HPC Market Update and Observations on Modern Data Workloads

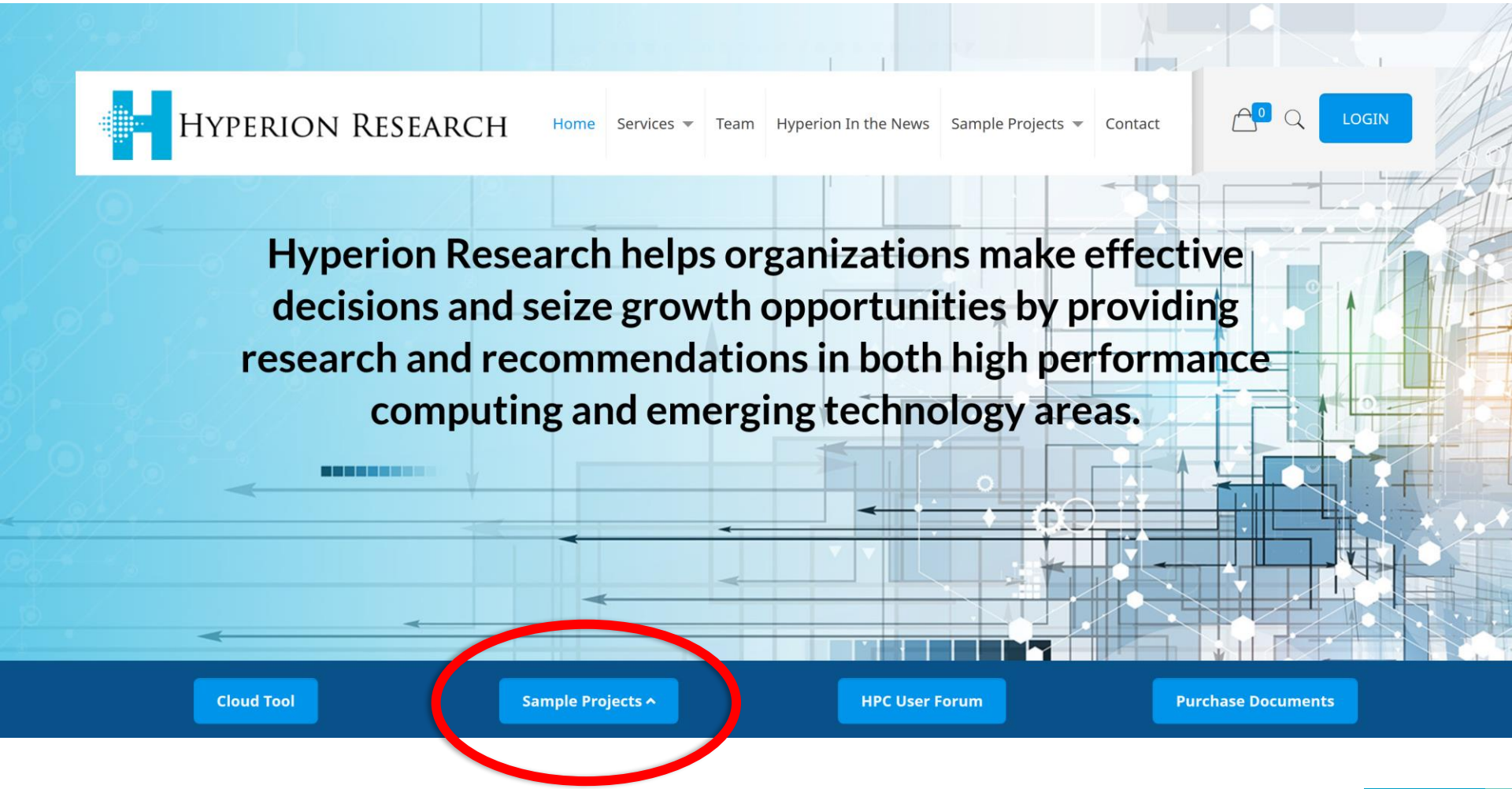
May 2021

www.HyperionResearch.com
www.hpcuserforum.com

Mark Nossokoff

Visit Our Website: www.HyperionResearch.com

Twitter: @HPC_Hyperion

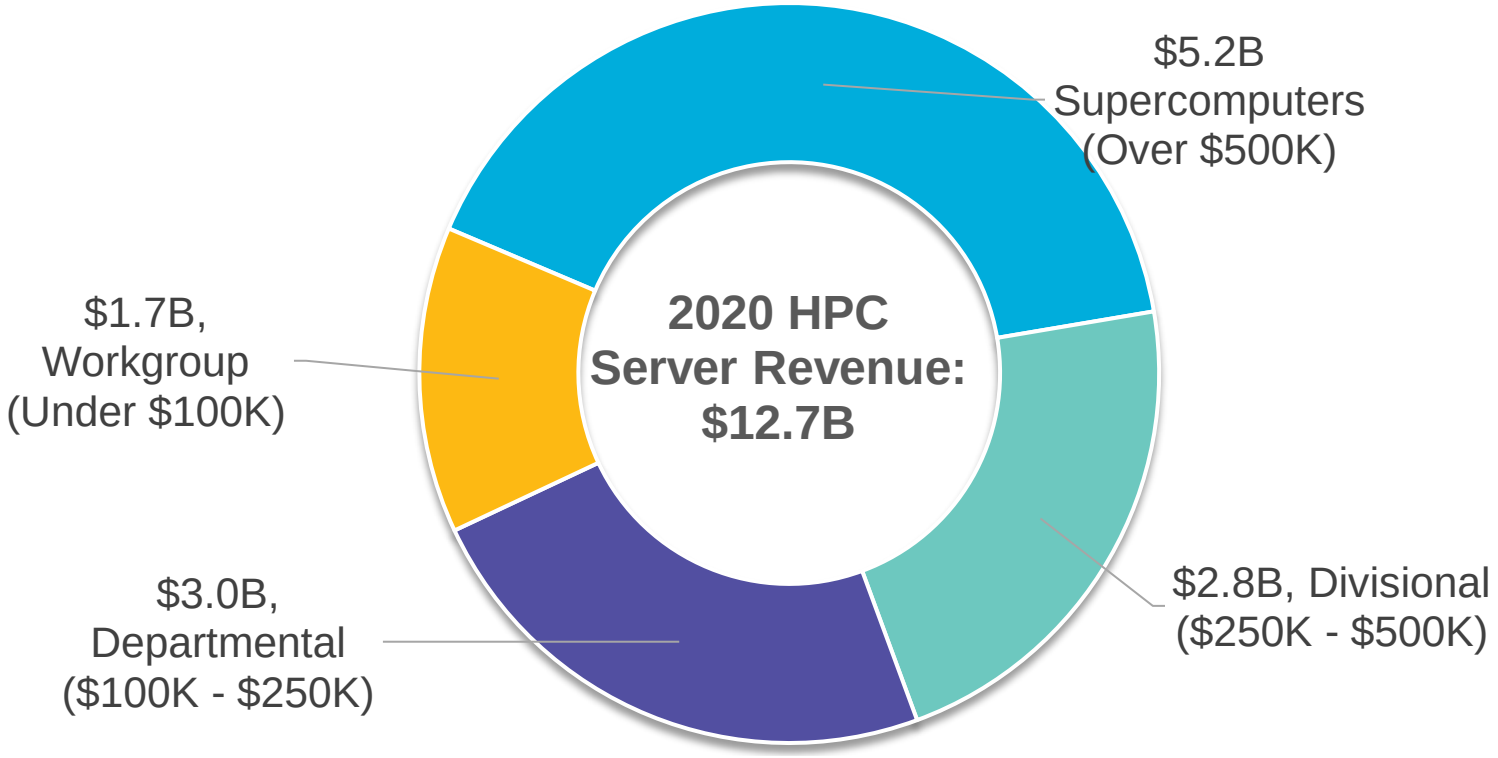


HYPERION RESEARCH

2020 Market Results

WW On-Prem HPC Server Market

$12.7 Billion (down 7.8% YOY) for 2020, using actuals for 3 quarters and estimates for Q4



