High-Performance Computing Applications' Transition to the Cloud in the Oil & Gas Industry

Suha N. Kayum
Saudi Aramco
Dhahran, Saudi Arabia
suha.kayum.1@aramco.com

Marcin Rogowski
Saudi Aramco
Dhahran, Saudi Arabia
marcin.rogowski@aramco.com

***Abstract*—In the cloud platform, High-Performance Computing (HPC) is meant to provide the capability of scaling to large numbers of tasks that run in parallel on-demand. However, it remains a dilemma for companies in the Oil & Gas (O&G) industry whether to transition HPC activities to the cloud or not. In this paper, the latest research studies are shared that shed light on some of the challenges and outlooks prevailing. The choice of which HPC applications should migrate to the cloud is shown to be application dependent and will be demonstrated in this paper with a case study assessing migrating reservoir simulation activity to the cloud. It is evident that a hybrid cloud solution for high-performance computing is a good starting point, as it mitigates a few challenges that entities in the O&G industry face today.**

Keywords—reservoir simulation, high performance computing, cloud computing

I. Introduction

Due to the world's eminent technological breakthrough, known as the Fourth Industrial Revolution, digital transformation and technology have emerged to be the driving core of businesses and organizations. Companies are continuously striving to incorporate the latest efficient technologies with the aim of meeting customer needs while reducing the overall operating expense. One of the most popular technologies, aiding the transformation towards automation and efficiency, is cloud computing. Most industries are highly benefitting from cloud technology due to the level of integration that the internet has brought to operations and transactions.

The use of HPC applications is increasing in many fields such as scientific research, academia, business, and data analysis [1]. Due to the gap between scientists' growing computational demands and limited local computing capabilities, cloud computing has become an alternative platform for running HPC applications [2]. Cloud computing service models can be categorized into three models: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS).

In addition, the cloud computing infrastructure deployment models can be categorized into three main models namely the Public, Private and Hybrid cloud It is important to distinguish between the different kinds of cloud deployment models. In the public cloud setup, the physical infrastructure is owned by the Cloud Service Provider (CSP). The cloud is open to the public and resources are shared among anyone who has access to the cloud. In the private cloud setup, the physical infrastructure may be owned by the CSP or the organization. However, the use of the cloud is by a single organization. The Hybrid Cloud environment can be a mix of two or more clouds setups. All of the participating clouds retain the status of a unique entity, but share standardized or proprietary technology [3].

Recently, because of the high computing needs that the scientific community is seeking, HPC as a service in the cloud has gained a great deal of attention and has begun to replace traditional cluster infrastructure. Traditional HPC clusters were mainly on-premise which made for an inelastic and inflexible model in terms of the number of resources and availability. However, due to the resources being inside an organizations firewall, i.e. on-premise, the highest data security is guaranteed. High Performance Computing Cloud (HPCC) generally would offer a more flexible and elastic environment, but however, because the cloud is public, i.e., outside an organizations firewall, accessible through the internet and maintained and supported by a third party, data security becomes a major concern.

Nonetheless, numerous entities in several industries have taken their computing needs or at least some of their demands to the cloud. For some scenarios, the transition of certain HPC application to the cloud can be so advantageous that the security risk could be overlooked. These benefits to HPCC have been introduced in the last few years such as elasticity of resources and elimination of cluster setup cost and time [1]. In addition, capacity and structure of HPC on cloud computing can be adjusted dynamically according to the requirements of customers [2].

The biggest advantages of HPCC is mitigating the cost, complexities, and resources that go into the establishment and configuration of in-house HPC infrastructure. Operations under cloud computing have ensured that organizations can store, access, and analyze massive data while reducing in-house server storage and compute resources. Another advantage is the flexibility that HPC on the cloud offers. Because of the high number of clients that the cloud companies tend to, a flexible model is created where near-infinite scalability is offered. In addition, customers are alleviated from the responsibility of managing hardware [4]. Furthermore, HPC applications need not only storage space and access to the data but the high-end infrastructure that may be heterogeneous; including graphics processing units and other hardware requirements. Such hardware may have high demand of electricity which is becoming an issue and increases the capital expenses of expanding on-premise HPC resources. Using HPCC services ensures the business focus on operational needs, rather than

978-1-7281-5020-8/19/$31.00 ©2019 IEEE

technical requirements that may involve budgeting and set up of data centers that are expensive in hardware and require the skill for proper implementation.

