



HYPERION RESEARCH

Recent Research on Using Public Clouds for HPC Workloads

Alex Norton

September 10, 2019

Definitions and Overview

- When we say HPC public cloud computing (or just cloud computing for short), we mean running HPC or HPC-enabled workloads, including AI, on third-party resources.
 - These resources may be owned and operated by a CSP, like Google, AWS, Azure, or others.
 - Or, they may be run by a vendor, like Dell, HPE, IBM, etc.
 - In addition, we also track PRIVATE CLOUDS, but as a separate category.

Major Trends

HPC in the Cloud



- **Over 70% of HPC sites run some jobs in public clouds**
 - Up from 13% in 2011
- **About 20% of all HPC jobs are now running in clouds**
 - A major uptick from the 10% figure two years ago
 - Key concerns: security and data loss, sometimes cost
 - Ultimate limiter of cloud growth may be data locality
- **Large public clouds are going heterogeneous – and are getting much better at running a broader set of HPC workloads**
 - AWS with FPGAs, Inferentia, GPUs, etc. Google with their TPU, Azure with bare metal Cray.

The Total HPC Market Including Public Cloud Spending

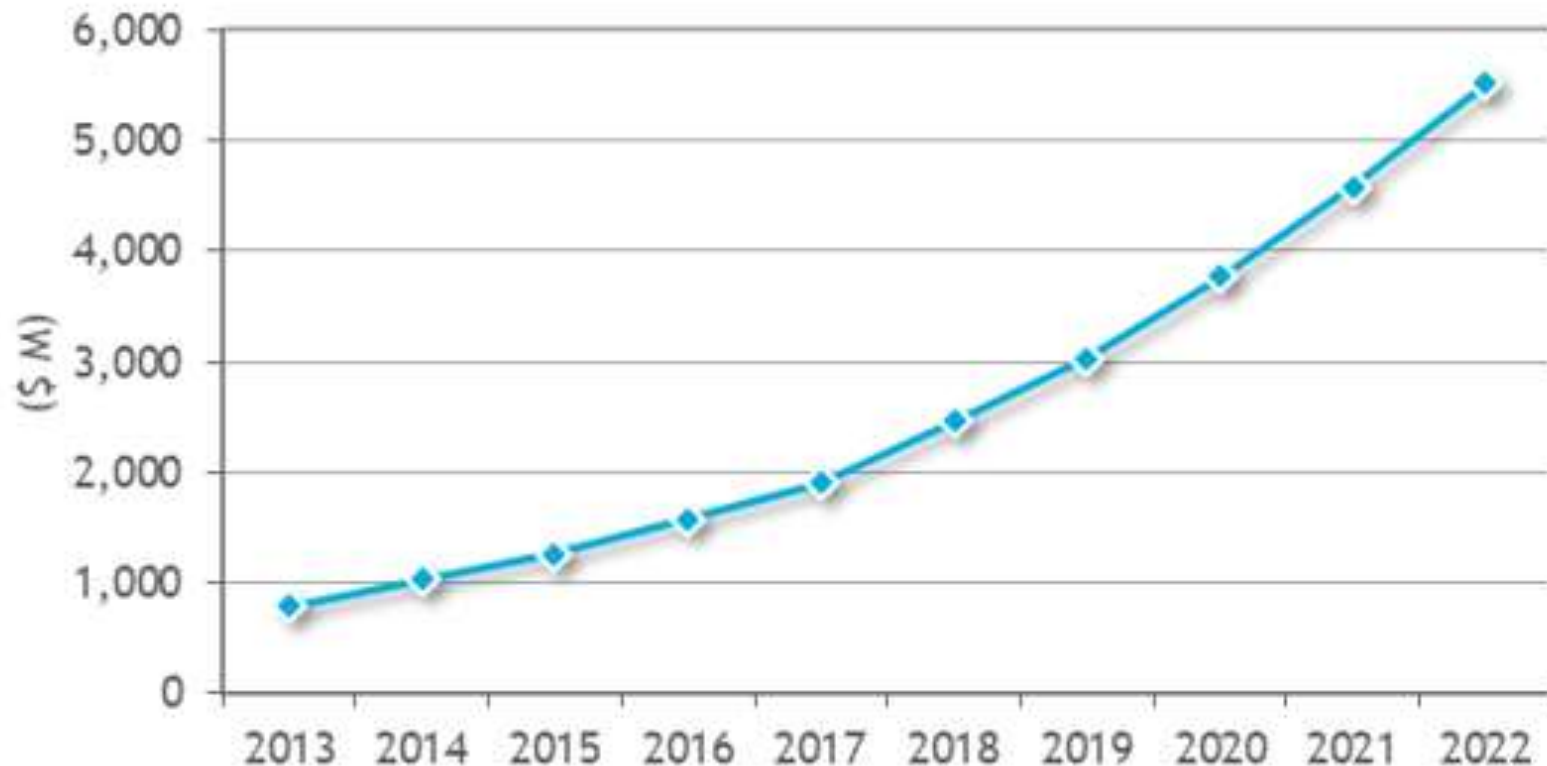
- TOTAL HPC spending grew from \$22B in 2013 to \$28B in 2018, and is projected to reach \$39B in 2023 - \$44B counting cloud usage

