# Remote Analysis Coordination

- Computing hardware is rather inexpensive
  - CPU and storage media are inexpensive
  - Small institutions can afford to own reasonable size computing facilities
- DØ collaboration is larger and more international
  - Most the collaborating institutions are remote
  - Code development can occur at remote stations
    - Exploit available man-hours for much needed software development
    - Give ownership to collaborators from remote institutes
  - Optimal and efficient access to data is of utmost importance to expedite analyses
  - Minimize travel around the globe for data access
  - Exploit existing but scattered computing resources
  - Sociological issue of HEP people at the home institutions
- It is quite certain that sharing a 15-20$fb^{-1}$ worth of raw and reconstructed data (~5-7PByte?) efficiently will be a big issue
- **Primary goal is empowering individual desktop users**

# What do we need?

- Remote DØ software development environment
  - Allow remote participation for code development which might soon be a bottleneck in expediting physics results
  - Allow remote analysis for histogram production
  - Allow remote reconstruction or production environment
- Optimized resource management tools
  - Allow to maximally utilize local resources
  - Allow to tap into available computing resources in other locations
  - Allow participation of remote resources for global reconstruction or production
- Efficient and transparent data delivery and sharing
  - Allow location independent access to data
  - Allow quicker access to sufficient statistics for initial analyses
  - Allow data sharing throughout the entire network of collaboration
  - Minimize central data storage dependence
  - Alleviate load for central data storage and servers

# What do we have for remote code development?

- Three usage categories of remote DØ code development system
  - Minimal executables and necessary configuration files: MC farms, releases done through very light mini-tar
  - Accepting binary only done via tar files, heavier than mini-tar. For local executable running ✍ Run-time environment effort will help this case
  - Full blown code releases; binaries via tar files and sources via ups/upd
- Both p and t releases are prepared for IRIX and Linux for remote stations (no other platform available for remote farms) weekly
- The remote release system is "PULL"
  - Minimize unnecessary network traffic of over 2GB releases to over 60 institutions every week
  - Minimize security issues to remote institutions
  - Give full control to the remote stations
  - Synchronization of the code is an issue

- Documentation and instructions for remote code development system setup are
  - Surprisingly well written, despite the fact that they are from a few years ago, but need updates
  - Explanation and instruction for setting up remote code release system are pretty accurate, but those little things ✍
- But obviously there are many issues (Don't ask me about video conferencing…)

# What has been done so far

- Have established a listserver distribution list, d0-remote-analysis
  - Consists of 50 subscribers who are institutional contacts so far
  - Will add a few mandatory names, such as Alan, Paul, Lee, Vicky, Heidi, Amber, Wyatt, Iain, etc
- One institutional Contact requested per each institution responsible for (People with more permanent position for continuity)
  - Setting up remote analysis stations
  - Keeping up releases
  - Answer institutional user inquiries
  - Channel through the inquiries that can't be answered locally to larger and more experienced crowd, d0-remote-analysis, to share experiences
  - Provide institutional supports for analysis or code development efforts
  - Participate in development of remote-analysis tools
  - Participate in testing and evaluating the tools
  - Establish necessary infrastructure for institutions (network, disk space, etc)

- Sent out Survey to gather information
  - How many were established
  - What is the depth of software download and installation (binaries, source)
  - Preference in Pull vs Push release system?
  - Biggest difficulties in establishing remote-sites?
  - What can be improved?
  - What can institutions offer?
  - Tasks for efficient remote analysis establishments?
  - Attend remote-analysis workshops?
  - Topics to be discussed in the workshops?
- 53/76 institutions (30/34US, 19/30 European, 5/7SA, 1/5A) assigned institutional contacts (45)
- 44/53 responded to the survey

THANK YOU!!!!

# Survey Results

- Workshop
  - Contacted three European institutions per Kors and Iain's suggestion
    - Marseille
    - Imperial College
    - Lancaster
  - None can do it in January, too short a time
  - Not many institutional representatives can attend
    - Due to time restriction
    - Due to budgetary constraints
  - Will have the first mini-workshop (1-2days) at around Feb. collaboration meeting in the US, most likely at Fermilab
  - Follow up a couple of months later

# Items to Tackle at the Workshop

- Primary goals:
  - Identify available resources within the collaboration
  - Sharing experiences
  - Understand the current status ✍ Exchange ideas
  - Identify missing or anticipated to be missing pieces for exploiting remote resources
  - Identify items that need to be prepared for expediting data analysis at remote sites
  - Identify necessary tools to empower desktop users
  - Set common goals, task lists and schedules
    - Distribute tasks
- Establishing clear road map for the future

- Remote site establishments
  - 13/28 established (6/17US, 3/6 Europe, 2/5SA)
- Pull vs Push
  - 17/20 prefer “pull” ✍ Full local control of the releases and versions
  - Prefer more automated system of “pull”
  - More efficient notification of build completion *(definition of a successful completion??)*
  - Synchronized update of other external products (heptuple, root, OpenInventor, etc)
  - There are a few utilities floating around ✍ John Ellison of UCR has agreed to provide the two scripts that he uses for general DØ software download and installation

- Depth of the code installation
  - 19/26 (12/17US, 5/5 E, 4/4SA) want full source code
- 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 along 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

- Offer for help
  - Some institutions volunteer for testing
    - Univ. of Wuppertal
    - Mainz
    - UTA
    - KSU
    - BU
  - Some institutions offer manpower in various forms
    - UTA
    - Langston
    - UC Fresno
    - LA Tech
  - Some offer specific tool developments
    - Build error information (Washington)
    - Pick-n-choose download and installation (UC Riverside)
    - Run time environment (Imperial College)
    - SAM & Condor batch submission (Imperial College)
    - PACMAN development (Michigan)
- But most institutions have hard time coming up with help

- Some suggestions from institutions
  - Agree upon a common OS for remote sites
  - Establish regional data analysis sites
  - Establish release procedure test bed
  - 24x7 coverage of d0mino or isolation from d0mino
  - Transparent delivery of data
  - Easier initial set up
  - GRID implementation of SAM at DØ
  - Automatic and simultaneous releases of external packages
  - Split releases in pieces so that installation is less susceptible to network interruptions
  - Easier sharing of experiences
  - Expedited update of Fermi RH versions

# So what should we do?

- Make the initial set up simpler and easier
  - Provide updated document and maintain them in regular bases
  - Provide initial set up script that needs only a push of a button
- Establish automatic "release-ready" notification system using the distribution d0-remote-analysis
  - Need to agree on a definition of "release-ready"
  - My suggestion is let the release managers define this
  - But provide sufficient information on the release (Gordon Watts – Release error log w/ dependencies; I will owe him a Sabboro beer when it's ready!!! )
- Provide tools for simplified (preferably one button web operation) "pull" based download and installation
  - UCR (John Ellison's) tool seems to be a good starting point
  - PACMAN?
  - Dylan Casey wrote a perl script ✍ Improve this?

- No other OS supports (IRIX, LINUX + minimal OSF)
  - Remote stations are responsible for local build for unsupported OS
- Start a bi-weekly (every on-week) remote-analysis meeting to share information and experiences
  - I will find a room and send out message to remote-site contact persons
  - Pick a few sites and get them through the set up process ✍ Refine the documents, tools, and procedures
  - Goal is to get all other institutions that want remote-analysis release system ready within 6 months or by next summer
- In the mean time, will prepare for workshop in Feb.
  - Alan, Iain, Kors, Lee, Vicky, Amber, Wyatt, Heidi, and many others' help are needed