



Technical Details of Setting up a RAC and associated IAC's

SAR Workshop
April 18-19, 2003
Arlington, Texas

Lee Lueking
Fermilab Computing Division, CEPA Dept.
DØ Liaison to PPDG
Batavia, Illinois

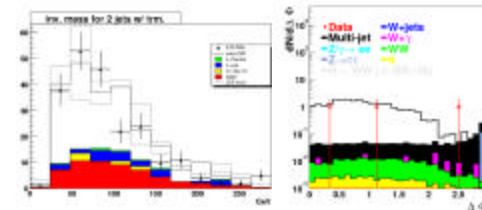
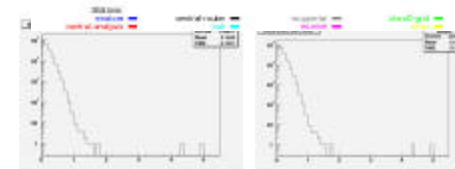
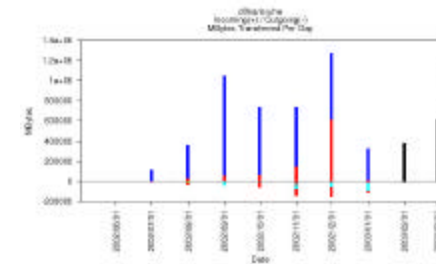


RAC Prototype: GridKa



- Overview: Aachen, Bonn, Freiburg, Mainz, Munich, Wuppertal
 - ◆ Location: Forschungszentrum Karlsruhe (FZK)
 - ◆ Regional Grid development, data and computing center. Established: 2002
 - ◆ Serves 8 HEP experiments: Alice, Atlas, BaBar, CDF, CMS, Compass, DØ, and LHCb
- Political Structure: Peter Mattig (wuppertal) FNAL rep. to Overv Board, C. Zeitnitz (Mainz), D. Wicke (Wuppertal) Tech. Adv. E reps.
- Status: Auto caching Thumbnails since August
 - ◆ Certified w/ physics samples
 - ◆ Physics results for Winter conferences
 - ◆ Some MC production done there
 - ◆ Very effectively used by DØ in Jan and Feb.

- **Resource Overview: (summarized on next page)**
 - Compute: 95 x dual PIII 1.2GHz, 68 x dual Xeon 2.2 GHz. D0 requested 6%. (updates in April)
 - Storage: DØ has 5.2 TB cache. Use of % of ~100TB MSS. (updates in April)
 - Network: 100Mb connection available to users.
 - Configuration: SAM w/ shared disk cache, private network, firewall restrictions, OpenPBS, Redhat 7.2, k 2.418, D0 software installed.

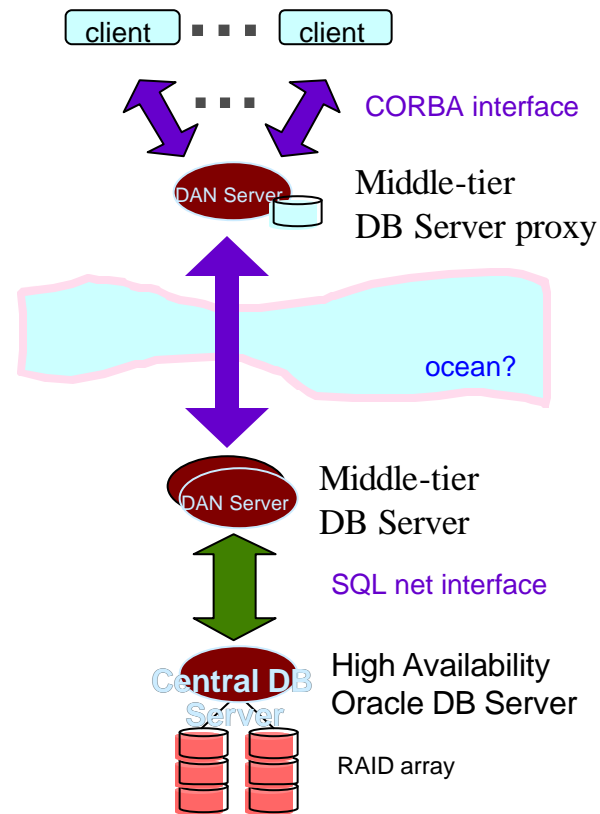
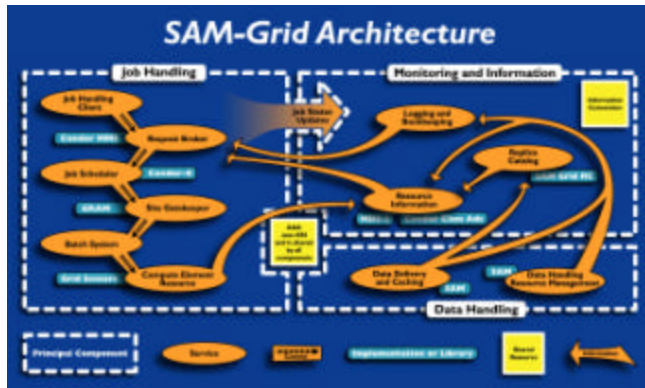




