Big Data Systems (BDS) at the ARM Archive II: Data Processing Opportunities

ORNL*, ARM Archive: Raymond McCord, Giri Palanisamy, Jay Manneschmidt, Pete Eby  PNNL: Matt MacDuff, James Mather

What? A system for large-scale ARM data processing. A new “dimension” for the ARM Climate Research User Facility
Why? Expedite creation and exploration of data products from new, very large data streams (radars and lidars)
Where? At ARM Archive; Adjacent to Archive Mass Storage
Who? Users include ARM infrastructure staff and Archive users (scientists)
How? Access with proposal submission (via Campaign proposal tool or Engineering Change Request)
When? Alpha testing now; starting Summer 2011

Why?
- Reduce “wall clock time” for large data processing
  - Near term:
    - Faster turnaround for Big Data VAP development
  - Long term:
    - Stabilize timely data production
    - Support reproprocessing and post processing
- Stepping stone to super-computers (if needed)?
- Engage more users with “new data” via complex visualization capabilities
  - Expedite identification of properties needing retrievals

Who?
- ARM Infrastructure
- Archive User Community
- By proposal
  - Batch data processing ➔ Field Campaign Proposal
  - Interactive visualization ➔ Identity confirmation by Archive staff
  - Software development ➔ Infrastructure procedures (ECR) or infrastructure sponsorship

What? - design & policy
- Looks like a “giant server” or small cluster
- Optimized for data I/O (Also substantial memory)
- Significant CPU within a single system
- Limited access (avoid task thrashing!)
  - 1 or 2 concurrent users (“get in & get out”)
  - Requires ORNL user account with ARM Archive sponsorship
  - Production-ready software
  - Data acquisition and storage by ARM Archive staff

What? - technology
- Hardware
  - “High-end” server with 4 CPU's, 6 cores each
  - 256 GB of memory
  - ~44 Tb of disks distributed across 6 RAID's
  - 10 GbE network interfaces / Tape library for local backup
- Software
  - RedHat Linux OS (and virtual machines)
  - Normal suite of compilers
  - IDL, MatLab, etc.
  - ARM - Integrated Software Development Environment (ISDE)

For more information or questions please contact: ARM Archive User Services (armarchive@ornl.gov)