Cloud Computing Techniques for Cloud Bursting

Godina Venkata Akhil Chandra
Dept. of Computer Science and Engineering
Lovely Professional University
Jalandhar, INDIA
example@xyz.com

Abstract

Recent years have seen the emergence of cloud computing, a brand-new paradigm that promises to deliver services online. Cloud bursting is a method used in a hybrid cloud environment that mixes internal (or local) organisational resources with public cloud resources. It was once used to handle work overload on local resources, speed up the execution of distributed applications while maintaining the necessary degree of QoS, and make efficient use of personal resources. The key considerations when using cloud bursting are deciding how many and what kind of resources will be provisioned.

The type of workload to be burst to the public cloud and the timing of resource release are also important factors that should be considered. Researchers are keen on tackling these problems. The purpose of this study is to evaluate recent studies on cloud bursting and resource provisioning. We will explore how each study address the problem.

1. Introduction

Cloud computing is the on-demand use of computer resources stationed at a distant data centre and provided by a cloud services provider, such as programmes, servers (both physical and virtual), data storage, development tools, networking capabilities, and more. To accomplish coherence, cloud computing depends on resource sharing, which commonly employs a pay-as-you-go model.

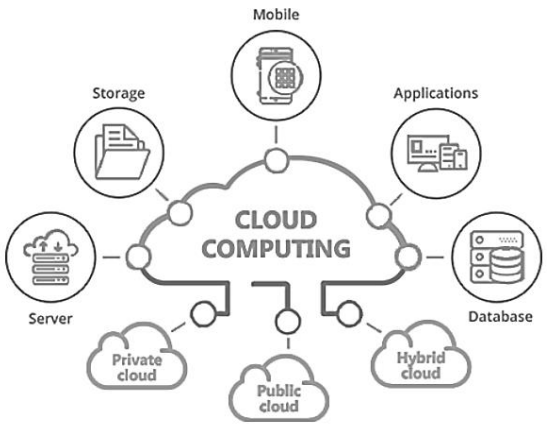

Fig 1.1: Visual representation of cloud computing

Following are the advantages of using cloud computing:

Agility:

You can easily access a wide variety of technologies with the help of the cloud, which allows you to innovate more quickly and create almost anything you can think of. You may instantly spin up resources as you require them, including Internet of Things, machine learning, data lakes, analytics, and infrastructure services like computation, storage, and databases.

Technology services may be deployed quickly, allowing you to move from idea to implementation much more quickly than in the past. This allows you the flexibility to try new things, test novel customer experience concepts, and reinvent your company.

Adaptability:

With cloud computing, you can handle future spikes in business activity despite having to over-provision resources now. As an alternate solution, you only provision the resources that you truly require. As your company's demands change, you may scale these resources up or down to immediately increase and decrease capacity.

Lower IT costs:

With the cloud, you can swap out fixed costs (such data centres and database hardware) for variable costs and only pay for the IT you truly use. Also, because of the economies of scale, the variable costs are considerably cheaper than what you would spend to do it yourself.

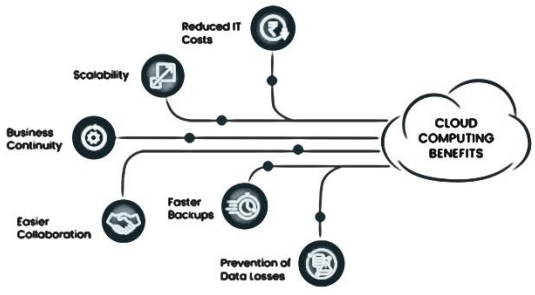

Fig 1.2: Benefits of cloud computing

2. Types of clouds

For clouds, there are three main types of deployment methods (Public, Private, Hybrid) and three main service models like Software-as-a-Service, Platform-as-a-Service, and Infrastructure-as-a-Service). In IaaS clouds that provide the basic IT resources, such as networking, processing power, and storage capacity, are the subject of this article.

