Hyperion Research HPC Market Update

November 2022

Earl Joseph, Bob Sorensen, Alex Norton,
Mark Nossokoff, Melissa Riddle,
Tom Sorensen and Jaclyn Ludema

www.HyperionResearch.com
www.hpcuserforum.com

About Hyperion Research

(www.HyperionResearch.com & www.HPCUserForum.com)

Hyperion Research mission:

- *Hyperion Research helps organizations make effective decisions and seize growth opportunities*
  - *By providing research and recommendations in high performance computing and emerging technology areas*

HPC User Forum mission:

- *To improve the health of the HPC/AI/QC industry*
  - *Through open discussions, information sharing and initiatives involving HPC users in industry, government and academia along with HPC vendors and other interested parties*

The Hyperion Research Team

Analysts

- Earl Joseph, CEO
- Bob Sorensen, SVP Research
- Mark Nossokoff, Research Director
- Alex Norton, Research Director
- Jaclyn Ludema, Analyst
- Melissa Riddle, Associate Analyst
- Thomas Sorensen, Associate Analyst
- Cary Sudan, Survey Specialist

Operations

- Jean Sorensen, COO

International Consultants

- Katsuya Nishi, Japan and Asia
- Jie Wu, China & Technology Trends

Global Accounts

- Rene Copeland, Dir. Business Development
- Mike Thorp, Sr. Global Sales Executive
- Kurt Gantrish, Sr. Account Executive

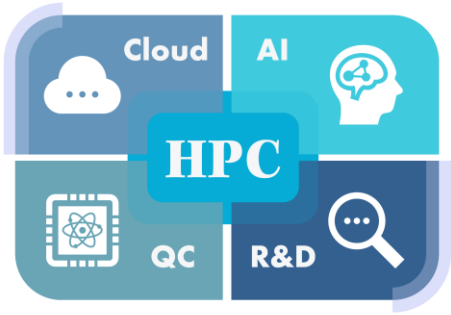
Data Collection

- Sue Sudan, Market Data Group
- Kirsten Chapman, KC Associates

Example Research Areas

(www.HyperionResearch.com & www.HPCUserForum.com)

- Traditional HPC
- AI, ML, DL, Graph
- Cloud Computing
- Storage & Data
- Interconnects
- Software & Applications
- Power & Cooling
- Tracking all Processor Types & Growth rates
- Quantum Computing
- R&D and Engineering -- all types
- Edge Computing
- Supply Chain Issues



Agenda

- HPC Market Update
- A Quick Update on Exascale Systems
- HPC in the Cloud Update
- Storage and Interconnects Update
- Multi-Client Study Results
- HPC Applications Update
- Global AI Update
- The Quantum Computing Sector
- HPC Innovation Awards
- Conclusions

HYPERION RESEARCH

HPC Market Update

Top Trends in HPC

The first half of 2022 was soft due to supply chain issues – $6.6 billion (US$) in revenues

- We still expect the second half of the year to do better
- 2021 grew by a strong 9%
- Supercomputers are growing better at 5.3% (1H2022)

There are many high growth areas:

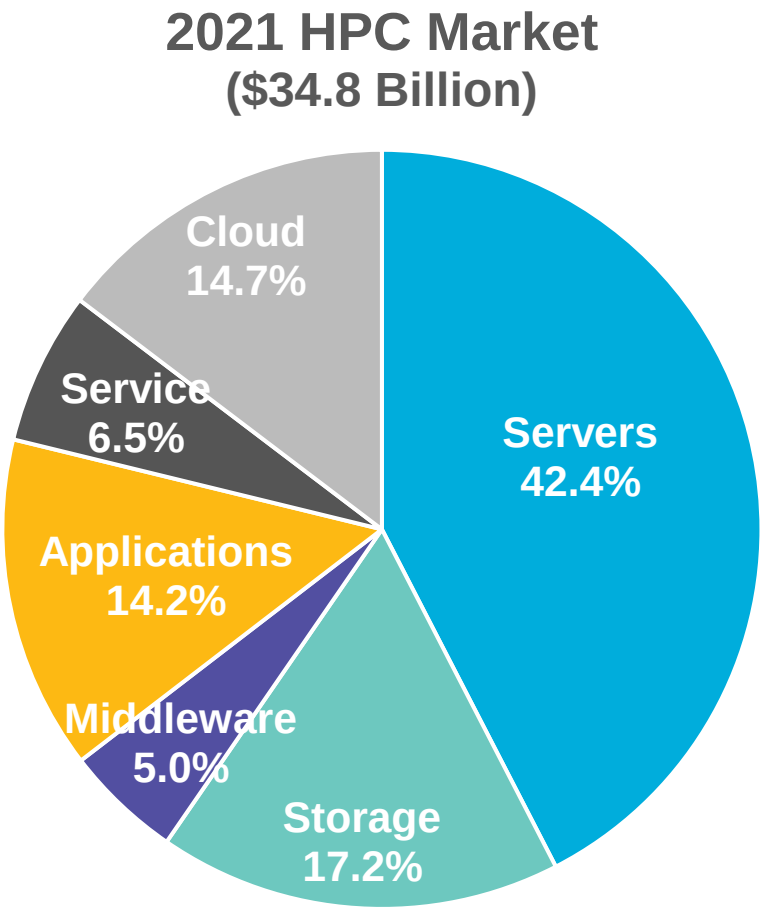
- AI, ML and DL
- Using clouds to run HPC workloads
- GPUs
- Storage

Software continues to be a major area that needs improvement

The lack of technical experts is creating a major constraint

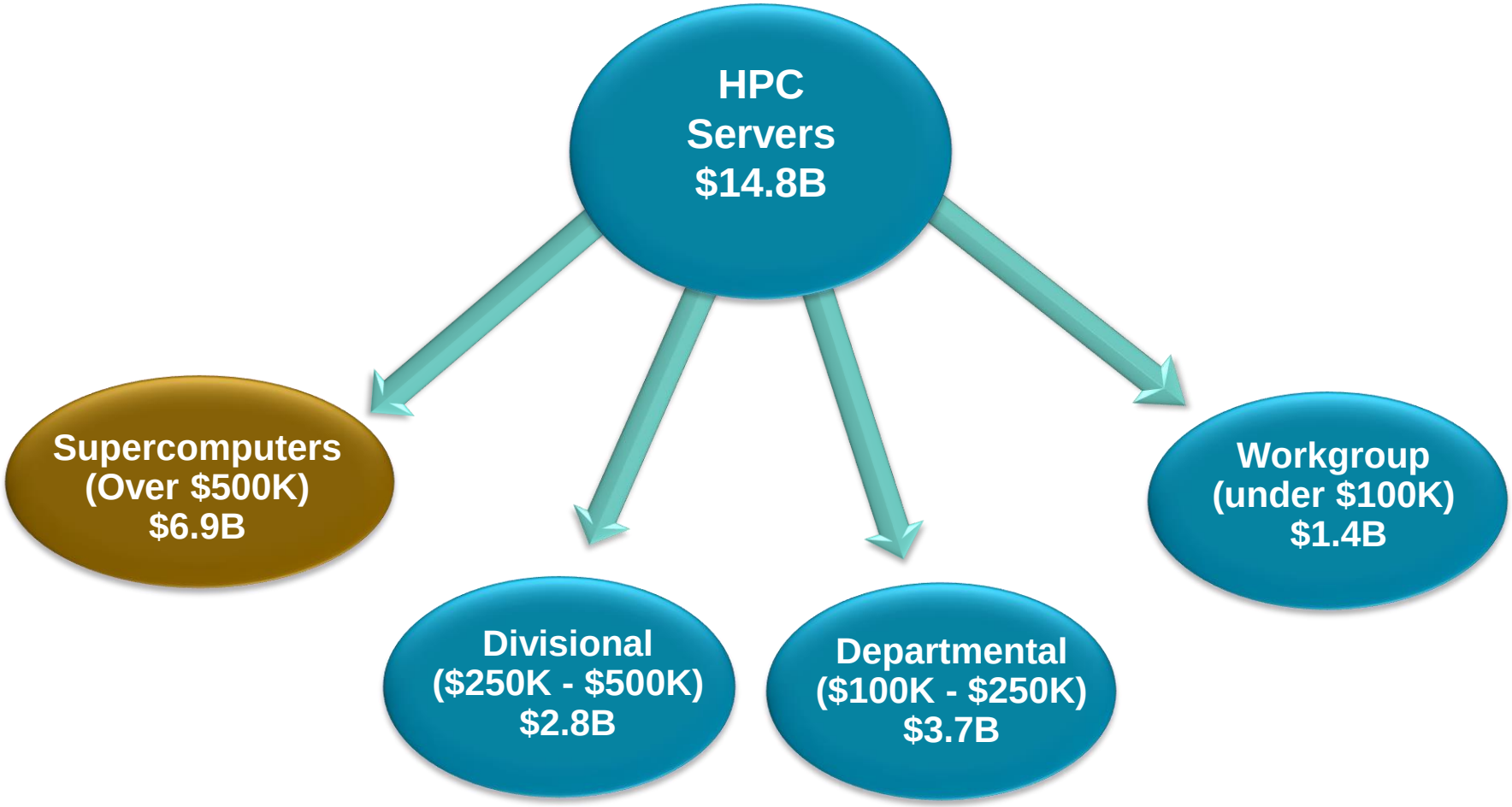
The Overall HPC Market in 2021

Looking at the overall HPC market, including servers, cloud usage, storage, software and repair services = $34.8 billion USD



The 2021 Worldwide On-Prem HPC Server Market: $14.8 Billion (up 9%)

2022 is projected to reach $16 to $17 Billion



WW HPC On-Prem Market By Vendor

($ Millions)

2021 On-Prem server Revenues By Vendor ($M)		
Vendor	**Server Revenues ($M)**	**Market Share**
HPE	5,050	34.2%
Dell Technologies	3,213	21.8%
Lenovo	1,174	8.0%
Inspur	993	6.7%
Atos	542	3.7%
Sugon	525	3.6%
IBM	463	3.1%
Penguin	378	2.6%
Fujitsu	176	1.2%
NEC	173	1.2%
Other	2,076	14.1%
Total On-Prem HPC	14,763	100.0%
Source: Hyperion Research, 2022		

WW HPC Market By Vertical ($ Millions)

2021 WW On-Prem High-Performance Systems Revenue by Applications ($M)	
	2021
Bio-Sciences	1,455
CAE	1,767
Chemical Engineering	177
DCC & Distribution	807
Economics/Financial	703
EDA / IT / ISV	849
Geosciences	1,010
Mechanical Design	59
Defense	1,552
Government Lab	2,866
University/Academic	2,637
Weather	681
Other	199
Total Revenue	14,763
Source: Hyperion Research, 2022	

The Broader On-premise Market Areas ($ Millions)

2021 total on-prem HPC spending reached $29.7B (excluding cloud spending, which brings it to $34.8B)

Revenues by the Broader HPC Market Areas			
	2019	2020	2021
Server	13,368	13,523	14,763
Storage	5,288	5,079	5,984
Middleware	1,572	1,491	1,731
Applications	4,569	4,315	4,952
Service	2,181	2,015	2,267
Total Revenue	26,979	26,423	29,697
Source: Hyperion Research, 2022			

HYPERION RESEARCH

HPC Market Forecasts

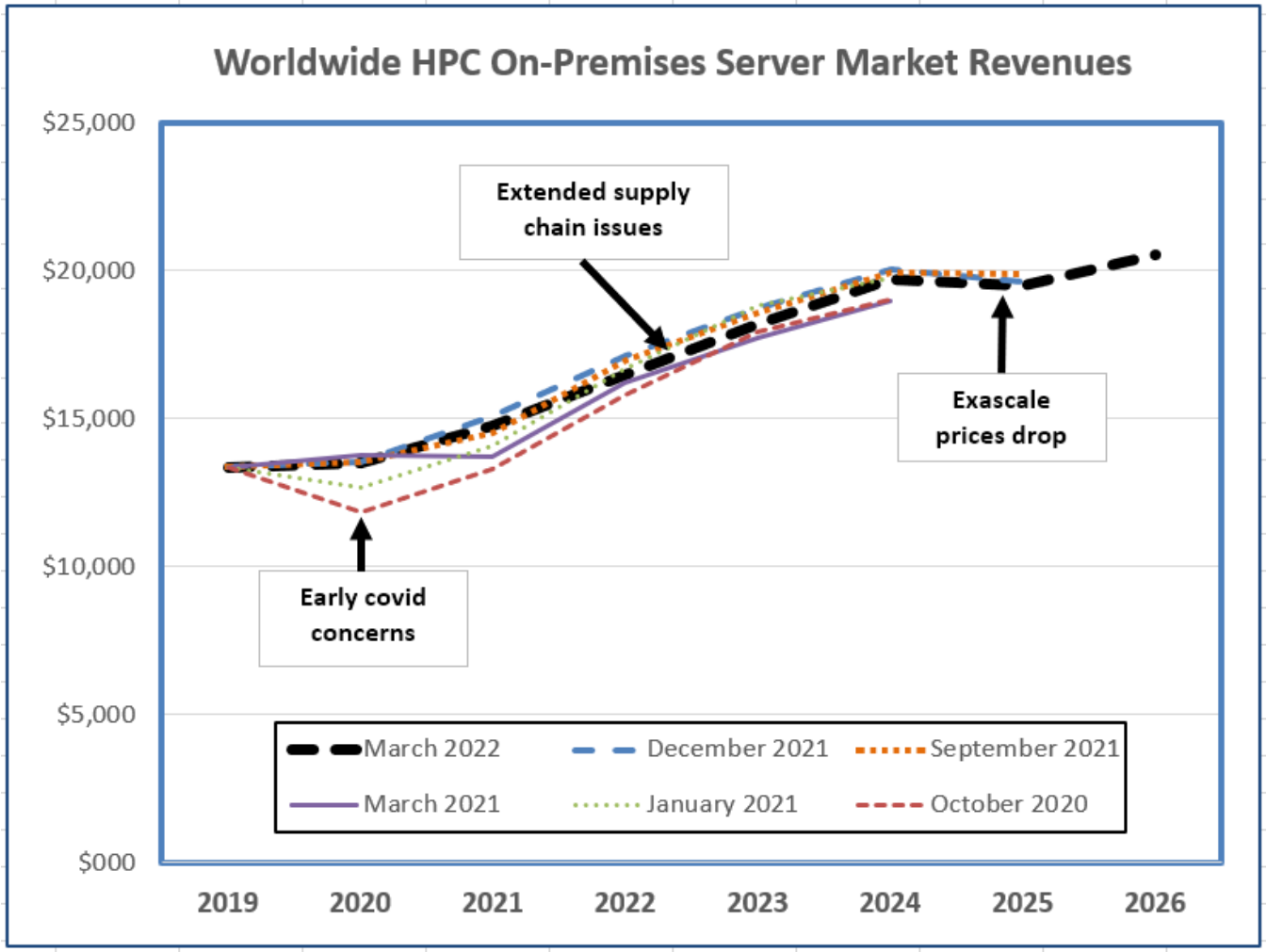
5-Year On-Prem HPC Server Forecast

6.9% growth over the next 5 years

On-Prem HPC Market Forecast by Competitive Segment							
$US Millions	**2021**	**2022**	**2023**	**2024**	**2025**	**2026**	**CAGR 21-26**
Supercomputer	6,926	8,125	9,023	9,846	9,661	9,543	6.6%
Divisional	2,803	3,077	3,397	3,681	3,677	4,397	9.4%
Departmental	3,648	3,909	4,316	4,645	4,704	5,137	7.1%
Workgroup	1,373	1,392	1,472	1,525	1,450	1,471	1.4%
Total	**14,763**	**16,503**	**18,208**	**19,697**	**19,492**	**20,549**	**6.9%**

