HYPERTION RESEARCH UPDATE:
Research Highlights In HPC, HPDA-AI, Cloud Computing, Quantum Computing, and Innovation Award Winners

SC19

Earl Joseph, Steve Conway, Bob Sorensen, and Alex Norton
Important Notes

- If you want a copy of the slides:
    - or *leave a business card*

- Check out our websites:
  - [www.hpcuserforum.com](http://www.hpcuserforum.com)
Hyperion Research helps organizations make effective decisions and seize growth opportunities by providing research and recommendations in both high performance computing and emerging technology areas.
Agenda

1. Recent Hyperion Research Results
2. Update on the HPC Market
3. Cloud Computing Update
   • We are declaring 2019 as the "Tipping Point" year for using public clouds for HPC jobs
4. Quantum Computing Update
5. The Exascale Race
6. The SC19 Innovation Award Winners
7. Conclusions and Predictions
Hyperion Research HPC Activities

• Track all HPC servers sold each quarter
  • By 28 countries
• 4 HPC User Forum meetings each year
• Publish 85 plus research reports each year
• Visit all major supercomputer sites & write reports
• Assist in collaborations between buyers/users and vendors
• Assist governments in HPC plans, strategies and direction
• Assist buyers/users in planning and procurements
• Maintain 5 year forecasts in many areas/topics
• Develop a worldwide ROI measurement system
• HPDA program (includes ML/DL/AI)
• HPC Cloud usage tracking
• Quarterly tracking of GPUs/accelerators
• Cyber Security
• Quantum Computing
• Map applications to algorithms to architectures

© Hyperion Research 2019
Important Dates For Your Calendar

- **2020 HPC USER FORUM MEETINGS:**
  - March 30–April 1 in Princeton, New Jersey
  - September 8-10 in Tucson, Arizona
  - International meetings will be set soon

- Register at [http://hpcuserforum.com/events.html](http://hpcuserforum.com/events.html)
We Want To Thank The HPC User Forum Steering Committee Members!

Paul Muzio  
Chairman, Industry Expert

Rupak Biswas  
NASA Ames  
Vice Chairman

Earl Joseph  
Executive Director, Hyperion Research

Vijay Agarwala  
Virginia Tech.

Alex Akkerman  
Ford Motor Company

Doug Ball  
HPC Expert

Mike Bernhardt  
Exascale Computing Project

Steve Conway  
Vice President, Hyperion Research

Steve Finn  
Emagine IT

Merle Giles  
Moonshot Research

Keith Gray  
BP

Arno Kolster  
Providentia Worldwide

Doug Kothe  
Oak Ridge National Laboratory

Jysoo Lee  
KAUST

David Martin  
Argonne National Laboratory

Jeff Broughton  
NERSC/Lawrence Berkeley National Lab

Paul Buerger  
Industry Expert

Clayton Chandler  
Credit Suisse Group AG

Candace Culhane  
Los Alamos National Labs

Sharan Kalwani  
Industry Specialist Consultant

Vince Scarafino  
Industry Expert

Suzy Tichenor  
Oak Ridge National Laboratory

Michael Resch  
HLRS, University of Stuttgart

Ryan Quick  
Providentia Worldwide

Stephane Requena  
GENCI

» SEE ALL BIO’s
Sample Research Results
ARM Processor Forecast For Use In HPC Servers

Note: This is our first full forecast for ARM-based processors – all comments and suggestions are welcomed!

© Hyperion Research 2019
Tipping Points: How Quickly HPC Buyers Can Change

Source: Hyperion Research, 2018
AI Is Still Near the Start

Today: Bounded Problems
- Many observations but few choices to make
- “One trick dogs”: 10 AI solutions in a box to solve 10 problems
- Already very useful:
  - Image & voice recognition
  - Advanced driver assistance
  - Reading an MRI

Future: Unbounded, Too
- Many observations, many choices to make
- Versatile decision-makers capable of serious experiential learning
- Examples:
  - Discerning human motivation
  - Fully automated driving
  - Diagnosing/“curing” a cancer

Low IQ (Weak Inferencing) High IQ (Strong Inferencing)

© Hyperion Research 2019
High Growth Areas: HPDA-AI

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

Table 1
Forecast: Worldwide HPC-Based AI Revenues vs Total HPDA Revenues (\$ Millions)

<table>
<thead>
<tr>
<th></th>
<th>2018</th>
<th>2019</th>
<th>2020</th>
<th>2021</th>
<th>2022</th>
<th>2023</th>
<th>CAGR</th>
</tr>
</thead>
<tbody>
<tr>
<td>WW HPC Server Revenues</td>
<td>13,706</td>
<td>14,495</td>
<td>15,780</td>
<td>17,376</td>
<td>18,983</td>
<td>19,947</td>
<td></td>
</tr>
<tr>
<td>Total WW HPDA Server Revenues</td>
<td>3,153</td>
<td>3,598</td>
<td>3,932</td>
<td>4,737</td>
<td>5,467</td>
<td>6,450</td>
<td>15.4%</td>
</tr>
<tr>
<td>Total HPC-Based AI</td>
<td>747</td>
<td>938</td>
<td>1,094</td>
<td>1,399</td>
<td>1,810</td>
<td>2,725</td>
<td>29.5%</td>
</tr>
</tbody>
</table>

