



HYPERION RESEARCH

Major Trends in the Worldwide HPC Market

HPC User Forum

April 2017

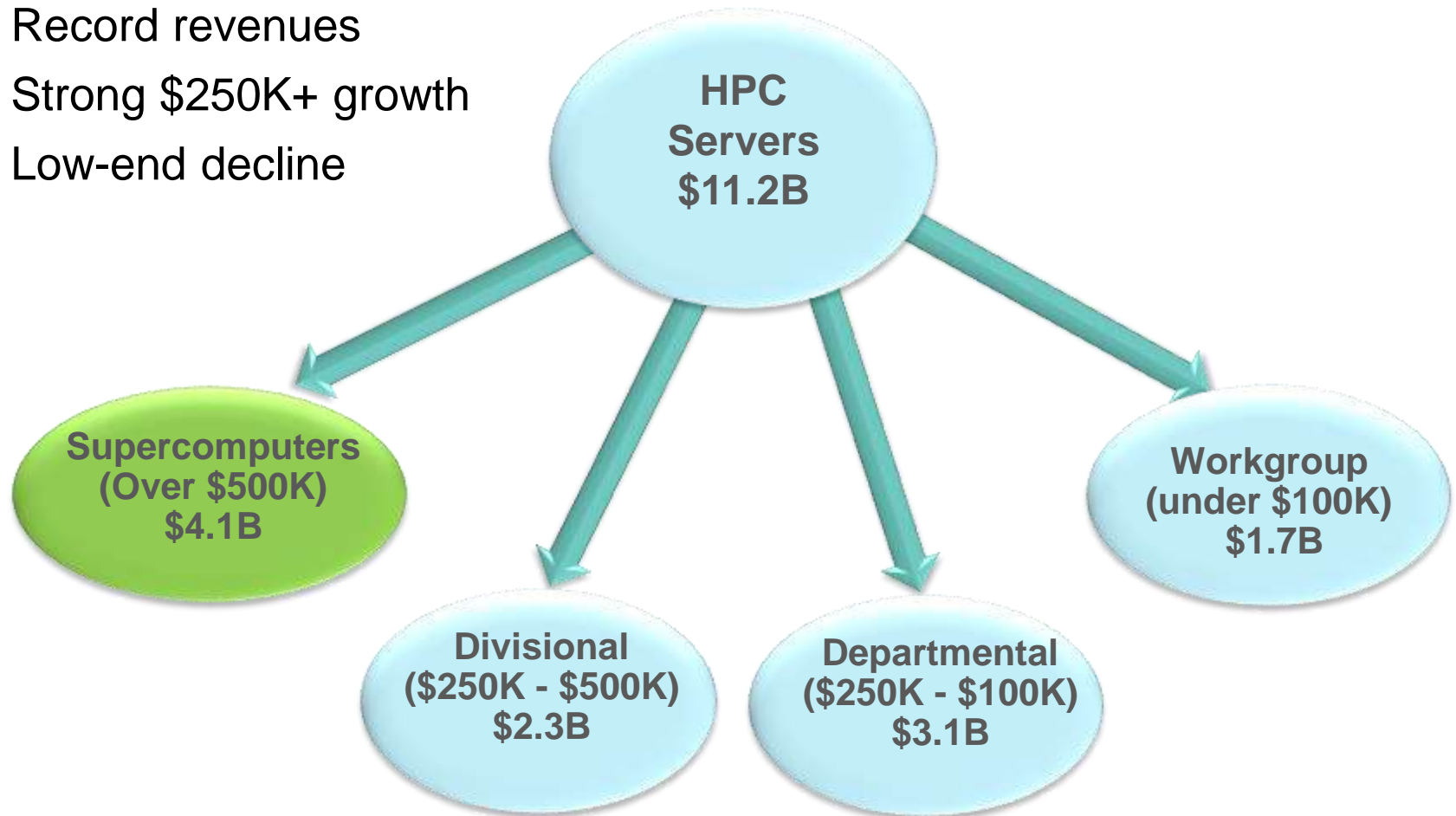
Earl Joseph and Steve Conway



THE 2016 HPC MARKET RESULTS, GROWTH PROJECTIONS AND TRENDS

The Worldwide HPC Server Market: \$11.2 Billion in 2016

- Record revenues
- Strong \$250K+ growth
- Low-end decline



2016 HPC Market By Vendor (\$ Millions)

| 2016 Market Results | |
|----------------------------|---------------|
| HPE/HP | 3,878 |
| Dell | 2,014 |
| Lenovo | 909 |
| IBM | 492 |
| Cray | 461 |
| Sugon | 315 |
| Fujitsu | 226 |
| SGI | 169 |
| NEC | 166 |
| Bull Atos | 118 |
| Other | 2,453 |
| Total | 11,200 |

2016 HPC Market By Vertical (\$ Millions)

| 2016 Market Results | |
|-------------------------------|---------------|
| Bio-Sciences | 1,049 |
| CAE | 1,251 |
| Chemical Engineering | 183 |
| DCC & Distribution | 692 |
| Economics/Financial | 624 |
| EDA / IT / ISV | 823 |
| Geosciences | 844 |
| Mechanical Design | 57 |
| Defense | 1,125 |
| Government Lab | 2,059 |
| University/Academic | 1,934 |
| Weather | 490 |
| Other | 70 |
| Total Revenue | 11,200 |