Not all users have the same criteria or restrictions for moving their HPC applications and data to the cloud. In the O&G, the requirements for HPC applications differ than other industries and could even differ within the O&G industry. To make that analysis, a better understanding of the transition of HPC applications [5] to the cloud has to be made. Initially, public cloud providers such as Amazon, Microsoft, and Google targeted clients related mainly to businesses and not the scientific community [6]. Nonetheless, that has changed in the last few years. The advantages and disadvantages that cloud computing has to offer need to be clear specifically in HPC application environment, which typically requires extreme levels of computational capacity.

This paper seeks to focus on the O&G industry and the transition of their HPC applications to cloud platforms effectively. The paper addresses the concern of moving HPC applications to the cloud beginning with the history of research studies concerning HPCC, which shed light on improvements, challenges, and outlooks prevailing in such works. The research focuses on improvements in cloud computing techniques involving network speed and latency, virtualization, service providing, scheduling, data management and security for HPCC. Some of the compelling matters are shared that may push entities to settle on cloud computing, versus in-house HPC systems. Finally, assessment methods for HPC applications in the cloud are discussed including an assessment tool that is used to aid in assessing HPC applications' feasibility and advantage of use in the cloud.

II. The History and Advancement of HPCC

According to [7], the trend in historic spending on HPCC has increased in the past few years as shown in Fig. 1.

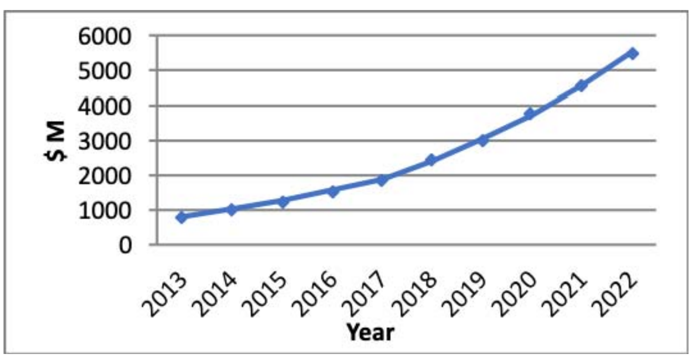


Fig. 1. Historic and Projected Spending on Public Cloud for HPC Applications [7]

According to their statistics, the number of HPC sites that are running some of their activities in the cloud has risen from 13% in 2011 to 70% in 2018. However, according to them, private and hybrid cloud use is growing even faster. The company used the historical data, as well as interviews and knowledge of the HPC ecosystem and the direction that the market is heading to project the future spending of HPCC. Based on their prediction, HPCC is expected to grow from a market size of a little over $3 billion in 2019 to over $5.5 billion in 2022. It is important to understand the history and advances cloud computing has made recently in the realm of HPC to paint a clear picture. There has been research and improvement in employing HPC on cloud environments and enhancing the performance of HPCC. In this section, a walkthrough of some of the latest advances in cloud computing and HPCC will be made. Five main aspects will be discussed which involve i) network speed improvement and latency reduction ii) virtualization, iii) service providing and resource management, iv) scheduling, and v) data management and security in the cloud as depicted in Fig. 2. For each component, the conducted literature review will be discussed.

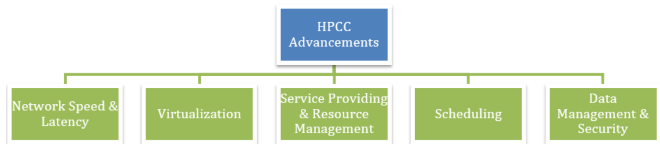


Fig. 2. Categories of Advancements of HPCC

A. *HPC Network Speed Improvement and Latency Reduction*

HPC applications require low latency and high bandwidth for inter-process communication. High network latency results in increased execution time for HPC applications, which are heavy on data communication such as Message Passing Interface (MPI). In cloud computing infrastructure, Ethernet is mainly used to communicate between virtual machines, which can cause a performance bottleneck of HPC applications [1]. According to TOP500 as of June 2019, 54.4% of supercomputers rely on Ethernet, whereas Infiniband, Omnipath and custom interconnect make up, 24.6%, 9.8% and 10.4% respectively [8].

A debate remains whether Infiniband versus Ethernet plays a huge role in the network speed and latency reduction in the cloud. For example, [9] and [10] proposed high-speed cluster interconnects InfiniBand to improve communication performance between cluster nodes, while other authors have argued that the difference is minimal and that optimization focus should be made elsewhere. Reference [11] argues that a suitable I/O virtualization support could serve as a bigger bottleneck than the network latency. This conclusion was made based on the results on increasing the high-speed network of Amazon EC2 cluster compute instances to 10 Gigabit Ethernet and still being penalized by the unsuitable I/O virtualization support.

