
# New Directions In HPC

October 2022

www.HyperionResearch.com
www.hpcuserforum.com

**Earl Joseph and Bob Sorensen**

HYPERION RESEARCH

# HPC
# High Growth Areas

# High Growth Areas

***These are redefining the HPC sector***

- **The use of external clouds for running HPC workloads**
- **AI, ML and DL**
- **New processor types and accelerators/GPUs**
- **Storage**
- **Exascale-class systems**
- **Quantum technologies**

# The HPC Cloud Market Will See Strong Growth in 2022

***The growth will build on the fundamental changes in buying behavior seen in 2021***

- **HPC & AI buyers around the world revealed for the first time that HPC buyers are planning to shift some of their on-premises budgets to spending in the cloud**
  - The shift is fundamental because up to 2021 very few sites were taking money from the on-premises budgets for cloud computing
- **End user spending on public cloud resources to run HPC workloads is projected to grow substantially in 2022, at a rate greater than 23%, and will exceed US $6.2 billion**
- **This major shift in buying behavior doesn't mean that on-premises HPC systems are going away**
  - The on-premises HPC server market is anticipated to exhibit healthy growth, 7%-8% a year, over the forecast period

# The HPC Cloud Market Will See Strong Growth in 2022

***The growth will build on the fundamental changes in buying behavior seen in 2021***

TABLE 2

**How will your plans for using external/public clouds affect the choice of your next on-premises HPC system?**

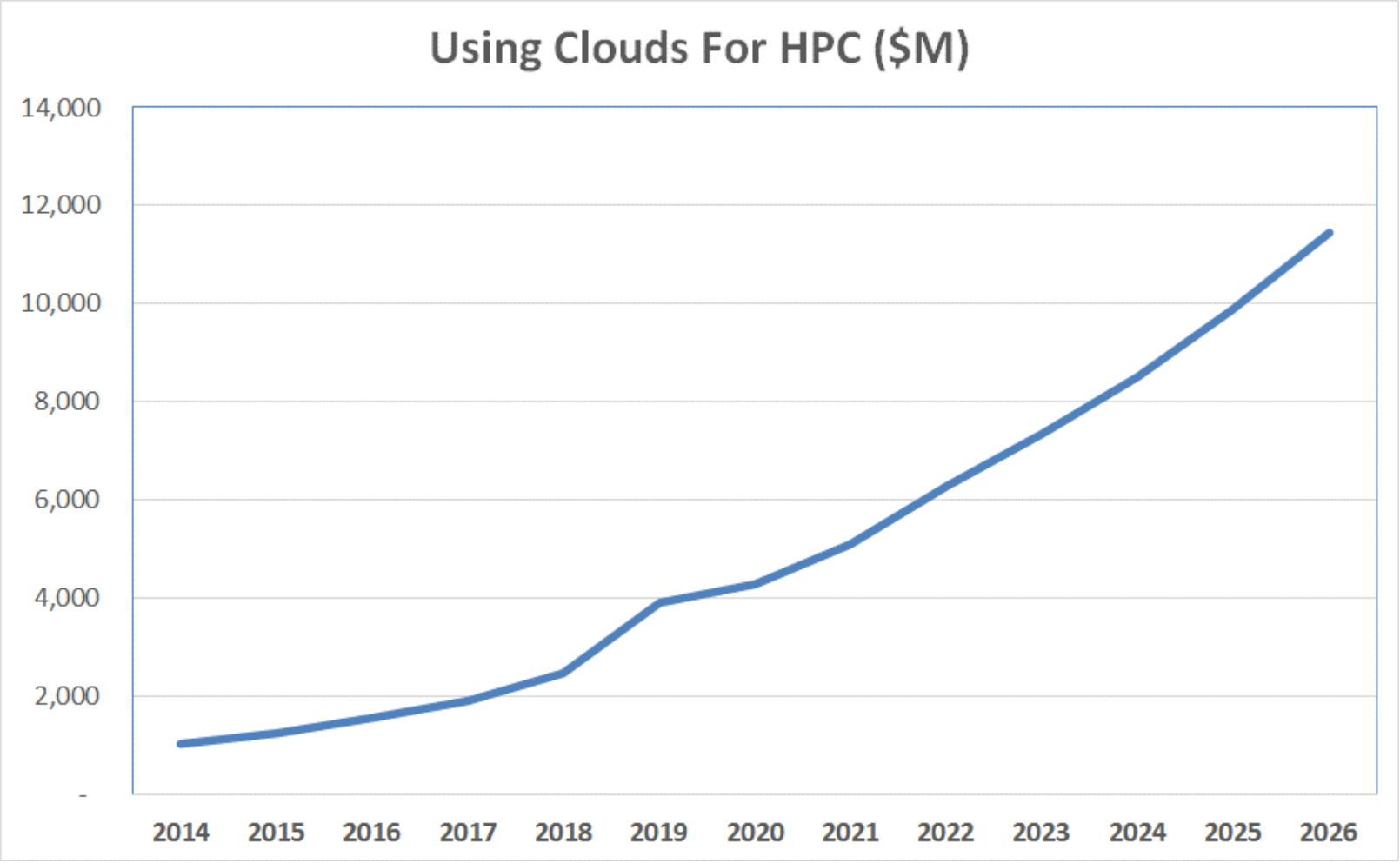
| Question Responses | Responses | Percent |
|---|---|---|
| I will stop buying any on-prem HPC resources | 7 | 5.0% |
| I will buy less on-prem HPC resources and use the extra money in the cloud | 42 | 29.8% |
| I will delay on-prem HPC purchases and use the extra money in the cloud | 19 | 13.5% |
| I will buy a different on-prem HPC system than previously planned because of my cloud usage | 16 | 11.3% |
| None of the above | 46 | 32.6% |
| Not certain/don't know | 32 | 22.7% |

