APEX

Alliance for application Performance at EXtreme scale

Initial Vendor Roadmap Presentation Guidance

October 20th 2014

High-level Design Philosophy

- Delivered application performance (as APEX suggests) is the primary driver in support of mission requirements
  - Peak FLOPS requirement will not appear in RFP
- APEX plans to purchase 2 platforms
  - ATS-3 and NERSC-9
- Both target delivery in FY20
- Advanced technology development is assumed to be necessary to meet mission needs
  - Accelerate development of yet to be identified key technologies
  - 3rd round of NRE – (Trinity/NERSC-8, CORAL, APEX)
- Considered pre-exascale platforms
  - MUST support path to exascale programming models
    - While supporting existing mission needs
  - Support MPI+OpenMP (threads)
    - Matured on Trinity/NERSC-8 and CORAL platforms
  - Additional support for other, yet to be identified, MPI+X programming models

Capability Improvement

- An increase in predictive capability requires increases in the fidelity of both geometric and physics models
    - This implies <u>usable</u> large platform memory capacity
- APEX must demonstrate a significant capability improvement
    - Improvement measured relative to Trinity (ATS1) and Cori (NERSC-8)
    - Improvement as a function of performance (total time to solution), increased geometries, increased physics capabilities, power/energy efficiency, resilience and other factors
- Previous DOE investments assumed to be an integral part of production computing for APEX.
    - Trinity/NERSC-8 NRE projects: Burst Buffer and Advanced Power management
    - Fast Forward and Design Forward

Facility, Power & Cooling

- ATS-3 will be located in the Nicholas C. Metropolis center (SCC) at Los Alamos National Lab
- NERSC-9 will be located in the CRT facility
- Estimated facility power and footprint
  - ATS3
    - 15MW
    - 8000 square feet
  - NERSC-9
    - Power and floor space likely not primary platform constraints
- Liquid cooled
  - Is our assumption correct?
  - Warm water or chilled ? Direct or indirect?

Guiding Questions

guiding not exhaustive

- Basically we want to understand your roadmap(s) in the timeframe we anticipate taking delivery (FY2020)
- Your roadmap presentations should NOT be limited to these guiding questions
- Tell us where and why our assumptions are wrong!
- We assume multi-level memory (storage) hierarchy
  - What will this look like?
  - Will it extend beyond the node?
  - Bandwidth and latency characteristics (between levels)?
  - Technologies?
  - Capacity?
  - Relative cost and energy trade-offs?
- What does a processor(s) look like on a node?
  - How many cores?
  - Heterogeneous or Homogeneous?
  - Core characteristics
  - NUMA characteristics?
  - Coherency?

Guiding Questions (cont)

- NIC
  - Integrated or discrete?
  - Injection bandwidth?
  - Message injection rate?
    - At what message size(s)?
  - Offload characteristics?
  - Access to memory?
- Interconnect
  - Topology?
  - Physical layer?
  - Bisection bandwidth?

Guiding Questions (cont)

- Software
  - Languages
  - Programming Environments
  - Programming Models
  - Profilers and Debuggers
  - Operating system(s)
  - Advanced Power Measurement and Control
  - RAS and/or System Management
  - Software to aid resiliency
  - Workload (and workflow) management

Guiding Questions (cont)

- What will the filesystem look like?
    - Integrated into memory hierarchy?
    - Is traditional application driven check point restart still required?
    - How can we optimize for analysis usage models?
- Support for task based programming model(s)?
- What are the advanced resilience mechanisms?
    - Hardware and/or software
- What is the optimal way to support emerging data intensive computing workloads on the same platform as 'traditional' HPC ones?
- Will you have early test platforms / proxies available that we can explore these issues with?
- What are your proposed NRE areas?
    - and required lead times for each
- How can APEX best influence your roadmap?

Schedule

- Week of January 19th targeted for first round of presentations

- Possible follow-up/deep dive presentations TBD

- Contact Jim Laros (jhlaros@sandia.gov) to schedule time-slot

- APEX team can be available by request at SC14 for questions and clarifications. (Jim Laros-jhlaros@sandia.gov, Jim Lujan jewel@lanl.gov, Nick Wright njwright@lbl.gov)