

Leveraging Power of Data Analytics to Enhance Efficiency of Resource Allocation in Cloud Computing Services - SaaS, PaaS and IaaS

¹Sujay V Shettar

Bachelor of Engineering,
Computer Science and Engineering,
Siddaganga Institute of Technology, Karnataka, India

Abstract: Cloud computing has evolved significantly through decades which has invariably enhanced the technology and the way resources are used for computing needs. Being a computing technology that is internet-based, it caters to the needs of consumers of the services – software, platform, infrastructure et al. providing computing resources in the pool to the users, it is a virtual pool of computing resources. Resource sharing being one of the most charismatic features of this paradigm, there arises a need of managing resources efficiently without causing downtime hence avoiding prodigality of resources and making the technology more cost-effective and providing seamless, user-friendly experience to the end users. This can be achieved by the approach of data analytics which helps in understanding the usage trends of a user and based on the analysis of the data gathered about the trends, make an efficient decision on allotting and revoking of critical resources so that it is back in the pool for essential disposal.

Index Terms: Data Analytics, Cloud Computing, Software as a Service, Platform as a Service, Infrastructure as a Service

I. INTRODUCTION

The exact definition of cloud computing is “A large-scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted, virtualized, dynamically scalable, managed computing power, storage, platforms, and services are delivered on demand to external customers over the Internet” [3]. The technology analysts at Gartner see cloud computing as a so-called “emerging technology” [2] on its way to the hype. Being an emerging computing paradigm, it aims at providing wide range of capabilities to its users including but not limited to resource pooling, elasticity, self-servicing capabilities, on-demand services, pay-as-you-go model, usability etc. The three major services of this paradigm being Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) have a very significant role to play in today’s cloud industry. With many offerings by the cloud and consequently massive number of live users, there arises a necessity of efficiently utilizing the resources at disposal and not merely allot them as per choice or deliberation of the user. This requires the provider to assess, analyze and subsequently make meaningful decisions in order to efficiently allocate resources. This involves analyzing the usage trends of the user as to how the user is utilizing the resources and to what extent.

To make informed decisions on the above-mentioned aspects, a powerful process called data analytics can be used which is the science of analyzing raw data to make conclusions about that information. Just like we use data analytics for the purpose of handling data that is derived from entities like social media, shopping details, demographic information, lifestyle details so on and so forth, we can also use the same methodologies to fetch various data from the cloud usage of the users to analyze the usage trend and effectively dispose the resources available in the pool so that users have a seamless experience while using the service – data like time of usage, amount of storage utilized, amount of memory utilized, units of CPU processing power requested and the process for which the resources are being requested for.

With more and more growing acceptance and adaptation of cloud-based services in both personal and public environments there is an invariable need of managing the resources very effectively to make the service cost-effective and enhance the performance as well. As this need has continued to grow so has the support and surrounding infrastructure needed to enhance the users, experience. In this scenario data analytics becomes a very powerful tool and a very significant solution for the problem.

II. LITERATURE SURVEY

Evolution of Cloud Computing

The power of cloud computing paradigm started to get recognized in the beginning of 21st century although most people were not fully aware of its benefits. As time progressed huge technologies started to realize the potential of cloud and started adapting their organizations to it and further developed and enhanced its capabilities and features to provide the services to their customers. With a colossal growth of understanding and acceptance of cloud-based applications ranging from social networking sites to business management and business intelligence related applications, more and more individuals are now highly interested in moving their