Note: n = 141, multiple responses were allowed.

Source: Hyperion Research, 2022

# HPC Cloud Usage Forecast

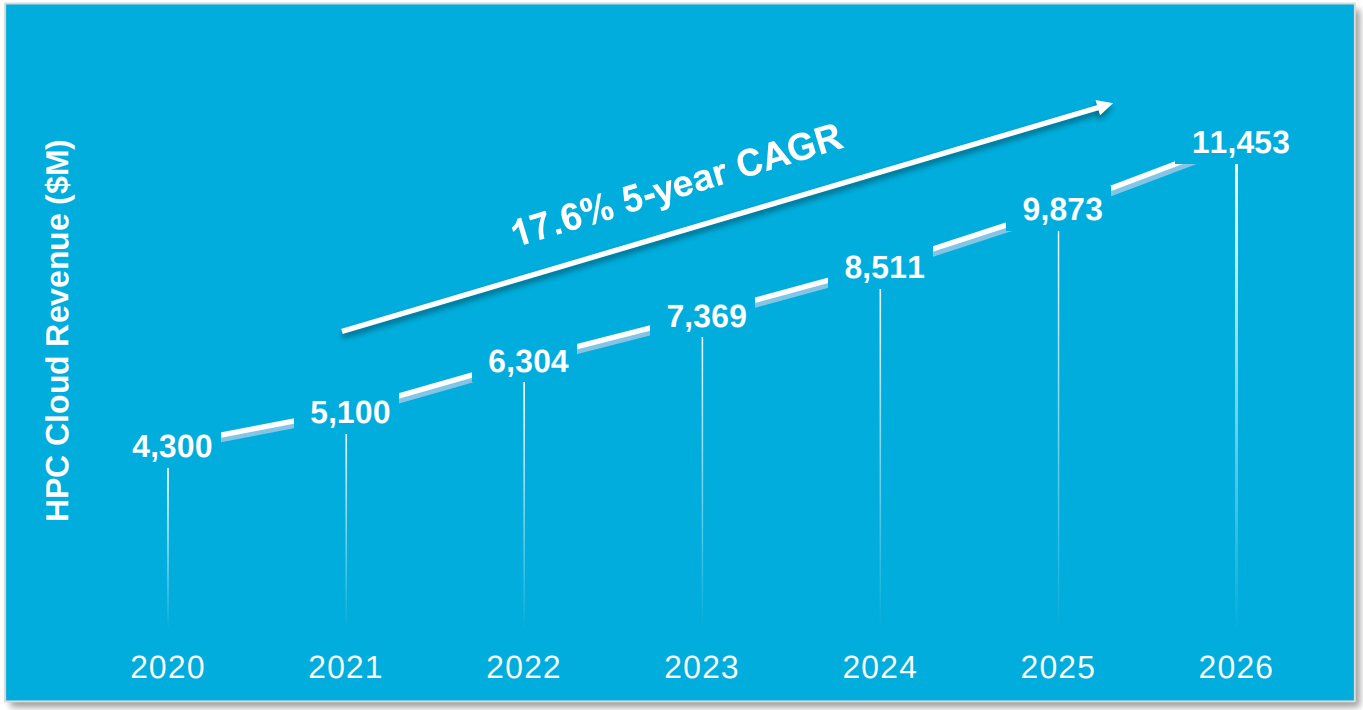
***17.6% growth over the next 5 years***



