



Catalog Archive Server

Development Plan

Outline



- Products
 - SkyServer, sqlLoader, sqlFitsCsv, Web Services
 - Downloadable clients (sdssQA, sqlcl)
- Development tracks
 - Serious Bugs
 - Required (missing) functionality
 - Performance Enhancements
 - Change Requests
 - Long-term development
- Time estimates

Serious Bugs



- Currently no showstoppers
- MS SQL bookmark lookup bug
 - SQL optimizer picks wrong plan when non-indexed fields involved in query
 - Millions of random disk accesses instead of sequential scan of table
 - Document bug, educate users on workarounds

Missing Features



- SkyServer
 - Interface to DAS
 - From form-based query tools (IQS/SQS)
 - From radial and rectangular search tools
 - Provide URLs to DAS files or tarballs
- sqlLoader
 - Backing out changes
 - Deleting a chunk and reloading it
 - Must be done manually, and very carefully!
 - Incremental loading

Change Requests



- More versatile crossID facility
 - Use query to narrow object selection further
 - Arbitrary selection of columns
- Form-based query enhancements
 - Imaging Query Server (IQS)
 - Spectro Query Server (SQS)
- Schema changes (post DR3)
 - Add galactic and survey coordinates

Performance



- Fully test and deploy CasJobs
 - Resolve linked server issues
 - Fast queries on one server, slower ones on another
 - Integrate with SQL search page
 - Route all queries through CasJobs
 - Queries longer than threshold resubmitted to CasJobs
- File Groups
 - Tested in DR1
 - Distribute tables on different disk volumes
- DB Partitioning
 - Partition data among cluster of servers
 - Horizontal partitioning of largest tables

Long-term Development



- Intelligent caching
 - Bypass caching appliance
 - Joint proposal with JHU CS department
 - Byte-yield metric to judge whether query to be cached
- Automated partitioning
 - Incorporate into loader publish step
- Integration with VO
 - Move from web pages to web services
 - Comply with VO standards
 - UCDs, VOTable

Effort estimates



- Missing features
 - Interface to DAS: 2 p-months
 - Incremental loading: 1 p-month
 - Back out and replace chunk: 1 p-month
- Change requests
 - IQS/SQS features: 2 p-months
 - Query in crossid: 2 p-weeks
 - Schema changes: 2 p-weeks
- Performance
 - File Groups: 1 p-month
 - CasJobs: 2 p-months
- Long Term
 - Data Partitioning: 6 p-months
 - Caching: 8–10 p-months