Source: Hyperion Research 2019

Table 2
Forecast: Worldwide ML, DL & Other AI HPC-Based Revenues (\$ Millions)

<table>
<thead>
<tr>
<th></th>
<th>2018</th>
<th>2019</th>
<th>2020</th>
<th>2021</th>
<th>2022</th>
<th>2023</th>
<th>CAGR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML in HPC</td>
<td>532</td>
<td>675</td>
<td>875</td>
<td>1130</td>
<td>1479</td>
<td>1940</td>
<td>29.5%</td>
</tr>
<tr>
<td>DL in HPC</td>
<td>177</td>
<td>216</td>
<td>301</td>
<td>392</td>
<td>510</td>
<td>665</td>
<td>30.3%</td>
</tr>
<tr>
<td>Other AI in HPC</td>
<td>38</td>
<td>47</td>
<td>66</td>
<td>80</td>
<td>95</td>
<td>120</td>
<td>25.9%</td>
</tr>
<tr>
<td>Total</td>
<td>747</td>
<td>938</td>
<td>1,242</td>
<td>1,602</td>
<td>2,084</td>
<td>2,725</td>
<td>29.5%</td>
</tr>
</tbody>
</table>

Source: Hyperion Research 2019
Coupled Environments

- **Automated Driving Systems**
  - Embedded processor for local control (car-car, car-environment)
  - Private cloud for citywide and beyond (“air traffic control”)

- **Healthcare/Precision Medicine**
  - Healthcare systems are already private cloud-based.
  - Future: couple in-office HPC decision-support engine to private cloud.

- **5G Will Reduce Local-Cloud Latency Issue**
Tracking New AI Hardware: Emergence of AI-Specific Hardware Ecosystem
Cloud Application Assessment Tool

This tool is designed to help engineers, IT professionals, HPC users and buyers, and business executives understand the extent to which their workloads are amenable to operation on a public cloud. The tool is broken down into a series of questions that delve into the specific characteristics of a particular workload, allowing a user to assign a numeric score for each, rated on a scale of 0 to 10.

For each characteristic, a value of 10 means it would be better suited to a public cloud, 5 mean no clear advantage on either platform, and a 0 would be better suited to an on-premise system. Each characteristic includes a short description of what a 10, 5, and 0 value entails.

The purpose of this tool is not to promote either option, but rather to provide the user with an unbiased assessment of where their application could be best operationally suited. A high score at the end of the assessment means that the application is friendly to the public clouds, albeit with on-premise systems still being a viable option. Likewise, an overall application score that leans towards the on-premise option, does not mean the application cannot go to the public clouds, but rather that a public clouds may not be the best suited option for that specific type of workload today.

*After you have filled in answers for all 13 characteristics, you will be presented with your score in a printable scorecard.*

Continue
Our Pulse Survey Methodology
The Pulse Survey Process

- Pulses are fast-out, lower cost, smaller sample size surveys
  - Designed to provide a quick read on major trends
  - (vs. more detailed statistical results)

- Hyperion Research works with the client to identify important topic(s) and then:
  - Develops the initial survey design
  - Reviews it with client
  - Runs the survey against a panel of custom-qualified respondents
  - Delivers slides with data tables and analysis
  - Synthesizes the pulse research results into a PPT summary of Pulse
What is the MOST IMPORTANT reason why your organization has run one or more MPI HPC workloads in a public cloud? (Choose ONLY ONE)

<table>
<thead>
<tr>
<th>Reason</th>
<th>Percent</th>
<th>Cloud Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>We don't run any MPI HPC workloads in public clouds</td>
<td>54.0%</td>
<td>n/a</td>
</tr>
<tr>
<td>Elasticity (ability to expand cloud resources)</td>
<td>14.0%</td>
<td>30.4%</td>
</tr>
<tr>
<td>Performance</td>
<td>6.0%</td>
<td>13.0%</td>
</tr>
<tr>
<td>Need to network to outside world</td>
<td>6.0%</td>
<td>13.0%</td>
</tr>
<tr>
<td>Fast access (avoid waiting in on-premise queue)</td>
<td>4.0%</td>
<td>8.7%</td>
</tr>
<tr>
<td>Price</td>
<td>2.0%</td>
<td>4.3%</td>
</tr>
<tr>
<td>Not sure/don't know</td>
<td>14.0%</td>
<td>30.4%</td>
</tr>
</tbody>
</table>

