ATLAS Software Infrastructure

Frederick Luehring
Indiana University
US ATLAS Tier 1&2 Meeting at Harvard
August 8, 2006

Introduction

- This talk is about the software infrastructure provided by the Software Infrastructure Team (SIT).
  - The SIT is convened by FL with lots of help from David Quarrie and Emil Obreshkov.
  - The most important (but certainly not only) SIT responsibility is building the ATLAS offline software and providing the distribution kit.
    - Doing this successfully requires a tremendous amount of help and feedback from the teams running the production and the testing and validation communities. The need for timely and copious feedback from the production team cannot be overstated.
- The SIT is not directly responsible for the tools that ATLAS uses on the GRID - for historical reasons this is separate.

SIT Responsibilities

- The SIT is responsible for the following areas:
  1. Compiler, OS, and platform support
  2. CVS, CMT, and the ATLAS offline software environment
  3. The tag collector
  4. Nightly and release builds
  5. Project building system
  6. Code Distribution (kit building scripts, Pacman, & mirrors)
  7. Systematic testing of the offline code (KV, ATN, & RTT)
  8. Discussion forums (HyperNews) - inc. help and prob. report
  9. Mailing lists (Simba2)
  10. Doxygen, web pages, and written documentation
  11. Bug reporting system (Savannah)
  12. Quality assurance (reviews, checkreq, RuleChecker)

US SIT People

- The following US people are involved with the SIT
  - Sebastian Binet (LBL): Release Coordinator
  - FL (IU): SIT Chair
  - David Quarrie (LBL): ATLAS computing project leader
  - Alex Undrus (BNL): Nightly Builds (NICOS) and Testing (ATN)
  - Martin Woudstra (LBL): Job transformations and geometry / conditions database release distribution and integration
  - Charlie Young (SLAC): HyperNews
  - Saul Youssef (BU): Pacman

  - Steve Goldfarb (Michigan) and FL (IU) have previously served as release coordinators.

Hot Issues for the SIT

- The release schedule (really controlled by SPMB)
- New Linux and compiler version: SLC4 & gcc345
- New platform support: AMD64, EM64T, and Mac
- New support people at CERN
- CVS repository clean-up
- Geometry / conditions database release distribution and integration
- Data security
- Kit improvements: nightly kits, reduced installation time, mirrors
- Space Management: how to provide storage for all the new users now joining ATLAS

Release Schedule

- The release schedule is set by the SPMB/CMB but responsibility for building the offline software rests with the SIT. There are frequent minor adjustments (i.e. delays) to the schedule when building.
- The current tight schedule for the rest of this year:

```
To: ATLAS SPMB <atlas-sw-spmb@cern.ch>
From: David Quarrie <DRQuarrie@lbl.gov>
Subject: Proposed schedule for CSC production
Date: Fri, 28 Jul 2006 07:50:55 +0200

Ian has made the following proposal for the CSC production schedule:

Sep 15  Start simul with 12.0.3. calibration samples first
Oct  6  Reco done on some samples with 12.0.3
Oct 25  Start reco again with calibrated 12.0.4
Nov 10  Decide if 12.0.5 is needed for reco.
Nov 24  (if not OK) start reco again with 12.0.5, at which point all the
        simulation should be done.
…
```

New Linux / gcc support

- CERN expects lxplus to be entirely converted to SLC4 (based on RHEL4) by the end of October.
  - The SIT is running nightlies for SLC4.
  - Release 13 (end Oct) is planned to be with both SLC3 (gcc323) and SLC4 (gcc345).
    - Release 13 will be the last release with SLC3/gcc323.
  - SLC5 (RHEL5) will be ready too late (2Q07) to be usable for LHC turn-on (assuming the current LHC schedule holds)
- SLC4 includes gcc345 - currently we are experimenting working with gcc344 and SLC3.
  - LCG_46 is now in development branch - it provides main externals (e.g. ROOT) that are compatible with gcc345.
  - After a long struggle, the ATLAS code is now compliant with the tighter syntax checks of the gcc34x.