Required Server Infrastructure



- SAM-Grid (SAM + JIM) Gateway
- Oracle database access servers (DAN)
- Accommodate realities like:
 - ◆ Policies and culture for each center
 - ◆ Sharing with other organizations
 - ◆ Firewalls, private networks, et cetera



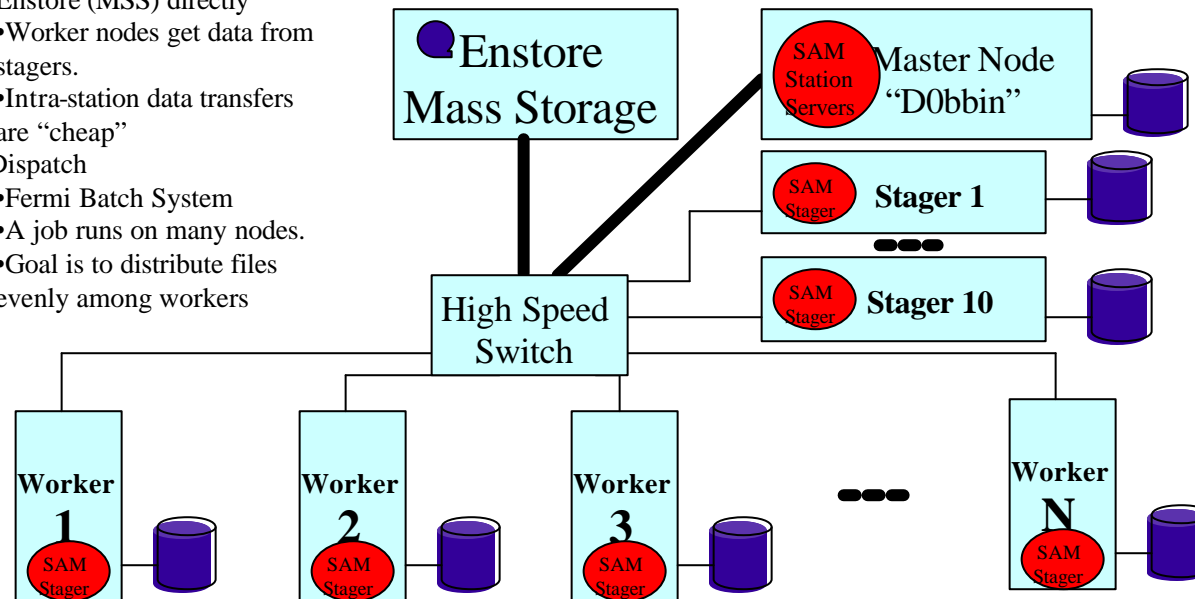
<http://d0db.fnal.gov/sam>



SAM Station: Dzero Distributed Cache Reconstruction Farm



- Network
 - Each Stager Node accesses Enstore (MSS) directly
 - Worker nodes get data from stagers.
 - Intra-station data transfers are “cheap”
- Job Dispatch
 - Fermi Batch System
 - A job runs on many nodes.
 - Goal is to distribute files evenly among workers

