HPC Market Update: HPC/AI Market Results, and High Growth Areas

SC23

www.HyperionResearch.com
www.hpcuserforum.com

Earl Joseph
(ejoseph@hyperionres.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
- Jaclyn Ludema, Analyst
- Melissa Riddle, Associate Analyst
- Thomas Sorensen, Associate Analyst
- Cary Sudan, Principal Survey Specialist

Operations

- Jean Sorensen, COO

International Consultants

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

Global Accounts

- Mike Thorp, Sr. Global Sales Executive
- Kurt Gantrish, Sr. Account Executive *

Data Collection

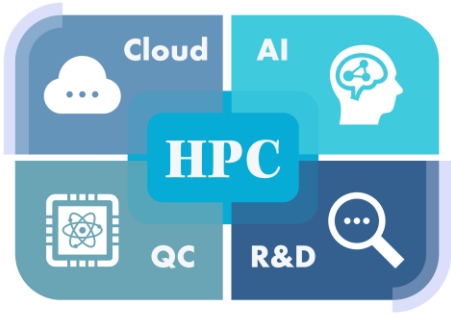
- Andrew Rugg, Certus Insights
- Kirsten Chapman, KC Associates

* primary individuals supporting this project

Our Research Areas

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

- **Traditional HPC**
- **AI: ML, DL, LLM & Large Scale AI**