Reference [12] introduced a communication topology-aware framework by pre-execution and multi-scale graph clustering of the cloud. The authors use MPI benchmark called NPB (NAS Parallel Benchmarks) [13] and their experimental results show that their method outperforms the existing methods that do not incorporate topology.

In 2014, [14] implemented an MPI version of parallel Radix sort. This effort was made to improve the performance and scalability of HPC application on Amazon's Cloud. A comparison of the performance was made between the cloud and a dedicated HPC cluster and revealed a significant improvement in the speedup of a reported 20% on the cloud.

978-1-7281-5020-8/19/$31.00 ©2019 IEEE

Recently, [15] attempted to restore the shared memory system and created cloud integration based on the latest versions of libvirt, OpenStack, Linux, and MPICH. Moreover, different factors that influence both TCP/IP and ivshmem communication were analyzed together with tuning techniques that enhanced the performance significantly.

B. HPC Virtualization

Of the many key components of facilitating HPC on the cloud, virtualization has been one of the most important. Virtualization allows HPC to leverage resources efficiently and allows further scalability in the cloud. Virtualization serves as a convenient way to package and deploy scientific applications across heterogeneous system all while surpassing a significant amount of overhead.

In contrast to traditional data centers, cloud computing model allows greater energy efficiency to be achieved by using techniques such as virtualization and consolidation of the workload, providing better control of computing resources in relation to their demand, and avoiding their waste [16]. In 2016, a Virtual Machine (VM) scheduler is presented by [17], for homogeneous HPC cloud platforms. The scheduler considers each VM workload type to decide the allocation. Moreover, the scheduler could prevent excessive consumption of energy, and reduce Service Level Agreement (SLA) violations, maintaining the performance by the simultaneous allocation of VMs that execute different tasks.

Recently, [18] proposed a cloud infrastructure, called nOSV, to serve the HPC applications running on the cloud. For HPC applications, nOSV developed an active isolated high-performance guest VM with fully committed resources. This research was done because limitations in virtualization-based cloud existed which affected the performance of HPC application on the cloud. The performance overhead was thought to have been caused by VM Monitor interceptions, virtualized I/O devices or cross-VM interference. In order to mitigate this overhead, nOSV was developed and prototyped. It uses a NUMA-aware fixed partition policy to allocate resources. Results show that nOSV could provide better performance for HPC applications when compared to other virtualization environment instances such as Xen or Docker.

C. HPC Service Providing and Resource Management

Due to the recent advancements in the cloud, it has become possible to allow VMs hosting HPC applications to link distributed cloud resources effortlessly. However, the management of resources plays a vital role in cloud computing. Hence, previous work employed for service providing and management of resources, is discussed in this section.

Reference [19] had improved the functionality of one of their software workflows by integrating it to use the remote cloud and HPC resources for data processing. They demonstrate how they have created a hybrid environment where users can access the software through a hybrid environment of either running it on a PC or server, a local cluster, a remote HPC cluster, and/or the Amazon cloud. It is demonstrated that this effort allows users to access the data quickly with less cost and lower staff requirements than other local installations. Later, [20] analyzed the efficiency of the price that would allow the customer to choose the provider's package. Moreover, the authors had investigated the relation that exists between prices, times and problem sizes/workloads from High-Performance Linpack on Amazon EC2.

One of the key challenges in HPC on the cloud is that the application availability, which depends on the software licensing, is less flexible than the hardware component in the HPCC. This is a result of the imbalance in supply versus demand of hardware, software licenses and complexity of access to the software. To mitigate these limitations and provide an on-demand service for end users, [21] proposed a SaaS technique to manage commercial applications of HPC as a Web-based service built on federated clouds.

Another attempt to ease the complexity of deployment to collaborating services on heterogeneous cloud environments was made by [22]. The motivation was to eliminate the manual configuration that required expert users and to create a compatibility environment for combining Cloud and HPC machines within the same design space. The authors attempted to exploit the parallels that exist to solve issues due to interoperability in the cloud. Here, the mOSAIC Ontology, pillar of the IEEE 2302, Standard for Intercloud Interoperability and Federation, was extended to develop CloudLightning Ontology, wherein the heterogeneous resources and HPC environments were incorporated in the Cloud.

D. HPC Scheduling

HPC on the cloud provides scalable and on-demand resources. Scheduling of jobs or other applications in the cloud platform is a complex task, as the resources and requirements of on-demand user applications are dynamic. Hence, the advancement of scheduling in relation to HPC on the cloud is discussed.

