CDF and the Grid

Presented on Behalf of CDF-UK and FNAL collaborators
Rick St. Denis, University of Glasgow

- Requirements and Anti-Requirements
- CDF-o-Centric View
- The Project
- The manpower
- Conclusion: CDF/D0 Deliverables

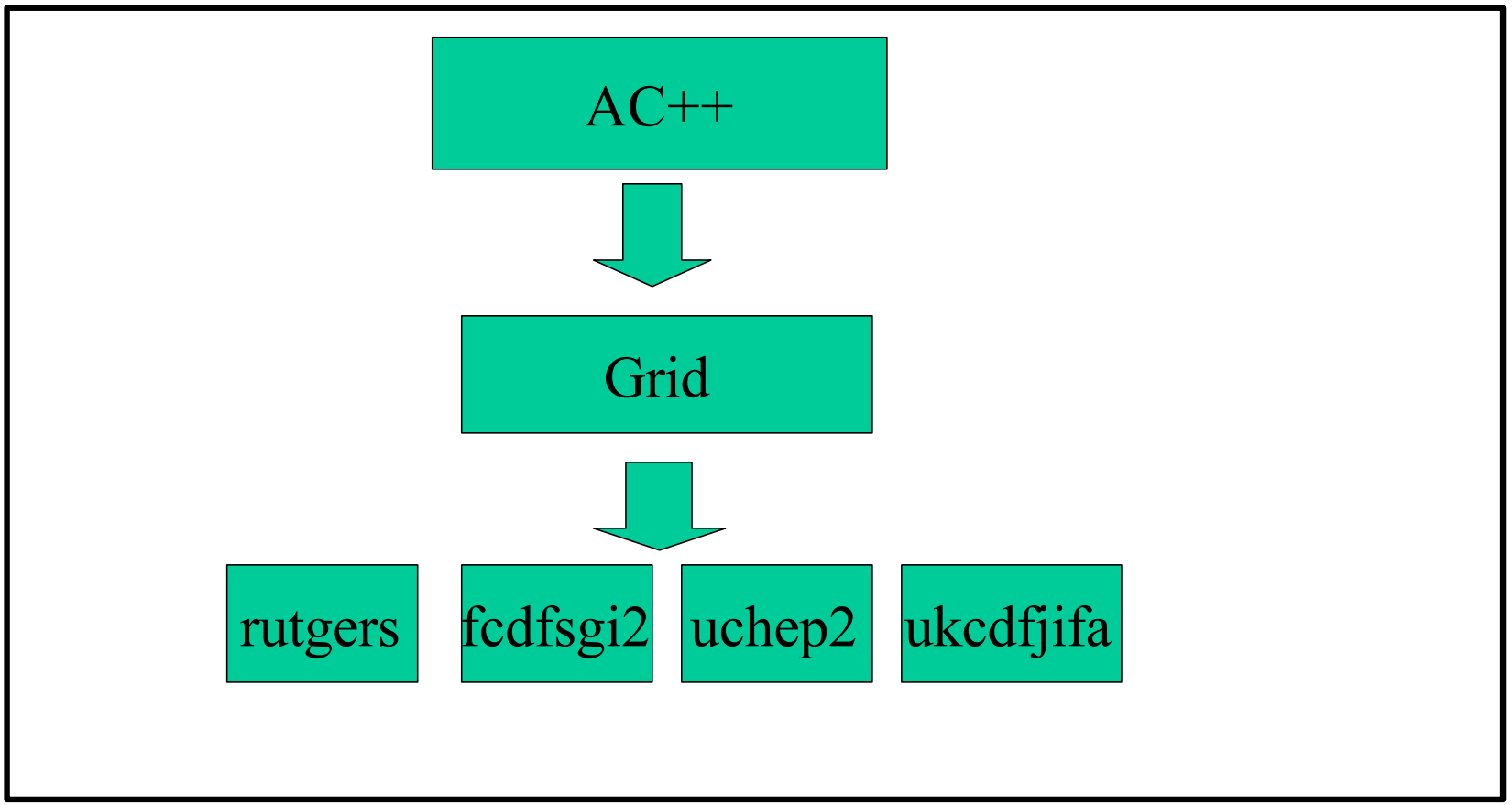
Basic Requirement

What you really really want:

- One place to log in and submit jobs
- Space for input
- Space for output
- Ability to find data
- Some else to worry about it (I DON'T mean D0!)
- Works to 2007 (or more?)