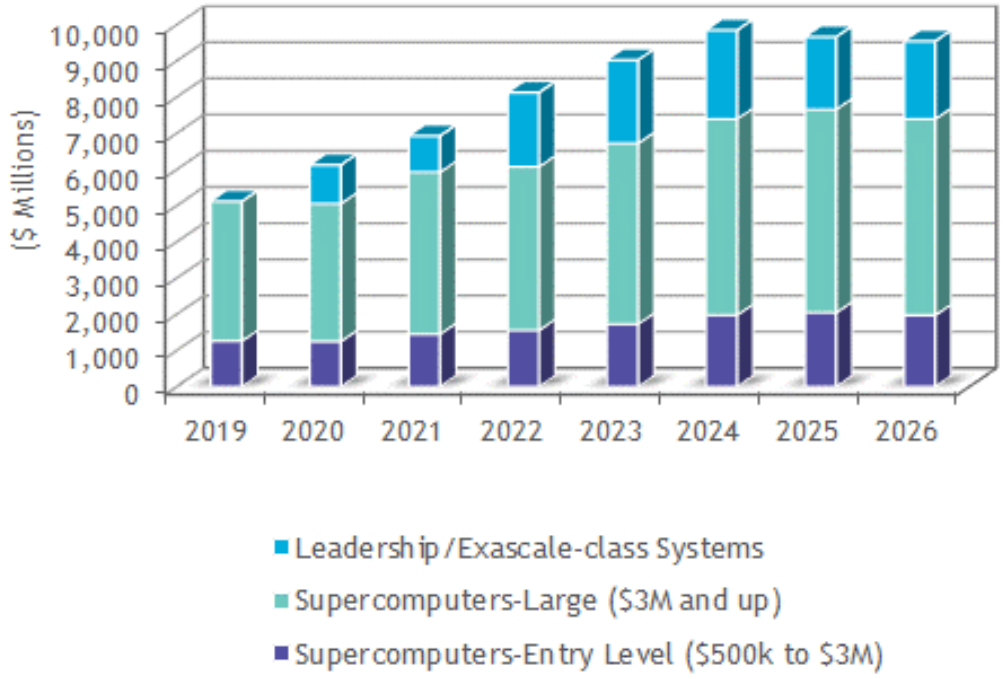
5-year HPC Server Forecast Changes

Quarterly adjustments due to covid



New Supercomputer Subsegments

New Supercomputer Subsegments									
$ Millions									
	2019	2020	2021	2022	2023	2024	2025	2026	CAGR 21-26
Leadership/Exascale-class Systems	0	1,065	1,000	2,050	2,300	2,450	2,000	2,150	16.5%
Supercomputers-Large ($3M and up)	3,858	3,808	4,477	4,516	4,995	5,420	5,615	5,419	3.9%
Supercomputers-Entry Level ($500k to $3M)	1,260	1,250	1,450	1,559	1,728	1,975	2,046	1,975	6.4%
Total Supercomputers ($500K and up)	**5,118**	**6,123**	**6,926**	**8,125**	**9,023**	**9,846**	**9,661**	**9,543**	**6.6%**
Source: Hyperion Research, May 2022									



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**

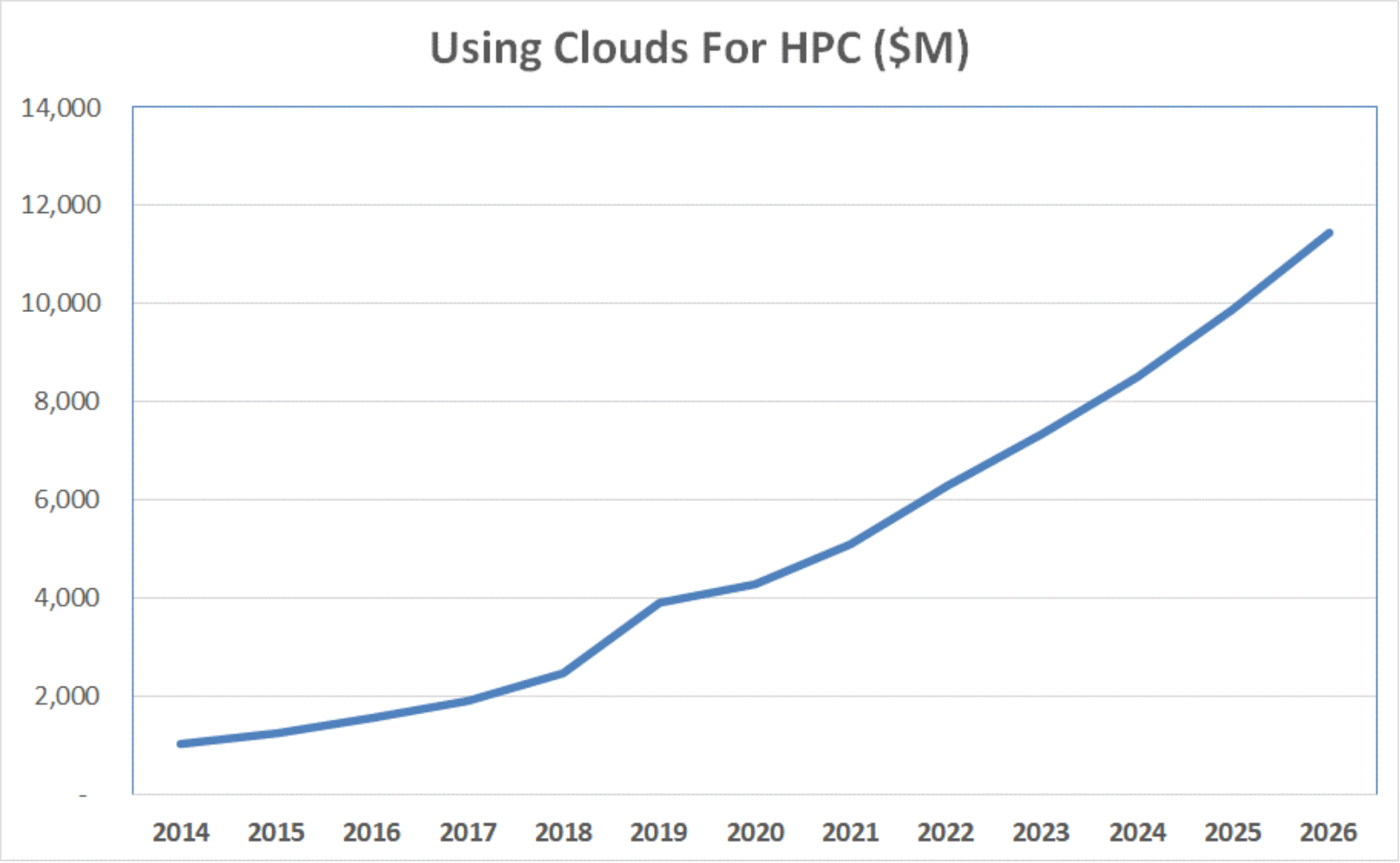
The HPC Cloud Market Will See Strong Growth in 2022