2016 HPC Market Forecasts (\$ Millions)

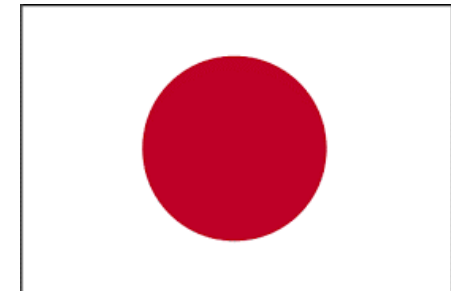
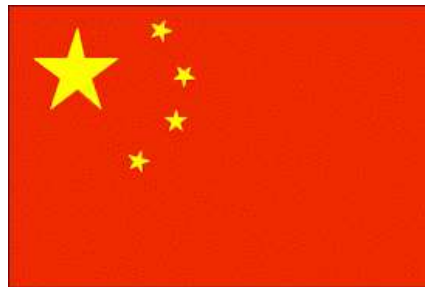
| | 2016 | 2021 |
|------------------------------|-------------|-------------|
| Supercomputer | 4,091 | 5,356 |
| Divisional | 2,273 | 2,902 |
| Departmental | 3,147 | 4,274 |
| Workgroup | 1,689 | 2,287 |
| Total | 11,200 | 14,819 |
| <i>Source: Hyperion 2017</i> | | |

Forecast: The Broader HPC Market (\$ Millions)

| | 2016 | 2021 | CAGR 16-21 |
|------------------------------|---------------|---------------|-----------------------|
| Server | 11,200 | 14,819 | 5.8% |
| Storage | 4,316 | 6,269 | 7.8% |
| Middleware | 1,277 | 1,786 | 6.9% |
| Applications | 3,739 | 5,071 | 6.3% |
| Service | 1,907 | 2,309 | 3.9% |
| Total Revenue | 22,439 | 30,253 | 6.2% |
| <i>Source: Hyperion 2017</i> | | | |

HPC Major Trend

The Global Exascale Race



IDC-Projected Exascale Dates and Suppliers

U.S.



- Sustained ES: 2023
- Peak ES: 2021
- Vendors: U.S.
- Processors: U.S.
- Initiatives: NSCI/ECP
- Cost: \$300-500M per system, plus heavy R&D investments

EU



- Sustained ES: 2023-24
- Peak ES: 2021
- Vendors: U.S., Europe
- Processors: U.S., ARM
- Initiatives: PRACE, ETP4HPC
- Cost: \$300-\$350 per system, plus heavy R&D investments

China



- Sustained ES: 2023
- Peak ES: 2020
- Vendors: Chinese
- Processors: Chinese (plus U.S.?)
- 13th 5-Year Plan
- Cost: \$350-500M per system, plus heavy R&D

Japan



- Sustained ES: 2023-24
- Peak ES: Not planned
- Vendors: Japanese
- Processors: Japanese
- Cost: \$600-850M, this includes both 1 system and the R&D costs...will also do many smaller size systems

IDC-Projected Exascale Investment Levels (In Addition to System Purchases)

U.S.



- \$1 to \$2 billion a year in R&D (including NRE)
- Investments by both governments & vendors
- Plans are to purchase multiple exascale systems

EU



- About 5 billion euros in total
- Investments in multiple exascale and pre-exascale systems
- Investments mostly by country governments with a little from the EU

China



- Over \$1 billion a year in R&D
- Investments by both governments & vendors
- Plans are to purchase multiple exascale systems each year
- Already investing in 3 pre-exascale systems by 2017/18

Japan



- Planned investment of just over \$1 billion* (over 5 years) for both the R&D and purchase of 1 exascale system
- To be followed by a number of smaller systems ~\$100M to \$150M each
- Creating a new processor and a new software environment

* Note that this includes both the system and R&D

HPC Major Trend

High Performance Data Analysis (HPDA): HPC-Big Data Convergence



HPC Moves into Industry (Examples)

MODELING & SIMULATION

DATA ANALYTICS



Government & Academic Research



The collage features logos for the following companies and organizations:

- Automotive:** Ford, Mercedes-Benz (DAIMLER), BOEING, AIRBUS, Panasonic
- Energy:** BÖRSE FRANKFURT, TOTAL, bp, Shell
- Retail:** Walmart
- Technology:** P&G, GEICO, Baidu, facebook, PayPal, Amazon, UNITED STATES POSTAL SERVICE
- Healthcare:** MERCK, SICOS, MAYO CLINIC
- Other:** NOVARTIS, DREAMWORKS ANIMATION SKG, Genentech

1970

1980

1990

2000

2010

Today

HPDA = Data-Intensive Computing Using HPC

Drivers:

- Competition
- Complexity
- Time

Modeling & Simulation

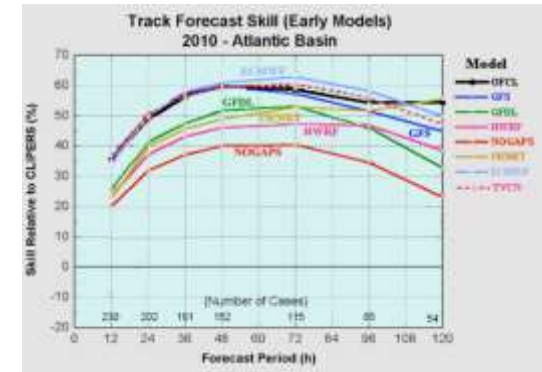
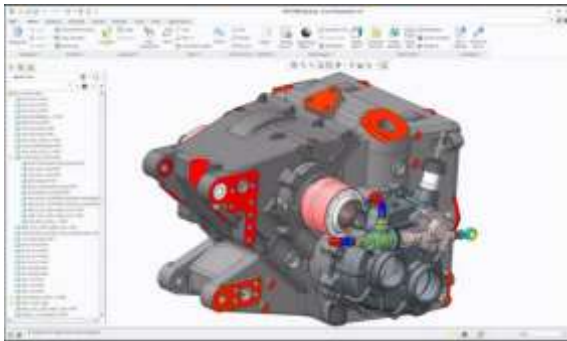
- Existing HPC users
 - Larger problem sizes
 - Higher resolution
 - Iterative methods
 - EP jobs to the cloud (Novartis)
- New commercial users
 - E.g., SMEs

