

What do we want DØ Regional Analysis Centers (DØRAC) do?

DØRAC Face-to-Face Meeting

Apr. 25, 2002

Jae Yu

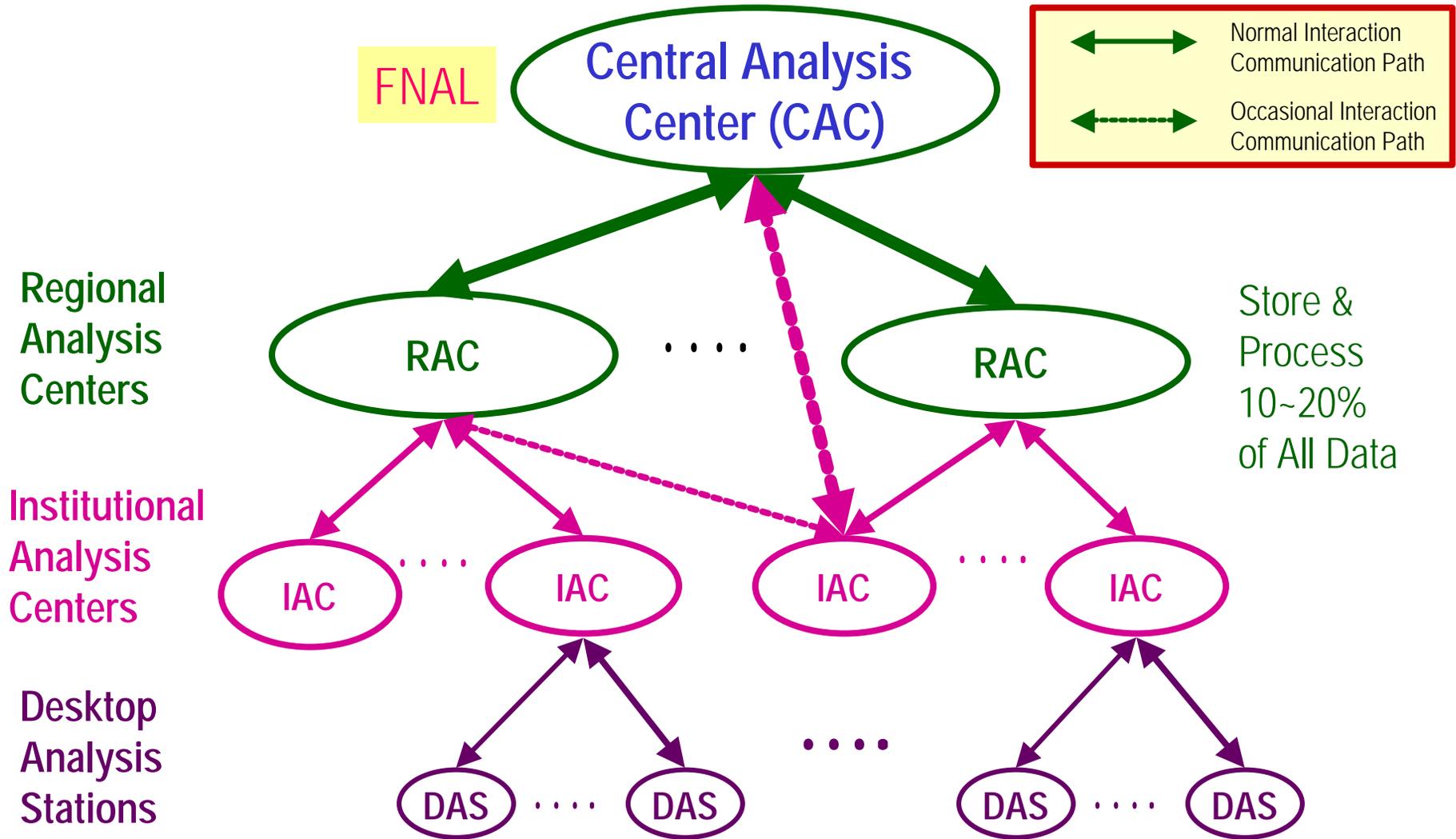
- Why do we need a DØRAC?
- What do we want a DØRAC do?
- What is involved in each of the functionality?

