



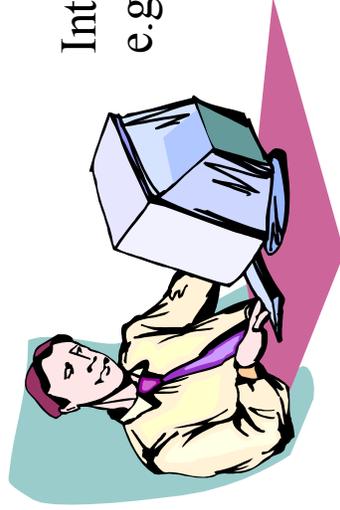
DIAL and Datasets

David Adams
PPDG All Hands
10th June 2003, BNL

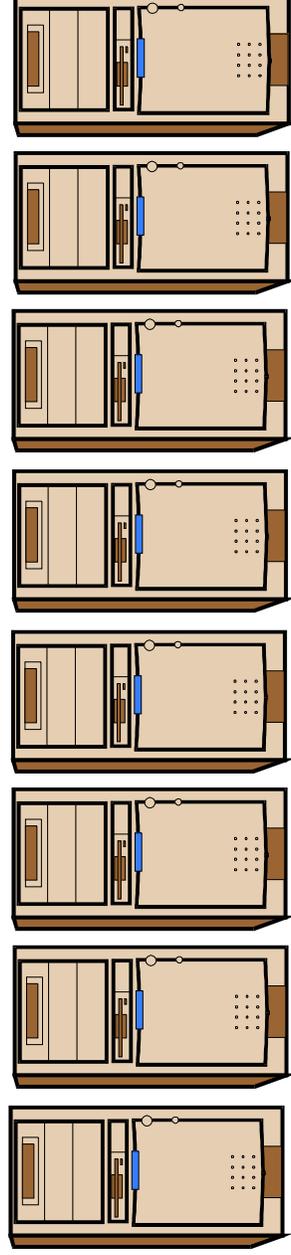
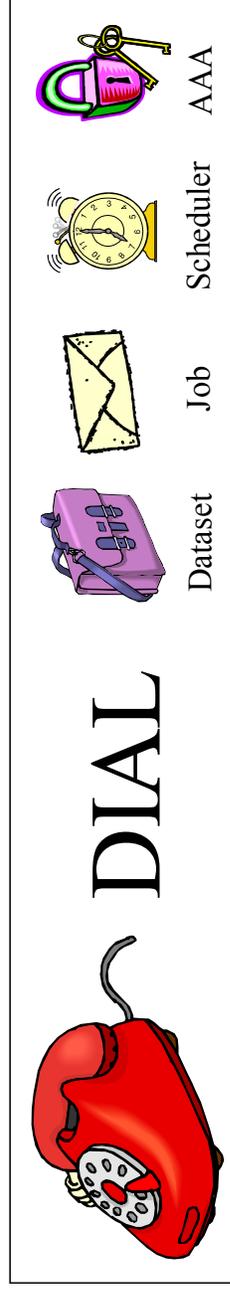


- Distributed Interactive Analysis of Large datasets
- Goals
 - Demonstrate feasibility of analyzing large dataset
 - Distributed data and processing
 - Set requirements for grid components and services
 - Provide useful distributed analysis tool for ATLAS
 - And other interested parties
- Design
 - Front end (interactive framework) provided externally
 - E.g. ROOT, JAS, SEAL-PI
 - Support back end application natural to the data of interest
 - Figure shows major components and their interactions
 - Another shows the scheduler hierarchy
 - Important component is the dataset

