

Welcome. The webinar will begin shortly.

Worldwide HPC and AI Market

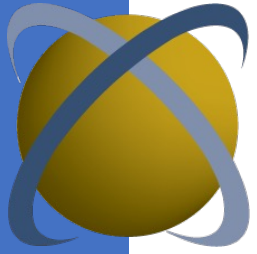
2022 Actuals, Change from 2021, 2023-27 Forecast

Including “Grand Unification”
of Data Center, Hyperscale, Cloud

May 2023



Intersect360
RESEARCH



Worldwide HPC and AI Market

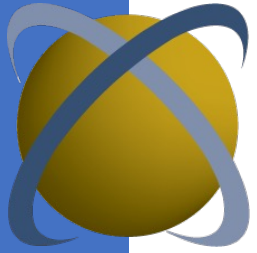
2022 Actuals, Change from 2021, 2023-27 Forecast

Including “Grand Unification”
of Data Center, Hyperscale, Cloud

May 2023



Intersect360
RESEARCH



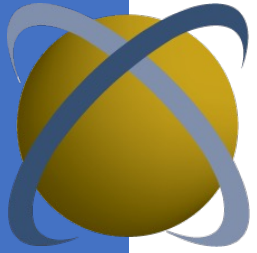
Intersect360 Research Momentum



Steve Conway
Senior Analyst
Joined December 2022



Frank Richardson
Director, Client Relations
Joined May 2023

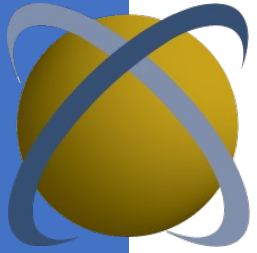


2023 Winter Classic Invitational Student Cluster Competition



- Virtual competition
- 12 teams from HBCUs and HSIs
- Mentors provide clusters & training

winterclassicinvitational.com

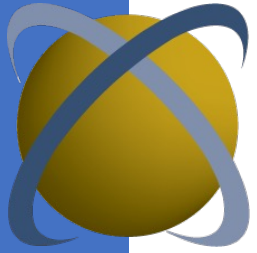


Upcoming HPC Market Reports

Intersect360 Research will publish full written research reports for the data in this presentation, in the following segmentations:

- Products and services
- Vertical markets
- Geographic regions
- Server vendor market shares*
- Server classes
- Software categories
- Services categories
- Cloud categories

* Market shares for 2021-22 only, not forecast.

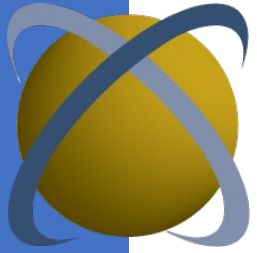


Data Center “Grand Unification”

All 2022 Data Center spending:
HPC, AI, Cloud, Hyperscale, Enterprise



Intersect360
R E S E A R C H

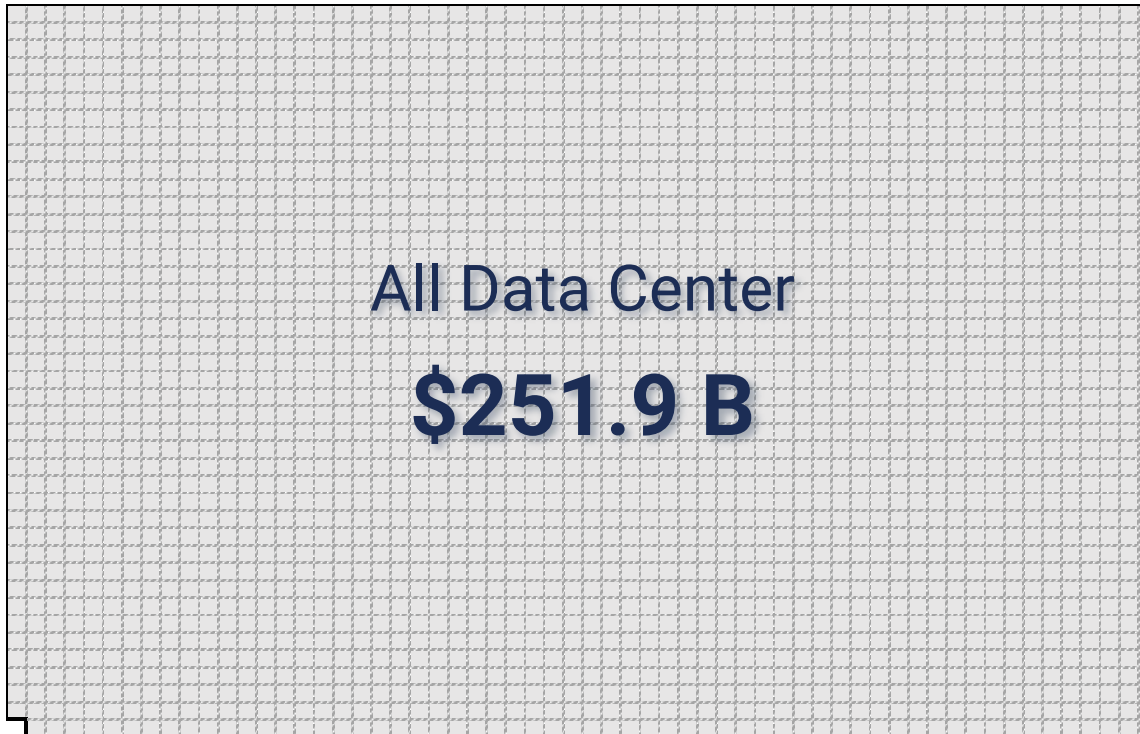


Why a Full Data Center Picture?

- Intersect360 Research has historically tracked both the HPC and Hyperscale markets, with clients primarily focused on HPC
- Hyperscale dominates the landscape and needs to be seen
- Especially important with AI and cloud adoption as prominent topics
- A unified view is required to understand the interrelated dynamics
- This view is independent work by Intersect360 Research. It is consistent methodologically with work published in prior years.



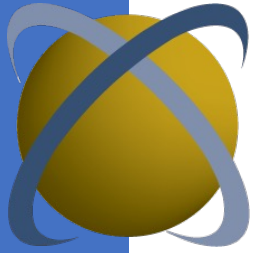
Data Center 2022: Grand Unification



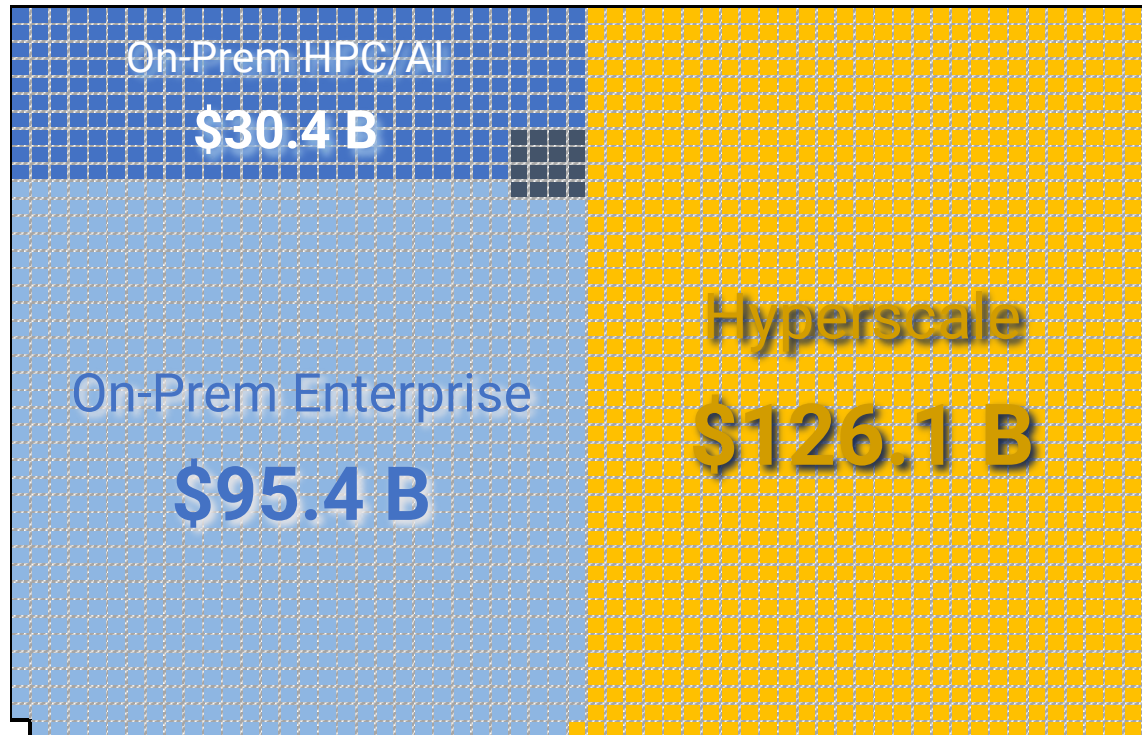
 = \$100 million

Total data center spending,
\$251.9 billion in 2022;
7.2% growth year-over-year

Includes both on-prem and
hyperscale data centers



Data Center 2022: Grand Unification

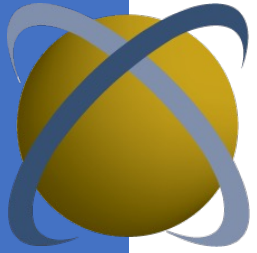


 = \$100 million

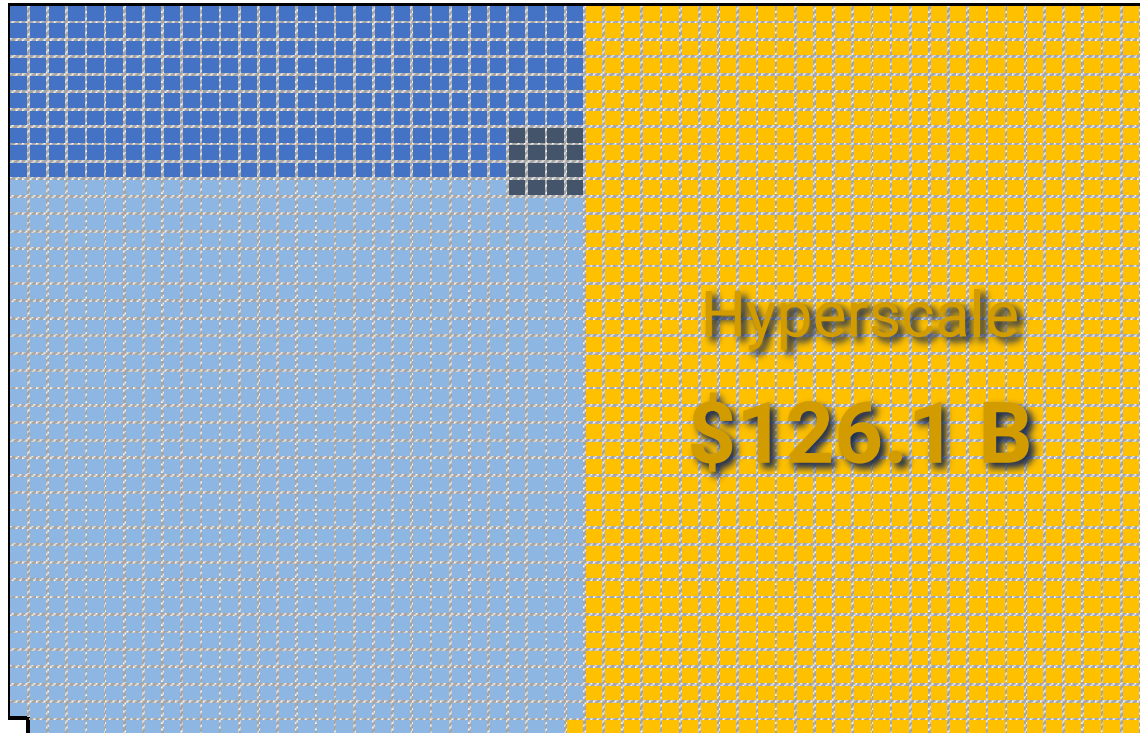
Hyperscale is now half (50.1%) of all data center spending

On-Prem HPC-AI includes both:

- Traditional HPC budgets (\$28.8B), usually mixed HPC-AI environments
- Pure AI, non-HPC (\$1.6B)



Data Center 2022: Grand Unification

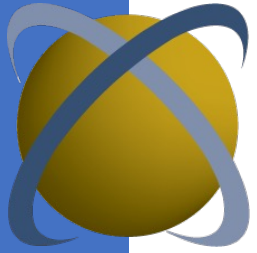


 = \$100 million

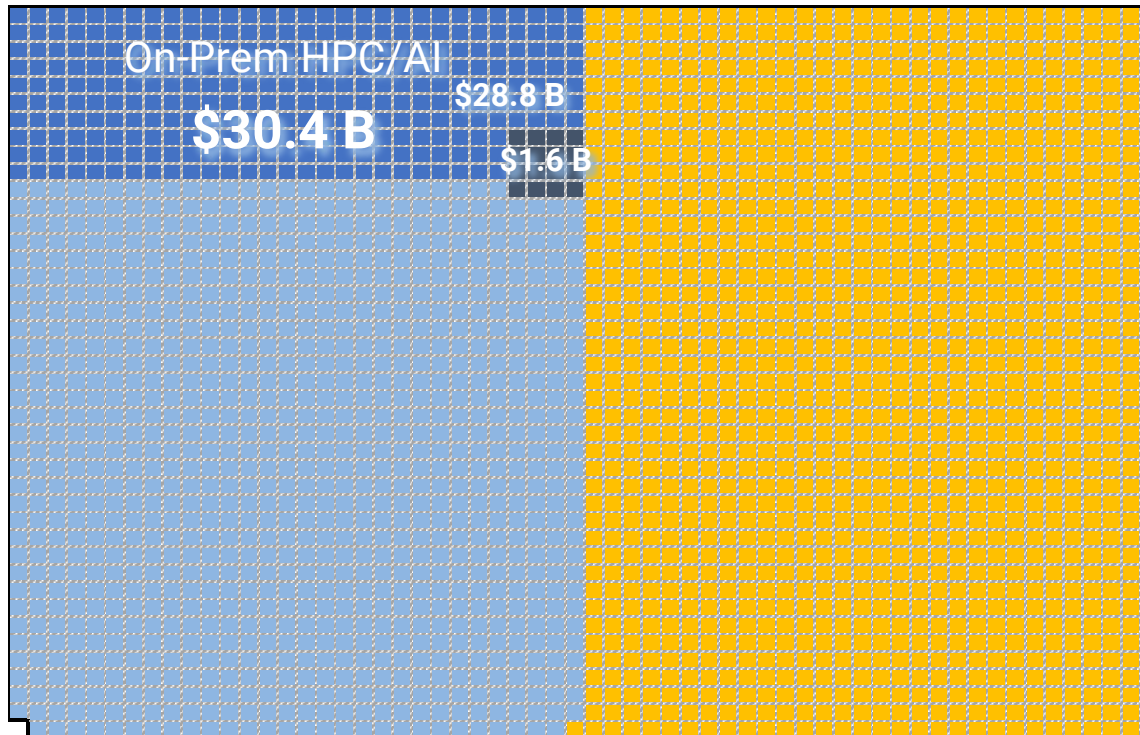
Hyperscale growth slowed down in 2022: 9.7% growth Y/Y, after >25% CAGR previous five years.

Top 10 companies account for about \$100B of spending.

Hyperscale is more than cloud; CSP is one line of business a Hyperscale company can pursue.



