User (My) experiences at the SLAC Western Tier 2

ATLAS Western Tier 2 Users Forum

David W. Miller

Stanford University
SLAC National Accelerator Laboratory

April 7, 2009

SLAC NATIONAL ACCELERATOR LABORATORY

Outline

1. *Introduction*
2. *Local computing and resources*
3. *GRID access and infrastructure*
4. *File transfer and registration*
5. *Performing analysis and looking ahead to first data*

SLAC NATIONAL ACCELERATOR LABORATORY

Outline

1 *Introduction*

2 *Local computing and resources*

3 *GRID access and infrastructure*

4 *File transfer and registration*

5 *Performing analysis and looking ahead to first data*

SLAC NATIONAL ACCELERATOR LABORATORY

What does it mean to be a "User"?

- In general, we as users are working within a computing and communication infrastructure which is designed to facilitate *many, many* different applications and goals. Sometimes it is useful to work backwards from our own particular usage of that system to understand how we are actually **using** it.
- I am going to try to lay out my usage of the system with both some general comments about what I have been using it for as well as some **very** specific examples.
- There are likely aspects of a real analysis that I am not covering, but these are typically the "User" aspects of what I have done with the SLAC computing and WT2 resources

Let me say right now that none of what I have done would have been possible without Wei, Peter, Stephen Gowdy, Neal, Su Dong, Andrew H. and many others...thank you! (And sorry for bugging some of you so much along the way).

The "User" part of "User Experiences"

- In the end, I want **plots**
- This means that ntuples are going to be produced in the last step for which being a "User" has any relevance
- However, there is much, much more behind those results than just a few AOD analysis jobs
  - GRID access and disk space for retrieving samples for local analysis or downloading analysis esults from the GRID
  - Good local interactive computing for testing analyses before GRID submission or running on locally produced data
  - All the necessary "stuff" for detailed ESD analyses (conditions database access, up-to-date geometry description, etc)
  - Private production (possibly the full chain) of samples for systematics or dedicated studies not supported/produced centrally

To be a "User" means that at some point I may need to execute any of the above steps on my way to a final analysis result.

SLAC NATIONAL ACCELERATOR LABORATORY

The "Experiences" part of "User Experiences"

Just to summarize before we get into the gritty details

- Batch jobs
  - Your average simulation / digitization / reconstruction job requires too much memory and CPU to be run interactively (expect maybe for just a couple of events)
  - Need to setup a batch submission script(s) that sets up the ATLAS environment, finds the right files (and probably copies them to the nodes), runs the job, moves output around, etc
  - This was not exactly easy to setup and everyone does it differently
  - Difficult for new people to jump right into
- XROOTD
  - XROOTD mostly served as static storage space from which I copy things in and out.
  - This is true except for those cases where XROOTD is mounted with **`XrootdFS`**) (but that certainly is not the case with batch nodes, so get familiar with **`xrdcp`**)
- GRID tools and usage
  - I have become ever more dependent on the GRID for running analysis
  - Use it like batch system that takes care of all the job submission and data-file manipulation for you (no horribly written batch scripts for Neal and Wei to help you debug!)
  - Only works if data on GRID, of course

SLAC NATIONAL ACCELERATOR LABORATORY

Outline

1 Introduction

2 Local computing and resources

3 GRID access and infrastructure

4 File transfer and registration

5 Performing analysis and looking ahead to first data

SLAC NATIONAL ACCELERATOR LABORATORY

What local computing resources do I regularly use?

...and how do I actually use them?

SLAC Batch

Neal already presented nearly all of the details, and I received a lot of help directly from him in setting up my own batch submission jobs. A few comments on how I use batch

- Top level script to split the full analysis/production into many batch jobs
- Calls submission script that executes **`bsub`**
- Actually sends the job script **`myjob.sh`** to the batch, which sets up ATHENA, copies input files to the batch node scratch space, calls the right job options (i.e., `SkipEvent=50/NJob`), etc

Also, specific to ATHENA and `dq2` (which has ***tons*** of data on XROOTD):

- ATHENA typically uses a lot of memory (especially if you're running pile-up!!)
- **`bsub -R "mem > 2000" -q xlong myjob.sh`**
- Can make exquisite use of **`xrdcp`** then, instead of **`cp`** to copy files in or out of the batch node scratch space (see Neal's sample script or an example batch script of my own (link))

What local computing resources do I regularly use?

