



SLAC

Interactive Dataset Analysis on the Grid

David A. Alexander & Bala Ananthan (Tech-X Corporation),
Tony Johnson & Victor Serbo (SLAC)

American Linear Collider Physics Group
Detector Simulation Workshop

Boulder, January 9-11, 2006

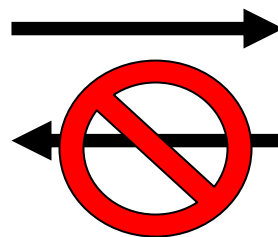
D.O.E supported research

- *Small Business Innovative Research* Grant from US Department of Energy Office of High Energy Physics
- Collaboration between *Tech-X Corporation* in Boulder, CO and *Stanford Linear Accelerator Center*
- “Interactive Dataset Analysis” on the Grid. Unique compared to common “Batch Processing”.

Features

- Desktop tools for **High Energy Physics** processing that can analyze large-scale datasets in **parallel**
- Grid service includes functions for **staging and handling** of remote datasets and analysis code
- Rich **graphical** client provides custom analysis coding capabilities and results visualization at **interactive speeds**

Analysis Code



DATA

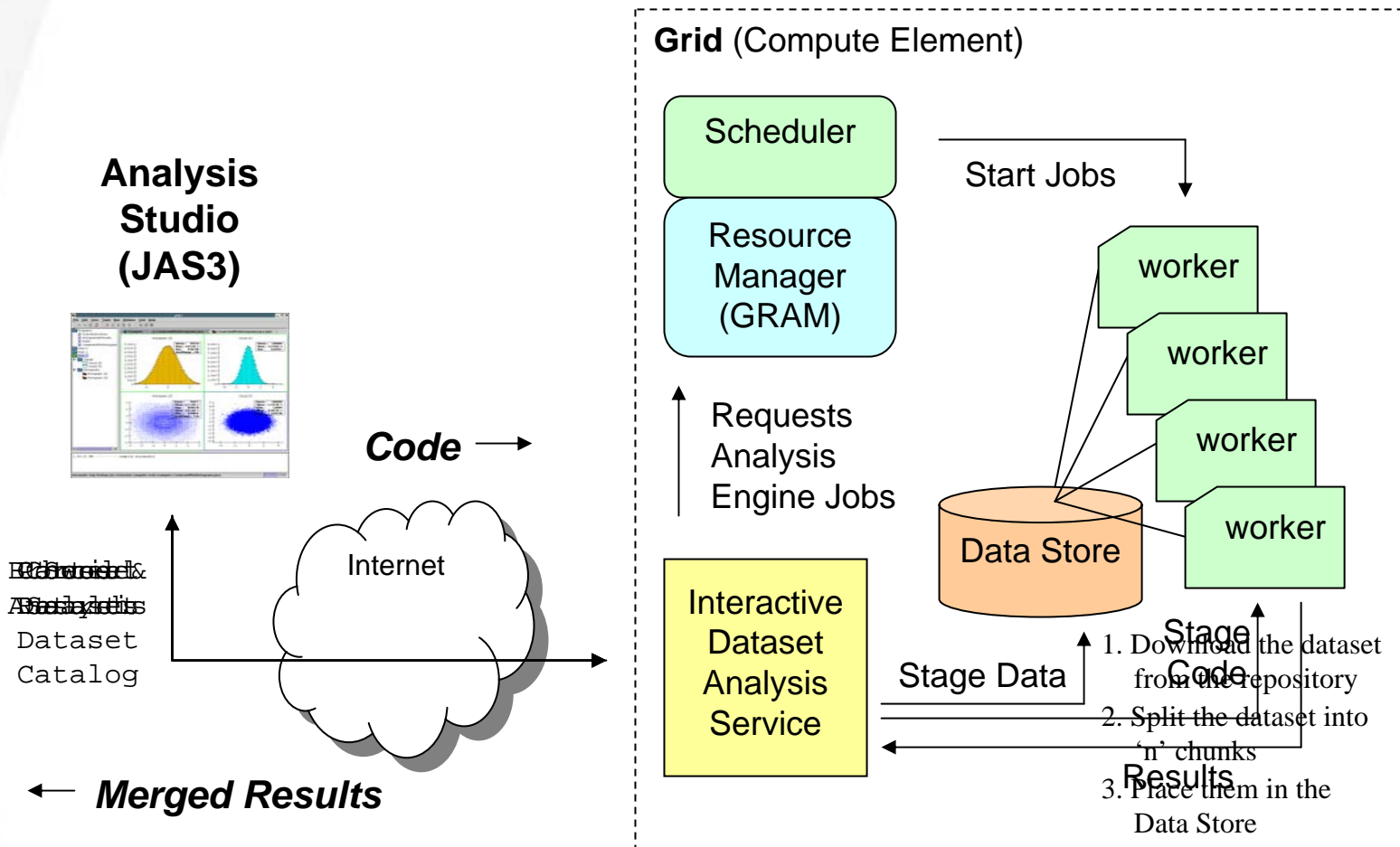
Features (cont...)