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

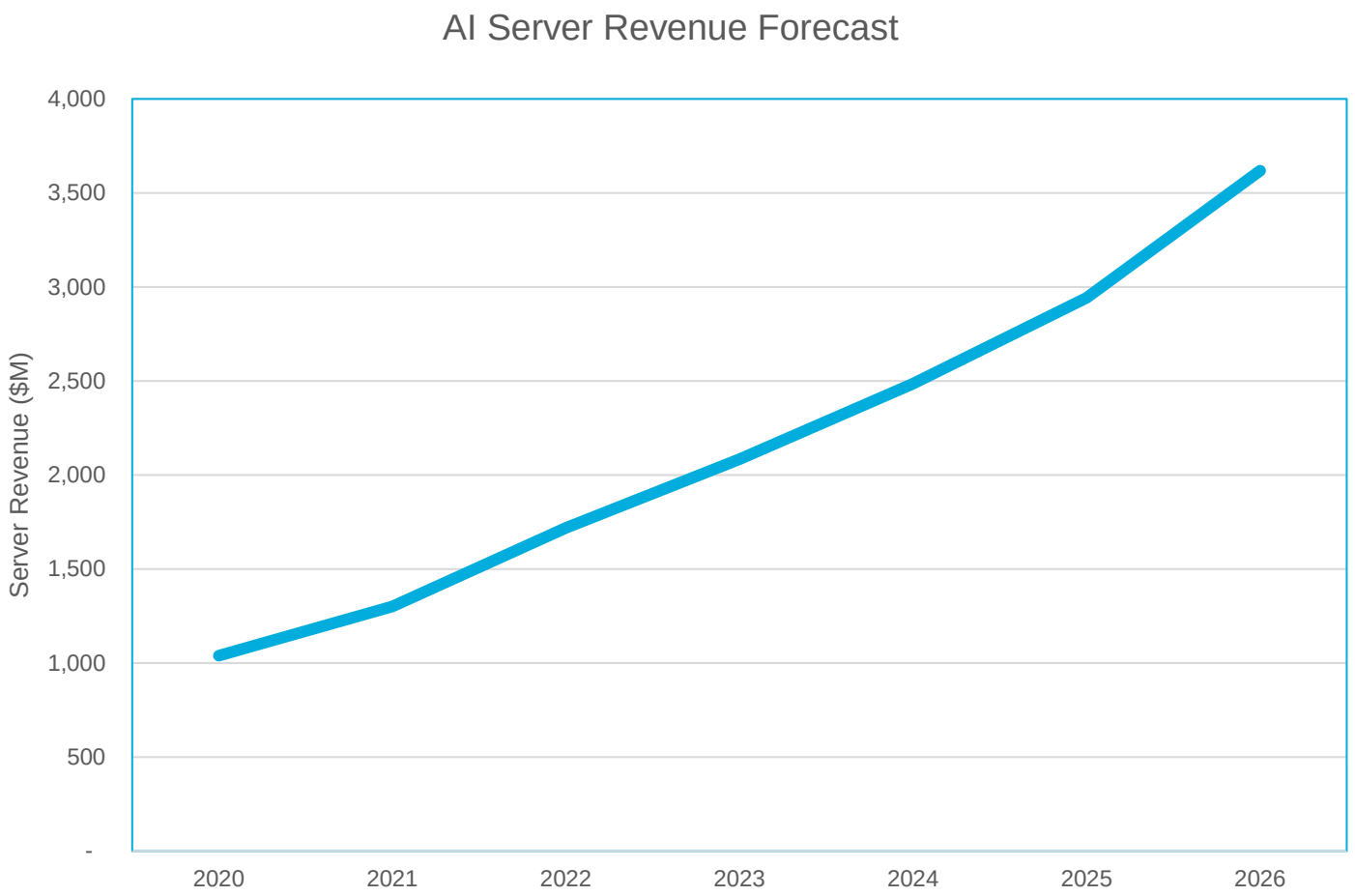
HPC Cloud Usage Forecast

17.6% growth over the next 5 years



HPC-enabled AI Server Forecast

5-year CAGR expected to exceed 22% growth



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		

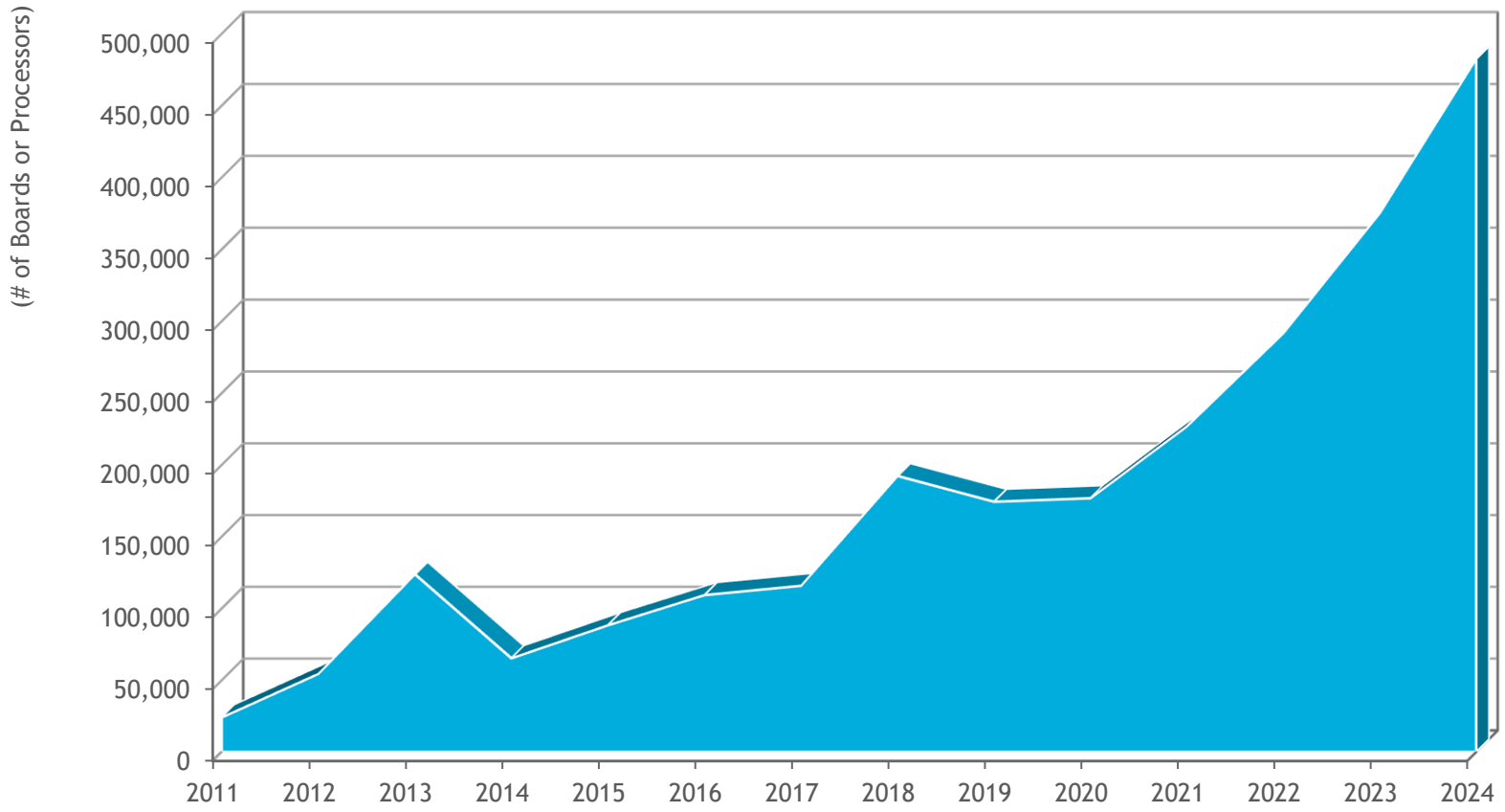
83% Of Sites Have Accelerators Or Co-processors Today

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		

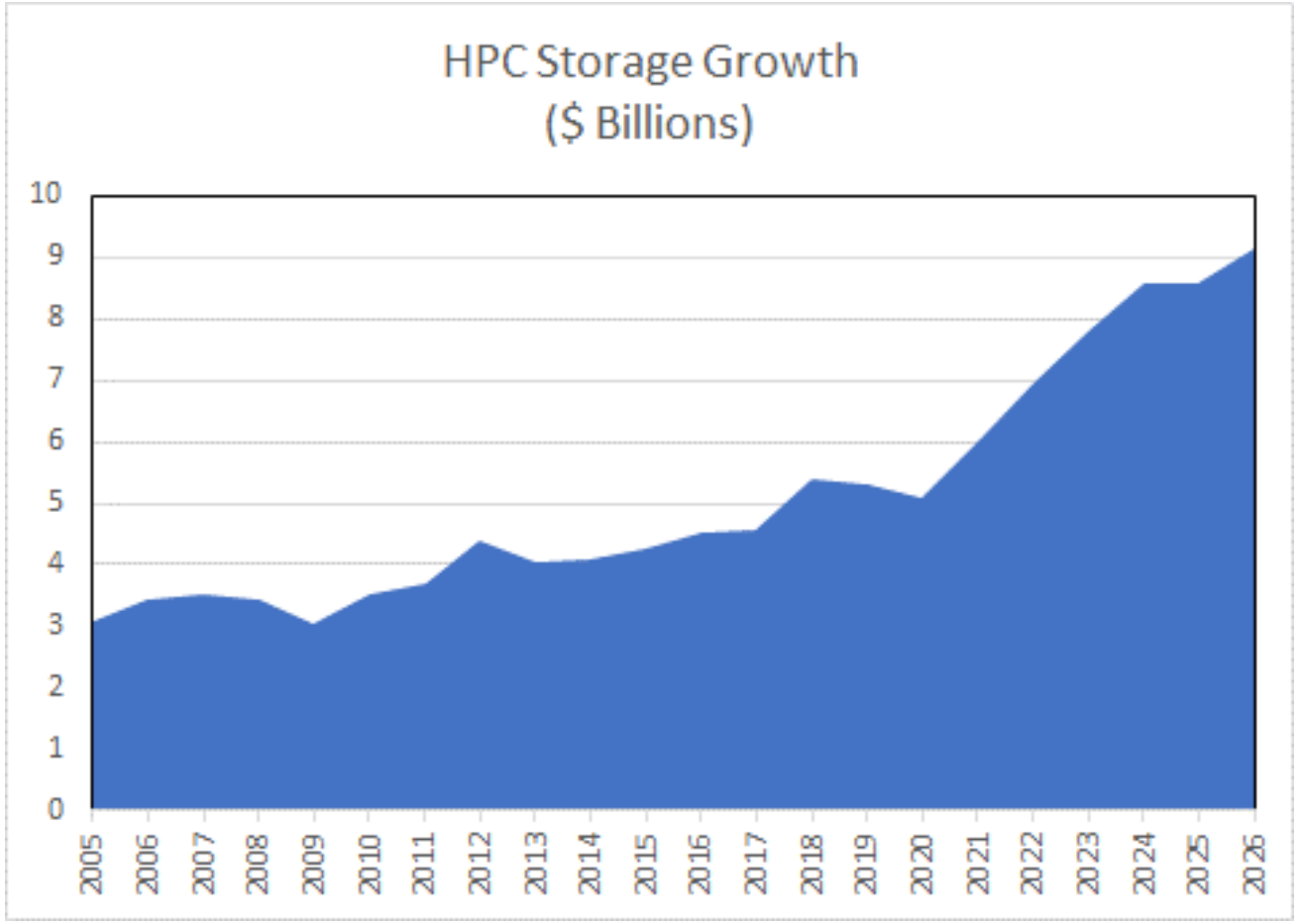
GPU/Accelerator Forecast

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



Agenda

- HPC Market Update
- A Quick Update on Exascale Systems
- HPC in the Cloud Update
- Storage and Interconnects Update
- Multi-Client Study Results
- HPC Applications Update
- Global AI Update
- The Quantum Computing Sector
- HPC Innovation Awards
- Conclusions

HYPERION RESEARCH

Questions?

We welcome questions, comments and suggestions

Please contact us at:
info@hyperionres.com

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

A Quick Update on Exascale Systems

SC 22

Bob Sorensen

www.HyperionResearch.com
www.hpcuserforum.com

Near-Term US Exascale Plans

Three systems over two years with budget of ~ $1.8 billion

- **Frontier: DOE Office of Science: Oak Ridge National Laboratory**
  - First US exascale system in US
  - June Top 500 List: Rpeak = 1.68 EFlops, Rmax = 1.1 EFlops
  - 21 MW to run Linpac
  - Cray Shasta with AMD EPYC CPU and AMD Radeon Instinct GPUs
  - Full user operations January 2023 (some delay)
- **Aurora: DOE Office of Science, Argonne National Laboratory**
  - **60MW**, ~ 1EF DP sustained
  - 08/21: Polaris testbed system (44PF DP and 1.5EF AI)
  - Cray Shasta architecture with Intel Xeons and Intel Xe GPU
  - Delivery in late 2022, acceptance in 2023 (delayed at least 12 months)
- **El Capitan: DOE NNSA's LLNL**
  - ~ 2EF
  - Cray Shasta architecture AMD EPYC processors, next generation Radeon Instinct GPUs
  - Fully deployed in 2023

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

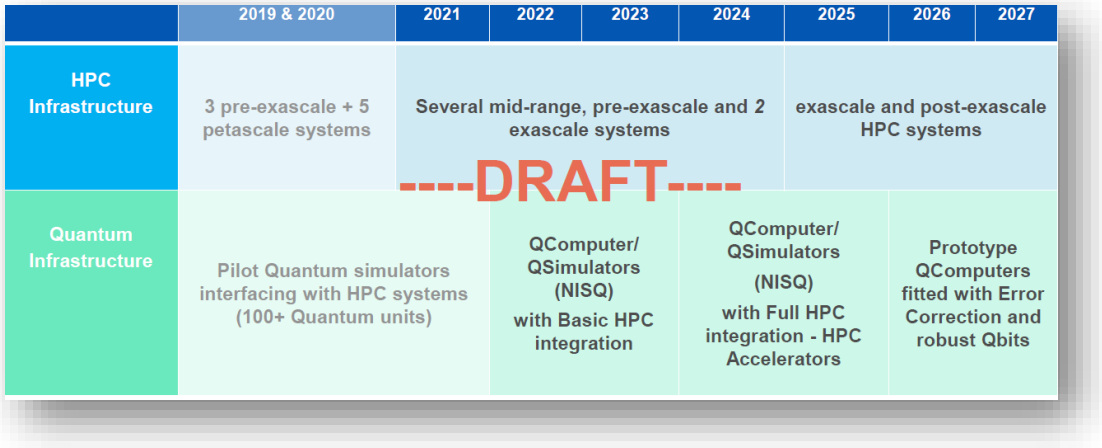
The unofficial reality?

- **Sunway Pro OceanLight**
  - Up and running since March 2021
- **Tianhe-3**
  - Up and running in last year (?)
- **Sugon**
  - Potentially delayed
- **No official announcements**
- **No entries for June 2021, November 2021, June 2022 Top 500 list**
  - Maybe this time around: likely not
  - Political decision
- **Strong evidence of at least five other Chinese systems that could make top 10 list today**

EU HPC Plans

Exascale plans going forward

- **EU plan calls for acquisition of two exascale systems in the 2021-2024 timeframe**
  - At least one to use European technology: specifically using an EPI-developed processor
  - Additional procurements in Germany in 2024, 2025
  - EU may include 2 additional ES systems in 2023-2026
- **Post Exascale System around 2027**
  - Plans call for integration and deployment of the first hybrid HPC/quantum infrastructure in Europe



Source: Leonardo Flores Añover, Senior Expert

DG CNECT, HPC & Quantum Technology Unit - European Commission 2021

SC22: What I'll Be Looking At

- **Mixed precision: not just AI, in traditional mod/sim**
- **Hybrid quantum/classical efforts**
- **Post exascale architectures**
- **Liquid cooling**
- **End of 'end of'**

HYPERION RESEARCH

QUESTIONS?

bsorensen@hyperionres.com

Float to the top or sink to the bottom. Everything in the middle is the churn. -Amos Burton, The Expanse

