Data Center & Large-Scale Systems
(updated)

Luis Ceze, Bill Feiereisen, Krishna Kant, Richard Murphy, Onur Mutlu, Anand Sivasubramanian, Christos Kozyrakis

ACAR Workshop – February 2010

Data Center Realities

- A supercomputer
  - 100s cores/chip, 10s nodes/rack, 10s racks/container, 100s containers/DC
  - Distributed memory & storage
  - Hierarchical network
- Programmed by many
- Used by billions
- Low cost

"Man on the Moon" Goals
(aka our 10-year deliverables)

- $1 and 1 watt / person for DC infrastructure
  - Assuming our whole life is on-line
  - Minimize capital and operational expenses
  - From HW architecture to automated DC mgmt
- One application program for all scales
  - From 1 node & 100 users to 1M servers / 1B users
  - What are the prog. models, system SW, and HW?

Research Directions for Computer Architects

Data center as a computer

1. Chip-level support for DCs
2. DC node architecture
3. DC memory/storage hierarchy
4. DC operating & runtime system

Crosscutting: energy efficient data center

Energy Efficient Data Center

- Minimize DC energy consumption
  - Energy efficiency & proportionality
  - >100x improvement in queries/watt
- Research questions
  - State aware scale down of nodes
    - Node vs component level power modes
    - Tradeoffs with availability and QoS
  - Energy efficient servers
    - Optimizing cores & memory for requests/Watt
    - Energy efficient software & algorithms
  - Environmental sustainability of DC HW

1. Chip-level Support for DC

- How do we optimize many-cores for DCs
- Research questions
  - Enabling fast messaging
  - Enabling remote memory access
  - Enabling memory scale-down
  - Enabling isolation & privacy
  - Enabling dynamic languages
  - Enabling end-to-end monitoring
  - Enabling managability
  - HW/SW interface, virtualization, scalability

2. DC Node Architecture

- Current HW is direct evolution of PC
  - Is this optimal for perf, energy, reliability, & cost?
- Research questions
  - Thin vs thick vs heterogeneous system?
    - Implications for language & mgmt system
    - Balancing throughput & QoS
  - DC unit: pizza box vs rack vs container?
    - Optimization opportunities at each granularity
    - Memory system organization
    - Integration of compute and networking
  - Role of SSDs, non-volatile memory, photonics?

4. DC Memory/Storage Hierarchies

- How do we provide fast, high BW access to terabytes?
- Research questions
  - Large-scale memory hierarchies?
    - Local & global HW architecture
    - Processor/memory balance
    - Role of SSDs, non-volatile memory, photonics?
  - Locality optimizations
    - Moving data to computer or compute to data
    - Interactions with availability and isolation issues
  - APIs and file-systems?
    - Consistency model and synchronization
    - Protection, security, …

5. DC OS & Runtime System

- What is the OS and runtime system for DC?
- Research questions
  - Scalability, availability, reliability
    - 1M nodes, 5 9s availability
  - Real-time (re)provisioning of resources
  - End-to-end monitoring, introspection, flow control
  - Graceful degradation and scale-down
  - Multi-tenancy, isolation, and security
  - Compliance

Methodology & Resources

- Methodology: back to the future
  - Full-system research + prototypes
  - Multi-disciplinary projects
    - Architecture, systems, languages, compilers, …
- Resources: what's missing
  - Experimental data center for systems research
    - Learn from other areas (e.g., GENI & Openflow)
  - Workloads & large-scale analysis methods
    - Collaboration with industry is critical

Funding Programs

- Horizontal programs are still useful
  - Need to educate panels for the needs of the area…
- Need larger scale efforts as well
  - NSF ERC, STC, MARCO center (Musyc), …
- Grand challenges program
  - Framed around one important application at the time
  - Help address its needs and make technology advances for DC
    - Provides focus, motivates full-system work, relevance of results
    - Avoids the burden of solving all problems
    - Take app-specific results and transfer to general purpose DC
  - Example: DC technology for neuro-engineering systems
    - Data capture & analysis of neural activity to control prosthetics
    - Challenge: a DC that fits in the closet of a biology lab

Roadmap

- Urgent: 2 year infrastructure program
- Two parallel thrusts
  - Using existing chips (system-level architecture & SW)
    - 5-year scope, >10x energy efficiency improvement
  - Using novel chips
    - 10-year scope, >100x energy efficiency improvement
- Open questions
  - NSF expedition?
  - International collaboration?