

Proposal for a DØ Remote Analysis Model (DØRAM)

DØRACE Workshop

Feb. 12, 2002

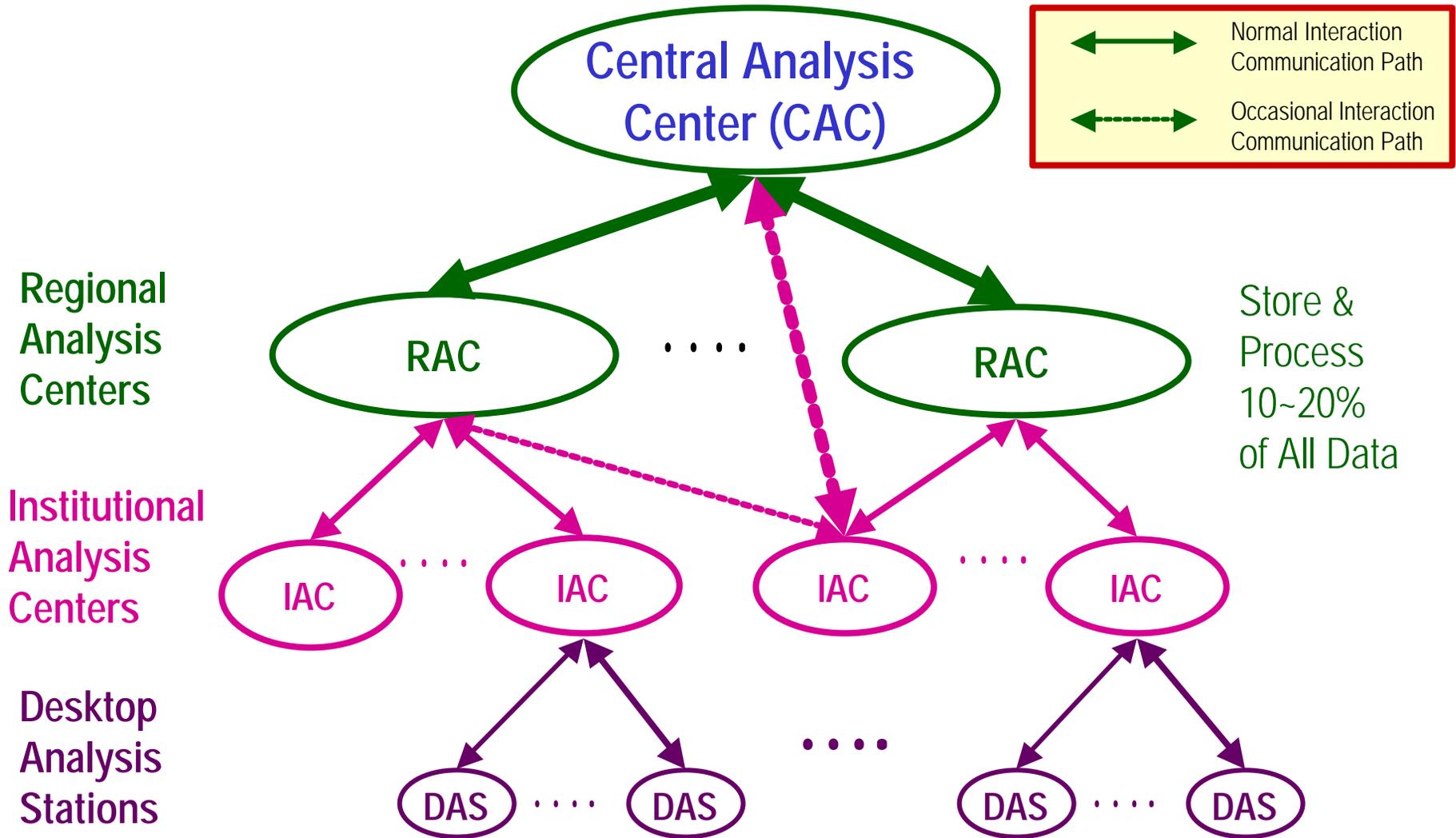
Jae Yu

- Introduction
- Remote Analysis Station Architecture
- Requirement for Regional Analysis Centers
- Suggested Storage Equipment Design
- What Do I Think We Should Do?
- Conclusions

Why do we need a DØRAM?

- Total Run IIa data sizes are
 - 350TB for RAW
 - 200-400 TB for Reco + root
 - 1.4×10^9 Events total
 - At the fully optimized 10sec/event reco. $\rightarrow 1.4 \times 10^{10}$ Seconds for one time reprocessing
 - Takes one full year w/ 500 machines
 - Takes ~8mos to transfer raw data for dedicated gigabit network
- Centralized system will do a lot of good but not sufficient (DØ analysis model proposal should be complemented with DØRAM)
- Need to allow remote locations to work on analysis efficiently
- Sociological benefits within the institutes
- Regional Analysis Centers should be established

Proposed DØRAM Architecture



Regional Analysis Centers

- A few geographically selected sites that satisfy requirements
- Provide almost the same level of service as FNAL to a few institutional analysis centers
- Analyses carried out within the regional center
 - Store 10~20% of statistically random data permanently
 - Most the analyses performed on these samples with the regional network
 - Refine the analyses using the smaller but unbiased data set
 - When the entire data set is needed → Underlying Grid architecture provide access to remaining data set

Regional Analysis Center Requirements

- Become a Mini-CAC
- Sufficient computing infrastructure
 - Large bandwidth (gagibit or better)
 - Sufficient Storage Space to hold 10~20% of data permanently and expandable to accommodate data increase
 - >30TB just for Run IIa RAW data
 - Sufficient CPU resources to provide regional or Institutional analysis requests and reprocessing
- Geographically located to avoid unnecessary network traffic overlap
- Software Distribution and Support
 - Mirror copy of CVS database for synchronized update between RAC's and CAC
 - Keep the relevant copies of data bases
 - Act as SAM service station

Suggestions and Comments

- Data characteristics
 - Specialized data set, in addition to service data set for reprocessing
 - Some level of replication should be allowed
 - Consistency of data must be ensured
 - Centralized organization of reprocessing
 - Book keeping of reprocessing
- Two staged approach:
 - Before Full gridification → All data kept in the CAC
 - After full gridification
 - Fully distributed within the network
 - Data sets are mutually exclusive

- In Europe, some institutions are already in the works to become an RAC
 - Karlsruhe (Germany)
 - NIKHEF (Netherlands)
 - IN2P3, Lyon (France)
- We want more US participation
- Agreed to form a group – committee if you wish, to formulate RAC more systematically → Write up a document
 - Functions
 - Services
 - Requirements
 - Etc.

Conclusions

- DØ must prepare for large data set era
- Need to expedite analyses in timely fashion
- Need to distribute data set throughout the collaboration
- Establishing regional analysis centers will be the first step toward DØ Grid
- Will write up a proposal for your perusal
- Will ask a few of you to serve as RAC committee members