Data Center 2022: Grand Unification

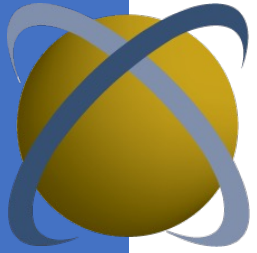


 = \$100 million

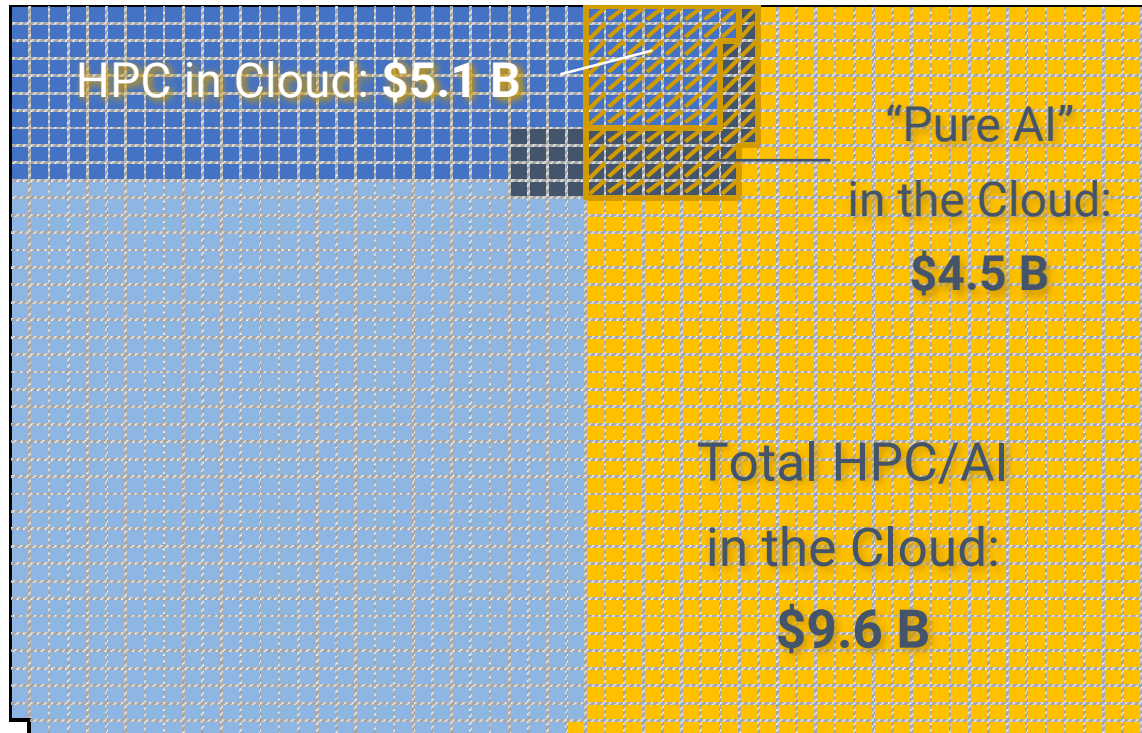
\$28.8 billion is associated with HPC budgets, conforming to our previous years' methodologies

This is the second year that on-prem "Pure AI" warrants tracking separately, \$1.6 billion.

We show both "traditional" and "expanded" views in these reports.



Data Center 2022: Grand Unification

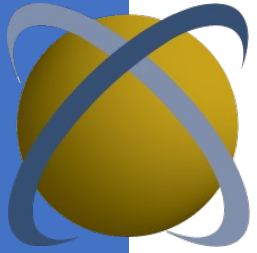


 = \$100 million

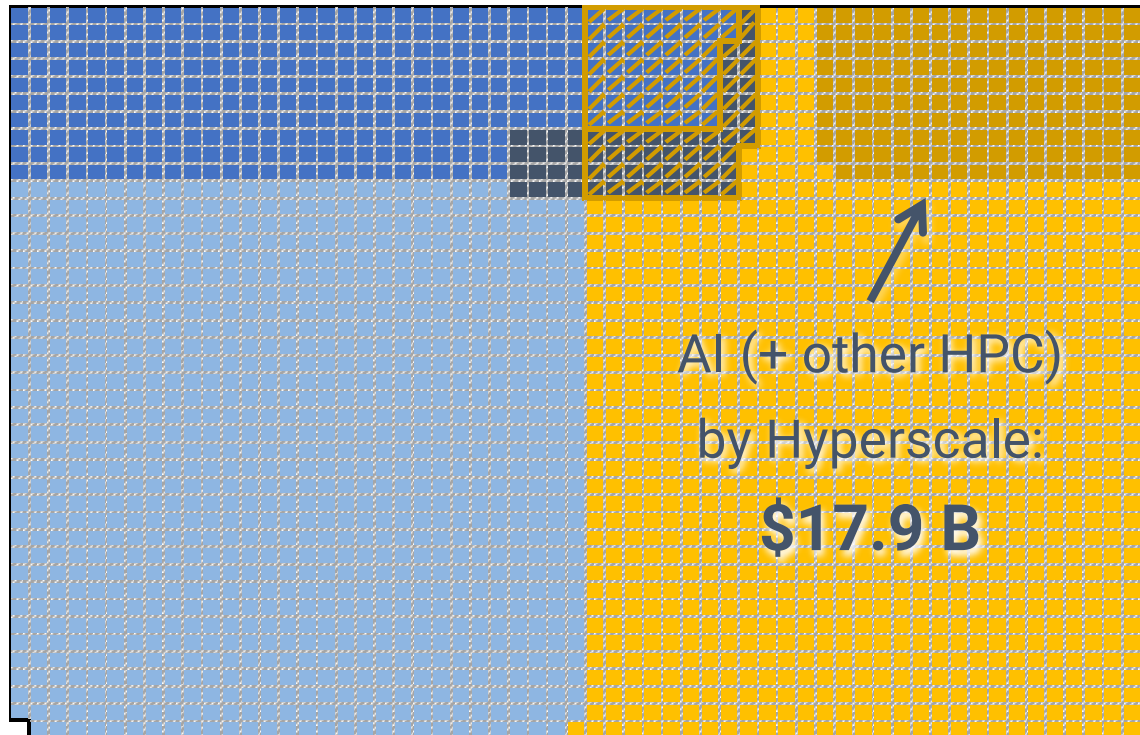
"Pure AI" budgets are much more cloud-centric than on-prem

Traditional HPC budgets are still mostly on-prem

After high growth during COVID-19 pandemic, cloud penetration is nearing its limit in HPC-AI



Data Center 2022: Grand Unification

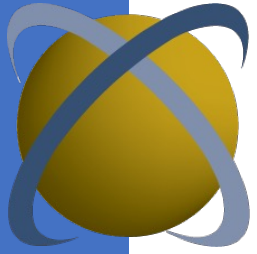


 = \$100 million

The primary driver of AI infrastructure is Hyperscale, for their own use.

More spending on AI in Hyperscale than in all other vertical markets combined.

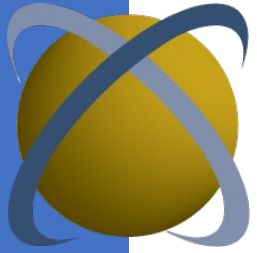
This spending is highlighted here but not included in most charts, so as not to distort the market opportunity.



Worldwide HPC/AI Market 2021-22 and 2023-27 Forecast



Intersect360
R E S E A R C H



What Is the HPC-AI Market?

Intersect360 Research has a proven history tracking high-performance, scalable data center technologies

Included: Multi-node, networked systems or cloud instances running parallel applications, generally requiring focus on performance or scalability in some dimension (e.g. processing, memory, I/O, networking); plus associated storage, software, services, etc.

Not included: Single-node desktops or workstations; embedded or edge devices



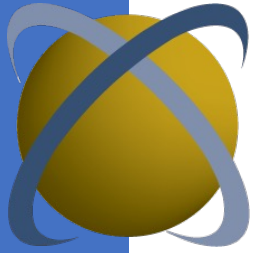
Two Market Views

“Traditional” HPC View

- HPC budgets
- Frequently (usually) includes machine learning workloads
- Can have increased budgets due to machine learning
- Conforms to previous years' methodologies

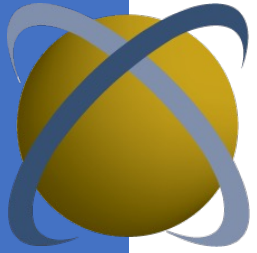
“Expanded” HPC + AI View

- Includes AI budgets in no way connected to HPC
- Conforms to how most technology suppliers view the market; basis of market shares
- Excludes hyperscale spending on AI, unless otherwise indicated



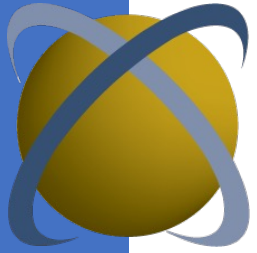
HPC + AI Market in 2022: Notes

- Now in stable growth mode following pandemic economic effects in 2020-21.
- Supply-chain issues continued to influence spending, limiting market bounce-back.
- Frontier supercomputer at ORNL was “official” start of Exascale era.
- Cloud continued fast growth through first half of 2022 before moderating. Cloud is now approaching its limit of penetration in HPC-AI.
- Major growth in Services categories reported across-the-board.
- On-prem “pure AI” deployments grew.
- Hyperscale is the dominant customer of AI; this is omitted unless specified.

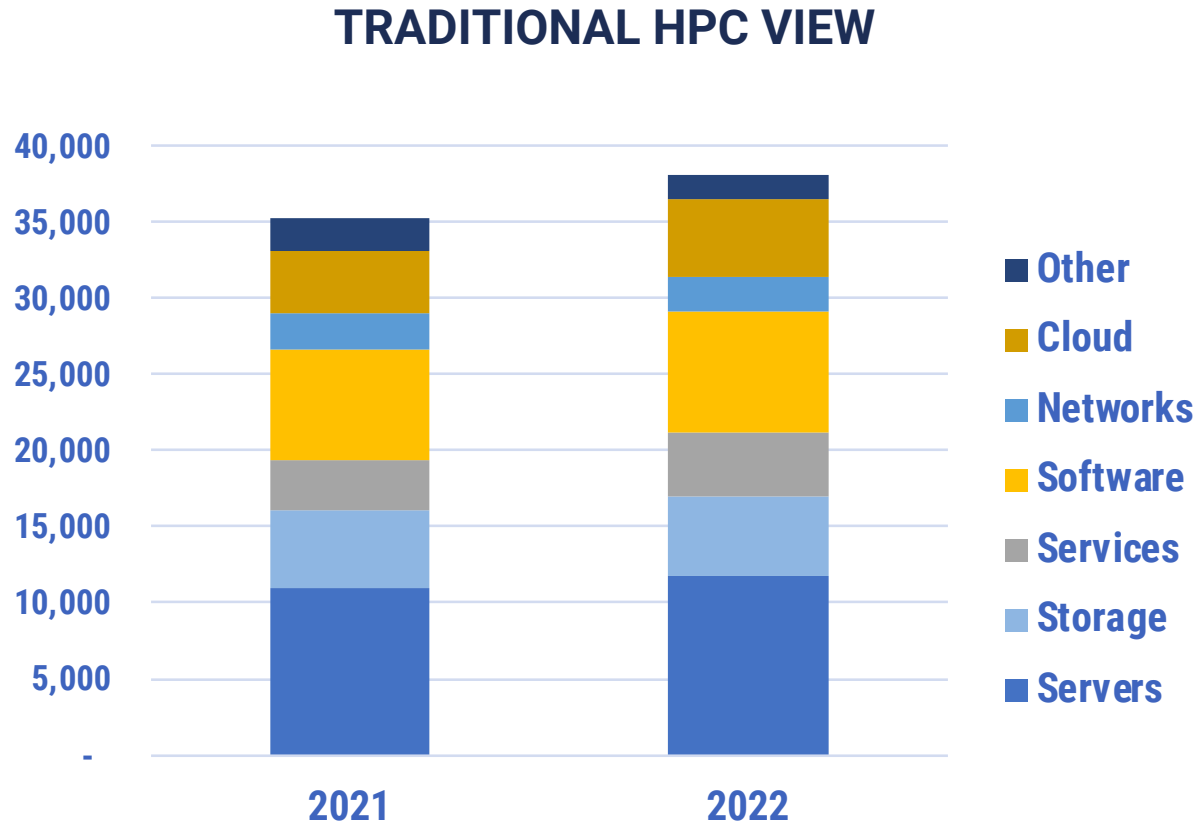


HPC + AI 2023-27 Market Forecast: Notes

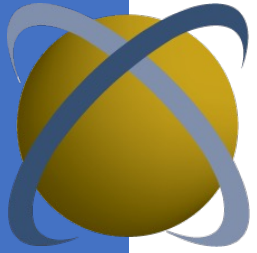
- Return to stable growth after stagnation during the COVID-19 pandemic.
- Traditional HPC market will grow at 6.1% CAGR to \$51.2 billion in 2027.
- Adding “pure AI” budgets, the combined HPC + AI market will grow 6.3% CAGR to \$60.4 billion, not including hyperscale market spending.
- 2023 is the final year of double-digit cloud growth in HPC-AI. Cloud will peak at about 15% of traditional HPC spending, or 23% of the broader HPC-AI market.
- Hyperscale companies are so dominant as to be unpredictable as to their long-term effect. For this forecast period, the data center is increasingly commanded by only a few companies. This is not long-term stable and should be monitored as a market risk.