DIAL design

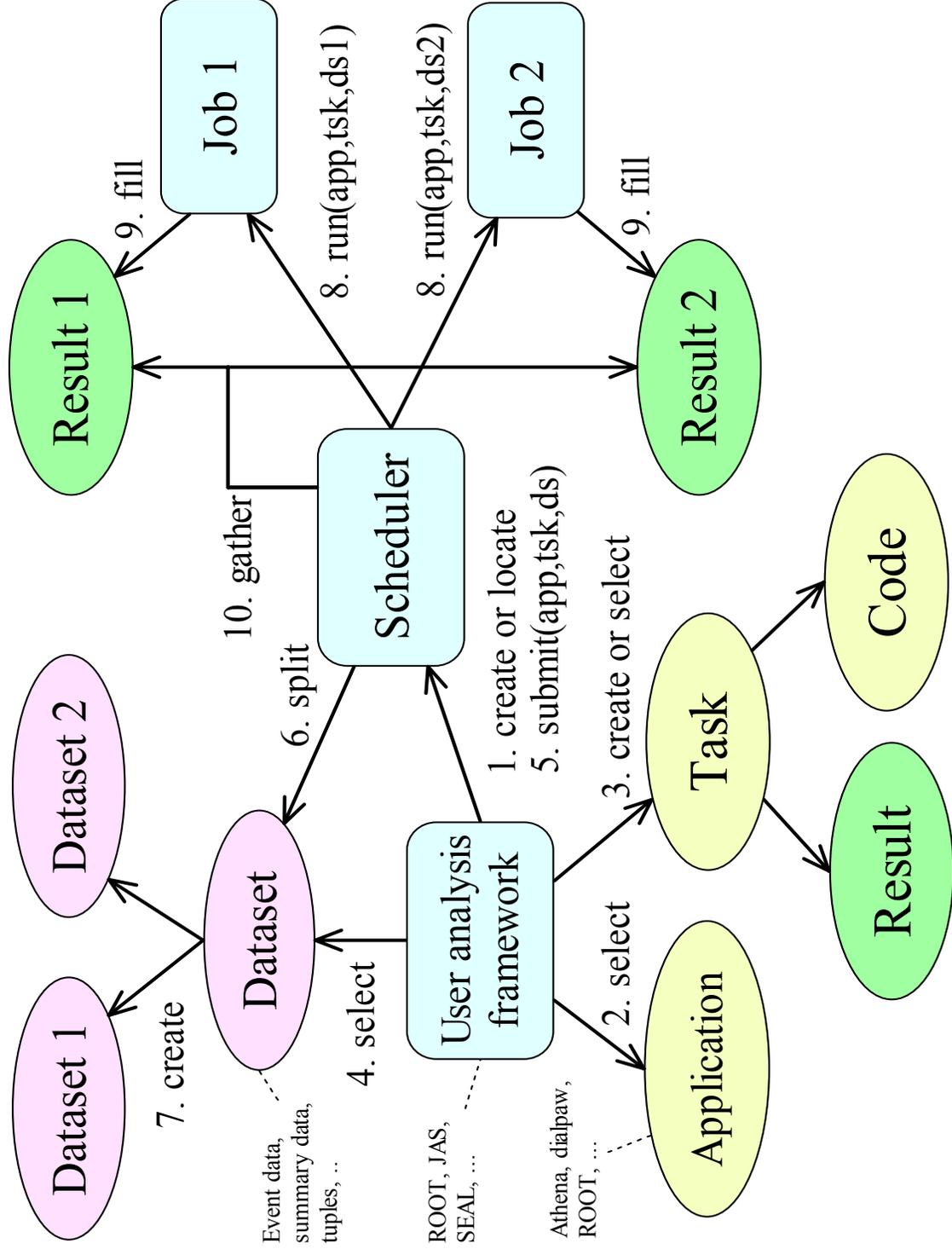


Interactive analysis
e.g. ROOT, JAS, ...

