Cloud computing

wznd85

December 8, 2017

1 Infrastructure Management

1.1 Answer

For this business, due to the critical nature of the infrastructure and the need for a custom software stack, I believe an Infrastructure as a Service (IaaS) model would be best to use. Furthermore, I would incorporate a hybrid cloud solution, with the aim of hosting the website on a public cloud; the reason being to reduce latency for users, and to deal with variable demand load while being infinitely scalable. The private cloud would be necessary to host sensitive data as this offers more security than the public cloud.

1.2 Deployment techniques

Cloud computing has become a corner-stone of successful businesses, offering up scalable on-demand self-service [1], and allowing the provisioning of computing resources - whether that be servers, storage, networking, etc - with minimal management effort [2]. It consists of three service models and four deployment models and it is essential to choose the appropriate combination of the two for this business.

The three service models are Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). Definitions vary but SaaS is generally considered to deal with ready-to-consume software applications which address a specific business function and offers limited customisation for the cloud consumer [3]. PaaS solutions provide the environment and software platforms to allow consumers to develop and host their own applications [3]. IaaS solutions provide virtual computing and storage resources to the consumers and offer the most control as the user must manage the content of the virtual resources [3]. A SaaS solution is not applicable in this scenario as the business needs to manage its own custom software stack. Furthermore, I believe IaaS would be the better option as it offers a higher level of scalability and reliability compared to PaaS.

The four deployment models in use are as follows: Public cloud which involves a cloud provider leasing its infrastructure to many customers; Private Cloud

which involves a company managing its own cloud; Hybrid cloud which involves using both a private and public cloud; Community cloud which involves organisations collaborating on a community-specific cloud [3]. For the purposes of this essay I will not consider community clouds as they are not widely used and do not offer any particular benefits over the other models.

Public clouds offer the benefit of low capital expenditure (CapEx) as no physical hardware is needed in the business however, the operating expenses (OpEx) can be high. Private clouds have a much higher CapEx but a lower overall OpEx. The benefits of a private cloud involve being able to use the same virtualisation technology used within a public cloud without the security risks that come from the multi-tenant environment created in a public cloud (as described in section 2). One downside to a private cloud is the lack of scalability when compared to a public cloud, and as this business requires an infinitely scalable solution a full private cloud solution is not appropriate. Hybrid clouds try to solve this issue by using cloud bursting to extend private clouds into public clouds at times of extreme loads, and they tend to support higher resilience, availability and reliability [3].

In this case, assuming that the business has a website component and a database component, I would use a hybrid solution, hosting the website in the public cloud (so as to not have latency problems for users) and hosting the database in a private cloud (particularly if it stores any sensitive user data).

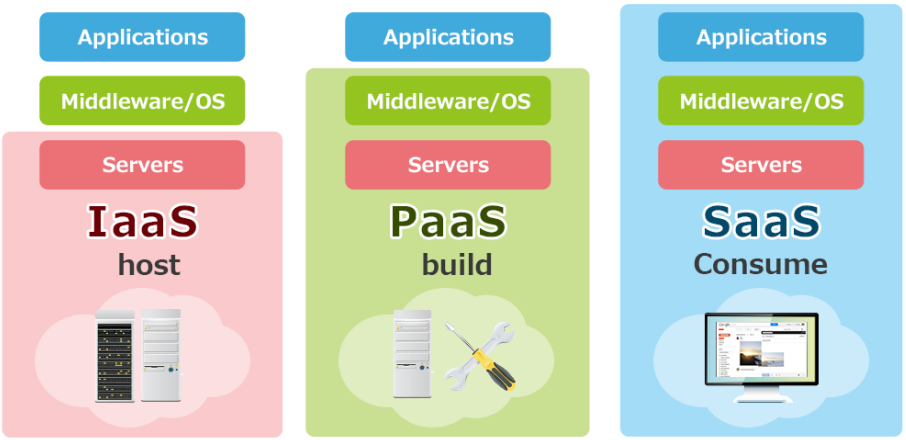

Figure 1: Difference between SaaS, PaaS and IaaS [4]

1.3 Autonomic Management

Clouds are complex [5], large-scale, distributed and dynamic; the management of such an elastic infrastructure is a challenging task requiring co-optimisation at multiple layers [6]. Quality of Service (QoS) requirements form the basis of Service Level Agreements (SLAs) and any violation of these results in a penalty for the provider so there is a need for a reliable, autonomic method

for provisioning resources to users. The idea of autonomic management is to provide self-monitoring, self-repairing, self-optimising, self-protecting, self-managing systems [7], [6], [8].