Why do we need a DØRAM?

- Total Run II data size reaches over multiple PB
 - 300TB and 2.8PB for RAW in Run IIa and IIb
 - 410TB and 3.8PB for RAW+DST+TMB
 - $1.0 \times 10^9 / 1.0 \times 10^9$ Events total
 - At the fully optimized 10sec/event (40 specInt95) reco. → 1.0×10^{10} Seconds for one time reprocessing for Run IIa
 - Takes 7.6Mo using 500 750MHz (40 specInt 95) machines at 100% CPU efficiency
 - 1.5 Mo with 500 4GHz machines for Run IIa
 - 7.5 to 9 Mos with 500 4GHz machines for Run II b
 - Time for data transfer occupying 100% of a gigabit (125Mbyte/s) network
 - $3.2 \times 10^6 / 3.2 \times 10^7$ seconds to transfer the entire data set (A full year with 100% OC3 bandwidth)

- Data should be readily available for expeditious analyses
 - Preferably disk resident so that time for caching is minimized
- Analysis processing compute power should be available without having the users relying on CAC
- MC generation should be transparently done
- Should exploit compute resources at remote sites
- Should exploit human resources at remote sites
- Minimize resource needs at the CAC
 - Different resources will be needed

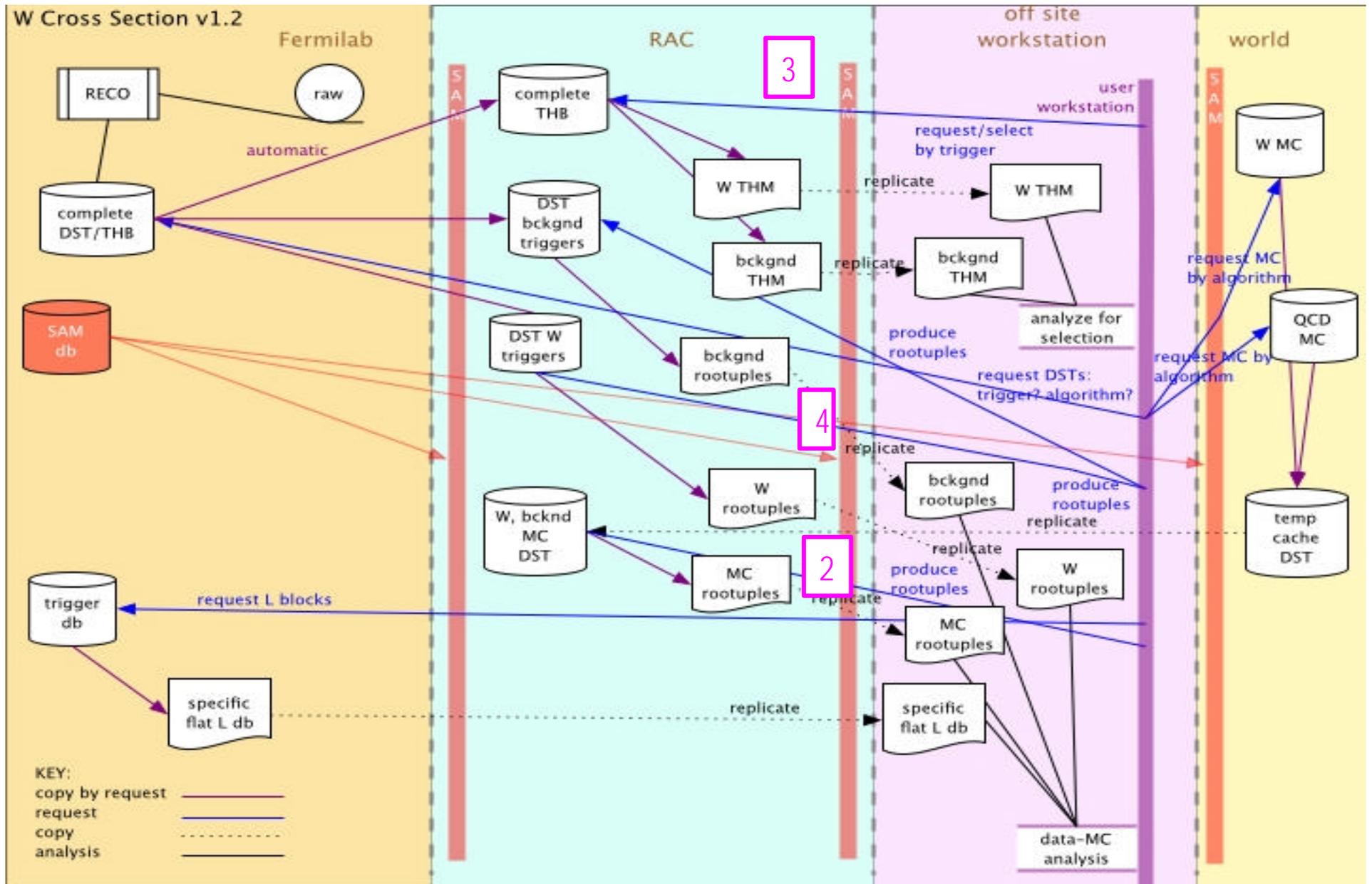
Proposed DØRAM Architecture



What is a DØRAC?

- An institute with large concentrated and available computing resources
 - Many 100s of CPUs
 - Many 10s of TBs of disk cache
 - Many 100Mbytes of network bandwidth
 - Possibly equipped with HPSS
- An institute willing to provide services to a few small institutes in the region
- An institute willing to provide increased infrastructure as the data from the experiment grows
- An institute willing to provide support personnel if necessary

Chip's W x-sec Measurement



What services do we want a DØRAC do?

1. Provide intermediary code distribution
2. Generate and reconstruct MC data set
3. Accept and execute analysis batch job requests
4. Store data and deliver them upon requests
5. Participate in re-reconstruction of data
6. Provide database access
