CLOUD-BASED HIGH PERFORMANCE COMPUTING

Gaining Value by Moving On-Premise Simulation to Cloud through POWER'BY

CLOUD-BASED HIGH PERFORMANCE COMPUTING

Gaining Value by Moving On-Premise Simulation to Cloud through POWER'BY

CONTENTS

INTRODUCTION

As enterprises of all sizes grapple with narrow margins in increasingly competitive international markets, they are constantly seeking ways to reduce costs to improve the bottom line. Chief Financial Officers, shareholders, investors and venture capitalists, are all demanding more for less. More from less staff, more from fewer suppliers, more from leaner infrastructure. Ultimately, seeking the highest possible income and growth, at the lowest possible cost.

This article focuses on an area of significant and regular capital expenditure, on-premise High Performance Computing (HPC). It examines transformational alternatives to the traditional purchase and usage model.

HPC is used for a wide-ranging and constantly increasing number of demanding corporate computing tasks. Within research, product design and engineering groups, it is often used for compute-intensive simulation and optimization tasks related to fluid flow dynamics, mechanical and structural stresses and electromagnetic behavior. It is also used in diverse area such as deep learning, drug discovery and financial forecasting.

The speed and capacity of installed HPC resources can strongly influence how quickly a new product is brought to market and its success against competition1. Corporations must regularly weigh the high cost of investing in new software, HPC hardware and related staffing, against the potential for increased productivity, reduced physical testing and improved competitive positioning those investments will bring.

The traditional enterprise model, up until about 15 years ago, was to buy and install computers and software for every computing requirement in the company.

Corporate email, databases, CRM, productivity tools, accounting, CAD/PLM/CAE/EDA, backup, software development, file storage, graphic design, marketing, VOIP–all had dedicated servers in various facilities across the company (figure 1). Each one required a housing facility, dedicated power & cooling, switches, UPS, firewalls, fire protection, cabling, racks and above all, significant manpower to manage, replace and upgrade, both the computers themselves and also the peripheral supporting infrastructure.

Every three to five years almost every part of the infrastructure was fully depreciated and required upgrades or complete replacements as new technology made items redundant or end-of-life. The associated expenses were widely seen as essential costs of doing business with little alternative.

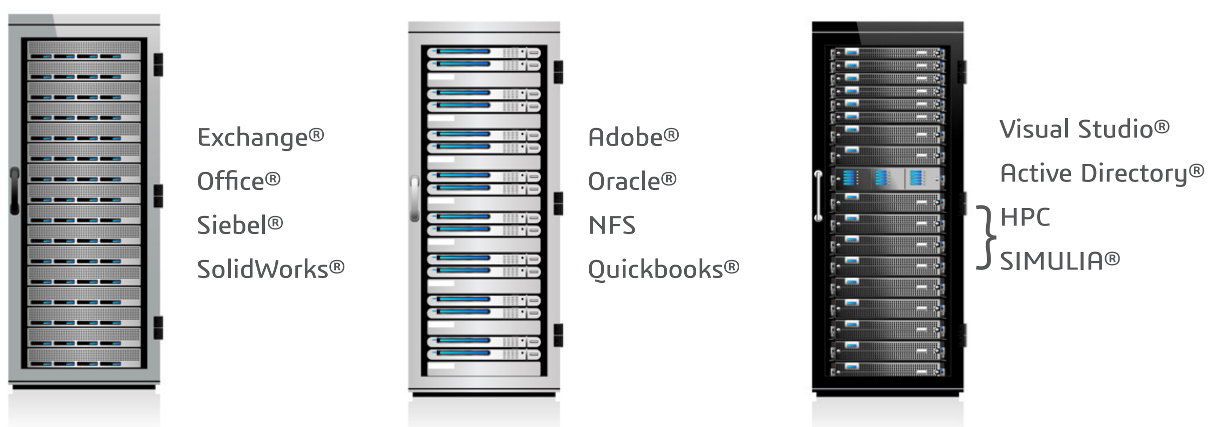

Figure 1: Representative on-premise server mix

EARLY DISRUPTION

In the early 2000's, Saleforce.com, Inc. started to disrupt the well-established Customer Relationship Management (CRM) database industry. Traditional CRM tools at the time had many features, but were large, complex and difficult to use. Access to the tools was typically slow and any locally installed version had to be regularly synchronized with a central server. Many sales professionals preferred to keep a simple local contact management database rather than deal with CRM headaches.