There have been a few proposed methods to provide autonomic management and they all rely on a way of monitoring the delivered services to ensure QoS requirements and avoid SLA violations.

[8] proposes an agent-based architecture where resources and virtual machines (VMs) are associated with worker agents that monitor changes in their local environments and take adaptive actions supervised by a network of management processes to achieve overall QoS. More recently, [5] proposed an SLA-aware solution (called STAR) which focuses on reducing the SLA violation rate by analysing cloud workloads to determine the feasibility of porting specific applications to the cloud, classifying the workloads based on security needs, network needs, variability of load and other QoS metrics, verifying availability of resources, and then assigning resources via IBM's autonomic model which consists of four stages: monitoring; analysing; planning; executing. Performance is monitored throughout execution to maintain efficiency by allocating more resources or re-negotiating SLAs depending on the scenario; this would be useful for this business as there will be varying demand load and a need for infinite scalability, meaning the provider will need an autonomic system to manage the resources and conform to the SLAs during runtime.

When it comes to deploying and executing non-native applications on the cloud, a problem arises as some applications may take days or even weeks to complete; monitoring is needed to ensure work is not lost just before completion. [7] explains how monitoring is usually realised by heartbeats (presence notification messages send from each virtual machine (VM) to a monitor).

1.4 Scalability under constraints

Typically load varies on a website or application due to a variety of factors including time of day and season and due to this there is a need to deal with varying capacity [9]. Rapid elasticity (also known as autoscaling) solves this problem and is one of the five essential characteristics that attracts users to the cloud [1]. Within a cloud, autoscaling must be realised to not only increase resources to meet peak demands but also scale down resources to optimise costs at times of low demand. An example of real-world autoscaling is Amazon EC2 in which images of client software can be created and loaded on to machines and when not required, the machines can be released. This leads to a potential problem known as thrashing in which due to frequent variation of workload, machines can be added and released too often, wasting resources.

A desirable solution for autoscaling would require the ability to predict the incoming workload and allocate resources pre-emptively. [9] proposes a model predictive control method to estimate system behaviour and predict future workload to acquire appropriate resources, while minimising cost.

The main problem concerning autoscaling is the ability for users to perform Distributed Denial of Service (DDoS) attacks in which they make legitimate requests for a service [10]; in this case, due to the requirement of infinite scalability, it would cause the service to cloud-burst in order to keep pace with the scale of requests, and the cost of sustaining operations may be so expensive that it threatens the economic sustainability of the business (known as an Economic Denial of Sustainability (EDoS) attack). [10] provides a mitigation method called In-Cloud Scrubber Service in which functionality is provided on a pay-per-use basis to generate and verify cryptographic puzzles in order to prove the legitimacy of service requests.