Advanced Analytics

- Existing HPC users
 - Intelligence community, FSI
 - Data-driven science/engineering (e.g., biology)
 - Knowledge discovery
 - ML/DL, cognitive, AI
- New commercial users
 - Fraud/anomaly detection
 - Business intelligence
 - Affinity marketing
 - Personalized medicine

HPDA Includes Cumulative Results of Iterative Methods

- Parametric modeling (product design)
- Stochastic modeling (financial)
- Ensemble modeling (weather/climate)



HPDA Market Drivers

- **More input data (ingestion)**
 - More powerful scientific instruments/sensor networks
 - More transactions/higher scrutiny (fraud, terrorism)
- **More output data for integration/analysis**
 - More powerful computers
 - More realism
 - More iterations in available time
- **The need to pose more intelligent questions**
 - Smarter mathematical models and algorithms
- **Real time, near-real time requirements**
 - Catch fraud before it hits credit cards
 - Catch terrorists before they strike
 - Diagnose patients before they leave the office
 - Provide insurance quotes before callers leave the phone



HPDA Analytics → New HPC Segments

1. Fraud and anomaly detection.

- Government (intelligence, cyber security)
- Industry (credit card fraud, cyber security)

2. Affinity Marketing.

- Discern potential customers' demographics, buying preferences and habits.

3. Business intelligence.

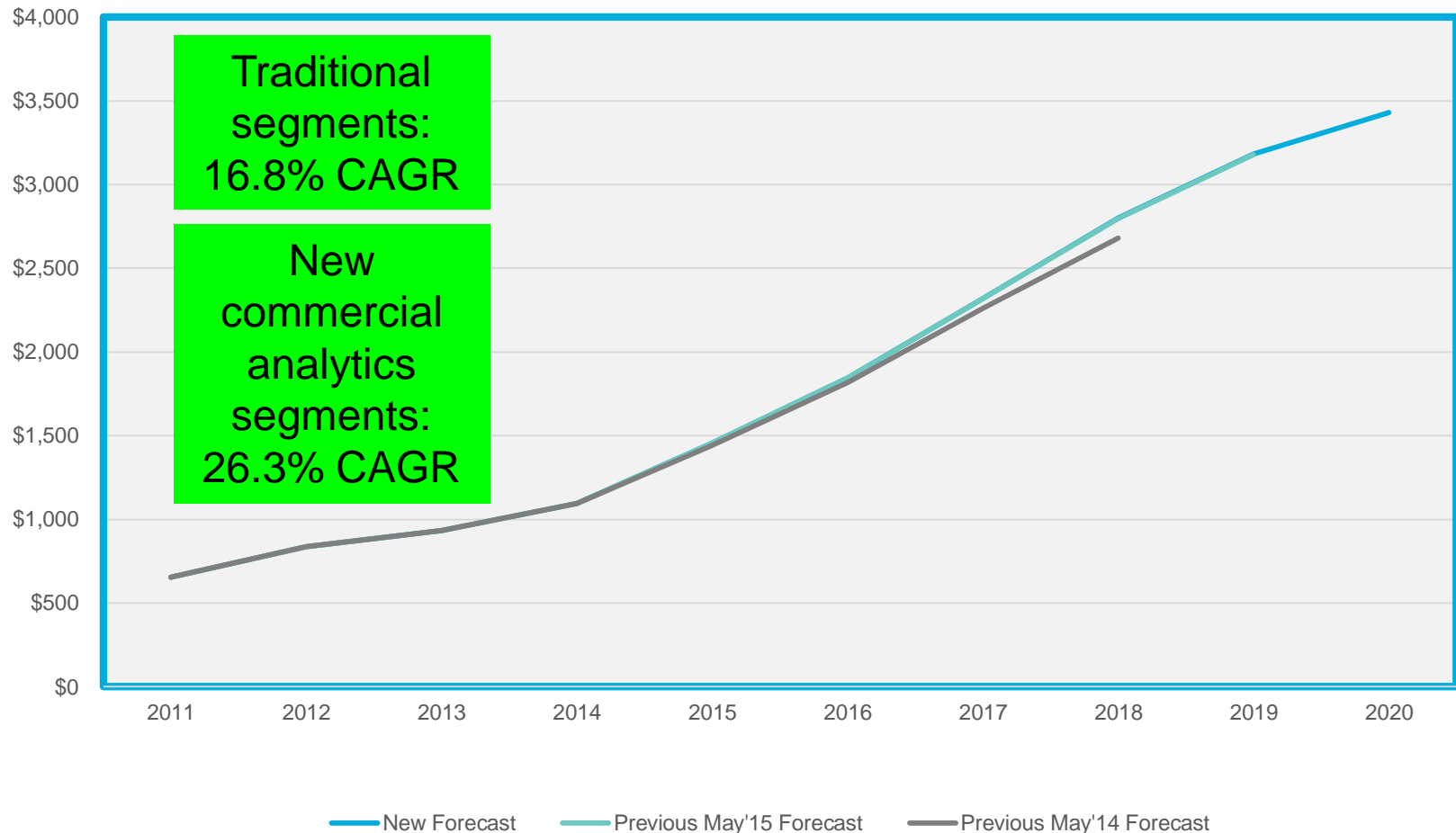
- Identify opportunities to advance market position and competitiveness

4. Precision Medicine

- Personalized approach to improve outcomes, control costs

HPDA Server Market Forecast to 2020

HPDA Server Forecast Comparisons (\$ Millions)



