

DØRACE Workshop

Jae Yu

Feb. 11, 2002

Fermilab

- Introduction
- DØRACE Progress
- Workshop Goals
- Arrangements

What is DØRACE, and Why Do We Need It?

- DØ Remote Analysis Coordination Efforts
- In existence to accomplish:
 - Setting up and maintaining remote analysis environment
 - Promote institutional contribution remotely
 - Allow remote institutions to participate in data analysis
 - To prepare for the future of data analysis
 - More efficient and faster delivery of multi-PB data
 - More efficient sharing processing resources
 - Prepare for possible massive re-processing and MC production to expedite the process
 - Expedite physics result production

DØRACE Cont'd

- Maintain self-sustained support amongst the remote institutions to construct a broader bases of knowledge
 - Alleviate the load on expert by sharing the knowledge and allow them to concentrate on preparing for the future
 - Improve communication between the experiment site and the remote institutions
 - Minimize travel around the globe for data access
 - Sociological issues of HEP people at the home institutions and within the field.
- **Primary goal is allow individual desktop users to make significant contribution without being at the lab**

From the Nov. Survey

- Difficulties
 - Having hard time setting up initially
 - Lack of updated documentation
 - Rather complicated set up procedure
 - Lack of experience → No forum to share experiences
 - OS version differences (RH6.2 vs 7.1), let alone OS
 - Most the established sites have easier time updating releases
 - Network problems affecting successful completion of large size releases (4GB) takes a couple of hours (SA)
 - No specific responsible persons to ask questions
 - Availability of all necessary software via UPS/UPD
 - Time difference between continents affecting efficiencies

What has been done?

- Regular bi-weekly meetings every on-week Thursdays
 - Remote participating through video conferencing (ISDN) → Moving toward switching over to VRVS per VCTF's recommendation
 - Keep up with the progress via site reports
 - Provide forum to share experience
- DØRACE home page established (<http://www-hep.uta.edu/~d0race>)
 - To ease the barrier over the difficulties in initial set up
 - Updated and simplified instructions for set up available on the web → Many institutions have participated in refining the instruction
 - Tools to make DØ software download and installation made available
 - More tools identified and are in the works (Need to automate download and installation as much as we can, if possible one button based operation)

- Release Ready notification system activated
 - Success is defined by institutions
 - Pull system → you can decide whether to download and install a specific release
- Build Error log and dependency tree utility in preparation (See Gordon's talk)
- Effort initiated by Alan and release managers for alleviating unstable network dependencies
- SAM station set up in progress in many institutions
- All speakers' talk slides have been put on the web prior to meetings, so far, every time
 - A more systematic way of doing this in the works (See Mark's talk)

DØRACE Strategy

- Categorized remote analysis system set up by the functionality

- Desk top only
- A modest analysis server
- Linux installation
- UPS/UPD Installation and deployment
- External package installation via UPS/UPD
 - CERNLIB
 - Kai-lib
 - Root
- Download and Install a DØ release
 - Tar-ball for ease of initial set up?
 - Use of existing utilities for latest release download
- Installation of cvs
- Code development
- KAI C++ compiler
- SAM station setup