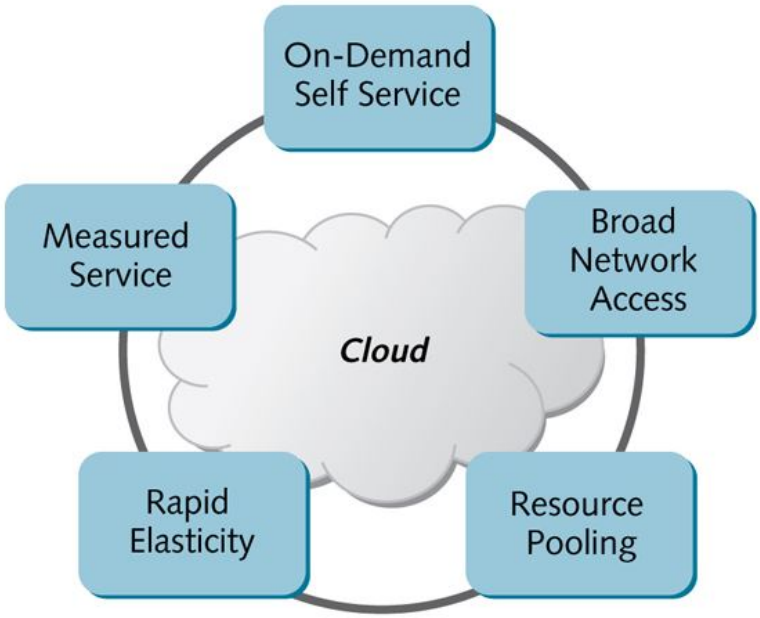

Figure 1.2 The essential characteristics of cloud computing

Figure 2: Shows rapid elasticity as one of the main 5 characteristics of the cloud [11]

1.5 Product offerings from different providers

There are many available products on offer from various cloud service providers (CSPs) but I will mainly focus on the major providers: Amazon, Google, Microsoft and IBM as these have been shown to work with many high-profile users and this business is migrating a business-critical infrastructure.

Amazon offers an elastic compute public cloud service, Amazon EC2 [12],

as part of its Amazon Web Services (AWS) system; this is an IaaS architecture that allows pay-per-second use of virtual resources. It has inbuilt autoscaling and allows you to increase or decrease capacity within minutes. Furthermore, it is integrated with most other AWS services such as S3 (Storage) so it is easy to configure multiple applications in the same cloud. The SLA commitment is 99.99% availability for each EC2 region.

Google also offers a public cloud IaaS solution in the form of Google Cloud Platform [13]. Similarly to Amazon EC2, it has autoscaling and uses a pay-per-second pricing model however, it has benefits over EC2 as it offers sustained-use discounts and also allows users to create custom machines which may be useful when hosting the custom software stack and can help to optimise costs by avoiding over-provisioning of resources. Google also offers Google Cloud Storage Nearline which is an archiving system that offers sub-second data availability which could be used to store a backup of the business's database in the public cloud.

IBM [14] offers a public IaaS solution but also focuses on private and multi-cloud solutions to offer flexibility, control, security and easy integration between clouds for a hybrid solution.

Microsoft Azure [15] focuses more on private and hybrid solutions (Azure Stack) but does offer public IaaS solutions, also with Azure autoscaling.

In terms of combining clouds for a hybrid solution, Eucalyptus [16] and Rackspace [**?**] are two popular services. Eucalyptus is open source software for building AWS-compatible private clouds and Rackspace incorporates RackConnect Global which provides private network connectivity between Rackspace and other data-centres and the cloud provider of your choice; including the main cloud providers mentioned previously. These two systems allow a hybrid solution to be created with ease so as to use public cloud for pay-as-you-go scalability and low-latency web hosting, and to use private cloud for enhanced security, control and to host sensitive data. Rackspace also offers support for creating OpenStack private clouds; OpenStack is open source and offers large pools of computing, storage and networking resources [17].

2 Security (include where responsibilities and liabilities are e.g Cloud Provider, Company or Consumer)

2.1 Data management and protection

Data management and protection is an important issue to consider when migrating this business infrastructure as it may deal with user data or business critical data and any data breaches would not only negatively impact user experience and potentially have legal ramifications but also lead to a negative reputation for the company, resulting further in reduced revenue which negates the cost benefits of using the cloud.