In short: Sufficient resources for CDF's Life

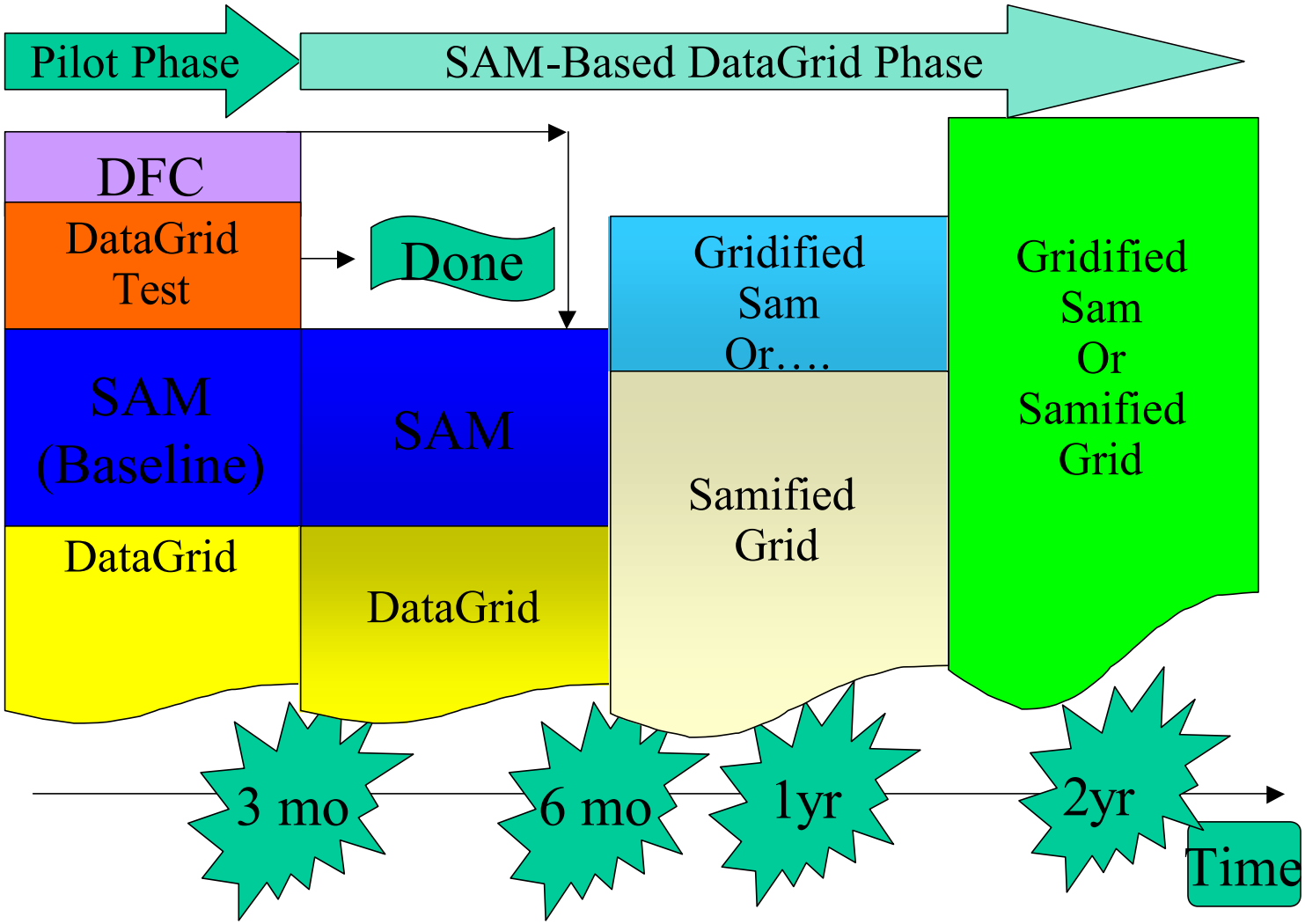
The CDF View of Grid



CDF Grid/Sam Project

- Three Project Phases:
  - »Pre-Pilot
  - Pilot (3mo)
  - Integration (3mo)
  - Gridification (Feature enhancement)
- Project end: Dec 31, 2004

CDF Grid Project Outline



Pre-Pilot Project

- AC++ access
- Distribution of SAM stations
- Database and Server installation
- Enstore access

Pre-Snake Oil Test

PrePilot 1. AC++ Access

- 1-1. Smooth installation of a sam station set up
- 1-2. Program connects to a project and serves files
- 1-3. CDF file declared to SAM: recon @ Glasgow
- 1-4.File moved between CDF institutes: UCL/Gla
  - The sam test file that is currently at Glasgow
  - The CDF file on the D0 database
- 1-5. SamInput module to read file
- 1-6. SamInput to read/process/read/process…
- 1-7. Run through multiple files
- 1-8. Declare more data and serve more files

PrePilot 1. AC++ Access

- Has been turned over to UCL for support, development and documentation.

