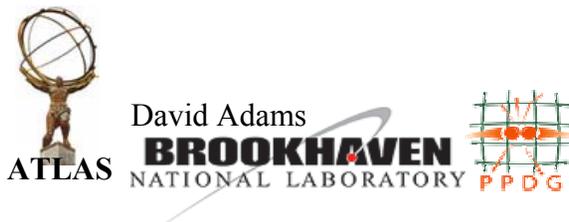


DIAL Status and Plans

ATLAS Software Workshop
Grid session
CERN

David Adams
BNL
September 25, 2003



Contents

Goals of DIAL

What is DIAL?

Design

Status

Development plans

Integration with GANGA

Goals of DIAL

1. Demonstrate the feasibility of interactive analysis of large datasets
 - How much data can we analyze interactively?
2. Set requirements for GRID services
 - Datasets, schedulers, jobs, results, resource discovery, authentication, allocation, ...
3. Provide ATLAS with a useful analysis tool
 - For current and upcoming data challenges
 - Real world deliverable
 - Like to add another experiment would show generality



What is DIAL?

DIAL provides a connection between

- Interactive analysis framework
 - Fitting, presentation graphics, ...
 - E.g. ROOT, JAS, ...
- and Data processing application
 - Natural to the data of interest
 - E.g. athena for ATLAS

DIAL distributes processing

- Among sites, farms, nodes
- To provide user with desired response time

Look to other projects to provide most infrastructure



David Adams

BROOKHAVEN
NATIONAL LABORATORY

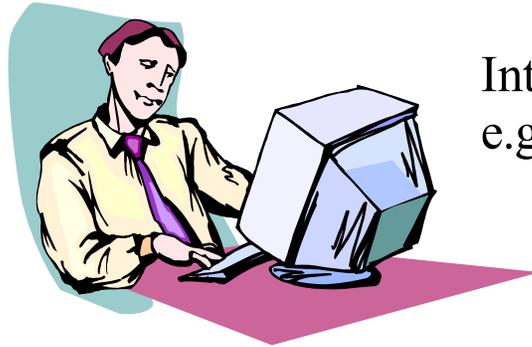


DIAL

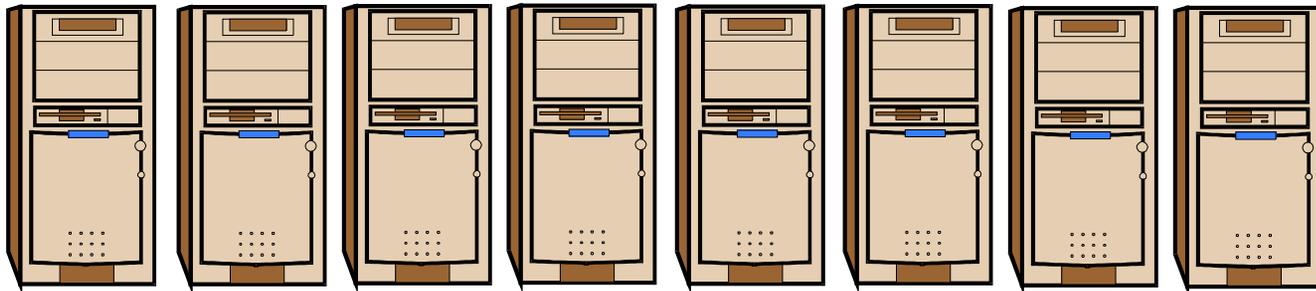
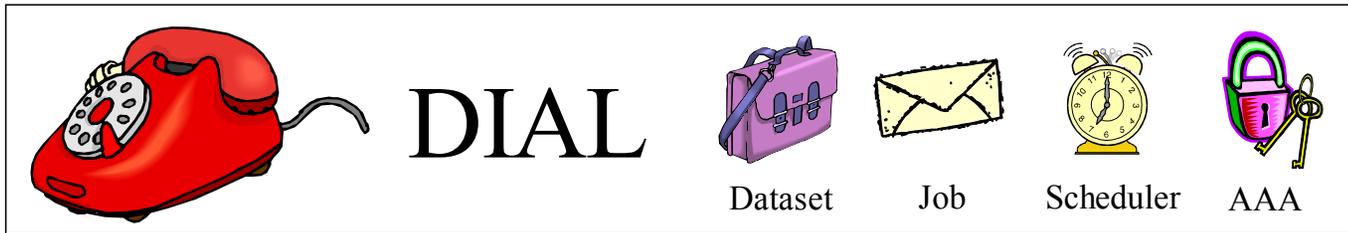
ATLAS SW - Grid