Commentary

- Elasticity is most often named as the MOST IMPORTANT reason for using public clouds to run MPI workloads.
For which general methodologies do you apply GPUs? (Choose ALL that apply)

- **Modeling and simulation**: 85.9% (73 responses)
- **Machine learning**: 64.7% (55 responses)
- **Deep learning using neural networks**: 58.8% (50 responses)
- **Graph analytics**: 21.2% (18 responses)
- **Other**: 15.3% (13 responses)
- **Not sure/don’t know**: 2.4% (2 responses)

Although M&S was the leading choice, ML and DL weren’t far behind and graphing was also important.
Recent Study Summary: Exploring Options for a Bespoke Supercomputer Targeted for Weather and Climate Workloads
Recent Study Summary

NASA’s earth system models have unique high performance computing (HPC) requirements, which can differ greatly from standard industry offerings.

- The gap between vendor HPC solutions and earth system models has been growing, such that these models can exploit less and less of the peak computing capability of current HPC systems.

The primary focus of the study was to gather key insights, through a series of surveys with weather and climate users and potential HPC suppliers, on options available to NASA, and others, to develop a bespoke HPC system specifically targeted for weather/climate research.
Study Structure

The study was divided into two major parts:

• The first phase centered on a series of interviews held with expert researchers and users in the HPC-based climate and weather community to collect thoughts and insights on current and future operation requirements as well as the specific HPC hardware, software, and architectures needed to meet those workloads
  ▪ For phase one, 15 different weather/climate organizations in the US and overseas were surveyed including ECMWF, LANL, NOAA, ORNL, UCAR, and the University of Delaware
• Phase two consisted of taking the results of phase one to generate a second survey of HPC suppliers and independent HPC designers to assess the challenges and opportunities of developing a bespoke HPC to meet the phase one requirements
  ▪ HPC suppliers providing input included: Cray Inc, Dell EMC, HPE, and IBM

© Hyperion Research 2019
Key Recommendations

- Improved Benchmarks Are Needed
- Embrace New Workloads
- Modernize and Refactor Codes
- Explore New Hardware and Software Options
- Conduct Codesign Conference
  - Generate relevant projections of advances in future HPC hardware and software could better align with NASA requirements but that are in keeping with larger commercial trends
  - Explore opportunities to leverage any NASA hardware or software developments into the larger weather/climate HPC ecosystem or even other HPC verticals to help vendors justify the use of non-COTS hardware or software
  - Enlist vendor insights or other USG organizations on the best ways for NASA to begin refactoring and modernizing their current software base to best take advantage of existing and planned commercial technology and HPC products
HPC Market Update
Top Trends in HPC

The first half on 2019 is looking strong with $6.7 Billion in HPC server sales

- 2018 was a very strong year with over 15% growth, $13.7 billion

Many new buyers from the enterprise world are moving into HPC

- For: Fraud/anomaly detection, Business intelligence, Affinity marketing, Personalized medicine, Smart cities, and IoT

The profusion of Exascale announcements are generating a lot of buzz

Big data combined with HPC is creating new solutions

- Adding many new users/buyers to the HPC space
- AI/ML/DL & HPDA are the hot new areas
The Worldwide HPC Server Market: $6.7 Billion in First Half 2019

- Strong revenues!

HPC Servers $6.7B

- Supercomputers (Over $500K) $2.6B
- Divisional ($250K - $500K) $1.2B
- Departmental (100K - $250K) $1.9B
- Workgroup (under $100K) $.9B

© Hyperion Research 2019
## 2018 HPC Market By Verticals

<table>
<thead>
<tr>
<th>WW HPC Systems Revenue by Applications</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bio-Sciences</td>
<td>1,245,865</td>
</tr>
<tr>
<td>CAE</td>
<td>1,521,850</td>
</tr>
<tr>
<td>Chemical Engineering</td>
<td>205,891</td>
</tr>
<tr>
<td>DCC &amp; Distribution</td>
<td>780,184</td>
</tr>
<tr>
<td>Economics/Financial</td>
<td>746,418</td>
</tr>
<tr>
<td>EDA / IT / ISV</td>
<td>984,887</td>
</tr>
<tr>
<td>Geosciences</td>
<td>1,029,041</td>
</tr>
<tr>
<td>Mechanical Design</td>
<td>63,137</td>
</tr>
<tr>
<td>Defense</td>
<td>1,403,164</td>
</tr>
<tr>
<td>Government Lab</td>
<td>2,616,822</td>
</tr>
<tr>
<td>University/Academic</td>
<td>2,420,440</td>
</tr>
<tr>
<td>Weather</td>
<td>560,631</td>
</tr>
<tr>
<td>Other</td>
<td>127,757</td>
</tr>
<tr>
<td>Total Revenue</td>
<td>13,706,088</td>
</tr>
</tbody>
</table>