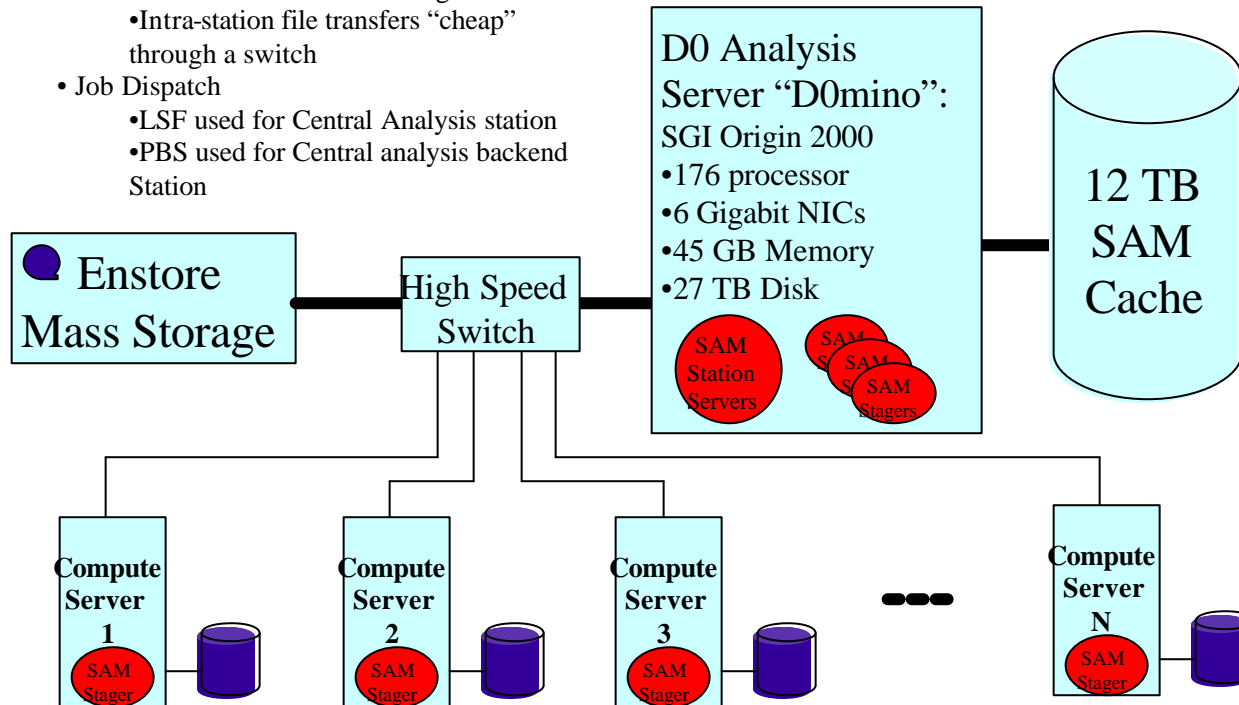




SAM Stations: Dzero Central Analysis and Central Analysis Backend



- Network
 - Access to Enstore is through D0mino
 - Intra-station file transfers “cheap” through a switch
- Job Dispatch
 - LSF used for Central Analysis station
 - PBS used for Central analysis backend Station

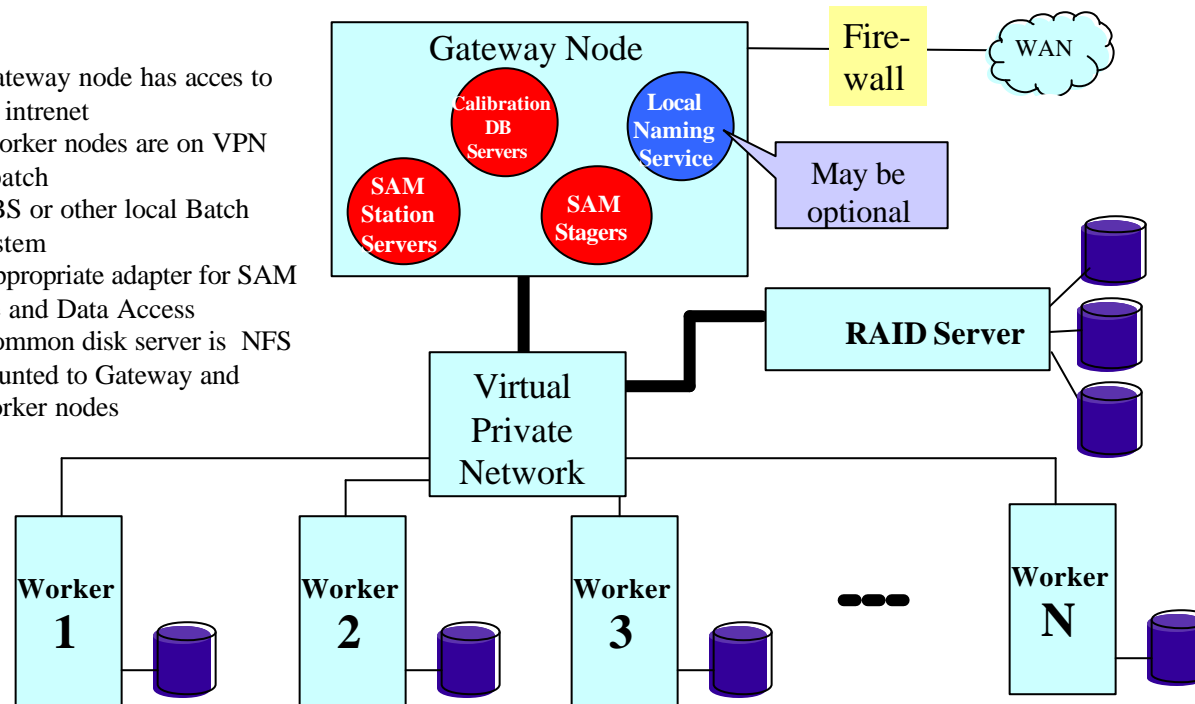




SAM Station: Shared Cache Configuration w/ PN (used at GridKa and U. Michigan NPACI)