Due to the characteristics of the cloud, applications and data have no fixed infrastructure and security boundaries; user data may then be accessed by unauthorised users [18]. To consider data protection and management, it is important to examine all stages of the life cycle of data within the cloud and explain the risks and mitigation techniques at each stage. The stages and risks are as follows:

- Generation: In traditional environments, organisations own and manage the data themselves, but in the cloud it needs to be considered how to maintain that ownership. Users are entitled to know what private data is being collected and so this needs to be tracked through the cloud.
- Transfer: Need to ensure data confidentiality and integrity when transferring data across enterprise boundaries.
- Use: Often the data used by cloud-based applications is not encrypted as it will lead to problems with indexing and queries. Due to multi-tenancy, this data is usually stored with other users' data which is a threat to security.
- Share: Sharing data between parties renders data permissions more complex.
- Storage: Data is often encrypted for storage, however must consider the computational efficiency of encrypting large amounts of data in the cloud, and also be aware of key management issues; cloud providers manage the encryption keys which is a complex process and can lead to some negative scenarios.
- Archival
- Destruction: Need to ensure deleted data is not recoverable

One of the biggest issues to a business in my opinion is storing and using unencrypted data in the cloud; in 2009 IBM developed an encryption scheme that allowed some data to be processed without being decrypted so this may be a solution [18]. In general, the key to data privacy protection is to separate sensitive data from non-sensitive data and then to encrypt the sensitive data; for this reason I believe when migrating this business infrastructure it is best to host any database system in a private on-site cloud.

[1] and [19] also explain data breaches and data loss or leakage as two of most important threats; to mitigate the effects of data breaches you could encrypt the data, however if you lose the key you lose the data. Data loss can occur due to malicious intent, accidental deletion or by a physical event leading to permanent loss; this can result in a loss of finance, loss of trust or even legal ramifications. One way to potentially avoid this is to archive data in a different cloud. For example, in this scenario, archiving the database that is hosted on the private cloud in another secure public cloud service, possibly encrypted.

2.2 Potential threat vectors (network, tenant, infrastructure, datacentre etc.) and Mitigation techniques (automatic and manual)

Privacy and security issues are often cited as the main obstacle to the adoption of cloud computing for enterprises [1]. Security risks are more numerous when dealing with IaaS as this gives more control of virtual resources to the consumer. IaaS clouds inherit security concerns from the technologies they use but also have new, and more challenging, ones created by multi-tenancy (large amounts of users sharing resources)[19]. It is important to note that a lot of the security risks that result from multi-tenancy are not applicable to a private cloud, hence why it is best to host private data in a private cloud for this business. [19] goes on to state how their are 3 domains within the IaaS model: machine virtualisation, network virtualisation, physical; each domain has its own risks and mitigation strategies. Throughout academic research, the same main threat vectors occur when considering an IaaS cloud as we are in this essay [19], [1]:

Threat	Description	Threat Vector
Abuse of cloud computing	Cloud resources can be used to break encryption, launch DDoS attacks, propagate malware and more	Tenant
Data breaches/Data loss	These risks have been explained in the previous section	Data-centre, Infrastructure, Network
Insecure APIs	Cloud security is dependent on the security of the APIs used to access the resources. Weak APIs expose the system to malicious or unidentified users	Infrastructure, Tenant
Account/Service Hijacking	A malicious individual can compromise confidentiality of services and also use well-known exploitation methods to eavesdrop on activities and transactions, and even manipulate data and responses	Network, Infrastructure
Malicious Insiders	An insider can have a huge impact on trust and finances.	Data-centre
Shared technology issues	IaaS vendors deliver scalability by sharing resources (CPU, caches etc.) which leads to potential problems involving multiple companies hosting applications on the same physical system.	Infrastructure, Tenant, Data-centre

To deal with these issues, [1] proposes two main methods: Defense in depth in which individual security controls are combined to create a more complete and robust solution; Honey pot in which a decoy system is set up with vulnerabilities with the aim of collecting information about any intruders that are attempting to attack VMs on the same system as the honey pot VM. [1] also mentions using monitoring of the network and available resources with anticipatory measures to protect against DoS attacks.

[19] explains a VM life cycle which involves definition, creation, customisation, transportation to a VMM, storage in a repository, deployment, running and undeployment; the most risky stages are transportation, storage, deployment and runtime. These stages correspond to setting up an OS, building images, and deploying the VM into a Virtual Machine Monitor (VMM).