# HPC Cloud Forecast

***HPC cloud revenue is expected to exceed $11 billion by 2026***

- **Storage-specific components comprise roughly 1/3 of cloud revenue for HPC**
- **AI and other data-intensive applications are a high growth segment for cloud adoption in HPC**



# Use Of Different AI/ML/DL Approaches

## *From our end-user MCS study*

| 36) Which categories will your top AI and/or data-intensive analytics applications fall under in the next 1 to 2 years? | | |
|---|---|---|
| | Responses | Percent |
| Machine learning | 99 | 70.2% |
| Deep learning | 86 | 61.0% |
| Graph analysis | 25 | 17.7% |
| Cognitive computing | 24 | 17.0% |
| Semantic analysis | 22 | 15.6% |
| Other big data/analytics | 41 | 29.1% |
| We don't plan to run applications of these types | 9 | 6.4% |
| n = 141 | | |
| Source: Hyperion Research, 2021 | | |

# HPC-enabled AI Forecast

## *5-year CAGR expected to reach over 22% growth*

| Forecast: Worldwide HPC server revenue breakout by compute-intensive and data-intensive focuses ($M) | 2020 | 2021 | 2022 | 2023 | 2024 | 2025 | 2026 | CAGR 2021-2026 |
|---|---|---|---|---|---|---|---|---|
| Worldwide HPC Server Revenue Forecast | 13,519 | 14,750 | 16,503 | 18,208 | 19,697 | 19,492 | 20,549 | 6.9% |
| Compute-Intensive Server Revenue | 10,020 | 10,848 | 12,103 | 13,280 | 14,177 | 13,586 | 13,993 | 5.2% |
| Data-Intensive Server Revenue | 3,499 | 3,901 | 4,400 | 4,928 | 5,519 | 5,906 | 6,555 | 10.9% |
| HPC-enabled AI (ML, DL & Other) Server Revenue | 1,039 | 1,300 | 1,718 | 2,083 | 2,484 | 2,941 | 3,619 | 22.7% |
| Traditional Data Science (non-AI HPDA) Focused Server Revenue | 2,460 | 2,601 | 2,682 | 2,845 | 3,036 | 2,965 | 2,937 | 2.5% |