September 25, 2003

4



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



Distributed processing running data-specific application



Design

DIAL has the following major components

- **Dataset** describing the data of interest
- **Application** defined by experiment/site
- **Task** is user extension to the application
- **Job** uses application and task to process a dataset
- **Result** is the output of a job
- **Scheduler** creates and manages jobs

Together these define a high-level JDL

- (job definition language)

Figure shows how these components interact →



David Adams

BROOKHAVEN
NATIONAL LABORATORY

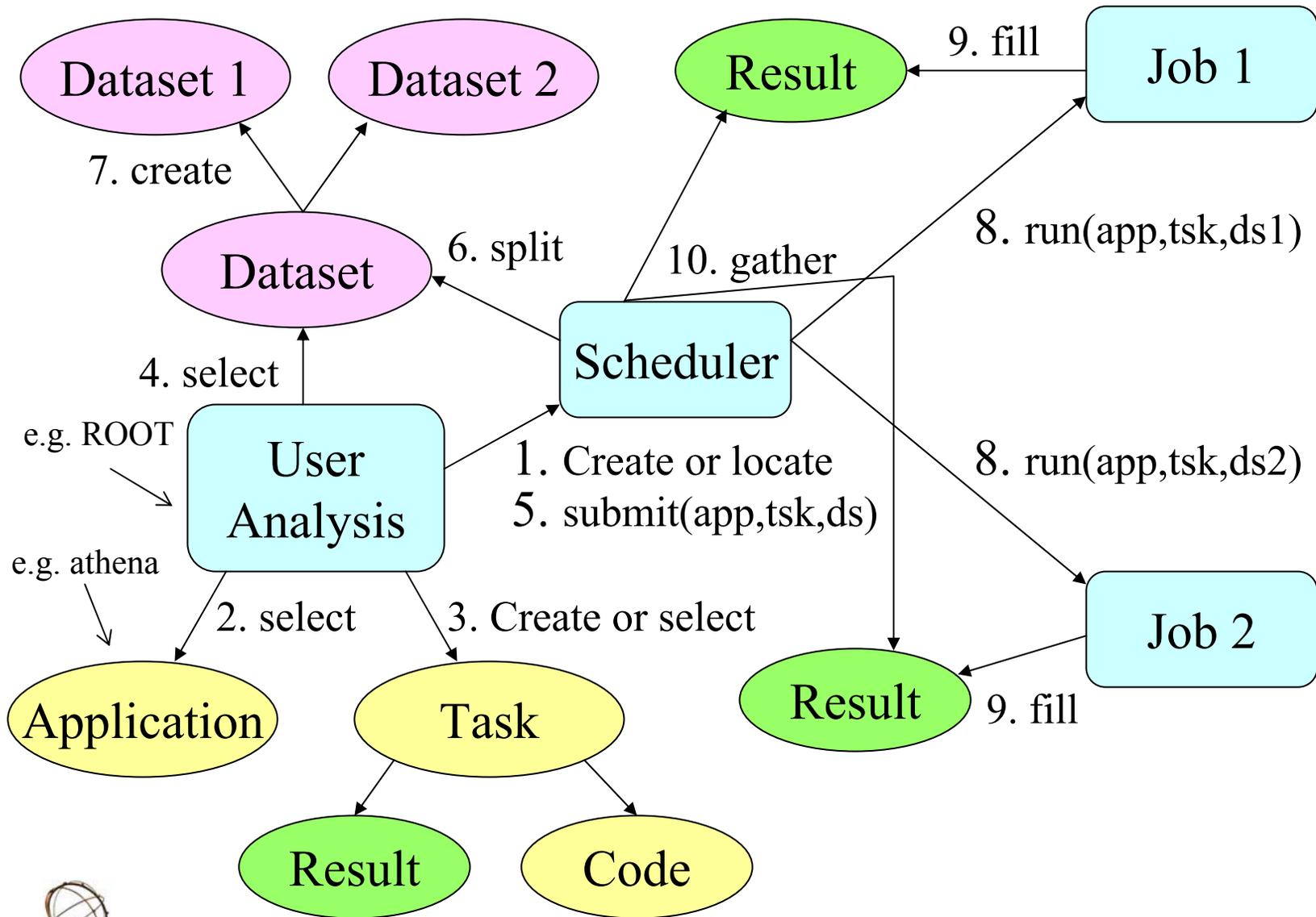


DIAL

ATLAS SW - Grid

September 25, 2003

6



Status

DIAL release 0.5 last week

User interfaces

- ROOT
 - Uses ACLiC to make all dataset/DIAL classes accessible
- Command line
 - Command dial_submit processes one job
 - Application, task and dataset taken from XML files

Schedulers

- LocalScheduler: single local job
- MasterScheduler: distributed processing at a site
 - Fork, LSF, lsrn or Condor
 - Run at both BNL and CERN



David Adams

BROOKHAVEN
NATIONAL LABORATORY



ATLAS SW - Grid

September 25, 2003

Status (cont)

Datasets

- CbntDataset describes a single CBNT hbook file
 - Inherits from EventDataset
- EventMergeDataset holds multiple EventDataset
- Together these can be used build dataset which corresponds to a collection of CBNT hbook files
- A few files with XML descriptions of such datasets are available for demonstration

Results

- HbookResult holds an hbook file assumed to hold a collection of histograms
- Merge method uses PAW to add histograms



David Adams

BROOKHAVEN
NATIONAL LABORATORY



DIAL

ATLAS SW - Grid

September 25, 2003