Source: Hyperion 2019
## HPC Market By Vendor Shares

<table>
<thead>
<tr>
<th>OEM</th>
<th>2018 Sales ($ Millions)</th>
<th>Share %</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPE/HP</td>
<td>4,766</td>
<td>34.8%</td>
</tr>
<tr>
<td>Dell EMC</td>
<td>2,857</td>
<td>20.8%</td>
</tr>
<tr>
<td>IBM</td>
<td>971</td>
<td>7.1%</td>
</tr>
<tr>
<td>Lenovo</td>
<td>957</td>
<td>7.0%</td>
</tr>
<tr>
<td>Inspur</td>
<td>788</td>
<td>5.8%</td>
</tr>
<tr>
<td>Sugon (Dawning)</td>
<td>462</td>
<td>3.4%</td>
</tr>
<tr>
<td>Cray</td>
<td>313</td>
<td>2.3%</td>
</tr>
<tr>
<td>Fujitsu</td>
<td>269</td>
<td>2.0%</td>
</tr>
<tr>
<td>Penguin</td>
<td>244</td>
<td>1.8%</td>
</tr>
<tr>
<td>NEC</td>
<td>201</td>
<td>1.5%</td>
</tr>
<tr>
<td>Atos</td>
<td>150</td>
<td>1.1%</td>
</tr>
<tr>
<td>Other</td>
<td>1,728</td>
<td>12.6%</td>
</tr>
<tr>
<td>Grand Total</td>
<td>13,706</td>
<td>100.0%</td>
</tr>
</tbody>
</table>
## SUPERCOMPUTER Market Trends by Vendor

<table>
<thead>
<tr>
<th>Mfr</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPE</td>
<td>1,202</td>
<td>1,248</td>
<td>1,202</td>
<td>1,402</td>
</tr>
<tr>
<td>Dell EMC</td>
<td>54</td>
<td>553</td>
<td>876</td>
<td>1,105</td>
</tr>
<tr>
<td>Lenovo</td>
<td>170</td>
<td>259</td>
<td>247</td>
<td>318</td>
</tr>
<tr>
<td>IBM</td>
<td>214</td>
<td>293</td>
<td>366</td>
<td>770</td>
</tr>
<tr>
<td>Cray</td>
<td>584</td>
<td>461</td>
<td>250</td>
<td>313</td>
</tr>
<tr>
<td>SGI</td>
<td>88</td>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fujitsu</td>
<td>106</td>
<td>169</td>
<td>185</td>
<td>228</td>
</tr>
<tr>
<td>NEC</td>
<td>138</td>
<td>58</td>
<td>81</td>
<td>112</td>
</tr>
<tr>
<td>Inspur</td>
<td>490</td>
<td>420</td>
<td>455</td>
<td></td>
</tr>
<tr>
<td>Penguin</td>
<td>136</td>
<td>159</td>
<td>244</td>
<td></td>
</tr>
<tr>
<td>Atos</td>
<td>42</td>
<td>64</td>
<td>91</td>
<td>117</td>
</tr>
<tr>
<td>Sugon</td>
<td>13</td>
<td>24</td>
<td>61</td>
<td>132</td>
</tr>
<tr>
<td>Wuxi</td>
<td></td>
<td>300</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>630</td>
<td>339</td>
<td>420</td>
<td>166</td>
</tr>
<tr>
<td><strong>Grand Total</strong></td>
<td><strong>3,241</strong></td>
<td><strong>4,438</strong></td>
<td><strong>4,358</strong></td>
<td><strong>5,362</strong></td>
</tr>
</tbody>
</table>

© Hyperion Research 2019
# SUPERCOMPUTER Market Trends By Region

<table>
<thead>
<tr>
<th>Data</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum of WW HPC server N.A. Rev.</td>
<td>1,793</td>
<td>1,977</td>
<td>2,101</td>
<td>2,551</td>
</tr>
<tr>
<td>Sum of WW HPC server EMEA Rev.</td>
<td>647</td>
<td>797</td>
<td>892</td>
<td>1,132</td>
</tr>
<tr>
<td>Sum of WW HPC server Asia/Pac Rev. (w/Japan)</td>
<td>811</td>
<td>1,513</td>
<td>1,303</td>
<td>1,629</td>
</tr>
<tr>
<td>Sum of WW HPC server ROW Rev.</td>
<td>21</td>
<td>56</td>
<td>61</td>
<td>50</td>
</tr>
</tbody>
</table>