Reference [5] presents a cloud platform Aneka, with flexible scheduling policies. The scheduler allows HPC applications to exploit the on-demand elastic capabilities provided by clouds. However, the author does mention that Aneka does allow third parties to seamlessly plug other scheduling algorithms into the model as well.

A technique, named CloudKon was introduced by [23] for the task execution such that the technique is compact, distributed, scalable, and light-weight. This distributed job management system is designed using cloud computing building blocks. CloudKon could support both HPC and Many-Task Computing workloads with millions of jobs. One of the major components of the system is Amazon Simple Queuing Service that functions as a content delivery service for the jobs/tasks so that the clients could communicate with the workers in a scalable manner. The technique is light-weight and offers high throughput and good load balancing. Later, [24] proposed a job scheduling scheme that exploits the resource heterogeneity in clouds to simultaneously provide high performance and fairness.

Reference [25] presented a cloud-based job execution prototype that uses static scheduling techniques to allow the user to choose from different scheduling options. The options are presented to the user based on the job execution time and price. This prototype was evaluated with Amazon EC2 cloud and was shown to result in good utilization of resources and provide the users with the optimal pricing model by quoting a tentative price for the resources.

Reference [26] then introduced a scheduler that is HPC-aware and takes into account the topology when running HPC applications. By utilizing benchmarking results, the network requirements of the HPC application can be identified and knowledge of the effect of the resource distribution can be known. This information can be used to optimize the scheduling further. Experiments using NPB with the topology-aware scheduler versus the unaware scheduler have shown performance improvement of up to 45%.

Reference [27] had presented CLOUD Resource Broker (CLOUDRB), to manage cloud resources and complete jobs efficiently for scientific applications within the deadline that is specified by the user. CLOUDRB was integrated with scheduling and resource allocation mechanisms, namely deadline-based job scheduling and Particle Swarm Optimization (PSO)-based resource allocation. The prioritization of jobs was done using PSO algorithm based on their deadlines and selection of resources such that makespan and cost were minimized within the deadline specified by the user. The target applications were either embarrassingly parallel or parameter sweep type.

E. HPC Data Management and Security

Scheduling of jobs in the cloud platform has been researched and optimized methods have been introduced. Data Management in the Cloud is another area that has been subject to research in the past few years. The large size of the software applications and data sets has been cited as a significant barrier for cloud computing [28] which is exacerbated by in the upstream O&G due to extreme data confidentiality. Data storage and movement is common in HPC environments. In the cloud, data storage, data movement and data visualization can be a bottleneck.

Data security is extremely important to the O&G industry specifically in upstream [29]. Reference [28] discusses several concerns that have arisen as the upstream O&G sector considers the cloud. Of the main concerns are data security and its storage outside the organizations' firewall. Reference [30] that there exist five essential traits distinguishing cloud computing related to the nature of services offered to the customer and the effects on the provider. The traits that are highlighted give the best service to the customer and reduce costs. Nonetheless, [30] mentions the security concerns and highlights it as one of the areas that need key analysis and assessment.

To address the solution of security in the O&G industry, [4] analyzes the move to the cloud in other sectors such as healthcare, retail and scientific research. One of the examples mentioned is TC3, a US-based healthcare services company. The company encrypts its sensitive data before uploading it to the cloud. This could be a solution for O&G but it would add some overhead to operations in O&G where data is of very large sizes. Reference [31] had investigated the usage of Hadoop to solve the problem that occurs due to moving large-sized data. Yahoo! has also used Hadoop successfully to manage moving up to 25 petabytes of enterprise data [32]. In addition, according to [4], the Medical College of Wisconsin developed a customized technical solution that allowed a fast way of moving large amounts of data to the public cloud. The solutions implemented pertaining to handling large data sizes could serve as solutions that could be adopted by the O&G industry.

Data moving during processing is required by most HPC application in the O&G industry. Users that run simulations need to visualize their data during a simulation run to analyze it. According to [33], data movement and visualization should be integrated and if so could allow users to analyze results via remote visualization in real time. This integration could enhance the user experience and reduce storage and data transfer costs. However, the author does conclude that this area is subject to further research and improvement.

III. Applicability of HPCC in the O&G Industry

In the previous section, the roadmap of the advancements in HPCC has been shared. The historical advancements give an explanation on the trend that is seen in the increase in historic spending and the projected spending shown in Fig. 1. The O&G industry is continuously advancing their technologies and optimizing their workflows. One of the latest dilemmas is whether moving high-performance computing applications to the cloud can result in significant benefits to companies in the O&G industry and specifically in the upstream sector. However, as a result of the shift in the O&G industry to digitization and automation, i.e. the transition to the Fourth Industrial Revolution, simulation, data processing and Artificial Intelligence requirements have increased. The reliance on HPC to handle these massive workloads and data has increased significantly.

