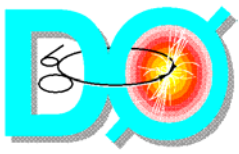# Sam-GRID Overview
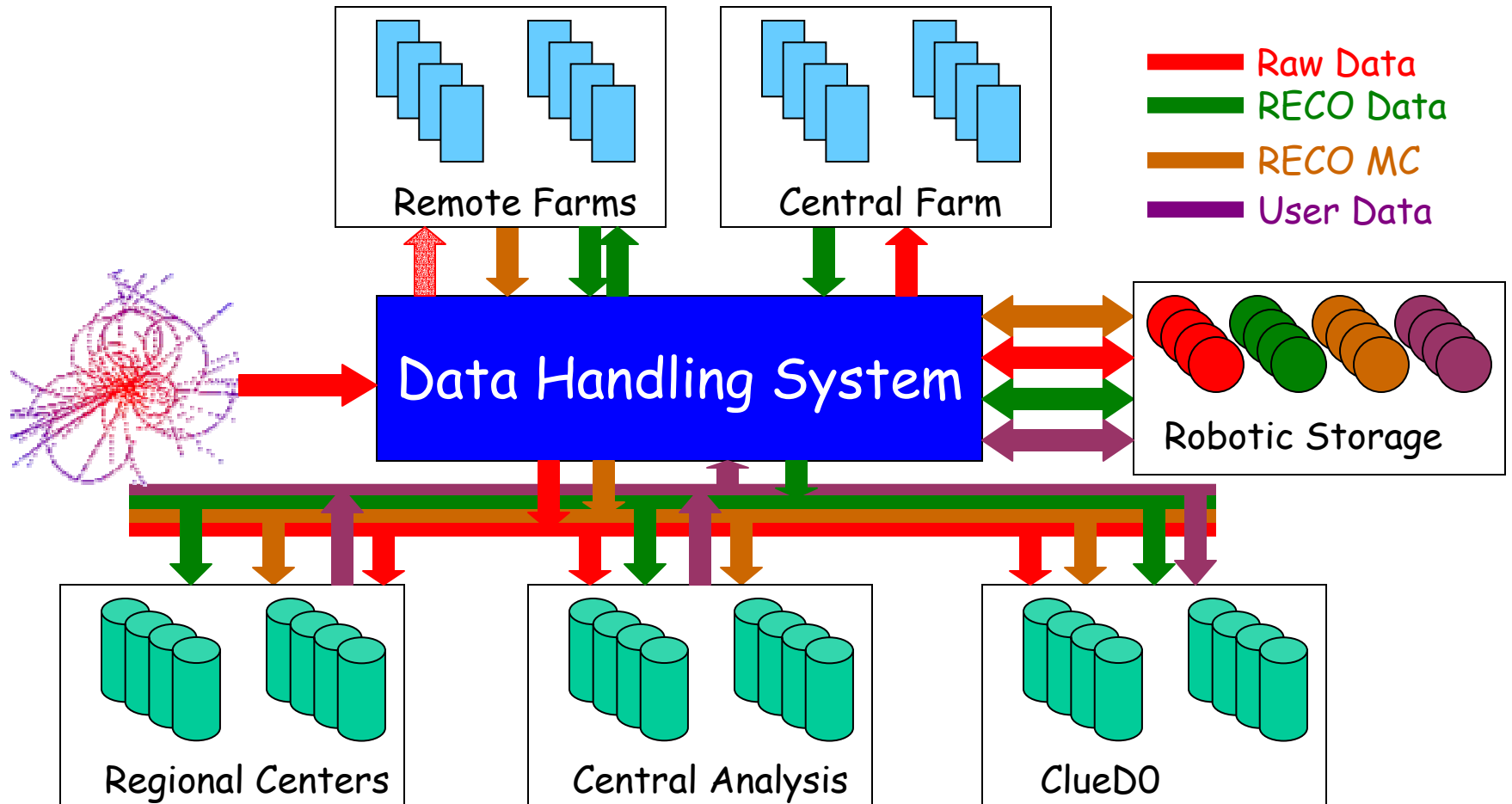
Amber Boehnlein
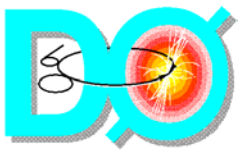
SAR Workshop