© Hyperion Research 2019
## The Broader HPC Market Forecast

<table>
<thead>
<tr>
<th>Revenues by the Broader HPC Market Areas</th>
<th>2018</th>
<th>2023</th>
<th>CAGR 18-23</th>
</tr>
</thead>
<tbody>
<tr>
<td>Server</td>
<td>13,706,088</td>
<td>19,979,016</td>
<td>7.8%</td>
</tr>
<tr>
<td>Storage</td>
<td>5,547,188</td>
<td>7,771,184</td>
<td>7.0%</td>
</tr>
<tr>
<td>Middleware</td>
<td>1,582,892</td>
<td>2,217,801</td>
<td>7.0%</td>
</tr>
<tr>
<td>Applications</td>
<td>4,627,492</td>
<td>6,413,592</td>
<td>6.7%</td>
</tr>
<tr>
<td>Service</td>
<td>2,229,921</td>
<td>2,858,820</td>
<td>5.1%</td>
</tr>
<tr>
<td>Total Revenue</td>
<td>27,693,580</td>
<td>39,240,413</td>
<td>7.2%</td>
</tr>
</tbody>
</table>

*Source: Hyperion 2019*
New HPC Cloud Computing Study: Key Findings
HPC Cloud Use Has Ramped Up Quickly and Has Substantial Room for Growth

- From a survey of HPC users that are using the cloud today:
  - On average, they run 33% of their HPC workloads in 3rd-party clouds. Extrapolation to entire HPC community: 20% of workloads
  - 40% of respondents say all their HPC workloads could be run in 3rd-party clouds.
    - Not the same as saying the should be run there.
  - 60% say some jobs aren’t appropriate to run in 3rd-party clouds (too mission-critical, secure, etc.)
  - Ultimate limit may be data locality: large data volumes stored near the jobs/computing resources needing them.
The HPC Cloud Computing Market Is Worth Special Attention from CSPs

- HPC has become a sizable market.
- HPC is at the forefront of R&D for economically important AI use cases.
  - HPC shows where the mainstream AI market is headed.
The HPC and Enterprise Markets Are Also Converging in the Cloud

Established HPC: mostly batch-oriented, upstream R&D

Enterprise HPC: near-real time business operations

<table>
<thead>
<tr>
<th>Where Do You Run HPC workloads? (Choose ALL that apply)</th>
</tr>
</thead>
<tbody>
<tr>
<td>On-premise HPC data center</td>
</tr>
<tr>
<td>On-premise enterprise data center (business operations)</td>
</tr>
<tr>
<td>More than one external cloud (e.g., Amazon, Google, Microsoft)</td>
</tr>
<tr>
<td>On-premise private or hybrid cloud</td>
</tr>
<tr>
<td>One external cloud (e.g., Amazon, Google, Microsoft)</td>
</tr>
<tr>
<td>Not sure/don’t know</td>
</tr>
<tr>
<td>Other</td>
</tr>
</tbody>
</table>
Most of the Surveyed HPC Sites Use Hybrid Clouds Linked to CSPs

- 58% of respondents exploit hybrid clouds for HPC.
  - The CSP end may be public or private.
  - Half of all new HPC private clouds are on 3rd-party clouds.
The Fastest-Growing Workloads Are in AI (Machine and Deep Learning)

- 87% of the surveyed CSPs and 94% of the cloud-offering system vendors said AI workloads are the fastest-growing for their cloud services.

- Where they say more investment is needed:
  - Data transport
  - Simulation for AI
HPC Cloud Forecast
Major Trend

HPC in the Cloud

- We are declaring 2019 as a major growth year for HPC in public clouds
- It represents a "tipping point" or an inflection point in the growth curve
- So we have updated our forecasts for HPC computing in public clouds …
This 2019 Cloud Forecast

- This forecast incorporates new data into the 2018 HPC Cloud forecast.
- This forecast DOES NOT include HPC users who are cloud-born, i.e., those who do not own or do any HPC work on-premise and use third-party hosted cloud resources only.
- When we say public cloud, we mean any third party hosted compute resources, this may include public clouds, private clouds hosted by a third party, or hybrid cloud environments.
- To be clear, this forecast is of HPC SPEND on third party cloud resources.
  - The amount end users spend in public clouds
  - It’s not a measure of the hardware within the clouds
Major Trend
New Public Cloud Forecast

© Hyperion Research 2019
Major Trend
New Public Cloud Forecast

<table>
<thead>
<tr>
<th>Application</th>
<th>2018</th>
<th>2023</th>
<th>5 Year CAGR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bio-Sciences</td>
<td>370</td>
<td>1,155</td>
<td>25.6%</td>
</tr>
<tr>
<td>CAE</td>
<td>296</td>
<td>894</td>
<td>24.7%</td>
</tr>
<tr>
<td>Chemical Engineering</td>
<td>62</td>
<td>180</td>
<td>23.9%</td>
</tr>
<tr>
<td>DCC &amp; Distribution</td>
<td>141</td>
<td>443</td>
<td>25.8%</td>
</tr>
<tr>
<td>Economics/Financial</td>
<td>197</td>
<td>625</td>
<td>25.9%</td>
</tr>
<tr>
<td>EDA</td>
<td>202</td>
<td>643</td>
<td>26.0%</td>
</tr>
<tr>
<td>Geosciences</td>
<td>210</td>
<td>700</td>
<td>27.3%</td>
</tr>
<tr>
<td>Mechanical Design</td>
<td>25</td>
<td>72</td>
<td>23.9%</td>
</tr>
<tr>
<td>Defense</td>
<td>296</td>
<td>863</td>
<td>23.9%</td>
</tr>
<tr>
<td>Government Lab</td>
<td>234</td>
<td>719</td>
<td>25.2%</td>
</tr>
<tr>
<td>University/Academic</td>
<td>345</td>
<td>907</td>
<td>21.3%</td>
</tr>
<tr>
<td>Weather</td>
<td>49</td>
<td>147</td>
<td>24.4%</td>
</tr>
<tr>
<td>Other</td>
<td>39</td>
<td>72</td>
<td>12.8%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>2,466</td>
<td>7,418</td>
<td>24.6%</td>
</tr>
</tbody>
</table>