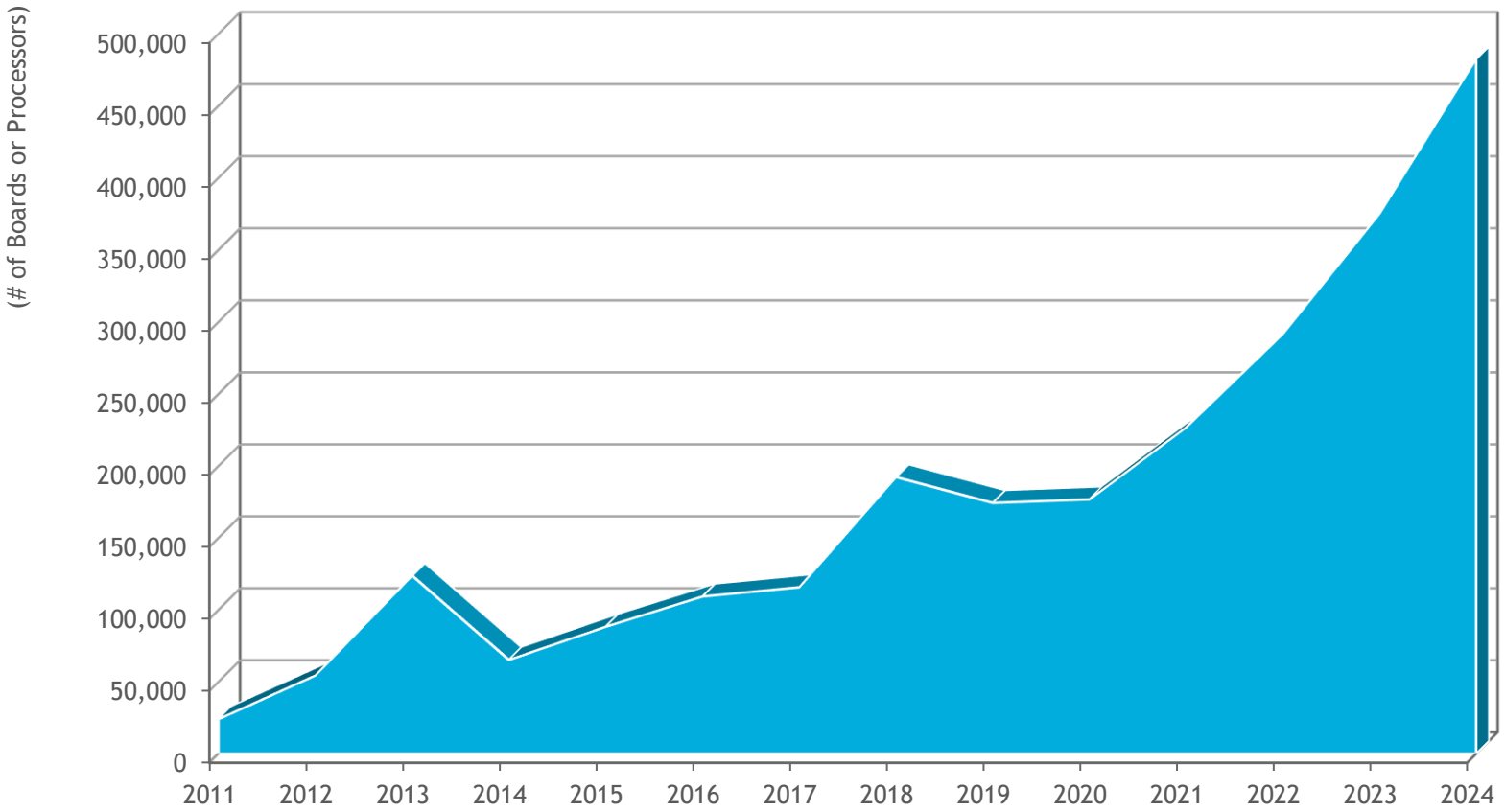
# 83% Of Sites Have Accelerators Or Co-processors Today

## *From our end-user MCS study*

| How many co-processors or accelerators are in your largest HPC technical server? | | |
|---|---|---|
| | **Responses** | **Percent** |
| None | 23 | 17.3% |
| Less than 32 | 28 | 21.1% |
| 32 to less than 64 | 18 | 13.5% |
| 64 to less than 100 | 19 | 14.3% |
| 100 to less than 500 | 18 | 13.5% |
| 500 to less than 1,000 | 11 | 8.3% |
| 1,000 to less than 5,000 | 10 | 7.5% |
| 5,000 to less than 10,000 | 4 | 3.0% |
| 10,000 or more | 2 | 1.5% |
| n = 133 | | |
| *Source: Hyperion Research, 2021* | | |