Salesforce tapped into that contact management demand with a simple, cloud based solution requiring no installation and no corporate IT involvement. A sales person could start using it at low cost with just an internet connection and few other barriers to entry. Over time, Salesforce increased the capabilities to replicate a fully-fledged and scalable CRM system. Today it and its competitors have redefined CRM into a dominant software as a service (SaaS) cloud business model. IT management no longer has to worry about dedicated CRM servers, complex installations and synchronizing multiple databases.

CRM on cloud is a huge success story, but there was resistance along the way. Sensitive data being outside of the corporate firewall being one of the largest concerns. Salesforce understood these concerns well and, along with technology partners, developed solutions that the market demanded, propelling CRM to become one of the largest worldwide software markets at around $40B a year in 2017[2].

CLOUD ADOPTION ACCELERATES

As the CRM market proved, the SaaS cloud model is attractive with low barriers to entry, zero user maintenance, rapid scalability and flexible pricing. It's collaborative by nature and maintains a single source of truth for shared data and projects. Over the past few years, and as broadband internet access has become ubiquitous, other corporate essentials have seen strong growth on cloud under a cloud SaaS model. Google with its G Suite of products, Microsoft with Office365, Adobe with Creative Cloud, marketing automation via Hubspot and Intuit with QuickBooks Cloud. Leveraging cloud for big-data analytics has also grown rapidly with companies such as Teradata, C3IoT and Cloudera, providing cloud based big-data solutions.

Software companies selling tools "cloud first" under a SaaS model have enjoyed rapid growth and the difference can be striking when comparing the growth of cloud first companies to those maintaining more traditional on-premise usage models (figure 2).

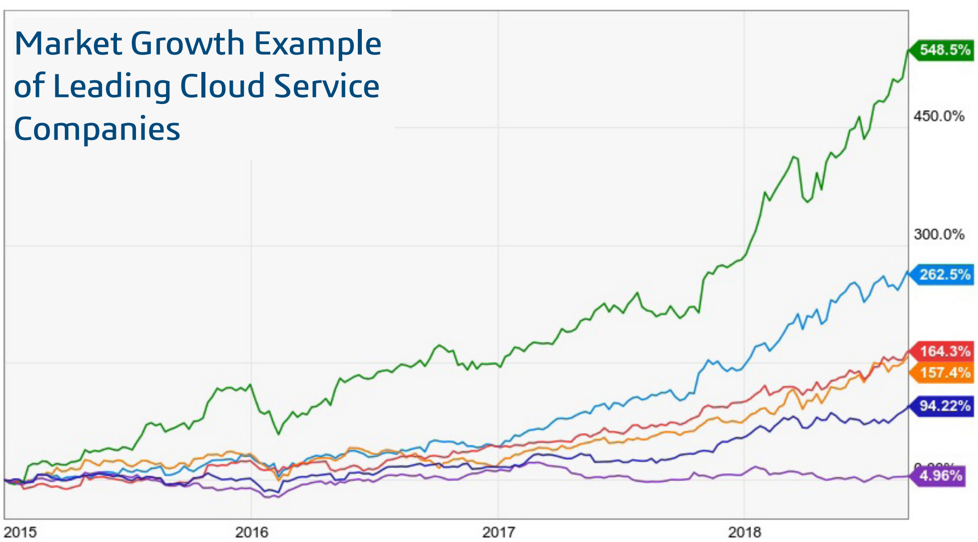

Figure 2: Cloud first companies achieve rapid growth

The reality today, is that it's now possible to run a thriving business with just a laptop for each employee and zero additional IT infrastructure. That means no server room, no dedicated power or AC, no upgrade cycle and no IT staffing requirement.

Some argue this model only works for the start-up or SMB, but rather than building new private datacenters, multi-billion dollar enterprises are already starting to migrate productivity and other key tools to cloud and over an extended period, retiring what amounts to the non-core business of managing IT.

In tandem with the huge growth of SaaS, perhaps not surprisingly, the cloud infrastructure to run it on has also seen rapid expansion. Amazon Web Services (AWS), Microsoft Azure and Google Cloud are currently 3 of the largest[3]. Typically referred to as infrastructure as a service (IaaS), this layer is often transparent to the end user who may unknowingly be using multiple cloud infrastructure, depending on the software in use. The SaaS developer will negotiate with the IaaS providers to find the option that best fits the requirements of the tool and offers them and their customers the best value.