*Source: Hyperion Research, 2019*
Summary of On-Going Quantum Computing Program
Research Goals and Objectives

• Locate major QC & QT programs around the world
• Track and access these quantum computing developments and quantum technologies around the world to:
  • Develop and document a worldwide R&D and market view
  • Track QC-related technologies in computers, communications, sensors, networks, metrology, fabrication, etc.
  • Track main research objectives, staff sizes, track records, areas that they are researching, etc.
  • Track global investments, including trends in spending and numbers of researchers, by geography
  • Collect and catalog real world use cases – and those planned
• Focus on international activities, e.g., Canada, China/Asia, Europe, Australia, Japan, Russia, etc.
  • Plus, smaller and new entrants
Comparisons of National Level of Effort

Quantum Computing and Communications R&D Publications 2008-2018

Quantum Computing R&D Publications 2008-2018

Quantum Communications R&D Publications 2008-2018

© Hyperion Research 2019
Developing a QT Ranking System

- Goal is to use multiple 1-5 point rating methods to assess the value of a particular QT effort spanning:
  - Key Researchers
  - Key Facilities
  - Key Technology Areas
  - Key Programs
  - Key Commercials Development

- Identify most promising QT R&D efforts
- Support assessments among and across national level activities
## Data Collection Success:
### Country Efforts Collected to Date (474)

<table>
<thead>
<tr>
<th>Country</th>
<th>Efforts Collected</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td></td>
</tr>
<tr>
<td>Russia</td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td></td>
</tr>
<tr>
<td>Switzerland</td>
<td></td>
</tr>
<tr>
<td>Austria</td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td></td>
</tr>
<tr>
<td>Finland</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td></td>
</tr>
<tr>
<td>EU</td>
<td></td>
</tr>
<tr>
<td>France</td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td></td>
</tr>
<tr>
<td>Spain</td>
<td></td>
</tr>
</tbody>
</table>

© Hyperion Research 2019
The Exascale Race
Our Forecast On When & Where Exascale Systems Will Be Installed

<table>
<thead>
<tr>
<th>Year Accepted</th>
<th>China</th>
<th>EU</th>
<th>Japan</th>
<th>US</th>
<th>Total Installations</th>
<th>Total Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020</td>
<td>1 (Post K Accepted)</td>
<td>1 pre-exascale</td>
<td>1 pre-exascale</td>
<td>1 pre-exascale</td>
<td>3-4</td>
<td>$750 Million</td>
</tr>
<tr>
<td>2021</td>
<td>1 pre-exascale</td>
<td>1 (Post K Accepted)</td>
<td>1 pre-exascale</td>
<td>4-5</td>
<td>$1,900 Million</td>
<td></td>
</tr>
<tr>
<td>2022</td>
<td>1 or 2 exascale</td>
<td>1 near-exascale</td>
<td>1 near-exascale</td>
<td>2 exascale</td>
<td>4-5</td>
<td>$1,700 Million</td>
</tr>
<tr>
<td>2023</td>
<td>1 exascale</td>
<td>1 exascale</td>
<td>1 near-exascale ($100 million)</td>
<td>1 or 2 exascale</td>
<td>4</td>
<td>$1,500 Million</td>
</tr>
<tr>
<td>2024</td>
<td>1 exascale</td>
<td>1 exascale</td>
<td>?</td>
<td>2 exascale</td>
<td>4</td>
<td>$1,400 Million</td>
</tr>
<tr>
<td>2025</td>
<td>2 exascale</td>
<td>1 or 2 exascale</td>
<td>1 near-exascale ($100 million)</td>
<td>1 exascale</td>
<td>5-6</td>
<td>$1,600 Million</td>
</tr>
</tbody>
</table>

Source: Hyperion Research 2019

And perhaps an early one in the UK!
Pre-Exascale and Exascale System Rollouts 2020-2025

Hyperion Research projects that major government-sponsored efforts will drive the development of about 26 near-exascale and exascale systems in the 2020 to 2025 time frame, with a total price tag of around $9 billion