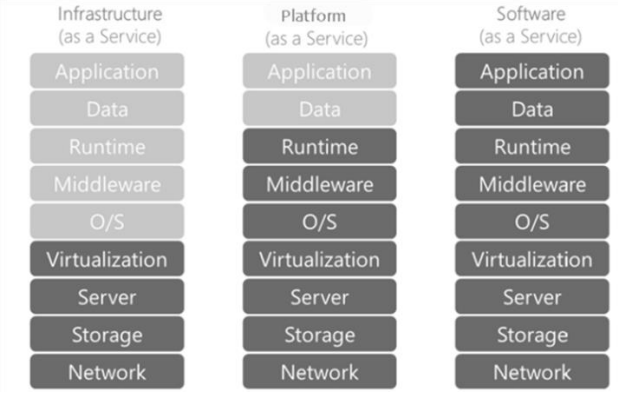

Fig 2.1: Types of service models IaaS, PaaS and SaaS

In 2006 the Elastic Cloud Computing (EC2) service was introduced by Amazon, making them the first provider of public IaaS. Numerous suppliers have since entered the market, and private

Cloud-based solutions also emerged. Private IaaS clouds allow businesses to leverage cloud computing using their own IT infrastructure. These software solutions typically assist creating a hybrid cloud environment by fusing the private and public clouds. The technique called cloud bursting allows the companies to dynamically extend their infrastructure by renting third-party resources.

3. Cloud Bursting

Cloud bursting is an application deployment technique in which an application that runs on onsite, private cloud or data centre and bursts into a public cloud when the demand for computing capacity spikes and reaches the threshold values. This deployment model gives an organization access to more computing resources when needed. Cloud Bursting was first proposed by Amazon's Jeff Barr.

Cloud bursting provides a company with more flexibility to handle spikes in IT demand when compute demand surpasses a private cloud's capacity. Cloud bursting also makes local resources available for other crucial applications.

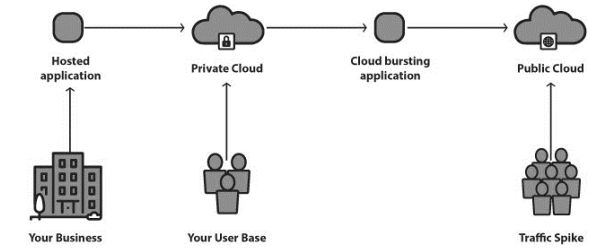

Fig 3.1: Cloud bursting architecture

4. Cloud Bursting Importance

In the past, businesses used private data centres or collocation facilities to store and operate their own computing infrastructure, including servers, storage units, and network hardware. Organizations can now employ publicly accessible computing infrastructure that is safe, readily scales up or down to meet workload demands, and is accessible in many places across the world thanks to the rise of third-party cloud providers like Amazon Web Services, Microsoft Azure, Oracle, etc. Using infrastructure that was entirely handled by others became more practical. To differentiate between internal infrastructure and external third-party cloud resources, the phrase "public cloud" was developed.

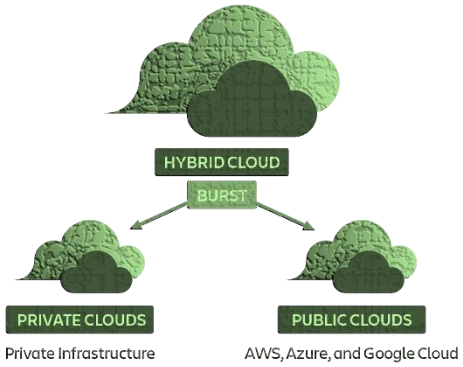

Fig 4.1: Cloud computing in a nutshell

Many organizations want to leverage both the advantages of the public cloud and their current on-premises computer infrastructure. When they run out of on-premises computing capabilities, they can implement a cloud-bursting hybrid cloud architecture to access public cloud resources.

Cloud bursts can be triggered either automatically based on high usage demands or manually via a request.

An organisation should consider its security requirements, any platform capabilities, and any compliance needs while adopting cloud bursting. Critical applications or data should not be used for cloud bursting

since the data will move across clouds, although private clouds are usually more secure than public clouds.

5. Cloud Bursting Requirements

To achieve the best implementation of cloud bursting, a list of prerequisites for both private and public clouds is provided. The business should offer a system to check the level of SLA required for service provisioning in the internal cloud as well as other ancillary needs like energy consumption. Before moving data and services to the public cloud, the internal cloud must establish the nature of its services, including crucially, data privacy and some restrictions. In public cloud, a correct management must be found to satisfy QoS which is agreed with internal cloud company, furthermore it should satisfy the capacity needed by internal cloud and execute dynamically.