HPDA As a % of Total HPC Utilization (All Systems)

| Of all the workloads on your systems, what percentage of your total system utilization is devoted to Big Data analytics (MapReduce/Hadoop, graph analytics, semantic analysis, etc.)? | | | |
|---|------------------------------------|----------------------------------|----------------------------------|
| | Percentage of System in Government | Percentage of System in Academia | Percentage of System in Industry |
| Currently used for Big Data analytics - Percent | 18.2% | 8.4% | 32.6% |
| Expected 6 to 18 months from now - Percent | 19.8% | 15.6% | 33.4% |
| N = 128 | | | |

Source: IDC, 2015

Where HPDA Workloads Are Run

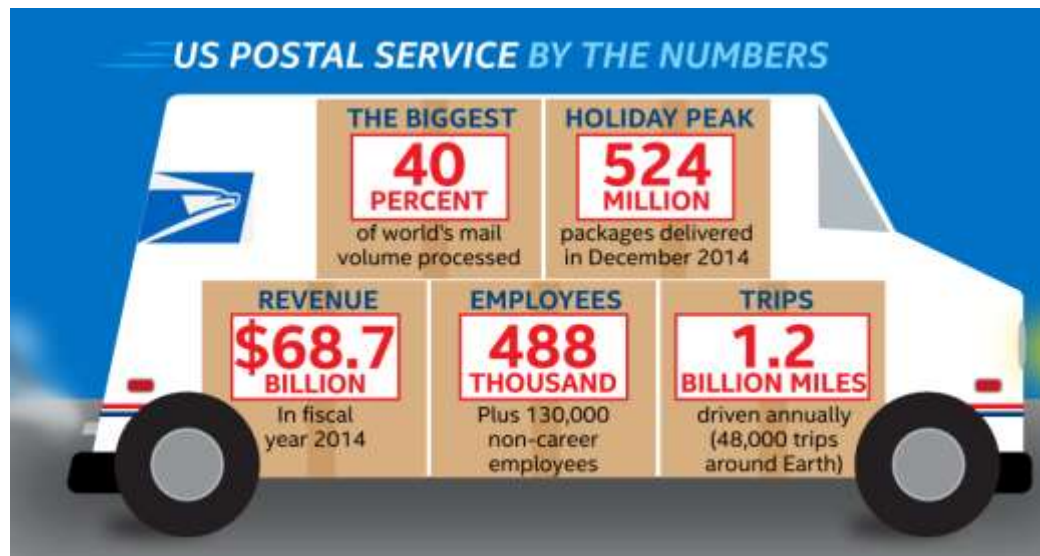
Type System Used for HPDA Workloads

| | Number of Responses | Percentage of Respondents |
|--|---------------------|---------------------------|
| We currently run these types of workloads on the same HPC system used for other types of workloads and we plan to continue as is for at least the next 6-18 months | 56 | 52.8% |
| We currently run these types of workloads on the same HPC system used for other types of workloads but we plan to migrate them over to a separate HPC system or Big Data appliance within the next 6-18 months | 25 | 23.6% |
| We currently run these types of workloads on a separate HPC system or Big Data appliance and we plan to continue as is for at least the next 6-18 months | 26 | 24.5% |
| We currently run these types of workloads on a separate HPC system or Big Data appliance but plan to migrate them over to the same HPC system used for other types of workloads within the next 6-18 months | 3 | 2.8% |
| Other | 3 | 2.8% |
| Total | 113 | |
| N = 106 | | |

Source: IDC, 2015

Processing in Memory

- **USPS reports processing very large data sets in memory can boost performance up to 6 orders of magnitude:**
- **Processing outside of memory can cut performance:**
 - ~50% moving data from one blade to neighboring blade
 - ~67% moving data from top blade to middle blade in a rack
 - 25x moving data from top blade to bottom blade in a rack
 - Much more when moving data between racks



Emerging HPDA Application: Smart City



HPDA Application:

Internet of Things (IoT)

■ IoT Network Management

- Wellness
- Security
- China's IoT plan

■ Dense local nodes

- Driverless traffic
- Smart power grid
- IoT edge computing

Not just the
Internet of Stupid
Things!



HPC Major Trend

Deep Learning, Machine Learning, AI



Machine Learning Healthcare Innovations



- Frauenklinik-Uni Mannheim (Germany): childbirth
- Childrens Mercy Hospital (U.S.): rare childhood diseases
- Frédéric Joliot Hospital (France): PET scan tracer optimization
- University of Oslo (Norway): bowel and prostate cancer
- Lobachevsky State University of Nizhni Novgorod (Russia): 3D virtual clones of patients
- University of Toronto SickKids Center (Canada): genomic medicine
- Neuroblastoma and Medulloblastoma Translational Research Consortium, Michigan (U.S.)
- University of Rochester Medical Center (U.S.): infant seizures
- Ariana Pharma (France): long-term cancer case management
- Victor Chang Cardiac Research Institute (Australia): heart arrhythmias

Deep Learning

IDC Research Findings



- The transition from ML (e.g., Jeopardy Watson) to deep learning (e.g., health care Watson) is difficult.
 - Baidu: Big difference between 95% and 99% accuracy. 99% needs ca. 25 exaops.
- The DL market is in research mode but expected to enter production mode in 18-24 months
- To be accurate, DL needs massive data volumes that aren't easily available yet in many markets outside of social media/Internet.
 - GlaxoSmithKline: "the healthcare data available today is miniscule"
 - Some organizations (e.g., UnitedHealth, IBM) are investing hundreds of millions (\$) in external data to build needed volumes
- To enter mainstream, must become interactive
- Use purpose-built workstation today to develop server-level data center solutions in 2-3 years.
- ML/DL software ecosystem is built around NVIDIA GPUs, but Intel Phi and FPGAs are gaining momentum (e.g., AWS F1)
- No standard benchmarks. Vendors need to express performance in terms of targeted domains:
 - E.g., cancer cell detections/second instead of flops or transactions/second
- AI is variously defined, but the forefront of AI research is moving to HPC
 - Facebook FAIR initiative: 22 supercomputers for 15 independent European research institutes, including TUM, Tübingen, IST Austria, ETHZ)