Sept 26, 2003

A Word of Thanks

- **The Southern Analysis Region is pioneering some key concepts for global analysis. Congratulations!**

- **Thank you for inviting me to participate and I regret that I could not attend the meeting.**

Amber Boehnlein, FNAL

Data Flow

Raw Data
RECO Data
RECO MC
User Data

Remote Farms

Central Farm

Data Handling System

Robotic Storage

Regional Centers

Central Analysis

ClueD0

~55000 FNAL projects and ~6000 remote projects have been run

Amber Boehnlein, FNAL

What's in the Blue Box

- **We classify and categorize our data, thinking in terms of datasets, and track its usage.**
  - **Use Oracle database as central metadata store**
- **We have the ability to distribute data from one location to another**
  - **Protocols include bbftp, GridFTP, rcp, encp, dccp**
- **We have systems for job processing**
  - **Sam submit, DOtools,MCRunJob and JIM (in testing)**
- **We interface to local storage and batch systems**
  - **HPSS, TSM, LSF, PBS, FBS, Condor and SGE**
- **We have stable operations**
  - **Operational load is still relatively high for small team of developers.**
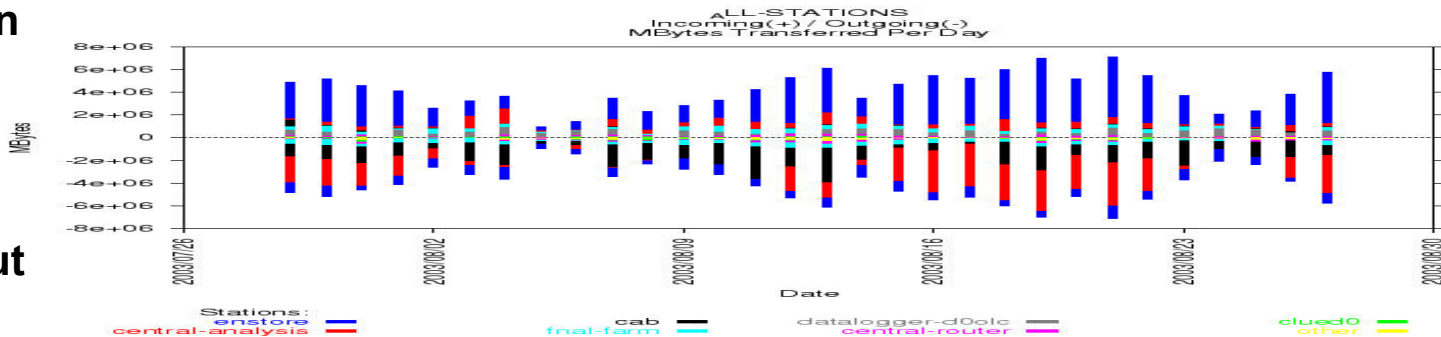 - **Need 24/7 shift support.**

Amber Boehnlein, FNAL
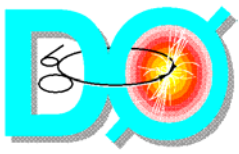
Data to tape as of Sept 8, 2003, ~ 5 GB lost

Library	Data Tiers	Storage
9940A	**Raw (May)** **DST (Feb)**	**207 TB**
9940B	**Raw** **DST**	**116 TB**
LTO	**MC Tiers** **Data TMB**	**94 TB**

Plot shows daily transfers in/out of SAM stations Colors represent stations- Blue is to/from tape and Red is to/from farm
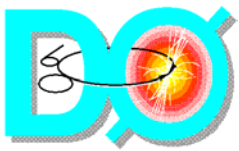
8TB in