6. How Cloud Bursting works

IT administrators help determine capacity thresholds for apps in the private cloud. When workload capacity nears its threshold, the used application automatically switches over into the public cloud and traffic is pointed towards it. As the spike in resource needs diminish, the application is relocated back to the private cloud or on-premises infrastructure.
Organization can take one of the following approaches to cloud bursting:

6.1 Distributed Load Balancer

Applications run between a public cloud and a data centre with distributed load balancing. When workload traffic reaches a certain threshold, a similar environment diverts it to a public cloud. This approach demands the deployment of an application both locally and on the public cloud, as well as the use of load balancing techniques to divide traffic.

Modern high-traffic websites must quickly and reliably respond to hundreds of thousands, if not millions, of concurrent user or client requests for the right text, photos, videos, or application data. Modern computing best practise typically necessitates the addition of more servers in order to cost-effectively scale to handle these massive volumes.

A load balancer serves as the "traffic police" in front of your servers, distributing client requests among all servers equipped to handle them in a way that maximises speed and capacity utilisation and makes sure that no server is overworked, which can result in performance degradation. The load balancer routes traffic to the active servers in case one server goes offline. The load balancer initiates request to a new server when it is added to the server group.

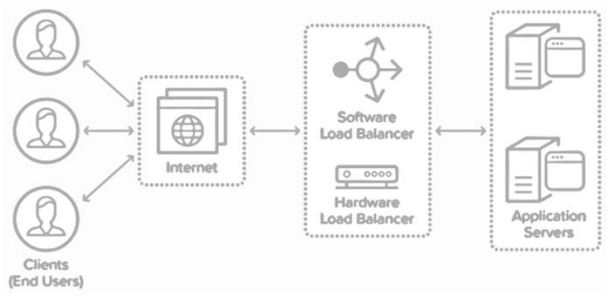

Fig 6.1: Load balancer visualized

6.1 Types of load balancer:

6.1.2 Round Robin Load Balancing

The most basic and widely used method for load balancing is round-robin load balancing. Application servers receive client requests in a straightforward rotation. The first client request is sent to the first application server in the list, the second client request is sent to the second application server, the third client request is delivered to the third application server, the fourth client request is sent to the first application server, and so on.

When predictable client request streams are distributed across a server farm with members who have roughly equal processing power and resource availability, round robin load balancing is the best option (such as network bandwidth and storage).

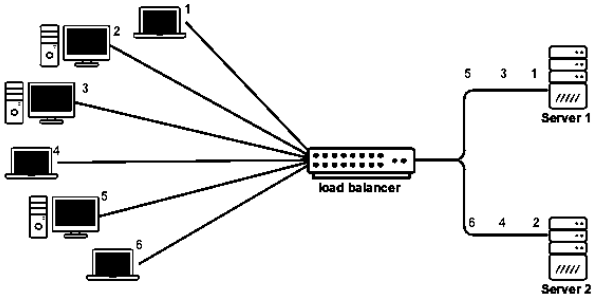

Fig 6.2: Round Robin Load balancer

6.2.3 Least Connections Load Balancing

A dynamic load balancing mechanism called least connection load balancing distributes client requests to the application server that has the fewest active connections at the moment, the client request is received. This algorithm takes the active connection load into account. When application servers have identical specs, one server may be overwhelmed due to longer-lasting connections. This method works well with incoming requests that have variable connection latency and a group of servers that have similar processing capabilities and resource availability.

6.2.4 Fixed Weighting Load Balancing Method

In order to indicate the relative traffic-handling capacity of each server in the server farm, the administrator assigns a weight to each application server using fixed weighting, a load balancing algorithm. All traffic will be sent to the application server with the highest weight. All traffic will be routed to the application server with the

next greatest weight if the application server with the highest weight fails. This approach is suitable for workloads where a single server can handle all anticipated incoming requests and there are one or more "hot spare" servers prepared to take over if the active server fails.

6.2.5 Source IP Hash Balancing Method
The source IP hash load balancing algorithm uses the source and destination IP addresses of the client request to generate a unique hash key which is used to allocate the client to a particular server. The client request is sent to the same server it previously used because the key can be created if the connection is terminated. When it's essential for a client to connect to the same server during each subsequent connection, this solution is best suited.

