

# SAM Overview (training session)

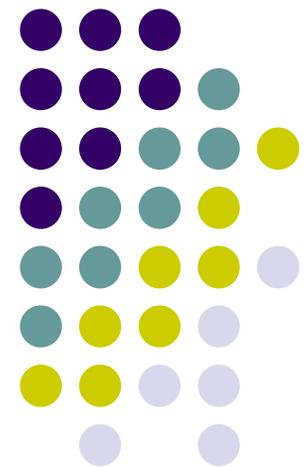
for CDF Users

Doug Benjamin

Duke University

Krzysztof Genser

Fermilab/CD

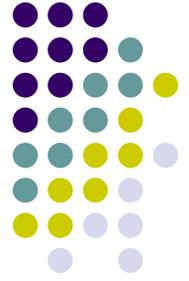




# Why SAM?

Why should CDF change its data handling model/methods?

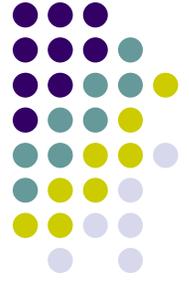
- No real choice...
  - CDF has lost and is losing the human resources required to maintain its own Data Handling system (Data File Catalog, DFC)
  - Computing Division is combining efforts to support the Tevatron Experiments (CDF/D0)



# Some SAM History

- SAM was designed and written as a joint project between D0 and Computing Division. D0 has been using it for many years now
  - Monte Carlo submission and scheduling (not a feature used in CDF)
  - Data Handling
    - Metadata catalogue (where to find the data)
    - Data delivery (SAM delivers files to users)
    - Monitoring/Tracking use of files through projects
- CDF started work on SAM some time ago
  - Several offsite locations have been using it for a few years now

# SAM @ CDF

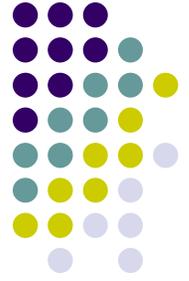


- CDF personnel who have worked on SAM integration and deployment
  - Valeria Bartsch, Doug Benjamin, Morag Burgon-Lyon, Gabrielle Compostella, Armando Fella, Krzysztof Genser, Randolph Herber, Suen Hou, Tsan Hsieh, Shih-Chieh Hsu, Elliot Lipeles, Donatella Luchessi, Ulrich Kerzel, Art Kreymer, Thoms Kuhr, Matt Norman, Fedor Ratnikov, Aidan Robson, Igor Sfiligoi, Stefan Stonjek, Alan Sill, Rick St. Denis, Bernd Stelzer, ... please let us know if we missed anyone)
  - and the SAM Users Committee  
David Dagenhart (editor), Ray Culbertson, Kenichi Hatakeyama, Daniel Whiteson, Konstantin Anikeev, Matt Herndon
- A lot of work by the SAM Team to accommodate CDF requests

# Sam @ CDF Support Model



- Each Physics Group has a power user who will help the members of the group.
- Wiki (Pasha's tiki) used for internal documentation. Power users able to make changes...
  - <http://www-cdf.fnal.gov/tiki/tikiindex.php?page=CdfSamUserDocumentation>
- Rick St. Denis assisting in answering all questions that the power users can't answer
- Doug Benjamin and Krzysztof Genser act as contacts between CDF and CD SAM team (as per agreement with Computing Division)



# SAM power users

- B group: Konstantin Anikeev ([anikeev@fnal.gov](mailto:anikeev@fnal.gov)) and Matt Herndon ([herndon@fnal.gov](mailto:herndon@fnal.gov))
- Electroweak: David Dagenhart ([wdd@fnal.gov](mailto:wdd@fnal.gov))
- Exotics: Ray Culbertson ([rlc@fnal.gov](mailto:rlc@fnal.gov))
- QCD: Kenichi Hatakeyama ([hatake@fnal.gov](mailto:hatake@fnal.gov))
- Top: Daniel Whiteson ([danielw@fnal.gov](mailto:danielw@fnal.gov))



# Sam @ CDF (Now)

- All DFC Metadata had been replicated automatically in SAM
- Production Farm Output (starting since June 2005 e.g. current 0h and 0i) only in SAM (not in DFC)
  - Soon to be applied to RAW Data as well
- Calibrations for the Farm Production Exe and DQM use SAM to read the data
- Stn & Top ntuples are being produced accessing data using SAM



# Going from DFC to SAM

- diskcache\_i interface to SAM and CAF make using SAM relatively transparent (not pain free though).
  - Some things change:
    - SAM delivers data via files (not filesets) specified by user defined datasets
    - SAM optimization not same as DFC
    - Files are NOT delivered to job sections in a fixed order (every time job runs sections can get them in a different order)
    - SAM provides good reporting for files read by jobs