7. Provide manpower support for the above activities

Code Distribution Service

- Current releases: 4GB total → will grow to >8GB?
- Why needed?:
 - Downloading 8GB once every week is not a big load on network bandwidth
 - Efficiency of release update rely on Network stability
 - Exploit remote human resources
- What is needed?
 - Release synchronization must be done at all RACs every time a new release become available
 - Potentially need large disk spaces to keep releases
 - UPS/UPD deployment at RACs
 - FNAL specific
 - Interaction with other systems?
 - Need administrative support for bookkeeping

Generate and Reconstruct MC data

- Currently done 100% at remote sites
- Why needed?
 - Extremely self-contained
 - Code distribution done via a tar-ball
 - Demand will grow
 - Exploit available compute resources
- What is needed?
 - A mechanism to automate request processing
 - A Grid that can
 - Accept job requests
 - Packages the job
 - Identify and locate the necessary resources
 - Assign the job to the located institution
 - Provide status to the users
 - Deliver or keep the results
- Perhaps most undisputable task but do we need a DØRAC?

Batch Job Processing

- Currently rely on FNAL resources
 - D0mino, ClueD0, CLUBS, etc
- Why needed?:
 - Bring the compute resources closer to the user
 - Distribute the computing load to available resources
 - Allow remote users to process their jobs expeditiously
 - Exploit the available compute resources
 - Minimize resource load at CAC
 - Exploit remote human resources

Batch Job Processing cont'd

- What is needed?
 - Sufficient computing infrastructure to process requests
 - Network
 - CPU
 - Cache storage
 - Access to relevant databases
 - A Grid that can:
 - Accept job requests
 - Packages the job
 - Identify and locate the necessary resources
 - Assign the job to the located institution
 - Provide status to the users
 - Deliver or keep the results
- This task definitely needs a DØRAC
 - What do we do with input? Keep them at RACs?