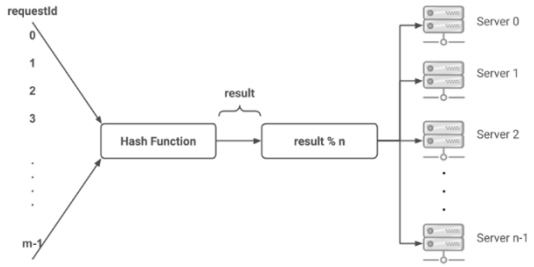

Fig 6.3: Source IP hash balancing

6.2 Manual Bursting
A company can manually deploy and deprovision cloud services and resources by using manual bursting. For transitory large cloud installations, when higher traffic is anticipated, or to free up local resources for mission-critical applications, manual cloud bursting is appropriate.

6.3 Automated Bursting
A company must establish policies to specify how to manage automated bursting. Once configured, a private cloud application can seamlessly transition to a public cloud. The application is changed automatically by software. A company can use this to supply cloud resources quickly and precisely.

7. Benefits of Cloud Bursting

Cost
Only when additional computing resources are required does an organisation pay for them. Like public cloud infrastructure, private cloud infrastructure costs can be reduced by using few resources.

Flexibility
Cloud bursting can easily adapt to changes in cloud capacity. Moreover, private cloud resources are freed.

Business Continuity
Without disturbing its users, a programme has the ability to burst over into the public cloud.

High traffic times
Cloud bursting can be utilised to accommodate any anticipated or unanticipated peaks in compute resource demands in the event that a company anticipates a sudden increase in traffic, such as during a holiday.

8.How to evaluate whether cloud bursting is the best strategy for your organization

Businesses can grow without building and managing additional resources on-premises by using cloud bursting. These assets are overhead capacity that you probably will not utilise frequently enough to justify the investment.

Bear in mind that by possibly distributing sensitive data to a public cloud, cloud bursting also poses security risks. When deploying cloud bursting, consider how important it is to keep your data's integrity while remaining cost-effective.

You can utilise cloud bursting to protect yourself against service interruptions brought on by demand spikes if your company offers fewer sensitive services or applications, or if you would mostly be unaffected when a public cloud provider is compromised.

Bursting might not be an option, though, if security is more important than cost-effectiveness. Some security-sensitive tasks can only be run locally.

9. When does an organization need cloud bursting?

For high-performance, non-critical applications that deal with non-sensitive data, cloud bursting is advised. You can either transfer an application to the public cloud to free up local resources for business-critical applications, or you can deploy an application locally and burst it to the public cloud to handle peak demand. Applications that do not require integration with other applications, systems, or data centre-internal components or a complicated application delivery infrastructure perform well when using cloud bursting.

A company must consider security measures and legal compliance requirements while adopting cloud bursting. For instance, cloud bursting is frequently mentioned as a practical solution for retailers who encounter demand peaks during the holiday shopping season. However, cloud computing service providers might not always provide a setting that complies with the Payment Card Industry Data Security Standard, and retailers might be placing sensitive data at risk by blasting it to the public cloud.

Software development, analytics, big data modelling, and marketing initiatives can all benefit from cloud bursting. Organizations that work with big data or machine learning, for instance, can employ cloud bursting to create models that are larger than the capacity of their private clouds. If a company anticipates a

significant increase in traffic as a result of a marketing campaign, they can deploy cloud bursting in conjunction with it. Cloud bursting is supported by Amazon Web Services, Google Cloud, and Microsoft Azure, three cloud service providers.

10. Leading Cloud Bursting Providers

- Amazon Web Service
- Microsoft Azure
- Cisco Intercloud Fabric
- Equinix
- VMware NSX

11. Challenges in Cloud Bursting

Although the idea behind this application-tier cloud bursting appears straightforward, there are some crucial technical and commercial factors to consider if you wish to make use of this architecture. They include dealing with the security, latency, and cost challenges associated with communication across public and private clouds as well as bridging the discrepancies among various clouds. While there are methods for dealing with these technical difficulties, many businesses opt for alternatives to the conventional perspective of cloud bursting that put the emphasis on moving workloads to the appropriate cloud at the level of an application portfolio or an application service.