Worldwide HPC Vendor Server Market Shares ($ Millions)

	2018	2019	3 Quarters 2020	2020 Shares
HPE	4,766	5,095	3,349	36.4%
Dell Technologies	2,833	3,012	2,078	22.6%
Lenovo	957	891	659	7.1%
IBM	970	498	321	3.5%
Cray	313	106	-	0.0%
Fujitsu	269	295	196	2.1%
NEC	201	216	140	1.5%
Inspur	781	860	708	7.7%
Penguin	244	249	172	1.9%
Atos	270	398	182	2.0%
Sugon	462	442	333	3.6%
Other	1,612	1,650	1,072	11.6%
Grand Total	**13,679**	**13,710**	**9,212**	100.0%

The Broader Market Areas ($millions)

The 2020 total spending exceeded $25 billion (excluding cloud spending)

Revenues by the Broader HPC Market Areas		
	2019	2020
Server	$13,710	$12,671
Storage	$5,425	$5,108
Middleware	$1,613	$1,501
Applications	$4,687	$4,347
Service	$2,237	$2,033
Total Revenue	$27,673	$25,660
Source: Hyperion Research, February 2021		

HYPERION RESEARCH

Updated On-Prem HPC Forecasts

HPC On-Prem Server Forecast
($millions)

The overall CAGR is now 7.6%

- **2020 is projected to <u>decline</u> by ~7% to 8%**
  - Forecasts now on a quarterly update cadence

World Wide Overall Technical Computer Market Revenue								
	2018	2019	2020	2021	2022	2023	2024	CAGR 19-24
New January 2021	$13,675	$13,710	$12,671	$14,097	$16,683	$18,813	$19,758	7.6%
October 2020	$13,675	$13,710	$11,846	$13,295	$15,817	$17,942	$19,044	6.8%
Pre-Covid May 2020	$13,683	$13,713	$14,484	$15,658	$18,457	$19,940	$20,844	8.7%
Previous June 2019	$13,706	$14,495	$15,780	$17,376	$18,983	$19,947		
Source: Hyperion Research, January 2021								

On-Prem Forecasts For The Broader Market Areas ($millions)

Storage is expected to grow the most at a 9.1% CAGR

Revenues by the Broader HPC Market Areas							
	2019	2020	2021	2022	2023	2024	CAGR 19-24
Server	$13,710	$12,671	$14,097	$16,683	$18,813	$19,758	7.6%
Storage	$5,426	$5,105	$5,737	$6,873	$7,945	$8,406	9.1%
Middleware	$1,613	$1,500	$1,671	$2,004	$2,275	$2,404	8.3%
Applications	$4,689	$4,345	$4,725	$5,540	$6,144	$6,339	6.2%
Service	$2,238	$2,032	$2,164	$2,492	$2,711	$2,742	4.1%
Total Revenue	$27,677	$25,653	$28,394	$33,592	$37,889	$39,648	7.5%
Source: Hyperion Research, January 2021							

HPC On-Prem Server Forecast By Application Area ($millions)

The overall CAGR is now 7.6%

WW High-Performance Systems Revenue by Applications							
	2019	2020	2021	2022	2023	2024	CAGR 19-24
Bio-Sciences	$1,457	$1,327	$1,304	$1,625	$1,830	$1,927	5.7%
CAE	$1,721	$1,572	$1,585	$1,964	$2,214	$2,327	6.2%
Chemical Engineering	$170	$155	$164	$194	$219	$228	6.0%
DCC & Distribution	$825	$746	$725	$897	$1,014	$1,054	5.0%
Economics/Financial	$709	$651	$664	$856	$966	$1,007	7.3%
EDA / IT / ISV	$821	$751	$741	$963	$1,087	$1,134	6.7%
Geosciences	$970	$872	$894	$1,068	$1,203	$1,269	5.5%
Mechanical Design	$052	$047	$052	$060	$068	$071	6.5%
Defense	$1,472	$1,374	$1,402	$1,777	$2,006	$2,100	7.4%
Government Lab	$2,417	$2,307	$3,528	$3,502	$3,946	$4,292	12.2%
University/Academic	$2,301	$2,130	$2,272	$2,800	$3,128	$3,160	6.5%
Weather	$639	$591	$606	$762	$859	$899	7.1%
Other	$155	$149	$160	$214	$275	$291	13.5%
Total Revenue	$13,710	$12,671	$14,097	$16,683	$18,813	$19,758	7.6%
Source: Hyperion Research, January 2021							

The Exascale Market (System Acceptances)

~30 systems and over $10 billion in value

Exascale and Near-Exascale Systems (2021 to 2026)							
28 - 38 Systems, ~$10-$15B in Value							
Year Accepted	**China**	**EU & UK**	**Japan**	**US**	**Other Countries***	**Total Systems**	**Total Value**
2021	1 or 2 near-exascale systems ~$400M each	1 pre-exascale system ~$150M	1 exascale system Fugaku ~$1B	--	--	3-4	$1.6B-$2.0B
2022	1 or 2 exascale systems ~$350M - $400M each	2 pre-exascale systems ~$150M each	1 near-exascale system ~$200M	2 exascale systems ~$550M each	--	6-7	$2.0B-$2.4B
2023	1 or 2 exascale system ~$350M - $400M each	1 or 2 exascale systems ~$375M	1 near-exascale system ~$150M	1 exascale system ~600M	--	4-6	$1.5B-$2.3B
2024	1 exascale system ~$350M - $400M each	1 or 2 exascale systems ~$375M	?	1 or 2 exascale systems ~$500M each	1 exascale system ~$250M	4-6	$1.5B-$2.4B
2025	2 exascale systems ~$350M - $400M each	1 exascale systems ~$375M	1 exascale system ~$150M	1 or 2 exascale systems ~$500M each	1 exascale system ~$200M	6-7	$1.9B-$2.5B
2026	1 or 2 exascale systems ~$350M - $400M each	1 or 2 exascale systems ~$375M each	?	2 exascale systems ~$500M each	1 or 2 exascale systems ~$200M each	5-8	$2.0B-$3.0B
Total	**7-11**	**7-10**	**4+**	**7-9**	**3-4**	**28-38**	**$10B-$15B**

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

Source: Hyperion Research, July 2020

Key Buying Requirements For HPC

***#1 = price/performance** (for running their specific applications) and performance on their specific applications*

Top Criteria For Next Purchase	
Price	83%
Application Performance	61%
Security	25%
Faster CPUs	25%
AI-Big Data Capabilities	22%
Interconnect Performance	16%
Quality	15%
Accelerators	14%
Storage	11%
Memory Bandwidth	10%
Compatibility with Current Systems	10%
Source of Open Source Software	4%
Other	3%
Source: Hyperion Research 2020	

Barriers For Buying More On-prem

#1 = budgets

Top Barriers to Expanding Purchases	
Financial barriers — budgets, system costs, other costs	81%
Power & cooling cost	43%
Space limitations	30%
Difficulties related to scaling/moving our work up to an HPC server	29%
Lack of knowledge, or skilled HPC/Technical computing support staff	25%
Lack of support by management	21%
Ease-of-use issues: e.g. lack of system management software	21%
3rd party applications costs	18%
Programming hurdles with hybrid environments	16%
Lack of application availability	7%
Other	9%
Source: Hyperion Research 2020	