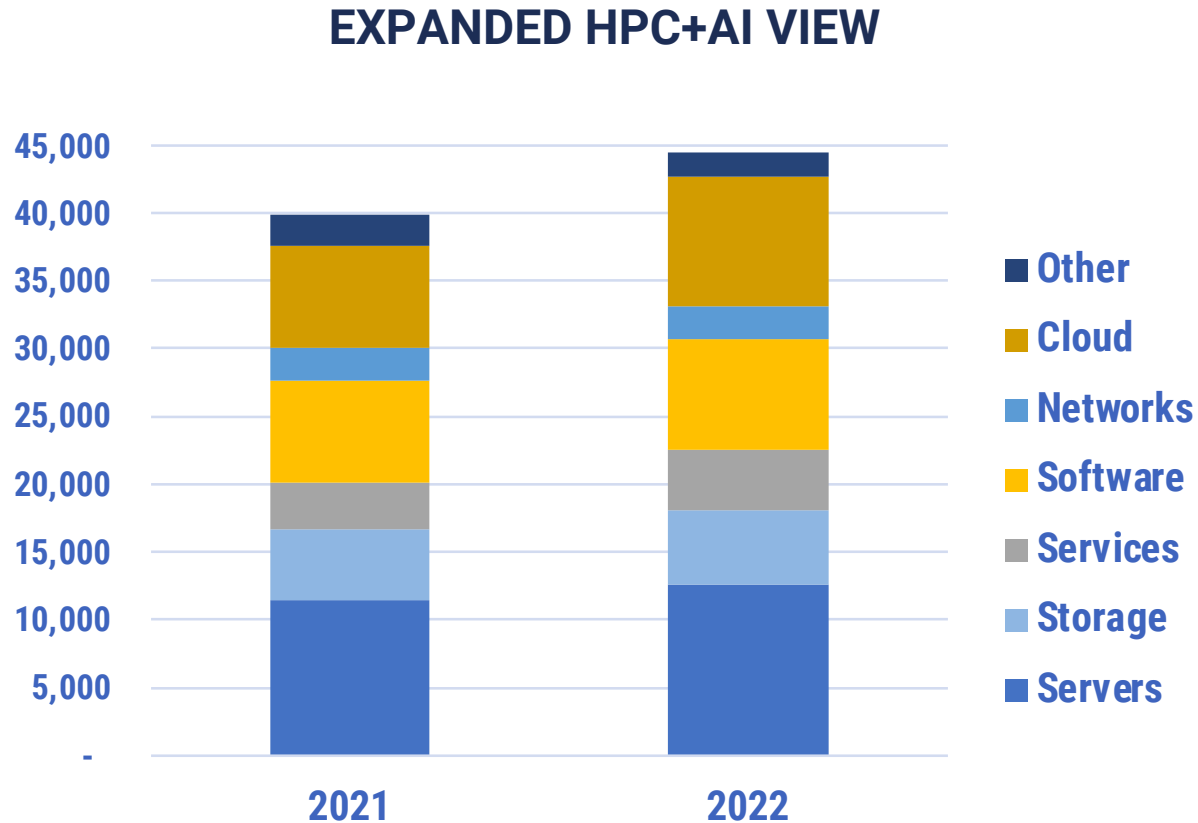
Traditional HPC Products & Services (\$M)



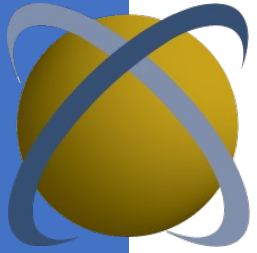
- 8.1% market growth year-over-year, to \$38.1 billion
- HPC servers up 7.5%
- Strongest growth in services (29.5%) and cloud (25.3%)
- Includes AI as part of mixed HPC-AI budgets



Expanded HPC-AI Products & Services (\$M)

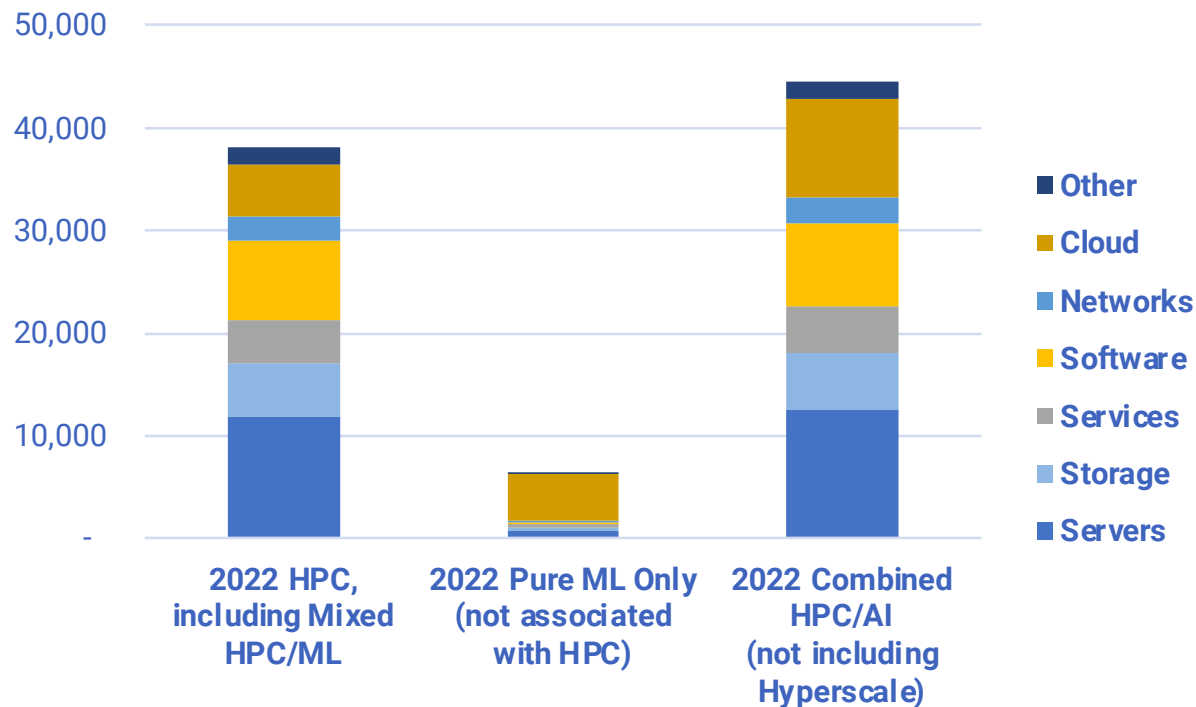


- 11.3% market growth year-over-year, to \$44.4 billion
- Cloud has much higher penetration including “pure AI”
- Strongest growth in services (31.5%) and cloud (26.2%)

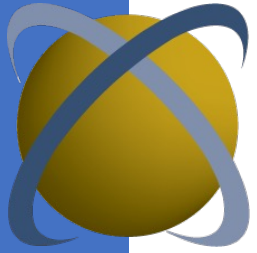


Traditional and Expanded View Comparison

EXPANDED HPC + AI TRAINING VIEW

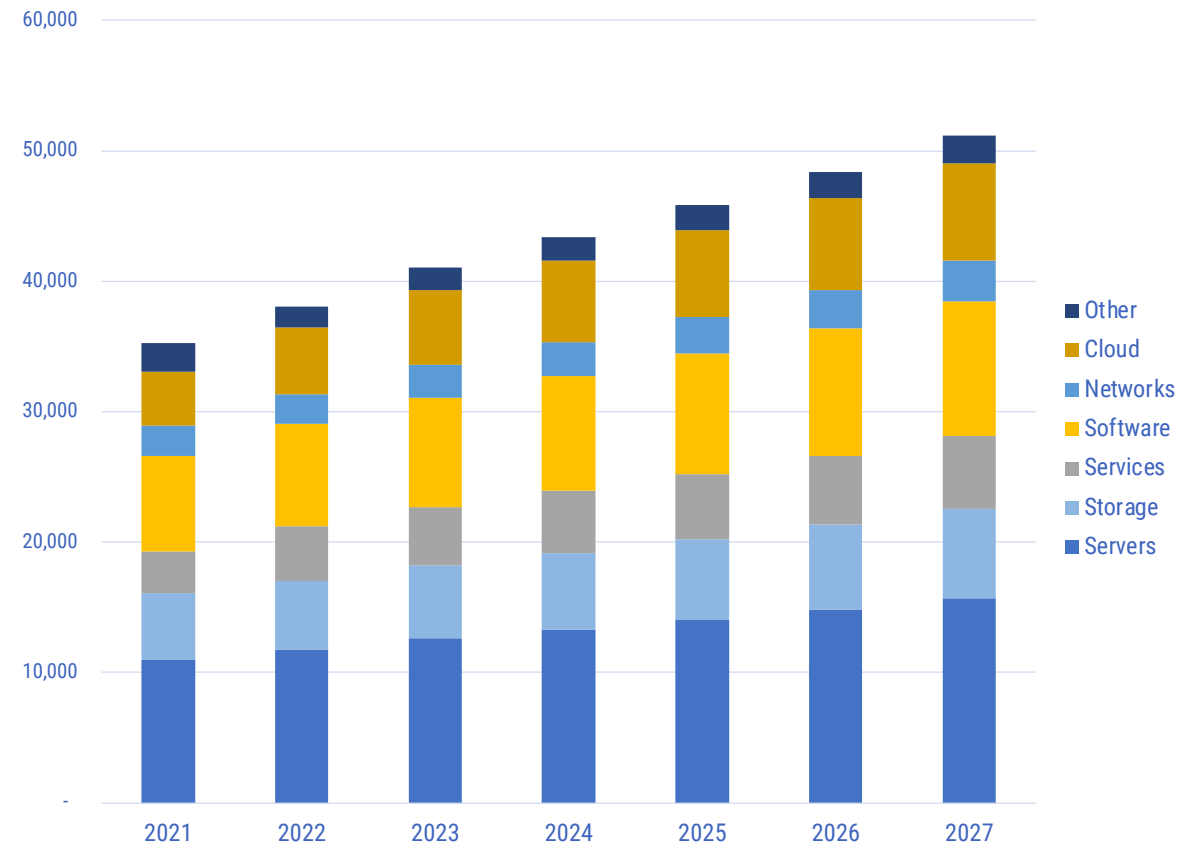


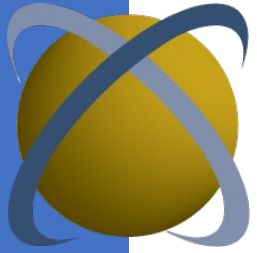
- “Pure AI” spending not associated with HPC adds \$6.3B to market
- 70.4% of this “pure AI” spending is cloud
- Combined, cloud is \$9.6 billion
- Does not include \$17.9 billion hyperscale spent on AI training



Traditional HPC Forecast: Products and Services (\$M)

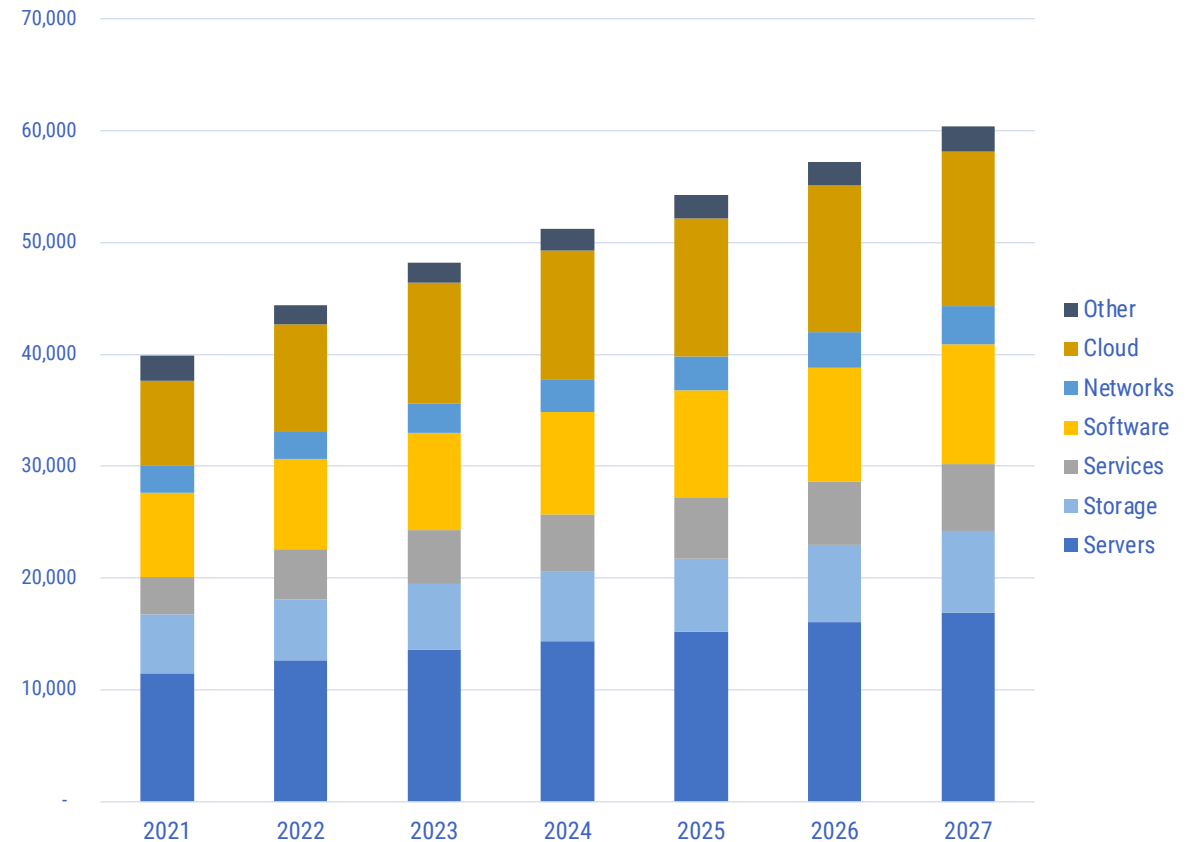
- Return to steady post-pandemic growth
- Total market 5.7% CAGR; HPC servers 5.6% CAGR
- Cloud achieves maximum penetration ~15% of HPC spending





Expanded HPC+AI Forecast: Products and Services (\$M)

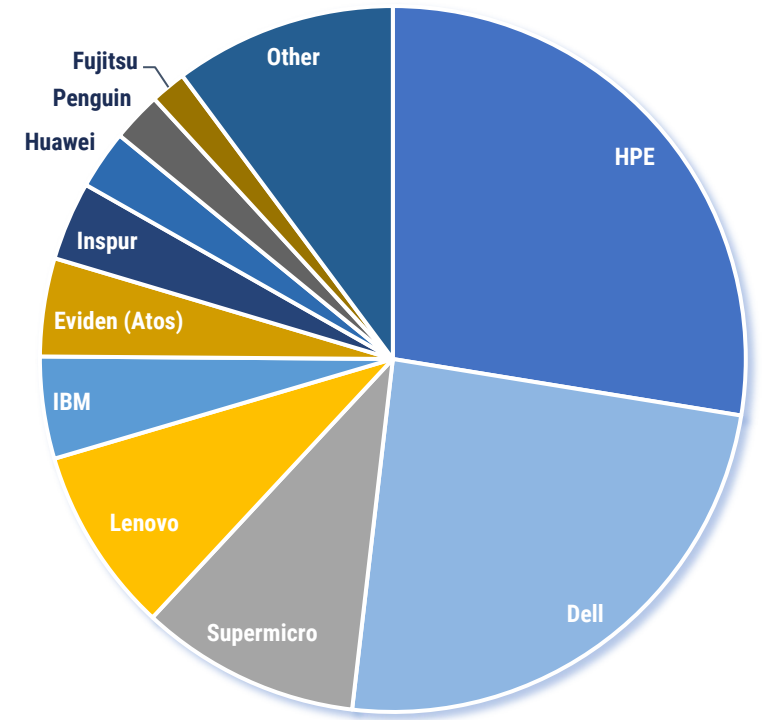
- Slightly higher CAGR (6.3%) when distinct “pure AI” budgets are included
- Cloud has a much higher penetration in AI
- Cloud nears maximum penetration ~23% of spending
- Remember this does NOT include spending by hyperscale companies

