

ADA User Scenario

ATLAS Physics Analysis Tools Meeting

David Adams
BNL
February 11, 2004



David Adams

BROOKHAVEN
NATIONAL LABORATORY



Contents

ADA model

Application

Task

Dataset

Scenario for the first release

Example

Comments on example

References



David Adams

BROOKHAVEN
NATIONAL LABORATORY



ADA scenario

APAT meeting

February 11, 2004 2

ADA model

ADA = ATLAS Distributed Analysis

Users interact with distributed (or non-distributed) processing system using a job-based model

- AJDL (Analysis Job Description Language)
- User input
 - Application, e.g. athena
 - Task, e.g. algorithm to fill histograms from AOD
 - Dataset, e.g. AOD data
 - Configuration (response time, where to place data, ...)
- Output
 - New dataset, e.g. collection of histograms



David Adams

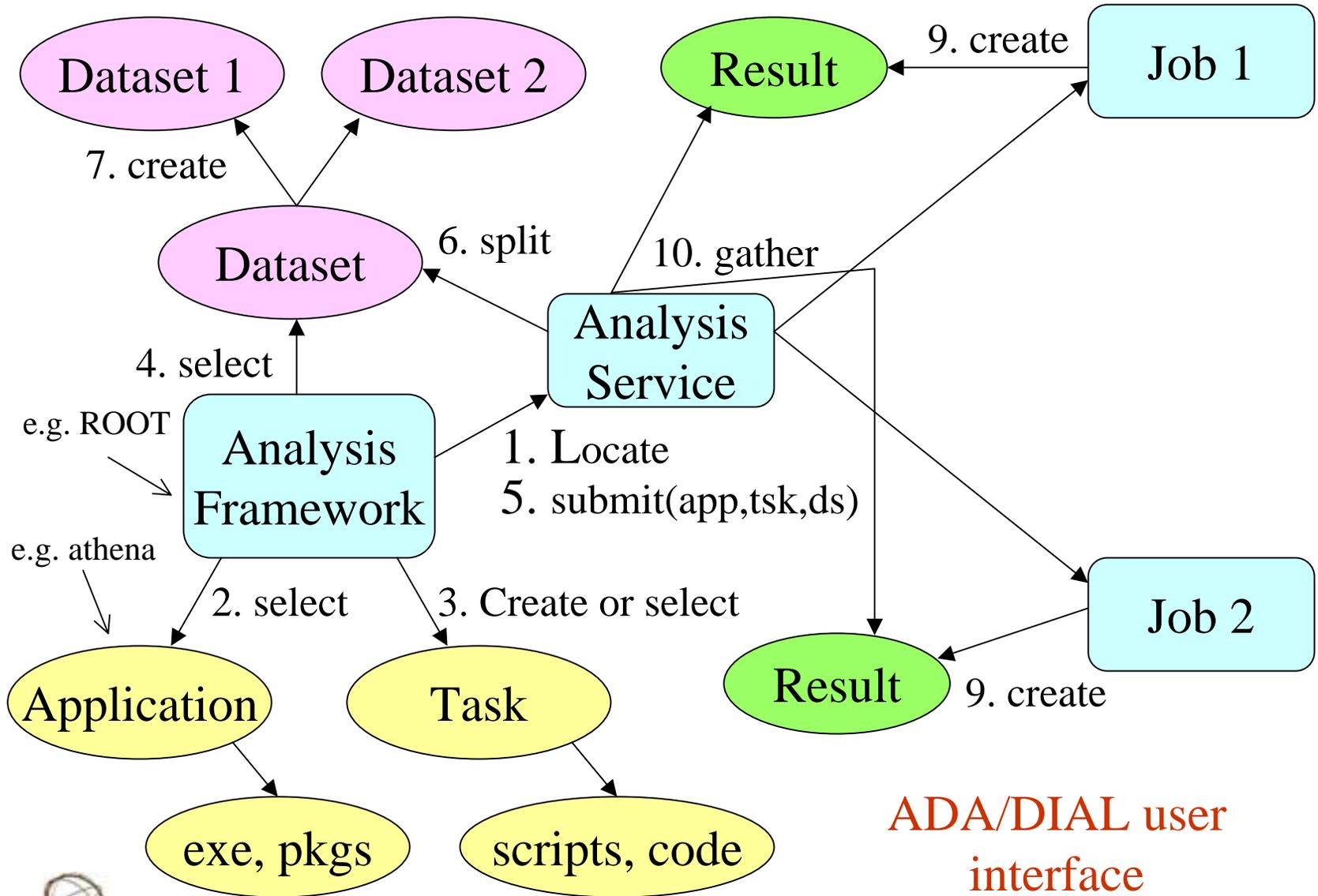
BROOKHAVEN
NATIONAL LABORATORY



ADA scenario

APAT meeting

February 11, 2004 3



David Adams



ADA scenario

APAT meeting

February 11, 2004 4

Application

Normally defined in advance by an expert

- User selects from menu of existing applications

Application provides entry points for

- Building a task
 - E.g., compiling algorithm in our example
- Processing a dataset using task
 - E.g., running athena and calling algorithm for each event
 - Result is another dataset

Application does *local* processing

- Distributed by calling application in multiple places

Application likely installed as a software package

- Or pre-installed by hand



David Adams

BROOKHAVEN
NATIONAL LABORATORY



ADA scenario

APAT meeting

February 11, 2004 5

Task

Used to configure the application

- Interpreted by the application
- Allows user to specify processing without creating a new SW package

Currently a collection of text files

- Copied to remote site
 - E.g. embedded in SOAP message