...and how do I actually use them?

XROOTD

I make extensive use of XROOTD (maybe too much?)

```
[atlint01] Tue Apr 07 @ 06:59:20:  ~ > ls /xrootd/atlas/usr/f/fizisist/
MYNTUP                                                  pile0sf00.lumioff.misal1_csc11.005011.J2_pythia_jetjet
WbbNp0.500evt.0skip.trigger_test.ESD.pool.root          pile0sf00.lumioff.misal1_csc11.005013.J4_pythia_jetjet
WbbNp0.50evt.123skip.trigger_test.ESD.pool.root         pile0sf00.lumioff.misal1_csc11.005014.J5_pythia_jetjet
WbbNp0_AOD_2K.pool.root                                 pile1sf01.low.misal1_mc12.005200.T1_McAtNlo_Jimmy
WbbNp0_ESD_2K.pool.root                                 pile1sf01.verylow.misal1_csc11.005001.pythia_minbias
dq2_logs                                                pile1sf01.verylow.misal1_csc11.005009.J0_pythia_jetjet
esdoutputV5.pool.root                                   pile1sf01.verylow.misal1_csc11.005010.J1_pythia_jetjet
mc08.008078.PythiaPhotonJet6_FIXED.digit.RDO.E322_s391_d87  pile1sf01.verylow.misal1_csc11.005011.J2_pythia_jetjet
mc08.008095.PythiaPhotonJet1_FIXED.digit.RDO.e306_s387_d87  pile1sf01.verylow.misal1_csc11.005012.J3_pythia_jetjet
mc08.008097.PythiaPhotonJet3_FIXED.digit.RDO.e306_s391_d87  pile1sf01.verylow.misal1_csc11.005013.J4_pythia_jetjet
mc08.105011.J2_pythia_jetjet.digit.RDO.e344_s479_d126   pile1sf01.verylow.misal1_csc11.005014.J5_pythia_jetjet
mc08.107681.AlpgenJimmyWenuNp1_pt20.digit.RDO.e368_s462_d126  pile1sf01.verylow.misal1_mc12.005200.T1_McAtNlo_Jimmy
mc08.107682.AlpgenJimmyWenuNp2_pt20.digit.RDO.e368_s462_d126  pile1sf05.verylow.misal1_mc12.005200.T1_McAtNlo_Jimmy
mc08.107683.AlpgenJimmyWenuNp3_pt20.digit.RDO.e368_s462_d126  test
mc08.107685.AlpgenJimmyWenuNp5_pt20.digit.RDO.e368_s462_d126  test_2K_NewProduction_trig_btag.root
misal1_csc11.005011.J2_pythia_jetjet.digit.RDO.v12003103      test_NewProduction_trig_btag.root
misal1_csc11.005014.J5_pythia_jetjet.digit.RDO.v12003105      user09.DavidWilkinsMiller.jetAlgs.v08.mc08.105011.J2_pythia_jetje
misal1_mc12.005200.T1_McAtNlo_Jimmy.digit.RDO.v12000701       user09.DavidWilkinsMiller.jetAlgs.v08.mc08.105011.J2_pythia_jetje
misal1_mc12.006280.AlpgenJimmyWbbNp0.digit.RDO.v12000605      user09.DavidWilkinsMiller.jetAlgs.v08.mc08.105015.J6_pythia_jetje
misal1_mc12.006281.AlpgenJimmyWbbNp1.digit.RDO.v12000605      user09.DavidWilkinsMiller.jetAlgs.v08.mc08.105015.J6_pythia_jetje
misal1_mc12.006282.AlpgenJimmyWbbNp2.digit.RDO.v12000605      user09.DavidWilkinsMiller.jetAlgs.v08.valid1.105011.J2_pythia_jet
misal1_mc12.006283.AlpgenJimmyWbbNp3.digit.RDO.v12000605      user09.DavidWilkinsMiller.jetAlgs.v08.valid1.105011.J2_pythia_jet
myFirstRegistration.001.J0.RDO.log                      user09.DavidWilkinsMiller.jetAlgs.v08.valid1.105015.J6_pythia_jet
pile0sf00.lumioff.misal1_csc11.005001.pythia_minbias    user09.DavidWilkinsMiller.jetAlgs.v08.valid1.105015.J6_pythia_jet
pile0sf00.lumioff.misal1_csc11.005009.J0_pythia_jetjet
```

- Used for data which I know might disappear (`RDO` datasets especially)
- Private production (see all those **`pile1sf*`**?)