- In 2021, the combined major HPC development entities, China, the EU, Japan, and the United States, could start installing 4 to 5 near-exascale or exascale systems costing a total of close to $2 billion, an historic high-water mark
- Subsequently, the sector will see an average of 4 to 5 new high-end HPCs per year out to 2025, with total annual revenues in the $1.4 to $1.7 billion range

It is unclear to what extent future (beyond 2025) large supercomputers will continue to be purchased in the $500 million plus price range or will return to more "modest" price levels in the $100 to $250 million price range per system
Near-term US Exascale Plans

The dust is settling on the first three US exascale systems to be rolled out in the next three plus years -- with a total budget of $1.8 billion:

- **Aurora:**
  - DOE Argonne National Laboratory (Intel Prime/Cray Sub)
  - Delivery in late 2021, acceptance in 2022
  - Cray Shasta architecture with Intel Xeons and Intel Xe GPU

- **Frontier:**
  - DOE Oak Ridge National Laboratory (Cray Prime)
  - Delivery in late 2021 and acceptance in 2022
  - Cray Shasta with AMD EPYC CPU and future Radeon GPUs

- **El Capitan:** (Cray Prime)
  - DOE NNSA’s LLNL
  - Delivery in late 2022, with full production targeted for late 2023
  - Based on Cray Shasta architecture
China Exascale Plans

Currently there are three Chinese exascale systems under development

- One or more prototypes may be selected for full-up production
- NUDT
  - Indigenous CPU, possibly Arm-based Phytium Xiaomi or Fujitsu A64FX
  - MT-2000+ NUDT accelerator (or follow-on)
  - 400 Gbps homegrown network
- Sugon
  - Heterogeneous architecture (2 CPU/2 DCU accelerators per node)
  - Hygon processor is licensed clone of AMD Gen 1 EPYC processor
  - Hygon-developed accelerator
  - Liquid immersion cooling
- Sunway (prototype specifications)
  - CPU-only SW26010 chip follow-on (260 cores @ 3Tflops per chip)

*China is driving to have the first exascale system in the world with delivery in late 2020 or 2021*
EU HPC Plans

- EuroHPC program stood up in 2018 to develop EU-wide HPC development program
  - 28 participating States + EU
  - Operational Duration: 11/2018-2026
- Three Sites Recently Selected for 150-200 Pflops systems
  - Kajaani Finland, Barcelona Spain, and Bologna Italy
  - Total Investment: 650 million Euros
    - 50% EU, 50% Consortium
- Five Sites Selected for Medium Range HPCs (~4Pflops)
  - Investment” ~180 million Euros
- Systems are owned by EuroHPC Joint Undertaking
- Installations to start 4Q2020
EU Exascale Plans Going Forward

- EU Plan calls for acquisition of two exascale systems in the 2022-2023 time frame
  - At least one to use European technology: specifically using an EPI-developed processor
- Post Exascale System around 2027
  - Plans call out for including integration and deployment of the first hybrid HPC/quantum infrastructure in Europe
The UK has not joined EuroHPC and will likely not be eligible to fully take part in EuroHPC projects or access calls when Horizon 2020 ends next year

- A Strategic Case was made in August 2018 to develop an Exascale Outline Business case for the UK Government
- Requirements Highlighted:
  - Support Mode/Sim and AI/DL
  - Support scientific and industrial users
- Operational by Q422/Q123
- Cost estimated to be £700m - £1.2 bn
- *Stand by for more details*
Japan’s Exascale System

Riken’s Fugaku (Post-K System) is currently in its early installation stage

- Uses Fujitsu A64 ARM processor
  - 48 compute cores with GPU-like vector extensions
  - HMB2 on package memory
  - On-chip networks and PCIe connections
- 150K+ single socket nodes
- Tofu-D bandwidth 10X total global CSP traffic
- Peak DP > 400PFs, Peak SP> 800Pf, Peak HP > 1600 Flops
- Typically 37X faster that predecessor K system on target co-design applications
- Targeting 1H 2021 full production capability

*It will likely have best efficiency rating of all 1st round exascale systems*
The HPC Innovation Award and Winners
Our Award Program:
https://www.hpcuserforum.com/innovationaward/

HPC Innovation Awards

The HPC Innovation Awards are given twice a year to organizations that have made outstanding achievements using high performance computing.

The three award categories showcase ROI and success stories showing HPC’s impact on increasing economic value, advancing scientific innovation and engineering progress, and on improving the quality of life worldwide.