HYPERION RESEARCH

Using Clouds For HPC Workloads

HPC Cloud Forecast

HPC cloud spend projected to reach ~$9B by 2024

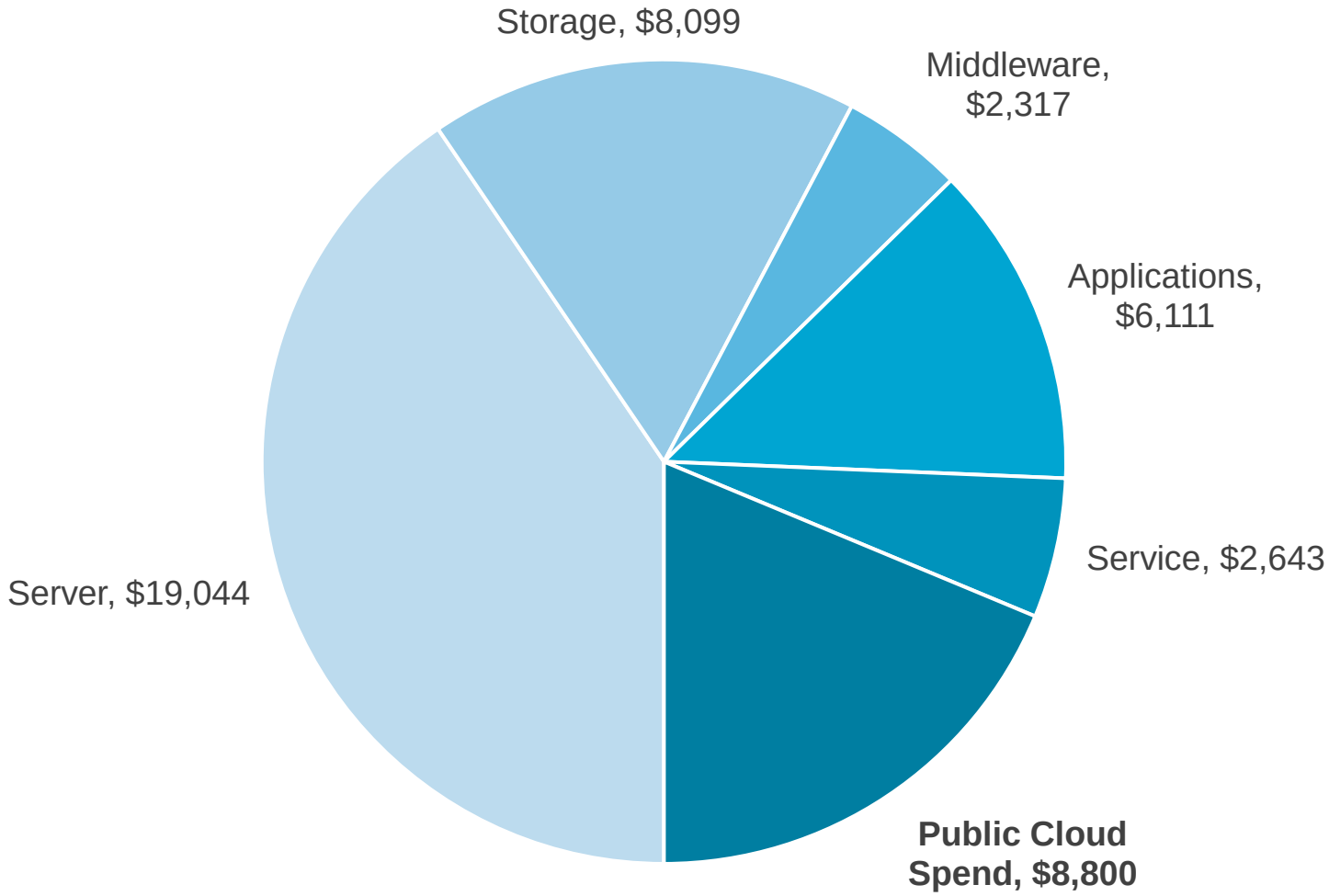
($M)	2018	2019	2020	2021	2022	2023	2024	CAGR '19-'24
NEW 2020 HPC Cloud Forecast	2,466	3,910	4,300	5,300	6,400	7,600	8,800	17.6%
2019 HPC Cloud Forecast	2,466	3,910	4,262	5,135	6,182	7,418	-	24.6% ('18-'24)

- **This forecast includes covid-19 impacts, which has accelerated cloud adoption even more**
- **HPC in the cloud is expected to grow more than 2.5 times faster than the on-prem HPC server market**

Note: This forecast is for public cloud computing, and is from the perspective of end-user spending in the cloud

2024 Total HPC Market with Cloud

Cloud spend brings 2024 HPC forecast to $47 billion

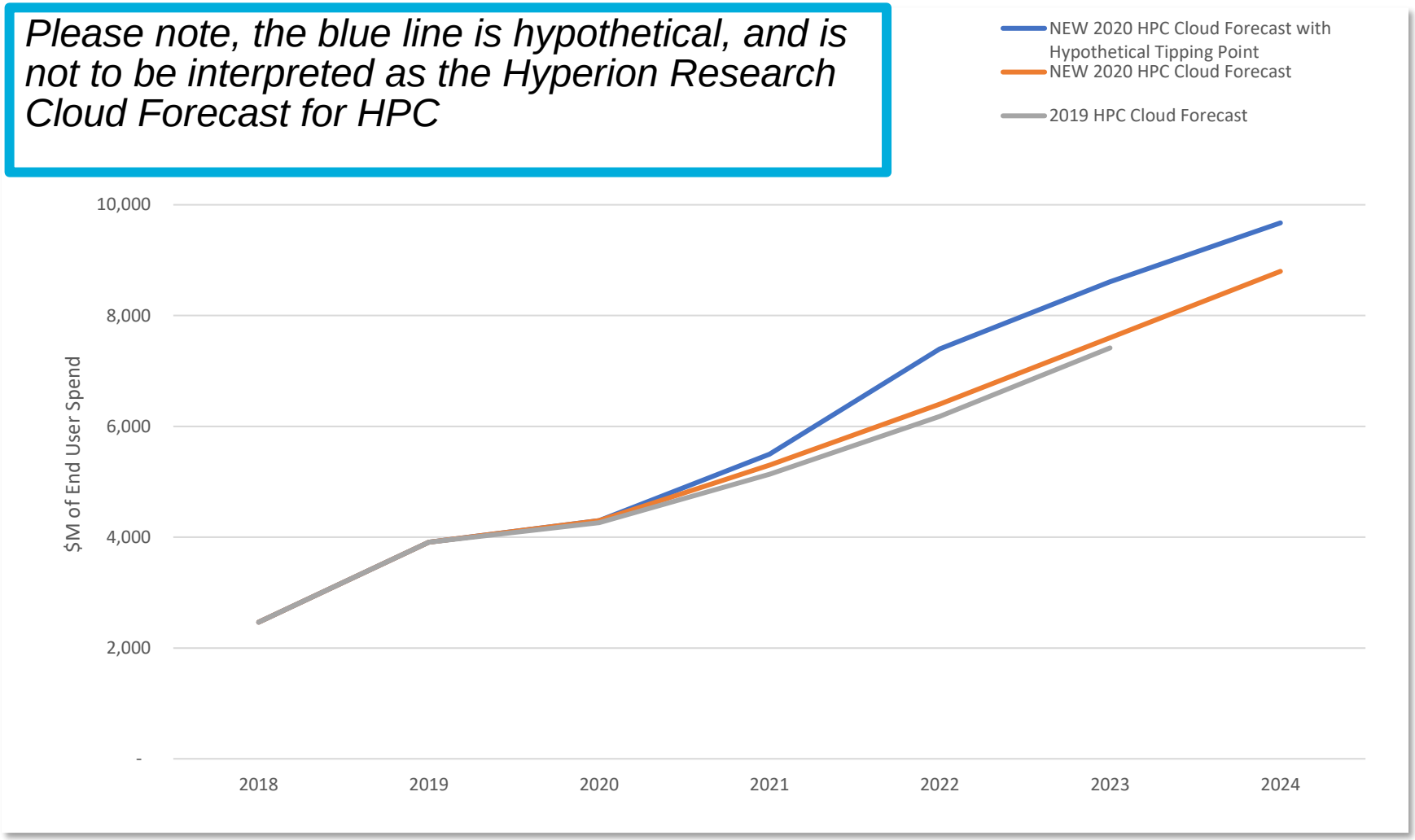


Note: All values are in millions

Cloud Forecast Scenarios

An example showing a potential new tipping point

Please note, the blue line is hypothetical, and is not to be interpreted as the Hyperion Research Cloud Forecast for HPC

