-based Kubernetes services.

Note: For the in-house server and cloud-based virtual machine approaches, assume that no containerization and associated runtime systems are used, the software would be installed directly on the operating system.

The Scenario

You are working for a medium sized organisation that specialises in the sale of visual art in various forms (paintings and prints, ornaments and statues, photo printing services, etc.). The organisation is looking to expand their business by adding a new web service where clients can request advice on how to maintain their visual art collections, however they expect to grow this in the future by expanding into consultancy and restoration services. Your task is to determine where/how this web service will be run.

For this purpose, a software development company has been contracted to build the new website using a microservices architecture. The contracted company will be responsible for deploying the website and verify the site is working properly before and after deployment, after which the contract will be considered finalised and the software development company will no longer be involved. The organisation maintains adequate infrastructure on-site to adequately support the new website, e.g., networking (including Internet), power, backup, datacentre space, etc., except for a server if an on-site deployment is chosen (see "Get Prepared", below).

Ongoing support will be provided in-house, for which the organisation has decided the site must have a 99.9% uptime/availability. The organisation currently employs six support staff to ensure at least one person is available on site at all times, however these staff are currently close to capacity and can only provide periodic support for deploying updates. Additional staff would be required if the new service requires more regular support. Any other software/infrastructure (racking, network, backup, air conditioning, etc.) is already licensed/owned by the organisation.

Currently, the developers specify a minimum of 2GB of RAM and 1TB of storage for the hosting site, however these requirements are expected to grow significantly after the website is launched.

Get Prepared

Start by obtaining the "task resources" file where you accessed this question from in OnTrack. This file contains a template answer for you to complete for this exercise.

For this task you will need to investigate the costings for the three alternative approaches. For in-house, you will need to cost an appropriate server computer, e.g., Dell. For the cloud-based solutions, you will need to investigate cloud costs for the required services. The major vendors (AWS, Azure, GCP, etc.) provide calculators to help estimate costings, however you need to consider how these services actually work, e.g., hosting a VM usually is usually broken down into sub-costs for compute, storage, networking, and so on.

The organisation requires 99.9% uptime, so you will need to consider the best way to solve this with the solutions you propose. In some cases, you may need to include the cost of additional staff to increase the size of the support team, for which you can assume a cost of $80,000 per person, per year (including all associated costs such as their workstation/laptop).

Your proposed solutions must be sensible and consider both capital costs and operational costs for each approach. Some hints are included in the task resources file, however you will need to customise these and may require less/more rows depending on your solution. All costs must be stated in Australian dollars (note that Google can convert currencies for you, e.g., search for "100 USD to AUD").

Note 1: it is not required to have a deep understanding of the intended application. If there is a need to know how a particular aspect would work, make a reasonable assumption and state this in your answer (there is a section for assumptions in the template).

Note 2: you do not require large numbers of staff to support the new service. Although the six staff indicated are currently at capacity, if a new staff member were to be appointed existing workload would be redistributed, eventually resulting in a relative even distribution of work.

Complete the Task

Page Limit: no limit for costs and assumptions (will most likely be around a page), 1 page of text formatted reasonably for the summary and analysis sections, e.g., 2cm margins, 11 or 12 point font, appropriate headings/spacing, etc.

A template for your answer is provided in the "task resources" associated with this question which you must use to complete your answer.

1. Complete the costings for each hosting alternative, as complete and as accurately as you can. The table included in the template (in the task resources) includes some examples to help get started (no page limit).
2. State any assumptions that you made in the costings (if relevant, no page limit).
3. Starting on a new page, provide a short discussion on the costs you have identified for each solution (approximately 0.2 page each). Some suggestions are provided in the template.
4. Recommend one of the three approaches you have examined here and justify your selection (approximately 0.4 page). As noted in the template, you must explain the reasons for your recommendation explain why the alternatives are poorer choices.

Submit Your Task

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

Taking it Further (Optional)

There is significant scope for further developing your knowledge in the areas addressed by this question. The task of identifying appropriate hardware and/or cloud services for delivering any network service may seem like a significant challenge, but there are a number of tools and techniques that can be used.

- There are many tools available for measuring the resource usage of applications, including profiling tools often used by software developers and various telemetry gathering systems used for monitoring system usage. Some hosting platforms also can provide an accurate review of resource usage, e.g., the public cloud providers themselves will always carefully account for the resources your application uses. You can also use tools to generate artificial load on your services as well, usually referred to as load testing tools or stress testing tools.
- There are a vast number of possible combinations of hardware that could be used to host different applications. Different combinations of CPU, memory, and even motherboard chipsets can make a significant difference to the performance of the computers used to host your services, and this is before you consider issues such as fault tolerance on the platforms. If you were working in operations and involved in planning hardware purchases you would need at least a basic awareness of these technologies, however understanding the relationship between how hardware choices and software performance quickly becomes critical.
- The alternative to locally hosted infrastructure is public cloud infrastructure. Understanding the available services is important, however there is much more complexity in this picture. For example, if you explore the compute services of the various providers, you'll usually see you can create virtual machines on-demand (instantly when needed), by reserving capacity ahead of time (pre-paid capacity), or using spot pricing (accessing cheaper resources which is sitting idle). You will also find that pricing varies by resource type and by region, which can significantly increase the complexity. There are also other solutions we haven't considered here, such as container services (without Kubernetes) or using a serverless architecture. You can even consider differences between a managed Kubernetes service versus creating your own Kubernetes cluster on VMs.
- Finally, have you really understood what would be required to host your own infrastructure locally? Ignoring pricing, start by just brainstorming all the infrastructure, hardware, and software you might need to build and run a data centre before you even add the first server. How many items have you been able to identify? How much do you think this costs?
- How do the alternatives presented in this question compare when you consider security? Which service would you consider to be more secure and why? How would you plan, implement, and verify the security for any of these alternatives?

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 website 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.