In this section, a discussion is made on the applicability of HPCC in upstream O&G HPC applications. In terms of the security concern mentioned in previous sections, the private cloud could pose as a viable solution. However, although it resolves the security issue, it does have a drawback such as an increase in cost associated with the private cloud compared to the public cloud. In addition, since the private cloud setup has a fixed amount of resources, it does limit the degree of scalability [34]. Therefore, the hybrid setup has emerged as a suitable solution for O&G. The major advantage of hybrid cloud is that it offers customers the flexibility of opting to pay for extra usage on a need basis that cannot be met by private cloud [34].

A. Assessing HPC Applicability in the Cloud

Reference [35] describes the use of HPCC in the oil and gas industry. The authors share a methodology for companies to use to make decisions of moving their HPC applications to the cloud. This framework was deduced from a project which entailed moving ECLIPSE, a reservoir simulator to Amazon HPC cloud. The framework is shown in Fig. 3 and is comprised of six steps.

978-1-7281-5020-8/19/$31.00 ©2019 IEEE

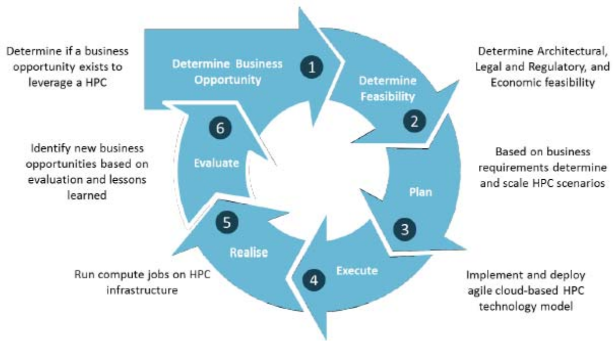


Fig. 3. Framework for assessing and implementing HPC opportunities on the cloud [35]

The framework is explained thoroughly in [35] and is composed of steps in which companies can pursue business opportunities that leverage the latest trends in the cloud in an agile manner. The method begins with identifying a business opportunity in which its move to the cloud could be advantageous, then determining whether the migration would be feasible in terms of both software and hardware requirements. After which, the security requirements are examined and regional restrictions and other such requirements are analyzed. The cost is then studied to determine the breakeven point and whether this move would be valuable to the organization. After the analysis steps result in a green light to move to the cloud, the planning, execution and realization steps would be implemented next. Finally, an evaluation step is conducted where the organization can post-analyze, evaluate, document the lessons learned and learn from the process, to excel in this framework over time and transfer the knowledge to assessing new opportunities over time.

Another method to assess HPC applications' applicability on the cloud is by using Hyperion Research's Cloud Application Assessment tool launched in February 2019 [7]. This tool was created with the intent of helping with the assessment of whether an organization can move all or parts of their HPC activities to the cloud. Several factors are important to assess before an organization can make a decision of whether their HPC activities should be moved to the cloud or not. The tool takes into account the most determining factors, which are listed in Table 1.

According to the tool, organizations that score high can move to the cloud, whereas organizations whose activities score low would be fit for on-premise High-Performance Computing resources. Organizations that score in the middle are more suitable for a hybrid environment. This tool could be added to the assessment framework as a pre-assessment step, or could be used independently. In the next section a case study will be conducted on evaluating the move of a reservoir simulation application to the cloud. This case study has common characteristics with a wide range of HPC applications used in the O&G upstream industry such as seismic applications or training of machine learning applications.

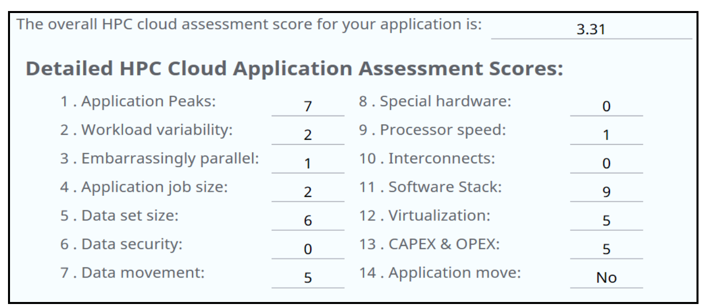
Table 1. Hyperion Research Public Cloud Scorecard

Specific Characteristics
Workload Peaks
Workload Variability
Workload Parallelism
Workload Size
Data Set Size
Data Security
Data Movement During Processing
Required Processors Speed
Interconnects Required
Application Software Stack
Application Virtualization Performance
CAPEX Considerations

B. *Case Study: Assessment of the Migration of Reservoir Simulation to the Cloud*

The scenario takes place in an O&G company where an HPC center is already set up with several thousand-compute nodes. Frequent reservoir simulation studies are conducted on the clusters that run an in-house developed reservoir simulator. The clusters are in high but consistent usage; as reservoir simulation studies are conducted continuously. However, it is in an environment where activity could spike due to abrupt shifts in simulation demands. The set-up is in a country that does not host its own cloud.