Examples of large batch production at WT2 (I)

Private pile-up digitization+reconstruction

In late 2007 and almost all of 2008 I needed (and still do!) large statistics pile-up samples that were not available centrally. This meant

- Locating `HITS` files for minimum bias and signal to combine into pile-up
  - Note that for a given number of signal events, you need at least a factor of 10 more minimum bias events in order to reliably simulate the pile-up
- Downloading these (extremely large!) HITS datasets to SLAC, because GRID production of pile-up was still in it's infancy and often not all the samples needed are available at a **single site** (crucial!)
- Access via interactive and batch jobs to the physical location of these datasets
- Large numbers of batch **digitization and reconstruction** jobs running simultaneously (at one time, 4000 jobs) because each single job can only process about 50 events reliably
- Storage of output ESD's and AOD's for later ntuple production
- Finally, if all that worked, produce ntuples from the modest amount of output data (500K events with pile-up)

SLAC NATIONAL ACCELERATOR LABORATORY

Examples of large batch production at WT2 (II)

Running ATHENA ntuple making in batch with XROOTD

The logical (only?) storage location for all of the output pile-up samples was of course XROOTD.

- Huge storage space, which at the time was difficult to read directly in an ATHENA job
  - "Optimized for reading and copying, not writing or constant file I/O"
- Ntuple jobs meant to read a lot of data (dump flat event contents) and write a lot of data
- Had to develop a system of shell scripts for copying in and out (**`xrdcp`**) the right files from XROOTD that would work in both an interactive job and in batch mode (*thanks Wei and Neal!*)

Once ntuples are produced, need to get them to my laptop (which may be anywhere in the world at the time) in order to make plots!

SLAC NATIONAL ACCELERATOR LABORATORY

Examples of large batch production at WT2 (III)

Transferring ntuples to my laptop for making pretty plots

As promised (and as you all know) the last step is to make all the pretty plots I will show you tomorrow. This means ntuples on my laptop:

- Recall that ntuples are resident on XROOTD
- Because of a change of guard in the GRID software at the time, **`dq2-put`** was not a viable option, and thus needed to take the brute force approach
- Make extensive use of XROOTD mounted file system (**`XrootFS`**), mounted on a single machine, and **`scp`** everything out of there!
  - This is a terrible approach, but it worked
  - Now have **`bbcp`** which works much & faster (see Wei's talk)
- Even with this approach, needed to be organized:
  - List of files to be transferred
  - Physical file locations (since of course, no "file discovery possbile" from coffee shop in North Carolina)

Once ntuples are produced, need to get them to my laptop (which may be anywhere in the world at the time) in order to make plots!

Outline

1 Introduction

2 Local computing and resources

3 GRID access and infrastructure

4 File transfer and registration

5 Performing analysis and looking ahead to first data

SLAC NATIONAL ACCELERATOR LABORATORY

Using the GRID and PATHENA at SLAC

- Setup is still slightly different from CERN (perhaps this has updated recently?)
- Other than that, **dq2** tools work as anywhere else (i.e. CERN)

SLAC Setup

```
if [ `uname` = 'Linux' ]; then
  uname -r | grep -q ^2\.6
  if [ $? -eq 0 ]; then  # kernel 2.6, RHEL4 or higher
    manpath="$MANPATH"
    . /afs/slac/package/vdt/wlcg-client/setup.sh
    . /afs/slac/g/atlas/packages/dq2-client/setup.sh
    export LFC_HOST=atl-lfc.slac.stanford.edu
    export DQ2_LOCAL_SITE_ID=SLACXRD
    export MANPATH="$MANPATH:$manpath"
  fi
fi
voms-proxy-init -voms atlas -valid 48:0
```