9

Status (cont)

Application

- dial_cbnt uses PAW to fill histograms using a user-supplied fortran function

Task

- For use with the above, task holds hbook file with empty histograms and fortran code to fill them

➔ Possible to do distributed processing of a collection of CBNT hbook files

See dial root demo 2



Development plans

Add dataset catalog

- Enable users to select dataset
- Fill with DC1 datasets
- See talk at DB meeting

Grid enable

- Condor-G
 - Scaling, reliability, response time issues
- Define DIAL web service
 - Client scheduler to connect from user interface
 - User could submit jobs to BNL or CERN from anywhere
- DIAL grid scheduler
 - Use above DIAL service and/or “conventional” grid WMS



David Adams

BROOKHAVEN
NATIONAL LABORATORY



DIAL

ATLAS SW - Grid

September 25, 2003 11

Development plans

Python user interface

- Python busses can be used as user interface
 - GANGA, ATLAS interactive, PI, ...

ROOT CBNT application

- Similar to HBOOK version

Athena application

- Enable DIAL users run distributed athena
- Initial application to fill histograms
- Later add support to create new dataset (production)
- Need athena-compatible dataset
 - ATLAS POOL event collection
 - Zebra, Athena-ROOT obsolete?



David Adams

BROOKHAVEN
NATIONAL LABORATORY



DIAL

ATLAS SW - Grid

September 25, 2003 12

Integration with GANGA

We identified the following goals and tasks for DIAL/GANGA integration:

- Provide dataset access from GANGA
 - Add DS catalog to DIAL
 - Import dataset classes to Python
- Enable DIAL job submission (CBNT datasets) from GANGA
 - Modify GANGA job interface to use datasets
 - Import DIAL classes into Python
 - Provide GANGA GUI interface to DIAL scheduler