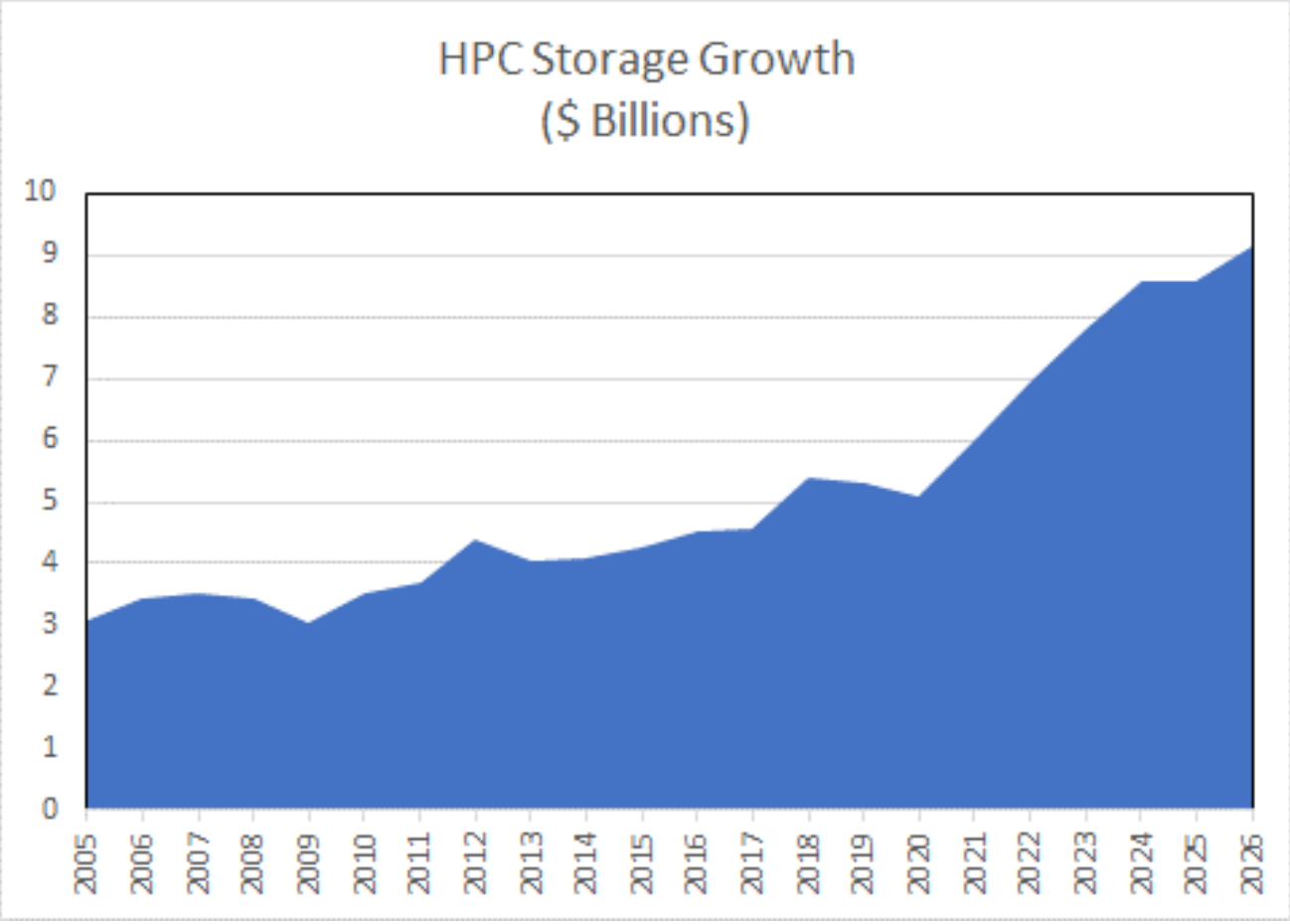
# GPU/Accelerator Forecast

## *Anticipated high growth for accelerators over next 5 years*



# Storage Growth Rates

***HPC storage is growing quickly, driven by AI, big data and growing modeling/simulation model sizes***



# The Exascale Market (System Acceptances)

## *Over 30 systems and over $10 billion in value*

**Exascale and Near-Exascale Leadership Systems (2020 to 2027)**

| Year Accepted | China | Europe | Japan | US | Other Countries* | Total Systems | Total Value |
|---|---|---|---|---|---|---|---|
| **2020** | | | 1 near-exascale system ~$1 B | | | 1 | $1.0B |
| **2021** | 2 exascale ~$350M each | 1 pre-exascale system ~$180M | ? | 1 pre-exascale system ~$200M | -- | 4 | $1.1B |
| **2022** | 1 exascale ~$350M each | 2 pre-exascale systems ~$190 each | 1 near-exascale system ~$150M | 1 exascale systems ~$600M | -- | 5 | $1.5B |
| **2023** | 1 exascale system ~$350M | 1 or 2 pre-exascale systems ~$150M each | 1 near-exascale system ~$150M | 1 or 2 exascale systems ~$600M each | -- | 4-5 | $1.8B - $2.4B |
| **2024** | 1 exascale system ~$350M | 1 exascale ~$500M, plus 1 exascale (or pre) systems ~$200 M | ? | 1 or 2 exascale systems ~$400M each | 1 exascale system ~$200M | 4-6 | $1.2B - $1.9B |
| **2025** | 1 or 2 exascale system ~$300M each | 1 or 2 exascale systems ~$350M each | 1 exascale system ~$150M | 1 or 2 exascale systems ~$350M each | 1 exascale system ~$150M | 5-8 | $1.3B - $2.3B |