[19] also explains how to use cryptographic and access control techniques to protect VMs at the storage/transportation stage. As security in the machine virtualisation domain, the VMM hides the fact that the VM is actually a VM by employing the use of a container. In the physical domain, the providers must secure many data-centres, distributed globally. IBM developed a virtual Trusted Platform Model (TPM) to provide basic security-related functions to the VMs. To reduce some of the inherent multi-tenancy vulnerabilities, IBM also uses a Trusted Virtual Data-centre (TVDc) approach to define which VMs can communicate and access resources.

VM image management is also an issue since VM images need to be moved from in-house trusted facilities to a cloud provider through unsecured networks [19]. A solution is to encrypt several small files that are concatenated and decrypted to form the VM image.

When considered which cloud service provider (CSP) to use for this business, it is important to consider how each CSP will cater to the security needs. [20] proposes a framework for ranking cloud security services using qualitative user requirements to avoid the user having to have detailed knowledge and extends SLAs to include security features, creating a secSLA. This may be useful to consider when choosing the final provider.

3 Identify and discuss the implications of other new trends that are emerging within Cloud Computing (focus on IaaS)

It is clear that cloud computing is here to stay, and 74% of Tech Chief Financial Officers say cloud computing will have the most measurable impact on their business in 2017 [**?**]. As more and more users turn to the cloud, big companies have started looking toward the future of cloud technologies and the features that will form the basis of the cloud.

[21] explains how containers will see more mainstream adoption in IaaS clouds. Cloud containers offer an alternative to virtual machines [22]. They are designed to virtualise a single application (e.g a MySQL container) by creating an isolation boundary at the application level rather than at the server level so that if anything goes wrong in a container, it does not affect the whole VM. As containers do not need a full OS to be installed more containers can be deployed on a server than VMs, meaning more scalability using less hardware can be achieved. Container virtualisation could be an alternative to reduce the

virtualisation overhead and therefore improve the resource usage in data-centres [23]. They can also reduce the start-up time of replicas, and once they carry both an application and all of their dependencies, it can be used to automate deployment and scaling of applications, databases and backend services which would improve business operations. All major cloud providers are starting to offer container services so it would be easy to update this solution at a later date.

[21] also mentions multi-vendor approach as one of the next big trends in cloud computing. A multi-vendor approach involves hosting different parts of the business infrastructure on different clouds. As mentioned earlier in this report, companies like Rackspace and Microsoft are already enabling the use of multiple cloud vendors to work in tandem across different deployments, and due to the beneficial increases in availability, reliability and security that arise from using multiple cloud vendors, this type of approach will only grow in popularity.

[24] mentions artificial intelligence (AI) and machine learning as one of the next big trends and the main cloud providers have already started creating AI-based cloud services to meet the growing demand in this area.

AWS includes machine learning services for image recognition (AWS Rekognition), text-to-speech deep learning (Polly) and the engine that powers Alexa (Lex). Google also offers a Cloud Machine Learning Engine which helps machine learning engineers build models based on its open source TensorFlow deep learning library, as well as APIs for things like natural language processing, translation and computer vision. Microsoft's Azure Machine Learning Studio allows specialist developers to write, test and deploy algorithms, as well as a marketplace for off-the-shelf APIs [25]. These examples show how the major cloud providers are already embracing artificial intelligence and will continue to do so as the demand increases; if the business wants to incorporate machine learning in the future it should be possible with any one of these providers.

The implications of these companies introducing more AI-based services is that machine learning can continue to grow and improve as due to the nature of the cloud, more data and computational power is available to perform the machine learning, thereby leading to more powerful solutions and from a business standpoint, creating more revenue.

References

[1] Communications Society, ed., *2009 IEEE 17th International Workshop on Quality of Service (IWQoS 2009): Charleston, SC, USA, 13 - 15 July 2009.* Piscataway, NJ: IEEE, 2009. OCLC: 930970845.

[2] A. P. Mell (NIST) and A. T. Grance (NIST), "https://csrc.nist.gov/publications/detail/sp/800-145/final," Dec. 2017.