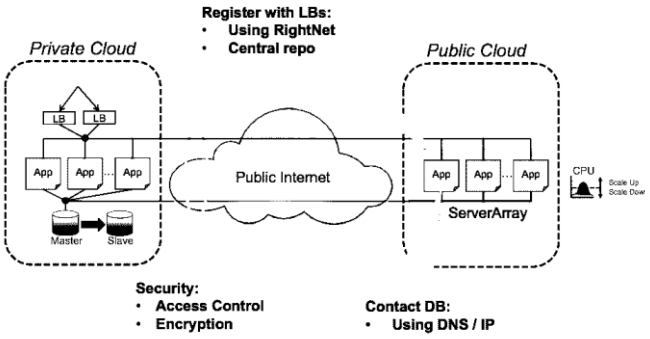

Fig 11.1: Security issues related to cloud bursting

The main challenges with cloud bursting are:

- Setting up and maintaining setups for various clouds.
- Establishing and maintaining a low-latency channel for cloud communication.
- Securing communications between the clouds.

11.1 Managing Configurations:

In an application, cloud bursting entails running the same types of resources - in this case, application servers - across different clouds. Nonetheless, there may be significant differences between the cloud service providers you select, necessitating the configuration and administration of your apps and their technology stacks across two distinct environments. A few of the variations include:

- Different hypervisor features and machine image formats.
- Several underlying hypervisors and hypervisor versions.
- Various clouds offer a variety of power and virtual machine kinds.
- Cloud APIs differ in behaviour as well as in the available calls.
- The different forms of storage that clouds provide, including object, block, permanent, and ephemeral storage, may behave differently or may not be offered by any other cloud at all.

- Network setups for clouds vary according to their nature. Others do not give the idea of availability zones. Some have subnets and ACLs, others have security groups, and some have both.

Also, this has knock-on consequences. For instance, having several storage subsystems and hypervisors typically necessitates using various base virtual machine images, which must be created for each cloud and updated with each new operating system and security patch.

11.2 Communication Latency

The application tier in the cloud bursting example above processes data that is being read from an on-premises database while being housed in the public cloud. Because the database access is not in a critical latency channel and the data is simply being read to generate thousands of static pages, this works pretty well.

Yet, if your application needs to transport a large quantity of data between your application tier and database when they are in different clouds, latency and throughput can become a problem. When communication takes place over the open Internet, the latency problem is very noticeable. Even tiny latencies can quickly accumulate since app-to-database communication can require numerous round trips for each front-end request.

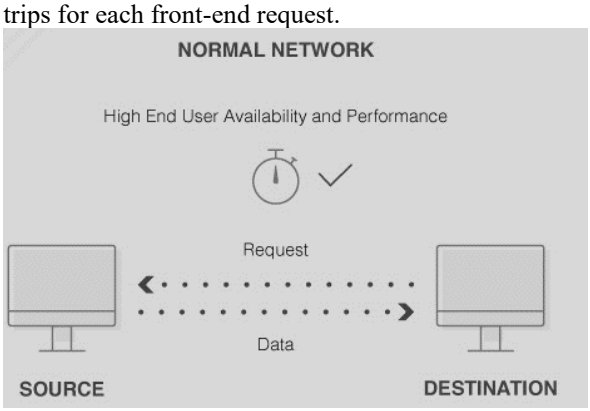

Fig 11.1: Communication without delay

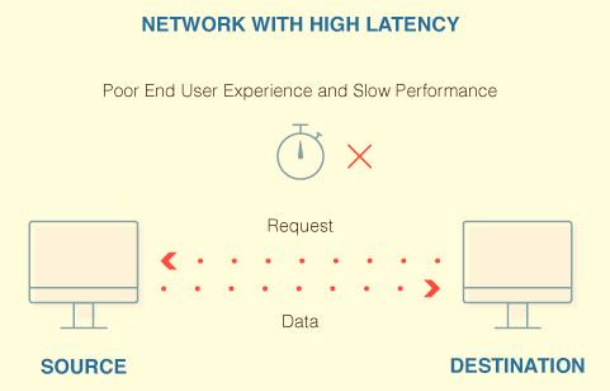

Fig 11.2: Communication with Delay

More bandwidth is needed for many applications since the volume and frequency of application-to-database interactions are not optimum. Whether the communication method uses a leased line or the public Internet, the cost of supplying it can be shockingly high. Some cloud service providers also charge for data intake or egress. In order to achieve your performance goals, this portion of the communication stream must be optimised in order to decrease latency and lower the cost of the bandwidth.

11.3 Security Handling

The issue of securing the communication line between the clouds is the last one to be addressed, which entails setting up encrypted channels, resolving the inescapable routing concerns, and meeting compliance and audit requirements.

These communication channels might need to be redundant at every level in order to meet application availability requirements, which raises the complexity of the routing process and raises the price of the equipment and provided pipes. Due to the requirement that new servers added to the array dynamically join your VPN, dynamic server allocation adds additional complexity.