CERN setup

```
source /afs/cern.ch/atlas/offline/external/GRID/ddm/DQ2Clients/setup.sh
voms-proxy-init -voms atlas -valid 48:0
```

SLAC NATIONAL ACCELERATOR LABORATORY

Outline

1 Introduction

2 Local computing and resources

3 GRID access and infrastructure

4 File transfer and registration

5 Performing analysis and looking ahead to first data

SLAC NATIONAL ACCELERATOR LABORATORY

The dirty work: file transfer, dataset registration

As mentioned before, the data might be at SLAC (even if just ntuples), whereas I am at CERN

Moving files to and from CERN

Wei discussed this in detail yesterday, but here's my take on it:

- You know where the files are at SLAC
- Make a list and copy them, after setting up an SSH-key

```
BASE=/xrootd/atlas/usr/f/fizisist
SERVER=atlint01.slac.stanford.edu
OUTLOG=/u2/fizisist
OUTDIR=/u2/data
dirs="
pile1sf01.verylow.misal1_csc11.005009.J0_pythia_jetjet/MYNTUP/NTUP
pile1sf01.verylow.misal1_csc11.005010.J1_pythia_jetjet/MYNTUP/NTUP
"
N=1
for i in ${dirs}; do
  ~/scripts/bbcp -z -T /usr/local/bin/bbcp -s 64 -P 2 -f -r
      fizisist@${SERVER}:${BASE}/${i} ${OUTDIR}/${i} >& ${OUTLOG}/get_$N.log &
    ((N += 1))
done
```

The dirty work: file transfer, dataset registration

This can be done slightly more elegantly if one registers the datasets to the GRID first with **dq2-put**

```
dsName=user.DavidWilkinsMiller.pile1sf01.verylow.
       misal1_csc11.005009.J0_pythia_jetjet.13003003.RDO
dsPath=/xrootd/atlas/usr/f/fizisist/
       pile1sf01.verylow.misal1_csc11.005009.J0_pythia_jetjet/RDO

dq2-put -a --long-surls -s ${dsPath} ${dsName} >&
        ~/datasets/register.${dsName}.log &
```

Outline

1 Introduction

2 Local computing and resources

3 GRID access and infrastructure

4 File transfer and registration

5 Performing analysis and looking ahead to first data

SLAC NATIONAL ACCELERATOR LABORATORY

The whole chain: what does it take to do analysis at SLAC?

What do I have that I will need?

- Easy way to locate the data I need: run lists from operations, dataset queries via AMI and PANDA
- The ability to locate that data on the GRID: **`dq2`**
- PATHENA submission scripts for running over that data when available: from PATHENA TWiki directly
- Resources and scripts to generate additional samples for systematics cross-checks, non-centrally produced data, etc: SLAC Batch System
- Disk space to store the analysis data either from the GRID or from the local private production directly: **`XROOTD and XrootdFS`**
- The ability to copy the results of the large-scale analysis back "to me" for my pretty (or maybe not so pretty) plots: **`bbcp, dq2-put`**

SLAC NATIONAL ACCELERATOR LABORATORY

The whole chain: what does it take to do analysis at SLAC?

*What do I **not** have that I will need?*

- Less cryptic commands for XROOTD when not mounted as **`XrootdFS`**: needed for batch jobs
- Less error prone usage of SLAC batch with ATHENA: is there any way to make the learning curve less steep? Standard ATLAS skeletons for batch jobs?
- Standardized way to transfer data from SLAC to CERN: we all do it all the time...is it possible to make it as efficient as possible?
- Accessibility of XROOTD from ATHENA: had many problems in the past with POOL access in conjunction with XROOTD...not sure if this is still a point of failure, but should be clarified
- More accessible communication of available releases (and thus databases, if any) at SLAC: SLAC ATLAS Dashboard?
- Easy way to access run data stored at SLAC: sure there is a way to do this with PANDA, but not exactly sure (could do it with FDR1/2 data, though
- PROOF: do we have working examples of PROOF analysis at SLAC? Optimized for usage of XROOTD data, and would be extremely helpful for analysis of very large datasets

SLAC NATIONAL ACCELERATOR LABORATORY