Introduction to the SAM System at DØ

Physics 5391

July 1, 2002

Mark Sosebee

U.T. Arlington

THE UNIVERSITY OF TEXAS
AT ARLINGTON

The Requirements

❑ An experiment like DØ collects an **enormous** amount of data!

❑ Must have a system to manage (catalog) the data, store it, and retrieve it efficiently.

❑ Enter SAM...

THE UNIVERSITY OF TEXAS AT ARLINGTON

What is SAM?

- ❑ SAM stands for "Sequential data Access via Meta-data.

- ❑ It provides the interface between the DØ data and users.

- ❑ "Meta-data" is a descriptive catalog that can be queried to extract sub-sets of the full data set.

THE UNIVERSITY OF TEXAS
AT ARLINGTON

SAM "Stations"

❑ **Station Master:** Overall station manager, manages disk cache, etc.

❑ **File Storage Server (FSS):** Invoked if storing files into the SAM system.

❑ **Stager(s):** Interface for communications with the various cache disks.

❑ **bbftp:** Transport protocol for moving data.

THE UNIVERSITY OF TEXAS
AT ARLINGTON

Glue That Holds Everything Together

❑ **Oracle databases:**

  ❑ nameservice

  ❑ logger

  ❑ dbserver

❑ As we'll see a bit later SAM queries ("project creation") are really Oracle db queries.

❑ Currently the db services are provided through servers at FNAL.

THE UNIVERSITY OF TEXAS
AT ARLINGTON

SAM Stations:

Server	Host	Version	Up Since
central-analysis	d0mino.fnal.gov	v4_1_1_4	26 Jun 2002 14:39:42
fnal-farm	d0bbin.fnal.gov	v4_2_0_3p1	29 Jun 2002 09:38:03
azsam1			
budzero			
cab	d0mino-sam.fnal.gov	v4_1_1_4	26 Jun 2002 14:40:14
caps10	caps10.phys.LaTech.edu	v4_1_0_9	28 Jun 2002 18:12:42
ccin2p3-analysis	ccd0.in2p3.fr	v4_1_0_0	29 Jun 2002 10:31:19
ccin2p3-production	unknown host	unknown	unknown
cdf-ucl	not registered		
central-archive	not registered		
central-compute	not registered		
chrisprd	not registered		
clued0	not registered		
d0-demo-station	not registered		
d0-test-station	not registered		
d02ka	not registered		
d0_main_analysis	not registered		
d0karlsruhe	d0.fzk.de	v4_0_0_3	11 Jun 2002 05:59:38
D0Mainz			
d0nevis-station	karthur.nevis.columbia.edu	v4_0_0_3	04 Apr 2002 13:40:07
d0small-01	not registered		
datalogger	not registered		
datalogger-d0ola	unknown host	unknown	unknown
datalogger-d0olb	d0olb-gb-1.fnal.gov	v4_1_0_0	28 Jun 2002 11:20:55
datalogger-d0olc	d0olc-gb-1.fnal.gov	v4_1_0_0	06 Jun 2002 09:17:42
droid	not registered		
droide	not registered		
hoeve	not registered		
imperial-test			
IndianaUniversity	d0svr2.physics.indiana.edu	v4_0_0_3	25 May 2002 16:15:17
lancs	fal000phys03.lancs.ac.uk	v4_1_0_9	11 Jun 2002 09:06:19
linux-analysis-cluster-1	d0lxac1.fnal.gov	v4_1_0_0	04 Jun 2002 14:57:47
luhep	not registered		
nsu	not registered		
munich			
nijmegen	doesburg.hef.kun.nl	v4_2_0_3	22 Jun 2002 14:09:20
ouhep	ouhep1.nhn.ou.edu	v4_1_0_9	27 Jun 2002 13:48:25
pctestfarm	not registered		
prague-test-station			
princeton-d0			
protofarm	not registered		
sam_kud0	not registered		
triviaal	triviaal.nikhef.nl	v4_0_0_3	27 Jun 2002 08:22:13
umdzero			
uta-analysis	hep.uta.edu	v3_2_18	26 Jun 2002 12:55:30
uta-d0-grid			
uta-hep	hepfm007.uta.edu	v4_1_0_0	28 Jun 2002 10:45:57
wuppertal	d0w.physik.uni-wuppertal.de	v4_0_0_3	27 May 2002 15:07:01

THE UNIVERSITY OF TEXAS AT ARLINGTON

Interested in More
SAM System Details??

❑ Large amount of documentation on the web:

    http://d0db.fnal.gov/sam/documents.html

    (see the section "Station Administration...")

❑ LISTSERV:

    SAM-ADMIN, SAM-DESIGN

❑ The SAM Team: Lee Lueking, Igor Terekhov, et al.

THE UNIVERSITY OF TEXAS
AT ARLINGTON

User Preliminaries

- ❑ Accounts on d0mino, clued0, (& UTA systems):

    http://www-d0.fnal.gov/computing/systems/comp_acc_sec.html

- ❑ Register as a SAM user:

    http://d0db.fnal.gov/sam_admin/cgi/autoRegister.py

- ❑ The usual directives (learn C++, Python, etc.)

THE UNIVERSITY OF TEXAS
AT ARLINGTON

How To Proceed

❑ Users create "Datasets" which define the scope of their searches.

❑ Two options:

   ❑ Web Interface

   ❑ Command Line

❑ We'll look at an example of each type.

THE UNIVERSITY OF TEXAS
AT ARLINGTON

Web-based Dataset Creation

❑ Start off by reviewing this URL:

http://d0db.fnal.gov/sam/doc/userdocs/creating_sam_projects.html

❑ Go to:

http://d0db.fnal.gov/sam_project_editor/DatasetEditor.html

❑ "Translate Constraints" tests the query

❑ "Create Dataset" is for real.

THE UNIVERSITY OF TEXAS
AT ARLINGTON

Command Line Dataset Creation

❑ Same idea, just typing in constraints by hand.
❑ Example:
  ➢ setup D0RunII *version*
  ➢ setup sam
  ➢ sam translate constraints –filename="%mcp08%UTA%" --datatier="reconstructed"

❑ If we liked this query (filter):
  ➢ sam define dataset –group=dzero –defname=uta-5391_demo \
  --filename="%mcp08%UTA%" –datatier=reconstructed"

❑ If successful a dataset would now exist we could use.

Mark Sosebee

THE UNIVERSITY OF TEXAS AT ARLINGTON

Use How??
(see http://d0db.fnal.gov/sam/doc/userdocs/running_sam_user_tips.html

- One possibility: incorporate into "your_package."
- The idea is to make a package "sam-aware" as part of the usual DØ software incantations:
  - newrel –t *version directory*
  - cd *directory*
  - addpkg *your_package*
- Add sam hooks to objects, libs, rcp files, etc.
- Build as usual:
  - d0setwa
  - gmake *your_package*.lib
  - gmake *your_package*.bin
- Create a script in *your_package*/bin which includes the dataset name

THE UNIVERSITY OF TEXAS
AT ARLINGTON

Conclusion

❑ SAM is the system DØ employs for data storage & retrieval.

❑ User interfaces are provided to allow for the creation of datasets (i.e., queried data samples).

❑ Not discussed here, but the SAM team is working on interfaces to "grid" technologies.

❑ Next: the "hands-on" session!

THE UNIVERSITY OF TEXAS
AT ARLINGTON