SHIFTING TO CONTAINERIZED DEVELOPMENT IN ORGANIZATIONS WITH CLOUD-SUPPORTED INFRASTRUCTURE

Natalya Bromall, Robert Morris University, bromall@rmu.edu
Kevin Slonka, University of Pittsburgh, kevinslonka@gmail.com
Peter Draus, Robert Morris University, draus@rmu.edu
Sushma Mishra, Robert Morris University, mishra@rmu.edu
Amit Pandey, FHLBank Pittsburgh, dramitpandey@outlook.com

ABSTRACT

Containerization is a new approach in developing and hosting applications; it allows running multiple applications on the same virtual machine, instead of virtualizing the entire operating system. It finds an unusually high level of acceptance in the cloud environment, increasing the speed and reducing the cost of running applications and services; however, as shown here, the most common reason for switching to containers is process automation. In this study, the researchers interviewed cloud service professionals in various organizations, attempting to understand what strategies and patterns they follow in utilizing the container software, and what their experiences are in doing so. Many of the participants made distinctions between external and internal usage of containers but a large minority has yet to fully implement containers for a variety of reasons, perceived security (lack of) being a big one. Approximately 40% of the participants reported no changes or problems with using containers and moving forward, most organizations are optimistic that the majority of services will be fully container-native. It was found that the organizations prefer a hybrid approach and are not yet ready to separate from internal physical infrastructure. Independently from the current use of containers, the experts see a great future for containerization and believe it will leverage the use of the cloud services.

Keywords: Containers, containerization, cloud services, Docker, Kubernetes, cloud strategy

INTRODUCTION

The history of the contemporary cloud technologies with the variety of services they offer started in 2006-2008 – the three years during which the top four cloud service providers, AWS (Amazon Web Services), Microsoft, Google, and IBM, launched their platforms (Foote, 2012). Within the subsequent years, cloud services gained tremendous popularity; at this point, there were very few businesses that utilized the IT services and had not yet established any cloud presence. Most consumer services, such as banking, education, and healthcare, have time-critical applications that run in the cloud (Gowri and Shanthi Bala, 2018). A quick analysis of 50 randomly selected companies showed that the top three most demanded cloud service types are virtual machines, virtual storage (especially web storage), and scaling tools, with networking and database services closely following (Bromall, Draus, Mishra, and Stewart, 2019).

Within the last few years, however, the situation started to shift; containerization platforms such as Kubernetes gained more and more popularity and were often named as one of the top cloud services trends in 2020 (Tsidulko, 2020). Containerized deployment is a type of serverless architecture that allows running certain functions or services without launching a virtual machine, which can be quite costly. An application is packaged in a container with all related files and information; the container agent then deploys this content. There are several container services, of which the first and still the most popular is Docker, followed by Kubernetes (Heusser, 2020). The top cloud providers brought these tools to their platforms, creating services such as Elastic Kubernetes Service (EKS) by AWS, Google Kubernetes Engine (GKE), and IBM Cloud Kubernetes Service. Containerization became a vital part of cloud services – to the

point when, as an example, AWS recently included this topic in its core certification exam (AWS Solutions Architect - Associate).

When moving their services to the cloud, organizations follow different strategies that fit their needs. For example, when Capital One adopted AWS cloud tools, it chose Elastic Container Services (ECS) with Docker and elastic load balancers to run a multitude of microservices with less delay than in an on-premise architecture (Stroppa, 2016). At the same time OpenSource, and Israel-based software developer, operates with Azure tools and uses Application Gateway and Azure Kubernetes Service (AKS) to retrieve the applications requested by its customers (Case Study, 2019). Each approach in transitioning to cloud container services is unique; there is no single dominant strategy for virtualization and the use of containers (Gowri and Shanthi Bala, 2019). In this study, the researchers conducted a series of interviews with IT and cloud service specialists to gain insight into what guides the organizational decision-making in the process of migrating to container services. The goal of this study is to understand the changes in current infrastructure due to containerization as well as the perceived future changes caused by such services. The research questions posed in this study are as follows.