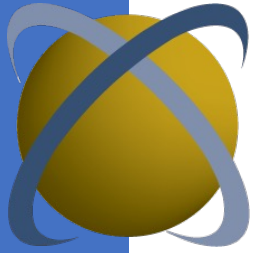




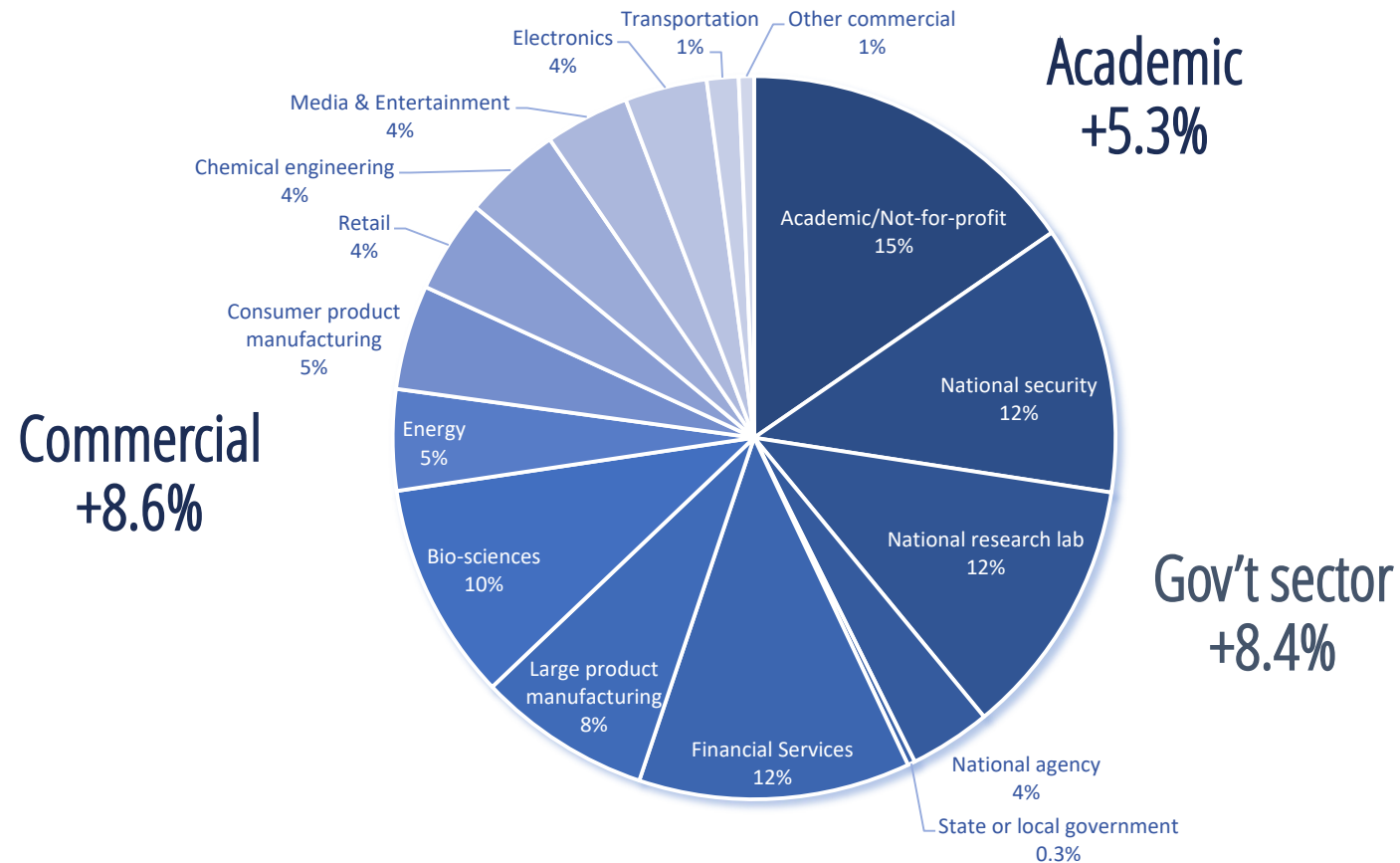
HPC + AI Server Revenue Share

- HPE (+7.3%) and Dell (+8.2%) are still the market leaders
- HPE Frontier system at ORNL is the "official" start of the Exascale era
- Major growth from Supermicro, overtaking Lenovo to become #3 vendor
- IBM also had strong growth thanks to AI contribution from new zSeries
- Notable "Others": NEC, Nvidia, Sugon
- Nvidia, if tracked, would be the #5 vendor – there is insufficient supply side guidance to plot Nvidia's non-Hyperscale HPC-AI system revenue with accuracy



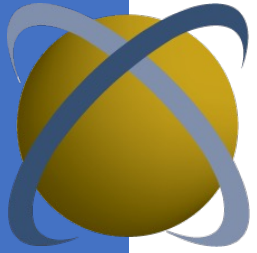


Traditional HPC Vertical Markets: 2022

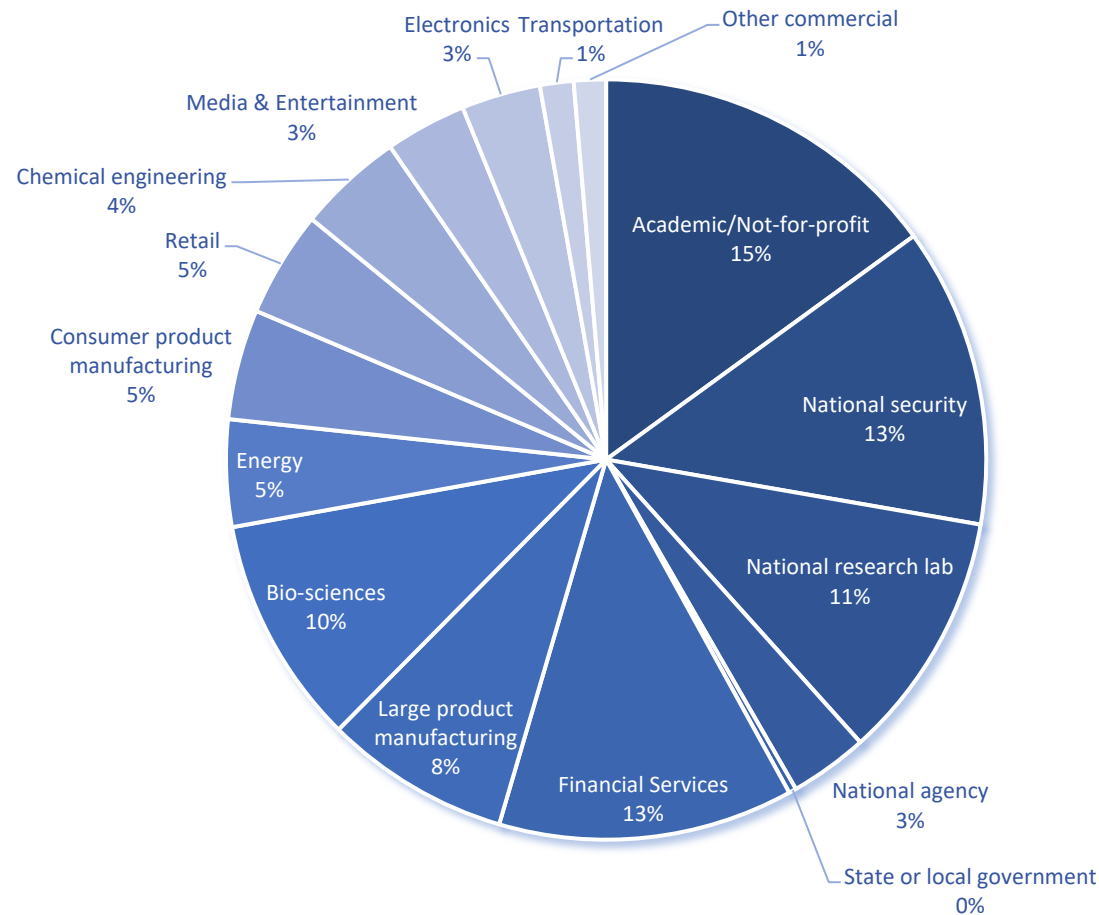


Some commercial segments – especially manufacturing and energy – experienced moderate post-COVID bounce-back with double-digit growth

National labs also had double-digit growth with installation of Frontier supercomputer

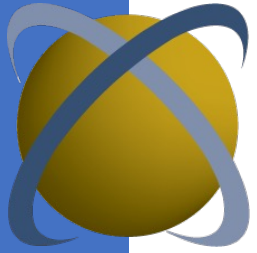


Expanded HPC + AI Vertical Markets: 2022

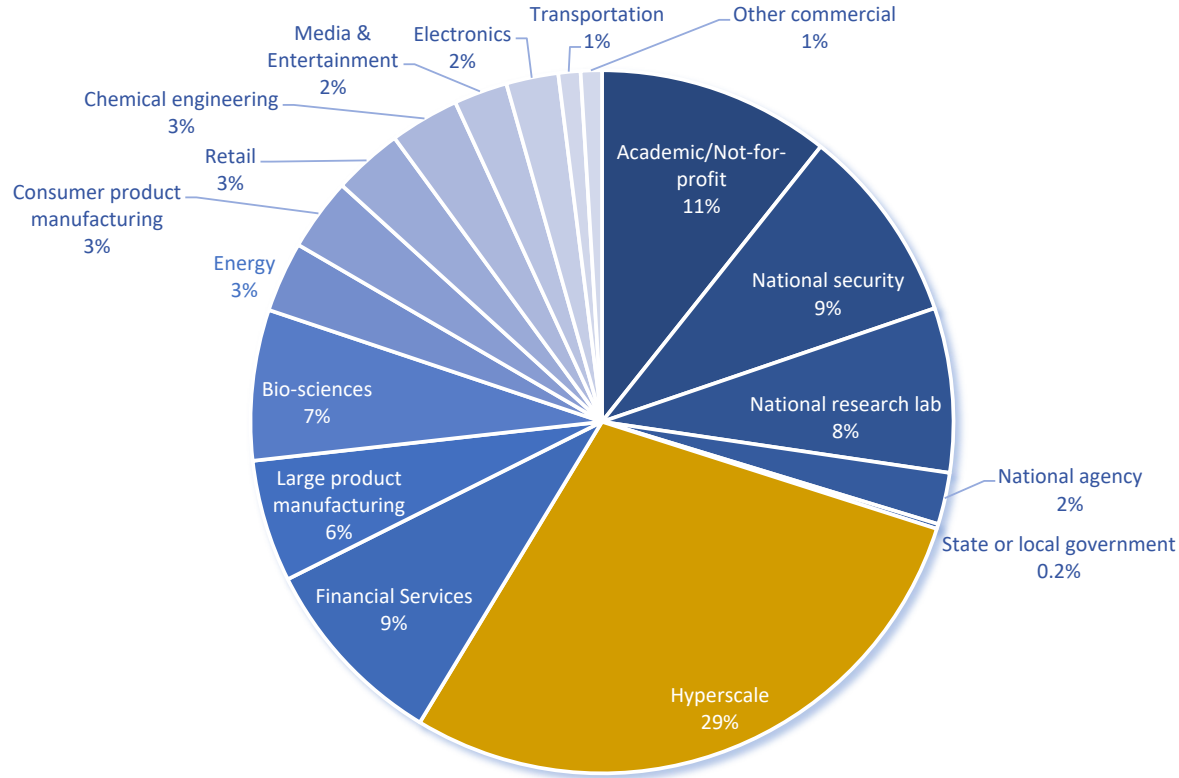


Not a major difference from "traditional" HPC chart, because pure AI spending is smaller

National security and financial services are major "pure AI" contributors

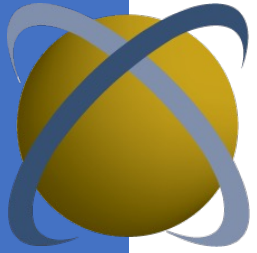


If Hyperscale Were Included in HPC + AI



Hyperscale, if included, would be the single biggest consumer of HPC-AI

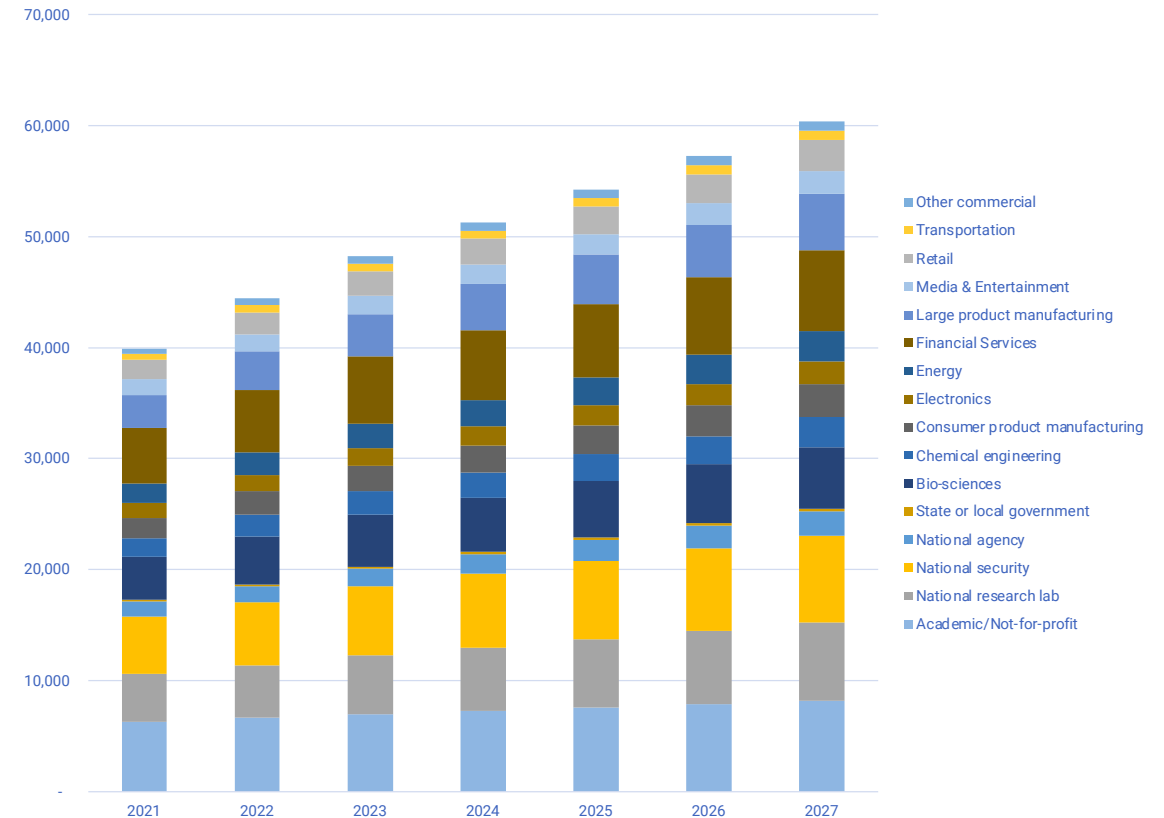
Nearly as much as entire public sector (academic + government) combined

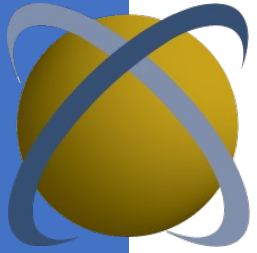


Expanded HPC + AI Forecast: Vertical Markets (\$M)

Return to stable post-pandemic growth model

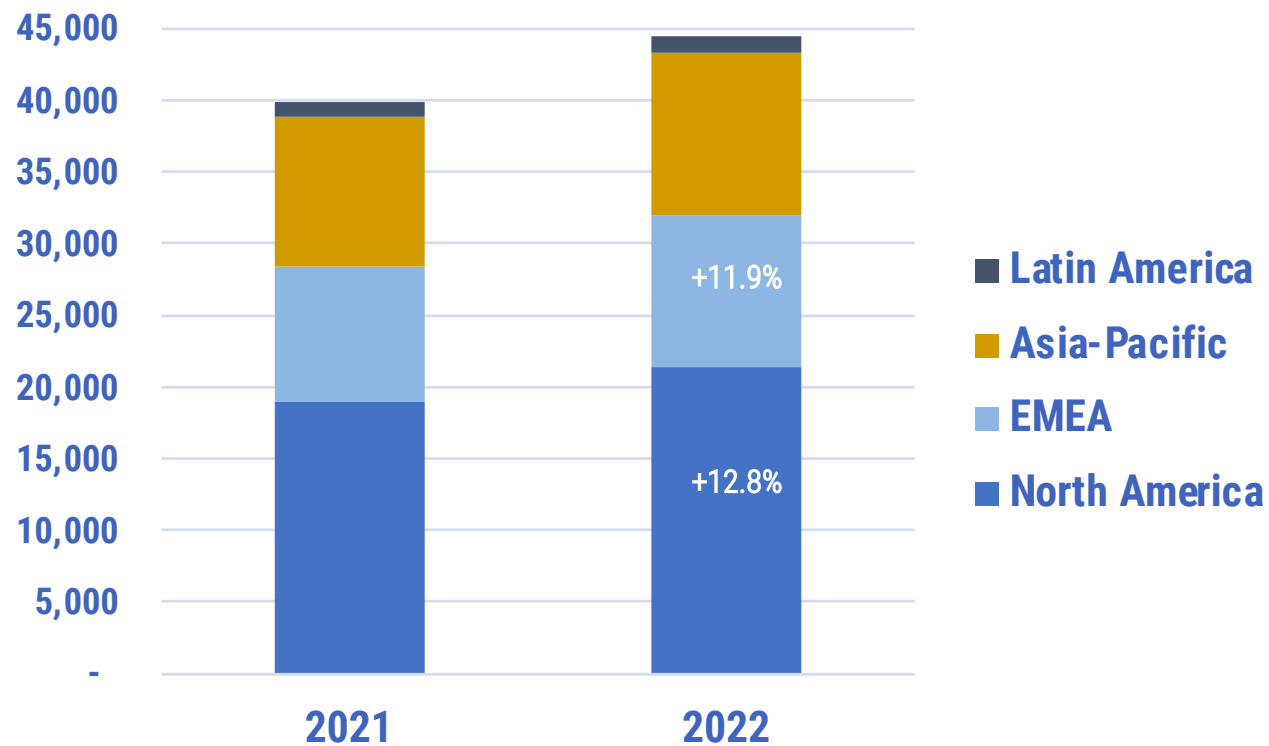
Government sector leads growth over the next five years, fueled by Exascale and increase in national security spending