Despite high growth, the SaaS and IaaS markets still have large expansion potential, only covering around 16% of the estimated $1.1 trillion total addressable enterprise market in 2017[4]. The market is booming for SaaS workloads that require scalable standard compute. Let's now consider high performance computing and why cloud is becoming an increasingly attractive option for demanding workloads.

COMPUTING CHALLENGES FOR SIMULATION AND ANALYSIS

Physics simulation, optimization, deep learning, life science simulation and financial modeling are all examples where specialized High Performance Computing is required in order to solve complex problems and complete them in a timely manner. The required architecture typically comprises multiple nodes (individual computers in a cluster), multiple computing cores per node, large shared memory, fast storage and high speed, low-latency interconnect. The cost and configuration options can be quite broad, ranging from $15K for a dual CPU, 32 core, 64 GB blade mounted in a server room rack, all the way to the latest multi-million dollar supercomputers as in the world's top five-hundred[5] .

Most commercial simulation and analysis algorithms, such as those developed by the SIMULIA brand of Dassault Systèmes, scale very well with the amount of hardware used. This means that while the larger systems will be the most expensive, they will also be the most productive. However, considering the highly variable, burst nature of HPC simulation workloads revolving around peak production cycles, an expensive system may be under-utilized in typical daily workload scenarios. Figure 3 shows how a large installed system can address the largest peak requirements, but be under-utilized on many other days. While, a lower-cost system might be fully utilized, the system could cause significant simulation job queuing bottlenecks due to lack of capacity on peak days.

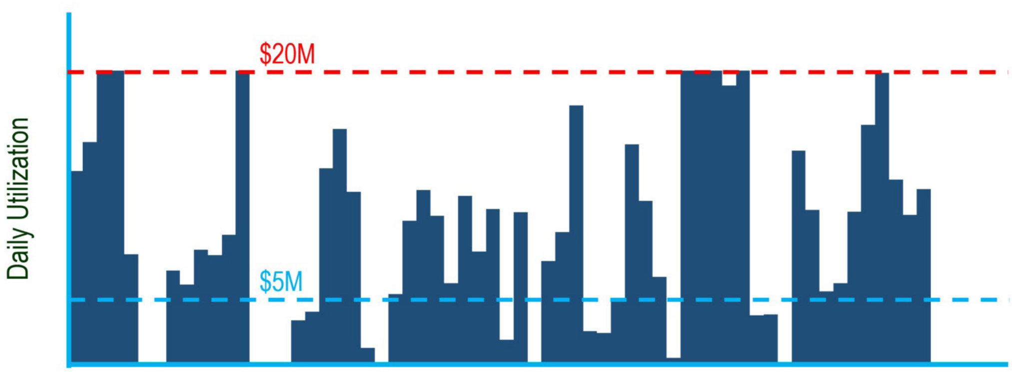

Figure 3: Comparing a fully utilized lower cost system requiring daily job queues to a more expensive system to accommodate peak loads, which may be severely under-utilized. Based on SIMULIA estimates.

Often regarded as a separate and specialized group within corporate IT, HPC management must try to ensure that they serve their internal engineering and scientific users with reliable, fast and sufficiently large, but highly utilized compute resources within a restricted budget. This is the first of several challenges when considering procurement of an HPC system.

HPC managers when considering purchasing hardware for their own server room or datacenter are looking at a broad price range, anywhere from $20K up to $500K for the raw hardware. Those looking at higher-end systems will often attend the annual Supercomputing event in the United States or International Super Computing (ISC) in Europe to look for their next system. In recent years however, and as demonstrated at those events, the availability of increasingly specialized processor types such as co-processors, GPUs and ASICs, and multiple interconnect technologies such as infiniband, Omnipath and NVlink (figure 4) have made the purchasing decision more difficult. Further compounding the problem is the lengthy in-service requirement and the increasing diversity of modern HPC workloads, which are often highly optimized for certain hardware configurations.