RQ1: What are the current practices in the utilization of containers in on-premise IT infrastructure?
RQ2: How does containerization reshape/influence the use of cloud computing in organizations?
RQ3: How does containerization reshape/impact on-premise IT infrastructure services in organizations?
RQ4: How will containerization impact both on-premise IT infrastructure and cloud operations in the future?

A critical literature review of relevant research in this area is presented next, followed by a description of the methodology used in this study for data collection and analysis—the results for each research question provided. Implications and future research directions are drawn, followed by the conclusion section.

BACKGROUND AND RELATED WORK

With more and more research focusing on the containerized application deployment and specifically its migration to the cloud, we can say that the field is moving toward container-based middleware and even container-based Platform as a Service (PaaS) (Pahl, Brogi, Soldani, and Jamshidi, 2019). Hypervisor-based virtualization, or full virtualization, means creating an entire virtual machine with its OS on the host (hypervisor). As opposed to full virtualization, containers allow multiple applications to share the same OS (Rad, Bhatti, and Ahmadi, 2017) and to avoid the complexity of creating a complete virtual machine for each task. Containers are considered agile alternatives to virtual machines (Zhou, Li, and Wu, 2019). Shown in Figure 1 is hypervisor-based virtualization vs. OS-based (or container-based) virtualization; however, there are several other less common types of container virtualization types (Gowri and Shanthi Bala, 2019).
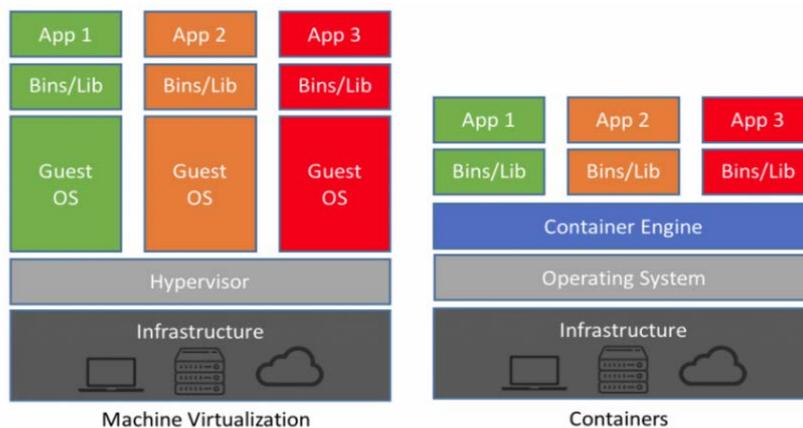

Figure 1. Machine virtualization vs. containers

Docker in Research

Docker is a common platform for building containers; Kubernetes is the system for orchestration of Docker containers and, in a way, is Docker's extension. These two systems are prevailing in the market of container tools and have been adopted by all top cloud service providers. Although there are many other container management tools, Docker remains the most popular in the market (Rad, Bhatti, and Ahmadi, 2017). There are several studies trying to understand the phenomenon of Docker and its popularity. According to some studies, it is not substantially better than other similar services; for example, Kozhirbayev and Sinnot (2017) compared Docker to Flockport and found that the former has no significant advantages compared to the latter in CPU performance, memory, networking capabilities, and some other parameters. In the same study, the authors admit that these conclusions are made for the case when Docker does not utilize any extra features.

Containerization in Research

The applications of container tools are well studied and presented in the literature. Pahl, Brogi, Soldani, and Jamshidi (2019) gave a comprehensive review of the studies in that area. They outlined three types of research: technology reviews, solutions, and use-case architectures.

A large group of studies provide technology reviews of cloud-based container tools and give evidence of improved performance, specifically with the agile approach and lightweight tasks. As an example, Yadav, Sousa, and Callou (2018) examined the features of Docker and ran a series of experiments showing that Docker containers required lower execution time against VMs. Furthermore, Rad, Bhatti, and Ahmadi (2017) concluded that in addition to performance-improvement, Docker containers significantly reduce having to rebuild the cloud development platform. At the same time, the latter study includes several disadvantages of Docker, such as incompatibility with older machines. A similar issue was brought up by the subjects in this study who experienced difficulties shifting between the Docker versions.