Phase 0
Preparation

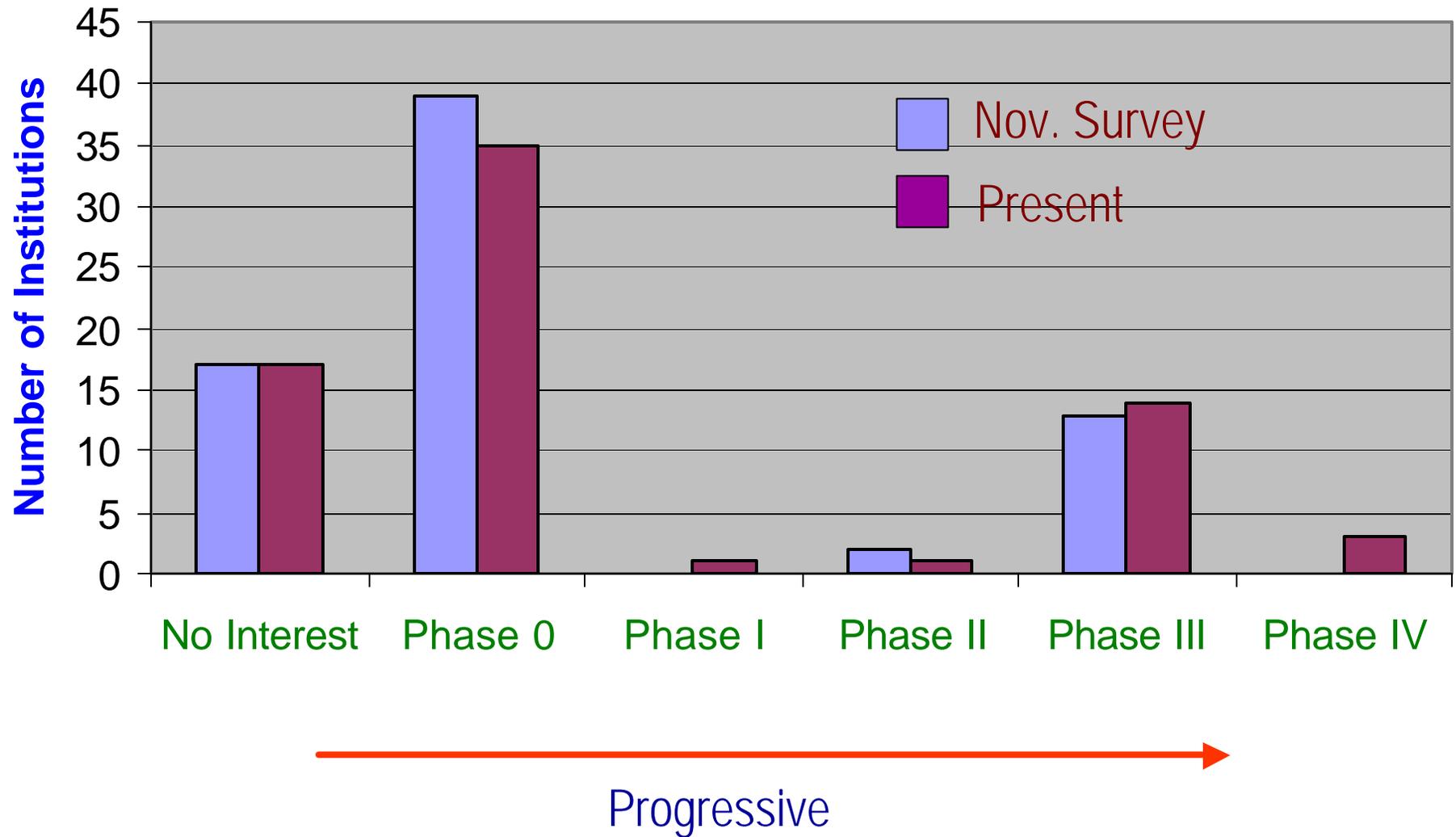
Phase I
Rootuple
Analysis

Phase II
Executables

Phase III
Code Dev.

