SAND2015-0985C

Data Architectures and Scientific Computing

Bruce Hendrickson

Director of Computing Research

Sandia National Laboratories, Albuquerque, NM

University of New Mexico, Computer Science Dept.

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Why We Should Care About Data Machines

4. Politics

3. Mission

2. Supply Chain

1. Technology synergy

4. Politics: "*Alignment*" is All the Rage

- **SEAB Report of the Task Force on HPC (exascale report):**
  - *Computational problems & data centric problems are coming together*
  - *We recognize and recommend a "new" alignment between classical and data centric computing*
- **ASCAC report on Data-Intensive Science & Exascale Computing**
  - *Should give high priority to investments that can benefit both data-intensive science and exascale computing*
- **NSCI plan is rumored to have similar language**

- **By embracing this vision of synergy, we connect to inter-agency efforts and gain broader advocacy for our program and our work**

3. Mission: Data Analytics is Important to Us

- **Goal of ASC is to build tools to help inform critical decisions**
- **Key capabilities include anomaly detection, UQ, QMU, etc.**
  - These are complex data-analytic challenges involving computational ensembles and experimental data
- **We have additional complexities from need for in-transit & in-situ analysis, check-pointing, memory & storage hierarchies, etc.**
- **Data community doing more computing, e.g. deep learning**
  - Our workloads increasingly look like theirs
- **Need to leverage ideas and capabilities from larger data-science community wherever possible**

2. Supply Chain: We're Bottom Feeders

- **Data opportunities are driving computing industry**
  - Commodity hardware will be designed for them
  - Our HPC machines will use these commodity parts

- **We need to understand the trends and consequences**
  - Optimize our NRE investments
  - Optimize our procurements
  - Leverage software stack investments where possible

Trends in Our Workloads

- **Higher fidelity simulations increasingly use:**
  - Unstructured and adaptive meshes
  - Multi-physics and multi-scale simulations
- **We're using simulation in more sophisticated ways**
  - Design optimization
  - Uncertainty quantification
  - In situ analysis and computational steering
- **These trends stress our memories and networks**
  - Computing is free
  - Data movement dominates performance and power consumption

Different …, But Not so Much

- **Data community emphasizes out-of-core computation**
  - Yes, but High Performance Data Analytics is an important subfield, and often in-memory
  - We're both interested in deep memory hierarchies
  - We can benefit from their expertise as we think about burst buffers and other uses of non-volatile memory technologies

- **Data community more focused on productivity than performance**
  - Yes, but we are increasingly emphasizing sophisticated runtimes and higher level programming models
  - Both communities feel need to hide architectural complexity from application programmers
  - We can learn from each other

1. Technology Synergy: We're More Similar Than we Admit

- **Trends in workloads:**
  - For scientific computing:
    - Data management and analysis is of growing importance (and difficulty)
    - Increasingly unstructured and dynamic
    - Data movement dominates runtime and energy consumption
  - For data computing:
    - Very large scale so scalability & resilience matter
    - Growth in computationally intensive applications like deep learning
    - Data movement dominates runtime and energy consumption

Common Technology Needs

- Faster and more energy efficient memory
- Better memory and storage hierarchies
- Memory-centric programming models
- Enhanced resilience
- Low-latency, low-power networking
- Improved programming models for heterogeneous architectures
- Improved packaging and administration for very large systems

- Most of the things we care about are also important to them. And they've got deeper pockets to help solve these problems!

Conclusions

- **Machines for data science intersect with us in multiple ways**

4. Politics

  – Alignment with broader agendas might help us

3. Mission

  – Data analytics is a growing component of our workflows

2. Supply Chain

  – Our machines will be built from data-computing components

1. Technology synergy

  – Our applications have a lot in common with data analytics

- **By understanding data science challenges, we can leverage and sometimes influence large investments towards our needs**