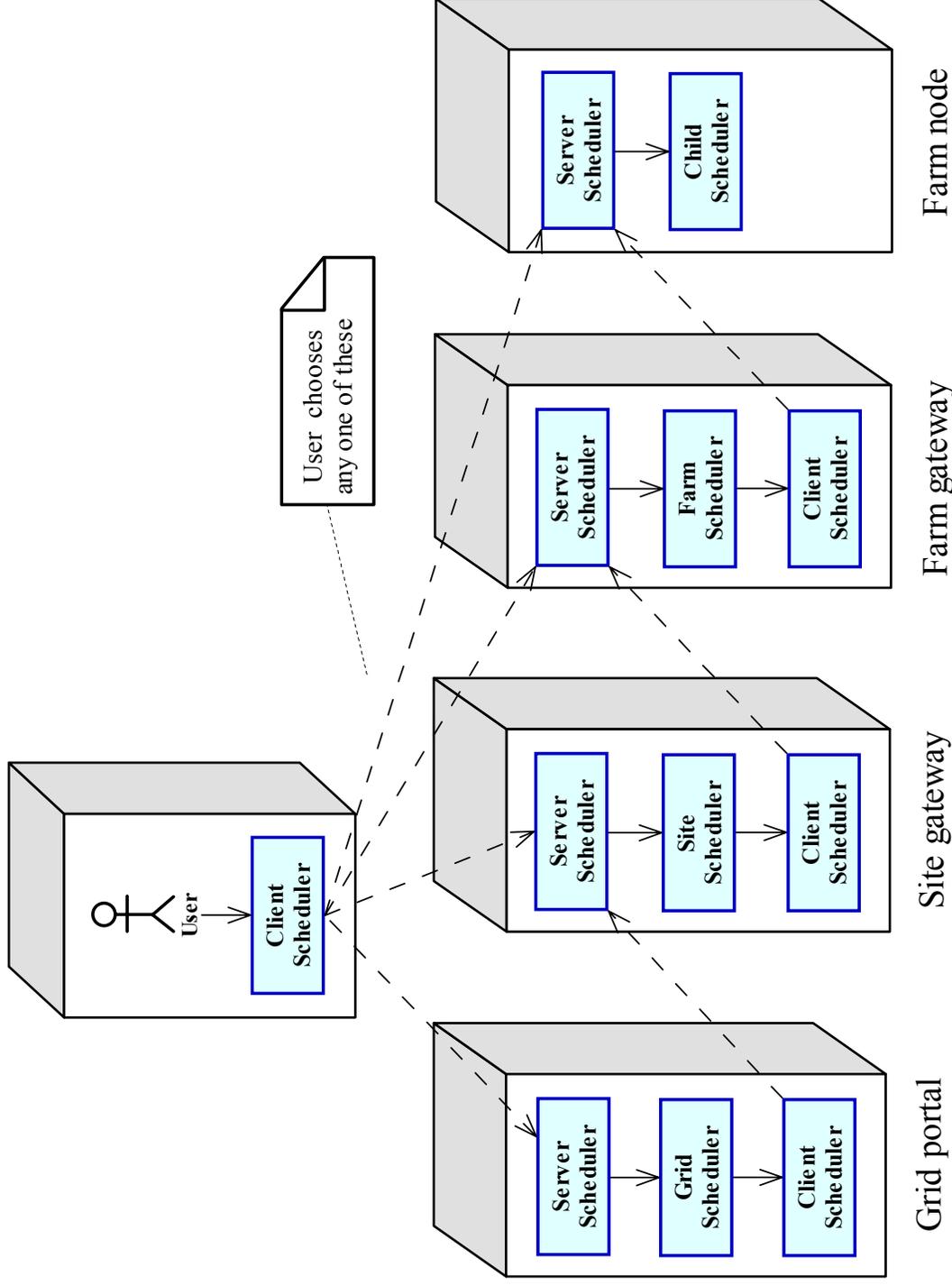


Distributed (data-specific) processing

DIAL components



Scheduler hierarchy



Deliverables and Milestones



- Talk and paper for CHEP 2003
 - Summary of DIAL design, status
 - Begins generation of grid requirements
 - Includes discussion of datasets
- Release 0.2 in February 2002
 - Interfaces and at least one concrete implementation in place for all components
 - Only scheduler starts process on current node
 - Move to gcc 3.2
 - All components imported to ROOT using ACLiC
 - ROOT can be used as DIAL front end

Deliverables and Milestones (cont)



- Release 0.3 expected in June or July 2003
 - ATLAS combined ntuple dataset
 - Application dialpaw enable use of PAW as back end
 - Demonstrate DIAL look and feel for ATLAS
 - (Processing not yet distributed)
- Fall 2003
 - Magda file catalog
 - ATLAS POOL dataset
 - Athena as application
 - Farm scheduler
 - Useful distributed processing for ATLAS
 - Very simple grid scheduler for DOE grid demo?

Deliverables and Milestones (cont)



- Winter 2004
 - Begin developing grid scheduler
 - Identify and integrate grid components
 - Replica catalog
 - Dataset catalog
 - Authentication and authorization
 - Resource location and allocation
 - Job submission and tracking
- Spring and summer 2004
 - Release system to support grid distributed interactive analysis for ATLAS data challenge 2
- Integration with other analysis projects
 - GANGA
 - DAWN

PPDG - Questions and Views



- How my work fits into the PPDG project
 - DIAL goals closely aligned with PPDG
 - ATLAS provides testbed for PPDG ideas and tool
 - Interactive analysis
 - GRID components and services
 - Dataset is understood as fundamental building block for distributed analysis (both DIAL and PPDG)
- My top 3 wishes for PPDG
 - Identify components and services required for distributed interactive analysis
 - Define interfaces for these components and services
 - Locate or provide reference implementations