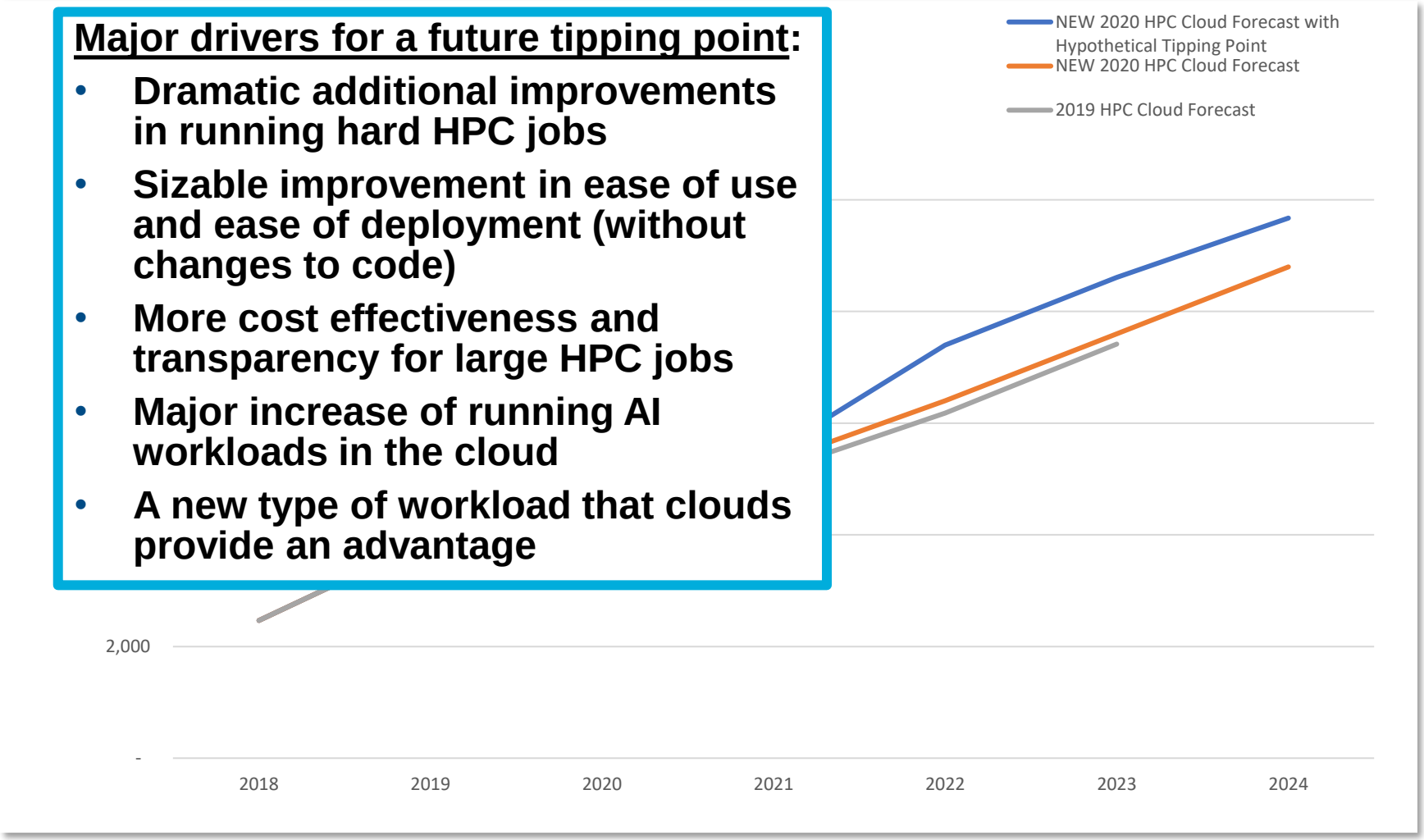


Cloud Forecast Scenarios

An example showing a potential new tipping point

Major drivers for a future tipping point:

- **Dramatic additional improvements in running hard HPC jobs**
- **Sizable improvement in ease of use and ease of deployment (without changes to code)**
- **More cost effectiveness and transparency for large HPC jobs**
- **Major increase of running AI workloads in the cloud**
- **A new type of workload that clouds provide an advantage**



Cloud Forecast by Vertical

Bio-sciences and CAE early cloud adopting verticals

($M)	2018	2019	2024	2019-2024 CAGR
Bio-Sciences	$778	$1,230	$2,453	14.8%
CAE	$469	$733	$1,540	16.0%
Chemical Engineering	$62	$98	$211	16.6%
DCC & Distribution	$141	$222	$519	18.5%
Economics/Financial	$123	$195	$430	17.2%
EDA	$178	$285	$677	18.9%
Geosciences	$148	$240	$660	22.4%
Mechanical Design	$12	$20	$44	17.5%
Defense	$185	$296	$705	18.9%
Government Lab	$173	$274	$625	17.9%
University/Academic	$123	$197	$528	21.8%
Weather	$26	$42	$220	39.0%
Other	$49	$79	$188	18.9%
Total	$2,466	$3,910	$8,800	17.6%

Cloud Forecast by Region and Competitive Segments

HPC Cloud Spending by Region

($M)	2018	2019	2020	2021	2022	2023	2024	CAGR '19-'24
North America	1,020	1,618	1,733	2,078	2,441	2,816	3,166	14.4%
EMEA	750	1,189	1,321	1,646	2,009	2,411	2,820	18.9%
APAC (with Japan)	618	979	1,098	1,380	1,699	2,055	2,424	19.9%
ROW	79	125	148	195	252	318	390	25.7%
Worldwide	**2,466**	**3,910**	**4,300**	**5,300**	**6,400**	**7,600**	**8,800**	**17.6%**

Source: Hyperion Research, March 2021

HPC Cloud Spending by Competitive Segment

($M)	2018	2019	2020	2021	2022	2023	2024	CAGR '19-'24
Supercomputer	429	681	753	933	1,134	1,354	1,576	18.3%
Divisional	719	1,140	1,247	1,529	1,837	2,170	2,499	17.0%
Departmental	679	1,076	1,177	1,443	1,733	2,047	2,357	17.0%
Workgroup	639	1,013	1,123	1,394	1,696	2,030	2,368	18.5%
Total	**2,466**	**3,910**	**4,300**	**5,300**	**6,400**	**7,600**	**8,800**	**17.6%**

Source: Hyperion Research, March 2021

New Drivers of Cloud Usage

Cloud cost-effectiveness increasing for some jobs

- **Cost has always been an issue, but as the cloud becomes more friendly to a wider set of HPC workloads, it has become more cost-effective to running certain workloads in cloud platforms**
- **A part of this cost analysis draws from characteristics of running workloads not tied to monetary cost, like:**
  - Queue times on-prem can be very long
  - Cloud platforms allow users to run workloads on a variety of hardware; some technologies are more efficient for a workload than on-prem deployments
  - Cloud platforms allow for scaling beyond what may be capable on-prem