- **Cloud Computing**
- **Quantum Computing**
- **Storage & Big Data**
- **Interconnects**
- **Software & Applications**
- **Power & Cooling**
- **The ROI and ROR from Using HPC**
- **Tracking all Processor Types & Growth Rates**
- **R&D and Engineering -- All Types of High Tech**
- **Edge Computing**
- **Staffing & Supply Chain Issues**



Today's Agenda

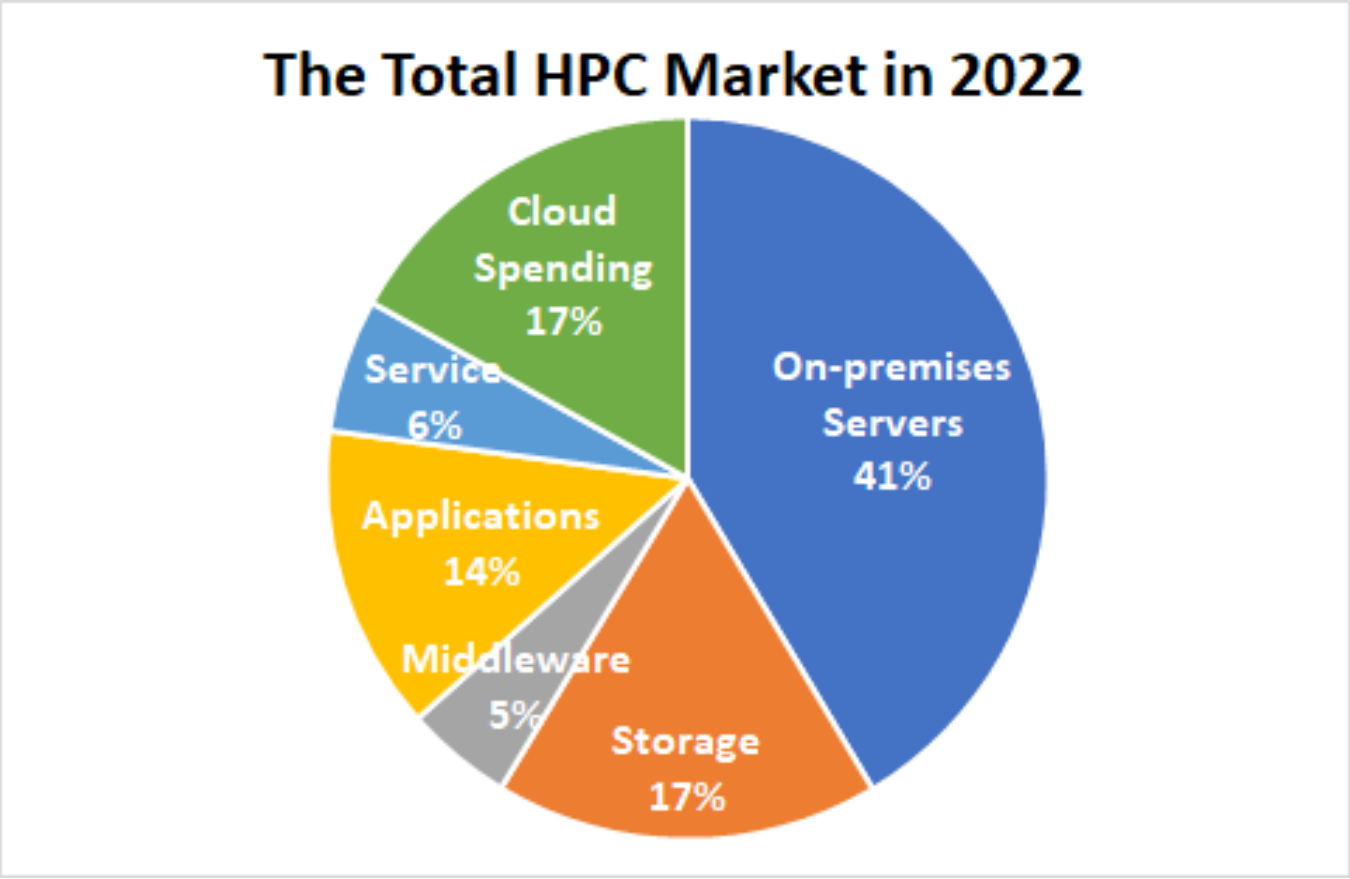
- Earl Joseph: Market Update
- Tom Sorensen: Update on the Intersection of AI and HPC
- Mark Nossokoff: State of HPC Cloud
- Jaclyn Ludema: Perspective on Sustainability in HPC
- Bob Sorensen: Exascale + Neo Exascale: What's Next?
- Mark Nossokoff: State of HPC Storage and Interconnects
- Bob Sorensen: The Global QC Market: Realistic and Steady Growth Ahead
- Melissa Riddle: HPC Applications and Verticals
- Earl Joseph: Conclusions