Learn More

© Hyperion Research 2019
Two Types Of HPC Innovation Awards

1. HPC end-user Innovation awards – showing the impact of their use of HPC
   • Two at SC19

2. For vendors for HPC innovations – that can make HPC systems better
   • Two at SC19
HPC Award Program Goals

- **#1 Help to expand the use of HPC by showing real ROI examples:**
  - Expand the “Missing Middle” – SMBs, SMSs, etc. by providing examples of what can be done with HPC
  - 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
  - Creating 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
Examples Of Previous Winners

© Hyperion Research 2019
SC19 Winners: HPC Innovation Excellence Award

Note: the award now includes a $1,000 prize
First time using a series of high spatial resolution physical based models for regional climate, surface hydrology and coastal flooding to build a dataset that provides estimates of changes in the risk of flooding, which was used by AT&T to develop a risk and resiliency tool.
Automated Segmentation of Knee Bone and Cartilage Combining Statistical Shape Knowledge and Convolutional Neural Networks

The Max Delbruck Center for Molecular Medicine provided an infrastructure for Zuse Institute Berlin to combine Convolutional Neural Networks and 3D Statistical Shape Knowledge to automate the identification of Osteoarthritis on a DDN A3I system.
For innovated AI hardware:
Cerebras has built the largest chip ever, which contains more than 1.2 trillion transistors and 400,000 compute cores to solve AI problems, that used to be solved in months, in minutes.
New Generation of Silicon Cube Blade System

Sugon

For innovative HPC system design:
Sugon has developed an efficient phase-change liquid cooling technology and combined it with their blade technology for a system with very high computing density and energy efficiency.
We Invite Everyone
To Apply For The Next Round Of
Innovation Awards!
In Summary:
Some Predictions
For the Next Year Or So
The Exascale Race Will Drive New Technologies

- The global ES race is boosting funding for the Supercomputers market segment and creating widespread interest in HPC
- Exascale systems are being designed for HPC, AI, HPDA, etc.
  - This will drive new processor types, new memories, new system designs, new software, etc.
- And (in some cases) that HPC is too strategic to depend on foreign sources
  - This has led to indigenous technology initiatives
The HPC-Enterprise Market Convergence Will Drive HPC Products into the Broader Enterprise Market

- Hyperion Research studies confirm that this convergence is occurring and speeding up
- Competitive forces are driving companies to aim more-complex questions at their data structures and push business operations closer to real time
- Important HPC capabilities: scalable parallel processing, ultrafast data movement and ultra-large, capable memory systems
- The HPC and commercial sectors are also converging around a shared need to extremely data-intensive AI-ML-DL workloads, both on the simulation and analytics side

© Hyperion Research 2019
Many New Processors/Accelerators Are on The Way

- Choices of processing elements (CPUs, accelerators) will increase
  - x86 will remain the dominant HPC CPU, but indigenous CPUs will gain ground
- In AI, startups and large companies are developing processors designed for analytics workloads
- NVIDIA is the dominant accelerator today
- Processors exploiting ARM IP are planned for Europe (EPI), Japan (Post-K computer) and China
Storage Systems Will Increasingly Become More Critical

- Data-intensive HPC is driving new storage requirements
- Iterative methods will expand the size of data volumes needing to be stored
- Future architectures will allow computing and storage to happen more pervasively on the HPC infrastructure
- Metadata management will deal with data stored in multiple geographic locations and environments
- Physically distributed, globally shared memory will become more important
- More intelligence will need to be built into storage software
Artificial Intelligence Will Grow Faster Than the HPC Market As A Whole

- The AI market is at an early stage but already highly useful (e.g., visual and voice recognition)
  - Once better understood, there are many high value use cases that will drive adoption
- Advances in inferencing will reduce the amount of training needed for today's AI tasks, but the need for training will grow to support more challenging tasks
- The trust (transparency) issue that strongly affects AI today will be overcome in time
- Learning models (ML, DL) have garnered most of the AI attention, but graph analytics will also play a crucial role with its unique ability to handle temporal and spatial relationships
Important Notes

- **If you want a copy of the slides:**
    - or *leave a business card*

- **Check out our websites:**
  - [www.hpcuserforum.com](http://www.hpcuserforum.com)
Thank You!

QUESTIONS?
INFO@HYPERIONRES.COM
THE ROI From HPC & Success Stories

Economic Models Linking HPC and ROI

ROI Study: Latest Results

These are the latest results of the ROI study that measures how HPC investments are related to improved economic success and increased scientific innovation.

HPC User Forum thanks DOE for its insights, guidance and funding of this research project.

<table>
<thead>
<tr>
<th>Sector</th>
<th>Sum of Total Jobs Added</th>
<th>Average of Revenue $ per HPC $</th>
<th>Average of Profit or Cost Saving $ per HPC $</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic</td>
<td>14</td>
<td>264.8</td>
<td>$110 M</td>
</tr>
<tr>
<td>Government</td>
<td>42</td>
<td>817.1</td>
<td>$121 M</td>
</tr>
<tr>
<td>Industry</td>
<td>2,279</td>
<td>468.2</td>
<td>$37 M</td>
</tr>
<tr>
<td>Grand Total</td>
<td>2,335</td>
<td>$458</td>
<td>$58</td>
</tr>
</tbody>
</table>

© Hyperion Research 2019