Workgroup User Cloud Adoption

Sizeable adoption increase over past few years

- **Workgroup users (machines under $100k) never fully recovered from 2008 economic recession and have not recovered to growth rates consistent with rest of market**
- **Recent studies suggest that the slow growth in the workgroup segment is due, in part, to their increased spending in the cloud**
- **The value proposition of cloud for the lower end of the market, both based on economics and the global pandemic situation, will likely propel further workgroup cloud adoption for HPC**

AI-Cloud Adoption

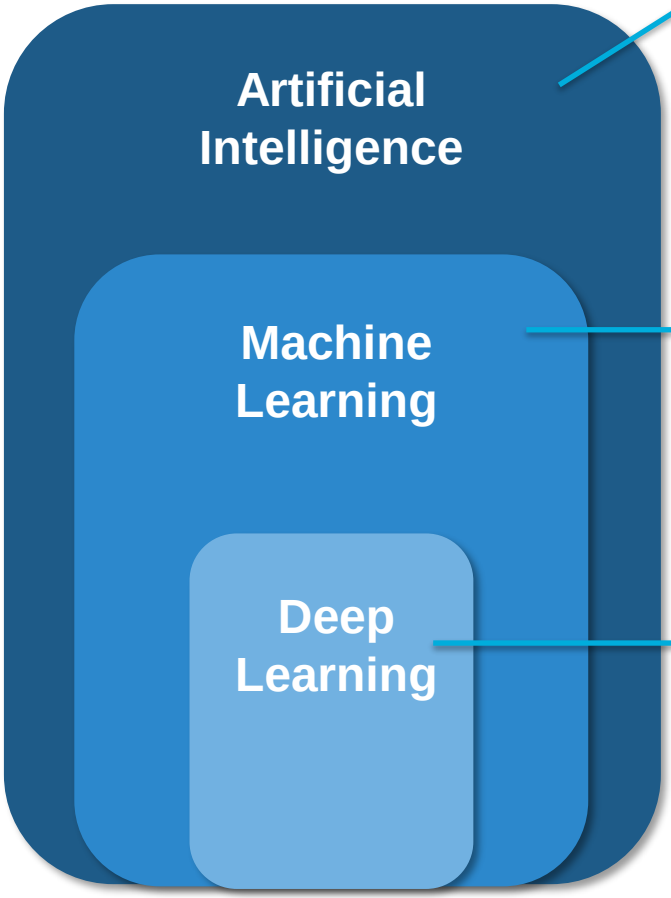
Users run many HPC-enabled AI workloads in the cloud

- **Recent data suggests that the average HPC cloud user anticipates running 1/5 of their HPC-enabled AI workloads in the cloud in the next year**
- **This increase is due to:**
  - Access to hardware & software not available on-prem
  - Access to data stored or collected in the cloud
    - This includes public data sets stored in the cloud, IoT and other sensor data collected in the cloud, or simulation-generated data stored in the cloud

HYPERION RESEARCH

HPDA/AI Market Update

Hyperion Research Definitions
AI: Machine Learning, Deep Learning



Artificial Intelligence (AI): a broad, general term for the ability of computers to do things human thinking does (but NOT to think in the same way humans think). AI includes machine learning, deep learning and other methodologies.

Machine learning (ML): a process where examples are used to train computers to recognize specified patterns, such as human blue eyes or numerical patterns indicating fraud. The computers are unable to learn beyond their training and human oversight is needed in the recognition process. The computer follows the base rules given to it.

Deep Learning (DL): an advanced form of machine learning that uses digital neural networks to enable a computer to go beyond its training and learn on its own, without additional explicit programming or human oversight. The computer develops its own rules.

High Growth Areas: HPDA-AI

HPDA is growing faster than overall HPC market; AI subset growing faster than all HPDA

Table 1

Forecast: Worldwide HPC-Based AI Revenues vs Total HPDA Revenues ($ Millions)

	2018	2019	2020	2021	2022	2023	2024	CAGR '19-'24
COVID-Impacted HPC Server Revenues	$13,679	$13,710	$12,671	$14,097	$16,684	$18,814	$19,758	7.6%
HPDA Server Revenues	$3,153	$3,598	$3,499	$4,500	$5,467	$6,650	$7,800	16.7%
HPC-Based AI (ML, DL & Other)	$747	$918	$1,039	$1,500	$2,010	$2,745	$3,800	32.9%

Source: Hyperion Research, 2021

Table 2

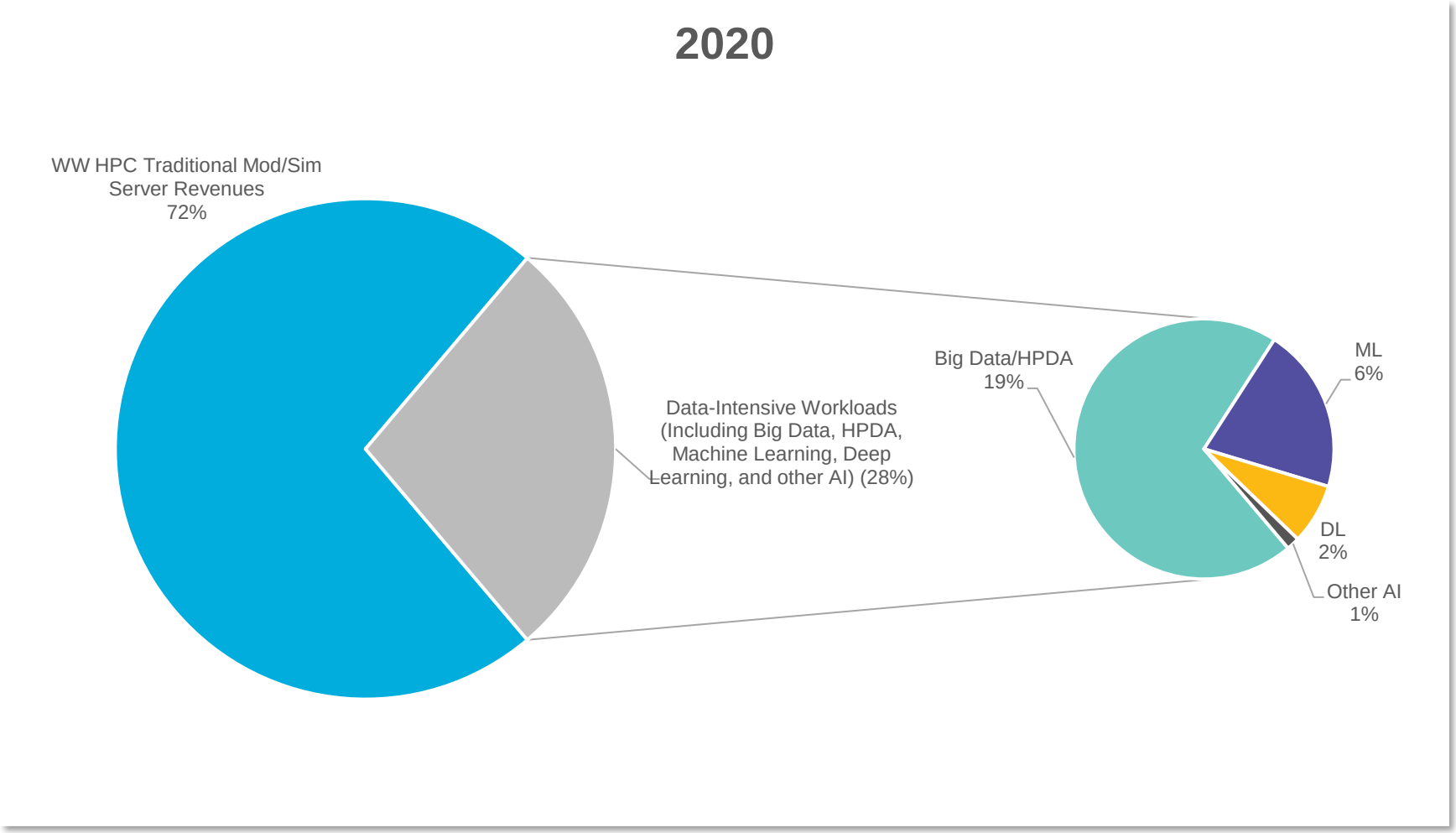
Forecast: Worldwide ML, DL & Other AI HPC-Based Revenues ($ Millions)