The studies from the second group contained specific solutions for improving container performance. As a newly emerged technology, containerization is often not used to its full potential. Correct placement of containers and testing container performance before deployment are the two problems frequently mentioned in the research. Wan et al. (2018) developed an algorithm to optimize container placement and task assignments. The testing problem can be solved by creating simulation environment models that help to model the deployment in the cloud (Piraghaj, Dastjerdi, Calheiros, and Buyya, 2016). Zhou, Li, and Wu (2019) expanded the problem and presented a possible efficient solution for organizing cloud container clusters. In addition to container organization, testing, and placement, network organization is an important task. Cziva and Pezaros (2017) gave an example of a network functions virtualization (NFV) model that improves running and orchestrating lightweight virtual network functions (VNFs).

The third group of studies includes case studies describing user experiences with specific cloud container solutions. Each study in this category reports a specific experience using such solutions without generalization to other (even similar) cases. Case studies show that containers became crucial mainly in two areas: IoT and big data analysis. Both fields became, during the past several years, extremely valuable to society. As an example, container-based infrastructures can facilitate the connection to a vast number of IoT devices (Khaur, Dhand, Kumar, and Zedally, 2017). The IoT-related research had a limitation, not being generalizable to infinitely large numbers of devices. Morabito (2017) noted this limitation of the earlier studies and discussed the possibility of deploying virtual instances for an unlimited number of IoT devices. In many case studies, the authors describe the cases of using containers not directly in IoT, but in its supporting infrastructure, for example, container-based fog computing (Yin, Luo, and Luo, 2018). Naik (2017) showed an example of a case where containers can host applications for processing the big data, in particular, generated by the IoT devices.

Pahl, Brogi, Soldani, and Jamshidi (2019), in their review of container-related research, discovered that there were more case studies than technology solutions. In other words, in many studies, the authors presented a single-use case. They evaluated the experience of containerization in that particular case, while fewer studies included a summative evaluation of container tools based on a group of experiences. Few attempts have been made to look at answers to the same questions across multiple case studies in various organizations.

METHODOLOGY

After a review of the research questions, the researchers looked for a data collection technique with high validity that would allow for data collection for today's complex and fluid environment of container usage; they settled on focused interviews. The following five question groups were used in the interview.

1. Group 1: Present IT strategy
   a. What container vendors are used at your organization?
   b. In what ways are containers used in your current internal IT infrastructure?
   c. What changes were made to your infrastructure and practices to implement containerization?
2. Group 2: Obstacles
   a. In what ways are containers hindering your organization's ability to work efficiently?
   b. What features are missing from current container implementations that would benefit your organization?
3. Group 3: Future IT strategy
   a. What are your views on the impact of containerization on IT services in the future?
4. Group 4: Present cloud strategy
   a. Describe your organization's current cloud strategy.
   b. In what ways are containers a part of your current cloud strategy?
   c. How does container usage influence your cloud strategy?
5. Group 5: Future cloud strategy
   a. What are your views on the impact of containerization on cloud strategies in the future?
6. Group 6: demographic questions (years of experience in IT and with the cloud platforms; industry type, current role, team and company size)

Ten interviews were conducted with a variety of seasoned IT professionals. Table 1, which lists the participant IDs and their current job titles, shows that the participants span the full range of duties within the IT field, from sysadmin to security to developer, and the full range of responsibilities, from entry-level to director-level. The participants were also plucked from a variety of business sectors, as shown in Figure 2, which groups the participants per their organization's North American Industry Classification System (NAICS) classification (Executive Office, 2017).

