HPC Market Update and Observations on Big Memory

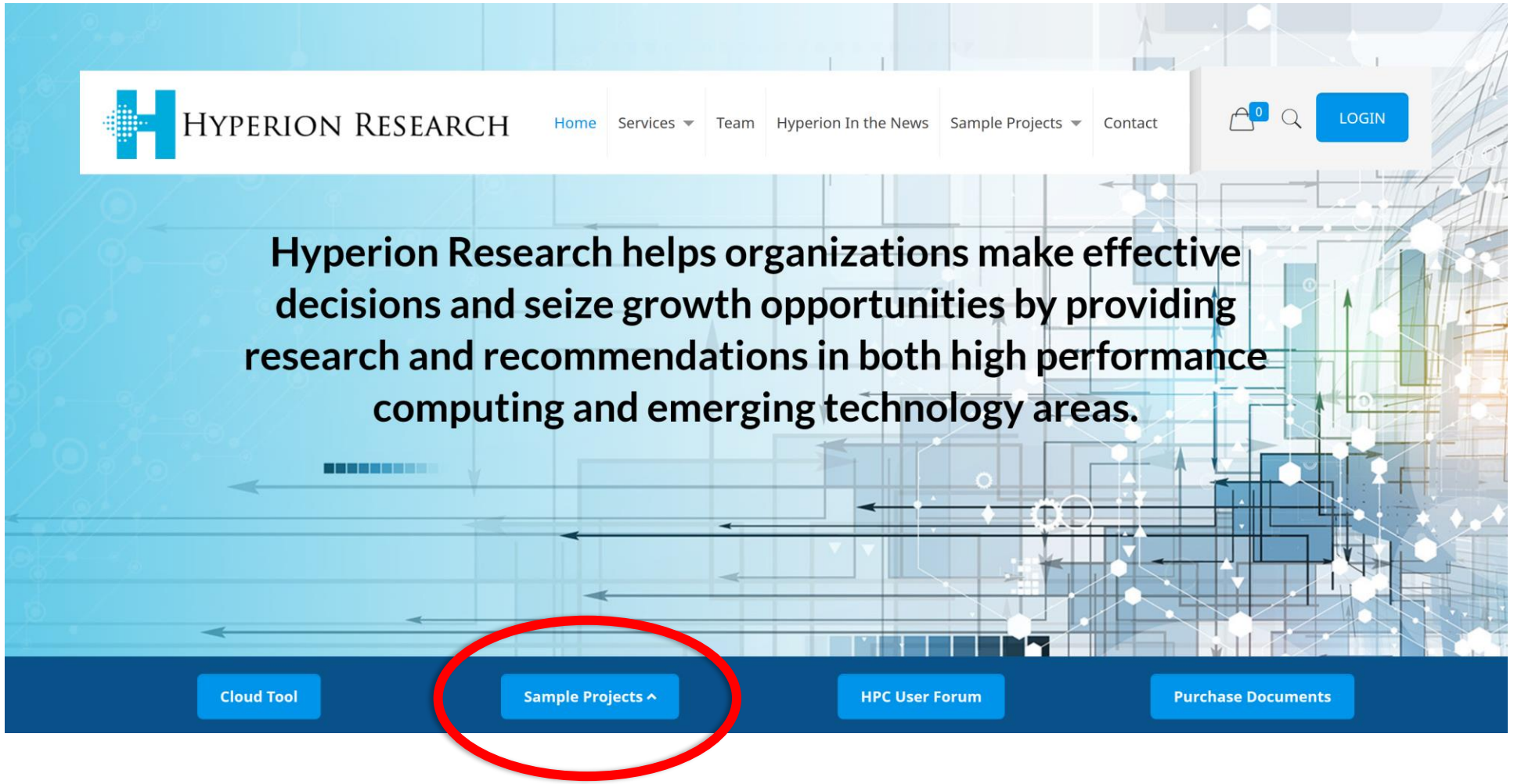
December 10, 2020

Mark Nossokoff
Senior Analyst, Lead Storage Analyst

www.HyperionResearch.com
www.hpcuserforum.com

Visit Our Website: www.HyperionResearch.com

- **Twitter: @HPC_Hyperion**



HYPERION RESEARCH

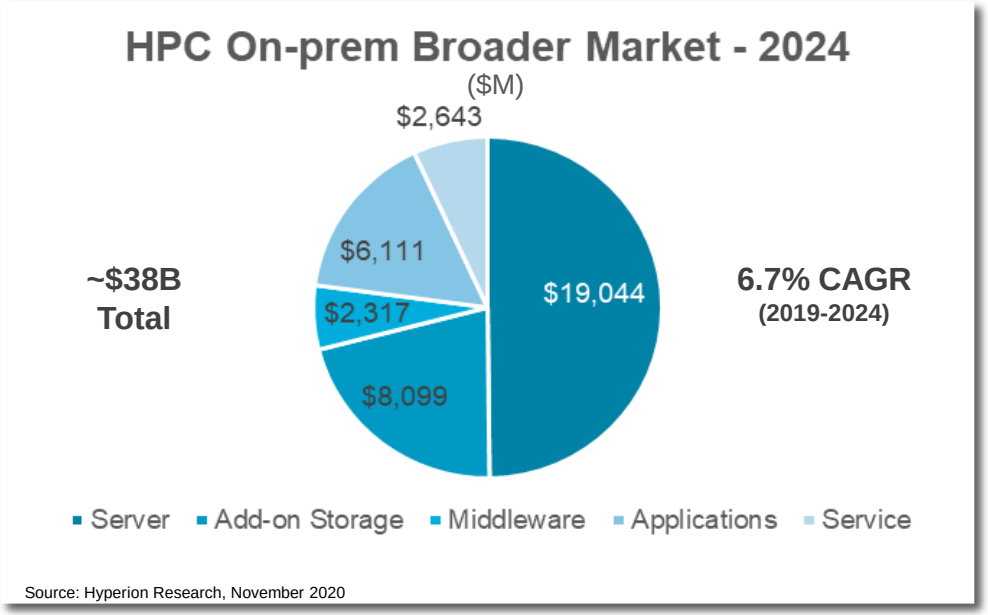
HPC Market Update

On-prem Broader Market Forecast

Storage is expected to grow the most at 8.3%

Market Area ($M)	2019	2020	2021	2022	2023	2024	CAGR 19-24