HYPERION RESEARCH

HPC Market Update

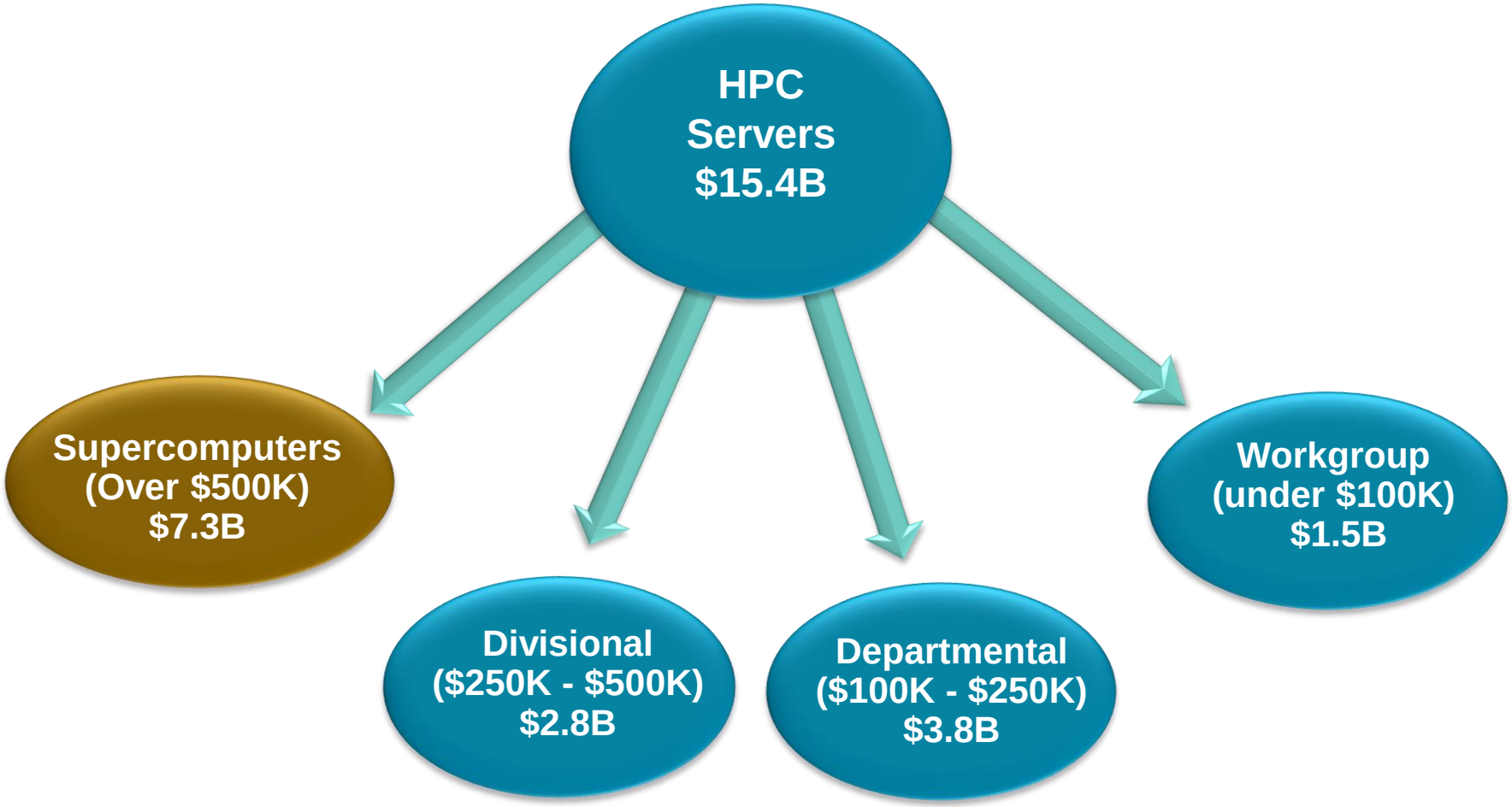
The Overall HPC Market in 2022

Looking at the overall HPC market, including servers, cloud usage, storage, software and repair services = $37.3 billion US dollars



The 2022 Worldwide On-Prem HPC Server Market: $15.4 Billion (up 4.3%)

2023 is projected to be around $16.5 Billion



2022 WW HPC On-Prem Market by Vendor and Sector ($ Millions)

HPC On-premises Server Market ($M)	
Vendor	**2022**
HPE	$5,137
Dell Technologies	$3,575
Lenovo	$1,201
Inspur	$1,073
Sugon	$603
IBM	$505
Atos	$480
Fujitsu	$230
NEC	$207
Penguin	$442
Other	$1,988
Total	**$15,441**
Source: Hyperion Research, 2023	

HPC On-premises Server Market ($M)	
Sector/Vertical	**2022**
Bio-Sciences	$1,449
CAE	$1,768
Chemical Engineering	$173
DCC & Distribution	$826
Economics/Financial	$757
EDA / IT / ISV	$873
Geosciences	$998
Mechanical Design	$57
Defense	$1,602
Government Lab	$3,342
University/Academic	$2,677
Weather	$700
Other	$221
Total	**$15,441**
Source: Hyperion Research, 2023	

First Half of 2023 WW HPC On-Prem
($ Millions)

9.5% growth for the first six months of 2023

First Half 2023 Growth Rate									
	2022				2023		First Half	First Half	Growth
$ millions	Q122	Q222	Q322	Q422	Q123	Q223	2022	2023	Rate
Supercomputer	$1,348	$1,533	$2,071	$2,267	$1,600	$1,729	$2,881	$3,330	15.6%
Divisional	$608	$656	$725	$815	$656	$673	$1,264	$1,329	5.1%
Departmental	$852	$900	$1,000	$1,073	$925	$886	$1,752	$1,811	3.4%
Workgroup	$337	$347	$412	$423	$355	$384	$684	$739	8.0%
Total Revenue	**$3,144**	**$3,437**	**$4,209**	**$4,578**	**$3,536**	**$3,673**	**$6,582**	**$7,209**	**9.5%**
Source: Hyperion Research, September 2023									