- Application used to build (e.g. compile) task at remote sites



David Adams

BROOKHAVEN
NATIONAL LABORATORY



ADA scenario

APAT meeting

February 11, 2004 6

Dataset

Used to describe input and output data

Properties

- Location, e.g. which files hold the data
- Content, e.g. AOD or histograms
- Compositeness: constituent datasets
- And more... (see “Datasets for the Grid”)

Virtual dataset has no location

- Replicas are cataloged
 - System can choose the best replica for a given job/site
 - System may generate data on demand
- Typical user to specifies virtual dataset by name
 - Catalogs aid users in selection



David Adams

BROOKHAVEN
NATIONAL LABORATORY



ADA scenario

APAT meeting

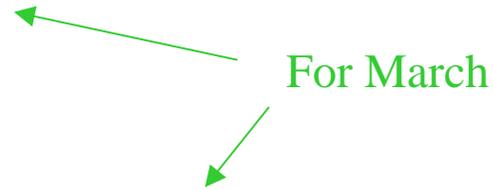
February 11, 2004 7

Scenario for the first release

Here is a scenario for user interaction with the first release of ADA

- Authenticate
 - Proxy from authentication service
- Choose application
 - E.g. PAW to process DC1 ntuples or
 - Athena to process DC2 AOD or
 - Athena reconstruction
- Define task
 - Analysis: provide code to define and fill histograms
 - Production: athena job options, maybe code
 - Perhaps select starting point from task catalog
- Select input dataset
 - From dataset (metadata) catalog service

For March



David Adams

BROOKHAVEN
NATIONAL LABORATORY



ADA scenario

APAT meeting

February 11, 2004 8

Scenario for the first release (cont)

- Create job configuration
 - Response time, role, optional splitter,...
- Locate processing service
- Submit job
 - Application, task, dataset, configuration
- While job is running
 - Query service for status and partial results
 - Examine partial results (e.g. histograms)
 - Kill job if results are bad
- When job is finished
 - Examine complete result
 - Modify task or select new dataset and repeat
- *Interactive* if job takes seconds or minutes