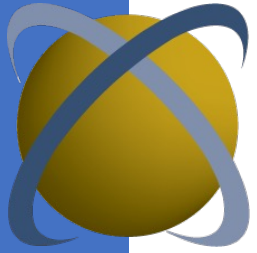


Expanded HPC+AI Global Regions

EXPANDED HPC-AI VIEW

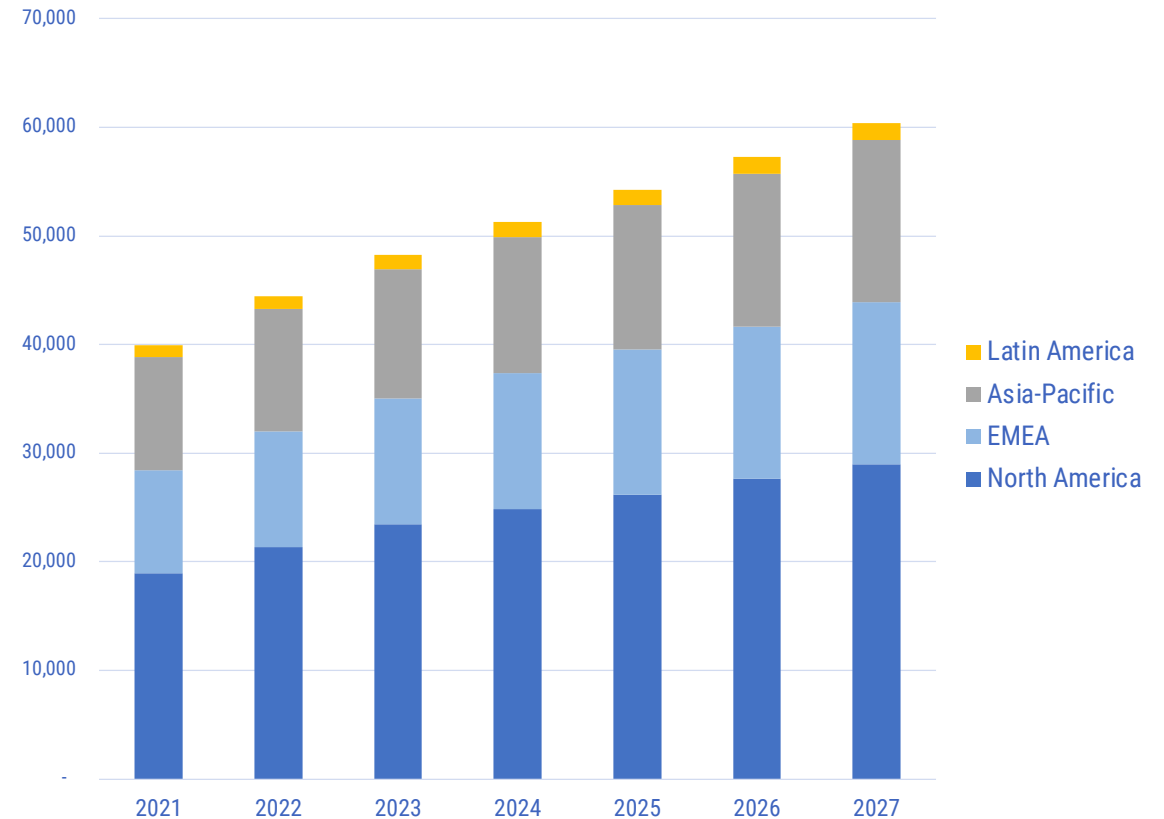


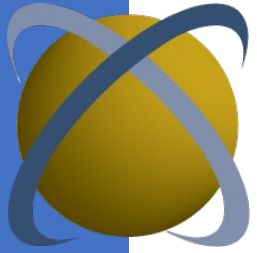
North America (+12.8%) and EMEA (+11.9%) were biggest gains, following two years led by Asia Pacific



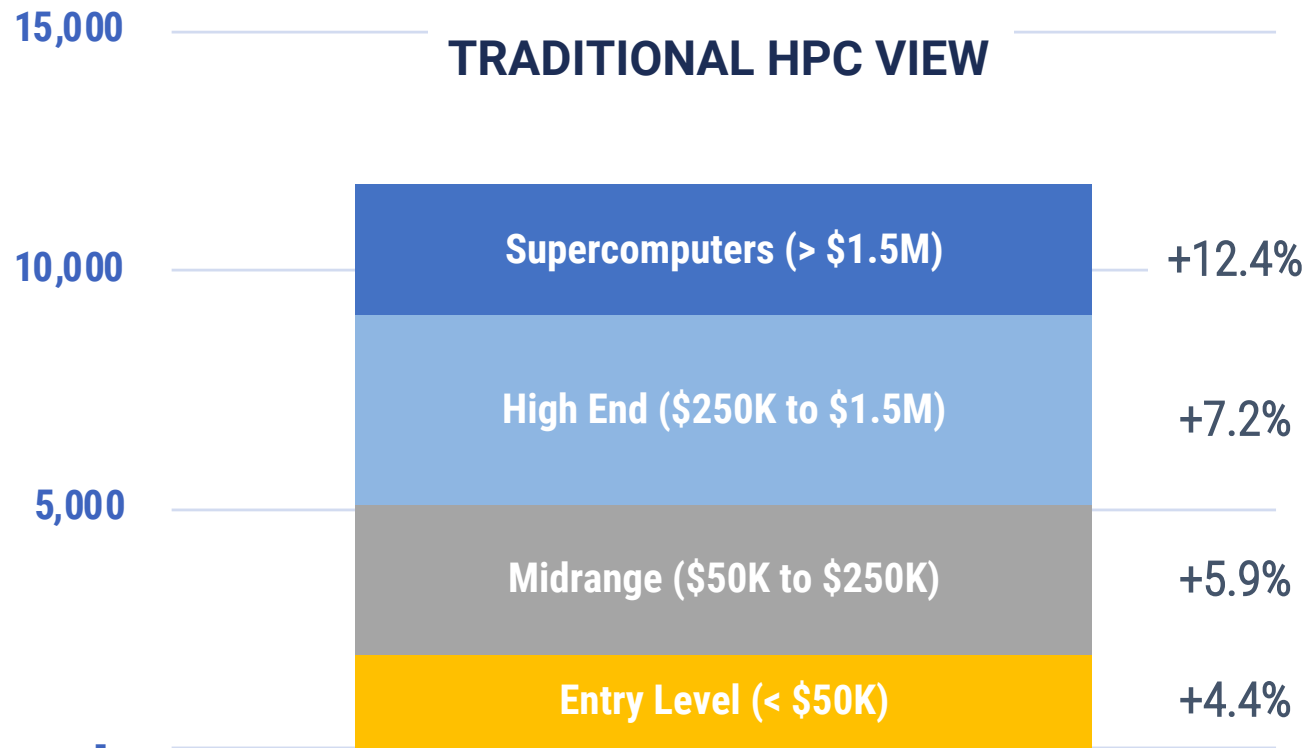
Expanded HPC + AI Forecast: Regions (\$M)

- Similar picture to traditional HPC model; all growth rates slightly higher
- In this view, EMEA crosses back ahead of Asia-Pacific in 2024





Traditional HPC Server Classes (\$M): 2022

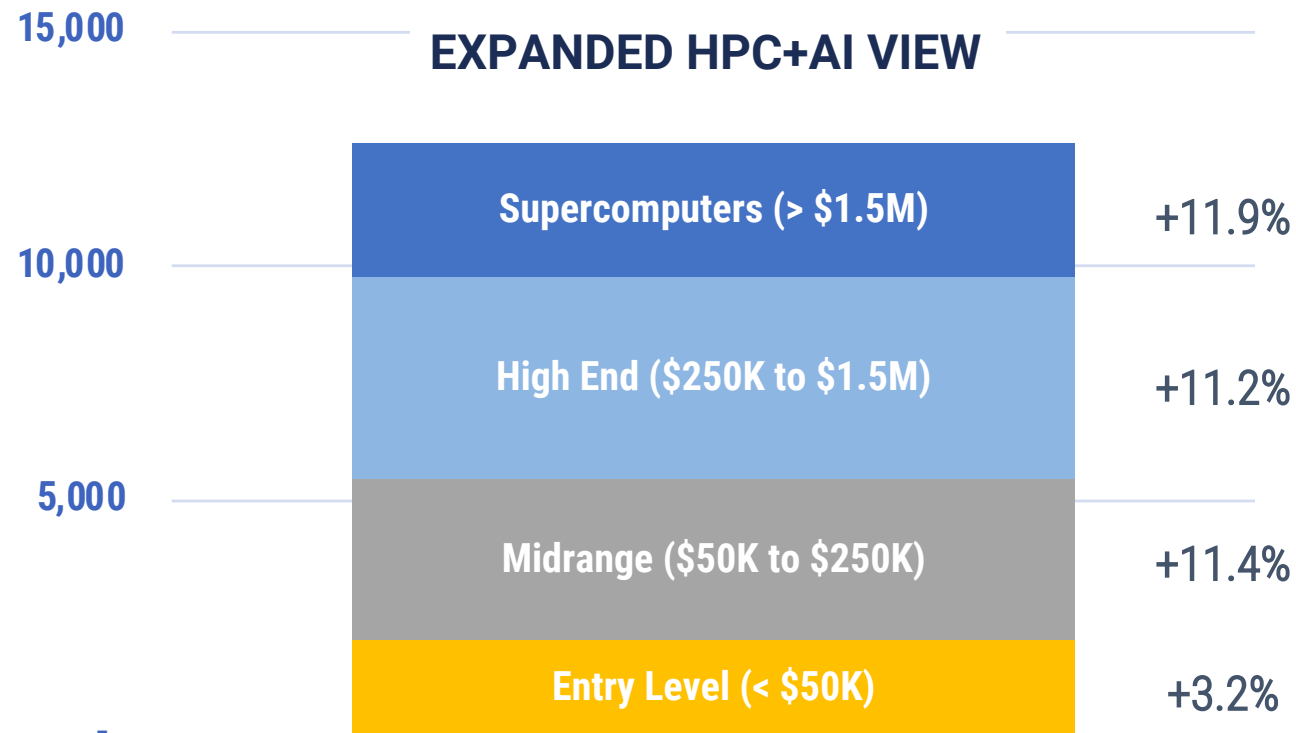


Supercomputers leading in growth. Cloud draws away more heavily from lower segments.

All HPC systems: +7.5%Y/Y

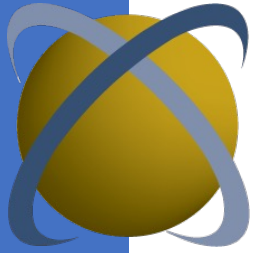


Expanded HPC+AI Server Classes (\$M): 2022



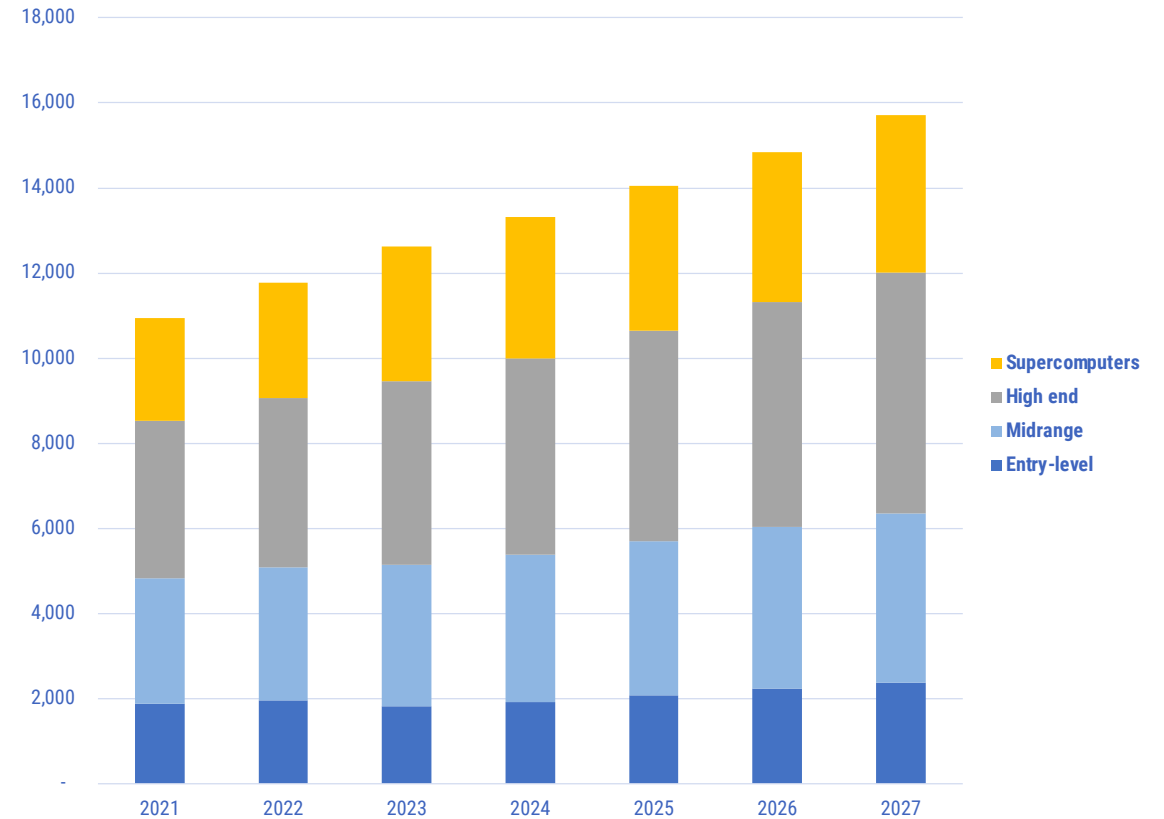
Little growth in entry level since most GPU-dense clusters do not fit in the price band.

