DIAL status

ATLAS software meeting

GRID session

David Adams

BNL

November 21, 2002

Goals of DIAL

Demonstrate the feasibility of interactive analysis of large datasets

- Large means too big for interactive analysis on a single CPU

Set requirements for GRID middleware

Provide ATLAS with a tool to analyze DC1 and DC2 event data

- More that just ntuples
- Large samples
- Distributed data and processing

What is DIAL?

Distributed

- Data and processing

Interactive

- Prompt response (seconds rather than hours)

Analysis of

- Histograms, event selection, …

Large datasets

- Any event data (not just ntuples or tag)

What is DIAL? (cont)

DIAL provides a connection between

- Interactive analysis framework
  - ROOT
  - Afrodita?
- Data processing application
  - Athena

User supplies task

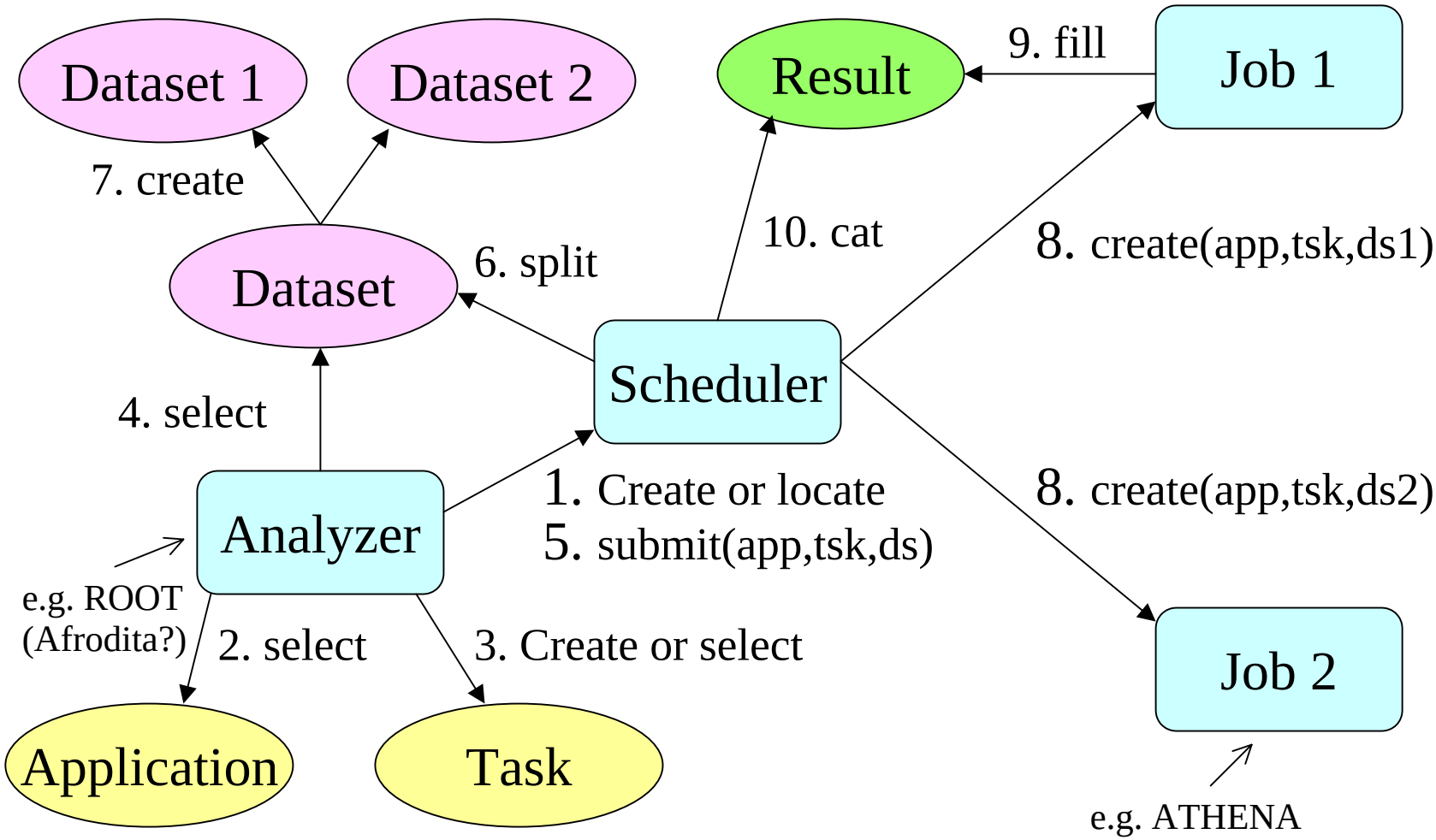
- Defines result
  - E.g. histogram
- C++ code snippet to fill result

What is DIAL? (cont)

Scheduler

- Accepts dataset, task and application from user
- Splits dataset along event boundaries
- Creates and submits a job for each sub-dataset
- Concatenates results from jobs
- Makes combined result available to the user
- Provides status reports
  - Fraction of events processed
  - Estimated time to completion
  - Partial results

DIAL interactions



Status

Datasets

- Generic implementation in place
  - http://www.usatlas.bnl.gov/~dladams/dataset
- Includes AthenaRoot MC track and vertex

DIAL

- All components in place
  - http://www.usatlas.bnl.gov/~dladams/dial
- Simple scheduler
  - local (same node)
  - creates a single processing job

Future

Dataset

- Zebra hits and digits (DC1-1)
- AthenaRoot reco objects (DC1-2)
- POOL EventCollection all objects (DC2)

Scheduler

- Remote processing
- Multiple jobs (splitting input dataset)
- GRID integration
  - Big deal