The cloud assessment tool is used that takes into account the different variables listed in Table 1. Fig. 4, illustrates the result of the assessment and will be used as a guide to the case study.


The overall HPC cloud assessment score for your application is: 3.31

Detailed HPC Cloud Application Assessment Scores:

1 . Application Peaks:	7	8 . Special hardware:	0
2 . Workload variability:	2	9 . Processor speed:	1
3 . Embarrassingly parallel:	1	10 . Interconnects:	0
4 . Application job size:	2	11 . Software Stack:	9
5 . Data set size:	6	12 . Virtualization:	5
6 . Data security:	0	13 . CAPEX & OPEX:	5
7 . Data movement:	5	14 . Application move:	No

Fig. 4. Results of Hyperion Research Assessment Tool for Reservoir Simulation Application

A score of 3.31 indicates that migrating to the cloud is not favorable for the reservoir simulation activity in this particular scenario but shows that a hybrid environment could be more suitable for this application. However, there are more variable reasons for moving to the cloud, therefore some of the determining characteristics are explored further.

1) *Application Peaks*

The day-to-day workload for reservoir simulation in the company is fairly consistent, however, there are occasionally urgent demands for conducting more runs than average. Such pressing runs will undeniably impact normal company activities by delaying the standard workload and creating a backlog of simulations. Having cloud computing as a supplementary resource during urgent demand will allow for more timely execution of all the tasks. In occasions where it is unyielding to attend to high demands, the alternative to using the cloud would require investing in more compute resources. This however, will lead to owning more resources than needed on average and the underutilization of resources during periods of standard workload.

2) Data Security

Security is often cited as one of the key concerns for the O&G industry as most of the data is highly confidential. However, as mentioned in the literature review and from surveying cloud service companies, it has become evident that cloud providers have advanced in that field by employing security experts to handle security vulnerabilities in the cloud. In addition, encryption when uploading to the cloud has been made available by most cloud providers. One limitation this creates is that the reservoir simulator in the scenario would need to be modified as such that the encrypted data can be executed and decrypted by the application. Since this case study is examining an in-house reservoir simulator, modifying it to handle the simulation of encrypted data could be feasible but costly. Another limitation that is of high concern is data jurisdiction. The cloud clusters are governed by the laws of the country that they are hosted in. In this scenario, the country does not have its own cloud. The concern is that data could be subpoenaed based on the jurisdiction of the country housing the cloud servers, and therefore hosting the cloud in the same country would be extremely advantageous.

3) Capital Expense

Capital expense is high when instantiating a new HPC center. The capital expense in this particular scenario is low as the computer center is already setup. However, in the situation of the spike in workload, the alternative scenario to using the cloud as a supplemental resource is to invest in more computer resources. This would result in complexities as it might entail expanding the on-premise center, developing new contracts to acquire more compute resources and even could result in establishing a new HPC center to host new hardware. The process is costly, tedious and could take years. Supplementing the on-premise resources with the cloud allows for avoiding increase in capital expense and increases flexibility. In addition, electric power is another factor to be considered, as increasing the hardware would also require increase in the power supply. Any increase in the power requirements must be planned and its feasibility must be studied. High power needs are becoming a challenge when planning HPC centers and its high cost is a deterring factor when increasing HPC resources. It is important to note that migrating to the cloud, although is faster than the process of setting up an on-premise center, is in fact not instant. The needed logistics for the particular application needs to be discussed with the CSP along with the relevant contractual agreements.

IV. Summary and Conclusion

Moving HPC applications to the cloud is still a debate for the O&G industry and in fact the O&G industry is amongst the slowest to make the transition. The literature review conducted on advances in cloud computing shows that significant research studies have been pursued focusing on the deployment of HPC applications in the cloud. The advancements have been made at a pace that results in expectation of a promising future for HPCC. History and advancements of HPCC in areas such as network speed and latency, virtualization, resource management, scheduling, data management, and security demonstrate that the future predicted trend is very attainable. An assessment tool is explored which aids in deciding on the cloud applicability of HPC applications and is used to assess a specific case study of moving reservoir simulation in an O&G company to the cloud.