HPC in the Cloud Update

SC22
November 2022

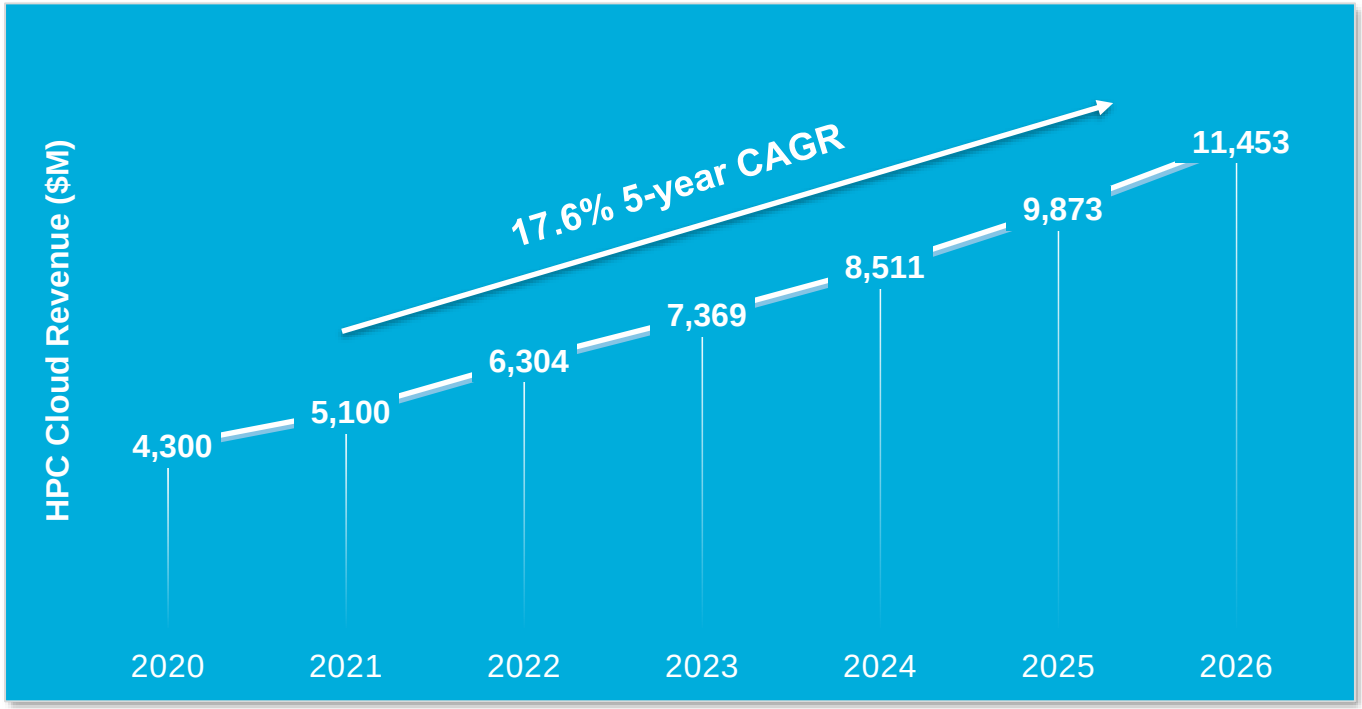
Alex Norton and Mark Nossokoff

www.HyperionResearch.com
www.hpcuserforum.com

HPC Cloud Forecast

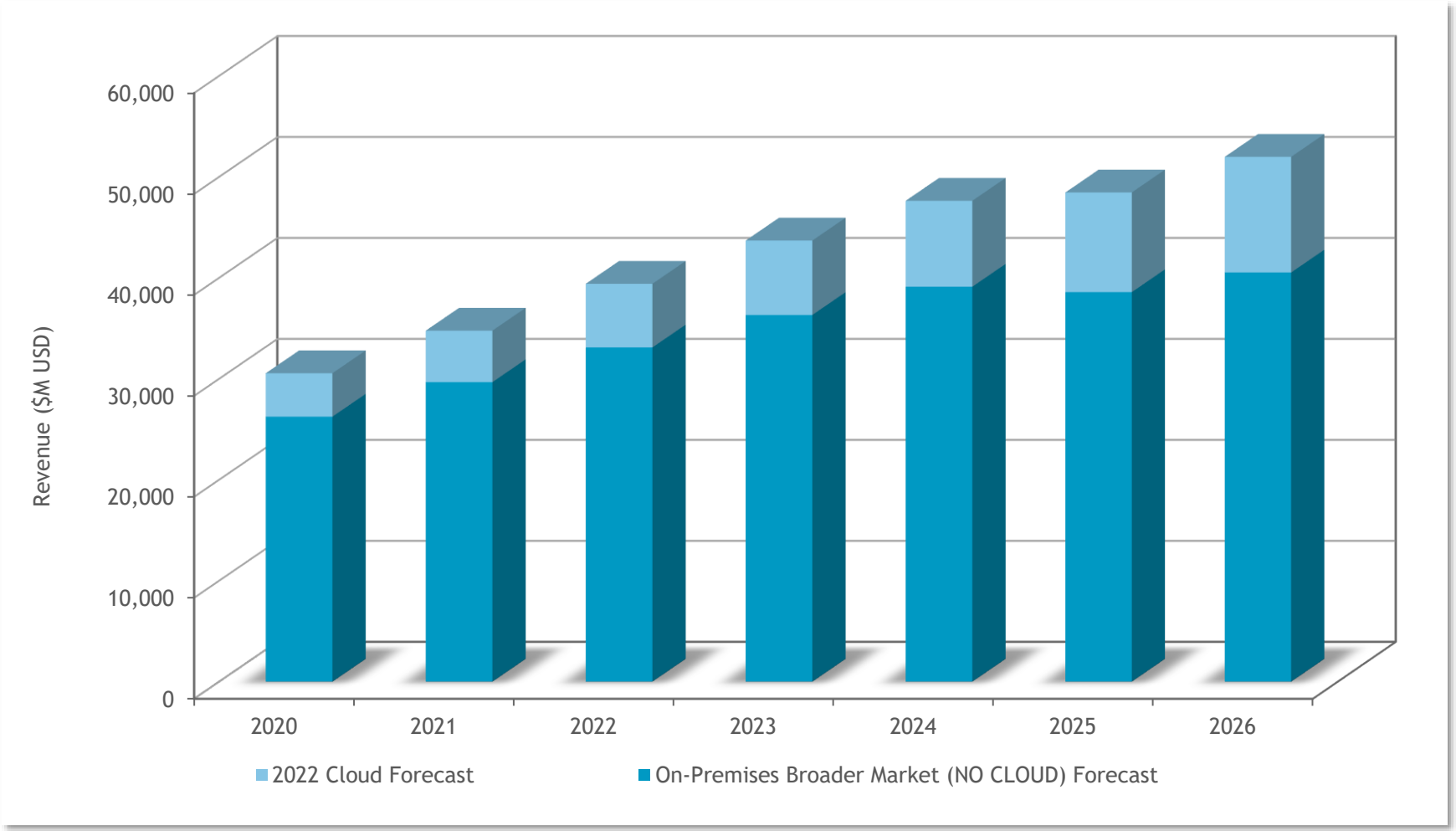
HPC cloud revenue is expected to exceed $11 billion by 2026

- **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**



The Total HPC Market: On-Prem and Cloud Computing

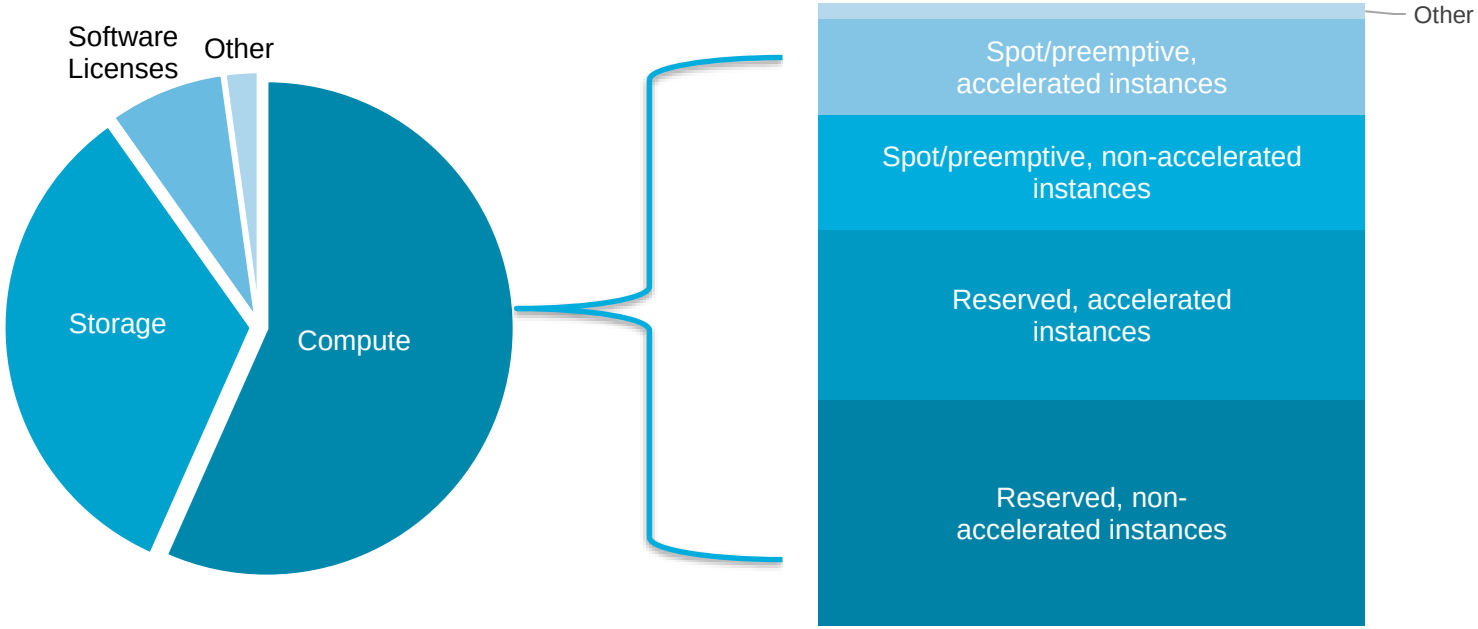
The cloud market is smaller, but growing faster



Cloud Spend Distribution

Compute comprises more than half of cloud spend

- **57% of cloud spend focused on compute instances:**
  - Roughly 50/50 split between accelerated and non-accelerated compute instances in the cloud
  - Slightly more spend on reserved instances over preemptive



Source: Hyperion Research Multi-Client End User Study 2022

Impact of HPC Cloud

Organizations are increasingly factoring cloud into future on-premises deployment plans

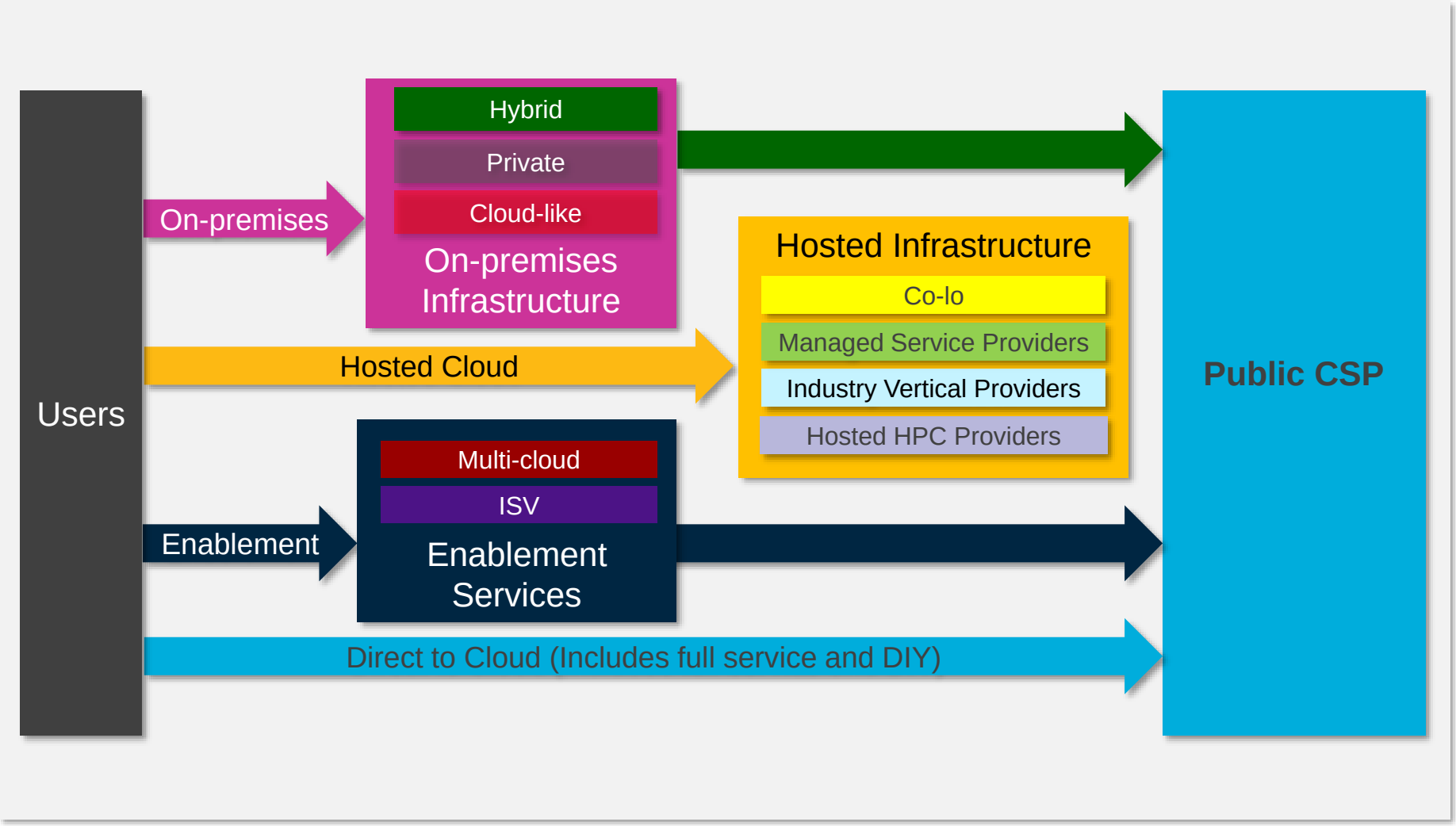
- **Public cloud resources have historically been seen as complementary to on-premises**
  - Many longitudinal studies show that cloud is used primarily for burst capabilities by many HPC users
  - This perception is shifting:
    - A recent study showed that almost 50% of the users are altering on-premises deployments due to cloud
- **Migrating HPC workloads to cloud platforms requires new skills for center managers & researchers**
  - Much of the education and training on using the cloud focuses on which workloads can and should be run in the cloud versus remain on-premises
  - IT departments are factoring in data movement and security as they expand their resource pools to consist of cloud resources

No One-Size-Fits-All

Cloud and on-prem decisions depend on variety of factors

- **Choosing where to run HPC applications depends on a wide set of factors:**
  - Internal skillsets and knowledge
  - Budget and cost
  - Performance characteristics
  - Hardware availability
  - Time constraints
  - Other
- **Understanding which applications can be run cost-effectively and performant in the cloud can aid in on-premises system design and optimization**
  - What technologies to invest in?
  - What scale to deploy on-premises?
  - Which middleware and software tools are necessary to optimally run HPC applications across platforms?
- **There is no single solution for all HPC user sites**

Models for Accessing HPC Cloud



HYPERION RESEARCH

Continue the conversation

anorton@hyperionres.com

mnossokoff@hyperionres.com

SC22 HPC Market Update - Storage and Interconnects

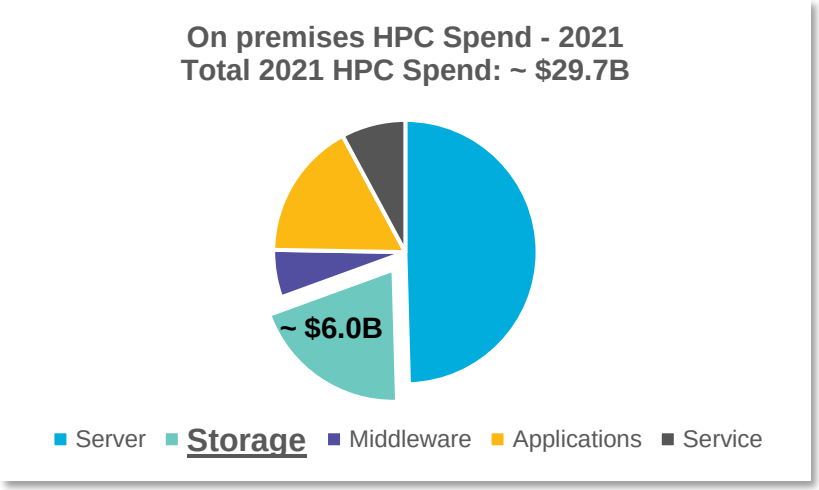
November 2022

www.HyperionResearch.com
www.hpcuserforum.com

Mark Nossokoff

HPC Storage Growth Continues

Demand increasing across all sectors and verticals



Source: Hyperion Research, 2022

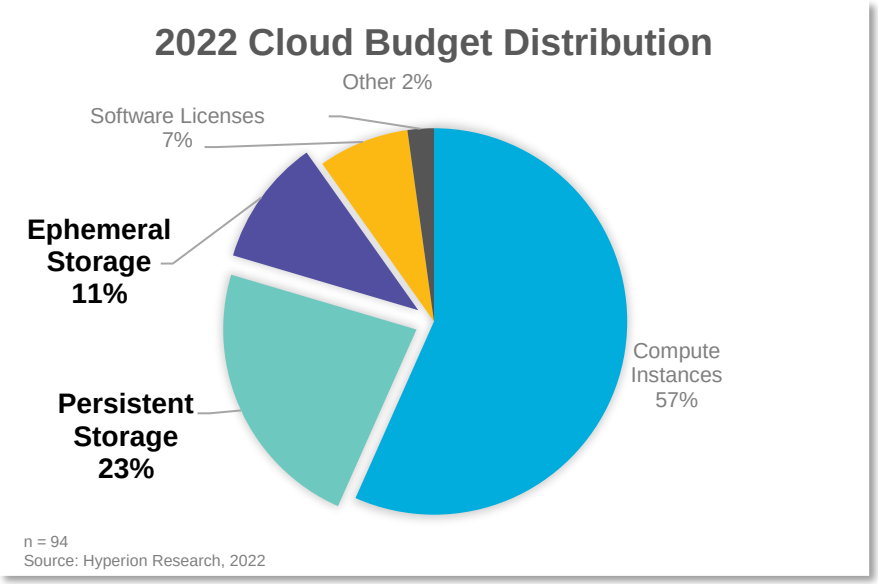
- **Storage historically the highest growth HPC element**
- **Storage represents ~ 20% of on-premises HPC spending and growing**
- **Almost half of sites surveyed expect their storage budgets to increase more than 5%**