Table 1. Participants by Occupation

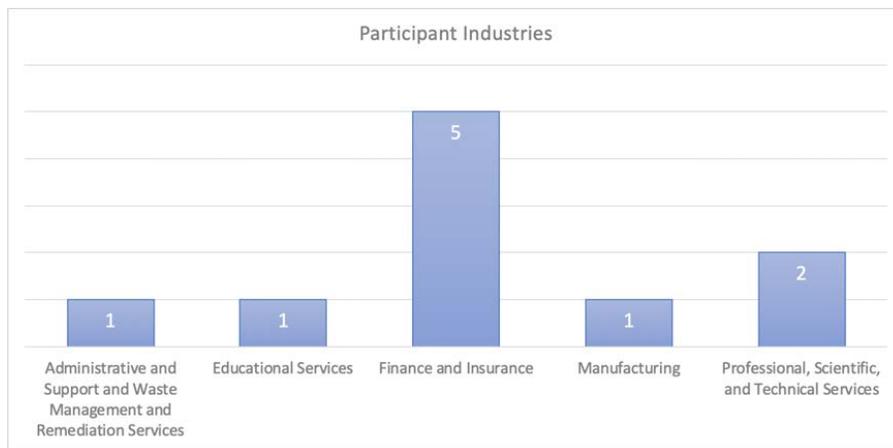| ID	Current Job Title
1	Systems Administrator
2	Security Engineer
3	Security Specialist
4	Senior Analyst
5	Cloud Architect
6	Solution Architect
7	Principal Developer
8	Director of Application Development
9	Technology Evangelist
10	Data Science Architect

Figure 2. Participant Industries

Additionally, as shown in Figure 3, the participants were distributed across not only the number of years in the IT industry but also the number of years working directly with cloud platforms. This allowed for keen insights and comparisons between those who have indoctrinated into the "pre-cloud" IT era and those who have recently entered the workforce and were introduced to cloud platforms from the beginning.
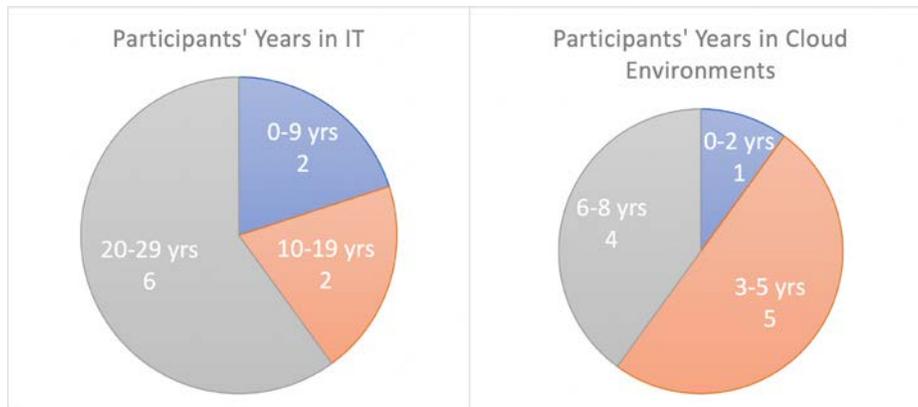

Figure 3. Experience in IT vs. Experience in Cloud Environments

As an additional step, the researchers ensured that the participants worked for organizations of varying sizes. It is a common perception in the IT industry that smaller organizations are more agile and quicker to adopt paradigm-shifting changes than large conglomerates, which usually have so many layers of red tape that it seems as if nothing ever gets accomplished. The participants for this study were distributed across all three categories of organization size (Gartner, 2020), as demonstrated in Figure 4. Although not noted in the graph, three of the Large organizations have 20,000+ employees.
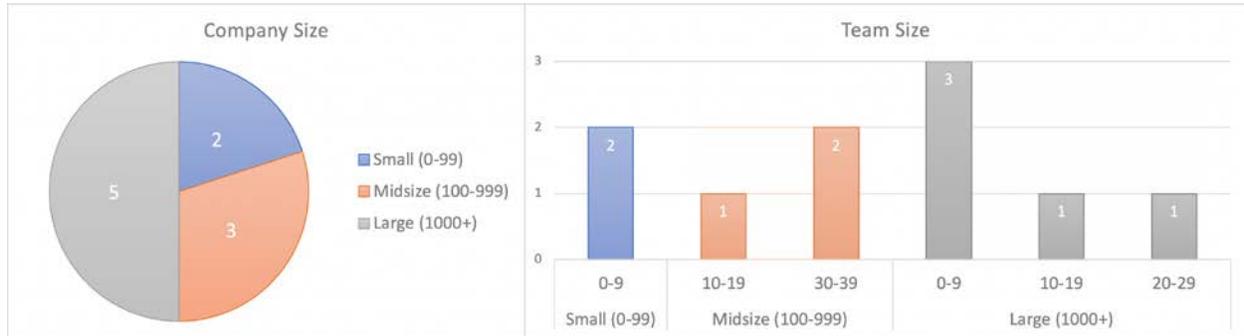

