



Hewlett Packard Enterprise

PUBLIC VERSION

October 11, 2024

Re: Comments on the Establishment of Reporting Requirements for the Development of Advanced Artificial Intelligence Models and Computing Clusters, Docket No. BIS-2024-0047, RIN 0694-AJ55

Hewlett Packard Enterprise (HPE) appreciates the opportunity to comment on the proposed **Establishment of Reporting Requirements for the Development of Advanced Artificial Intelligence Models and Computing Clusters**. HPE is committed to the responsible development and deployment of AI. We are also committed to our partnerships across the U.S. Government to protect our national and economic security and to advancing innovation that can help solve our nation's greatest challenges, including national defense.

HPE is a global market leader in compute and supercomputing. The latest [TOP500](#) list which verifies and ranks the world's most powerful supercomputers revealed that four of the top 10 systems were built by HPE and the top two of them are exascale machines (Frontier at Oak Ridge National Lab and Aurora at Argonne National Lab). As a leader in the compute infrastructure that powers AI, we are deeply interested in working with BIS to address national security concerns arising from AI and advanced computing.

Definitions

“Acquiring, developing, or coming into possession”:

The proposed rule lists applicable activities in 2(a)(ii) as “acquiring, developing, or coming into possession of a computing cluster” but does not define the terms acquiring, developing, or coming into possession. This creates some ambiguity regarding their scope as well as with respect to the intended impact on those who manufacture, integrate, and sell large-scale computing clusters. Are sellers required to report such sales, or is the rule intended to obligate only the purchasers of such systems?

HPE is a leading global provider of high-performance computers (HPC). We have built the fastest and largest systems currently in operation—these leadership systems are below the threshold levels set in EO 14110 for large computing clusters. Eventually, leadership HPC systems will meet the threshold levels, so HPE needs clarification what its obligations are in the role of manufacturer, integrator, and seller of the systems.

While HPE can report sales of leadership HPC systems meeting the thresholds at the time of sale, HPE also sells AI servers for smaller commercial deployments and may not have full visibility into the size of a customer's existing computing cluster and whether the clusters together would meet the thresholds. For example, customers may choose to build computing resources or clusters incrementally over time by buying equipment at different time intervals and from various vendors. Customers can modify hardware by adding GPUs and other components at a later date thus increasing the size of the cluster or system. For these reasons, we recommend that the entities operating these computing clusters be primarily responsible for reporting clusters that meet the set thresholds. This will minimize the duplicative, redundant, and/or potentially irrelevant

reporting of commercially sensitive information. Entities operating these clusters will have more accurate visibility and information on the amount of total computing power available in each cluster.

Nevertheless, if BIS desires to have some redundancy and overlap in reporting requirements between entities, then HPE proposes clearly defining a manufacturer/integrator/seller's obligation with respect to this rule. Accordingly, HPE proposes defining a role of "Seller" within the current framework of "acquiring, developing, or coming into possession of" and limiting its reporting obligation to systems that meet the thresholds for a large-scale computing cluster at the time of sale. Additionally, such obligation should be clearly limited to sales to U.S. covered persons and should exclude sales to the U.S. government.

HPE proposes to add a new defined term "Seller" and modify section 702.7(a)(ii) as follows:

Seller means a Covered U.S. person who acquires, develops, or comes into possession of a Large-scale computing cluster for the purpose of selling the cluster to a Covered U.S. person. The U.S. government and agencies thereof are not considered Covered U.S. persons in this context.

§ 702.7 Special requirements for on-going reporting regarding the development of advanced artificial intelligence models and computing clusters.

(a) *Reporting requirements.* (1) Covered U.S. persons are required to submit a notification to the Department by emailing ai_reporting@bis.doc.gov on a quarterly basis as defined in paragraph (a)(2) of this section if the covered U.S. person engages in, or plans, within six months, to engage in 'applicable activities,' defined as follows:

- i. Conducting any AI model training run using more than 10^{26} computational operations (e.g., integer or floating-point operations); or
- ii. Acquiring, developing, or coming into possession of a computing cluster that has a set of machines transitively connected by data center networking of greater than 300 Gbit/s¹ and having a theoretical maximum greater than 10^{20} computational operations (e.g., integer or floating point operations) per second (OP/s) for AI training, without sparsity; wherein a Seller of such a computing cluster is only required to report the sale thereof when made to a Covered U.S. person and the computing cluster being sold meets the requirements for a Large-scale computing cluster at the time of sale.