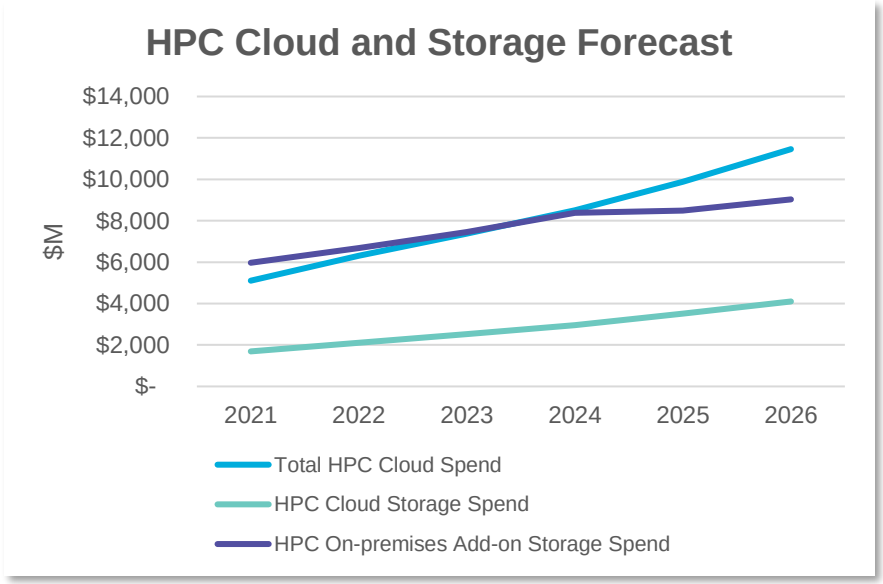
Area ($M)	2021	2022	2023	2024	2025	2026	CAGR 21-'26
Server	$14,748	$16,077	$17,738	$19,565	$19,495	$20,481	6.8%
Add-on Storage	**$5,971**	**$6,677**	**$7,457**	**$8,388**	**$8,491**	**$9,027**	**8.6%**
Middleware	$1,729	$1,863	$2,030	$2,212	$2,172	$2,268	5.6%
Applications	$4,948	$5,302	$5,731	$6,195	$6,089	$6,326	5.0%
Service	$2,266	$2,316	$2,389	$2,468	$2,336	$2,296	0.3%
Total Revenue	$29,662	$32,236	$35,345	$38,828	$38,584	$40,398	6.4%

Source: Hyperion Research, 2021

HPC Storage and the Cloud

Cloud adoption for storage remains strong and growing



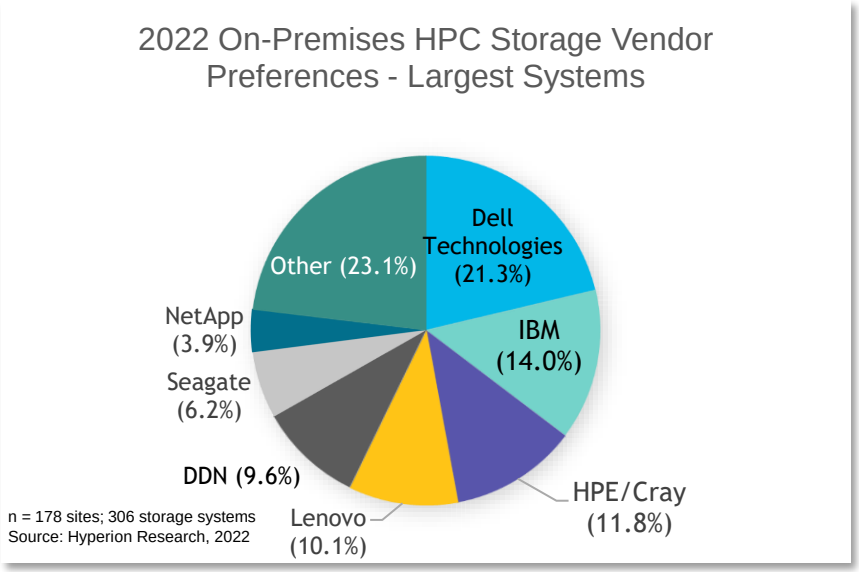


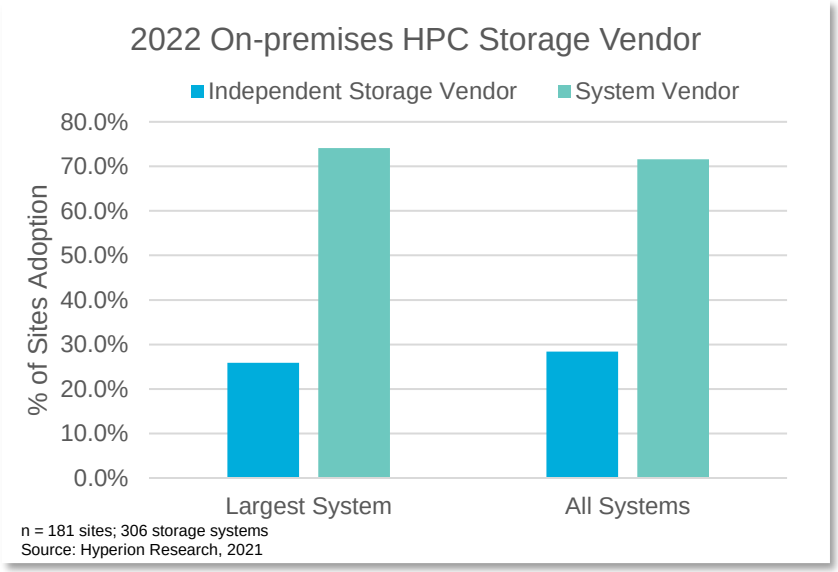
- **Storage ~ 1/3 total HPC spending in the cloud**
- **Spending on persistent, durable storage 2x greater than ephemeral, temporal storage**

- **~ $1.7B cloud storage spend in 2021**
- **Cloud storage growth ~ 2.3x on-premises storage growth**
- **Total cloud spending projected to overtake on-premises storage spending in 2024**

2022 HPC Storage Vendor Preferences

Dell Technologies continues as top preferred storage vendor





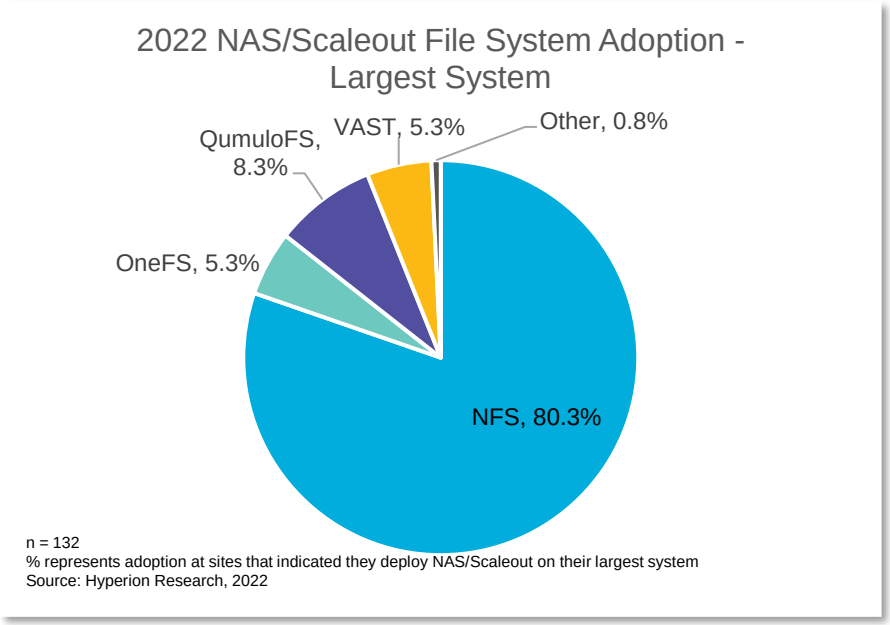
- **Dell Technologies first overall; preferred by Industry & Academia**
- **IBM second preferred overall, tied with HPE/Cray in Government and Academia**
- **DDN 4th continues as preferred independent storage vendor**
- **Users continue to prefer sourcing storage from their system providers**

HYPERION RESEARCH

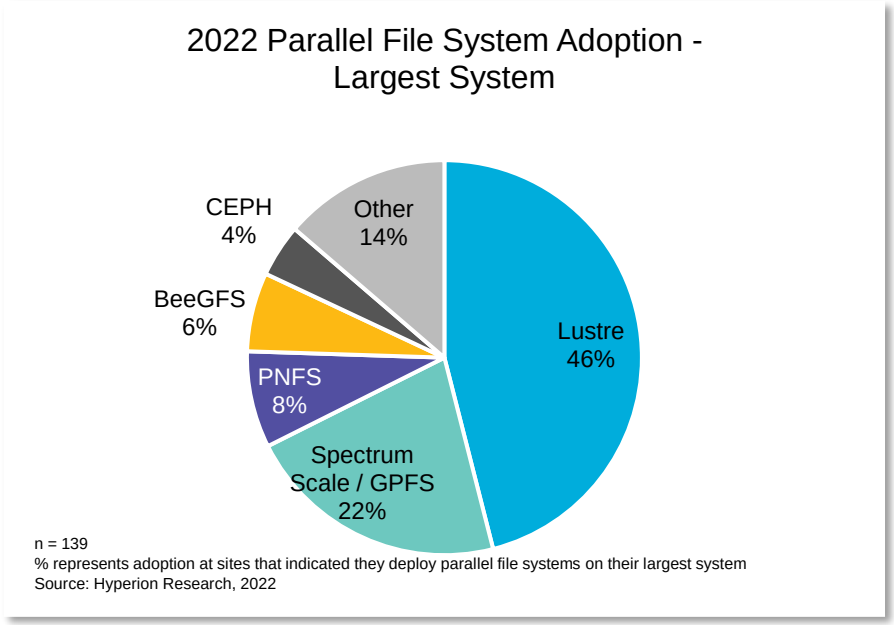
Storage Technology Ecosystem

File System Preferences – Largest System

Large % of sites adopt both NAS/scaleout and parallel file systems on their largest systems



n = 132
% represents adoption at sites that indicated they deploy NAS/Scaleout on their largest system
Source: Hyperion Research, 2022



n = 139
% represents adoption at sites that indicated they deploy parallel file systems on their largest system
Source: Hyperion Research, 2022

Interconnect Architecture

Subtle shift expected in interconnect implementation

Network Architecture	Current	Next Procurement
Independent Networks	54.1%	47%
Converged Networks	45.9%	53%

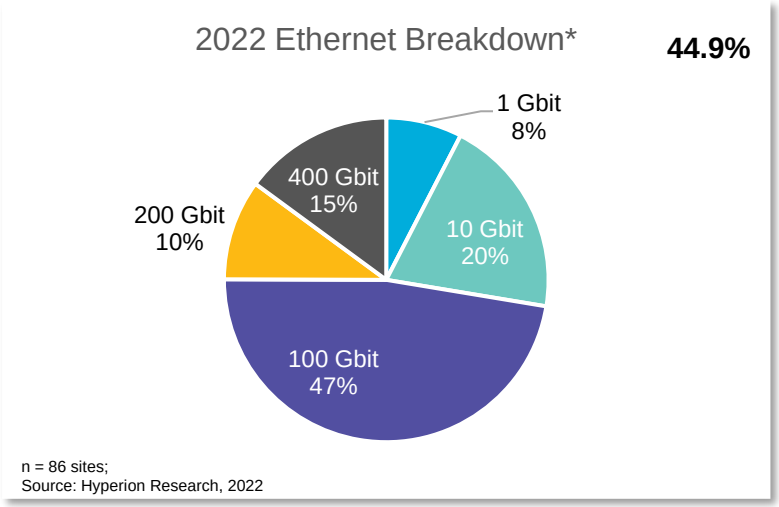
n = 94
Source: Hyperion Research, 2022

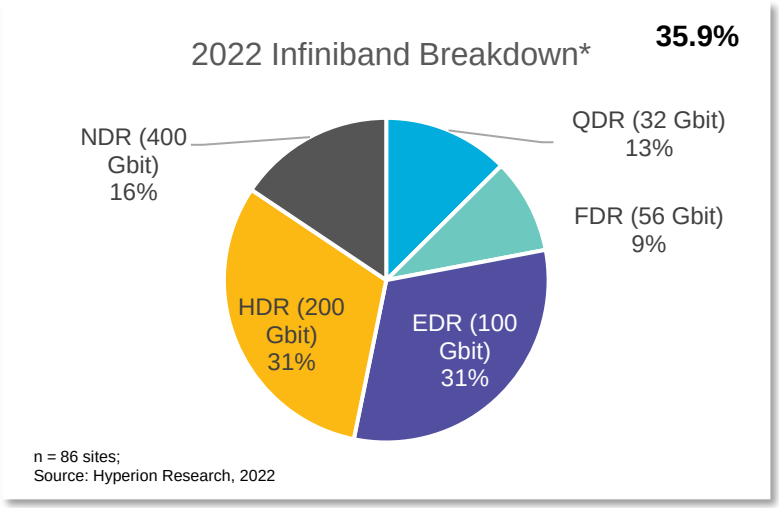
- **Site's largest systems architecture preference today skew towards independent system-system and system-storage**
- **Site's expressed preference to skew toward converged system-system and system-storage network in next procurement**
- **What's happening:**
  - Networking rates are increasing to the point of being able to support the demands of converged networks
  - Features are being implemented to optimize overall performance
  - Requirements to support the mixes workloads of compute-intensive and data-intensive workloads are being addressed to alleviate the need and expense of separate networks