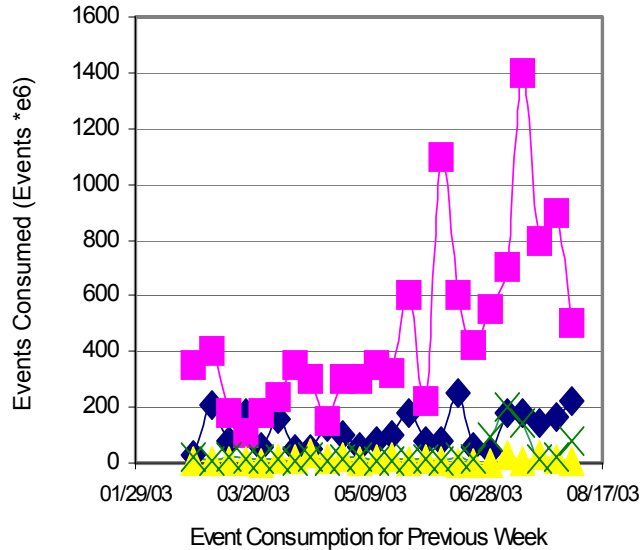**8TB out**

August, plotted daily

Note: daily transfers doubled
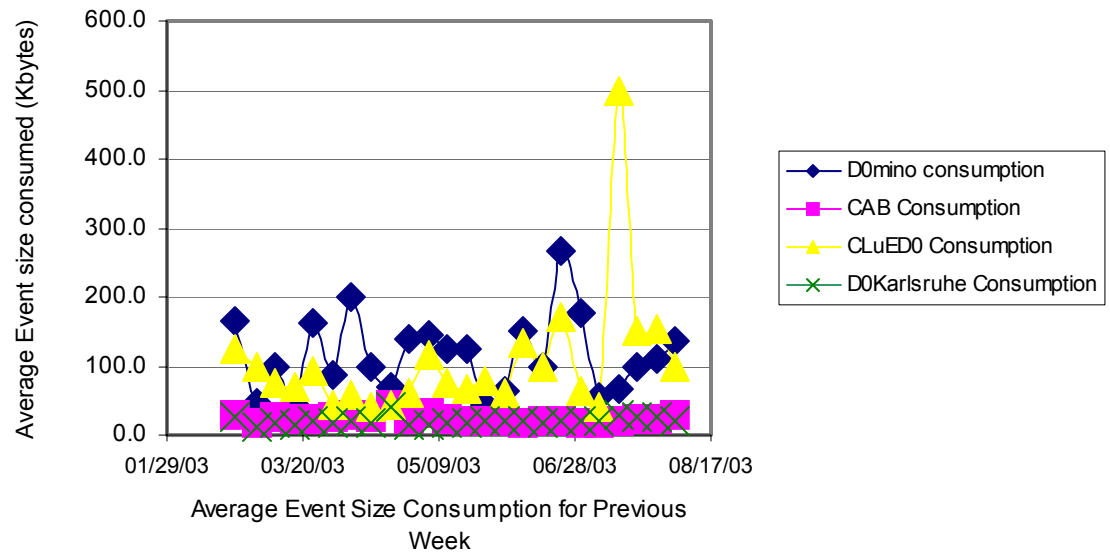Since March

