**Edge and Cloud: Finding the Right Balance for Data Science Workloads**

October 2019

**Questions posed by:** Z by HP  
**Answers by:** Tom Mainelli, Group Vice President, Devices & AR/VR

---

**Q.** Many organizations have shifted some percentage of their data science workloads to the cloud. Is this always the best option in terms of both workload (performance and security) and cost efficiency?

**A.** Many companies moved to embrace the public cloud with all types of workloads, including data science workloads. Today, many of those organizations are reevaluating if the public cloud is always the best place for these workloads. For some, a closer look often reveals a poor fit for development, debugging, and exploration, with services that overserve their performance requirements at a higher cost than they had anticipated. Other organizations come to realize that early decisions around security have to be rethought because their literacy around these topics has increased.

As a result, some companies have begun the process of reviewing which workloads they should keep in the public cloud, which workloads they should move to private cloud, and which workloads should be run locally on devices such as high-powered, purpose-built workstations. According to a recent IDC survey of IT decision makers (ITDMs), 85% of respondents said their companies are actively repatriating public cloud workloads. The top reasons for doing so are security (19%), performance (14%), and cost (12%).

**Q.** Do you advocate a purely cloud-based workload strategy, an on-premise workload strategy, or a hybrid of the two for most companies?

**A.** Every company is different, but generally speaking, a one-size-fits-all approach is never the best; there are reasons to consider public cloud, private cloud, on-premise, and on-device workloads on a case-by-case basis. Too often, sweeping decisions are made without consulting the right cross-section of executives, managers, and rank-and-file employees. Those who make this effort will often find that the diverse needs of the organization are much better served by a hybrid model that embraces not only public and private cloud options but also robust on-premise and on-device workload options.
It is equally important to acknowledge that needs change over time, and workloads that once seem ideally suited to the cloud may be better suited to a local device today. With experience, many users may simply prefer the efficiency and peace of mind of running their complex workloads, such as data science workloads, on a workstation placed deskside. Furthermore, many users may find that they prefer to begin locally — that is, initially testing a hypothesis or running a "quick and dirty" job on a local device — before moving to the cloud for production or when the job expands to require it.

Q. In an ideal world, what percentage of applications/workloads would most IT decision makers like to have run locally instead of in a noncloud environment?

A. Each organization will be different, but a recent IDC survey showed that ITDMs currently run more than a quarter of their applications on-premise in noncloud instances such as local workstations. When we asked those same respondents what proportion of the public cloud applications installed today will move to a private cloud, hosted private cloud, or noncloud environment over the next two years, the answer was a significantly large 50%. Smart organizations know that situations change, client and employee needs evolve, and workloads shift over time.

Moreover, today's workstation offerings are radically different from those of just a few years ago, offering notably better performance at more competitive prices, with next-generation designs that appeal to a wide range of employees. At the end of the day, they're hardware and software certified, ensuring both IT and end users a significant level of reliability.

Q. The workstation market is growing fast. Is there a connection between this growth and companies' ongoing need to do local processing?

A. While the broader PC market has experienced negative to flat growth in recent years, the workstation market enjoyed double-digit year-over-year growth in 2017 and 2018. IDC's Workstation Tracker expects the growth to continue through 2019, topping out at over 11% over the previous year. One of the fastest-growing application use cases for the workstation is scientific workloads, which grew 34% in 2018 and is forecast to grow another 26% at the worldwide level in 2019. While the workstation market is growing in the double digits from a shipment perspective, average selling prices continue to decline.

IDC's recent survey of ITDMs showed that about 8% of companies currently have workstations installed, but a whopping 64% of all companies said they expect to expand their workstation deployments in the near term. When asked what was driving the expansion, more than 45% of respondents said it was related to data science/machine learning/artificial intelligence workloads. We expect more companies to adopt workstations in the coming years as 20% of ITDMs indicated they would make inaugural deployments in the future.
Q. When a company begins exploring workstation options, does it want a full solution or just the hardware?

A. Each company has its own IT environment; some will have workstation experts on staff, while many will not. Even organizations that have experience supporting workstations may not have the right expertise required to configure, acquire, deploy, and support the right hardware to support the applications needed for data science workloads — for example, applications such as Google’s TensorFlow, an open source library that enables machine learning; and NVIDIA’s Deep Learning GPU Training System, which enables the training of deep neural networks (DNNs) and NVIDIA’s RAPIDS suite of software libraries, built on CUDA-X AI, which gives data scientists the freedom to execute end-to-end data science and analytics pipelines entirely on GPUs.

If an organization lacks the skill set to support these types of applications ad hoc, then a full solution will often make a great deal of sense. Such solutions align software requirements with hardware capabilities, ensuring the system utilizes the right type of CPU and GPU, offers adequate RAM and storage as well as the required networking and expansion capabilities, and supports NVIDIA GPU-accelerated databases. After all, there’s no point in taking a step toward enabling your data scientists with a high-end workstation only to find out it doesn’t meet their specific requirements.

About the Analyst

Tom Mainelli, Group Vice President, Devices & AR/VR

Tom Mainelli manages IDC’s Devices group, which covers a broad range of hardware categories, including PCs, tablets, smartphones, wearables, AR/VR, thin clients, and displays. In his role as group vice president, he works closely with company representatives, industry contacts, and other IDC analysts to provide in-depth insight and analysis on the always-evolving market of endpoint devices and their related services.
About Z by HP

The Z by HP Data Science Solution, powered by NVIDIA® Quadro RTX™ GPUs, provides Data Science teams with unprecedented processing capabilities in an on-premise, single-box solution that combines powerful GPU + CPU compute with the NVIDIA GPU-accelerated data science AI software stack. From data wrangling to artificial intelligence to neural network implementations, Z is the affordable, future-ready solution with optional cloud connectivity to solve data challenges and enable Data Science teams to deliver the insights needed to move business forward.

Learn more