Independent Networks: System-System

Ethernet preferred over Infiniband for system-system

- **Ethernet**
  - 44.9% of sites surveyed with independent networks
  - 100Gb overtakes 10Gbit as most widely deployed within Ethernet sites
- **InfiniBand**
  - 35.9% of sites surveyed with independent networks
  - HDR 200 catches up with EDR 100 Gbit as most widely deployed at Infiniband sites
- **Other**
  - Omni-Path = 4.5%
  - Other = 4.5%



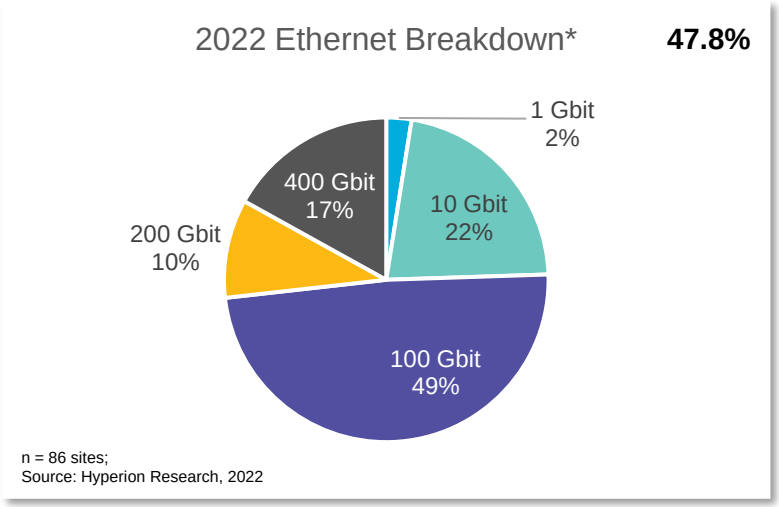


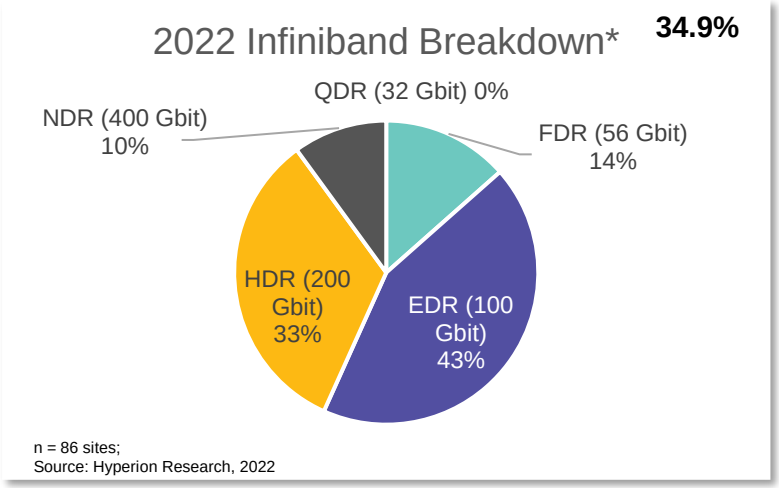
* % of site adoption Ethernet and Infiniband, respectively

Independent Networks: System-Storage

Ethernet preferred over Infiniband for system-storage

- **Ethernet**
  - 47.8% of sites surveyed with independent networks
  - 100Gb most widely deployed at almost half of Ethernet sites
- **InfiniBand**
  - 34.9% of sites surveyed with independent networks
  - EDR 100 Gbit most widely deployed at Infiniband sites
- **Other**
  - Omni-Path = 3.5%
  - Other = 1.2%



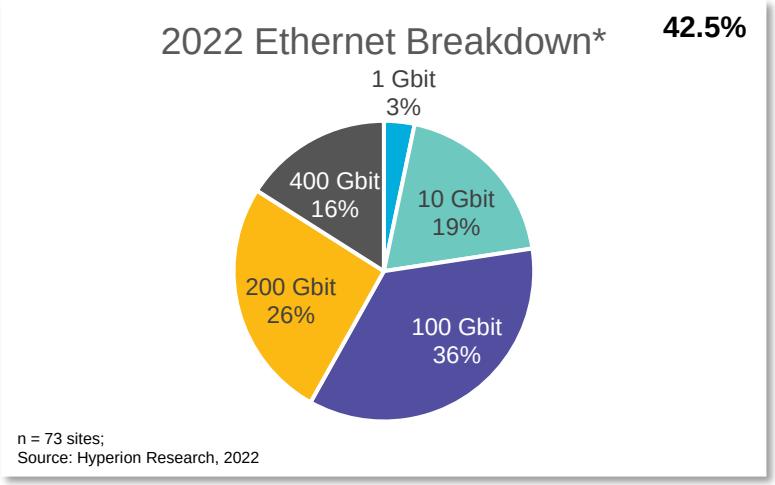


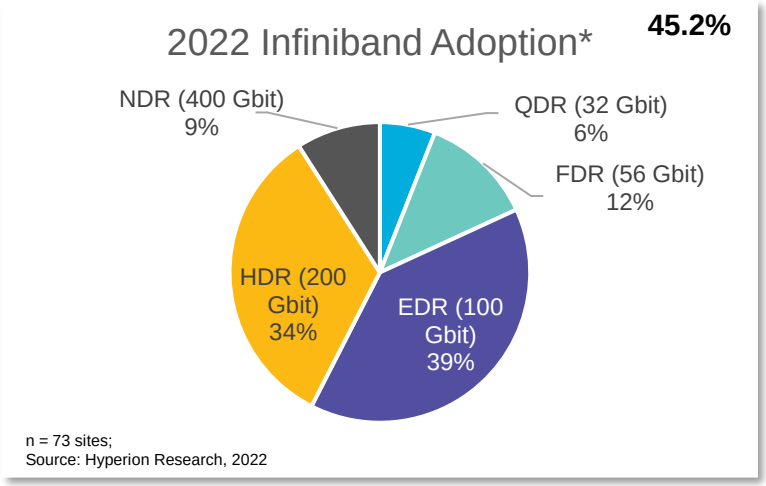
* % of site adoption Ethernet and Infiniband, respectively

Converged Networks

Infiniband preferred over Ethernet for converged networks

- **Ethernet**
  - 42.5% of sites surveyed with independent networks
  - 100Gb most widely deployed within Ethernet sites
- **InfiniBand**
  - 45.2% of sites surveyed with independent networks
  - EDR 100 Gbit most widely deployed at Infiniband sites
- **Other**
  - Omni-Path = 4.1%
  - Other = 8.2%





* % of site adoption Ethernet and Infiniband, respectively

HYPERION RESEARCH

Future Research Direction

Future Research Direction

Broad range of topics across diverse storage ecosystem

- **Grow on-premises census data including HDD, solid state and tape**
  - Use cases
    - Temporal, durable
    - File, block, object
    - Scratch, user, home directory, project, campaign, archive
  - Capacity
  - Internal to servers and compute nodes
- **Expand cloud storage coverage**
  - Storage landscape model
  - Workload usage and requirements
  - Impact of containers
- **Edge computing implications on storage**
  - Capacities
  - Architectures
  - Computational storage
- **Intranode interconnects and protocols**
  - NVMeoF
  - CXL
  - Memory pooling
  - Chiplet standards and adoption
    - UCIe
    - OCP Bunch of Wires (BoW)
- **Memory topics**
  - Big memory
  - In-and-near memory computing

HYPERION RESEARCH

PLEASE SHARE YOUR THOUGHTS!

info@hyperionres.com

mnossokoff@hyperionres.com

HYPERION RESEARCH

2022 Multi-Client Study Results

HPC Users Are Optimistic About Budgets, Prioritizing Hardware and Storage

- **Half of 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
- **Most users reported 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
- **Average site storage capacity increased significantly over the past year among Industry sites**
- **A quarter of overall public cloud spending in HPC (23%) is persistent storage**
- **Ethernet continues 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**
- **Top limitations for increasing public cloud use are budgets 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**
- **Following Red Hat's change to the CentOS business and support model, Ubuntu Linux has risen in popularity as a free alternative CentOS usage declines**

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

- **AI popularity continues to rise**
  - Virtually all users report plans to use AI methodologies
  - In Industry, AI is overtaking traditional mod/sim as % of workload
  - AI-specific software licenses such as Databricks and Anaconda are almost exclusively being purchased by Industry sites
- **A quarter of users (28%) 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 interest and attention, as are DPUs**

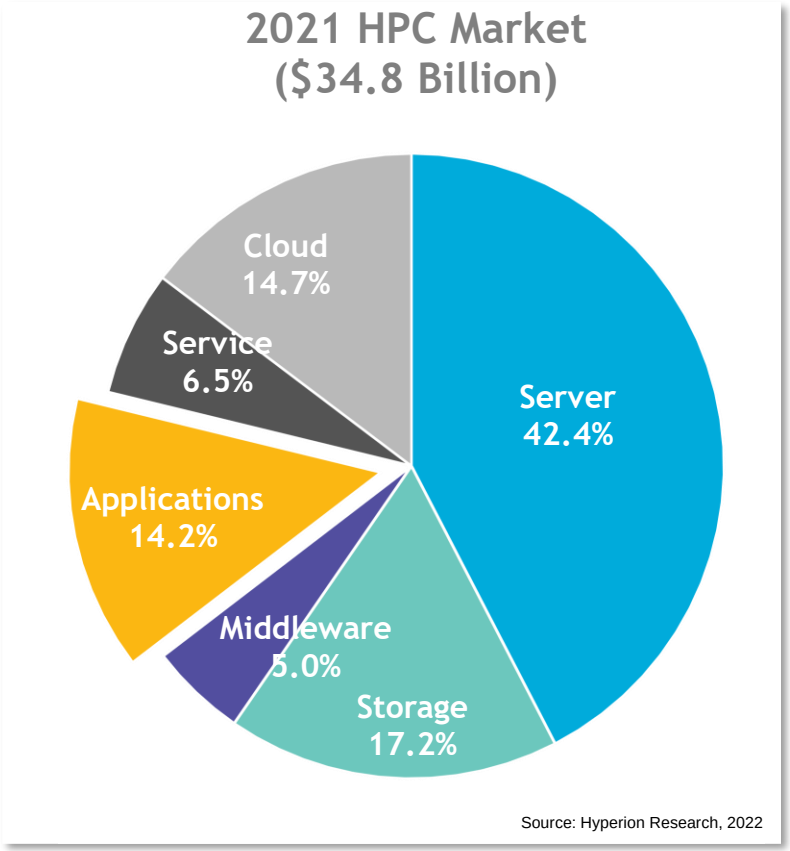
SC22 HPC Applications Update

November 2022

www.HyperionResearch.com
www.hpcuserforum.com

Jaclyn Ludema

HPC Applications Market View



- **$ 4.95 billion (14.2% of the broader market)**
- **For every dollar spent on servers, applications add 33 cents to the price tag**
- **The applications market exceeded growth expectations from 2020 to 2021. Last year it was expected to grow by 6.8%, and we're seeing 14.8% growth so far**

Revenues by the Broader HPC Market Areas

	2019	2020	2021
Server	13,368	13,523	14,763
Storage	5,288	5,079	5,984
Middleware	1,572	1,491	1,731
Applications	4,569	4,315	4,952
Service	2,181	2,015	2,267
Total Revenue	26,979	26,423	29,697

Source: Hyperion Research, 2022

Migration of CentOS Users

Supporting numbers of the trend to migrate away from CentOS

- **Only 37.6% of MCS 2022 respondents report using CentOS this year, with 72.1% of those users reporting plans to migrate from CentOS to another operating system**
- **Ubuntu Linux is gaining popularity**

Migration from CentOS, what OS are you migrating to?