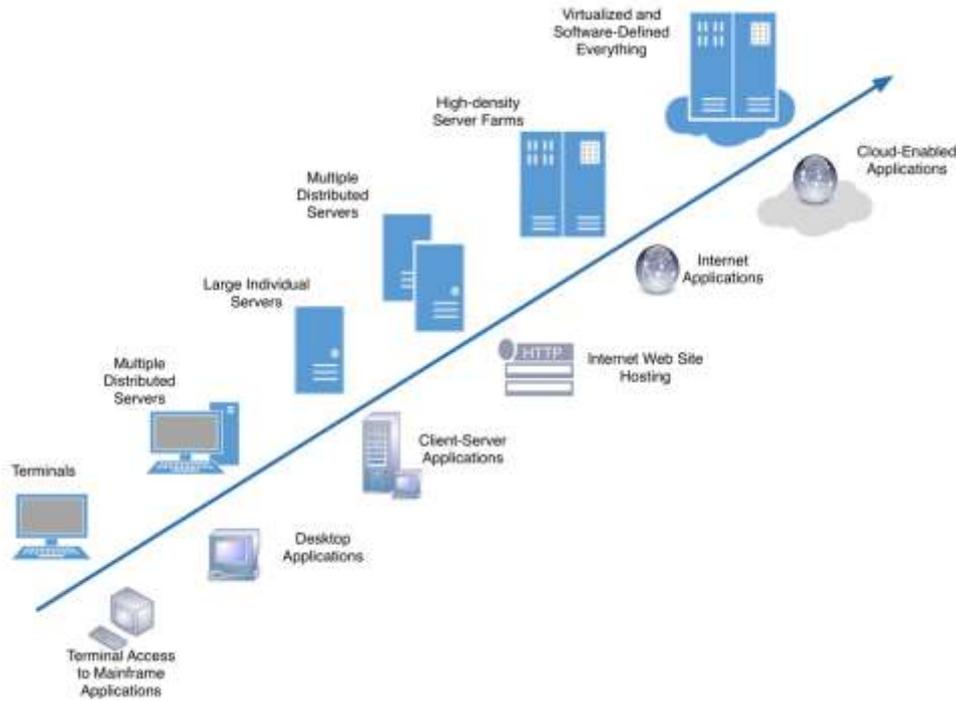


Figure 1: Evolution of Cloud Computing [4]

personal and business operations to a cloud computing environment. Big players in cloud like Google, Microsoft and Amazon have exponentially increased the quality and the power of their infrastructure and platform. Various applications from social networking sites like Facebook to cloud storage providers like OneDrive and dropbox utilize cloud for operations. There are various other critical applications as well which are hosted on cloud environment. Therefore, to ensure zero downtime and seamless experience for the users there is a need of analyzing the usage trends and allocating the resources contingent on requirement.

Figure 1 [4] illustrates the evolution of cloud computing from the times of usage of terminals until cloud-embedded applications. Initial usage of mainframe applications providing terminal access developed into desktop applications hosted on multiple distributed servers. With usage and requirements growing, the desktop applications evolved into client-server applications. Then started the expansion to the global scale which resulted in Internet website hosting creating internet applications that run from the server farms. But this proved to be an inefficient and an expensive usage methodology. Hence the technology evolved to become cloud-embedded applications that are virtualized and software-defined. With this paradigm started the concept of resource sharing.

Power of Data Analytics



Figure 2: Various Features of Cloud Analytics [5]

Data analytics is the process of converting raw data into meaningful information which can be leveraged to make highly informed decisions in various contexts. Since data analytics helps us arrive to an informed conclusion from the raw data that has been provided

it becomes a very important paradigm because it helps us to exponentially enhance the performance – be it businesses or in this case cloud services. Figure 2 [5] illustrates the various features of cloud analytics.

For example, we can measure the runtimes and downtimes when certain processes running and based on the data that we obtain we can come to certain conclusions for both runtime and downtime. We can understand the list of resources that a process requires and the quantity of these resources to the process requested for when the downtime occurred and based on this conclusion, we can make informed decisions when the same or a similar process is running and ensure that the resources are allocated in the required quantity hence proactively ensuring zero downtime. Similarly, we can follow the same process for understanding the performance of application during runtime and based on the metrics we can devise a method to enhance the performance of the application during current time as well.

Data Analytics in Cloud Computing

The realization of immense power that data analytics has to offer when it comes to cloud computing paradigm, has made us exploit this capability and leverage this power to draw conclusions from the information collected to make meaningful decisions that help us drive the process of resource allocation very efficiently and effectively providing a seamless experience to the users. Leveraging capabilities of data analytics has proved to be highly impactful and has created major success stories in today's technology era. Some of the great well-known and amazing success stories are Amazon salesforce Alibaba and many more. Amazon uses data analytics in various ways which helps it in making informed decisions that helps them make great profits. They collect data from the users about the products they are interested in, the necessity of those products and many more information of this sort which helps them recommend and provide the most precise product that the customer requires.