Revenues by the Broader HPC Market Areas			
	2018	2023	CAGR 18-23
Server	13,706,088	19,979,016	7.8%
Storage	5,547,188	7,771,184	7.0%
Middleware	1,582,892	2,217,801	7.0%
Applications	4,627,492	6,413,592	6.7%
Service	2,229,921	2,858,820	5.1%
Total Revenue	27,693,580	39,240,413	7.2%
Source: Hyperion 2019			

HPC cloud (CSP) usage raises forecast to \$44 billion

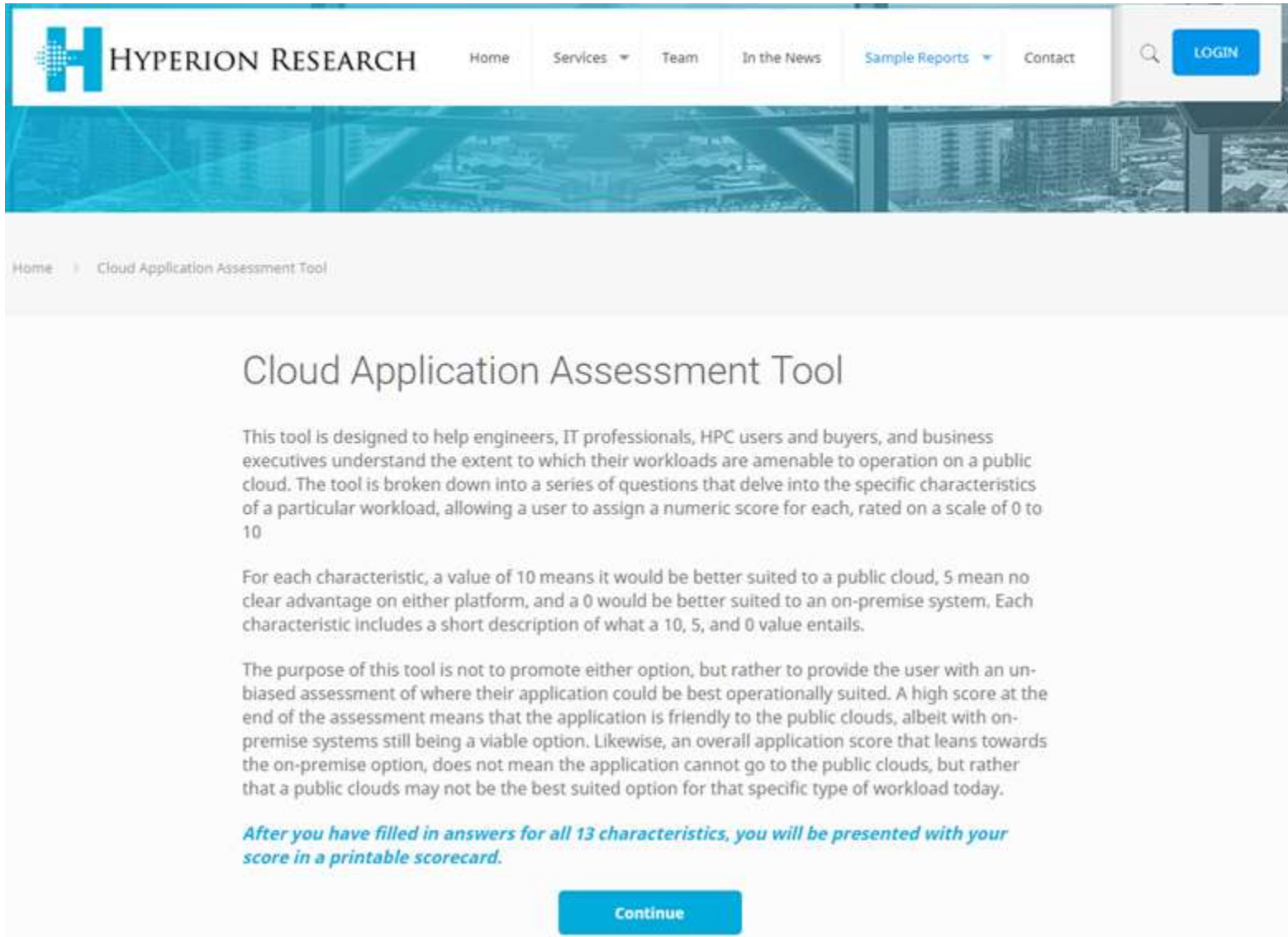
HPC Public Cloud Forecast

Projected Spending in Public Clouds for HPC Applications



A New Cloud Application Assessment Tool

<https://hyperionresearch.com/cloud-application-assessment-tool/>



The screenshot shows the top navigation bar of the Hyperion Research website. The logo is on the left, followed by a menu with items: Home, Services (with a dropdown arrow), Team, In the News, Sample Reports (with a dropdown arrow), and Contact. On the right side of the navigation bar, there is a search icon and a blue 'LOGIN' button. Below the navigation bar is a large blue-tinted image of a modern building. Underneath the image is a breadcrumb trail: Home > Cloud Application Assessment Tool. The main heading is 'Cloud Application Assessment Tool'. The text below explains the tool's purpose: it helps engineers, IT professionals, HPC users, and business executives understand if their workloads are suitable for public clouds. It is a series of questions with scores from 0 to 10. A score of 10 is best for public cloud, 5 is neutral, and 0 is best for on-premise. The tool is unbiased. A blue button labeled 'Continue' is at the bottom.

HYPERION RESEARCH

Home Services Team In the News Sample Reports Contact

LOGIN

Home > Cloud Application Assessment Tool

Cloud Application Assessment Tool

This tool is designed to help engineers, IT professionals, HPC users and buyers, and business executives understand the extent to which their workloads are amenable to operation on a public cloud. The tool is broken down into a series of questions that delve into the specific characteristics of a particular workload, allowing a user to assign a numeric score for each, rated on a scale of 0 to 10

For each characteristic, a value of 10 means it would be better suited to a public cloud, 5 mean no clear advantage on either platform, and a 0 would be better suited to an on-premise system. Each characteristic includes a short description of what a 10, 5, and 0 value entails.

The purpose of this tool is not to promote either option, but rather to provide the user with an unbiased assessment of where their application could be best operationally suited. A high score at the end of the assessment means that the application is friendly to the public clouds, albeit with on-premise systems still being a viable option. Likewise, an overall application score that leans towards the on-premise option, does not mean the application cannot go to the public clouds, but rather that a public clouds may not be the best suited option for that specific type of workload today.

After you have filled in answers for all 13 characteristics, you will be presented with your score in a printable scorecard.

Continue

Results From a Recent Study