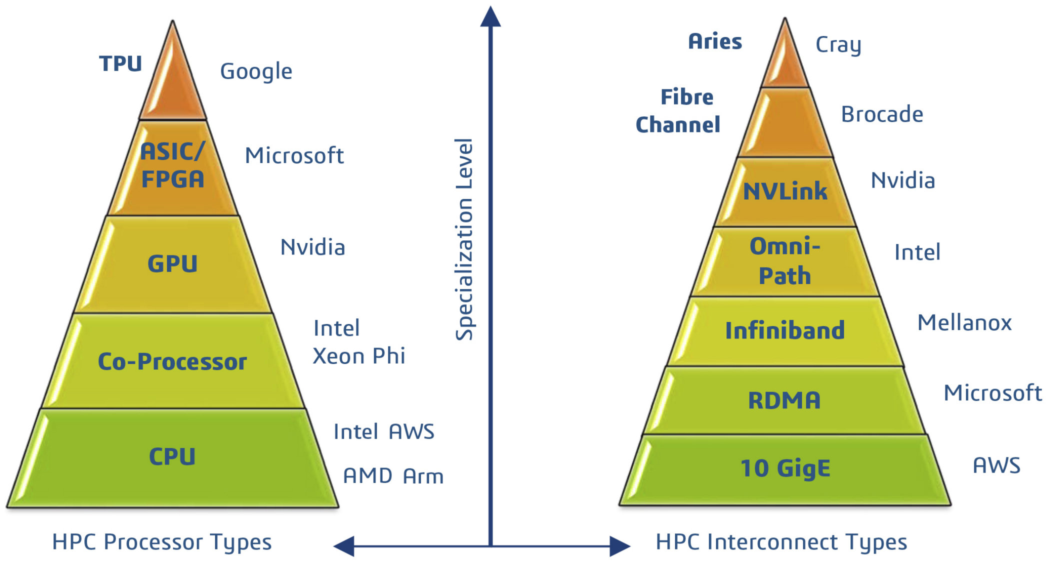

Figure 4: Modern HPC architecture diversity

The HPC infrastructure required for heavy-duty simulation and other demanding workloads is a significant investment for the enterprise and due to continued technology advances, will continue to be so. Unfortunately, this also leads to rapid depreciation and replacement cycles – there is little market for 4 or 5-year-old HPC equipment, particularly if it's based on highly specialized architecture, where technology ages rapidly.

In addition to the raw hardware, there are the many associated peripheral costs as mentioned in the introduction and as shown in figure 5, which have their own expensive upgrade and replacement cycles.

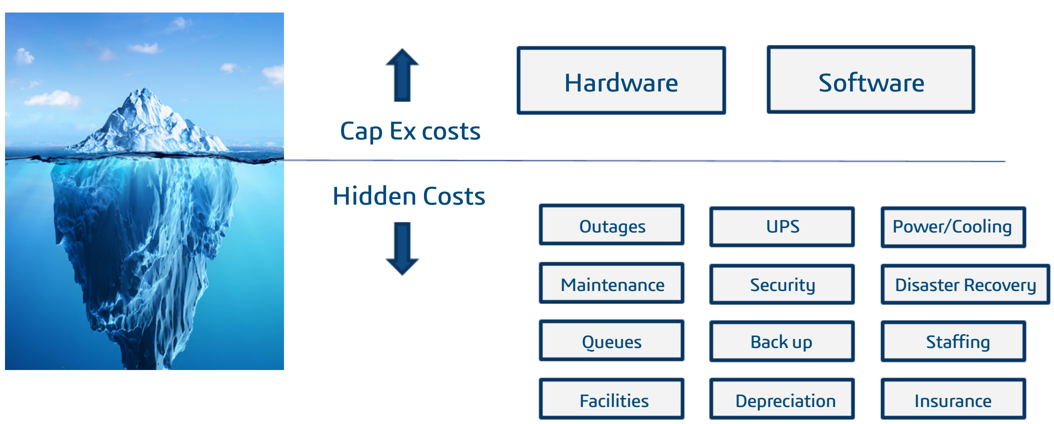

Figure 5: Hidden costs of on-premise HPC.

THE REAL COST OF OWNING HARDWARE

A few years ago, IDC interviewed over 300 HPC managers and CIOs across a range of industries and platforms and asked them about their 3-year infrastructure cost breakdown[5] (figure 6).

The result was surprising in that the software and hardware itself was only 14% of the total 3-year cost. Staffing was the major cost over the period, at more than half of the total expenditure.

Consider a typical 256 core HPC cluster with 1 TB RAM. Such a system will have a list price of around $100K in mid-2018 and can be purchased from HPE, Dell and others. Additional hardware such as switches, interconnect and storage will increase this to around $120K. Based on the IDC data, this system will conservatively have a total 3-year ownership cost of $1M. Assuming a 100% utilization rate, the cost per core per hour over the 3-year period is $0.15.

