

Data Handling Overview

Status, Plans, Manpower *etc.*

4th March 2005

General Observations

- We have a lot of areas where only one person understands/maintains component
 - Should try to improve documentation
 - Confluence is a good tool for this
 - We don't need whole books, just an overview of key areas and technologies so others can come up to speed
 - Should try to spread knowledge among team
- Will try to set aside 2 days towards end of March for “developers workshop”
 - Aimed at developers (us), hopefully we already have good idea of user requirements
 - Chance to review existing tools/technologies
 - Chance to “brainstorm” on future work, next generation of pipeline, and ideas for data server backend
 - Hope to come away with specific tasks assigned to specific people

Pipeline

- Current (perl + j2ee) system in use and will remain so for remainder of this year
- Ongoing: (Dan+Matt)
 - Operate pipeline, handle user question, problems. (+?)
 - Add support for prod, dev, test databases. Clarify use, recovery, tagging etc.
 - Add support for cross-task datasets
 - Add support for task versioning (see Warren's presentation last week)
 - Archiving
 - Parallelization
 - Understand issues limiting number of simultaneous jobs (Richard needs to generate 50million DC2 events). Perhaps switch to using Navid's job submission interface.
 - Web interface improvements (more filtering, task editing)
 - Improved web interface to catalog (especially for DC2 MC) (Karen)
 - Request to handle input tar-balls from Bryson
 - Elapsed time plotting, more web queries, Etc...
- Future:
 - Continue to develop front-end using existing technologies (hibernate, spring, cold-fusion, etc). (Matt + others). Hopefully this will become a showcase for the technologies Matt has been pursuing, plus core of future pipeline.
 - Will need to put substantial manpower into next generation pipeline (Matt, Dan, Navid, New Hire, Mark, ?) . Can initially put it into production in parallel to current system.

System Tests

- JSP web front-end has been put together over last 2 months (**Max, Tony**)
- Ongoing:
 - **Karen** is adding ability to annotate plots and set user preferences via web.
 - Will soon need to integrate with Matt's permissions scheme (via web service?) (**Matt, Tony**)
- Future:
 - Add correlation plots (**Max**)
 - Caching/Memory issues (**Tony**)
 - Add capability of marking failed plots as "ok really"
 - Review implementation as time/experience allows

Data Server

- Ongoing: Developing JSP Web Interface + high speed initial metadata search (**Jean-Paul**)
 - Will continue to try MySQL and Oracle in parallel for now, using DC2 size data samples
 - Need quick interface for DC2 background studies
 - Striped down JSP page to submit request
 - **Navid** has set up Java interface for job submission
 - **Richard** has Root script to apply cuts.
 - Write output to FTP space (cron job cleans up)
- Future:
 - Continue to develop web front-end (**Jean-Paul**)
 - Automate ingest of DC2 data from pipeline
 - Try using xrootd+“event pump”(+tera-memory machine) to access data, apply cuts, transform result, stream to user. (**Jean-Paul, Tony, Matt**)
 - If this doesn't work use boring old batch jobs
 - Use WIRED for event based event browser (**Mark**)
 - Use AIDA for plotting cut distributions, simple analysis? (**Max**)
 - Have something available for DC2 use this summer.

Miscellaneous

- **Max, Tony and Matt** will review aidatId branches, agree on future directions.
- Get assistance from **TechPubs** to design “look and feel” for data handling pages.
- **Karen** will be helping ISOC (**Jim Lemon**) to improve trending database. **Rob Cameron** will provide detailed user requirements. **Karen** will start (with help from **Max**) to set up AIDA plotting. Will investigate use of Tomcat to integrate .jsp pages with existing .cfm pages.
- Citrix Farm (**Matt+Rodney**)
- Jira+Confluence
- GUI front-end for system tests?