Phase IV
Data Delivery

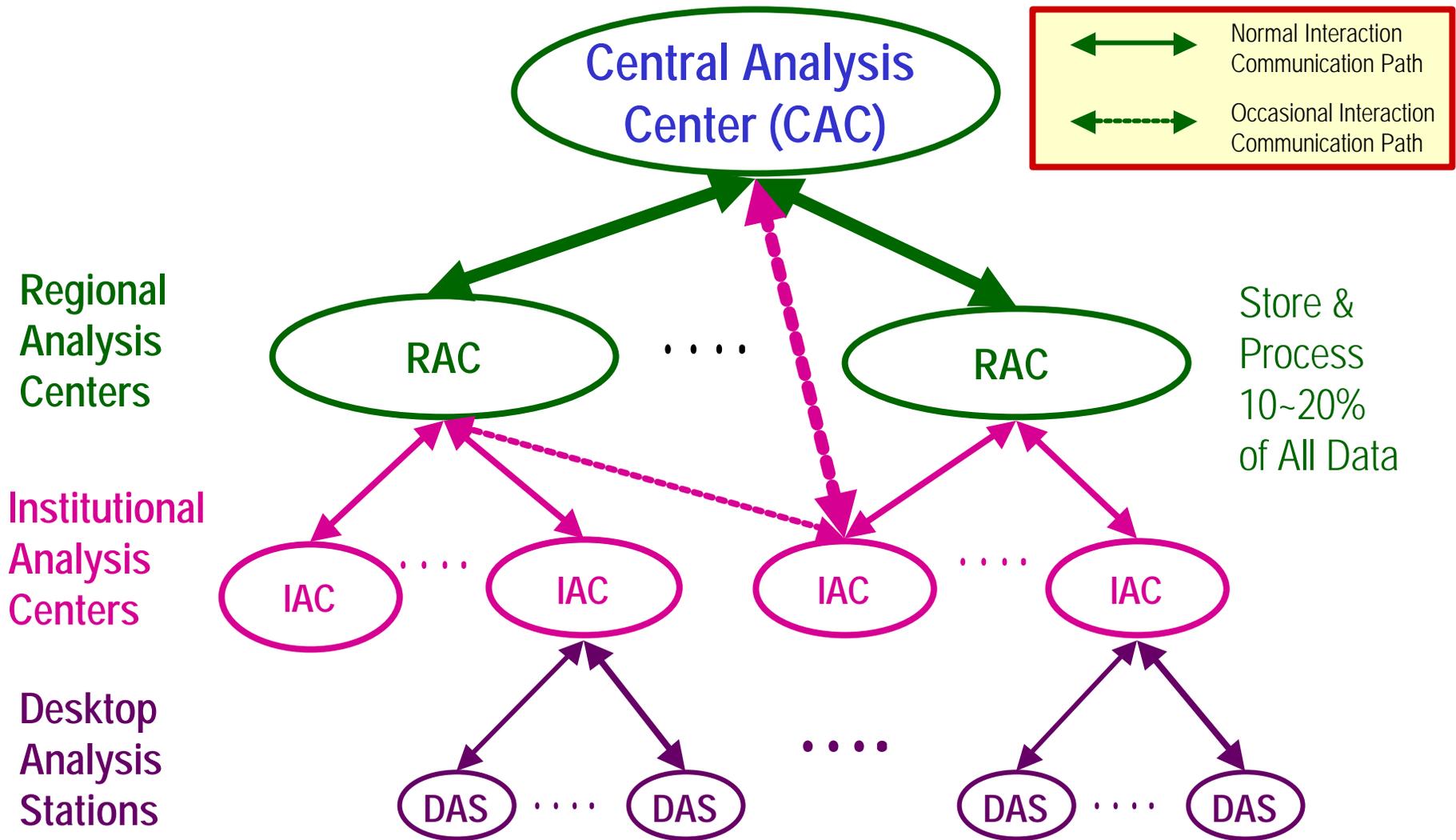
DØRACE Status by Setup Phases



Where are we?

- DØRACE is entering into the next stage
 - The compilation and running
 - Active code development
 - Propagation of setup to all institutions
 - Instructions seem to take their shape well
 - Need to maintain and to keep them up to date
 - Support to help problems people encounter
- There is the DØGRID group
 - Need remote institutions' participation
 - Prepare SAM and other utilities for this
 - More efficient resource sharing necessary
- We will need to establish regional analysis centers

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 (gigabit 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

What Do I Think We Should Do?

- Most the students and postDocs are at FNAL, thus it is important to provide them sufficient computing and cache resources for their analysis. → The Current suggestion for backend analysis clusters should be built!!
- In the mean time, we should select a few sites as RACs and prepare sufficient hardware and infrastructure
 - My rough map scan gives FNAL+3RACs in the US, and a few in Europe
- Software effort for Grid should proceed as fast as we can to supplement the hardware
 - We cannot afford to spend time for Test beds
 - Our set ups should be the Test Bed and the actual Grid
- A committee to determine number of RAC sites, their requirements, and select RACs within the next couple of months.

Workshop Goals

- Primary goals:
 - Establish as many institutions as possible with remote analysis, code development, and data delivery as possible through the afternoon “hands-on” session.
 - Identify available resources within the collaboration
 - Identify what each institution are working on for better coordination
 - Identify missing or anticipated to be missing pieces for exploiting remote resources
 - Identify tasks that need to be prepared for expediting data analysis at remote sites
 - Set common goals, task lists and schedules
 - Distribute tasks
 - Identify issues in communication and in sociological barriers and come up with possible solutions
- Establish clear road map for the future

Some Workshop Arrangements

- Today's afternoon session will be held at the computer training center located at the old Fixed target operation's center at 2pm.
 - Your package has the map to the center
 - There are 16 computers + 15 Ethernet cables in the small room
 - Additional 10 Ethernet cables + Wireless access point in the large room next to the training room
 - Instruction booklet are made and ready for your use at the session
 - Coffee break at the training center at 3:30pm
 - Parking is limited. Please share the rides as much as you can.
- Fermi RH 7.1 CD-Rom to be available at the registration desk. One per institution please
- MC Farm regular meeting will be in the RACE track (WH7X) beginning 4pm
- The banquet is at 7pm, Tokyo Restaurant. See the map and the directions included in your packet. → You need to pay now at the registration for convenience (Tip is included)
- Tomorrow morning's session begins at 9am here in 1West. Afternoon sessions are in 1 West and 1 North.