Actual utilization rates of such systems are lower – HPC workloads tend to fluctuate around product lifecycles and utilizations rates are typically in the range of 40% at smaller companies up to 70% at larger enterprises. Figure 7 shows how lower utilization rates increase the effective core/hour cost significantly.

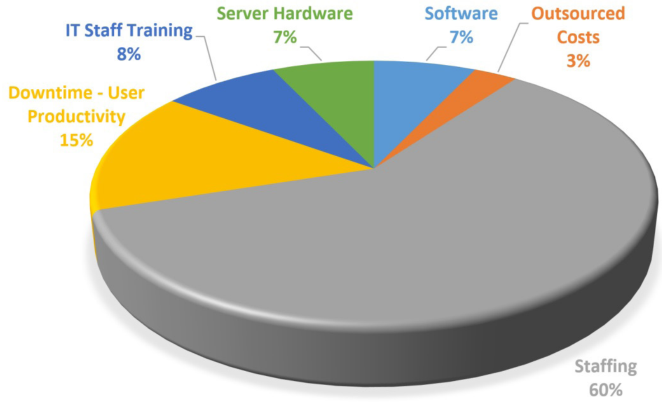

Figure 6: Typical three year cost to own on-premise Infrastructure

HPC WORKLOADS ON CLOUD

There are an ongoing and increasing number of challenges to selecting and owning HPC hardware. Some of the reasons mentioned for strong adoption of SaaS models for standard compute tasks also apply to HPC and in certain situations, can be even more compelling.

Consider figure 3 again and compare it to the cloud version in figure 8. The elastic and on-demand nature of cloud allows it to fully scale up to the service level requirement, but critically, scale back down on non-peak days. This means that the total 4 year hardware cost in this example will be in the region of $5M in accumulated operating expenses, but delivering the service level of a $20M on-premise system.

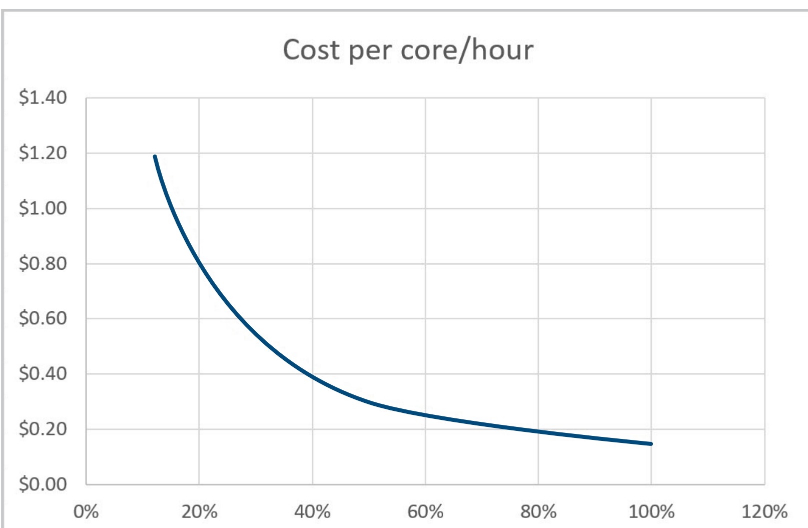

Figure 7: Effective cost per core/hour with utilization rate

Despite the advantages, HPC on cloud has only recently started to generate excitement and interest. As SaaS did in its early days, cloud HPC has raised fears at buyers and even sellers. In cloud HPC, there is no "one size fits all" and each type of analysis may perform better on one type of architecture than another. Even within physics simulation applications, each solver algorithm can be highly architecture performance-dependent. While sales people may be wary of the perceived complexities, buyer's concerns center around having full access to the variety of technology shown in figure 4 without being locked-in, security for critical IP related data, virtualization overhead, remote visualization and cost.