	2018	2019	2020	2021	2022	2023	2024	CAGR '19-'24
ML in HPC	$532	$667	$719	$1,039	$1,366	$1,816	$2,445	29.7%
DL in HPC	$177	$209	$263	$390	$560	$804	$1,200	41.8%
Other AI in HPC	$38	$42	$57	$71	$84	$125	$155	29.8%
Total	$747	$918	$1,039	$1,500	$2,010	$2,745	$3,800	32.9%

Source: Hyperion Research, 2021

HPDA/AI Server Market within HPC

Traditional mod/sim servers make up more than 70% of market



HPDA Systems

1/5 use a separate system today, ¼ plan to use a separate system in 6-18 months

Systems Used for HPDA Applications

Q: For your CURRENT AI and Big Data analytics workloads: What do you run these workloads on? IN THE NEXT 6 to 18 MONTHS: What do you plan to run your AI and Big Data analytics workloads on?

System Type	Today (% of Respondents)	Next 6-18 Months (% of Respondents)
The same HPC system used for simulation workloads	68.0%	64.4%
A separate HPC system or Big Data appliance	20.6%	26.3%
Not currently planned	11.3%	9.3%

n=194

Source: Hyperion Research, 2020

Important Commercial Use Cases

Most will take longer to mature than previously thought

Precision Medicine

Automated Driving Systems

Fraud and Anomaly Detection

Affinity Marketing

Business Intelligence

Cyber Security

IoT/Edge/Smart Cities

AI is Still Near the Start

- **Today: Special (Weak) AI**
  - Many observations but few choices
  - “One trick dogs”: 10 AI solutions in a box to solve 10 problems
  - Rudimentary training/inferencing
  - Short on real-world data
  - Examples:
    - Image & voice recognition
    - Early automated driving
    - Reading an MRI
- **Future: General (Strong) AI**
  - Many observations, many choices
  - Versatile decision-makers capable of serious experiential learning
  - More intelligent training/inferencing
  - High-volume synthetic data
  - Examples:
    - Discerning human motivation
    - Mature automated driving
    - Diagnosing/”curing” a cancer

Key Takeaways #1

- **AI is about machines guessing (inferencing) much faster than humans**
  - Human intuition is far better at this but far slower
  - For the foreseeable future, machines will mainly carry out the tedious AI work and hand off the challenging work to humans
- **AI ethics and liability activities center around the position of humans vs. machines and the HMI**
- **Current issues (transparency, biased input) could slow but not stop AI momentum**
  - Our studies show almost all HPC sites are involved in AI

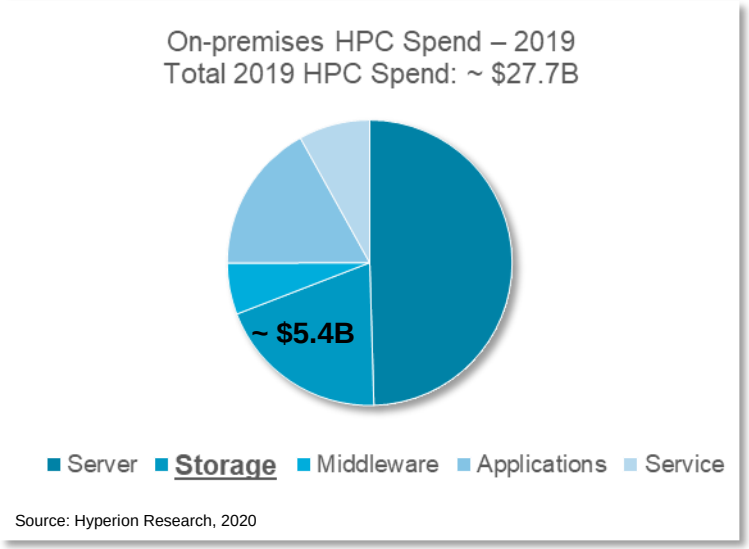
Key Takeaways #2

- **HPC is crucial at the forefront of AI R&D**
  - HPC market growth + AI potential is motivating vendors
- **HPC innovations heavily influence mainstream AI:**
  - Algorithmic sophistication
  - Parallelization
  - Clustered servers (“clusters”)
  - CPU-accelerator processing
  - Ultrafast system data rates
  - Capable memory subsystems
- **AI is exiting the peak of the hype cycle**
  - Vendors are less often setting unrealistic expectations
- **HPC data center & enterprise deployments are different**
  - HPC data center: monolithic, standalone upgrade
  - Enterprise data center: integrate into existing infrastructure and workflow

HYPERION RESEARCH

Storage and Interconnects Market Update

HPC Storage is an Attractive Market



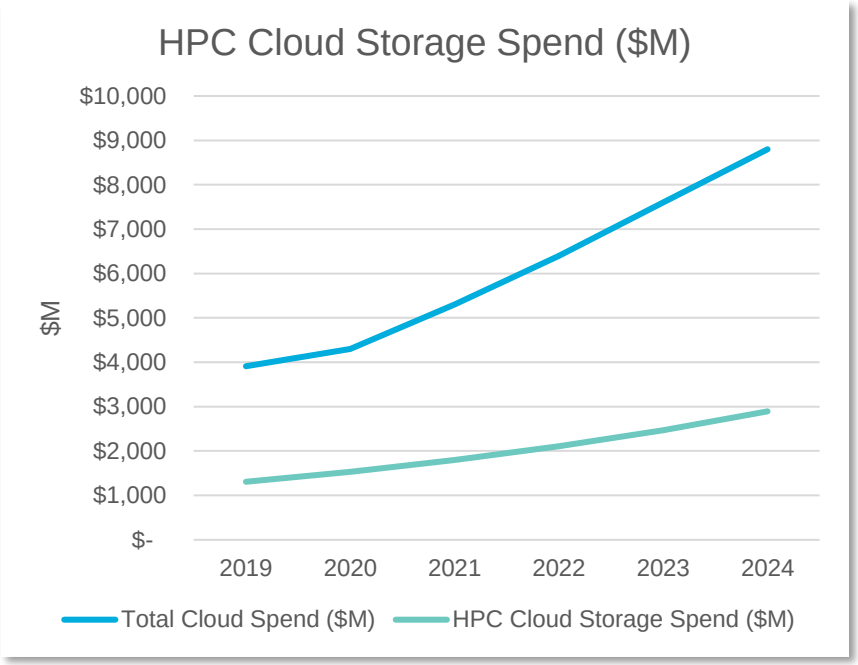
- **Storage historically the highest growth HPC element**
- **Storage represents ~ 20% of HPC spending**
- **For every $1 spent on compute, ~ $0.40 is spent on storage**