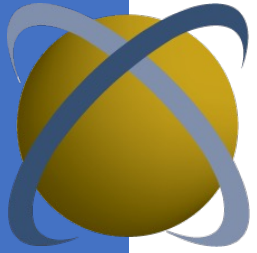
All HPC-AI systems: +10.1%Y/Y



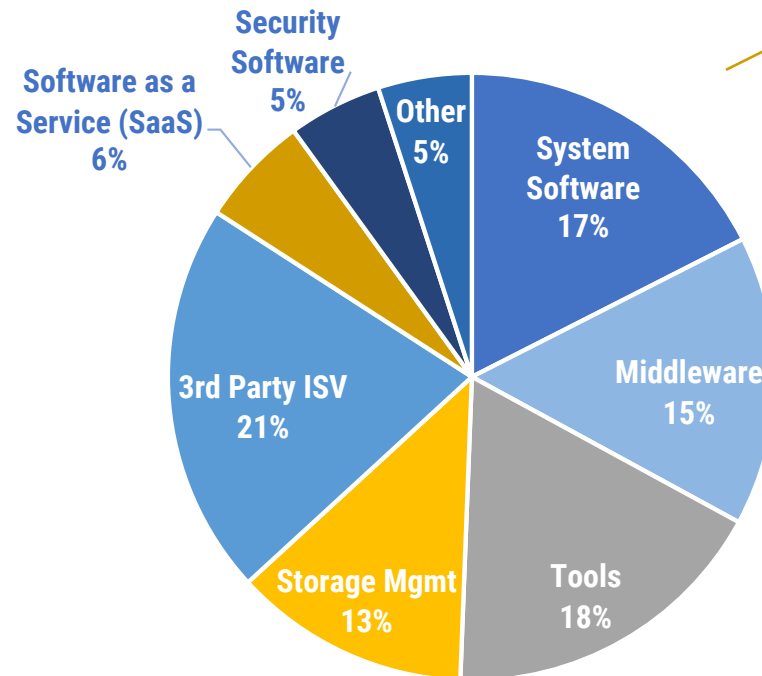
Traditional HPC Forecast: Server Classes (\$M)

- HPC servers will grow to \$15.7 billion in 2027
- Growth is concentrated in the top half of the market, due both to Exascale and effects of cloud
- Entry-level systems growth resumes after cloud reaches maximum penetration



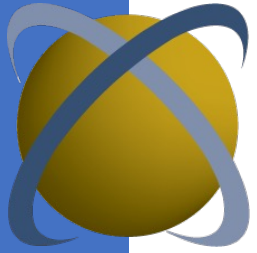


Traditional HPC Software: 2022



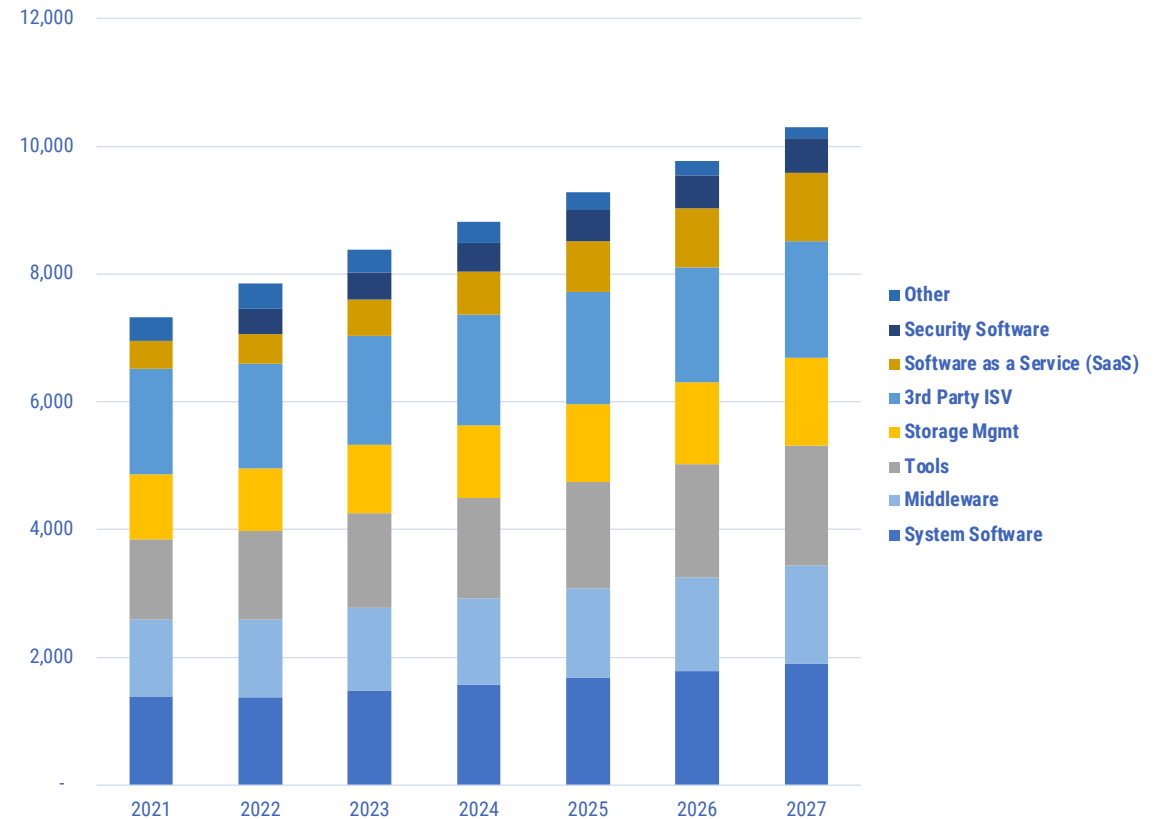
HPC software grew 7.1% Y/Y

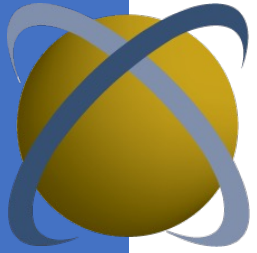
Security software:
New segment
added in 2022.



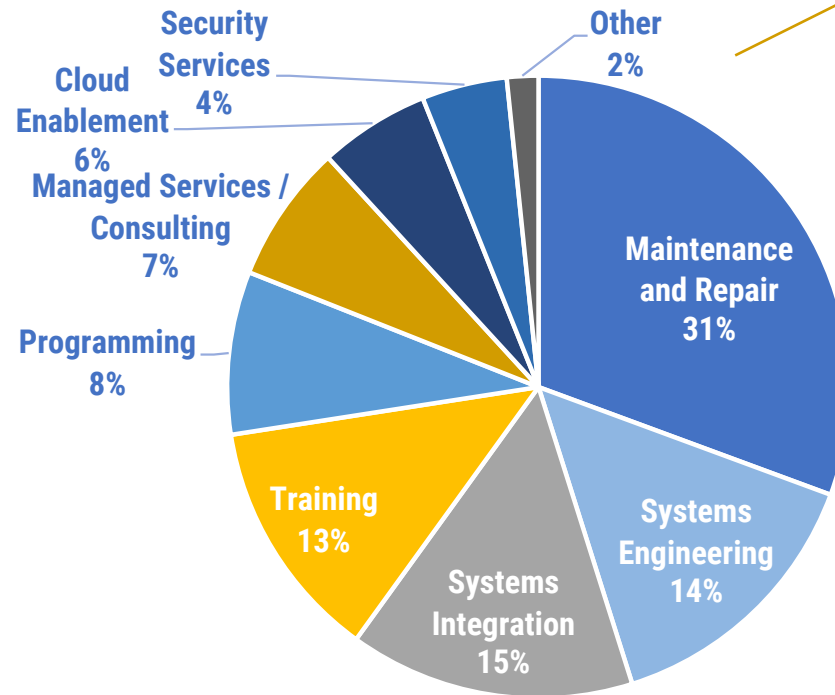
Traditional HPC Forecast: Software Segments (\$M)

- SaaS continues to have double-digit CAGR even beyond cloud penetration maximum
- Third-party ISV licensing is nearly flat over the forecast period, as usage is declining with cost models that rise with changes in processing architectures and cloud





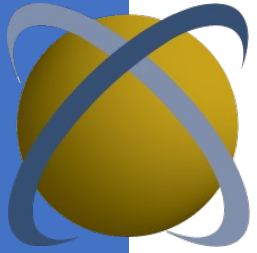
Traditional HPC Services: 2022



HPC Services grew 29.5% Y/Y

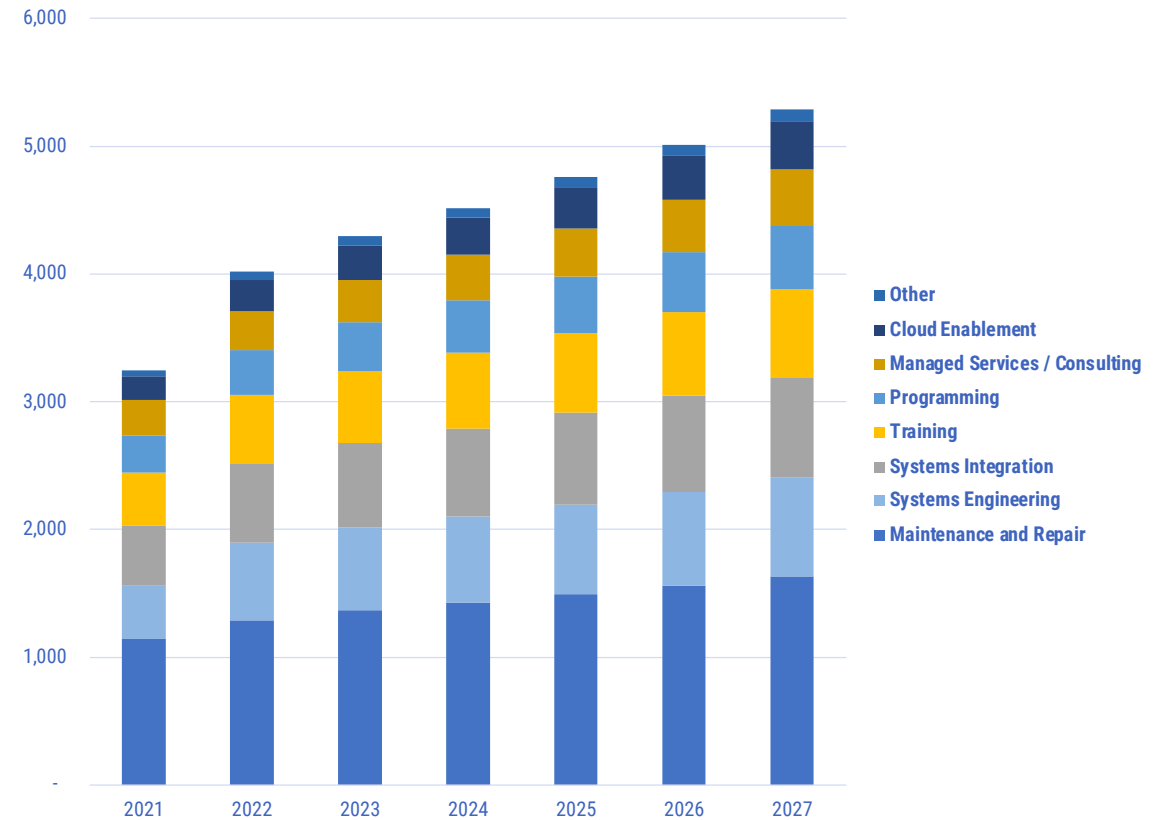
Large increase in reported spending on HPC services across the board.

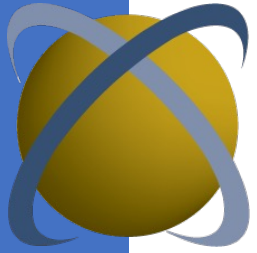
Security services:
New category in 2022



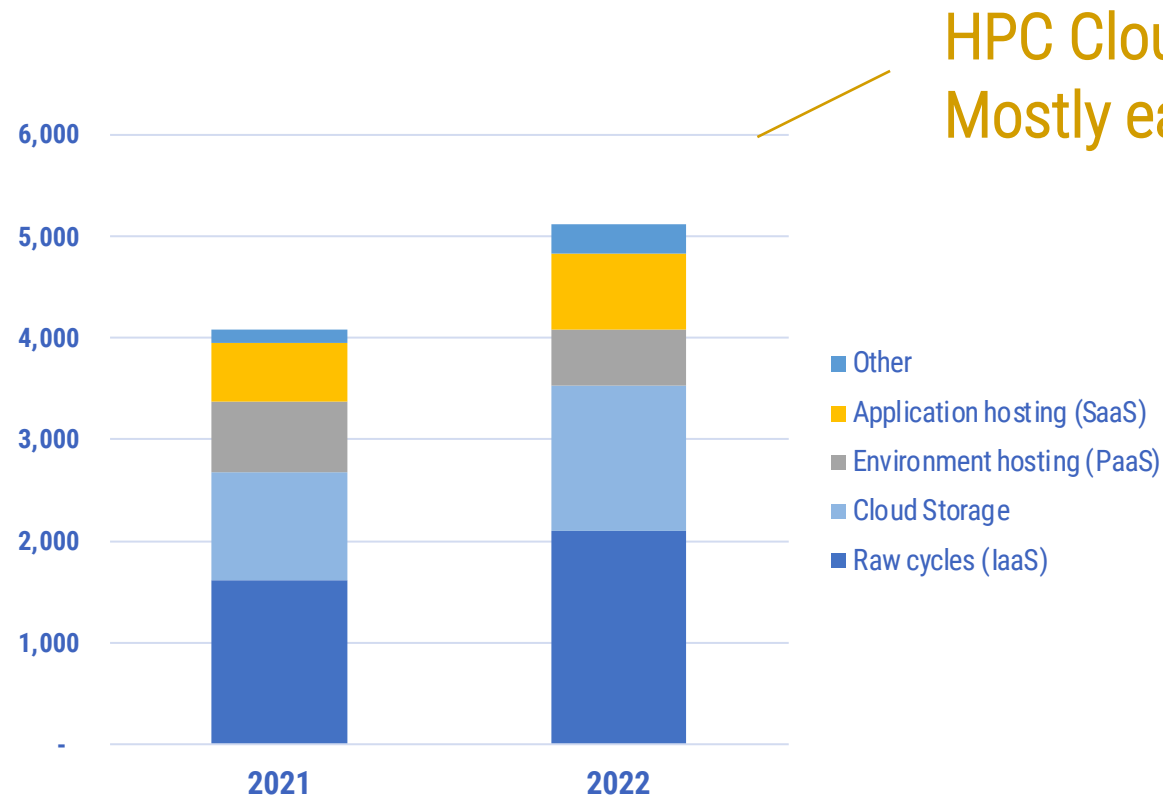
Traditional HPC Forecast: Services Segments (\$M)

- High growth in 2022 could set new baseline
- Forecasting return to stable growth for forecast period