- Can analyze **any type** of dataset - as long as Datasets are **Events Oriented** (Records based).
 - .stdhep or .slcio or .root ...

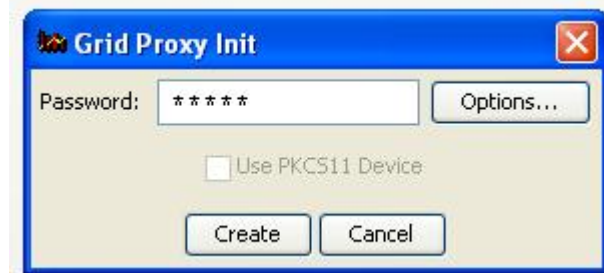
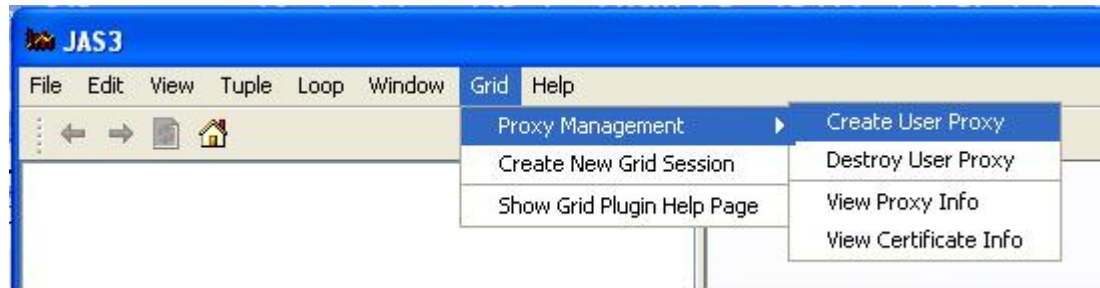
How to Use It?

- Just use JAS normally
- Install DAGS plugin
- Select and stage any Dataset (that are Events based)
- Analyze it with exactly the same Analysis Code that you would use to analyze the Dataset locally.