5-Year On-Prem HPC Server Forecast

8.0% yearly average growth over the next 5 years
LLMs and other AI are driving growth increases

On-Prem HPC Server Revenue Forecast by Competitive Segment								
$ millions	2021	2022	2023	2024	2025	2026	2027	CAGR 22-27
Supercomputer	$6,971	$7,219	$7,958	$8,839	$9,577	$10,390	$11,360	9.5%
Divisional	$2,783	$2,805	$3,067	$3,371	$3,580	$3,897	$4,182	8.3%
Departmental	$3,614	$3,826	$4,020	$4,397	$4,660	$4,981	$5,336	6.9%
Workgroup	$1,412	$1,519	$1,441	$1,507	$1,553	$1,626	$1,707	2.4%
Total	**$14,781**	**$15,369**	**$16,486**	**$18,113**	**$19,369**	**$20,894**	**$22,586**	**8.0%**
Source: Hyperion Research, September 2023								

On-Prem Broader Market Forecast

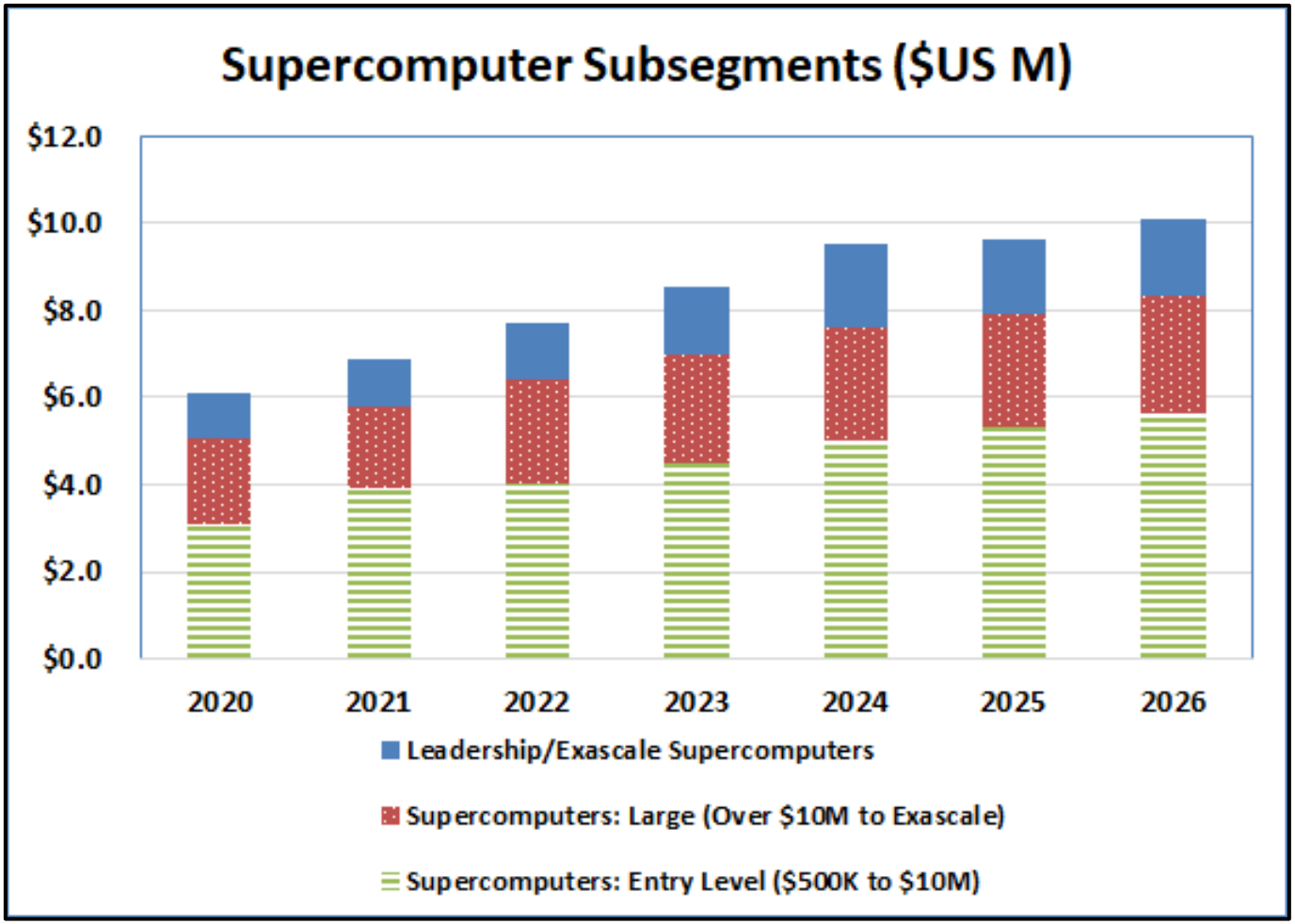
Overall market to exceed $44 billion by 2027