Server	$13,710	$11,846	$13,295	$15,817	$17,942	$19,044	6.8%
Storage	$5,427	$4,772	$5,410	$6,519	$7,577	$8,099	8.3%
Middleware	$1,613	$1,402	$1,576	$1,902	$2,171	$2,317	7.5%
Applications	$4,689	$4,062	$4,455	$5,258	$5,862	$6,111	5.4%
Service	$2,239	$1,899	$2,040	$2,366	$2,587	$2,643	3.4%
Total Revenue	**$27,678**	**$23,981**	**$26,774**	**$31,862**	**$36,138**	**$38,214**	**6.7%**

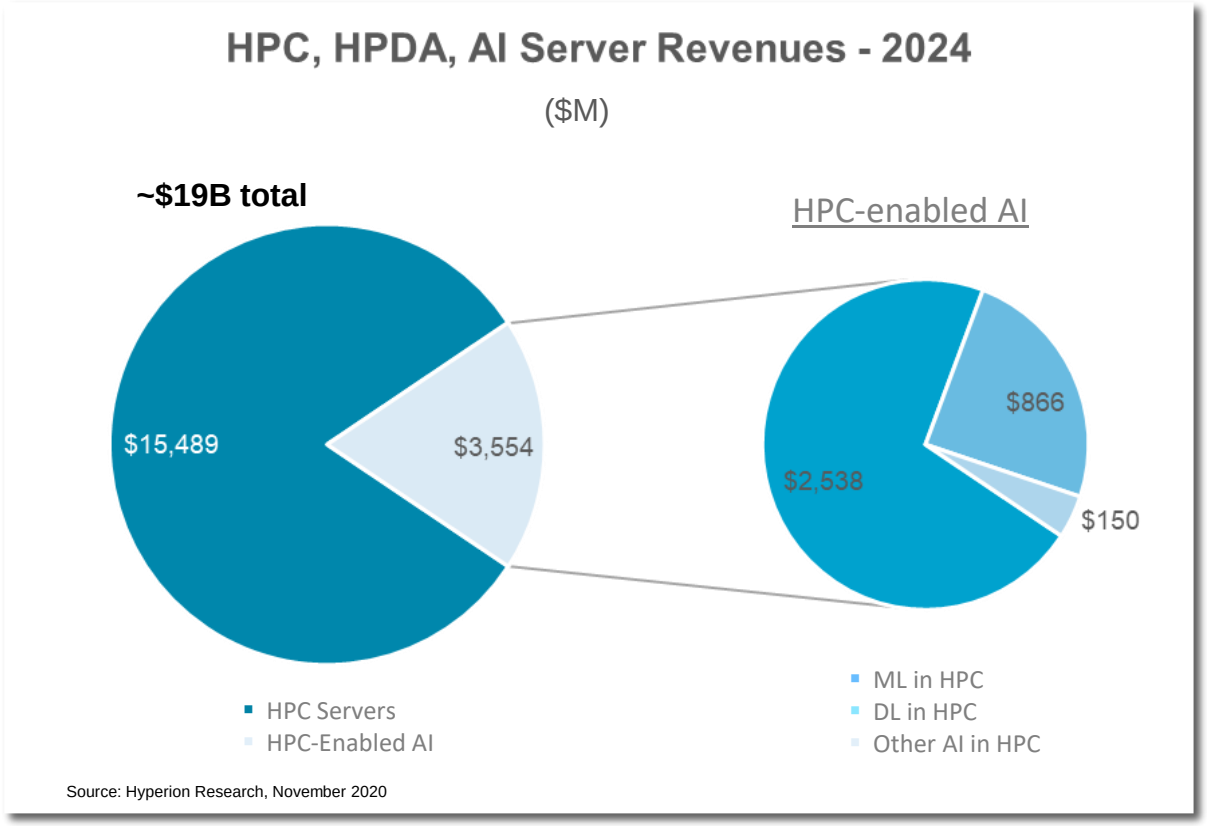
Source: Hyperion Research, November 2020



