Sam for Users

November 17, 1998

Outline

- How Sam would be used by a physics group to make a W ->eν sample.
  - **Setup**
  - **Concept of Project**
  - **Run**
  - **Evaluation**
- Other use patterns

SAM

User Job

- Any user job has
  - **job scripts**
  - **an executable**
  - **input file lists (input projects)**
  - **output protocols**
  - **metadata describing the job**

SAM ‘Project_W’

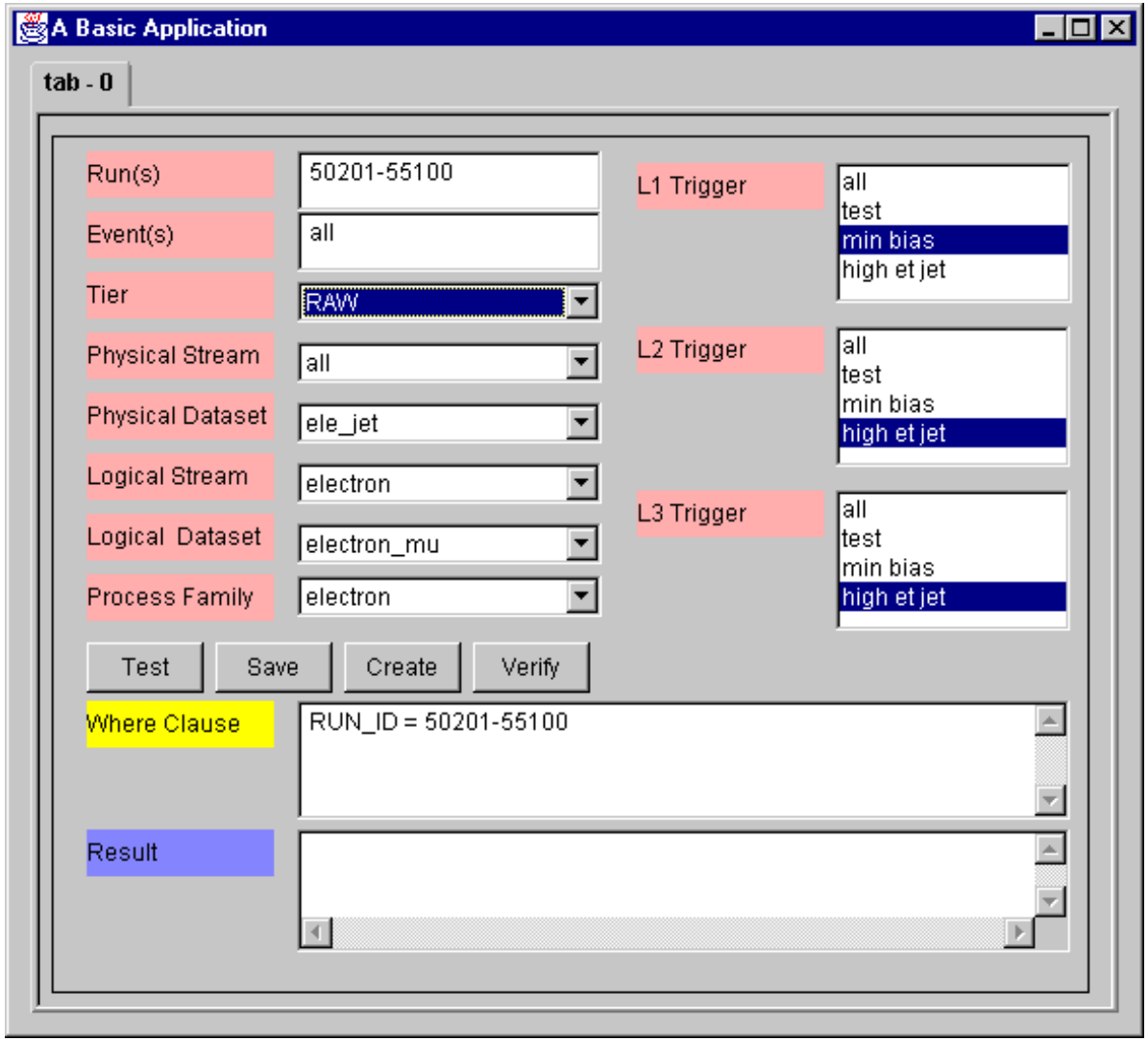
Say a group of users decide to make a W sample which requires processing 50,000 files.

