SIRIUS Kickoff meeting

10/19/2015

Present: Scott Klasky, Hasan Abbasi, Dave Pugmire, Mark Ainsworth, Qing Liu, Feiyi Wang, Carlos Maltzahn., James Kress, Jay Lofstead, George Ostrouchov, Manish Parashar, Matt Curry, Jeremy Logan

Scott

* Exascale computing initiative
  + ~90 million / year
  + managed by ANL (Paul Messina)
  + Czar at each lab
  + details are unknown
* SCIDAC world is about to implode – all CS will be in one SCIDAC
* CoDesign
* Funding to begin in December (Universities too)
* Project mgmt.
  + At least one call per month
  + Executive call every other week
* Demos
  + Duct tape and band-aids
* Web page
  + Logo?
* Git (will use the same repository that the proposal was in)
* Deliverables
  + Will be a lead for each deliverable, but others may participate
* Memory arrangement example (PIC code)
* Goals
  + New APIs
  + New layer for “Big Data”
  + Prototypes focused on important users
  + Fundamentally change the nature of SSIO for HPC
* Research questions
  + Abstractions for “data utility”
  + Faster time to solution
  + Trade accuracy for performance
* Demo

Hasan

* Interesting topics
  + Data description and application interface
  + Data refactoring
  + Migration
  + Resource Management
* Utility
* Chunking

Mark

* Data vs. Information
  + Shannon information content
  + More random the data, the higher the entropy
  + Limits the amount of compression that can be done.
* Scientific data can’t be compressed very much (losslessly)
* Tradeoff for mapping to storage hierarchy
* Example of dealing with CS Chang’s data
  + Using an interpolation auditor

Matt

* Sirocco
* Static organization of data on the file system causes problems
* POSIX is bad
* Two tier mechanism
  + Lower level object store
  + Upper level consists of a set of object-store clients that provide various services (such as POSIX, HDF5, S3, etc.)