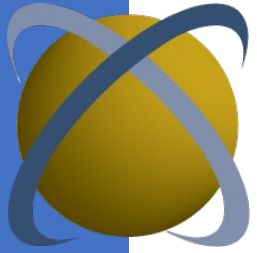




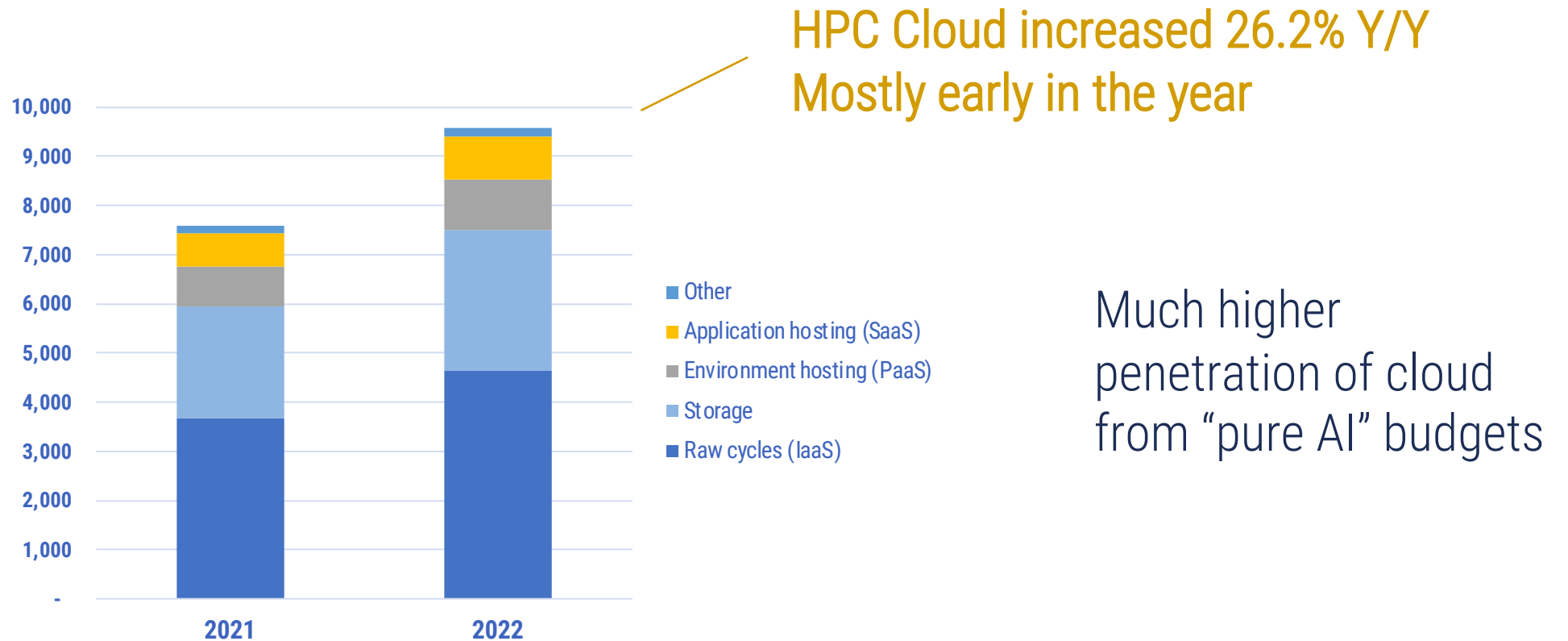
Traditional HPC Cloud: 2020 vs. 2021



Decline in “PaaS” (and growth in “Other”) could mean market moving away from that model and terminology



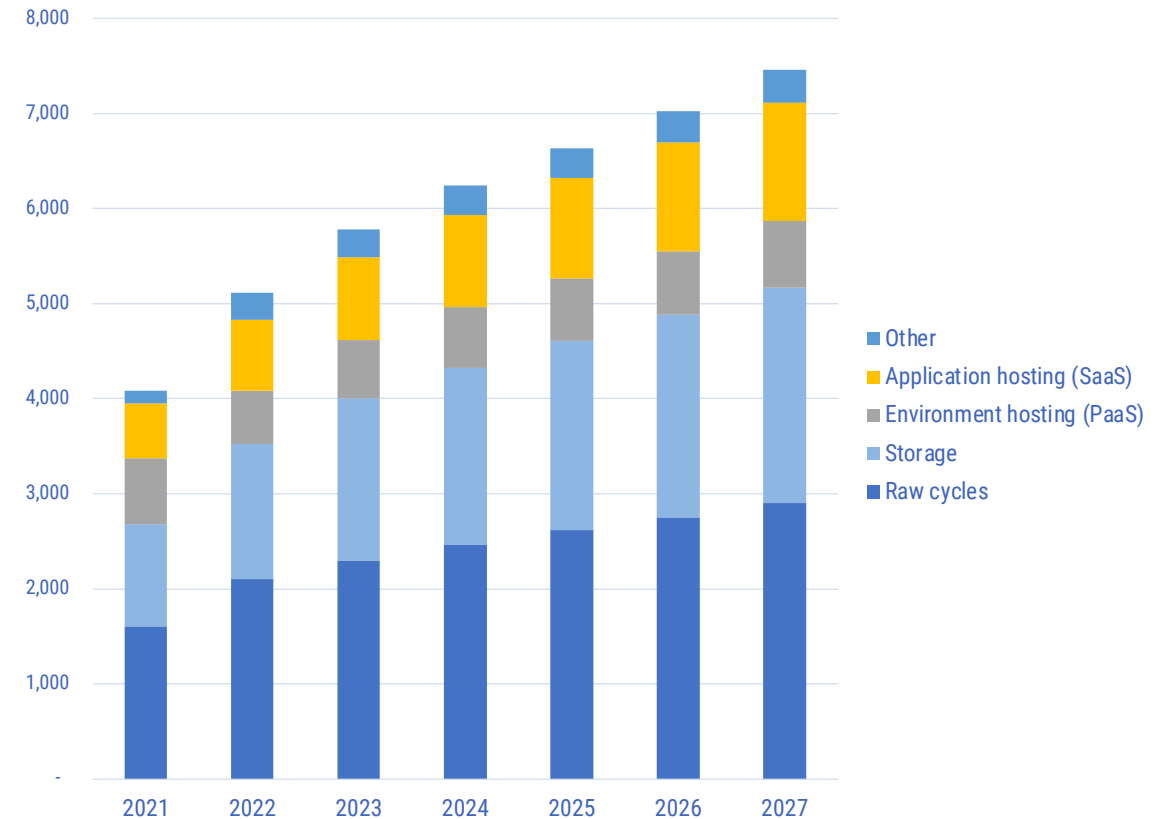
Expanded HPC+AI Cloud: 2021 vs. 2022

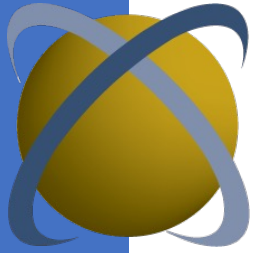




Traditional HPC Forecast: Cloud Segments (\$M)

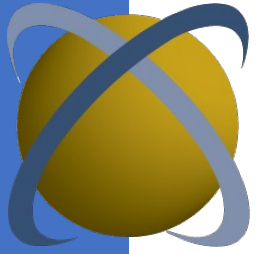
- 2022 was eighth-straight year of double-digit growth in cloud
- Beyond 2024, cloud grows along at market rates, having achieved near-peak penetration
- SaaS has highest growth rate through the forecast period





Conclusions

- HPC (including pure AI budgets) is a long-term growth market, returning to stability
- Exascale and AI are supporting long-term growth trends.
- Pure AI budgets, completely independent from HPC, are adding to the market—mostly in cloud, but with enough on-prem presence to warrant tracking separately for now. This may become a meaningless distinction by the end of the decade.
- Cloud computing is nearing its maximum penetration across HPC.
- **There is significant risk in this forecast.**
 - Assuming continuation of “normal” post-pandemic economic activity.
 - Risk of significant worsening of geopolitical environment (tariffs, wars, etc.).
- Contact Intersect360 Research for additional inquiries.

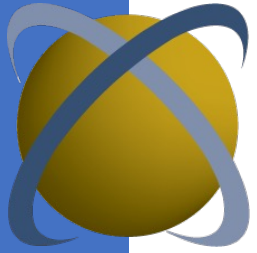


Machine Learning and AI

Workloads, Frameworks, Data Types
Configurations, Cloud Usage, Business Outcomes

February 2023

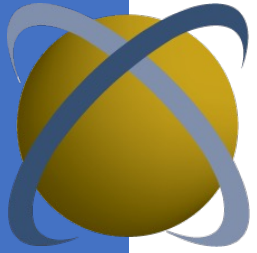




*Studied machine learning initiatives across
HPC-using organizations worldwide*

*Budgets, initiatives, workflows, data types
and sources, configurations, cloud usage,
and business outcomes*

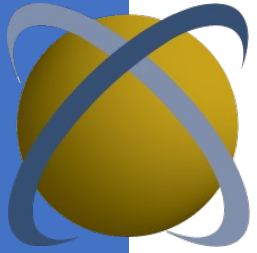
- *Survey conducted Q4 2022 – Q1 2023*
- *152 qualified respondents*



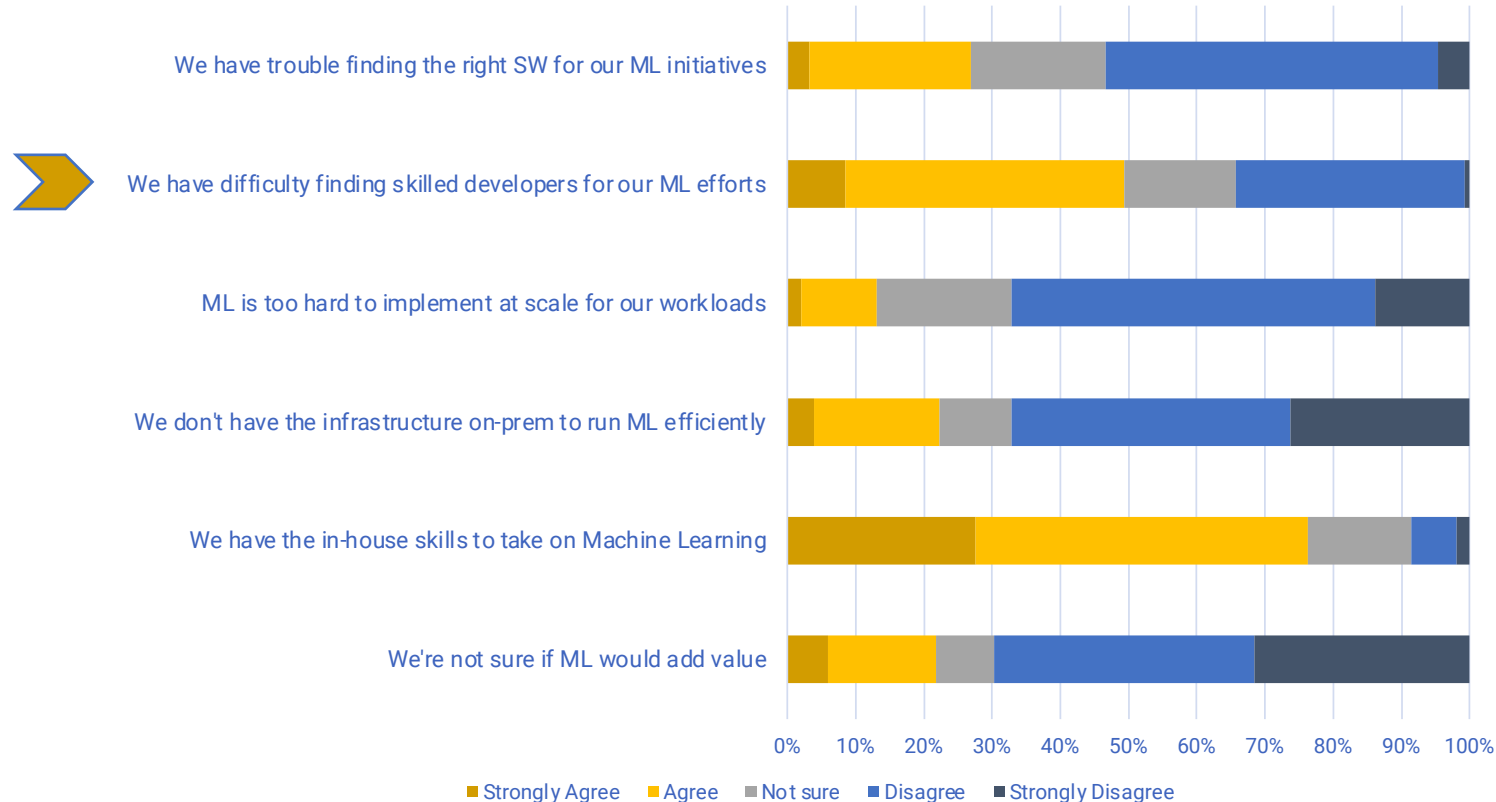
New Survey: Machine Learning

Comprehensive study of machine learning usage, strategies, and outcomes at HPC-using organizations worldwide

- Intentions of machine learning initiatives across 17 vertical markets and application domains
- Types of learning, with stages of implementation from feasibility studies to production
- Data types, sources, sizes, and growth; software strategies and frameworks used
- Computing architectures and configurations; scalability by framework
- Supervised vs. unsupervised; frequency of retraining
- Cloud and a-a-Service strategies and vendors
- Business outcomes: ROI, competitive advantage, etc.

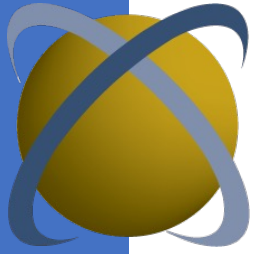


Considerations for Implementing ML

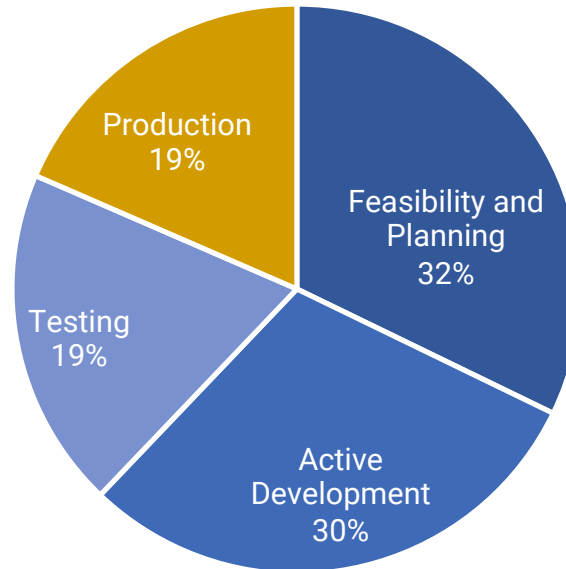


Roughly half of respondents say they have difficulty finding skilled developers for ML.

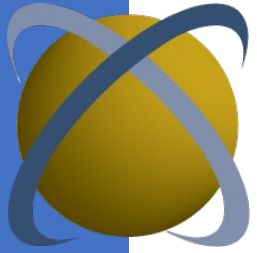
This increases to **61%** among respondents from commercial organizations.