Figure 4. Organization Sizes

What is interesting is the distribution of team sizes within organizational sizes. One might expect that large organizations would have the largest teams dedicated to the cloud or development in general, but that is not the case. Only the midsize organizations have team sizes greater than 30, while the large organizations have the highest number of small teams. This purposeful selection ensured that all participants could properly inform the research questions and help the researchers better understand the phenomenon being studied: container usage.

The ten interview questions focused on three areas: Current container usage, problems they encounter with containers, and their views on the future of containers in their industry. The answers to the ten interview questions were written down and transcribed, and then themes were coded from their answers for each item. Finally, the themes were collapsed and reduced into the four research questions.

RESULTS AND DISCUSSIONS

RQ1: What are the current practices in the utilization of containers in on-premise IT infrastructure?

The themes that come out of the interviews for this section are shown in Table 2. The actual results indicated Docker and Kubernetes were the primary commercial products used (80% of respondents), but there were five others that were mentioned as being used by the participants. Some respondents admitted using open source platforms (20%) or a commercial PKS platform (20%).

Table 2. Interview Themes – RQ1

Commercial Products used (80%)
Wide range of usage from testing to everything
Wide range of integration from none through everything

There was a wide range of users across all of the participants, even those in the banking sector where there were four participants. An interesting aspect that was noted by the researchers was that many of the participants made distinctions between external and internal usage of containers. Another interesting point that arose was the amount of change in the IT operations that occurred when implementing containers. The participants reported no changes all the way through significant changes that occurred or needed to be made to implement container technology.

Process automation efforts seem to lead the efforts in the utilization of containers in on-premise infrastructure. These efforts require a considerable amount of knowledge creation opportunities in the existing IT practices, as acknowledged by our participants. Training for employees in this field is required. Such changes at the infrastructure level require prioritization of goals and resources, and this helps organizations in aligning with overall digital strategies of the future.

RQ2: How does containerization reshape/influence the use of cloud computing in organizations?

The themes that come out of the interviews for this section are shown in Table 3.

Table 3. Interview Themes – RQ 2

More cloud usage as more container usage (50%)
Lift and Shift to the cloud (70%)

The slight majority are using containers for their organizations' primary services, whether internal or external. A large minority has yet to fully implement containers for a variety of reasons, perceived security (lack of) being a big one. The use of containers, however, seems to drive usage of the cloud versus internal infrastructure in those organizations that have not yet fully embraced the cloud. For others, the type of work done by the organization is contract-based, so the usage of containers is dictated by each contract. In summary, our data suggest that most organizations are using a hybrid approach when it comes to containers and the cloud, where some services remain on internal infrastructure, and others are pushed to the cloud.

While not a major theme, a few participants noted the connection between containers and their switch to an agile development framework.

> "Developers: everything is pushed to Docker containers because they are all working on different platforms and different environments." – Participant 5

> "No container at this time. In the development phase, we are still learning, setting up the process; our main goal is to meet the agility and automation principles we established in the Bank." – Participant 6

> "Platform services that run department, test, assessment, and monitoring" – Participant 3

> "Using different containers for different purposes: there is a data phasing container, many platform containers for different services." – Participant 10

Our data suggest that majority of organizations do not have an enterprise strategy for containers but are using it project-wise. The usage is primarily in a non-production environment. Applications are cloud-based and are headed to a single-command-push to a production setup. There are many examples of a cautious hybrid approach of using containerization. The containerization strategy is focused on application enhancement of on-premise IT infrastructure. An essential requirement for the complete integration of on-premise and cloud strategy is understanding and meeting regulatory compliance needs.