- They build and debug an executable
- They then query the File Database to find which files contain the appropriate triggers.

This set of files is what we call a Project

SAM

User Query



SAM Why Project instead of Dataset?

A Project is an Object not just a list of files

- Constructors
  - **Produce a Project Definition via simple SQL queries on Run, Trigger …**
  - **Produce a Project Snapshot with real file names**
  - **Produce an Optimized Project Snapshot for fast access to tape**
- Accessors
  - **Deliver the files to the user**
  - **Receive files from the user**
- Persistency
  - **The list of files delivered can be associated with information about executables, luminosity, success in delivery - stored in a database.**

Running Project W

- The W group query the database with trigger and run information to get a Project Definition(SQL) called 'Project_W'
- They test and document their code
  - **CVS tag of executables and scripts**
  - **who is running the job**
  - **comments**
- The database is queried again to produces a Project Snapshot (current optimized list of files) 'Project W-Xmas'

Submitting the Job

- They determine what job parameters they needed to run the job:
  - **Number of files(disk and tape) - From SAM**
  - **CPU time/file - From their tests**
  - **Size of available input buffers - From SAM or local machine**
  - **Size of available output buffers - From SAM or local machine**
- They submit the job to an appropriate batch queue or run interactively on a “Station” with sufficient IO/CPU resources.
- The Project instance is further defined, with user, code and a jobid - Project_W.Xmas.joeW.Wcode.run43

Running the Job

- The Job
  - **The Project starts on the Station and delivers the first files**
  - **The executable starts and consumes input files via a Consumer**
  - **When the executable is 'done' with an input file it is released.**
  - **Output files are written back into the SAM system.**
- At the end of the job, this Instance of 'Project_W' is complete and now consists of a list of delivered (and undelivered) files. Project_W.Xmas.joeW.Wcode.run43

SAM

Doing the Physics

Six months later, the editorial board asks about the dataset

- **Helen W decides to reproduce the analysis with a new W sample making executable.**
- **She asks for Project Definition 'Project_W' based on runs and triggers.**
- **She creates a current snapshot 'Project_W.July4' and can compare the files available to those available when Joe W ran 6 months ago.**
- **She can decide to run only on those files actually used by Project_W. Xmas.joeW.Wcode.run43 or run on the newer list. Either way, any differences are understood.**

SAM

Other Access Modes

- Freight - Train
  - **Several Users agree that they need the electron triggers and start a Freight-Train Project**
  - **The Project allows several Consumers to use files and does not erase them until all registered Consumers have used them or timed out.**
- Farm/Parallel Analysis
  - **Several processes register with a Project as a single Consumer, and alternately receive files. Each file is delivered to only one process.**
- Multiple Stream
  - **A single user decides to overlay events from HERWIG with zero-bias data. This may involve 2 input Projects!**

SAM What if the system dies?

If the system dies halfway through Project W, Joe W has several options:

- The Old-Fashioned way: Start over from scratch by submitting a new job with new ID and reprocessing all files.
- Start up where he left off by asking SAM to revive Project_W.Xmas.joeW.Wcode.run43 and deliver any files not yet flagged as processed.
- If 'Project_W.Xmas' was shared with other users, and Joe W was on vacation and missed the restart,he can also rejoin the other users when he gets back from Aruba and pick up the files staged while he was gone later.

Conclusions

- Projects are an interface to SAM (Sam-lite?)
- They are largely self-documenting
- They are flexible enough handle event overlays, freight-trains and farms.
- They are flexible enough to deliver data to multiple UNIX/NT platforms in a consistent fashion.