On-Prem Revenues by the Broader HPC Market Areas								
$ millions	2021	2022	2023	2024	2025	2026	2027	CAGR 22-27
Server	$14,781	$15,369	$16,486	$18,113	$19,369	$20,894	$22,586	8.0%
Storage	$5,985	$6,380	$6,924	$7,759	$8,434	$9,161	$10,007	9.4%
Middleware	$1,733	$1,781	$1,887	$2,049	$2,160	$2,316	$2,503	7.0%
Applications	$4,960	$5,069	$5,320	$5,729	$6,045	$6,446	$6,935	6.5%
Service	$2,272	$2,214	$2,220	$2,286	$2,323	$2,344	$2,508	2.5%
Total Revenue	**$29,731**	**$30,813**	**$32,836**	**$35,936**	**$38,331**	**$41,161**	**$44,539**	**7.6%**

Source: Hyperion Research, September 2023

Supercomputer Subsegments

The market for systems under $10M US is very large



High Growth Areas

Relative Growth Rates

The use of LLMs and other AI will likely increase these rates

- **8.0% On-premises HPC/AI servers**
  - 7.6% HPC/AI broader on-premises market
- **17.2 % GPU boards/accelerators**
  - With increased prices, GPU revenues are growing at a considerably higher rate
- **17.9% Running HPC workloads in the cloud**
- **22.7% AI systems**
  - 32.2% for DL

The Exascale Market (System Acceptances)

Exascale and Near-Exascale Leadership Systems (2020 to 2028)							
Year Accepted	**China**	**Europe**	**Japan**	**US**	**Other Countries***	**Total Systems**	**Total Value**
2020			1 near-exascale system ~$1.1B			1	$1.1B
2021	2 exascale ~$350M each	1 pre-exascale system ~$180M	--	1 pre-exascale system ~$200M	--	4	$1.1B
2022	1 exascale ~$350M	2 pre-exascale systems ~$390M **total**	--	1 exascale system ~$600M (2/3 accepted 2022)	--	4	$1.1B
2023	1 exascale system ~$350M	1 or 2 pre-exascale systems ~$150M each	1 near-exascale system ~$150M	1 exascale system ~$600M + remaining 1/3 of Frontier system	--	5-6	$1.5B - $1.6B
2024	1 exascale system ~$350M	1 exascale ~$350M, plus 1 exascale (or pre) system ~$200M	?	1 exascale system ~$600M	1 pre-exascale system ~$125M	5	~$1.6B
2025	1 or 2 exascale systems ~$300M each	2 or 3 exascale systems ~$350M each	1 exascale system ~$200M	1 or 2 exascale systems ~$350M each	1 near-exascale system ~$125M	6-9	$1.7B - $2.7B
2026	2 exascale systems ~$300M each	2 or 3 exascale systems ~$325M each	?	1 or 2 exascale systems ~$325M each	1 or 2 exascale systems ~$150M each	6-9	$1.7B - $2.5B
2027	2 exascale systems ~$275M each	2 or 3 exascale systems ~$300M	1 exascale system ~$150M	1 or 2 exascale systems ~$275M each	2 or 3 exascale systems ~$130M each	8-11	$1.8B - $2.5B
2028	2 exascale systems ~$250M each	2 or 3 exascale systems ~$275M	1 or 2 exascale systems ~$150M each	1 or 2 exascale systems ~$275M each	2 or 3 exascale systems ~$125M each	8-12	$1.7B - $2.6B
Total	**12-13**	**14-19**	**5-6**	**8-12**	**7-10**	**47-61**	**$13.4B - $16.8B**

