BDEC2
BIG DATA AND EXTREME-SCALE COMPUTING 2

BDEC2 BoF
SC18, Dallas
November 13, 2018

BIG DATA AND EXTREME-SCALE COMPUTING

International Exascale Software Project Meetings

Overall goal:

- **Bring together the international community** to explore plans and identify gaps for **producing a software infrastructure** capable of supporting exascale applications

Meeting history:

1. Santa Fe, NM, US, April 2009
2. Paris, France, June 2009
3. Tsukuba, Japan, October 2009
4. Oxford, UK, April 2010
5. Maui, HI, US, October 2010
6. San Francisco, US, April 2011
7. Cologne, Germany, October 2011
8. Kobe, Japan, April 2012

SC08 (Austin), SC09 (Portland), ISC09, SC10 (NOLA), ISC10, SC11 (Seattle)

www.exascale.org: White Papers and Slides

IESP Roadmap (2009 – 2012)

The IESP Roadmap presented a multidimensional analysis of the major challenges to be overcome in order to create a software infrastructure capable of supporting exaflop performance on next generation systems, and made a cogent case for the urgency of starting that work as soon as possible.

Spurred in some degree by the work of the IESP and its Roadmap, the United States, the European Union, and Japan have, in the past three years, moved aggressively to develop their own plans for achieving exascale computing in the next decade

IESP Roadmap Components

THE INTERNATIONAL JOURNAL
HIGH PERFORMANC
COMPUTING
APPLICATIONS

The International Exascale Software Project roadmap

The International Journal of High Performance Computing Applicatio
25(1) 3–60

Reprints and permission:
sagepub.co.uk/journalsPermissions.n
DOI: 10.1177/1094342010391989
hpc.sagepub.com
SAGE

Jack Dongarra, Pete Beckman, Terry Moore, Patrick Aerts, Giovanni Aloisio, Jean-Claude Andre, David Barkai, Jean-Yves Berthou, Taisuke Boku, Bertrand Braunschweig, Franck Cappello, Barbara Chapman, Xuebin Chi, Alok Choudhary, Sudip Dosanjh, Thom Dunning, Sandro Fiore, Al Geist, Bill Gropp, Robert Harrison, Mark Hereld, Michael Heroux, Adolfy Hoisie, Koh Hotta, Zhong Jin, Yutaka Ishikawa, Fred Johnson, Sanjay Kale, Richard Kenway, David Keyes, Bill Kramer, Jesus Labarta, Alain Lichnewsky Thomas Lippert, Bob Lucas, Barney Maccabe, Satoshi Matsuoka, Paul Messina, Peter Michielse, Bernd Mohr, Matthias S. Mueller, Wolfgang E. Nagel, Hiroshi Nakashima Michael E Papka, Dan Reed, Mitsuhisa Sato, Ed Seidel, John Shalf, David Skinner, Marc Snir, Thomas Sterling, Rick Stevens, Fred Streitz, Bob Sugar, Shinji Sumimoto, William Tang, John Taylor, Rajeev Thakur, Anne Trefethen, Mateo Valero, Aad van der Steen, Jeffrey Vetter, Peg Williams, Robert Wisniewski and Kathy Yelick

Abstract

Over the last 20 years, the open-source community has provided more and more software on which the world's hig performance computing systems depend for performance and productivity. The community has invested millions dollars and years of effort to build key components. However, although the investments in these separate software el ments have been tremendously valuable, a great deal of productivity has also been lost because of the lack of plannin coordination, and key integration of technologies necessary to make them work together smoothly and efficiently, bo within individual petascale systems and between different systems. It seems clear that this completely uncoordinate development model will not provide the software needed to support the unprecedented parallelism required for pet exascale computation on millions of cores, or the flexibility required to exploit new hardware models and features, su as transactional memory, speculative execution, and graphics processing units. This report describes the work of th community to prepare for the challenges of exascale computing, ultimately combing their efforts in a coordinated Inte

IESP ➡ BDEC (2013 – today)

BDEC derived much of its impetus from the earlier work

- International Exascale Software Project (IESP)
- European Exascale Software Initiative (EESI)
- European Exascale Software Initiative 2 (EESI2)
- European eXtreme Data and Computing Initiative (EXDCI)