| **2026** | 1 or 2 exascale system ~$300M each | 1 or 2 exascale systems ~$325M each | ? | 1 or 2 exascale systems ~$350M each | 1 or 2 exascale systems ~$150M each | 4-8 | $1.1B - $2.2B |
| **2027** | 1 or 2 exascale systems ~$250M each | 1 or 2 exascale systems ~$300M | ? | 1 or 2 exascale systems ~$300M each | 1 or 2 exascale systems ~$150M each | 4-8 | $1.0B - $2.0B |
| **Total** | **8-11** | **8-12** | **4** | **7-12** | **4-6** | **31-45** | **$10B - $14B** |

* Includes S. Korea, Singapore, Australia, Russia, Canada, India, Israel, Saudi Arabia, etc.

*Source: Hyperion Research, July 2022*

# China Exascale Prototypes

## *In 2019/2020*

- Sunway prototype
  - Manycore-based system optimized for multiple objectives
  - Processor composed of big cores and small cores, organized in groups, **heterogeneity on chip**
- Sugon prototype
  - Heterogeneous accelerated architecture using Hygon x86 processors and Hygon DCU accelerators
  - Node composed of CPU and GPU, **heterogeneity on node**
- Tianhe prototype
  - Flexible architecture to meet the requirement of different applications
  - Software defined configurations, **heterogeneity in system**


# China Exascale Status

*The official situation*

- **Sunway Pro OceanLight**
  - ~1.3 EFlops Rpeak, ~1.05 EFlops Rmax
  - 35 MW, 38 million cores
  - ShenWei post-Alpha CPU
  - National Supercomputing Center-Wuxi
- **Tianhe-3**
  - Dual-chip FeiTeng ARM and matrix accelerator nodes
  - ~ 1.7 EFlops Rpeak, 1.3 EFlops Rmax
  - NSCC-Tianjin
- **Sugon**
  - Hygon processors (low confidence), may go AMD Zen4
  - NSCC-Shenzhen