Amber Boehnlein, FNAL
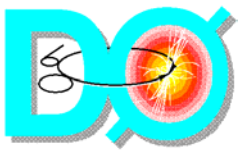
Data Handling Metrics
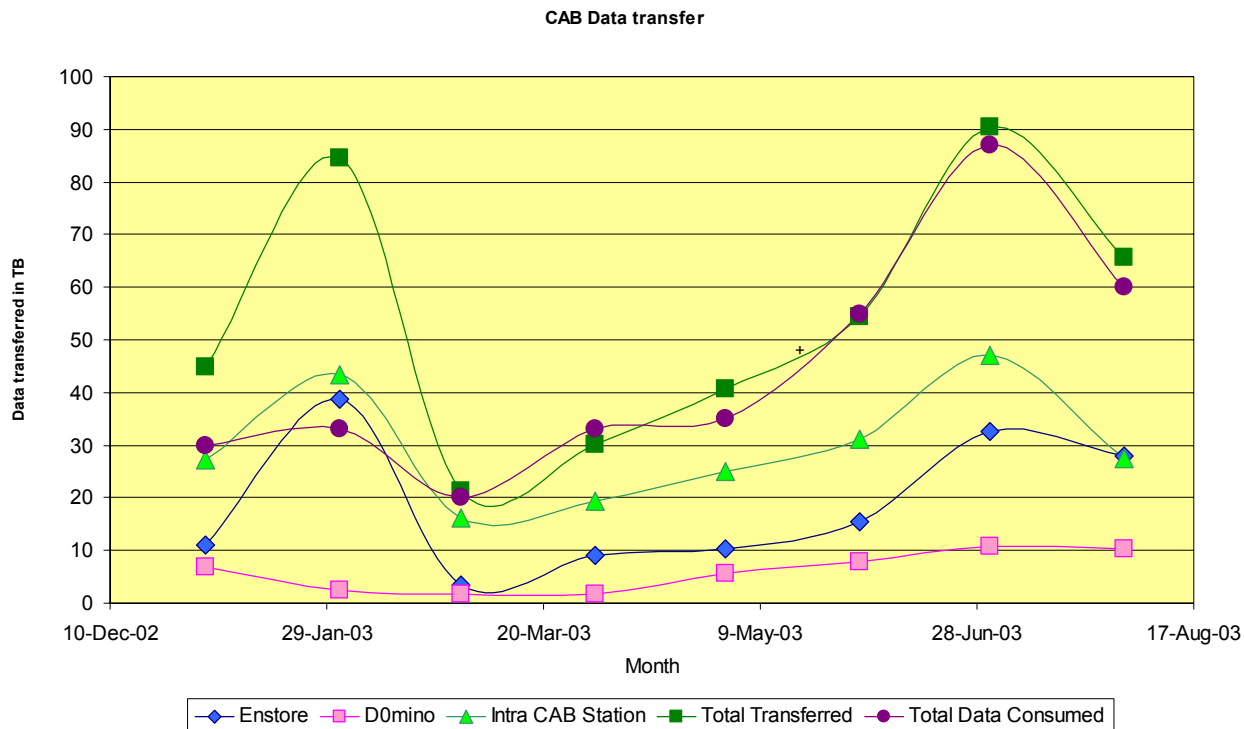
Event Consumption on Analysis Stations

Consumption plot gives information
On analysis uses cases.
Use more varied on desktop, Domino
Than on CAB or DOKarlsruhe

Average Event Size Consumed on Analysis Stations
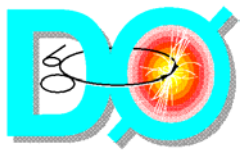

Amber Boehnlein, FNAL

CAB File Transfers

CAB Data transfer

Major components of data transfer on CAB
Slight excess transfers is fine—jobs crash while SAM delivers
Station problem in Feb now fixed.

Amber Boehnlein, FNAL

Data Handling Extensions
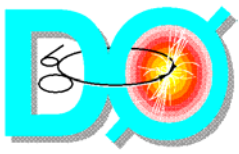
- **Improved operations**
  - **Investigate slow transfers**
  - **Continue to improve tools, documentation metrics**
- **Short Term Extensions**
  - **Job broker working with MC (JIM)**
  - **Support for Remote Systems**
  - **Cache working group--dCache**
  - **Linux Fileservers**