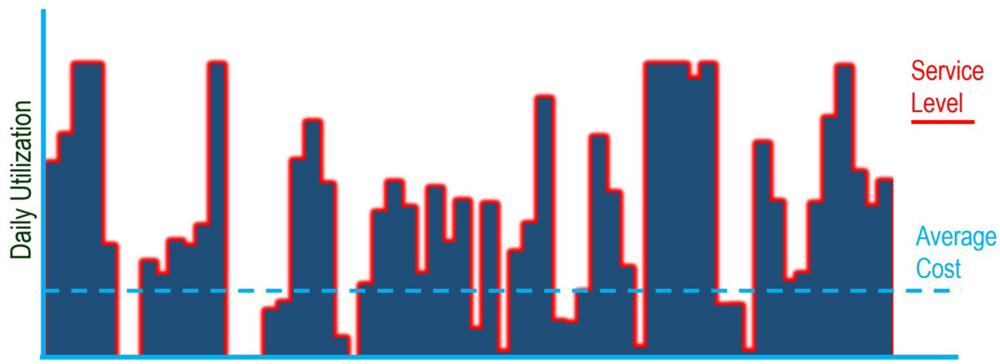

Figure 8: Cloud HPC cost efficiency for variable workloads

The good news for all stakeholders is that many of these concerns are no longer justified with the very latest hardware now routinely available on cloud and accessible with fast remote visualization to nearly replicate the local experience. Benchmarking and individual proofs of concept (POCs) have proliferated to provide technical reassurance in the sales and solution process. Security is rapidly becoming less of an issue as stakeholders gain confidence and understanding around certifications, end-to-end encryption and public cloud provider's standards for data protection.

Cost is the one standout and a contentious issue that is vigorously debated. Compared to traditional SaaS, HPC cloud workloads are demanding and there are many cost elements to consider in an on-demand SaaS type model–run times, hardware, memory, storage and data transfer including visualization. Additionally, HPC managers must consider whether to use the public cloud directly and deal with associated costs of the compute stack or middleware to manage their usage as well as tool installation, updates and license hosting. Alternatively, they might consider a third party hosting company working together with public cloud providers and independent software vendors (ISVs) to manage those items for them – at a cost.

In the section below we'll breakdown and compare costs or total cost of ownership (TCO) for cloud HPC.

COMPARING CLOUD AND ON-PREMISE COSTS FOR HPC

When considering cloud as an alternative to on-premise infrastructure, HPC management often ask how much it costs to run in the cloud on a core hour basis, so that they can compare to the cost of a particular core in their datacenter. As stated previously however, the total ownership cost may not be fully broken down. Purely on an initial purchase basis, the $120K system outlined above, would show a core/hour cost for a 3-year period of $0.018, almost 10x lower than the more realistic, all-inclusive cost of $0.15. With economies of scale–for very large installations, the on-premise price can be reduced to below $0.09.

Even considering all the available data, senior management may still underestimate the productivity losses caused by owning an insufficiently sized HPC infrastructure. When jobs queue for extended periods, product releases can be delayed causing downstream losses in time-critical markets such as high-tech.

Figure 9 shows various representative public cloud prices vs. the on-premise price for a system similar to that described above, with all systems shown being based on the Intel Skylake microarchitecture. The best cloud option will depend on commitment level, the service level required and whether a fully managed option is desired.

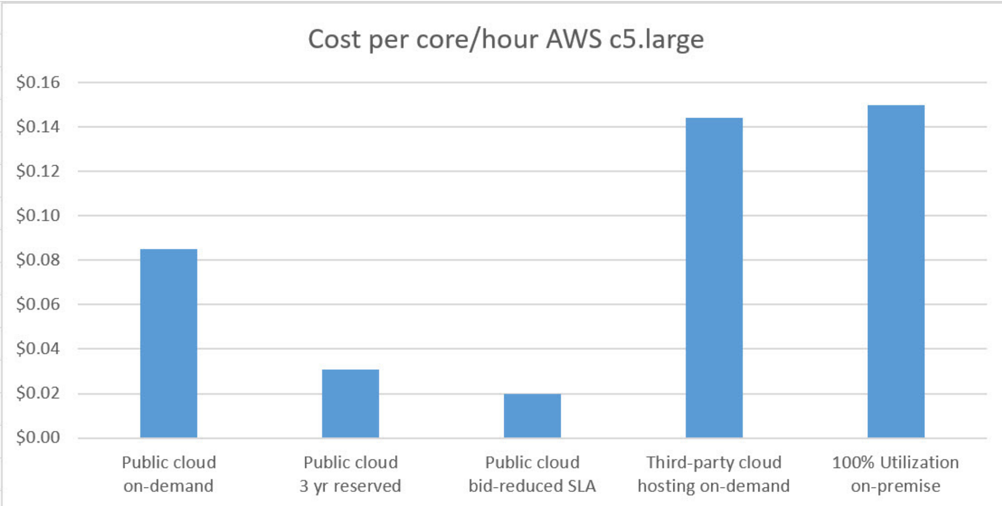

Figure 9: Cost per core/hour of various public cloud hardware models vs on-premise.