- Hyperion Research conducts many large-scale cloud studies each year.
- The study was designed to understand more deeply the use of the cloud from the end user perspective, as well as dive into the ways in which providers are helping their customers to achieve their goals.

Jobs in the Cloud

- Our most recent study showed that the percent of applications moving to the cloud is continuing to increase:
 - Current cloud users reported that on average 33% of their jobs were running in the cloud.
 - If we extrapolate that out to the larger HPC community, we get a figure of roughly 20% of all HPC jobs are currently running in the cloud.

Hybrid Cloud Use

- Over half of the users in the study reported the use of a hybrid cloud environment (an environment that creates a layer of connection between a third-party cloud and an on-premise system).
 - Interestingly, burst computing was not the most common use of hybrid environments
→ experimentation was the most common use case.

Cloud Native HPC Users

- Cloud native users do not own any on-premise HPC hardware and run HPC jobs strictly in the cloud:
 - 40% of these users believe every job can be run in the cloud.
- One-third of the CSPs said there are certain jobs they advise against running in a cloud – mainly because of data locality

HPC applications or job types not run in an external cloud

	Number of Responses	Percentage of Sample
No, all applications could be run in an external public cloud	61	39.6%
Yes, some of our applications would not run well in an external public cloud	93	60.4%

n = 154

Source: Hyperion Research, 2019

In Conclusion

An Approaching Growth Elbow

- By sector: Industry is the fastest adopter of HPC cloud computing, followed by government and then academia.
 - Security issues tend to limit government customers
 - The price/cost is too high for some customers, especially in academia
- A growth inflection point is likely coming soon.
 - The data in this recent study showed that more workloads are moving to the cloud over the next 18 months.
 - Major growth will happen when more HPC jobs perform better in public clouds.
 - A major jump in adoption will also come when TCO analysis run by end users includes more costs than just the system cost (e.g. power, staff, and hard to quantify costs like the loss of work due to long queue times or the inability to use most effective hardware for their application).

Cloud Computing For HPC Workloads Will Grow Fast, Via Tipping Points



- Running HPC workloads in CSP environments will likely grow in step function-like leaps
 - As clouds get better at running HPC workloads
 - And as the costs are better understood
- HPC on premise and cloud environments will more closely resemble each other
 - *E.g. The containerization of HPC*
- 5G will be important for reducing latency in AI use cases that rely on coupled local-cloud environments
 - Such as automated driving systems, precision medicine, fraud detection, cyber security and Smart Cities/IoT
- Hybrid cloud environments will grow to be a highly viable option for HPC users

QUESTIONS?



anorton@hyperionres.com

ejoseph@hyperionres.com

sconway@hyperionres.com

bsorensen@hyperionres.com