C: Impact of changing architectures on data-processing (in-situ + post): Salon 12

(Alok Choudhary)

Participants: Pls email me to include the name in the list

choudhar@eecs.northwestern.edu

Impact of changing architectures on data-processing: Salon 12

- Exascale architectures
- Changing commodity CPU architectures
  - Parallelism, Reliability
  - Heterogeneous cores
  - 100 - 1000 cores
- Constrained memory footprint
  - Memory/core reduced
- NVRAM
  - Extension of the memory hierarchy
  - Impact on reading/writing
  - Speed and capacity, iops
- I/O wall, limited I/O bandwidth
  - Bytes/flops
- Power considerations of communication, reduced communication algorithms
  - Data transport, storage power costs
- Resiliency/Fault-Tolerance
- GPU, accelerator technologies

Exascale

- Current state of the art
  - typically store to disk and post process. some work on in-situ
- Issues with current approach
  - will not work on anticipated exascale systems
- Identify needs and consider tasks, linkages, metrics for success, …
  - making viz/analytics "work" at exascale (and with apps)
  - optimizing viz/analytics at exascle
- Identify opportunities for co-design and cross-cutting issues
  - Co-Design: programming models, runtime, provisioning, I/O and storage, structure/configuration, provisioning and use of the memory hierarchy
  - Cross Cutting Issues: power, resilience, fault-tolerance
- What specific interface issues and interdependencies will be addressed?
  - Examples: Fault tolerance, power, performance, I/O
  - see above – co-design aspects
- Identify opportunities for partnerships
  - scientists, industry/vendors, PRAC awardees, system-level researchers
- Highlights/takeaways
  - power, memory and IO constraints, will require fundamental research and rethinking at all levels – algorithms, programming models, runtime, etc…
  - traditional I/O post-processing models will not scale – new model and mechanisms are necessary
  - understand cross-cutting issues and unique requirements for data analytics and visualization is critical for effective co-design

Impact on Visualization/Analytics

- Issues specific to Analytics and visualization
  - viz/analysis is more data intensive compared to computation
  - compute vs data movement
  - data movement minimizing algorithms
  - standard for (in-situ) analysis/visualization
  - dynamic partitioning of (processor/memory hierarchy) resources between compute and in-situ viz/analytics
  - maximize memory reuse between compute and viz/analysis
  - visualization is occasionally interactive
  - multiscale, multiresolution analytics
  - Data models and formats

Impact on Visualization/Analytics

- Algorithms/Formulations
  - resilient, asynchronous, de-coupled
  - Impact of ensemble application formulations

- Programming models (probably crosscutting)
  - support in-situ and concurrent data processing – decoupled, asynchronous
  - take advantage of the number of cores
  - expose parallelism hierarchically, incremental
    - redefine data structures
  - hybrid programming models
  - manage hardware/system level controls

- Effects on the use of NVRAM
  - data re-organization, data reduction, time series analysis, indexing
  - development of out-of-core algorithms (NVRAM)
  - exploitation of read/write asymmetry
  - co-scheduling of analytics
  - persistence of data in NVRAM
  - supporting “in-situ” workflows - use NVRAM, programming models, provisioning, runtime…?

Impact on Visualization/Analytics

- I/O Wall – limited I/O bandwidth
  - pressure to move towards data volume reduction in-situ before IO
  - development of algorithms for data volume reduction
  - asynchronous I/O co-scheduling

- Fault tolerance, resilience
  - analytics/viz. support application level fault tolerance
  - analytics/visualization methods should be fault-tolerance

- Exascale analytics/post-processing
  - reading/querying (subsets) abstractions
  - provenance
  - impact on write mechanisms/formats, data models/formats
  - indexing and storage formats for effective querying

- Power
  - address power trade-off in data-movement and data storage
  - what controls are anticipated to be available at the application/systems levels?
  - what explicit controls will be available?

Success Metrics

- Analytics Benchmarks for
  - Power, performance and scalability evaluations
- Usability
- Develop quantifiable measures (requires more work!)