- Network
 - Gateway node has access to the internet
 - Worker nodes are on VPN
- Job Dispatch
 - PBS or other local Batch System
 - Appropriate adapter for SAM
- Software and Data Access
 - Common disk server is NFS mounted to Gateway and Worker nodes





More Details



- SAM is distributed to clients via fnal ups/upd products distribution and versioning.
- Gateway runs sam servers, special setup, user sam account.
- Runs GridFTP demon for parallel transfers. Needs service certificates (KCA for FNAL transfers).
- Experience with the SAM shared cache configuration is:
 - ◆ It is great in environments where nodes are shared, but...
 - ◆ Can be NFS and RAID server bottlenecks, doesn't scale easily.
- Calibration DB servers are caching proxies connected through primary servers at FNAL to the central data base. Needed for RAW Reconstruction.
- Interface to tape storage system is still customized for each site (GridKa had to do this)
- Testing new “network file access” station feature at Lyon (CCIN2P3) which allows access to files not in local cache (eg. IN2P3 uses rfio).
- Making the D0 code distribution work caused some delays at GridKa



<http://d0db.fnal.gov/sam>



Specific Ports for Firewalls



- If the gateway node is behind a firewall, ports > 1024 need to be unrestricted to d0ora1.fnal.gov, d0ora3.fnal.gov, and d0mino.fnal.gov.
- This is what is used for the nameservice:
 - ◆ d0ora3> ups inquire sam_config -q ns_dev
 - ◆ SAM_NAMING_SERVICE=d0db-dev.fnal.gov:9000
 - ◆ <d0ora1> ups inquire sam_config -q ns_prd
 - ◆ SAM_NAMING_SERVICE=d0db.fnal.gov:9010
 - ◆ <d0ora1> ups inquire sam_config -q ns_int
 - ◆ SAM_NAMING_SERVICE=d0db-int.fnal.gov:9005
- gridftp 4567 to all station hosts.
- We are not enforcing optimizers to run on any particular port.
- Additional ports for JIM- Globus & Condor-G





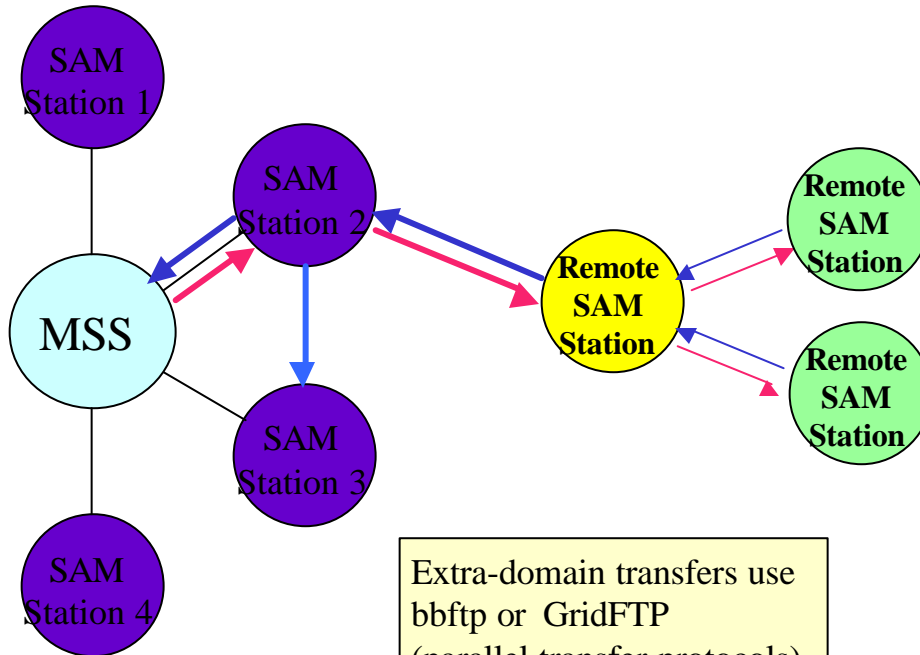
Data to and from Remote Sites



Data Forwarding and Routing

Station Configuration

- **Replica location**
 - Prefer
 - Avoid
- **Forwarding**
 - File stores can be forwarded through other stations
- **Routing**
 - Routes for file transfers are configurable



Extra-domain transfers use bbftp or GridFTP (parallel transfer protocols)