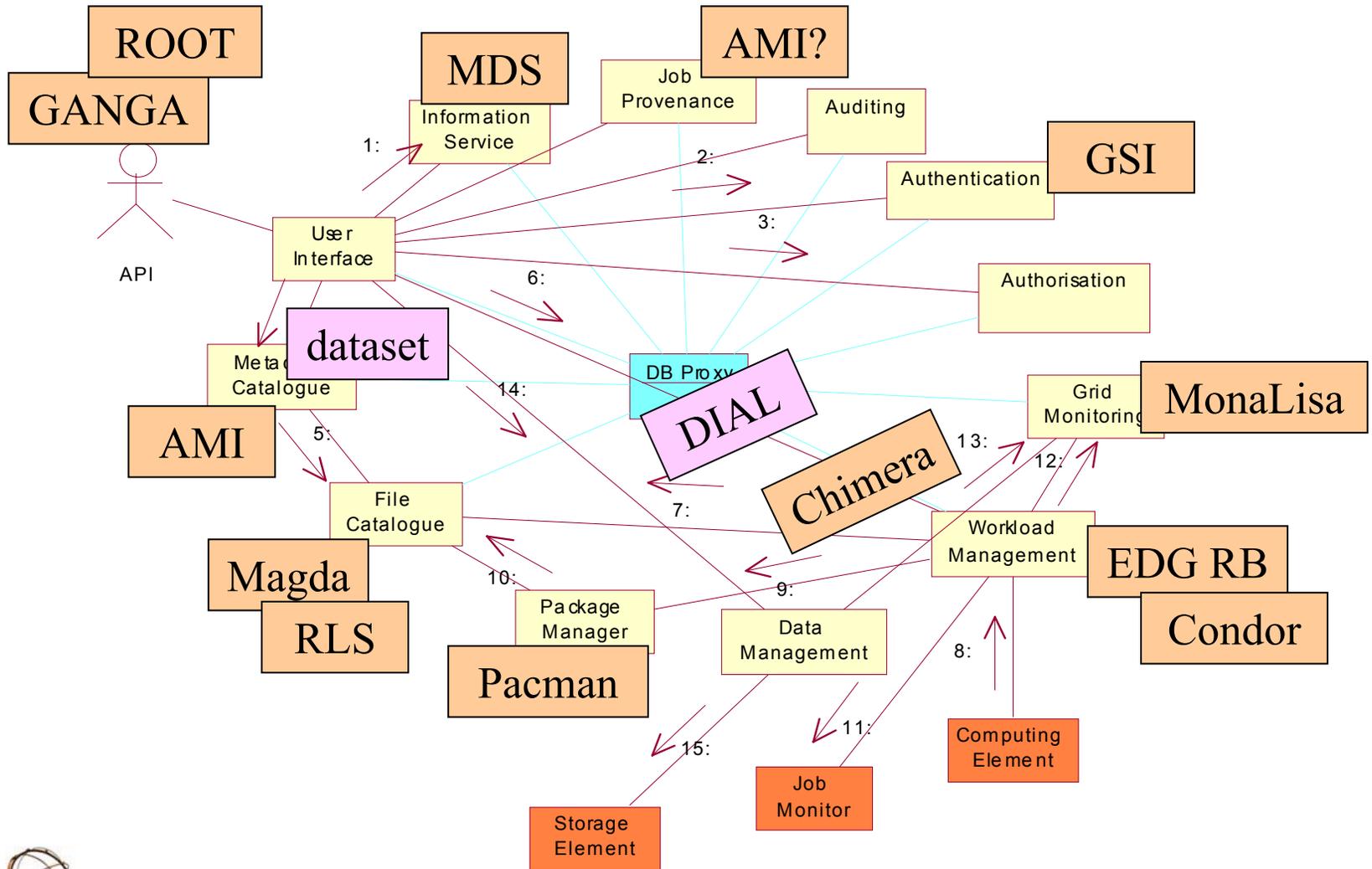


Integration with GANGA (cont)

- Enable GANGA job monitor to track DIAL jobs
 - Define common XML job description
 - Incorporate in GANGA and DIAL
- Enable GANGA to use DIAL for distributed production with athena
 - Add athena application support to DIAL
 - Add GANGA GUI for this job submission



What about ARDA?



What about ARDA? (cont)

Architectural Roadmap towards Distributed Analysis

- LCG RTAG

DIAL dataset

- More general than HEPCAL/ARDA dataset
- Datasets and their catalogs are part of the metadata and data management services identified by ARDA

DIAL components

- Define language by which user interacts with workload management service
- May also intrude into this service
 - At what point in the grid do we turn DIAL datasets into files?
 - How do we get interactive response?



David Adams

BROOKHAVEN
NATIONAL LABORATORY



DIAL

ATLAS SW - Grid

September 25, 2003 16