David Adams

BROOKHAVEN
NATIONAL LABORATORY



ADA scenario

APAT meeting

February 11, 2004 9

Example

```
mynode> grid-proxy-init Authenticate  
Your identity: /DC=org/DC=SomeOrg/OU=People/CN=Joe Physics 1234  
Enter GRID pass phrase for this identity: *****  
Creating proxy.....Done  
Your proxy is valid until: Fri Feb 13 13:00:00 2004  
mynode> root.exe  
*****  
* * Choose analysis service  
* WELCOME TO ROOT *  
... Choose application  
root[0] GridScheduler sch("http://www.cern.ch/atlas/analysis_service");  
root[1] Application app("athena_histomaker", "0.40");  
root[2] Task tsk;  
root[3] .sh emacs algo.cxx Define task  
root[4] tsk.insert("algo.cxx");
```



David Adams

BROOKHAVEN
NATIONAL LABORATORY



ADA scenario

APAT meeting

February 11, 2004 10

Example (cont)

```
root[5] DatasetSelectionCatalog dsc("http://ami.fr/atals/dsc");
```

```
root[6] DatasetId did = dsc.id("multi_lepton_stream");
```

```
root[7] DatasetDb ddb("http://ami.fr/atlas/dsc");
```

```
root[8] Dataset* pdst = ddb.extract(did);
```

```
root[9] cout << *pdst << endl;
```

VirtualEventDataset 10003-4447 with parent 10003-4338 is locked

Dataset has 32247893 events

Content: RAW, ESD, AOD, TAG

```
root[10] JobConfiguration cfg;
```

```
root[11] cfg.set("response_time", "300");
```

```
root[12] cfg.set("output_site", "nikhef");
```

```
root[13] JobId jid = sch.submit(app, tsk, dst, cfg);
```

```
root[14] cout << jid << endl;
```

422-135592

Select virtual
dataset by name

Job configuration

Submit job



David Adams

BROOKHAVEN
NATIONAL LABORATORY



ADA scenario

APAT meeting

February 11, 2004 11

Example (cont)

```
root[15] cout << sch.job(jid) << endl;
```

Display job status

Job is running

athena_histomaker 0.30

Task 101-11776

Dataset 10003-4447 with 32247893 events

create time: 2004 February 11 15:35:22

start time: 2004 February 11 15:35:29 (7 sec elapsed)

update time: 2004 February 11 15:36:24 (62 sec elapsed)

Events processed: 7844262 (24%)

Run directory: /cern/home/dial/jobs/master/rgdy834ne_uww3

There 1128 subjobs

345 running

277 done

0 failed

0 killed

253 included in result



David Adams

BROOKHAVEN
NATIONAL LABORATORY



ADA scenario

APAT meeting

February 11, 2004 12

Example (cont)

```
root[16] Dataset& newdst = sch.job(jid).result();
```

Fetch output

```
root[17] cout << newdst << endl;
```

RootHistogramDataset 10019-678173 with parent 10003-4447 is locked

Dataset has 77 histograms

```
root[18] RootHistogramDataset& rdst = newdst.cast<RootHistogramDataset>();
```

```
root[19] rdst.get_histograms();
```

```
root[20] electron_1_pt.Draw();
```

Display histograms

```
root[21] electron_2_pt.Draw();
```

```
root[22] electron_3_pt.Draw();
```

```
root[23] electron_4_pt.Draw();
```

```
root[24] four_electron_mass.Draw();
```



David Adams

BROOKHAVEN
NATIONAL LABORATORY



ADA scenario

APAT meeting

February 11, 2004 13

Comments on example

Previous is approximate

- Gives look and feel of current system
- Parts still under development

Similar syntax in Python and maybe command line

GANGA will add nicer user interface

- Menu driven
- Graphical job monitoring



David Adams

BROOKHAVEN
NATIONAL LABORATORY



ADA scenario

APAT meeting

February 11, 2004 14

References

ADA (ATLAS Distribute Analysis) is described at

- <http://www.usatlas.bnl.gov/ADA>

Documents link on that page includes

- “Catalog services for ATLAS”, version 0.10
- “AJDL: Analysis Job Description Language”, version 0.20
- “Datasets for the Grid”, version 5



David Adams

BROOKHAVEN
NATIONAL LABORATORY



ADA scenario

APAT meeting

February 11, 2004 15