THE DASSAULT SYSTEMES 3DEXPERIENCE PLATFORM APPROACH TO CLOUD HPC

Dassault Systèmes is the **3D**EXPERIENCE® company providing businesses and people with virtual universes. Its **3D**EXPERIENCE platform leverages the Company's world-leading 3D software applications to transform the way products are designed, produced, and supported. Dassault Systèmes is in a unique industry position offering a complete suite of connected tools, all the way from early stage product conception, through physics based optimization to manufacturing simulation and in-service operation. Rather than forcing customers to seek a third party hosted cloud HPC solution, it's cloud platform approach allows turnkey access to software and compute infrastructure, offering an all-in-one solution for cloud based simulation.

As physics simulation has grown to become a critical part of the product and system design and optimization phase, the SIMULIA brand of Dassault Systèmes has focused on delivering a robust portfolio of multiphysics / multiscale simulation applications. Those include mechanics, fluids and electromagnetics, vibroacoustics, multibody simulation and design optimization. However, a few years ago, visionaries at Dassault Systèmes realized that it didn't make sense for all of this technology to exist in isolation–the end user working in a silo with arduous team collaboration and dealing with unreliable or slow data translation between department tools. Simulation was particularly isolated and widely seen as such a highly specialized analyst role that it was off limits to CAD designers or PLM specialists for example. Even analysts working in different areas of physics were isolated from each other, despite those physics often being interdependent.

Their solution was to create a comprehensive, fully collaborative, cloud native platform that manages the tools, data, hardware and licensing in a cohesive modern framework, offering SaaS, IaaS and PaaS (platform as a service). All of these on-demand services working seamlessly to enable turnkey, scalable, cross department workflows. Much like modern CRM, cloud based data and results can be selectively shared and worked on by a team worldwide while maintaining a traceable single source of truth. For corporations with globally distributed design and manufacturing centers, this collaborative approach offers faster time to market and lower costs through improved project organization, communication and productivity. Gone are the typical errors and delays associated with transferring, converting and versioning files in a multi tool, multi-location environment.

In many companies, data, files and results are stored and replicated multiple times, both intentionally for backup or localization purposes and unintentionally due to poor organization–all leading to excess and increasingly unmanageable disk usage. This is particularly pertinent for simulation data where very large results files can be generated. The **3D**EXPERIENCE platform eliminates unnecessary replication and improves efficiency and data backup and disaster recovery are built in.

Multiphysics and multiscale simulation is now a key element of successful and competitive product development and by democratizing it through the **3D**EXPERIENCE platform to both specialist and non-specialist alike, via tailored roles and apps, a much wider corporate group is now able to incorporate the insight and benefits that simulation provides.

With the move to cloud being a major undertaking for large corporations reliant on simulation and who already have investements in on-premise infrastrucutre, the **3D**EXPERIENCE platform allows organizations to work in a hybrid fashion – for each simulation, choosing either to run on-premise or use the cloud, but benefiting from cloud collaboration in either case.

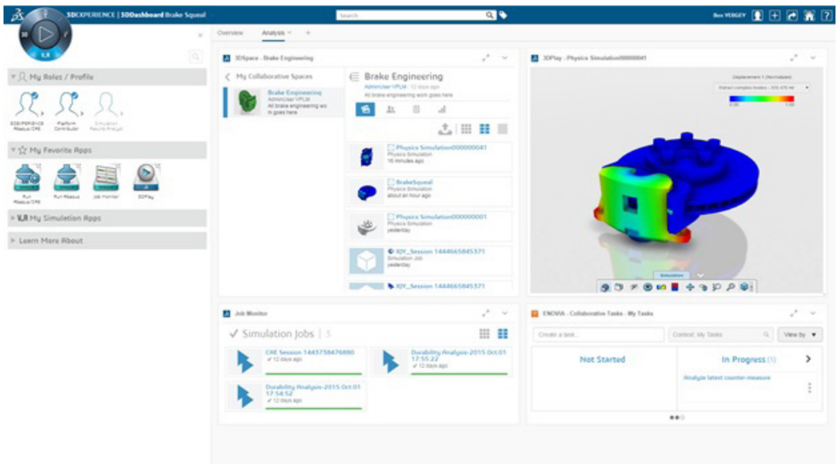

Figure 10: **3D**EXPERIENCE platform

Adding HPC on cloud to the platform has created greater flexibility for product manufacturers. In a one-stop-shop they can access software, hardware and all the peripheral services, on-demand in a burst or sustained manner. Small companies can maintain or move to a zero IT footprint and large enterprises can try cloud HPC and work in a hybrid fashion with their own datacenter, as they build confidence in cloud solutions.