PrePilot 2. Station Distribution

- 2-1. Upgrade upd server for RH7.1
- 2-2. Oxford station
- 2-3. Chicago station
- 2-4. MIT station
- Station Distribution is still Awkward: works with a person at FNAL on video who can call experts in for immediate consulation

PrePilot 3. Database and Server

- 3-1. SAM into the CDF dev database.
- 3-2 Middle tier servers on linux; suite of checks to be run. (not on DB server)
- 3-3. Use the CDF database;repeat 1-5.
- 3-4. Estimate table sizes.
- Could follow “Standards”: Integration then Production Database (in next week)

PrePilot 4. Enstore

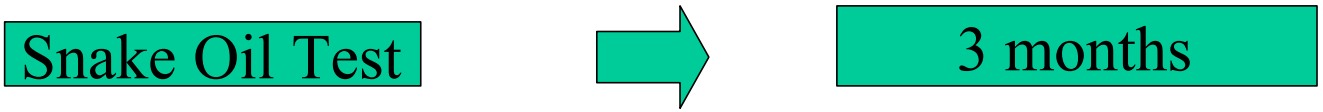
- 4-1. SAM station on machine with enstore access.(Rob Kennedy)
- 4-2. Access files from Enstore.
- 4-3. Present a plan for cutting of SAM into CDF production.

PrePilot 5. The part I forgot

- GUI and database entry: We have to adapt tools
- The CDF database Browser is getting a SAM section(Randy Herber)

Pilot Project

- Set up the initial fabric.
- Deploy SAM as well as standard Grid tools for evaluation.
- Evaluate usage patterns and user requirements.



Integration

- Deployment and support of SAM stations in UK. PrePilot long way to this: Glasgow, UCL Oxford support.
- Adapt CDF software to function with SAM.
  - CDF data handling tools: translate cataloguing information to SAM. Key component: interfacing of the Database schema compatible with SAM (or even in common with DØ). Randy and Dmitri working on this.
  - Data input modules:Differences with CDF/D0 in Event Data Model and how data files are catalogued. PrePilot long way to this: handed over to UCL.

Gridification (with D0)

1. Standard middleware for interoperability.
   i. Globus security and Kerberos, CAS in future.
   ii. GridFTP.
   iii. Globus job submission as a supported system.
   iv. Condor and extensions as a supported system.
   v. Publish SAM station resources/services availability using emerging Grid middleware.
   vi. Publish catalogue of data files using emerging PPDG/DataGrid standards.Start communication with WP2: Replica catalogue

Gridification (with D0)

2. Grid functionality for Job spec., submit, tracking
   i. Condor: migration, checkpointing, CDF/D0/Condor mods.
   ii. Incremental enhancement: Job spec, submission language,co-location of job execution/data files. Query optimization algorithms under study with simulation of CDF CAF (Will Bell, Glasgow)
   iii. Reliable, dependencies between job steps. (DAGMAN starts).
3. Monitoring, diagnostic, logging: use Grid Mon.Tools
4. Fabric Management: UK CDF JIF , the UK DØ JIF computers, Lancaster farms, plus future facilities such as Mosix or SMP clusters

Manpower and Projects

- Sam Product support
- Database Schema
- Query Language for Metadata; Browsing
- Performance and Monitoring
- Simulation
- Batch
- Machines and communication

Sam Products

- - distribution
- - support
- - deployment
- - Station distribution: Alan, Paul, Kevin, UCL

Database Schema

- Will, Gavin, UCL, Randy, Sill, Ombretta
- Size estimate and growth monitoring
- performance monitoring and simulation
- schema cuts, improvements,coordiation with d0 (includes scripts for select from insert into)
- Api design, jdbc interaction
- Design issues
- Interface with grid and dfc: Randy, Dmitri

Database Query Layout

- Will, Gavin, Randy
- SAM dimensions, interaction with input module (Paul)
- /grid (Kevin)
- webdav (Alan)

Monoitor/Simulate/Browse

- Performance monitoring- CAF and SAM: Alan
- Simulation Will/David CAF and SAM
- Batch: Rod and Alan - Start with SAM implementation/ then grid implementation/absorption: Condor
- Browser: user interfaces: Randy with Will, Gavin, Kevin

Machines

- Coordinate with RAL:Extend to other labs. Need to share information: Bugzilla (or jitterbug)
- Integration, dev, prod. Make the Glasgow machine
- An integration machine, one RAL machine for Glasgow production
- Need a pc for development
- Spec maintenance with JIF, include development, integration, and production for CAF

Conclusion

- Using AC++/SAM to analyze is just around the corner
- Station Distn should go quickly (UK)
- CDF Development database running. Sizing needed for production. Enstore under way.
- Grid started with GlwUK: Replica catalog and query optimization.
- Interfacing with Condor/Batch group comingb

CDF/D0 Deliverables

Month 3	CDF Pilot project ends with test of Globus tools and SAM. A deployed SAM station at one or more CDF institutes should be available to transfer files for data analysis.
Month 6	Integration of Globus Security infrastructure and GridFTP into SAM and deployment at several UK stations interoperating with Fermilab and other SAM stations. First demonstration of MC production system using Request interface and automated job submission to one SAM station with limited intelligence in job distribution and load balancing.
Month 12	Fully commissioned MC Production System with reliable execution of jobs, splitting into sub-jobs as necessary, intelligent job distribution and load balancing, taking into account the economics of data movement versus job movement.
Month 24	A fully robust production quality system, excellent monitoring and interoperability with other Grid projects and EU DataGrid with sharing of some resources. Updated Fabric capable of handing the data on each experiment.