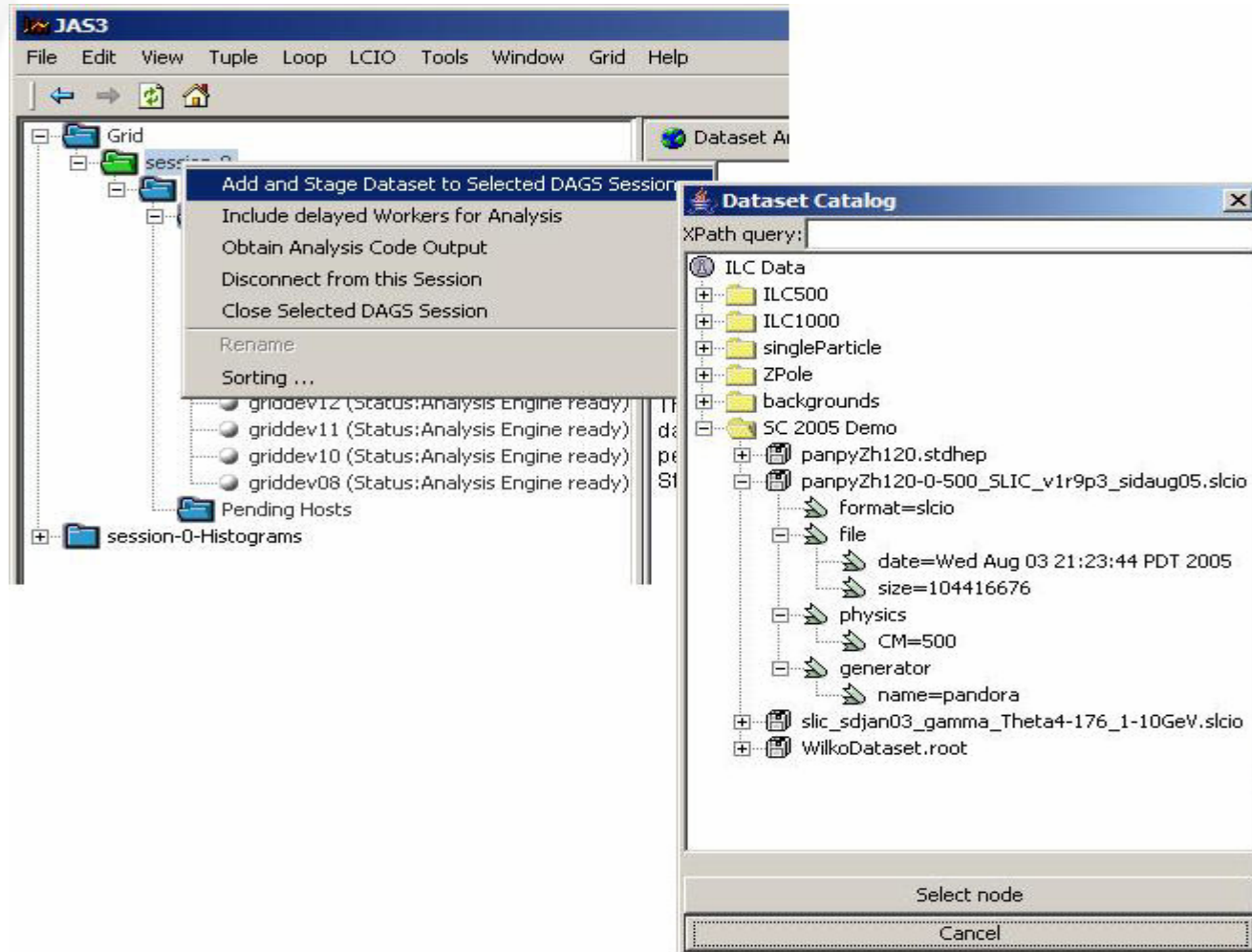
Sending Code and Merging Results



Secure Login



Catalog Service to Browse and Search Datasets



Integrated Development Environment

The screenshot shows the JAS3 IDE interface. On the left, a file explorer displays a tree structure for a Grid session named 'session-0'. Under 'Host Information', there is a list of 'Included Hosts' (griddev07 through griddev08) and 'Pending Hosts'. A file named 'panpyZh120-0-500_SLIC_v1r9p3_sidaug05.slcio' is selected. The main window shows a code editor with the following Java code:

```

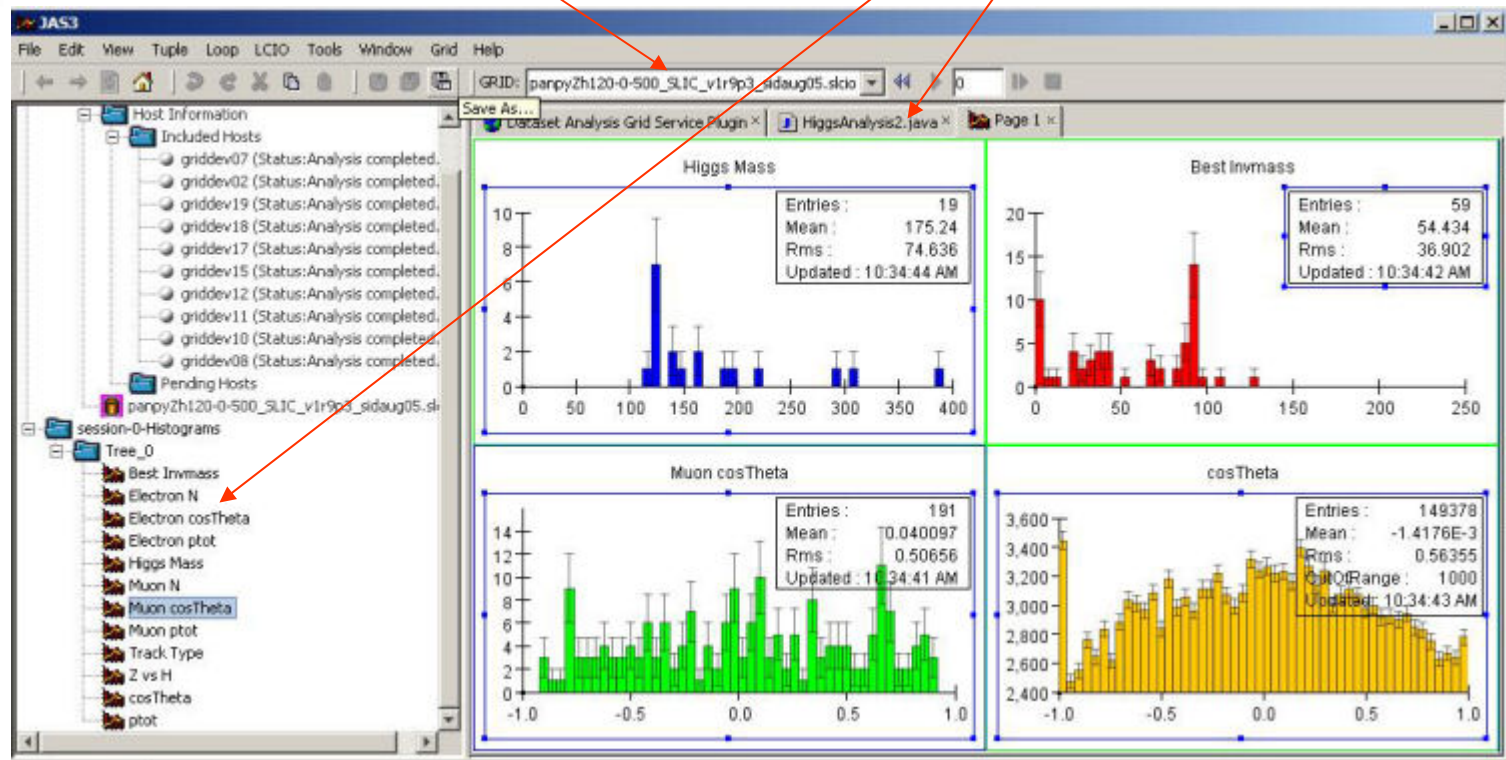
1 import hep.aida.IAnalysisFactory;
2 import hep.aida.IHistogram1D;
3 import hep.aida.IHistogram2D;
4 import hep.aida.IHistogramFactory;
5 import hep.physics.vec.VecOp;
6 import java.util.ArrayList;
7 import java.util.List;
8 import org.lcsim.event.EventHeader;
9 import org.lcsim.event.MCParticle;
10 import org.lcsim.util.Driver;
11
12 public class HiggsAnalysis2 extends
13 {
14     private static double mZ0 = 91.18;
15     private static double energyCM =
16     private MCParticle z1, z2;
17     enum PType {MUON, ELECTRON}
18
19     private IHistogram1D h_ptot;
20     private IHistogram1D h_cosTheta;
21     private IHistogram1D h_muonPtc;
22     private IHistogram1D h_muonCo;
23     private IHistogram1D h_electronf;
24     private IHistogram1D h_electronCosTheta;
25     private IHistogram1D h_trackType;
26     private IHistogram1D h_muonN;