Area ($M)	2019	2020	2021	2022	2023	2024	CAGR '19-'24
Server	$13,710	$12,671	$14,097	$16,683	$18,813	$19,758	7.6%
Add-on Storage	**$5,426**	**$5,105**	**$5,737**	**$6,873**	**$7,945**	**$8,406**	**9.1%**
Middleware	$1,613	$1,500	$1,671	$2,004	$2,275	$2,404	8.3%
Applications	$4,689	$4,345	$4,725	$5,540	$6,144	$6,339	6.2%
Service	$2,238	$2,032	$2,164	$2,492	$2,711	$2,742	4.1%
Total Revenue	$27,677	$25,653	$28,394	$33,592	$37,889	$39,648	7.5%

HPC Cloud Storage Forecast

Approaching $3B in 2024

- **Total cloud spend of $3.9B in 2019**
- **~1/3 is storage**
- **17.3% increase in storage spend in 2020**



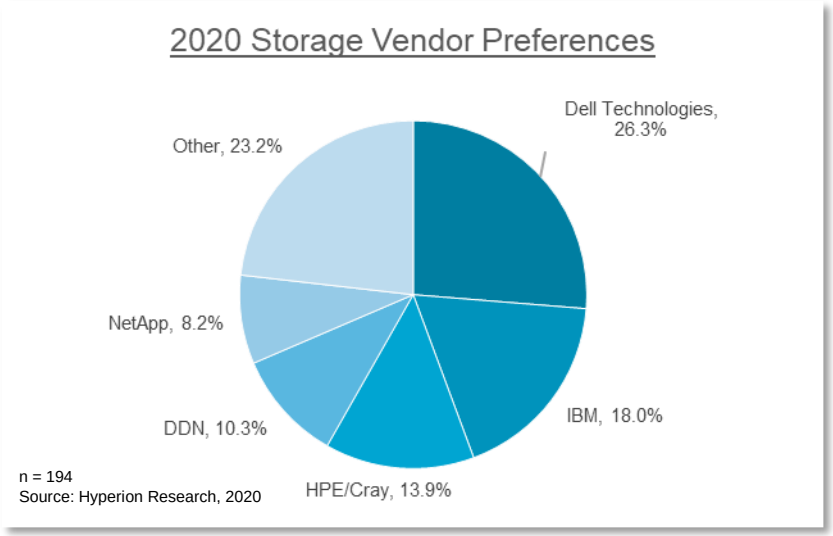
Source: Hyperion Research, 2020

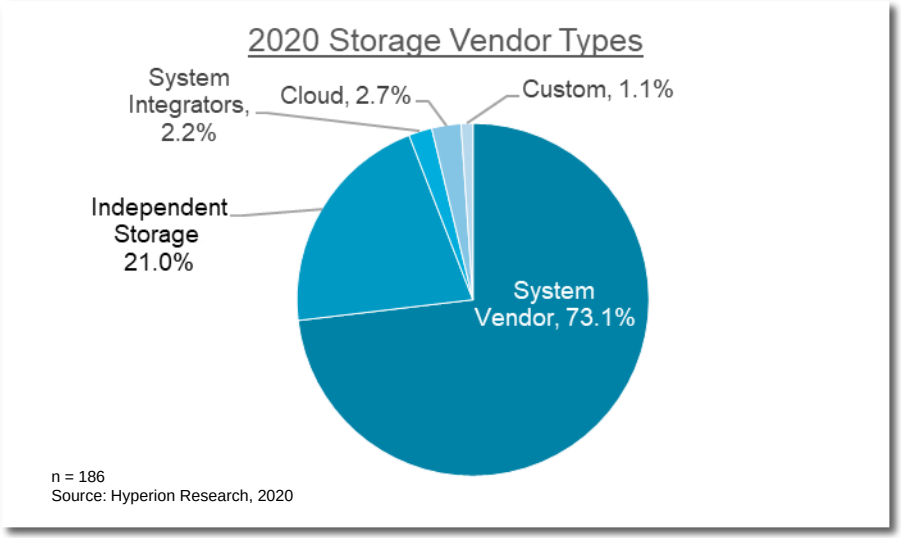
	2019	2020	2021	2022	2023	2024
Total Cloud Spend ($M)	$ 3,910	$ 4,300	$ 5,300	$ 6,400	$ 7,600	$ 8,800
HPC Cloud Storage Spend ($M)	$ 1,303	$ 1,529	$ 1,793	$ 2,104	$ 2,467	$ 2,894

Source: Hyperion Research, 2020

2020 HPC Storage Vendor Preferences

Dell Technologies top preferred HPC storage vendor

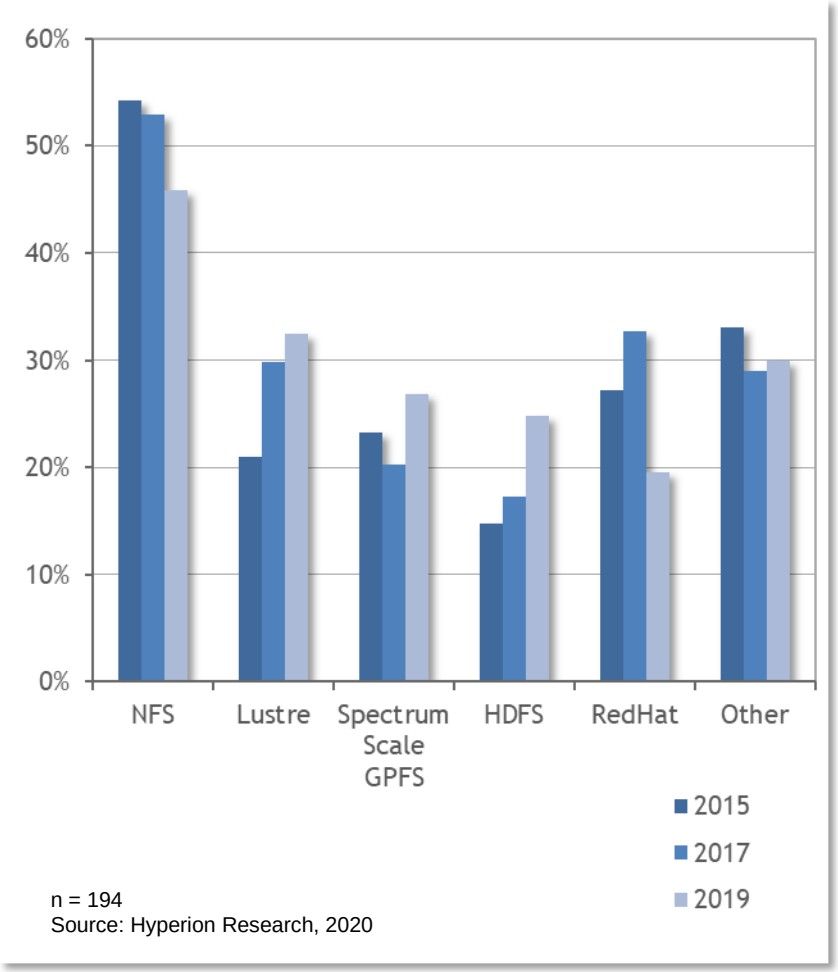




- **Dell Technologies also preferred most in Academia and Industry sectors**
- **IBM second preferred overall**
- **HPE/Cray preferred 3rd overall but tops in Government**
- **DDN 4th overall and top independent storage vendor**