Netflix another popular software as a service cloud-based application utilizes is the full power of data analytics by tracking the movies users watch and recommends others they might enjoy, providing a service to clients while supporting the use of their product. All user information is remotely stored remotely on cloud, so users' preferences do not change from computer to computer. Netflix retains all the users' preferences and tastes in movies and shows, enabling them to create a television show that statistically appealed to a large portion of their audience based on their desired taste. [7]

The above-mentioned approach used by big firms like Amazon and Netflix can be utilized by adapting it to the process of allocating resources for cloud services based on what the users prefer for services. This will enable us to gather information and analyze the cloud usage trends of the user and ensure high-performance of the applications during runtime alongside maintaining zero downtime to the users so that they have a seamless experience while operating using the cloud computing paradigm.

III. PROBLEM STATEMENT

Maximizing profits is the main intention of their businesses. To maximize the profits, they need to devise an efficient strategy and to help them navigate through the strategy businesses have been using data analytics since sometime. Data analytics helps us minimize uncertainty by tracking the data patterns analyzing the trends and help devise meaningful business tactics. Data analytics not only helps us make meaningful decisions of about the future strategies, but also helps us understand how we can enhance and improve the current strategies.

We can utilize this understanding about the cloud computing paradigm and the power of data analytics that are at our disposal to solve the problem of allocating resources in efficiently with the users of cloud services request for them from the resource pool. We have to devise a mechanism of how we can collect the data of users, the trends of their usage, the applications they run, the resources that applications require, the quantity of resources in the resource pool, the rate of utilization of resources, the actual required amount of resources, and the requested amount of resources, the duration of utilization of the resources, the issues that occur during runtime and the issues that occur during downtime, the peak hours of utilization of resources, the time during which resources are utilized very less, the applications that are using resources not required for them to function, the applications that are unable to get resources allocated for them when requested and many more parameters.

IV. IMPORTANCE OF RESOURCE ALLOCATION

Cloud computing is the highly emerging paradigm having a huge resource pool that caters to users' needs where the users use these and pay as per their usages. Cloud computing is based on the virtualization technology, grid computing and service-oriented architecture (SOA). The cloud computing offers Infrastructure as a Service (IaaS) which is also a pay as you use model which allocates infrastructure resources based on the requirements of the users. [6]

We need to understand the significance of efficient and effective resource allocation in the present-day technology era of cloud computing. It is a process of allocating resources from resource pool to the applications hosted by the clients into the cloud. If the resources are not efficiently and effectively allocated/managed, then the resources are exhausted and the service experiences downtime resulting in a not so good experience to the users. We propose to leverage the power of data analytics for this purpose of efficient and effective allocation of resources to the users to ensure zero downtime and to give the users a seamless experience and thereby making the service cost effective.

V. PROPOSED DESIGN AND IMPLEMENTATION

Due to the ever-increasing demands of the users for cloud services, it results in the requirement of massive number of resources, and consequently becomes difficult to allocate resources efficiently to meet the user demands in order to satisfy their requests and to take care of the Service Level Agreements (SLA) provided by the service providers [1].

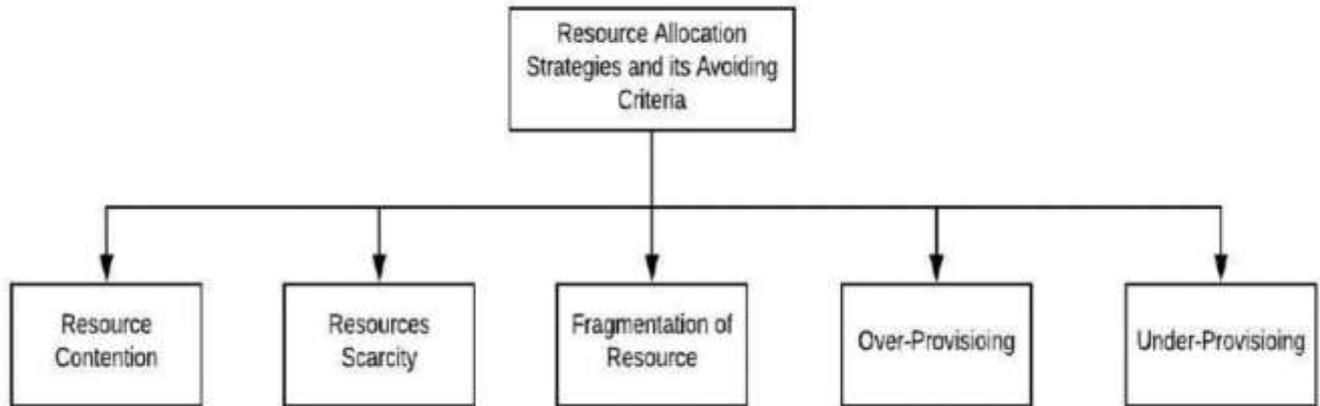


Figure 3: Resource Allocation Strategies and its Avoiding Criteria [6]