HPC Major Trend

HPC in the Cloud: Real-World Status and Prospects



Major Trend

HPC in the Cloud



- **64% of HPC sites run some jobs in public clouds**
 - Up from 13% in 2011
- **But only 7-8% of all their jobs (average)**
 - Not much changed since 2011
 - Public clouds are cost-effective for some jobs, but up to 10x more expensive for others
 - Key concerns: security, data loss
- **Private and hybrid cloud use is growing faster**
- **Big public clouds are going heterogeneous**
 - AWS with Ryft FPGAs, Google with NVIDIA GPGPUs

ROI Associated with HPC Investments (3-Year Global Study for DOE-Science, DOE-NNSA)



 **IDC**
Analyze the Future

SPECIAL STUDY

Creating Economic Models Showing the Relationship Between Investments in HPC and the Resulting Financial ROI and Innovation — and How It Can Impact a Nation's Competitiveness and Innovation

Earl C. Joseph, Ph.D. Steve Conway
Chirag Dekate, Ph.D.

IDC OPINION

The authors want to thank the U.S. Department of Energy (DOE) for its insights and guidance on and funding of this research project. This study is based upon work funded by the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research, under award number DE-SC0008540.

This pilot study investigates how high-performance computing (HPC) investments can improve economic success and increase scientific innovation. This research is focused on the common good and should be useful to DOE, other government agencies, industry, and academia. The study has created two unique economic models and an innovation index:

8200 F. 508.9315.4015 www.idc.com

Grant References

- **The authors thank DOE for its insights and guidance on and funding of this grant-based research project**
 - This study is based upon work funded by the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research, and the National Nuclear Security Administration, under award number DE-SC0012576.
- **DOE Program Managers:**
 - Christine Chalk, 301-903-5152, christine.chalk@science.doe.gov, and Barbara J. Helland, 301-903-3127, barbara.helland@science.doe.gov, U.S. Department of Energy Office of Science, Germantown Building, 1000 Independence Avenue, S.W. Washington, D.C., 20585-1290, and Doug Wade in National Nuclear security Administration (NNSA).
- **Administrator/Contracting Officer:**
 - Warren Riley, 630-252-2485, warren.riley@ch.doe.gov, U.S. Department of Energy Office of Acquisition and Assistance, 9800 South Cass Avenue, Argonne, Illinois, 60439
- **IDC Reporting:**
 - Principal investigator: Earl C. Joseph, Ph.D., 612-812-5798, ejoseph@idc.com
 - Senior technical project manager: John Daly, 508-935-4643, jdaly@idc.com

Rationale for Study

Absent the Cold War and with major economies still hurting, science and national security are necessary but no longer sufficient reasons for government investments in leadership-class supercomputers

- Leadership-class supercomputers can cost up to \$500 million (vs. \$30 million ca. 1990).
- Need for additional argument: economic competitiveness (“to out-compute is to out-compete”)
- Need for scientific and industrial/economic ROI to be quantified, no longer just assumed.
 - IDC developed new macroeconomic models to do this.

Current Database Demographics

- Total ROI & ROR examples (673):
 - Financial ROI examples = 148
 - Innovation ROR examples = 525

| Accomplishment Type | Count of Accomplishment Type |
|----------------------------|-------------------------------------|
| Financial ROI | 148 |
| Innovation | 525 |
| Grand Total | 673 |

Note: This is up from July 30th, 2015 when we were at 329

Note: We are finally at a stage in the research that we feel comfortable starting the publication process!

Current Database Demographics

- By country:

| Country | Count of Accomplishment Type |
|--------------------|------------------------------|
| China | 126 |
| France | 17 |
| India | 1 |
| U.S. | 47 |
| UK | 120 |
| US | 257 |
| Italy | 9 |
| Germany | 10 |
| Canada | 1 |
| Slovenia | 2 |
| Netherlands | 1 |
| Australia | 2 |
| Japan | 79 |
| South Africa | 1 |
| Grand Total | 673 |



Current Database Demographics

- By sector:

| Sector <input type="button" value="▼"/> | Count of Accomplishment Type |
|---|------------------------------|
| Academic | 362 |
| Government | 76 |
| Industry | 235 |
| Grand Total | 673 |



HPC ROI

Latest Findings: ROI from HPC is Very High

Results indicate high ROI returns resulting from investments in HPC

On average, from the latest data:

- \$551 dollars on average in revenue per dollar of HPC invested.
- \$52 dollars on average of profits (or cost savings) per dollar of HPC invested.