94.3% of Sites Have Accelerators in Their Largest System Today

Up from 82.7% having accelerators in 2021

In Mid 2021

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		

In Late 2022

Largest System Accelerator Count		
Q: How many compute-oriented accelerators/co-processors are in your largest on-premises HPC technical server?		
		Overall Percent
None		5.7%
Less than 32		24.4%
32 to less than 64		15.3%
64 to less than 100		12.5%
100 to less than 500		13.1%
500 to less than 1,000		7.4%
1,000 to less than 5,000		7.4%
5,000 to less than 10,000		2.8%
10,000 to less than 50,000		2.3%
50,000 to less than 100,000		4.0%
100,000 to less than 250,000		3.4%
250,000 to less than 500,000		0.6%
750,000 to less than 1,000,000		0.6%
1,000,000 to less than 5,000,000		0.6%
n = 176; 104; 20; 52		
Source: Hyperion Research, 2023		

Accelerator Plans for Next Purchases

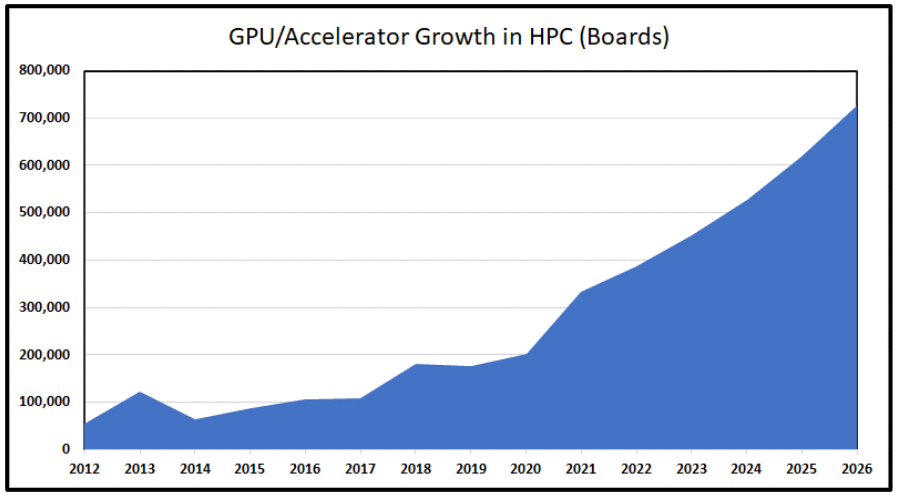
From our recent end-user MCS study

Planned Processing Elements by Sector				
Q: In the next 12 – 18 months, which of these processing elements do you expect will be incorporated into your HPC/AI/HPDA compute resources? Select all that apply:				
	Overall Percent	**Industry Percent**	**Government Percent**	**Academia Percent**
GPUs	74.0%	67.9%	85.0%	82.7%
TPUs (tensor processing units)	24.3%	27.5%	25.0%	17.3%
FPGAs	22.7%	28.4%	15.0%	13.5%
Single-purpose AI processors	11.0%	12.8%	5.0%	9.6%
ASICs	8.3%	11.9%	0.0%	3.8%
Neuromorphic processors	7.7%	9.2%	10.0%	3.8%
eASICs	2.2%	3.7%	0.0%	0.0%
Other	2.8%	2.8%	0.0%	3.8%
None	5.5%	7.3%	5.0%	1.9%
n = 181; 109; 20; 52				
Source: Hyperion Research, 2023				