To address this issue we have to devise a mechanism to collect the data of users, the trends of their usage, the applications they run, the resources that applications require, the quantity of resources in the resource pool, the rate of utilization of resources, the actual required amount of resources, and the requested amount of resources, the duration of utilization of the resources, the issues that occur during runtime and the issues that occur during downtime, the peak hours of utilization of resources, the time during which resources are utilized very less, the applications that are using resources not required for them to function, the applications that are unable to get resources allocated for them when requested and many more parameters. Figure 3 [6] illustrates the avoiding criteria for resource allocation strategies. For all the avoiding criteria mentioned in the figure, below we have described how we can address each of those using data analytics.

For the data collected we create models and machine learning algorithms to run on this huge dataset that we collect to convert it into meaningful information for us to draw conclusions and make informed decisions based on it. For downtime issues we can fetch information like the number of specific resources that have been requested for and the reason for failing to allocate those resources and consequently make decisions as to how we can prevent such scenarios in the future. For runtime issues we can fetch information pertaining to performance of the application running and analyze if the application has been running at its full performance in an efficient way. If not, we try to understand what resources have been requested for the application and if that has been allocated as required or not.

For issues pertaining to resource contention, we analyze the number of resources all the applications that have been requesting and try to conclude as to why the resource contention occurred. This can be figured out by analyzing which applications have been requesting resources that are more than required and ensure that the applications are provided with resources as much as needed and not beyond what is required to maintain the efficiency and performance for all the users. The issues pertaining to resource scarcity can be addressed by analyzing the total amount of resources that is required for all the applications and processes that are running on a specific platform or infrastructure of cloud and based on that conclusion we can make an informed decision as to how much number of resources are to be maintained in the resource pool at any given point of time.

Over provisioning and under provisioning of resources can be avoided by analyzing the data and understanding the applications and processes that are requesting for resources more than required or the ones that are getting allocated less than requested. From this conclusion we can ensure and adopt a resource allocation strategy to make sure the applications or processes request for only as much sure resources as required so that the excess of resources can be freed up and allocated for the ones that are in need.

VI. CONCLUSION

Cloud computing being an emerging technology faces new challenges as it evolves. Numerous researchers are working to solve these problems that cloud faces and huge investment is being made on the same. With a wide acceptance and adoption of cloud computing paradigm there is a need to solve these constraints as early as possible to minimize losses and maximize performance. The author illustrates the use cases that are currently poorly supported in both infrastructure and platform of the cloud and thereby proposes the approach of data analytics and leveraging the power of this methodology to ensure seamless experience to users by efficiently and effectively devising strategies to allocate resources dynamically based on the usage trends of the users.

References

- [1] Sudeepa R, Dr. H S Guruprasad, 2014, Resource Allocation in Cloud Computing, International Journal of Modern Communication Technologies & Research (IJMCTR), ISSN: 2321-0850, Volume-2, Issue-4, April 2014.
- [2] Fenn, Jackie, Nikos Drakos, Whit Andrews, et al. 2008. Hype Cycle for Emerging Technologies, 2008. In Research, edited by Gartner: Gartner.

- [3] Z. Xiao and Y. Xiao, "Security and Privacy in Cloud Computing ",IEEE Commun.Surveys and Tutorials, vol. 15, no.2, pp.843 - 859, Second quarter 2013.
- [4] <https://mycloudblog7.wordpress.com/2015/05/29/the-evolution-to-cloud-computing-how-did-we-get-here/>
- [5] What Is Cloud Analytics? - https://www.splunk.com/en_us/data-insider/what-is-cloud-analytics.html
- [6] Prasad, Vivek & Nair, Anuja & Tanwar, Sudeep. (2019). Resource Allocation in Cloud Computing, https://www.researchgate.net/publication/334836317_Resource_Allocation_in_Cloud_Computing
- [7] https://technologyadvice.com/wp-content/uploads/2013/05/Data-Analytics-in-Cloud-Computing_TechnologyAdvice.pdf