Financial ROI Model – By Industry

Results continue to indicate very substantial returns for investments in HPC:

- 2,335 jobs were created across these financial ROI projects

| Q3: Industry | Average of Revenue \$ per HPC \$ | Sum of Total Jobs Added | Average of Profit or Cost Saving \$ per HPC \$ | Count of Accomplishment Type |
|--------------------|----------------------------------|-------------------------|--|------------------------------|
| Academic | \$1,150 M | 14 | \$44 M | 21 |
| Defense | \$75 M | - | \$5 M | 2 |
| Financial | \$834 M | 602 | \$81 M | 31 |
| Government | \$1,206 M | 42 | \$141 M | 15 |
| Insurance | \$71 M | 5 | | 1 |
| Life Sciences | \$160 M | 48 | \$41 M | 13 |
| Manufacturing | \$83 M | 678 | \$20 M | 43 |
| O&G | \$419 M | 100 | \$46 M | 10 |
| Retail | \$30 M | 49 | \$12 M | 3 |
| Telecomm | \$10 M | 420 | \$10 M | 2 |
| Transportation | \$1,804 M | 377 | \$16 M | 7 |
| Grand Total | \$551 M | 2,335 | \$52 M | 148 |

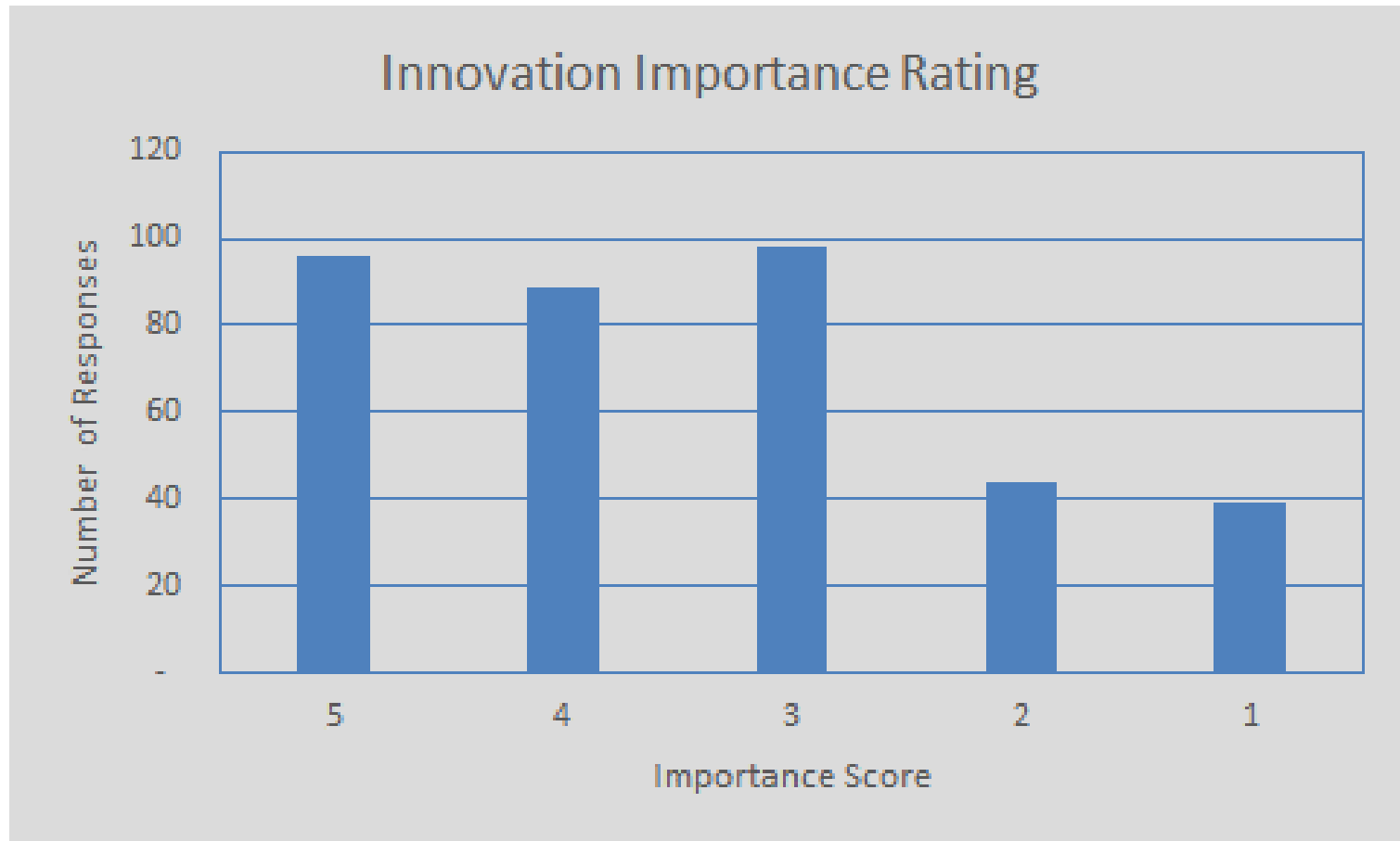


New Findings: Innovations – Basic vs. Applied

| Sector ▼ | Count of Q8: Basic | Count of Q8: Applied | Count of Q8: Basic/ Applied |
|--------------------|-------------------------------|---------------------------------|--|
| Academic | 265 | 80 | 341 |
| Government | 37 | 27 | 61 |
| Industry | 29 | 92 | 123 |
| Grand Total | 331 | 199 | 525 |