32 vs. 64 Bit Support & Mac

- Nightlies are running that test native 64 bit versions of the ATLAS software and its externals.
  - The first stage is to build both 32-bit and 64-bit versions under SLC4 / gcc345.
    - Nightlies already in place - <1% of packages failing to compile in 64-bit mode
  - 32-bit mode SLC4 / gcc345 essentially functional (needs validation).
  - Goal is to have 64-bit mode functional by end of Oct 2006 and validated by end 2006.
  - All of the 64 bit SW is for AMD64 / EM64T and not Itanium.
- We are also thinking of porting to Mac OS/Intel
  - Lower priority with no dedicated workers
  - No plan to use the Mac for production

New SIT People at CERN

- Thanks to new M&OA money, new infrastructure people are starting at CERN.
  - Emil Obreshkov (SW Librarian) and Vasily Kabachenko (part time SW librarian) have been carrying most the SIT load at CERN (along with David Quarrie).
  - Igor Kachaev will assist the librarians .
  - Krzysztof Ciba will work on platform validation.
  - Toni Coarasa will work on getting the full software chain (including RTT) to work reliably.
- All of these people will contribute to user support.
- There are also new SIT people to work part time on documentation issues.
  - Stephen Haywood (RAL) has accepted the Documentation Coordination role - please support his efforts.

Cleaning the CVS Repository

- The ATLAS offline CVS repository has never been cleaned. We plan to clean the repository soon.
  - Our original plan was to use subversion but this was deemed “too risky” this close to turn on by SPMB/CMB.
  - One of the new SIT people will be assigned to the CVS cleaning task and will either recreate the CVS repository in a new instance or create a new directory tree in CVS with only the active packages.
  - This work will require much interaction with developers to identify which packages and versions should be kept and which should be removed.
  - It would be great if this could happen by the end of the year but it’s a BIG job.
  - It is still foreseen to switch to subversion after the ATLAS detector is commissioned.

Conditions Database Versioning

- It has become apparent in the last few months that it is urgent to cleanly define the version of the conditions database within the software distribution system.
  - Lots of time has been wasted by people using the wrong version of conditions database data.
  - Work is underway by Vakho Tsulaia and Martin Woudstra to define a clean set of job transformations and database tags that the production team can use with confidence that they are correct.
  - The tag collector team is working to provide a way to tag the database version in the tag collector so that it will be apparent to everyone.

ATLAS Data Security

Kit Issues

- Alex Undrus is working with Emil Obreshkov and Grigori Rybkine to provide a kit for every nightly.
  - Needed to run RTT tests on the grid.
  - Also provides potential to reduce problems caused by AFS.
  - BUT building and installing the kit lengthens the nightly build and potentially delays the start of the RTT tests.
- Grigori has put much work into changing the scripts that build the kits so that there is one Pacman files per project and not one per package.
  - This reduces the kit install time to less than two hours.
- Saul Youssef has been working to setup a Pacman based mirroring system for kit distribution.
  - Needed to avoid overloading the CERN Pacman cache.

Installing the ATLAS offline SW

- How should Tier 1/2/3 sites be installing the ATLAS offline SW locally? Answer: Pacman Mirrors
  - Currently Xin et al. are installing the software at the production sites directly from CERN. This is time consuming.
  - We are currently recommending trying Pacman mirrors.
    - Saul is the expert on this - please ask him detailed questions.
    - The official installation instructions are at:

    https://twiki.cern.ch/twiki/bin/view/Atlas/DraftNewInstallForWB
    - Use a cron job to check for SW updates frequently (~once per hour). Only the changes to the CERN cache will be downloaded when the mirror is updated.
    - The initial download to a mirror from the CERN cache of a new release does take about the same amount of time (~2 hours) that directly installing the cache currently takes BUT local installations will be much faster (~10-15 minutes).
  - Do NOT forget to run Kit Validation after installing.

More on Mirrors

- There are a number of advantages of mirrors:
  - Reduced installation time for users at each site with a mirror.
    - Should be able to locally install a release in ~10-15 minutes.
  - Reduced load on the central CERN site.
    - There already indications that the CERN Pacman cache is overloaded when many people try to use it to download a new release immediately after it becomes available.
  - Ability to install multiple releases in the same directory tree.
  - Simplified installation of the software.
  - Useful to keep grid machines running the same version.
- Currently BU, BNL, and IU have mirrors installed.
  - Other sites should follow this example so that we have an appropriate geographic distribution of mirrors.
  - It makes sense even for a Tier 3 site to have a mirror for local users.

Conclusion

- Time is tight! We will be turning on the detector soon.
  - It will be a stressful time for the SIT (& Tier 2!) activity.
- Feedback from the sites & production team is critical.
  - For proven problems please use Savannah!
    - Private emails are very dangerous - do not use them for problems.
  - Discussion on the HyperNews groups has proven useful.
    - Please use HyperNews for discussion. I am glad to point people to the correct forum to use. REMEMBER if you prove a problem in a HyperNews discussion - submit it to Savannah.
  - Please subscribe to the Releases and Distribution Kit & Offline Announcements HyperNews.
- Did I mention that you should report all problems in Savannah?