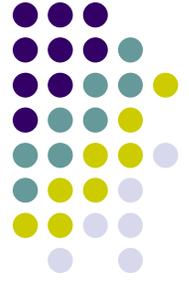


# SAM @ CDF (soon)

- Gen 6 MC will be uploaded into SAM only
  - The tools exist (sam\_upload)
    - They were developed and tested for B physics data skimming
    - Tested for upload of MC data from Toronto last week

# SAM@CDF

## (future enhancements)



- More advanced handshaking between CAF & SAM to facilitate more robust bookkeeping & error recovery (by the recovery projects)
- Improved interface between SAM & dCache

# SAM@CDF

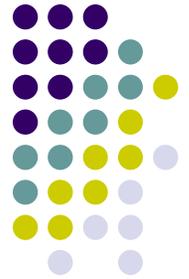
## (more use cases)



- Ntuples....
  - Ntuples can be loaded into SAM
  - Ntuple files need to be sufficiently large (> 1 GB)  
(small files are bad for pnfs/dCache/enstore)
  - Need Meta data associated with Ntuples
  - A root macro example using SAM:
    - <http://www-ekp.physik.uni-karlsruhe.de/~tkuhr/TSam/index.html>
  - Needs further testing/development
  - must use SAM when running on ntuples in the CAF; using the old methods is no longer an option due to administrative and scalability issues
  - ***Need help from the collaboration to solve this problem...***