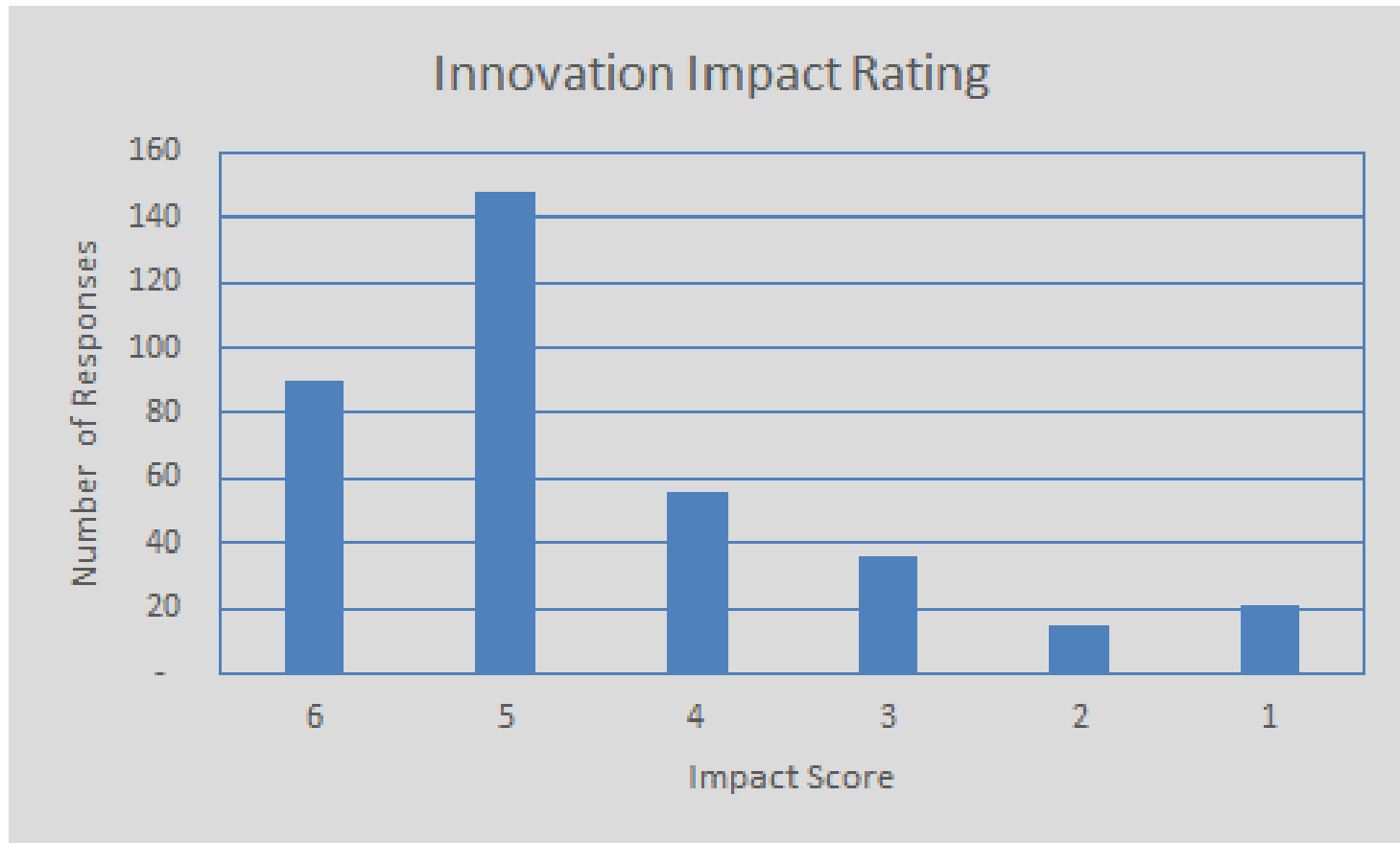
The Innovation **IMPORTANT**CE Index



- 5. One of the top 2 to 3 innovations in the last decade
- 4. One of the top 5 innovations in the last decade
- 3. One of the top 10 innovations in the last decade
- 2. One of the top 25 innovations in the last decade
- 1. One of the top 50 innovations in the last decade



New Findings: The Innovation IMPACT Index



6. It is useful to over 50 organizations
5. An innovation that is useful to 10 to 49 organizations
4. An innovation that is useful to 6 to 10 organizations
3. An innovation useful to 2 to 5 organizations
2. An innovation only useful to 1 organization
1. An innovation that is recognized ONLY by experts in the field