- **Forecast incorporates Covid-19's impact**
  - Downside pressure
    - Delayed product shipments
    - Delayed revenues
    - Delayed orders
    - Decline of 11.5% in first half of 2020
    - Forecasting Y/Y decline of 14% for 2020
  - Upside momentum
    - Demand to combat Covid-19
    - Increase in HPC workloads running in the public cloud
    - Expected recovery in mid 2021

HPC-enabled On-prem AI Server Forecast

HPC-Enabled AI Growth ~ 5x Overall HPC Server Growth 2019-2024



HPC Servers	• 6.8% growth
HPDA Servers	• 15.8% growth
HPC-enabled AI Servers	• 31.1% growth

HPC On-Prem Server Forecast By Application Area

Government, Academic, CAE/Manufacturing and Bio-sciences >50% of market

$M	2019	2020	2021	2022	2023	2024	CAGR 19-24
Bio-Sciences	$1,457	$1,239	$1,226	$1,536	$1,739	$1,850	4.9%
CAE	$1,721	$1,468	$1,492	$1,859	$2,110	$2,242	5.4%
Chemical Engineering	$170	$145	$154	$185	$209	$220	5.2%
DCC & Distribution	$825	$696	$681	$857	$970	$1,017	4.3%
Economics/Financial	$710	$608	$623	$818	$924	$972	6.5%
EDA / IT / ISV	$822	$702	$696	$918	$1,037	$1,091	5.8%
Geosciences	$969	$815	$843	$1,010	$1,151	$1,231	4.9%
Mechanical Design	$52	$044	$049	$057	$065	$068	5.6%
Defense	$1,472	$1,284	$1,317	$1,692	$1,916	$2,027	6.6%
Government Lab	$2,418	$2,161	$3,352	$3,314	$3,759	$4,127	11.3%
University/Academic	$2,301	$1,993	$2,141	$2,647	$2,981	$3,053	5.8%
Weather	$639	$553	$570	$724	$819	$866	6.3%
Other	$155	$139	$151	$202	$261	$279	12.5%
Total Revenue	$13,710	$11,846	$13,295	$15,817	$17,942	$19,044	6.8%

