

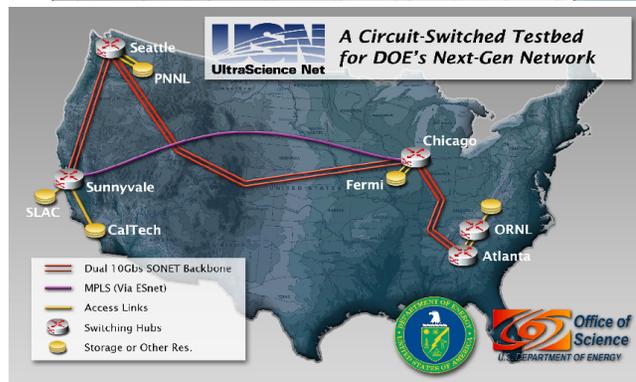
# Oak Ridge National Laboratory

## Computing and Computational Sciences

### HEC FSIO 2010 Workshop

Futures Panel

Presented by  
Steve Poole



August 04, 2010



Managed by UT-Battelle for the  
U. S. Department of Energy

Stephen Poole – 2010 - ORNL



# Posed Questions

- **1) What do you \*think\* parallel FS (hw and sw) will look like in 10 years?**
- **2) What do you \*wish\* parallel FS (hw and sw) will look like in 10 years?**
- **3) What do you think will be the top 3 challenges in the next 10 years to provide IO and associated infrastructure into Exa and enormous data analytics?**
- **4) Where will HPC be able to find good external tools to leverage (e.g. Google, LSST, etc)?**
- **5) How does the hype about clouds and virtualization add challenges or opportunities to file system and storage SLAs both at future exa as well as at current more moderate scale?**
- **6) Is it time to toss the file systems concept entirely and go to some other kind of abstraction?**

# Caveat (Read the others first)

- **Garth**
- **Peter**
- **Roger**
- **Rob**

# Answers to Posed Questions

- **What do you \*think\* parallel FS (hw and sw) will look like in 10 years?**
  - **Assumption #1 NO serious Government funding**
    - **Current implementations / evolutions + research activities**
      - Will this answer the needs in the PB/EB range ? (analytics/normal)
    - **PFS concepts from folks like Google/Yahoo/Micro\$oft**
      - Proprietary, we follow and not lead ?
    - **Has government funding given us the \*best\* or even our monies worth ?**
  - **Assumption #2 sufficient Government funding**
    - **We define and ENFORCE Open Source/Open Development project(s)**
    - **Government shares all IP rights**
    - **Government enables an environment to test at scale**
    - **We have funded several companies, do they REALLY answer our needs ?**
    - **One winner after (5) years next (5) to commercialize**
    - **Assume “morphable” from the onset.**

# Answers to Posed Questions

- **What do you \*wish\* parallel FS (hw and sw) will look like in 10 years?**
  - **Not a Parallel DISK File System (fewer assumptions of underlying technology OR HUGE assumptions ? (come on 512B blocks ?)**
  - **It is another level in the memory hierarchy (VMM is a KILLER)**
    - **What semantics (PUT/GET ?)(Not everything is local/nor remote)**
    - **VERY topo aware, how hard is self describing FS/elements ?**
    - **Extended POSIX (we have made “trivial” progress)(MD Tools)**
    - **What would we do with meta-data ?**
    - **Could make TLB’s fun, assume NO help from CPU/system vendors**
    - **Potential HUGE OS impact**
  - **Any Miracles required ? ;-) Any Research ?**

# Answers to Posed Questions

- **What do you think will be the top 3 challenges in the next 10 years to provide IO and associated infrastructure into Exa and enormous data analytics?**
  - Funding, Funding... It needs to be properly funded (10 years)
  - Wide govt/commercial acceptance (HPC types)
  - Test facilities / systems
    - Vendor involvement...
- **Where will HPC be able to find good external tools to leverage (e.g. Google, LSST, etc)?**
  - **\*IF\*** we do not develop, we will have to rely on **\*what\*** they decide to release. How fare behind are we ?
  - Funded tool developers (Universities, Labs...)
  - Involved commercial folks ?

# Answers to Posed Questions

- **How does the hype about clouds and virtualization add challenges or opportunities to file system and storage SLAs both at future exa as well as at current more moderate scale?**
  - **Current, probably too late(S==\$\$, easy equation) (Java is Great ???)**
    - If you like the Kool-Aide, you will figure it out (Does Cloud/Grid == HDFS ?)
    - For some scale of problems they will do well (Cloud-I (Grid) is still around)
    - How is the security model doing ? (current and future)(Lights-out ?)
  - **Future**
    - Too few @ Scale environments, apps to tell (how many have been enabled ?)
    - What is the conversion cost (how long will it last ?)
    - What is the replication (HDFS) cost and how will it impact LARGE systems ?
    - How will it handle QoS, SDC, potential latency issue, RAM impact for HDFS
      - 600 bytes (1 file object + 2 block objects) to store an average file in RAM of the NameNode
      - To store 100 million files, a NameNode requires at least 60 GB of RAM.
      - With 100 million files, each having an average of 1.5 blocks, there will be 200 million blocks in the file system

# Answers to Posed Questions

- **Is it time to toss the file systems concept entirely and go to some other kind of abstraction?**
  - **Yes, especially DISK file systems**
  - **Have seen “trivial” examples of holographic storage. NOT ready for prime time yet.**
  - **Moving “programmable functions” closer to storage.**
    - **Think collectives/AMO’s, analytics**
  - **I/O and Networking as “first class” citizens**
  - **Remember, not everything is CkPtRst / Write is NOT the only process**
    - **Seismic, Bio, Real time saves...**
  - **Assume “currently” co-design is a myth**
  - **3D**

# Acknowledgements



**This work was supported by the United States Department of Defense & used resources of the Extreme Scale Systems Center at Oak Ridge National Laboratory.**