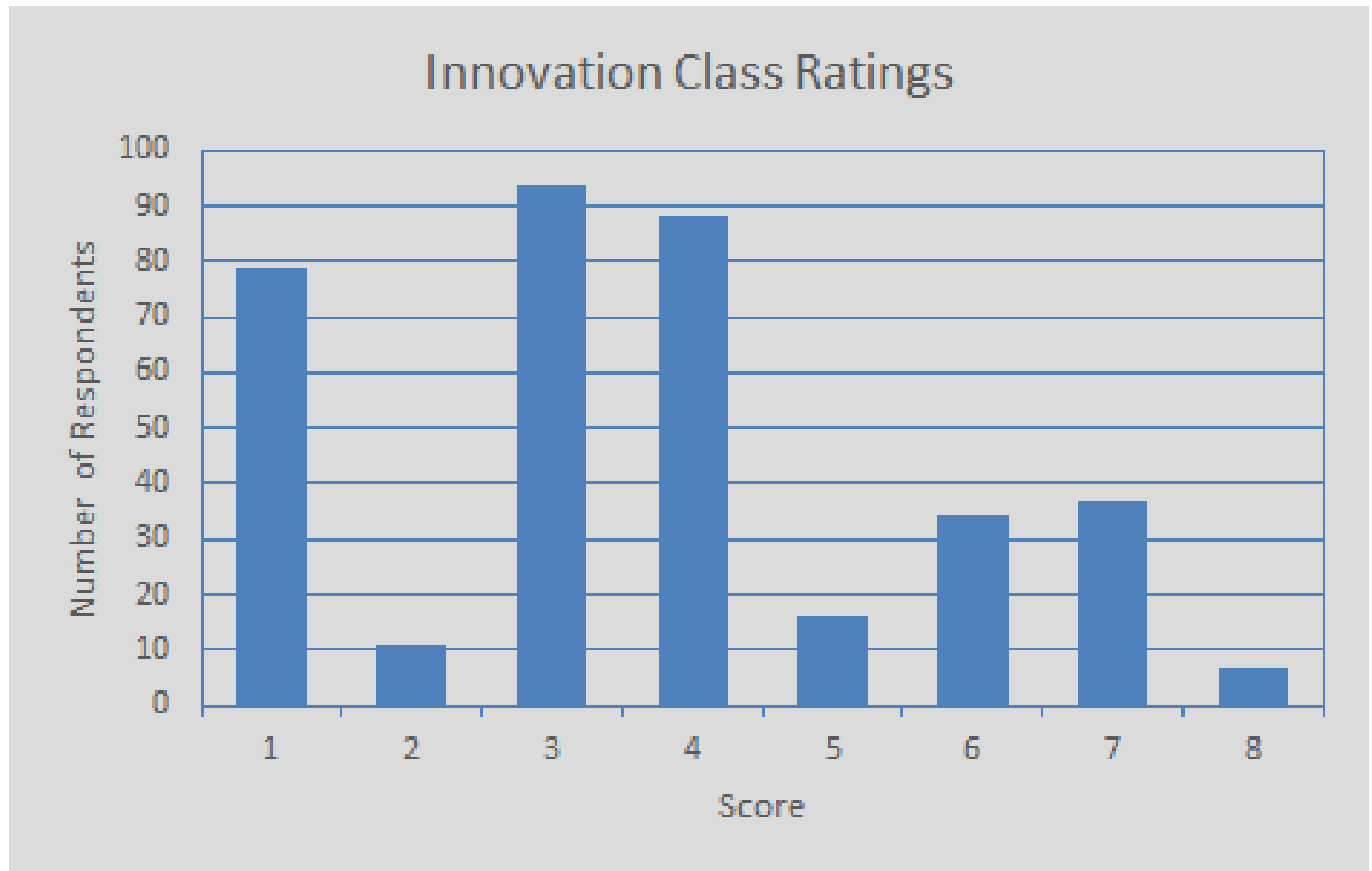


Changes Made in 2016: New Innovation “Class” Scale

- Expanded the innovation indexes by adding a new innovation “Class” scale:
 - Class 1 innovations – One of the top 2-3 innovations in a field over the last ten years PLUS useful to over 10 organizations
 - Class 2 innovations -- One of the top 5 innovations in a field over the last ten years PLUS useful to over 10 organizations
 - Class 3 innovations – One of the top 5 innovations in a field over the last ten years PLUS useful to at least 5 organizations
 - Class 4 innovations – One of the top 10 innovations in a field over the last ten years PLUS useful to at least 5 organizations
 - Class 5 innovations – One of the top 25 innovations in a field over the last ten years PLUS useful to at over 10 organizations
 - Class 6 innovations – One of the top 25 innovations in a field over the last ten years PLUS useful to at least 2 organizations
 - Class 7 innovations – One of the top 50 innovations in a field over the last ten years PLUS useful to at least 2 organizations
 - Class 8 innovations – The rest of the innovations in the study



The New INNOVATION CLASS Index



New Findings: The INNOVATION CLASS Index

Average innovation Class rating = 3.64

| Innovation Class | Count of Accomplishment Type | Average of HPC \$M per Innovation | Sum of Total Jobs Added |
|--------------------|------------------------------|-----------------------------------|-------------------------|
| 1 | 79 | 33.2 | 1,449 |
| 2 | 11 | 125.9 | 28 |
| 3 | 94 | 30.6 | 266 |
| 4 | 88 | 1.1 | 434 |
| 5 | 16 | 1.7 | 14 |
| 6 | 34 | 4.0 | 15 |
| 7 | 37 | 21.7 | 1,063 |
| 8 | 7 | 0.8 | - |
| Grand Total | 366 | 19.1 | 3,268 |

New Findings: Success Stories

Note that an additional outcome of this research is an expansive list of HPC success stories

- These can be used to help explain the importance of HPC to funding bodies, key decision makers and the broader public
- IDC will be writing up a number of them for broader dissemination



Download Results: www.hpcuserforum.com/ROI



IDC 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.

[Latest ROI Data](#)

[ROI Slide Deck](#)

ROI Pilot Study Overview

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

The study included creating two unique models:

1. A macroeconomic model which depicts how HPC investments result in economic advancements in the form of ROI, growth and jobs.
2. An "Innovation Index" that measures and compares innovation levels, based on the level of applying HPC computing resources towards scientific and technical advancement.

[The Full Report](#)

[The ROI Models](#)

[HPC Success Stories](#)

IDC thanks DOE for its insights, guidance and funding of this research project.

This study is based upon work funded by the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research, and the National Nuclear Security Administration, under award number DE-SC0008540.

In Summary

Conclusions

- **HPC is still expected to be a growth market**
 - Growing recognition of HPC's strategic value
 - HPDA, including ML/DL, cognitive and AI
 - HPC in the Cloud will lift the sector writ large
- **Vendor share positions shifted greatly in 2015 & 2016 and continue to shift**
 - E.g., HPE acquisition of SGI
- **The HPDA market will expand opportunities for vendors**

But There are Still Major Customer Pain Points

Software is the #1 roadblock

- Better management software is needed
- Parallel software is lacking for most users → Many applications will need a major redesign

Storage access time is becoming the #2 pain point

- Areal density has improved much faster than access density

Clusters are still hard to use and manage

- System management & growing cluster complexity
- Power, cooling and floor space are major issues
- Storage and data management are becoming new bottle necks
- Lack of support for heterogeneous environment and accelerators

... Some good news in that there are new technologies in Big data, accelerators, clouds, etc.

QUESTIONS?



ejoseph@hyperionres.com

sconway@hyperionres.com

bsorensen@hyperionres.com

kmonroe@hyperionres.com

mthorp@hyperionres.com