Source: Hyperion Research, November 2020

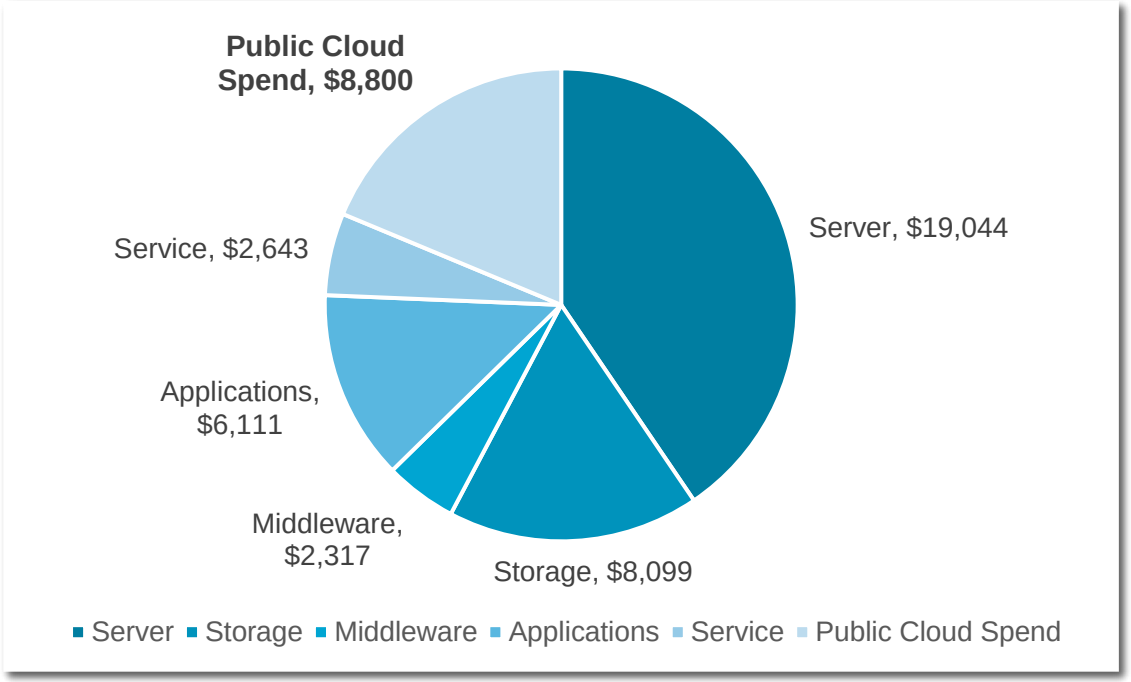
HPC Usage in the Cloud

Expected to incrementally add $8.8B to on-prem HPC spend in 2024

Market Area ($M)	2019	2020	2021	2022	2023	2024	CAGR 19-24
Server	$13,710	$11,846	$13,295	$15,817	$17,942	$19,044	6.8%
Storage	$5,427	$4,772	$5,410	$6,519	$7,577	$8,099	8.3%
Middleware	$1,613	$1,402	$1,576	$1,902	$2,171	$2,317	7.5%
Applications	$4,689	$4,062	$4,455	$5,258	$5,862	$6,111	5.4%
Service	$2,239	$1,899	$2,040	$2,366	$2,587	$2,643	3.4%
Public Cloud Spend	**$3,910**	**$4,300**	**$5,300**	**$4,600**	**$7,600**	**$8,800**	**17.6%**
Total On and Off Prem Revenue	**$31,588**	**$28,281**	**$32,076**	**$36,462**	**$43,739**	**$47,014**	**8.3%**

Source: Hyperion Research, November 2020

2024 Broader Market Forecast - ~$47B



Source: Hyperion Research, November 2020

Key Buying Requirements For On-prem HPC

Price/performance and overall performance on specific applications the top items

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%
Backwards Compatibility with Current Systems	10%
Source of Open Source Software	4%
Other	3%

HYPERION RESEARCH

Observations on Big Memory and HPC

What is Big Memory?

High capacity, performant, resilient data via memory footprint and accessibility

	Historic perspective on memory	HPC Requirements	Big Memory
Cost	Expensive		Less expensive
Capacity	100s GB memory per server		100s TB memory per server
Resiliency	Volatile		HA Tier
Relationship to Storage	Extension of memory		Data is in memory

- **Persistent Memory + Memory Virtualization Software**

Data Access	Type	Form Factor
	Integrated	n/a
Hot/Active	DRAM	DIMM
	Persistent Memory	DIMM
Warm	SSD	AIC, U.2, M.2, EDSFF
	HDD dual actuator	3.5”
	HDD	3.5”
Cold	Tape	