File System Landscape is Changing

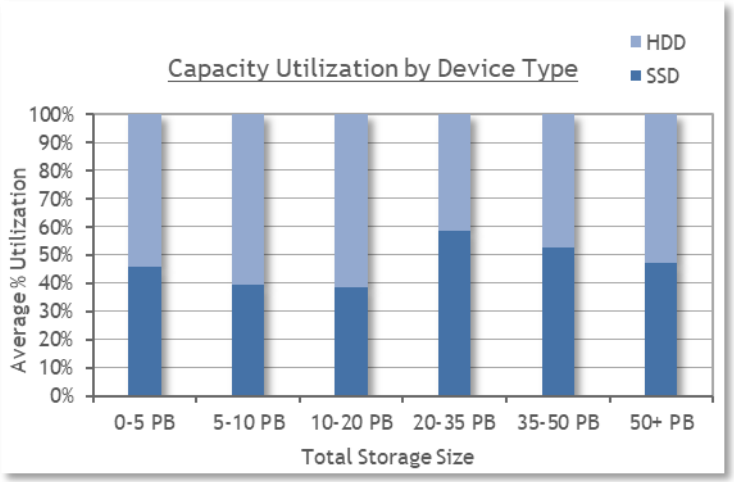
Lustre continues its adoption growth

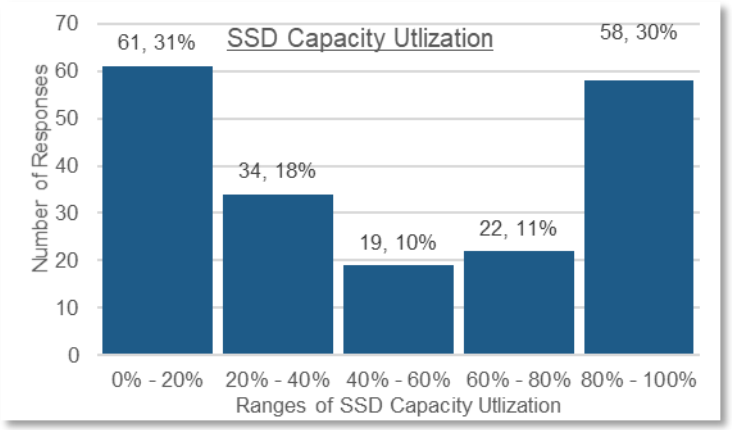


- **Academia and Government tend towards Lustre**
- **Industry tends towards NFS**
- **GFS, CEPH and PanFS in "Other"**
- **Expect shifts to continue as enterprises increasingly adopt HPC-enabled AI**

SSDs Becoming Storage Medium of Choice

Surveyed sites deployed 45% of capacity on SSDs in 2019



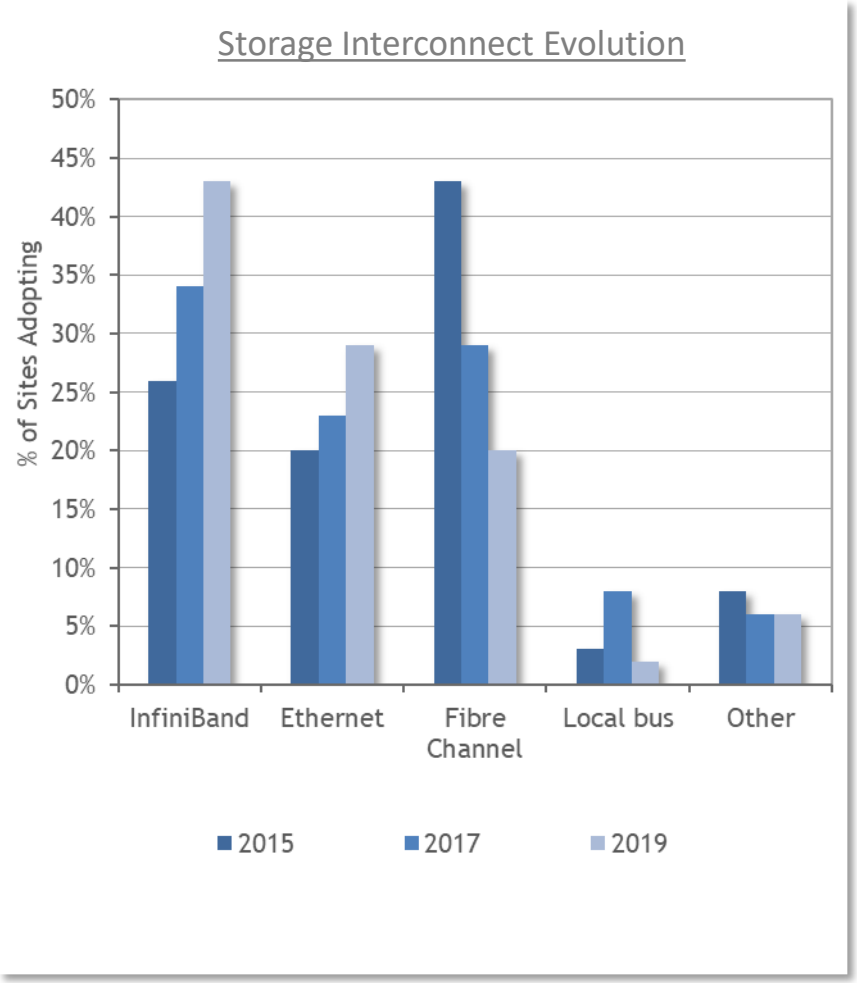


n = 194
Source: Hyperion Research,2020

- **86% of sites deployed flash-based storage**
- **Wide range of media mix**
  - 15% were all flash
  - 71% were hybrid
  - 14% were all HDD
- **~ 60% of the sites have strong tendency (> 80% of respective capacity) towards either SSD or HDD**

Storage Interconnects Entering Spotlight

InfiniBand the leading storage interconnect



- **InfiniBand increasingly being adopted**
- **Ethernet also exhibiting growth**
- **Fibre Channel continues decline**
- **Interconnects becoming increasingly important to deliver the bandwidth required to keep GPU and other accelerator-based nodes fully utilized**

HYPERION RESEARCH

Edge Computing: The Role of HPC

Edge Computing

A Newer Paradigm for Highly Distributed Computing

- **A relatively new approach where some or all of the necessary computation is done directly at or near data sources.**
  - Vehicles and traffic sensors
  - Medical devices
  - Product manufacturing lines
  - Military sites
  - Other data-generating locations
- **Contrasts with norm of sending Big Data to data centers or cloud computing platforms.**

Benefits of Computing at the Edge

- **Faster responses to local issues (lower latency)**
  - Identifying lawbreakers in time for apprehension
  - Enabling vehicle-vehicle communication (e.g., truck fleets)
  - Monitoring patients in hospitals
  - Operating smart homes and factories
  - Reacting quickly to changing battlefield conditions
- **Lower costs**
  - Minimizes moving and storing Big Data in distant facilities
- **Higher autonomy, reliability and (potentially) security**
  - No need to share resources (co-tenancy)
  - No network switches to fail and cause disruptions
  - Less data to transfer over vulnerable networks (but edge devices themselves need to be secure)
- **Scalability**
  - Adding low-cost edge devices can efficiently handle growth in source data.

HPC's Role in Edge Computing

When Wide-Area Situation Awareness/Control Are Needed

- **When wide-area situation awareness/control are needed. Examples:**
  - Smart city functions (traffic congestion, smart power grids)
  - Battlefield operations

- **Generally, only a small subset of data needs to be moved for deeper analysis from the edge to HPC systems in data centers of clouds.**
  - A subset from millions of devices can still be large.
  - "Fog computing" usually refers to small clusters close to the edge.

Observations on Modern Data Workloads

Platform considerations becoming increasingly complex

Area	Traditional	Modern
Workload type	Compute-intensive	Data-intensive
Workload profile	Single workload; multiple, similar workloads	Heterogeneous workloads
Locality	Consolidated on-premises	Distributed between on-prem, cloud, and edge
HW architecture	Single HW architecture	Diverse HW architectures
Users	Technical researchers, engineers, HPC datacenters	Data scientists, enterprise datacenters

HYPERION RESEARCH

THANK YOU!

mnossokoff@hyperionres.com