Collection Thresholds

Large computing cluster:

HPE suggests amending the definition of large computing cluster to clarify how "transitively connected" should be calculated, as follows:

"a computing cluster that has a set of machines transitively connected by data center networking of greater than 300 Gbit/s (measured as the ratio between the network bisection bandwidth and the number of machines in the clusters) and having a theoretical maximum greater than 10^{20} computational operations (e.g., integer or floating-point operations) per second (OP/s) for AI training, without sparsity"

¹ Please see next section for suggested clarifications to the technical definition of a computing cluster.

“Transitively connected” is not an industry standard term and by itself does not permit capability comparisons between systems. For accurate and consistent reporting, we recommend that the rule clarify and define the term. Without a definition, the current threshold could capture a broad set of systems because reaching an arbitrarily high bandwidth (certainly above 300 Gbps) for a single flow is relatively easy. For example, a system could be built where the top-level switch only sustains 300 Gbps and hits the threshold.

HPE recommends that the rule define transitively connected as bisection bandwidth divided by end points or number of nodes. This would capture cumulative bandwidth for all the servers in a cluster, from different partitions, to communicate with each other at the same time. This bisection bandwidth is what differentiates a leadership system from a commodity cluster. Moreover, bisection bandwidth/number of nodes is an easy ratio to measure (both numbers are part of the specifications of large systems), and captures the cumulative network speed, which is key to AI training performance.

Technical updates:

The proposed rule states that BIS will update technical conditions that trigger reporting requirements as appropriate. HPE urges BIS to set a cadence for updates to the technical parameters in the final rule. If BIS does not set a process from the outset, there is a risk that the technical parameters will not keep up with the pace of innovation in this field.

Annual re-evaluations and updates would be reasonable to keep up with innovation in processor architecture and training methodologies. Further, an annual review and update of technical parameters that trigger reporting requirements would help ensure that the reporting remains focused on the most advanced frontier models and leadership systems as mandated in EO 14110. BIS could drive the annual review and update by leveraging the information reported in the preceding reporting periods, explicitly assigning an advisory role to the existing Technical Advisory Committee (e.g., ISTAC), and/or creating a specific Technical Advisory Committee for these issues.

Notification Schedule

The time to build, deliver, and deploy commercial small-medium size HPC clusters is typically [], but these systems are well below the rule’s threshold. For leadership class systems with increased size and complexity that approach the rule’s threshold (like the Frontier and Aurora systems deployed at U.S. DOE Labs), this deployment time increases []. Furthermore, building the next generation of leadership class supercomputers is approximately [], and HPE would know the technical specifications when a contract is awarded, which is []. These cycles also tend to be aligned to major tech transitions (e.g., new PCIe or DRAM standard generation) so they tend to be more or less synchronized. Accordingly, HPE submits that a quarterly reporting requirement is not needed for BIS to receive early notification of such systems. HPE proposes that a six month or annual reporting requirement would meet BIS’s objective.

Collection and Storage

As the NPRM acknowledges, the information sought under the rule is proprietary and extremely sensitive. For both the manufacturers of large compute clusters and for the entities acquiring them, their existence, timing, and location is confidential and competitive information. As such, information reported to BIS should be adequately protected and access to it should be restricted. HPE recommends that the final rule specify how reported information will be protected and clarify that information provided will be exempt from Freedom of Information Act requests.

COMPANY CERTIFICATION

I, {Maria Medrano}, acting as {Director of Corporate Affairs}, for Hewlett Packard Enterprise Co., certify that I prepared or otherwise supervised the preparation of the attached submission entitled “**Comments on the Establishment of Reporting Requirements for the Development of Advanced Artificial Intelligence Models and Computing Clusters**” filed on **October 11, 2024**, in **Reporting Requirements for the Development of Advanced Artificial Intelligence Models and Computing Clusters**, Docket No. BIS-2024-0047, RIN 0694-AJ55. The bracketed information in the business confidential version of the submission, and redacted from the public version, constitutes confidential trade secrets and commercial or financial information which is exempt from disclosure under 5 U.S.C. 552(b)(4).



Signature: _____

October 11, 2024

Date: _____