27     private IHistogram1D h_electronN;
    
```

A context menu is open over the code editor, listing actions such as Save, Refresh, Undo, Redo, Cut, Copy, Paste, Load This Class in GRID (selected), Compile, and Run. The 'Load This Class in GRID' option is pointing to the 'session-0' folder in the file explorer.

Interactive Graphical Results in Seconds

Controls to Start/Stop Analysis

Can change analysis code, clear graphs, and restart



DEMO

- Using a small set of farm machines at SLAC.
- Machines used are part of Open Science Grid

Other Grid Sites

- We are looking for Grid Sites interested in hosting interactive analysis services.
- Requirements include:
 - GRAM, GridFTP servers up and running
 - Jobs should be started quickly (relatively)
 - Few ports should be opened that are not blocked by the firewall.

Possible Improvements

- Support use of Clouds
- Continue to add support for more dataset formats
- Continue to add support for more code types in addition to Java and Pnut Scripting (Python scripting, etc.)

Key New Features

(summary)

- Assisted **parallel processing** on the Grid.
- Grid Service provides **secure connection**, dataset catalog, & staging functions
- Client provides **interactive controls** for the analysis session and tools to visualize the results

More Info

- General info about DAGS
- More presentations and documents
- Plugin help
- DAGS Javadocs
- Requirements to run DAGS at a Grid Site

<http://grid.txcorp.com/confluence/display/DAGS/>