HPC On-Prem Server Forecast By Application Area

Government, Academic, CAE/Manufacturing and Bio-sciences >50% of market

Application	Revenue ($M)			Processors Shipped (estimated)		
	2019	2024	CAGR 19-24	2019	2024	CAGR 19-24
Bio-Sciences	$1,457	$1,850	4.9%	425,956	534,882	4.7%
CAE / Manufacturing	$1,721	$2,242	5.4%	502,965	648,452	5.2%
Chemical Engineering	$170	$220	5.2%	49,796	63,591	5.0%
DCC & Distribution	$825	$1,017	4.3%	241,401	294,212	4.0%
Economics/Financial	$710	$972	6.5%	206,904	281,127	6.3%
EDA / IT / ISV	$822	$1,091	5.8%	240,322	315,575	5.6%
Geosciences	$969	$1,231	4.9%	283,098	355,851	4.7%
Mechanical Design	$52	$068	5.6%	15,166	19,748	5.4%
Defense	$1,472	$2,027	6.6%	430,349	586,136	6.4%
Government Lab	$2,418	$4,127	11.3%	785,793	1,193,592	8.7%
University/Academic	$2,301	$3,053	5.8%	672,908	882,790	5.6%
Weather	$639	$866	6.3%	186,845	250,432	6.0%
Other	$155	$279	12.5%	45,191	80,660	12.3%
Total	**$13,710**	**$19,044**	**6.8%**	**4,086,694**	**5,507,047**	**6.1%**

Source: Hyperion Research, November 2020

Most amenable to Big Memory

Likely amenable to Big Memory

- **Core counts growing faster than memory capacities**
- **Memory amount per core decreasing**
- **Can memory be efficiently and effectively pooled and utilized?**

Key Buying Requirements For On-prem HPC

Price/performance and overall performance on specific applications the top items

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%
Backwards Compatibility with Current Systems	10%
Source of Open Source Software	4%
Other	3%

Potential areas Big Memory can address

HPC and HPDA/AI Workloads

HPDA/AI workloads changing the status quo of data access

Workload	Use Case	Description
Traditional HPC	Project	• Sometimes referred to as home directories or user files • Used to capture and share final results of the modelling and simulation • Mixture of bandwidth and throughput needs, utilizing hybrid flash, HDD storage solutions
	Scratch	• Workspace capacity used to perform the modelling and simulation • Includes metadata capacity (high throughput [IOs/sec] and flash-based) and raw data capacity and checkpoint writes for protection against system component failure during long simulation runs (high bandwidth [GB/s], traditionally HDD-based but now largely hybrid flash and HDDs
	Archive	• Long-term data retention • Scalable storage without a critical latency requirement • Largely near-line HDD-based systems with a growing cloud-based element. • Typically file or object data types
HPDA/AI	Ingest	• Quickly loading large amounts of data from a variety of different sources such that the data can be tagged, normalized, stored and swiftly retrieved for subsequent analysis • Very high bandwidth (GB/s) performance at scale to sustain retrieving data rates, typically object-based, high-capacity HDD-based and increasingly cloud-based.
	Data Preparation	• Often times referred to as data classification or data tagging, requires a balanced mix of throughput and bandwidth (hybrid flash and HDD storage systems)
	Training	• Utilizing Machine Learning (ML) and/or Deep Learning (DL) to build an accurate model for researchers, engineers and business analysts to use for their research, design and business needs • Requires high throughput (IOs/sec) and low latency for continuous and repetitive computational analysis of the data, typically flash-based storage.
	Inference	• Utilizing the model for experimentation and analysis to derive and deliver the targeted scientific or business insights • Also requires high bandwidth and low latency and typically flash-based, often with a caching layer
	Archive	• Long-term data retention Scalable storage without a critical latency requirement • Largely near-line HDD-based systems with a growing cloud-based element. • Typically file or object data types

- **Traditional HPC**
  - Metadata
    - Small block, random
    - Focus on latency, IOPs
  - Simulation data
    - Large block, sequential
    - Focus on GB/s
  - Historically separate data stores

- **HDPA / AI**
  - Heterogenous I/O profiles
  - Interspersed transfer sizes, access patterns and performance focus
  - Growing dataset sizes

Most amenable to Big Memory

Likely amenable to Big Memory

Closing Observations on Big Data, Big Memory and HPC

HPDA requires massive growth in data consumption and memory sizes

"Traditional" Memory
- Node-based
- Ephemeral
- Transient
- Byte addressable
- Lowest latencies

Opportunity

"Traditional" Storage
- Add-on
- Persistent
- Resilient
- Block addressable
- Longer Latencies

- **Conventional thoughts on memory**
  - Limited amount, expensive, persistent
  - Plentiful, less expensive, but not persistent
- **Consistent feedback from HPC users for most new technologies**
  - Is there enough *[insert resource]* for my *[insert task]?*
  - Is there enough *memory* for my *working dataset size?*
  - How much will my "time to results" be improved?
  - Will it simplify (at least not complicate) system management, data management and workflow?
  - Do I need to change any code?
  - Can I afford the amount of memory I need for my HPC workloads?

HYPERION RESEARCH

QUESTIONS?

Questions or comments are welcome.

mnossokoff@hyperionres.com