The Future is for the Weird

Lee Ward

9 April, 2014
HPC for closely coupled simulations will continue along the current vector

Hybrids (I hope I’m wrong about this one)

More cores per node; A lot more cores?
  - Will need separate coherency domains?
    * Not really a problem, we are already used to partitioning across nodes

More volatile memory per node

Non-Volatile memory on node
Non-Volatile Memory on Node

• Hmm… Interesting
• Why, this will make a nice low-latency, high-bandwidth storage device for my favorite node OS!
  – I can use it as a backing store, large node-local scratch, etc
• Every node gets a private “burst buffer”!
• Nope, that’s nuts
• It’s *memory*, use it like memory
  – It sits on the memory bus as a first-class citizen
    • Might be fronted by a volatile memory with, at least, hardware assisted synchronization to the non-volatile part
• Except it’s persistent
  – Want snapshots, not cycle-granular perseverance
  – Want access to snapshots so long as the NIC is powered
Where does the Storage Live?

• Storage? We don’t need no stinkin’ storage.

• NVM on the bus provides a load-store interface
  – There’s no bloody open, close, read, *or* write
  – It’s just memory, in the state that the named snapshot saved
    • Teensy issue, need to reconnect everything to continue
    • OK, and we need a path to the NVM part on “dead” nodes

• These nodes are modified Harvard Architecture
  – Who said we had to limit ourselves to two address spaces?
  – Shared read/write thingies live in an internode NUMA region
    instead of, yucky, “files” and it becomes a linker problem

• How to import/export data?

• I dunno, Lustre maybe?
  – A seriously less intense Lustre deployment though 😊