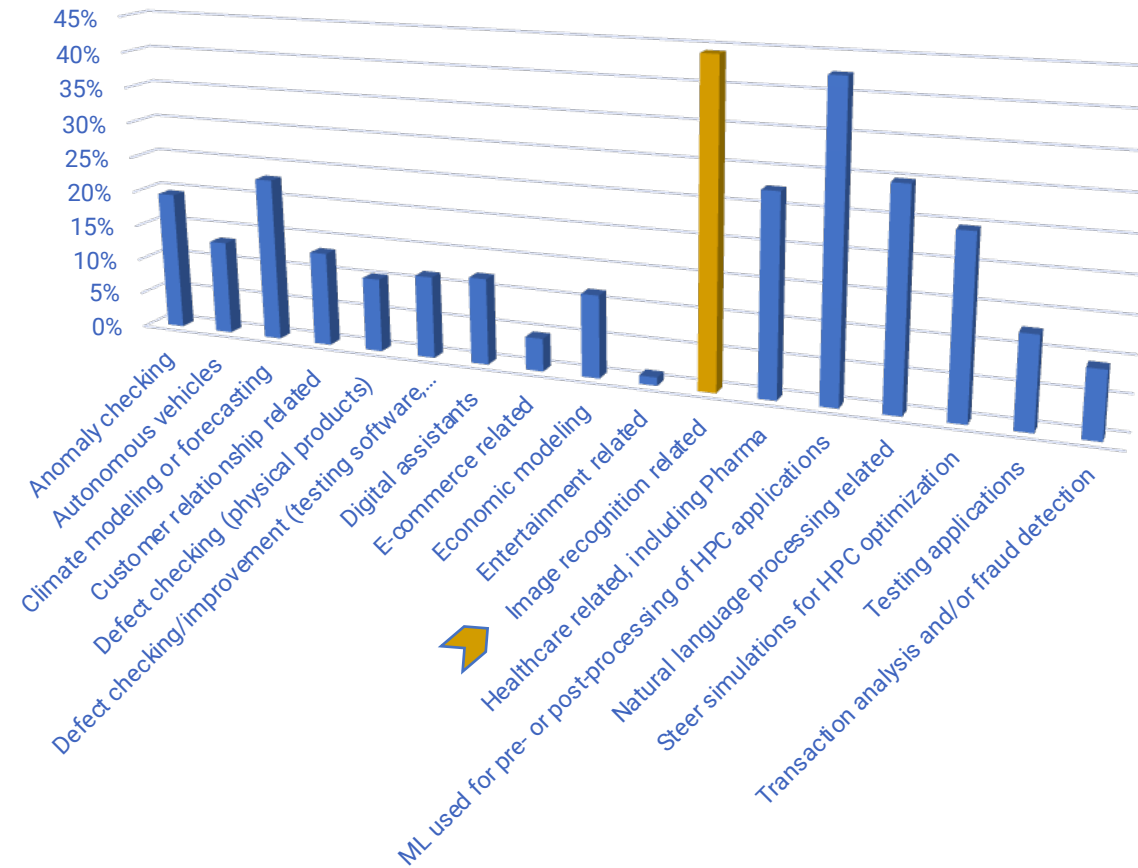
ML Application Development Stages



Less than one in five respondents reported having machine learning applications in full production.



What ML Is Used For

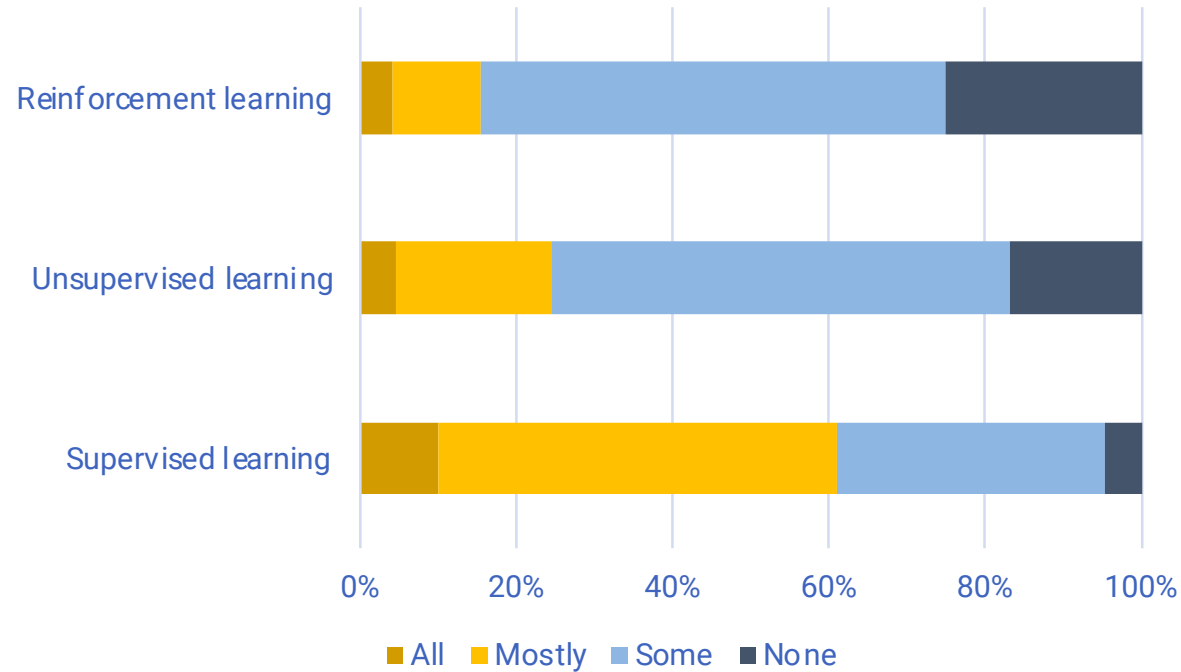


Surveyed organizations covered a wide range of application domains.

ML is broadly used for image recognition, including in blended HPC-AI environments

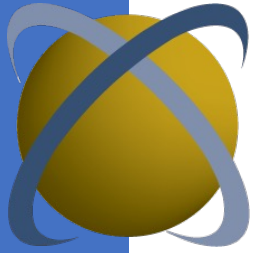


Supervised vs. Unsupervised Learning

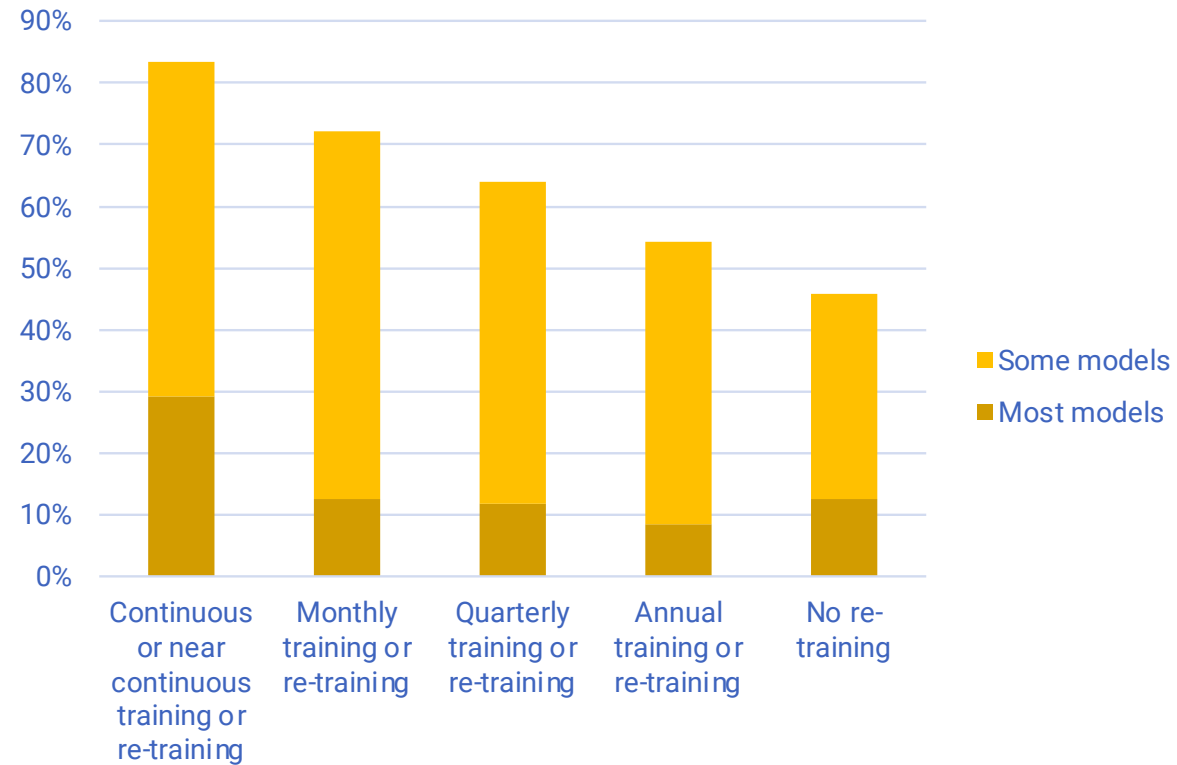


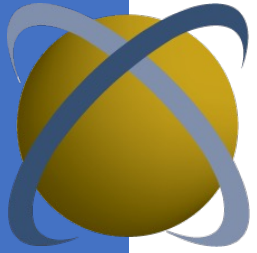
Supervised learning is much more common than unsupervised.

[Chart ignores "Don't know" responses]

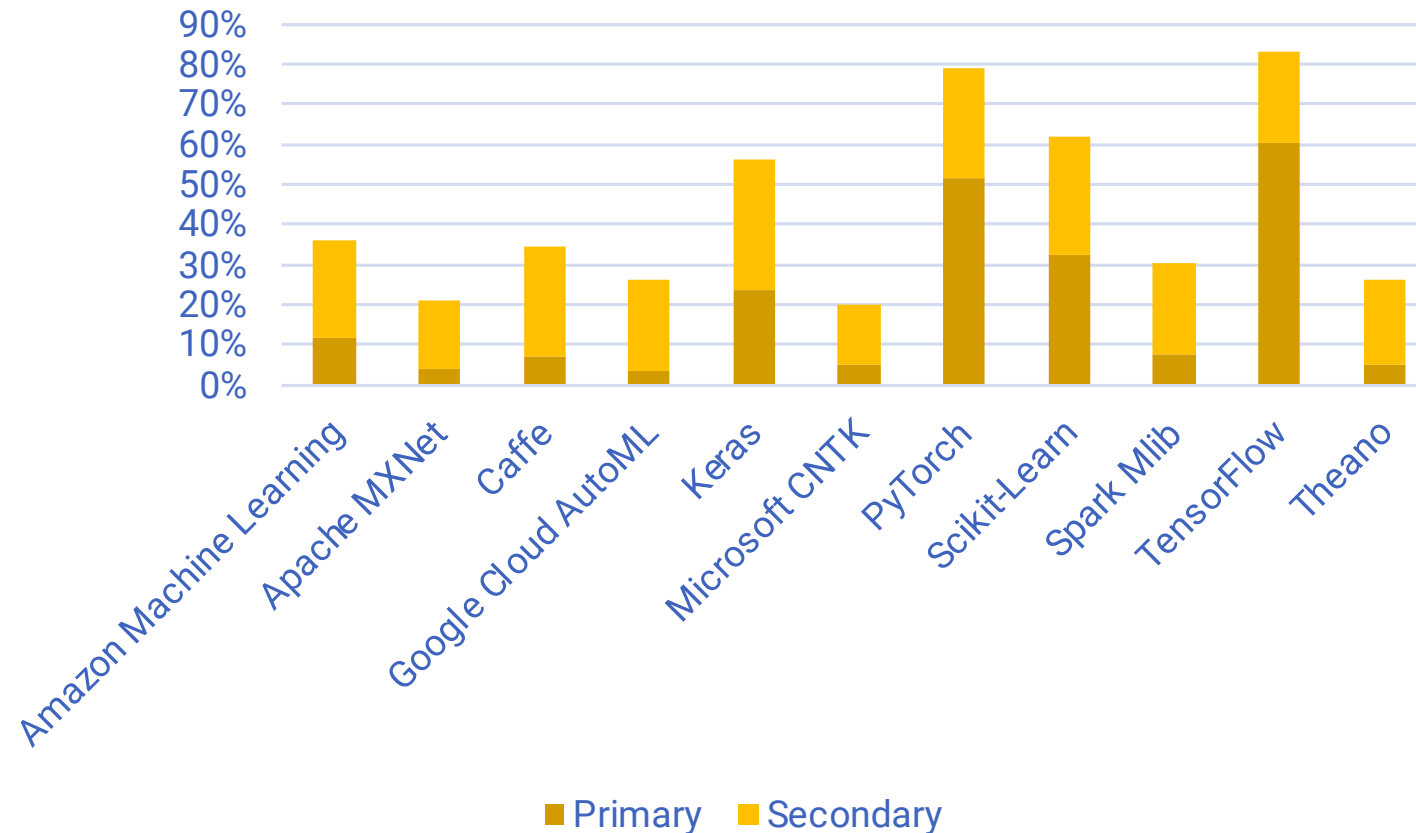


Re-training Frequency

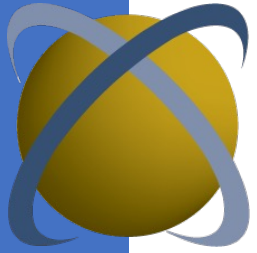




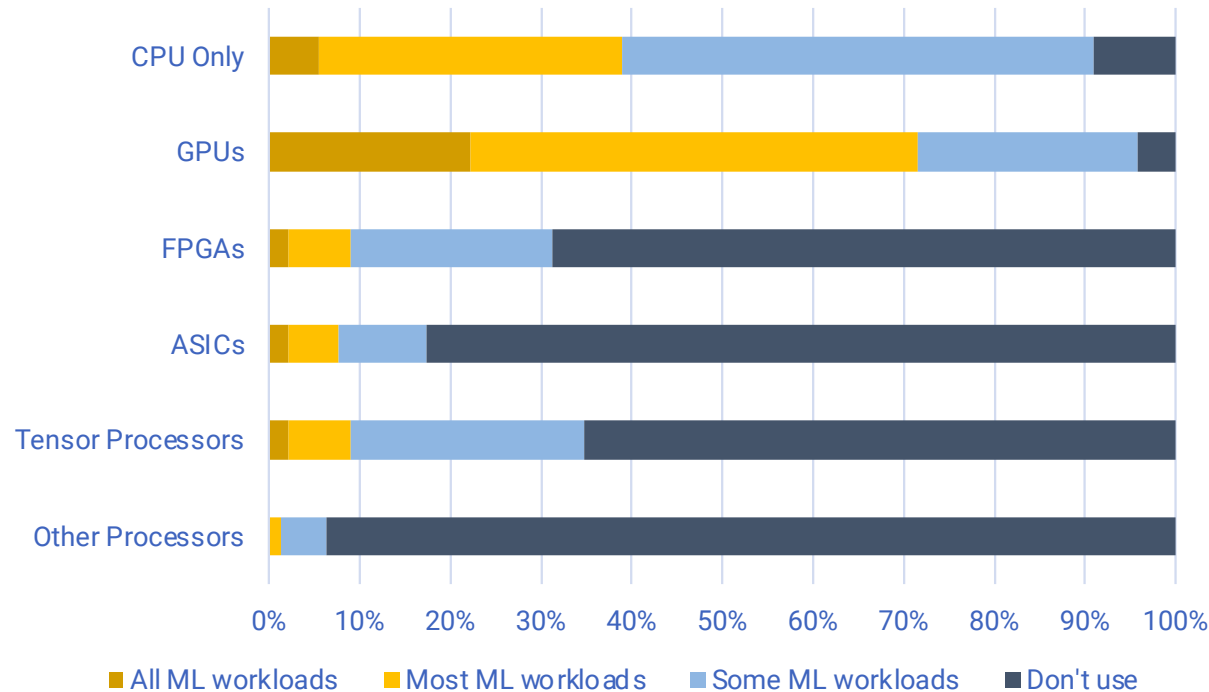
Frameworks Currently in Use



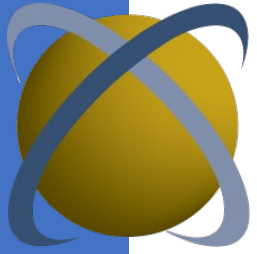
Most sites use multiple frameworks, including multiple that are considered “primary”



Processing Elements for ML



ML has a high affinity to GPUs; most sites use GPUs for most ML workloads. Some are run CPU-only. Other processing elements are uncommon.



Intersect360 Research at ISC23

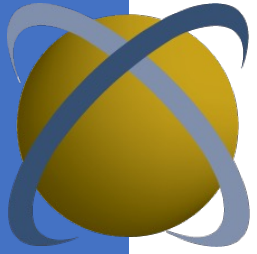
Vendor Showdown

Fishbowl Panel



SME Panel

HPC Trends and
Solutions Showcase



Thank you!

Q&A



Intersect360
R E S E A R C H