	Overall Percent	Industry Percent	Government Percent	Academia Percent
Red Hat Linux	23.5%	31.6%	12.5%	13.6%
Rocky Linux	20.6%	15.8%	37.5%	22.7%
Ubuntu Linux	19.1%	21.1%	25.0%	13.6%
Alma Linux	4.4%	2.6%	-	9.1%
SUSE Linux	2.9%	5.3%	-	-
Windows	1.5%	2.6%	-	-
I will continue to use CentOS and not migrate	27.9%	21.1%	25.0%	40.9%

Source: Hyperion Research, 2022

Accelerated Applications

Finding meaningful uses for GPUs and accelerators is a priority

- **Use of GPUs and accelerators are on the rise among HPC users**
- **On average, MCS 2022 respondents spend a third (32%) of their HPC cycles on accelerated applications**
  - 87.8% of respondents have at least one accelerated critical application
  - The average site has 7.5 accelerated applications
  - 74.6% of respondents use coprocessors/accelerators on their most important or most used application (at least some of the time)
    - Industry respondents lead accelerator/coprocessor usage at 82%

Application Run Time

Government is speeding up

- **From the MCS 2022, 29% of the most important or most used applications have a typical runtime of 24 hours or more**
  - This is a remarkable shift from last years study, where 40% of respondents reported typical runtimes of 24 hours or more for their most important/most used applications
  - The government sector has the most notable change, where last year 61% of most important/most used applications were 24 hours or more, and this year that number has shrunk to 35%
- **Growing run time was a key finding of last year's MCS**

HYPERION RESEARCH

QUESTIONS?

info@hyperionres.com
jludema@hyperionres.com

Global AI Update

SC22

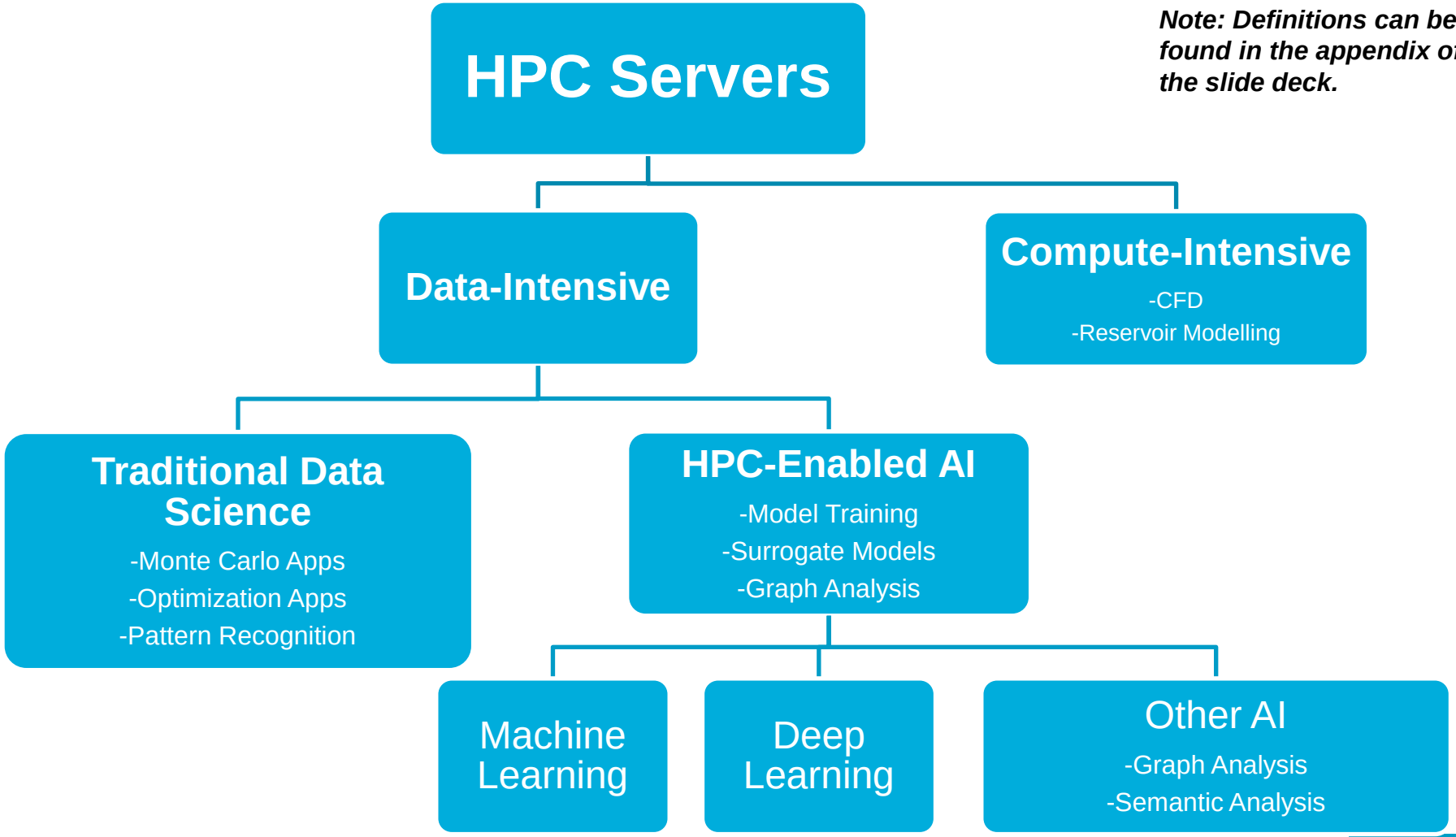
November 2022

Alex Norton and Tom Sorensen

www.HyperionResearch.com
www.hpcuserforum.com

On-Prem HPC Market Segmentation

Server classification based on end-user application



Note: Definitions can be found in the appendix of the slide deck.

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%

Future HPC System Design

AI and HPDA workloads pushing sites to consider new system architectures

- **As workloads become more diverse, system designs have shifted:**
  - Some sites are building single, large, heterogeneous systems to address a wide variety of applications
  - Some sites are building out multiple, smaller systems to handle different workloads specifically
  - Cloud resources are growing in utilization to address data-intensive workloads
- **Technology options have diversified as well**
  - New accelerator options, including AI-specific ASICs
  - Various memory, interconnect, and storage solutions
- **Compute resource allocation should be treated as an optimization problem:**
  - Find a balance among diverse technology options
  - Optimize for key workloads

Intersection of HPC and AI

Modeling and simulation workloads working in harmony with AI techniques

- **AI applications growing in the HPC space:**
  - Stand-alone AI models
  - AI incorporated into traditional simulation workloads:
    - Surrogate models
    - Data preparation and cleansing
    - Simulation steering with trained AI models
- **Mod/sim workloads benefiting from AI**
  - Acceleration of time to solution
  - Exploring new solution spaces
  - Parsing sparse matrices of data
- **AI benefiting from mod/sim workloads**
  - Generation of large synthetic datasets for training
  - Verification and testing of trained models in simulation

The Role of Explainability

For optimization, engagement, and compliance

- **Reproducibility and transparency as optimizers**
  - Automated monitoring hastens and ameliorates training
  - Timesaving capabilities (automated reporting, bias or drift detection) ease developer load and free valuable time
    - More models are developed overall
    - More models make it to production
- **Explainability drives engagement**
  - Many application spaces highly value auditability
  - Reassurance for previously hesitant domains
  - Contributes positively to development of AI workforce
- **Growing efforts to regulate and standardize**
  - Bolstered public knowledge and trust
  - Auditability as a legal obligation
  - Explainability tools mitigate regulatory fines

HYPERION RESEARCH

What did we miss?

tsorensen@hyperionres.com

anorton@hyperionres.com

The Quantum Computing Sector (in 420 seconds or less)

SC 22

www.HyperionResearch.com
www.hpcuserforum.com

Bob Sorensen
Chief Analyst for Quantum Computing

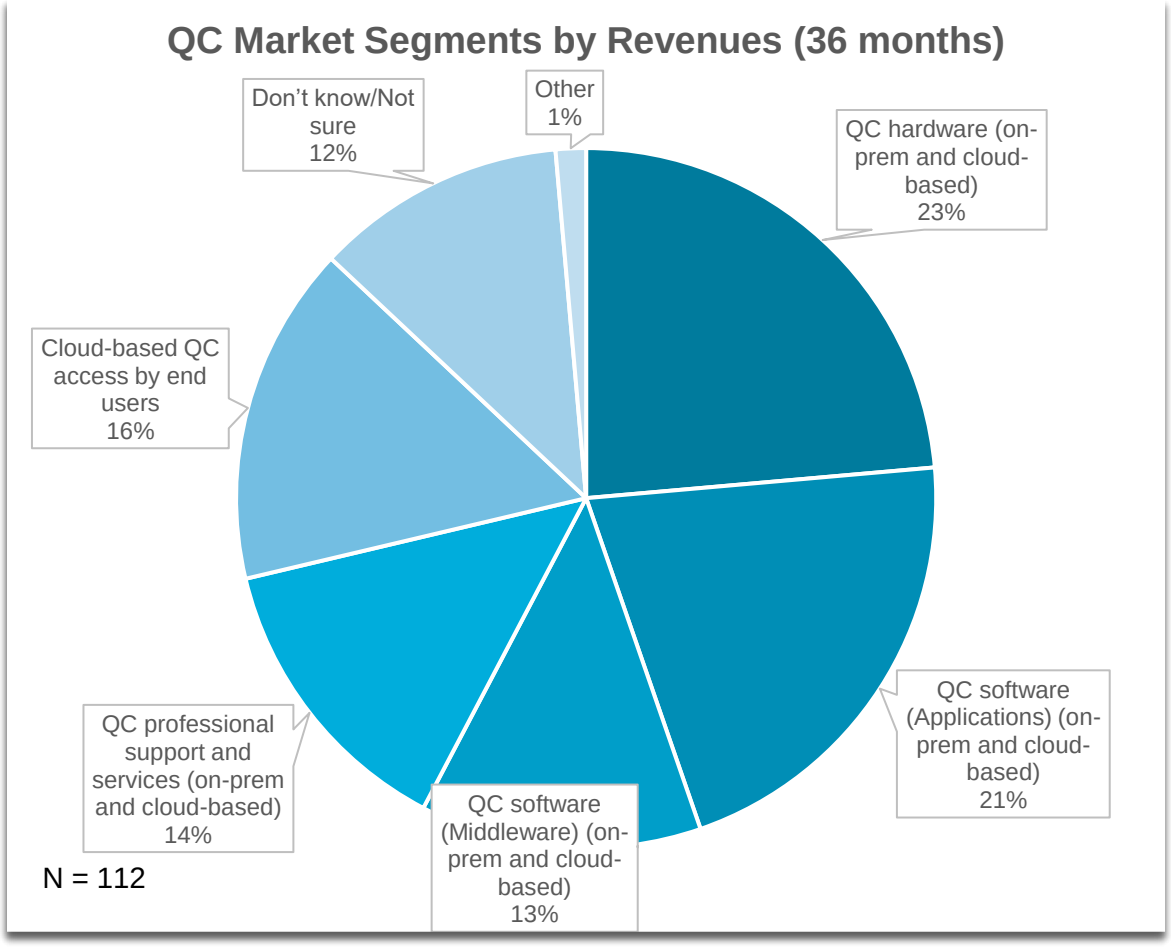
QC Market Summary

Continued strong growth seen for global QC sector

- **Based on a study of 112 QC suppliers from around the world, the estimated global QC market was worth about US $490 Million in 2021**
  - The anticipated CAGR for the sector between 2021 and 2024 is 21.9%
  - The global QC market in 2024 is estimated to be approximately US $900 Million
- **Current QC supplier base dominated by a few players**
  - 49% of companies < US $500K, 7% > US $10 Million
- **Cloud access model dominates for next three years**
  - All cloud (43%) + hybrid (21%) = 64%
- **Most Promising Market Segments**
  - QC, Cybersecurity, Financial, Academic, and Chemical/Chemistry

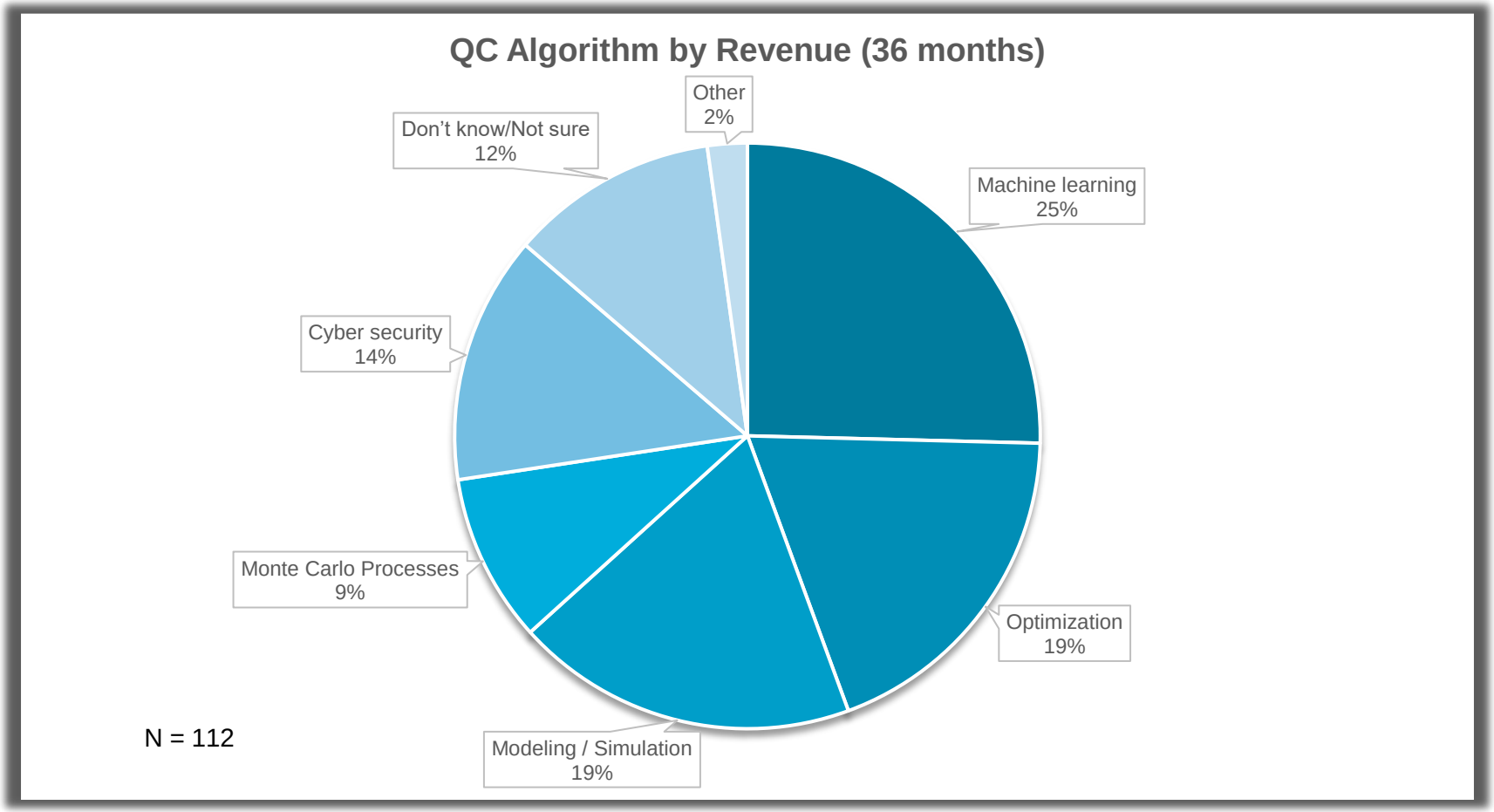
QC Market Segments by Revenue

In next 36 months, QC hardware/software near parity



QC Algorithm by Revenues

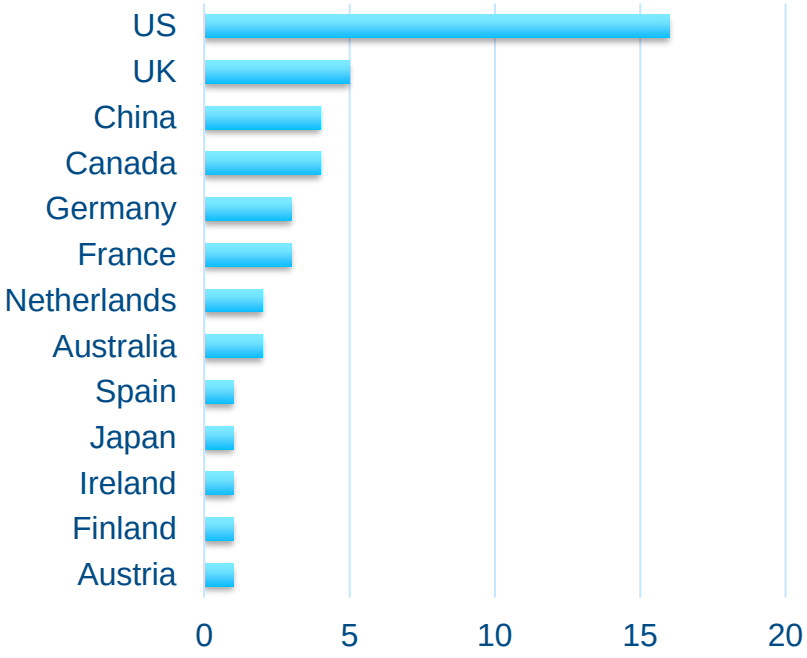
ML & optimization, M/S and cyber close behind