CLOUD OFFER

To enable existing customers using traditional on-premise products to benefit from the **3D**EXPERIENCE platform right away at minimum investment, Dassault Systèmes provides a POWER'BY option. POWER'BY allows non-platform products to run in the **3D**EXPERIENCE platform environment and take advantage of all the platform features such as collaboration, data management, and on-demand cloud execution, as well as access to native **3D**EXPERIENCE apps.

To allow straightforward and low cost access, three initial cloud compute pack offers have been developed around the POWER'BY Structural Mechanics Analyst role that include all the pre-requisites to run on the platform and on the cloud. These turnkey offers include software and hardware computation hours, storage, data transfer and core platform features, including advanced post-processing and visualization for packs 2 and 3. (figure 11).

SUMMARY

Cloud HPC is now a viable alternative to on-premise HPC and enterprises have options to go all-in and run with a "zero IT" footprint, or work in a hybrid on-premise/cloud environment, to capture the optimum cost/benefits.

With highly diverse and "peaky" corporate workloads and increasingly specialized architecture requirements, almost all companies using HPC can benefit from the diverse, scalable and on-demand nature of cloud.

There are many elements of a successful cloud HPC deployment and cost is an overriding concern. Having a detailed breakdown and forecast of on-premise HPC requirements and downstream effects of insufficient capacity, often shows cloud HPC costs to be lower over an extended period.

	Cloud Compute Pack 1	Cloud Compute Pack 2	Cloud Compute Pack 3	Add on Pack
Expected compute time	100 hours	250 hours	500 hours	100 hours
Included storage	100 GB	200 GB	1 TB	x
Structural Mechanics Analyst Role (web-based)	✓	✓	✓	x
3DEXPERIENCE platform and pre-requisites	✓	✓	✓	x
High performance cloud post-processing	x	✓	✓	x

Figure 11: Cloud compute packs to enable **3D**EXPERIENCE Platform cloud HPC

The Dassault Systèmes **3D**EXPERIENCE platform is a cloud-first example of a fully collaborative, scalable, on-demand environment that is inherently hybrid. CIOs, HPC managers and end-users should evaluate such an environment as they consider cloud HPC alternatives and their overall product development workflows.

REFERENCES

1 https://ec.europa.eu/digital-single-market/en/news/high-performance-computing-factsheet

2 https://www.gartner.com/newsroom/id/3871105

3 https://www.marketwatch.com/story/amazon-rules-the-public-cloud-but-google-microsoft-alibaba-are-growing-faster-2017-12-20

4 http://www.bain.com/Images/BAIN_BRIEF_The_Changing_Faces_of_the_Cloud.pdf

5 https://www.top500.org/

6 IDC 2007, Typical 3 year Server TCO based on over 300 interviews conducted across numerous platforms. Presented in composite form.

Our **3D**EXPERIENCE® platform powers our brand applications, serving 12 industries, and provides a rich portfolio of industry solution experiences.

Dassault Systèmes, the **3D**EXPERIENCE® Company, provides business and people with virtual universes to imagine sustainable innovations. Its world-leading solutions transform the way products are designed, produced, and supported. Dassault Systèmes' collaborative solutions foster social innovation, expanding possibilities for the virtual world to improve the real world. The group brings value to over 210,000 customers of all sizes in all industries in more than 140 countries. For more information, visit **www.3ds.com**.

3DEXPERIENCE®

DASSAULT SYSTEMES | The **3D**EXPERIENCE® Company

Americas
Dassault Systèmes
175 Wyman Street
Waltham, Massachusetts
02451-1223
USA

Europe/Middle East/Africa
Dassault Systèmes
10, rue Marcel Dassault
CS 40501
78946 Vélizy-Villacoublay Cedex
France

Asia-Pacific
Dassault Systèmes K.K.
ThinkPark Tower
2-1-1 Osaki, Shinagawa-ku,
Tokyo 141-6020
Japan