RQ3: How does containerization reshape/impact on-premise IT infrastructure services in organizations?

The themes that come out of the interviews for this section are shown in Table 3 above. The actual results are shown in Table 4 below, but it is interesting to note that approximately 40% of the participants reported no changes or problems with using containers.

> "Everything previously was done on OpenStack VMs because they didn't want to wait for the IT helpdesk to provision VMs in company infrastructure. " – Participant 1

> "No real Changes… only changes based on cost optimization (instances type and size, etc.)" – Participant 3

> "Can't make one-time/adhoc changes (this needs to be done frequently)." – Participant 1

> "Scaling in cross regions (load balancing between regions)" – Participant 3

Table 4. Interview Themes – RQ 3

Lack of a single OS means operations traditionally done at the OS level; security, backup, logs, are challenging (20%)
Flexibility and move to agile methodology (100% mentioned one or both)

For some, the downsides of containers keep them using the internal, physical infrastructure. Such items as the inability to make ad-hoc changes to containers, accessing local network resources, or the perceived lack of security to container/cloud architecture are too risky at this time to fully adopt containers in the cloud. This leads to a hybrid approach to containers, whereby many organizations are merely testing the waters with containers in the cloud, slowly moving services outward while keeping some services on internal infrastructure. Organizations will need time before fully separating from physical internal infrastructure.

Containerization is being used for the application development process. Application delivery is a critical aspect of moving application software from development to various test environments to staging and, finally, the production environment. The promotion of application software, along with associated configuration, requires a significant length of planning and collaboration with various teams in organizations. Containerized application image, along with related configurations, streamlines the application delivery process. The containerized application images are immutable and stateless. The stateless nature of containers decouples data from the container images themselves. This immutable nature of container images guarantees that containers are not modified or changed during their lives. This implies that there have been no updates, no patches, and no configuration changes on the containers. These characteristics of containers help deploy the same container image in various environments, including production and make them as identical as possible.

RQ4: How will containerization impact both on-premise IT infrastructure and cloud operations in the future?

All of the participants were enthusiastic about the use of containers expanding in the future, even if they could not be more specific of that "more" usage. Of all respondents 100% envision less on-premise and more cloud operations, and an increase in flexibility of their IT architecture; 20% mentioned inter-vendor operability as another future advantage of containerization.

Table 5. Interview Themes – RQ 4

Less on-premise and more cloud operations.
Flexibility
Inter-vendor operability

The themes related to these questions that emerged from the interviews are shown in Table 5. Moving forward, most organizations are optimistic that the majority of services will be fully container-native, with some even going as far

as expecting fully automated workflows that will test, secure, and publish containers with a "push of a button." It is the consensus that container usage will have a significant impact in the future, affecting the way most companies conduct business. Some are hesitant, however, due to the costs associated with some online container platforms.

> "I think we will definitely have everything with containers. For us, it is so easy to grab an image and have it immediately ready to go" – Participant 10.

> "All the companies will slowly move from old way of development to a new way of development." - Participant 7.

Containers are portable. The immutable and stateless nature of containers provides flexibility to deploy containers in cloud-native infrastructure and helps reduce the organization's risks from vendor locking. Containers will have a tremendous impact on organizations' ability to decide where to run their containerized applications on cloud or on-premise IT infrastructure. There would be a drastic change in on-premise IT infrastructure. The addition of vendors' container solutions such as VMWare Enterprise PKS or RedHat Openshift would enable the management of containerized applications across on-premises and public-cloud environments. This will help organizations to harness the benefits of cloud computing when it decides to deploy its containerized application to the cloud.

CONCLUSION

This study explores the current usage and future impacts of containerization on cloud and on-premise IT infrastructure. It provides insights into changing trends of usage of containerization and adds to the body of knowledge in cloud research. Ten IT professionals were interviewed about the state of containerization. The results suggested that containerization provides flexibility, portability, and seamless integration in infrastructure. The hybrid approach was preferred by organizations due to several issues with the complete adoption of cloud services. Practitioners can benefit from this study by understanding how to use containerization for their organization's benefit. Future research is required to understand how barriers in cloud adoption can be addressed.