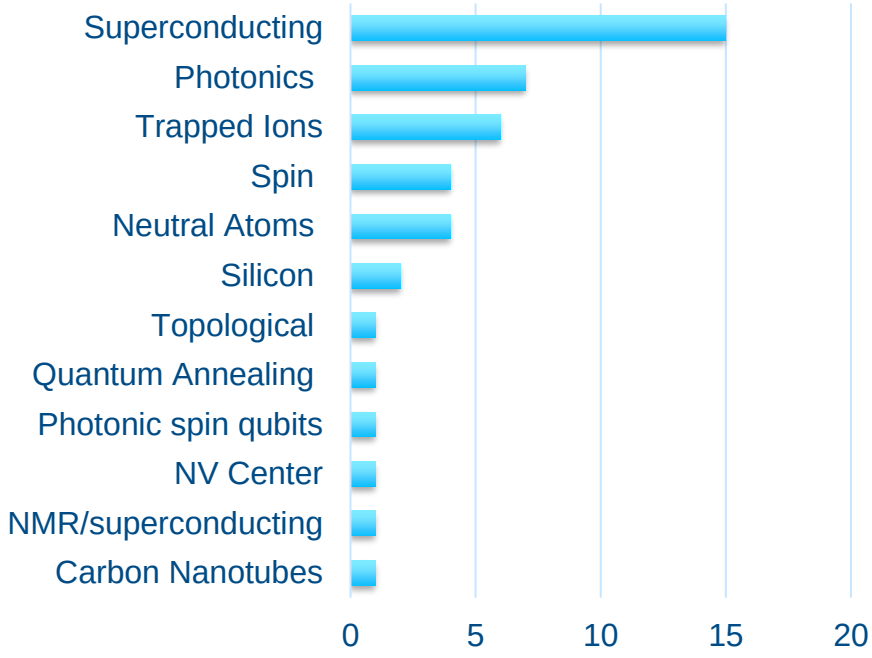


Current Field of Commercial QC Hardware Suppliers

Tracking the visible players from a global perspective

- **44 identified QC hardware developers**
- **12 quantum modalities under consideration**





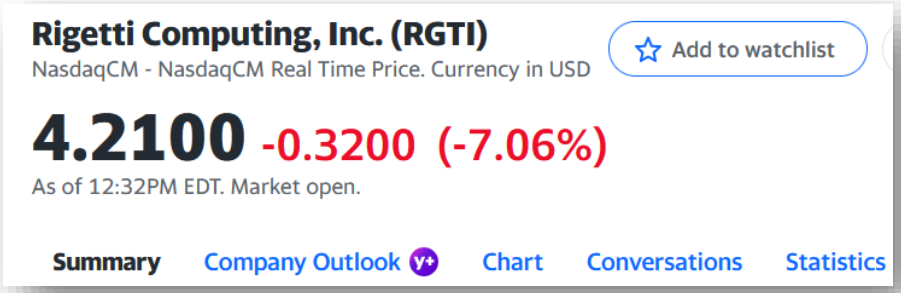
QC End User Study Key Takeaways

Commercial end users' interest in QC is high

- **Almost 70% of 415 companies surveyed worldwide have some in-house QC program**
  - Additional 20% plan to do so in the next few years
- **QC technology is seen as offering a range of corporate-level benefits including improved research capabilities and increased revenue**
- **End users are looking for solutions in optimization, factory processes, scheduling, etc.**
- **Every vertical surveyed had a significant number of organizations currently involved in some level of QC activity**
- **Most interested companies already involved in data analysis, ML/DL, optimization, mod/sim, material science—some of the most promising areas of QC today**

The QC Sector Matures

Transitioning from an R&D to market sector



Investing in Quantum Computing Stocks

An in-depth look at the best quantum computing stocks in the U.S stock market this year.

HYPERION RESEARCH

QUESTIONS?

bsorensen@hyperionres.com

Insufficient facts always invite danger.

- Spock, *Stardate: 3141.9.*

HYPERION RESEARCH

The HPC Innovation Award and Winners

Examples of Previous Winners

The Trophy for Winners

HPC Award Program Goals

- **#1 Help to expand the use of HPC by <u>showing real ROI examples</u>:**
  1. Expand the “Missing Middle” – SMBs, SMSs, etc. by providing examples of what can be done with HPC
  2. Show mainstream and leading edge HPC success stories

- **#2 Create a large database of success stories across many industries/verticals/disciplines**
  - To help justify investments and show non-users ideas on how to adopt HPC in their environment
  - Create many examples for funding bodies and politicians to use and better understand the value of HPC → to help grow public interest in expanding HPC investments
  - For OEMs to demonstrate success stories using their products

Users Submit the Value of the Accomplishment

- **Users are required to submit the value achieved with their HPC system, using 3 broad categories, following a very specific set of guidelines:**

  a) Dollar value of the HPC usage

  - e.g., made $$$ in new revenues, saved $$$ in costs, made $$$ in profits, etc.

  b) Scientific or engineering accomplishment

  - e.g., discovered how xyz really works, developed a new drug that does xyz, etc.

  c) Value to society as a whole

  - e.g., ended nuclear testing, made something safer, provided protection against xyz, etc.

  … and the investment in HPC that was required

The Judgment Process – Clear, Fair And Transparent

- **The ranking of the accomplishments are done by only HPC USERS, following very specific rules**
- **A three-step process is used:**
  1. First the submission has to be complete with a clear “value” shown
     - Some submissions were good, but needed a little more information
  2. Secondly, the HPC User Forum Steering Committee evaluates if the submission presents a realistic assessment of the value/returns
  3. Then, in cases where the value isn’t clear or a deeper technical depth is required, the final evaluation is by experts in the specific area/discipline

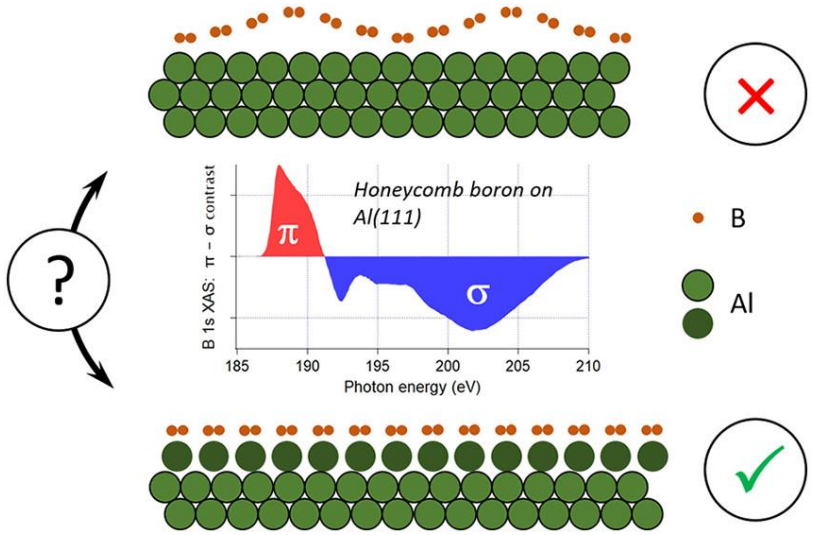
HYPERION RESEARCH

SC22 Winners:
HPC User Innovation Awards

HPC-driven design of innovative functional materials for catalysis, energy conversion, and storage

Andrey Lyalin, Hokkaido University ICReDD

- **Successfully predicted a variety of unique materials for catalysis, energy conversion, and storage applications which were then experimentally confirmed**
- **Developed novel computational methods in quantum chemistry, materials informatics, and machine learning that can be used to discover further materials for energy and environmental applications**



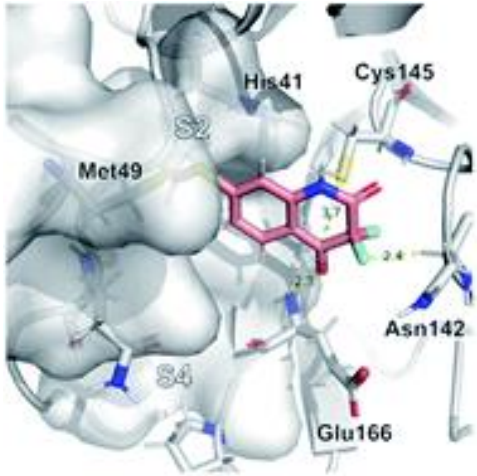
HOKKAIDO UNIVERSITY

ICReDD
Institute for Chemical Reaction Design and Discovery
HOKKAIDO UNIVERSITY

Fast-track computationally-driven discovery of new SARS-CoV-2 Main Protease (Mpro) inhibitors: from HPC to experimental drug candidates

Jean-Philip Piquemal, Sorbonne Université and Qubit Pharmaceuticals

- **Synthesized 2 new families of compounds in less than 3 months that were experimentally validated**
- **Produced a high-resolution molecular dynamics simulation of the SARS-CoV2 Main protease protein**
  - Modeled interaction with water
  - Unraveled potential druggable sites
  - Identified potential drug candidates
  - Predicted absolute free energies of binding with potential active molecules
- **Opens routes for further drug discovery and design**



SORBONNE UNIVERSITÉ

Qubit pharmaceuticals

An automated agent-based simulation method for the prediction of intensive care unit (ICU) occupancy due to COVID-19

Ralf Schneider, High Performance Computing Center Stuttgart

- **Weekly production of 4-week forecasts of ICU occupancy in Germany due to COVID-19**
  - This is the first time that a federal HPC-center in Germany delivers such a service in production
  - This proved for the first time that HPC for Global System Science can support decision makers in critical situations and can make a difference for society.
- **Forecasts are simulated by a model initially developed by the Federal Institute for Population Research**

H L R S

High-Performance Computing Center Stuttgart

AI-based emulator of Computational Fluid Dynamic simulations for urban wind flow and gas dispersion

Francisco Ramirez Javega, Bettair Cities SL

- **Air pollution is the single largest environmental health risk in Europe and a major cause of premature death and disease, but most urban air quality problems are hyper-local (e.g., land use, traffic)**
- **Ran CFD simulations of wind flow and dispersion patterns on a large data set and trained a DNN to simulate the results at a fraction of the time and cost**
  - This research empowers local communities, municipalities, and regional governments with accessible, actionable information about local emissions in real-time
- **This work has been funded within the FF4EuroHPC European project, Experiment 1012: Improving Bettair Air Quality Maps**
- **Part of the project was performed by a team from the Barcelona Supercomputing center using ALYA**

HYPERION RESEARCH

Thank you!

For questions:
info@hyperionres.com

Hyperion Research HPC Market Update: Conclusions

November 2022

Earl Joseph, Bob Sorensen, Alex Norton,
Mark Nossokoff, Melissa Riddle,
Tom Sorensen and Jaclyn Ludema

www.HyperionResearch.com
www.hpcuserforum.com

Conclusions

- **2021 was a strong year with a 9% increase**
  - 2022 is also expected to be a growth year
    - Exascale systems will help drive growth in 2022 to 2024
    - AI, HPDA/big data are hot areas
    - But supply chain issues continue
- **New technologies are showing up large numbers:**
  - Processors, AI hardware & software, memories, new storage approaches, etc.
  - Quantum technologies
- **The cloud has become a viable option for many sites**
  - HPC in the cloud is lifting the sector writ large
- **Storage will likely see major growth driven by AI, big data and the need for much larger data sets**
- **There are growing concerns about the workforce**

A Concern: HPC Expertise Shortage

The growing scarcity of HPC experts to implement new technologies is the number one roadblock for many HPC sites

- **Two major trends:**
  1) A shrinking HPC workforce
  2) A massive increase in system complexity

- **HPC experts are an aging workforce**
  - The pipeline of new HPC staff entering the workforce does not adequately match the outflow of retirees
  - Competition for HPC staff will intensify
- **Increasingly complex workloads are more difficult to manage**
  - Increasing HPC systems per site
  - Augmenting traditional modeling/simulation with AI and big data
  - Incorporating multiple processor types, co-processors, accelerators, and other specialized hardware
  - Balancing on-prem and cloud
  - And Enterprise IT users are entering HPC space, and need HPC expertise
- **HPC users need major improvements in ease-of-use, ease-of-selection, & ease-of-optimization**

Future Meetings

- **April 18 - 19, 2023 – Princeton, NJ**
  - In person HPC User Forum meeting
- **May 23, 2023 – ISC23 Breakfast Briefing**
  - In person at the Grand Elysée Hamburg
- **September 6 - 7, 2023 – Tucson, AZ**
  - In person HPC User Forum meeting
- **October HPC User Forum in Europe**

HYPERION RESEARCH

Questions?

We welcome questions, comments and suggestions

Please contact us at:
info@hyperionres.com