# China Exascale Status

*The unofficial reality?*

- **Sunway Pro OceanLight**
  - Up and running since March 2021
- **Tianhe-3**
  - Up and running in last six (?) months
- **Sugon**
  - Potentially delayed
- **No official announcements**
- **No entries for June 2021, November 2021 Top 500 list**
  - Maybe this time around…will find out soon
  - Some hesitancy by Chinese leadership
- **Strong evidence of at least five other Chinese systems that could make top 10 list today**

HYPERION RESEARCH

# Quantum Computing Trends

# Quantum Computing Research, Development, And Commercial Activity Occurring On A Global Scale

- **Quantum computing is considered by most major nations to be critical to both national defense and economic security, and they have acted on that realization with technical and financial support**
- **Different programs stress different priorities on the military-civil spectrum**
  - Most, however, see the need to secure a domestic QC supply chain
- **The commercial quantum computer sector is large, diverse, and growing**
  - Aspiring developers range from large players such as IBM and Quantinuum to small start ups with zero revenues to date
  - There are at least 44 different QC commercial hardware developers vying for market share

# End Users Already Actively Pursuing QC As A Solution To Their Most Vexing Computational Problems

- **At the corporate level, QC is seen as the latest accelerator for advanced computing workloads**
  - But most are concerned about cost, complexities with integration, and a lack of in-house QC experts
- **Of the nineteen commercial verticals considered to assess interest in quantum computing, spanning advanced manufacturing, aerospace, chemical, defense, healthcare, insurance and pharmaceuticals, and telecommunication, each were already involved in some level of QC activity**
  - Industry wide, on-going efforts ranged from entry-level activities such as exploring options and monitoring technology development to the use of QC technology for one of more business processes

# QC Five Years Out

***The sector is in transition and there is considerable uncertainty across a broad spectrum of issues***

- **Some demonstration of quantum advantage soon, but no one knows exactly where and when**
- **At least 12 different QC modalities now**
  - Only a few may still be around in five years
- **The QC sector will seek to transition from a quantum processor to quantum computers**
  - Will require a host of new technologies for quantum networking, quantum memory, quantum/classical interfaces, etc.
- **National level policy will increasingly impinge**
  - Supply chain concerns, funding commitments, export control issues, varying national security and economic agendas
- **Whither the VC community: are they in it for the long term?**

HYPERION RESEARCH

# Interesting Trends From Our Recent Studies

# HPC Users Are Optimistic About Budgets, Prioritizing Hardware and Storage

- **Most surveyed users (52%) expect HPC budgets to increase by at least 5% over the next year**
  - Top categories included HPC hardware, add-on storage, and public cloud
- **Many users reported a willingness to pay a 10-15% premium for their desired system attributes**
  - Top categories included better processors, larger/faster memory, higher performance external I/O and storage, and better density/power/cooling
- **Ethernet is taking share from Infiniband**

# Other Priorities: Acceleration, Data Locality and HPC/AI Expertise

- **Many sites report using accelerators on their most used or most important application**
- **Most cloud compute instances (64%) are reserved**
  - And many (43%) are accelerated
  - A quarter of overall public cloud spending in HPC (23%) is persistent storage
- **Top limitations for increasing public cloud use are costs and data locality/speed**
  - Among Industry sites, data security is #1 concern
- **Expertise is a top concern for both HPC and AI, outranked only by budget concerns**

# Emerging Technologies Are Gaining Traction: AI, Edge Computing, And Composable Infrastructures

- **Most HPC sites plan to use AI methodologies**
- **Most users expect to employ edge computing within 2 years**
  - Top motivators include improving real-time data collection/processing, accelerating HPC applications, access to IoT devices for data collection, and a wider range of sensor data
- **Composable infrastructures are gaining a lot of interest and attention, as are DPUs**

HYPERION RESEARCH

# Questions?

**We welcome questions, comments and suggestions**

**Please contact us at:**
**info@hyperionres.com**