Data Caching and Delivery

- Currently only at FNAL
- Why needed?
 - Limited disk cache at FNAL
 - Tape access needed
 - Latencies involved, sometimes very long
 - Delivering data within a reasonable time over the network to all the requests is imprudent
 - Reduce resource load on the CAC
 - Data should be readily available to the users with minimal latency for delivery

Data Caching and Delivery cont'd

- What is needed?
 - Need to know what data and how much we want to store
 - 100% TMB
 - 10-20% DST?
 - Any RAW data at all?
 - What about MC? 50% of the actual data
 - Should be on disk to minimize data caching latency
 - How much disk space? (10TB for Run IIa TMB)
 - Constant shipment of data to all RACs from the CAC
 - Constant bandwidth occupation (14MB/sec for Run IIa RAW)
 - Resources from CAC needed
 - A Grid that can
 - Locate the data (SAM can do this already...)
 - Tell the requester about the extent of the request
 - Decide whether to move the data or pull the job over

Data Reprocessing Services

- These include:
 - Re-reconstruction of the data
 - From DST?
 - From RAW?
 - Re-streaming of data
 - Re-production of TMB data sets
 - Re-production of roottree
 - *ab initio* reconstruction
- Currently done only at CAC offline farm

Reprocessing Services cont'd

- Why needed?:
 - The CAC offline farm will be busy with fresh data reconstruction
 - Only 50% of the projected capacity is used for this but ...
 - Going to be harder to re-reconstruct as more data accumulates
 - We will have to
 - Reconstruct a few times (>2) to improve data
 - Re-stream TMB
 - Re-produce TMBs from DST and RAW
 - Re-produce root-tree
 - It will take a many months to re-reconstruct the large amount of data
 - 1.5 Mo with 500 4GHz machines for Run IIa
 - 7.5 to 9 Mos for full reprocessing Run IIb
 - Exploit large resources in remote institutions
 - Expedite re-processing for expeditious analyses
 - Cutting down the time by a factor of 2 to 3 will make difference
 - Reduce the load on CAC offline farm
 - Just in case the CAC offline farm is having trouble, the RACs can even help out with *ab initio* reconstruction

Reprocessing Services cont'd

- What is needed?
 - Permanently store necessary data, because it would take a long time just to transfer data
 - DSTs
 - RAW
 - Large date storage
 - Constant data transfer from CAC to RACs as we take and reconstruct data
 - Dedicated file server for data distribution to RACs
 - Constant bandwidth occupation
 - Sufficient buffer storage at CAC in case network goes down
 - Reliable and stable network
 - Access to relevant databases
 - Calibration
 - Luminosity
 - Geometry and Magnetic Field Map

Apr. 25, 2002

Why DØRAC?
DØRAC FTFM, Jae Yu

16

Database Access Service

- Currently done only at CAC
- Why needed?
 - For data analysis
 - For reconstruction of data
 - To exploit available resources
- What is needed?
 - Remote DB access software services
 - Some copy of DB at RACs
 - A substitute of Oracle DB at remote sites
 - A means of synchronizing DBs

Reprocessing Services cont'd

- Transfer of new TMB and Roottrees to other sites
- Well synchronized reconstruction code
- A grid that can
 - Identify resources on the net
 - Optimize resource allocation for most expeditious reproduction
 - Move data around if necessary
- A dedicated block of time for concentrated CPU usage if disaster strikes
- Questions
 - Do we keep copies of all data at the CAC?
 - Do we ship DSTs and TMBs back to CAC?
- This service is perhaps the most debatable one but I strongly believe this is one of the most valuable functionality of RAC.