It is demonstrated in the case study that there are various potential benefits for HPCC in the O&G. If O&G companies aim at keeping up with the latest hardware, then moving all or part of their HPC applications to the cloud would be a viable approach. Technology is continuously progressing and the agility cloud computing provides allows organizations to be adaptable to the changing hardware, environment and resource demand. Furthermore, the jurisdiction in which the cloud data falls under, and the dependency of its security on the laws and regulations of the country hosting the cloud leads to a conclusion that for O&G industry it is safest to host the cloud in the organizations' corresponding country. However, when examining the cost of HPCC, it is determined that for a consistent activity such as reservoir simulation studies, establishing an on-premise center, although high in capital expense, is more economical for the long run. If CSP do not reduce their costs, cloud usage will be mostly confined to handling spike demand and for that, the hybrid cloud environment is the optimal solution.

References

[1] A. Gupta and D. Milojicic, "Evaluation of HPC Applications on Cloud," in *2011 Sixth Open Cirrus Summit*, 2011, pp. 22–26.

[2] Y. Hu, X. Long, and J. Zhang, "Enhance Virtualized HPC System Based on I/O Behavior Perception and Asymmetric Scheduling," in *2012 IEEE 14th International Conference on High Performance Computing and Communication 2012 IEEE 9th International Conference on Embedded Software and Systems*, 2012, pp. 169–178.

[3] M. Ali, S. U. Khan, and A. V. Vasilakos, "Security in cloud computing: Opportunities and challenges," *Inf. Sci. (Ny).*, vol. 305, pp. 357–383, 2015.

[4] R. K. Perrons and A. Hems, "Public, Private, or Hybrid? What the Upstream Oil & Gas Industry Can Learn from Other Sectors about 'the Cloud,'" *Abu Dhabi International Petroleum Conference and Exhibition*. Society of Petroleum Engineers, Abu Dhabi, UAE, p. 8, 2012.

[5] C. Vecchiola, S. Pandey, and R. Buyya, "High-Performance Cloud Computing: A View of Scientific Applications," in *2009 10th International Symposium on Pervasive Systems, Algorithms, and Networks*, 2009, pp. 4–16.

[6] M. Efthymiadou and A. Oberhauser, "The Transition of High Performance Computing Applications from Supercomputers and Clusters to the Cloud," 2014.

[7] Hyperion Research (2019, Aug.). *Cloud Application Assessment Tool* [Online]. Available: https://hyperionresearch.com/cloud-application-assessment-tool/.

[8] TOP500 Supercomputer Sites (2019, Jun.). *TOP500 June 2019 List* [Online]. Available: http://www.top500.org/.

[9] V. Mauch, M. Kunze, and M. Hillenbrand, "High Performance Cloud Computing," *Futur. Gener. Comput. Syst.*, vol. 29, no. 6, pp. 1408–1416, Aug. 2013.

[10] J. Jose, M. Li, X. Lu, K. C. Kandalla, M. D. Arnold, and D. K. Panda, "SR-IOV Support for Virtualization on InfiniBand Clusters: Early Experience," in *2013 13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing*, 2013, pp. 385–392.

[11] R. R. Expósito, G. L. Taboada, S. Ramos, J. Touriño, and R. Doallo, "Performance analysis of HPC applications in the cloud," *Futur. Gener. Comput. Syst.*, vol. 29, no. 1, pp. 218–229, 2013.

978-1-7281-5020-8/19/$31.00 ©2019 IEEE

[12] P. Fan, Z. Chen, J. Wang, Z. Zheng, and M. R. Lyu, "Topology-Aware Deployment of Scientific Applications in Cloud Computing," in *2012 IEEE Fifth International Conference on Cloud Computing*, 2012, pp. 319–326.

[13] NASA Advanced Supercomputing Division (2019, Mar.). *NAS Parallel Benchmarks* [Online]. Available: https://www.nas.nasa.gov/publications/npb.html.

[14] R. Hassani, M. Aiatullah, and P. Luksch, "Improving HPC Application Performance in Public Cloud," *IERI Procedia*, vol. 10, pp. 169–176, 2014.

[15] P. Ivanovic and H. Richter, "OpenStack cloud tuning for high performance computing," in *2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, 2018, pp. 142–146.

[16] M. Cardosa, M. R. Korupolu, and A. Singh, "Shares and utilities based power consolidation in virtualized server environments," in *2009 IFIP/IEEE International Symposium on Integrated Network Management*, 2009, pp. 327–334.