GPU/Accelerator Forecast

Anticipated high growth for accelerators over next 5 years

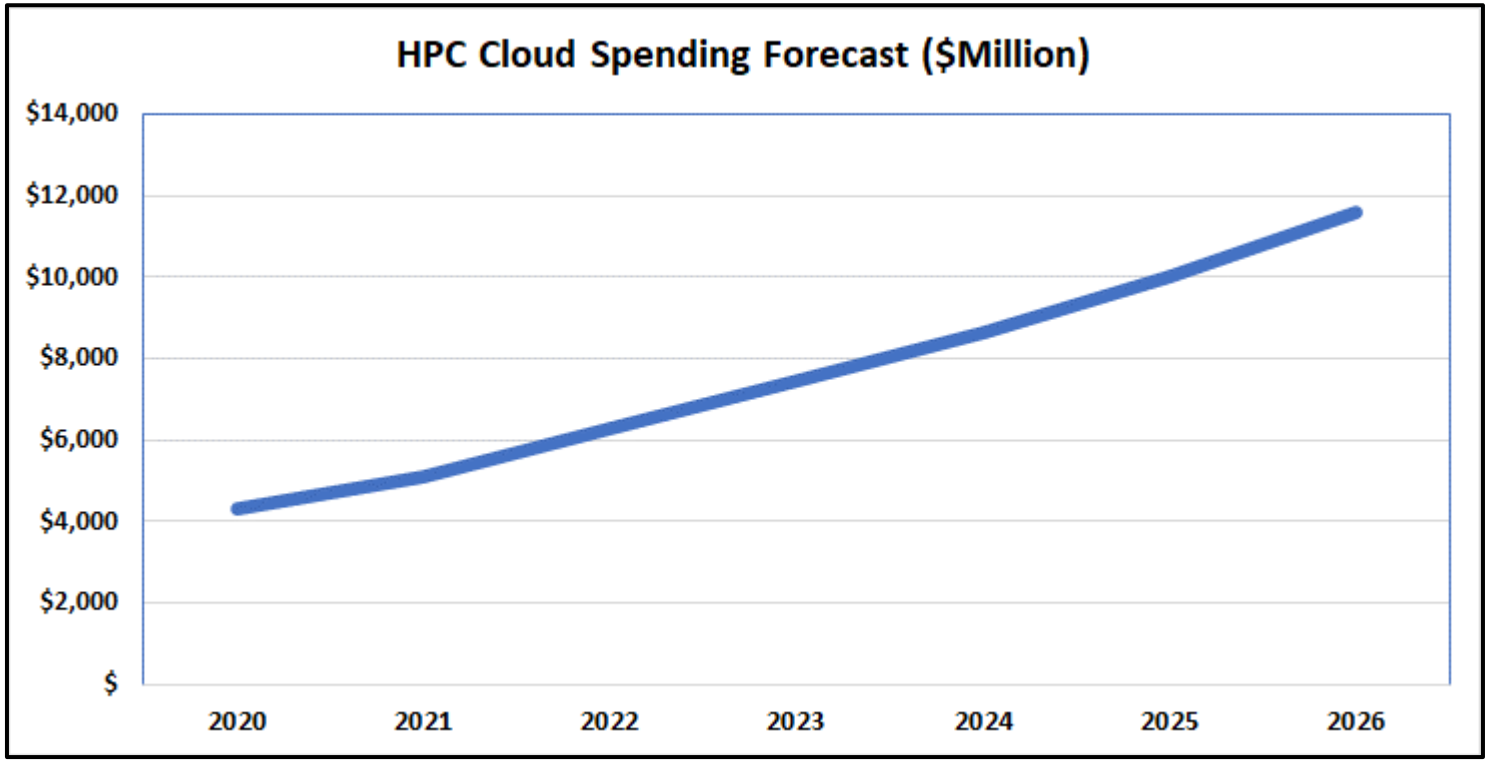
GPU & Accelerators Forecast								
	2021	2022	2023	2024	2025	2026	2027	CAGR 22-27
GPC & Accelerator Boards	334,037	386,035	476,371	546,438	639,405	742,853	854,280	17.2%
Source: Hyperion Research, September 2023								