[3] I. M. Abbadi, *Cloud management and security*. Chichester, West Sussex ; Hoboken, NJ: John Wiley & Sons Inc, 2014.

[4] B. Lamouchi, "Cloud Computing: Comparing SaaS, PaaS and IaaS," Apr. 2017.

[5] S. Singh, I. Chana, and R. Buyya, "STAR: SLA-aware Autonomic Management of Cloud Resources," *IEEE Transactions on Cloud Computing*, vol. PP, no. 99, pp. 1–1, 2017.

[6] R. Buyya, R. N. Calheiros, and X. Li, "Autonomic Cloud computing: Open challenges and architectural elements," in *2012 Third International Conference on Emerging Applications of Information Technology*, pp. 3–10, Nov. 2012.

[7] A. F. Leite, V. Alves, G. N. Rodrigues, C. Tadonki, C. Eisenbeis, and A. C. M. A. De Melo, "Autonomic Provisioning, Configuration, and Management of Inter-cloud Environments Based on a Software Product Line Engineering Method," in *Cloud and Autonomic Computing (ICCAC), 2016 International Conference on*, pp. 72–83, IEEE, 2016.

[8] Q. Liu, D. D. Silva, G. K. Theodoropoulos, and E. S. Liu, "Towards an agent-based symbiotic architecture for autonomic management of virtualized data centers," in *Proceedings of the 2012 Winter Simulation Conference (WSC)*, pp. 1–13, Dec. 2012.

[9] N. Roy, A. Dubey, and A. Gokhale, "Efficient Autoscaling in the Cloud Using Predictive Models for Workload Forecasting," in *2011 IEEE 4th International Conference on Cloud Computing*, pp. 500–507, July 2011.

[10] M. N. Kumar, P. Sujatha, V. Kalva, R. Nagori, A. K. Katukojwala, and M. Kumar, "Mitigating Economic Denial of Sustainability (EDoS) in Cloud Computing Using In-cloud Scrubber Service," in *2012 Fourth International Conference on Computational Intelligence and Communication Networks*, pp. 535–539, Nov. 2012.

[11] "https://www.datamation.com/cloud-computing/what-is-cloud-service.html," Dec. 2017.

[12] "https://aws.amazon.com/ec2/," Dec. 2017.

[13] "https://cloud.google.com/products/compute/," Dec. 2017.

[14] "https://www.ibm.com/cloud/," Dec. 2017.

[15] "https://azure.microsoft.com/en-gb/," Dec. 2017.

[16] "eucalyptus: Eucalyptus Cloud-computing Platform," Dec. 2017. original-date: 2011-05-24T20:36:49Z.

[17] "https://www.openstack.org/," Dec. 2017.

[18] D. Chen and H. Zhao, "Data Security and Privacy Protection Issues in Cloud Computing," pp. 647–651, IEEE, Mar. 2012.

[19] L. M. Vaquero, L. Rodero-Merino, and D. Morn, "Locking the Sky: A Survey on IaaS Cloud Security," *Computing*, vol. 91, pp. 93–118, Jan. 2011.

[20] A. Taha, R. Trapero, J. Luna, and N. Suri, "A Framework for Ranking Cloud Security Services," in *2017 IEEE International Conference on Services Computing (SCC)*, pp. 322–329, June 2017.

[21] "http://www.silicon.co.uk/cloud/trends-expect-cloud-2017-203891," Dec. 2017.

[22] "http://searchcloudsecurity.techtarget.com/feature/Cloud-containers-what-they-are-and-how-they-work," Dec. 2017.

[23] E. F. Coutinho, F. R. d. C. Sousa, P. A. L. Rego, D. G. Gomes, and J. N. d. Souza, "Elasticity in cloud computing: a survey," *annals of telecommunications - annales des tlcommunications*, vol. 70, pp. 289–309, Aug. 2015.

[24] "https://www.insight.com/content/insight-web/en_us/learn/content/2017/02132017-3-iaas-cloud-computing-trends-to-watch.html," Dec. 2017.

[25] "//aws.amazon.com/compliance/shared-responsibility-model/," Dec. 2017.