12. Conclusion

In cloud computing, the term "cloud bursting" describes the provision of public cloud resources with private resources to handle surges in workload or to speed up application execution. The difficulties in accomplishing the cloud bursting goal are in selecting the best workload to burst and the optimum resource to provision. This term paper surveys past and recent solutions to these problems and groups them according to their goals. A quick summary of an idea, main goal, a method of operation, and a benefit of each research are described.

13. Acknowledgement

I would like to express our deepest gratitude to all those who have contributed to the successful completion of this term paper.

First and foremost, I would like to thank our teacher for providing us with invaluable guidance and support throughout the entire process. Your expertise and insights have been invaluable in shaping the direction and focus of this work. Our gratitude also goes to the various scholars and researchers whose published work has informed our research, and to the various institutions and organizations who have provided us with access to data and resources. Finally, I would like to thank our friends and family for their constant encouragement and support. Your love and understanding have sustained us through the challenges of this project and beyond. Thank you all for your contributions, both large and small, towards the successful completion of this paper.

14. References

1. Michael Beham, Marius Vlad, and Hans P Reiser. Intrusion detection and honeypots in nested virtualization environments. In Dependable Systems and Networks (DSN), 2013 43rd Annual IEEE/IFIP International Conference on, pages 1–6. IEEE, 2013.

2. Alex Fishman, Mike Rapoport, Evgeny Budilovsky, and Izik Eidus. Hvx: Virtualizing the cloud. In Presented as part of the 5th USENIX Workshop on Hot Topics in Cloud Computing. USENIX, 2013.

3. Tian Guo, Upendra Sharma, Timothy Wood, Sambit Sahu, and Prashant Shenoy. Seagull: Intelligent cloud bursting for enterprise applications. In Proceedings of the 2012 USENIX Conference on Annual Technical Conference, USENIX ATC'12, pages 33–33, Berkeley, CA, USA, 2012. USENIX Association.

4. HeeSeok Choi, TaeMuk Lyoo, JongBeom Lim, Daeyong Jung, Jihun Kang, Taeweon Suh, and Heonchang Yu. A study on performance comparison of cloud architectures using nested virtualization. In Ubiquitous Information Technologies and Applications, pages 77–84. Springer, 2014.

5. https://www.techtarget.com/searchcloudcomputing/definition/cloud-bursting

6. https://aws.amazon.com/what-is-cloud-computing/

7. R. N. Calheiros, C. Vecchiola, D. Karunamoorthy, and R. Buyya, "The Aneka platform and QoS-driven resource provisioning for elastic applications on hybrid Clouds," Future Gener. Comput. Syst., vol. 28, pp. 861-870, 2012.

8. P. Amiri, S. Rad and F. Isfahani, "Providing a Solution to Improve Pre-Copy Method for Migrating Virtual Machines in Cloud Infrastructure", *Journal of Theoretical & Applied Information Technology*, vol. 92, no. 2, Oct 2016.

9. N. Xue, H. Haugerud and A. Yazidi, Towards a Hybrid Cloud Platform Using Apache Mesos, Cham:Springer International Publishing, pp. 143-148, 2017.

10. https://kemptechnologies.com/load-balancer/load-balancing-algorithms-techniques

11. https://www.cloudflare.com/learning/performance/types-of-load-balancing-algorithms/

12. Mahantesh Birje, Praveen Challagidad, R.H. Goudar and Manisha Tapale, "Cloud computing review: Concepts technology challenges and security", *International Journal of Cloud Computing*, vol. 6, no. 32, 2017.

13. https://www.nginx.com/resources/glossary/load-balancing/

14. *Blog: Real Use cases: Why 50% of enterprise are choosing hybrid cloud*, [online] Available: https://www.cloudcomputing-news.net/news/2015/jun/25/why-are-50-of-enterprises-choosing-hybrid-cloud-real-use-cases.

15. https://www.itpro.com/cloud/cloud-computing/357078/what-is-cloud-bursting

16. T. Bicer, D. Chiu, and G. Agrawal, "Time and Cost Sensitive Data-Intensive Computing on Hybrid Clouds, " in Cluster, Cloud and Grid Computing (CCGrid), 2012 12th IEEE/ACM International Symposium on, 2012, pp. 636-643.

17. https://circleci.com/blog/what-is-cloud-bursting/