Data Handling System

dCache Write

Robotic Storage

Central Farm

Application

Amber Boehnlein, FNAL
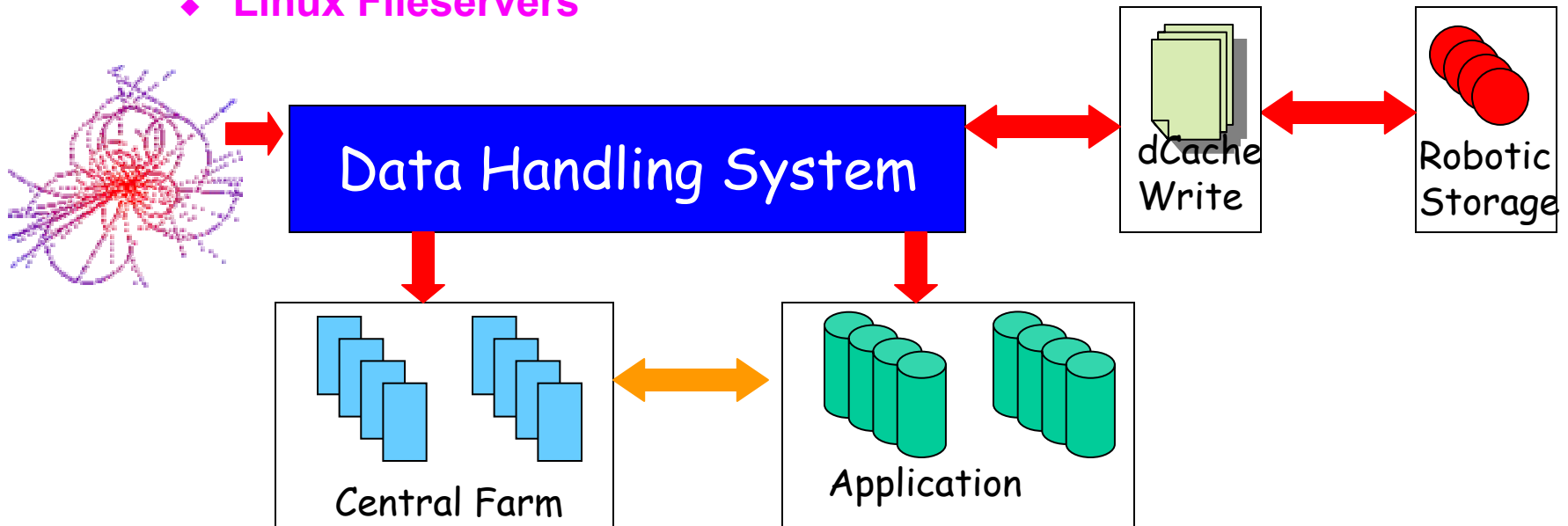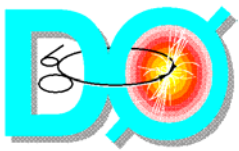
Longer Term

- **SAM-Grid has a task list which includes roughly 2 years worth of functionality and operational improvements.**
  - ◆ **Enhance job broker for re-processing**
  - ◆ **Schema updates to include processing requirements for jobs**
  - ◆ **Revise caching strategies**
  - ◆ **Implement Virtual Organization tools (with CMS?)**
  - ◆ **Implement improved monitoring and information server on the data handling side**
  - ◆ **Sort out issues with parallel jobs for analysis**
  - ◆ **Provide for distributed database, separating file location and processing information**
- **Team is small, is likely to get smaller, and CDF is still not using SAM centrally— but project cannot extend infinitely!**

Amber Boehnlein, FNAL

D0 Grid Meeting

- **Having a " DO Grid" strategy involves both technical and sociological challenges:**
  - **Learning to build a robust and performant grid**
    - **How to leverage all available effort by working as partners**
    - **Understanding the use cases, the error handling and systems in a profound way to better outline the interfaces and boundaries and constraints.**
    - **System and user Monitoring designed in—not an after-thought.**
    - **Learning to bring products in on time.**
    - **Avoid re-engineering the other guys' stuff.**
    - **Curing "FermiLab-centric" and other insular thought patterns**
    - **Understand and mitigate security risks**
  - **Learning how to install, deploy, and operate it.**
    - **Understand what the physics needs are, and tailor development along those lines**
    - **Training the users, installers, operators**
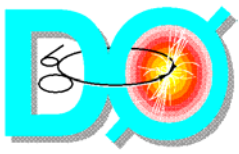
Conclusions

- **Data handling for D0 is very successful**
  - ◆ **Our ability to do remote computing is predicated on the thought that went into SAM at the design stage**
  - ◆ **We are taking the next logical steps**
- **SAR is a pioneer in concepts necessary to make for a successful grid strategy**
- **I hope that many of you will contribute to the DO Grid Meeting on Oct. 7**

Amber Boehnlein, FNAL