Europe-USA-Asia International series of Workshops on Big Data and Extreme Computing

http://www.exascale.org/bdec/

Overarching goal:

1. **Create an international collaborative process** focused on the co-design of software infrastructure for extreme scale science, **addressing the challenges of both extreme scale computing and big data**, and supporting a broad spectrum of major research domains,
2. Describe **funding structures and strategies** of public bodies with Exascale R&D goals worldwide
3. **Establishing and maintaining a global network** of expertise and funding bodies in the area of Exascale computing

1 – BDEC Workshop, Charleston, SC, USA, April 29-May 1, 2013

2 – BDEC Workshop, Fukuoka, Japan, February 26-28, 2014

3 – BDEC Workshop, Barcelona, Spain, January 28-30, 2015

4 – BDEC Workshop, Frankfurt, Germany, June 15-17, 2016

5 – BDEC Workshop, Wuxi, China, March 9-10, 2017

With BoF's at SC and ISC meetings

1. **BDEC2 (Next Generation) Planning, Chicago, IL, March 26-28, 2018**

BDEC Report to be Published in IJHPCA

Journal Title
XX(X):1–40
©The Author(s) 2018
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

BIG DATA AND EXTREME-SCALE COMPUTING: PATHWAYS TO CONVERGENCE. Toward a Shaping Strategy for a Future Software and Data Ecosystem for Scientific Inquiry*

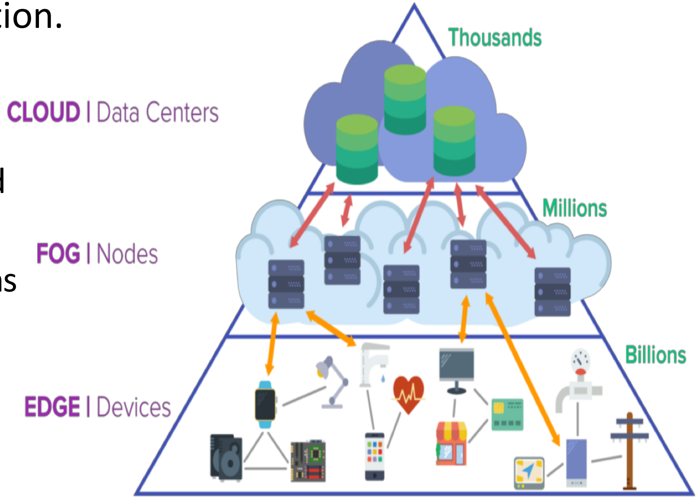
M. Asch, T. Moore, M. Asch, R. Badia, M. Beck, P. Beckman, T. Bidot, F. Bodin, F. Cappello, A. Choudhary, B. de Supinski, E. Deelman, J. Dongarra, A. Dubey, G. Fox, H. Fu, S. Girona, W. Gropp, M. Heroux, Y. Ishikawa, K. Keahey, D. Keyes, W. Kramer, J.-F. Lavignon, Y. Lu, S. Matsuoka, B. Mohr, D. Reed, S. Requena, J. Saltz, T. Schulthess, R. Stevens, M. Swany, A. Szalay, W. Tang, G. Varoquaux, J.-P. Vilotte, R. Wisniewski, Z. Xu and I. Zacharov

Shorten version of the January report:

http://www.exascale.org/bdec/sites/www.exascale.org.bdec/files/whitepapers/bdec2017pathways.pdf

Scientific Computing is Changing

- In the past, we moved experimental data to the centralized servers, which provided bulk storage and computational resources for analysis and simulation.
- Three things have changed:
  - **CPU advances** have enabled edge/IoT devices to be small parallel computers with real operating systems and multithreaded programming models (CUDA, OpenMP, TensorFlow, etc.).
  - **Machine learning** and AI has helped create a new class of algorithms that can sift through massive amounts of experimental data, and then pushing to the data center only the relevant results
  - **Edge devices** and scientific instruments are rapidly expanding, creating a new class of "edge software defined instrument" that must connect to cyberinfrastructure.



- These three changes are forcing us to rethink the central services model of HPC and embrace a new model where network infrastructure computes-along-the-way.
- To realize that goal, we need a new conceptual model for programming this new end-to-end infrastructure.