HPC Cloud Usage Forecast

17.9% growth over the next 5 years

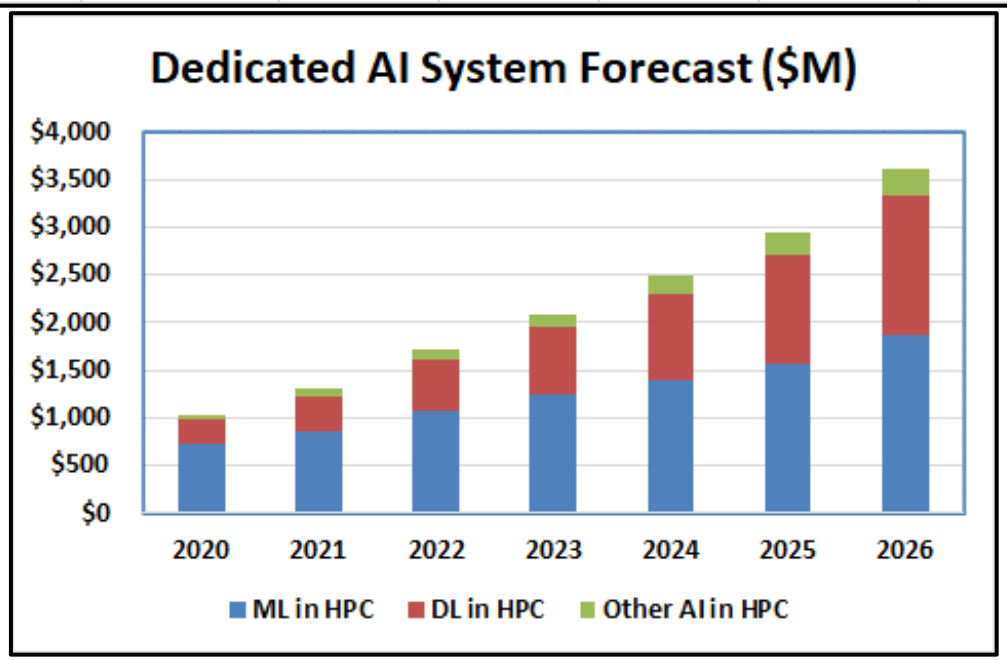
HPC Cloud Spending ($ Million)								
	2020	2021	2022	2023	2024	2025	2026	CAGR 21 to 26
HPC Cloud Spending	$4,300	$5,100	$6,304	$7,472	$8,630	$10,011	$11,613	17.9%
Source: Hyperion Research, 2023								



AI Forecast

22.7% growth over the next 5 years

Worldwide HPC-Enabled AI Forecast (ML, DL, & Other AI) Server Revenue ($M)								
	2020	2021	2022	2023	2024	2025	2026	CAGR 21-26
ML in HPC	$719	$861	$1,081	$1,243	$1,391	$1,568	$1,859	16.6%
DL in HPC	$263	$364	$532	$708	$919	$1,147	$1,468	32.2%
Other AI in HPC	$57	$75	$104	$132	$173	$226	$292	31.3%
Total AI Server Revenue	$1,039	$1,300	$1,718	$2,083	$2,484	$2,941	$3,619	22.7%
Source: Hyperion Research, 2023								



Today's Agenda

- Earl Joseph: Market Update
- Tom Sorensen: Update on the Intersection of AI and HPC
- Mark Nossokoff: State of HPC Cloud
- Jaclyn Ludema: Perspective on Sustainability in HPC
- Bob Sorensen: Exascale + Neo Exascale: What's Next?
- Mark Nossokoff: State of HPC Storage and Interconnects
- Bob Sorensen: The Global QC Market: Realistic and Steady Growth Ahead
- Melissa Riddle: HPC Applications and Verticals
- Earl Joseph: Conclusions