[17] F. Fernandes, D. Beserra, E. D. Moreno, B. Schulze, and R. C. G. Pinto, "A virtual machine scheduler based on CPU and I/O-bound features for energy-aware in high performance computing clouds," *Comput. Electr. Eng.*, vol. 56, pp. 854–870, 2016.

[18] J. Ren, Y. Qi, Y. Dai, Y. Xuan, and Y. Shi, "Nosv: A lightweight nested-virtualization VMM for hosting high performance computing on cloud," *J. Syst. Softw.*, vol. 124, pp. 137–152, 2017.

[19] D. C. Trudgian and H. Mirzaei, "Cloud CPFP: A Shotgun Proteomics Data Analysis Pipeline Using Cloud and High Performance Computing," *J. Proteome Res.*, vol. 11, no. 12, pp. 6282–6290, Dec. 2012.

[20] P. Prukkantragorn and K. Tientanopajai, "Price efficiency in High Performance Computing on Amazon Elastic Compute Cloud provider in Compute Optimize packages," in *2016 International Computer Science and Engineering Conference (ICSEC)*, 2016, pp. 1–6.

[21] Z. Hou, Y. Wang, Y. Sui, J. Gu, T. Zhao, and X. Zhou, "Managing high-performance computing applications as an on-demand service on federated clouds," *Comput. Electr. Eng.*, vol. 67, pp. 579–595, 2018.

[22] G. G. Castañé, H. Xiong, D. Dong, and J. P. Morrison, "An ontology for heterogeneous resources management interoperability and HPC in the cloud," *Futur. Gener. Comput. Syst.*, vol. 88, pp. 373–384, 2018.

[23] I. Sadooghi *et al.*, "Achieving Efficient Distributed Scheduling with Message Queues in the Cloud for Many-Task Computing and High-Performance Computing," in *2014 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, 2014, pp. 404–413.

[24] G. Lee, B.-G. Chun, and H. Katz, "Heterogeneity-aware Resource Allocation and Scheduling in the Cloud," in *Proceedings of the 3rd USENIX Conference on Hot Topics in Cloud Computing*, 2011, p. 4.

[25] T. A. Henzinger, A. V Singh, V. Singh, T. Wies, and D. Zufferey, "Static Scheduling in Clouds," in *Proceedings of the 3rd USENIX Conference on Hot Topics in Cloud Computing*, 2011, p. 1.

[26] A. Gupta, L. V Kalé, D. Milojicic, P. Faraboschi, and S. M. Balle, "HPC-Aware VM Placement in Infrastructure Clouds," in *2013 IEEE International Conference on Cloud Engineering (IC2E)*, 2013, pp. 11–20.

[27] T. S. Somasundaram and K. Govindarajan, "CLOUDRB: A framework for scheduling and managing High-Performance Computing (HPC) applications in science cloud," *Futur. Gener. Comput. Syst.*, vol. 34, pp. 47–65, 2014.

[28] J. Feblowitz, "Oil and Gas: Into the Cloud?," *J. Pet. Technol.*, vol. 63, pp. 32–33, 2015.

[29] H. Yuan, M. Mahdavi, and D. Paul, "Security: Digital Oil Field or Digital Nightmare?," *J. Pet. Technol.*, vol. 63, no. 08, pp. 16–18, 2011.

[30] P. Black, "Cyber Security, the Cloud, and Oil and Gas," *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, Dallas, Texas, USA, p. 8, 2018.

[31] B. Jia, T. W. Wlodarczyk, and C. Rong, "Performance Considerations of Data Acquisition in Hadoop System," in *2010 IEEE Second International Conference on Cloud Computing Technology and Science*, 2010, pp. 545–549.

[32] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The Hadoop Distributed File System," in *2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)*, 2010, pp. 1–10.

[33] M. A. S. Netto, R. N. Calheiros, E. R. Rodrigues, R. L. F. Cunha, and R. Buyya, "HPC Cloud for Scientific and Business Applications: Taxonomy, Vision, and Research Challenges," *ACM Comput. Surv.*, vol. 51, no. 1, pp. 8:1--8:29, Jan. 2018.

[34] A. Kumar *et al.*, "It's Raining Barrels: Cloud Computing in the O&G Industry," *SPE Middle East Artificial Lift Conference and Exhibition*. Society of Petroleum Engineers, Manama, Bahrain, p. 8, 2018.

[35] M. E. Eldred, A. Orangi, A. A. Al-Emadi, A. Ahmad, T. J. O'Reilly, and N. Barghouti, "Reservoir Simulations in a High Performance Cloud Computing Environment," *SPE Intelligent Energy Conference & Exhibition*. Society of Petroleum Engineers, Utrecht, The Netherlands, p. 8, 2014.