REFERENCES

ABN Asia (2019). Case study: WhiteSource simplified deployment using Azure Kubernetes Service. Retrieved on March 20, 2020 from http://abnasia.org/case-study-whitesource-simplifies-deployments-using-azure-kubernetes-service/.

Bromall, N., Draus, P., Mishra, S., and Stewart, J. (2019). A Case Study-Based Analysis of Innovative Cloud Technologies Implementation: Challenges, Tools, and Benefits of Amazon AWS. Northeast Decision Science 2019 Conference, Philadelphia, PA.

Executive Office of the President: Office of Management and Budget. (2017). *North American Industry Classification System: United States, 2017.* https://www.census.gov/eos/www/naics/2017NAICS/2017_NAICS_Manual.pdf.

Foote, K.D. (June 22, 2017). A brief history of cloud computing. Dataversity. Retrieved on 11/20/2018 from http://www.dataversity.net/brief-history-cloud-computing/

Gartner. (2019). Small and Midsize Business (SMB). In *Gartner Glossary.* Retrieved May, 10, 2020, from https://www.gartner.com/en/information-technology/glossary/smbs-small-and-midsize-businesses.

Gowri, A.S., Shanthi Bala, P. (2019). Impact of virtualization technologies in the development and management of cloud applications. International Journal of Intelligent Systems and Application in Engineering, 7(2), 104-110.

Heusser, M. (2020). 30 essential container technology tools and resources. TechBeacon, retrieved on 4/1/2020 from https://techbeacon.com/enterprise-it/30-essential-container-technology-tools-resources-0.

Kaur, K., Dhand, T., Kumar, N and Zedally, S. (June 2017). Container-as-a-Service at the edge: trade-off between energy efficiency and service availability at fog nano data centers. IEEE Wireless Communications, 48-56.

Morabito, R. (2017). Virtualization of Internet of Things edge devices with container technologies: a performance evaluation. IEEE Access, 5, 8835-8850.

Naik, N. (2017). Docker container-based big data processing system in multiple clouds for everyone. IEEE International Systems Engineering Symposium proceedings.

Pahl, C., Brogi, A., Soldani, J. and Jamshidi, P. (2019). Cloud container technologies: a state-of-the-art review. IEEE Transactions on Cloud Computing, 7(3), 677-792.

Piraghaj, S.F., Dastjerdi, A.V., Calheiros, R.N. and Buyya, R. (2017). ContainerCloudSim: an environment for modeling and simulation of containers in cloud data centers. Software: Practice and Experience, 47, 505-521.

Rad, B.B., Bhatti, H.J. and Ahmadi, M. (2017). An introduction to Docker and analysis of its performance. International Journal of Computer Science and Network Security, 17(3), 228-235.

Stroppa, D. (2016). Run containerized microservices with Amazon ECS and application load balancer. AWS Compute Blog. Retrieved on 3/30/2020 from https://aws.amazon.com/blogs/compute/microservice-delivery-with-amazon-ecs-and-application-load-balancers/

Tsidulko, J. (November 14, 2019). 10 emerging cloud computing trends to watch in 2020. CRN, retrieved on 4/1/2020 from https://www.crn.com/news/cloud/10-emerging-cloud-computing-trends-to-watch-in-2020

Wan, X., Guan, X., Wang, T., Bai, G. and Choi, B.-Y., (2018). Application deployment using Microservice and Docker containers: framework and optimization. Journal of Network and Computer Applications, 119, 97-109.

Yadav, R.R., Sousa, E.T.G. and Callou, G.R.A. (2018). Performance comparison between virtual machines and Docker containers. IEEE Latin America Transactions, 16(8), 2282-2288.

Yin, L., Luo, J. and Luo, H. (2018). Tasks scheduling and resource allocation in fog computing based on containers for smart manufacturing. IEEE Transactions on Industrial Informatics, 14(10), 4712-4721.

Zhou, R., Li, Z. and Wu, C. (2019). An efficient online placement scheme for cloud container clusters. IEEE Journal on Selected Areas in Communication, 37(5), 1046-1058.