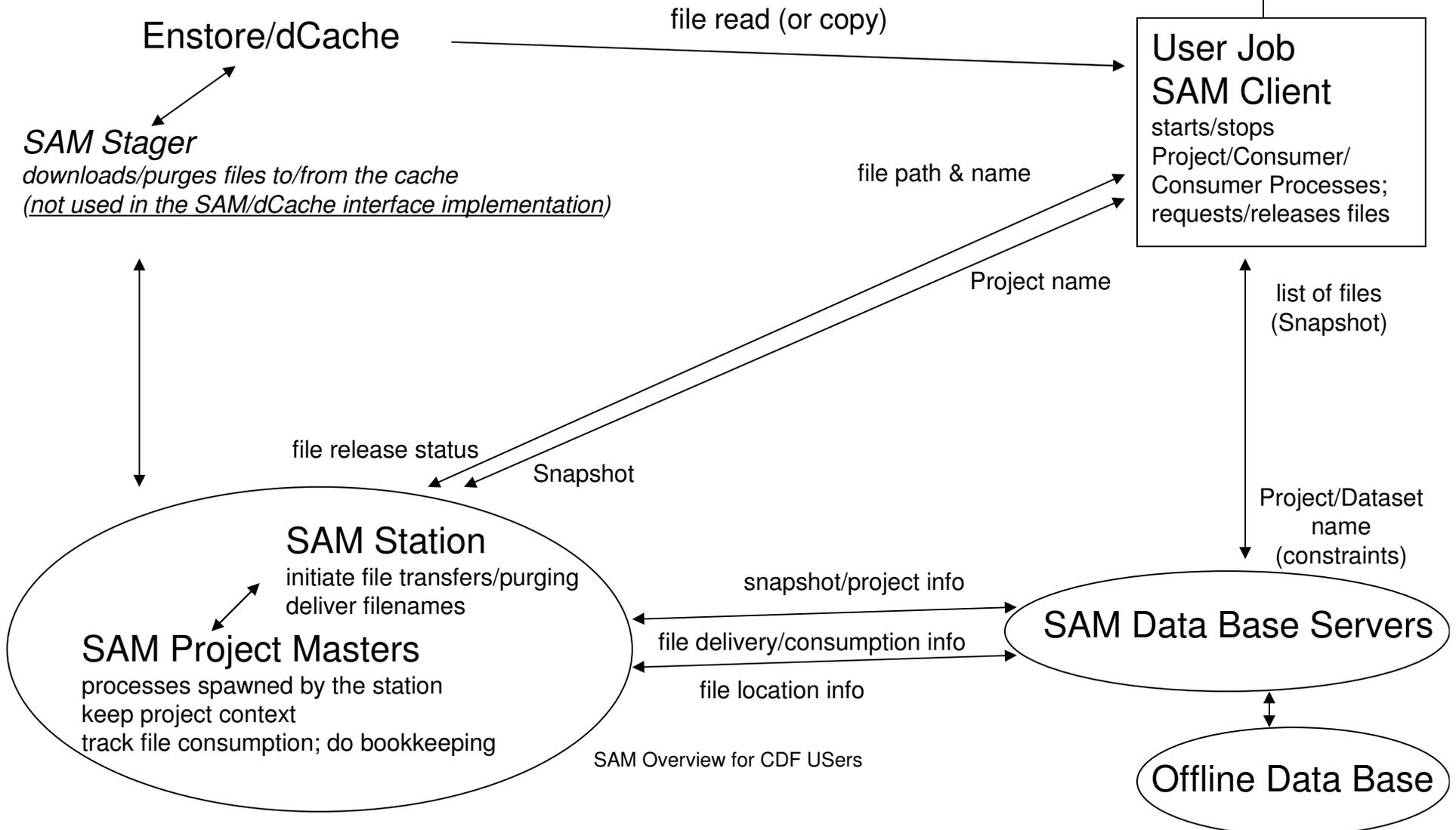
# SAM@CDF

## (for offsite use and installation)



- Please look at the following documentation
  - <http://projects.fnal.gov/samgrid/cdf/cdf.html>
  - <http://projects.fnal.gov/samgrid/shift.html>
- SAM can be used to fetch data from FNAL to offsite and handle it
- Datasets available at remote cdfs/SAM stations
  - <http://cdfsam-prd.fnal.gov/~sam/Datasets/>

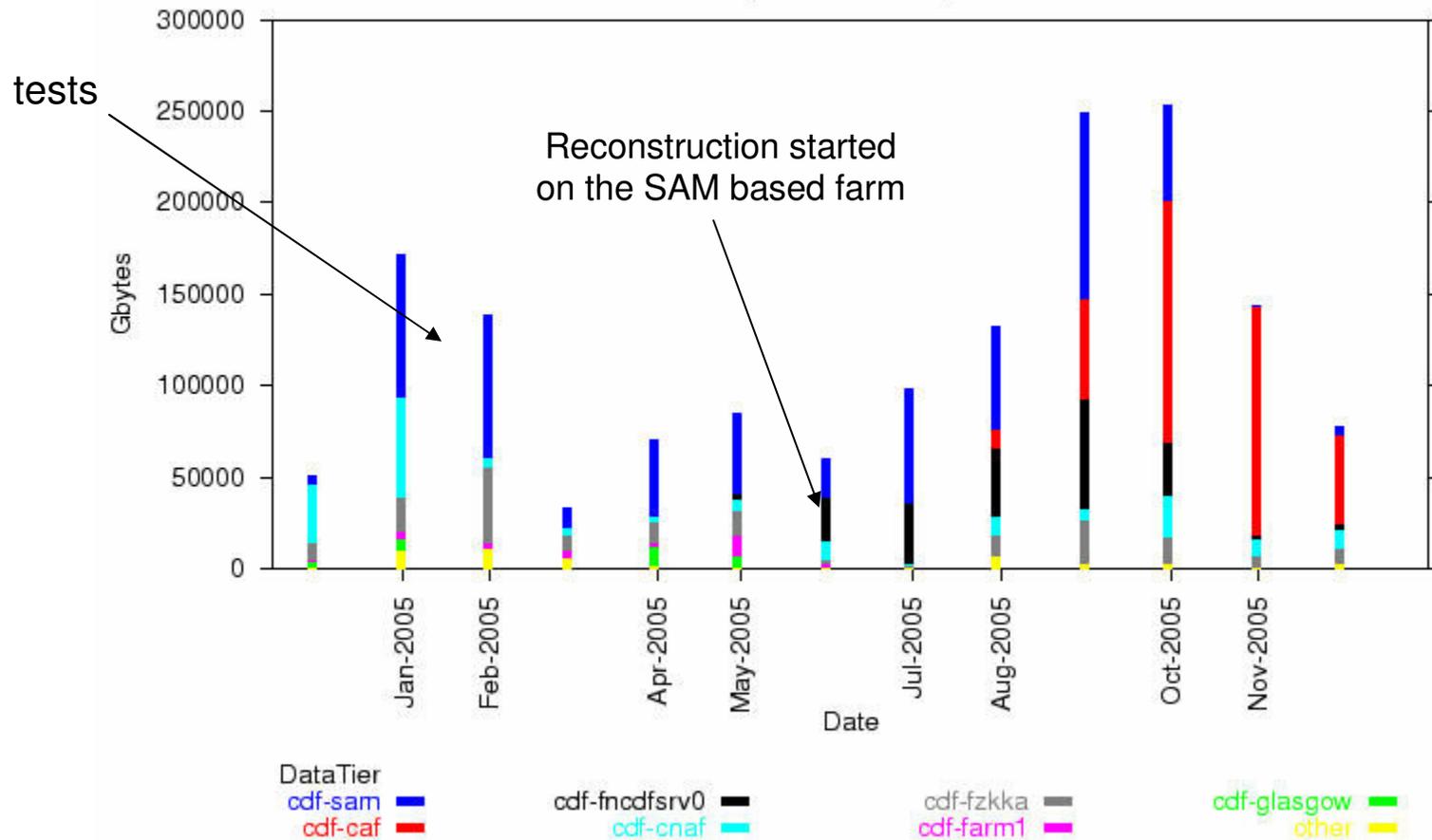
# SAM & Data Servers (used in data reading)



# SAM@CDF



Gbytes Consumed per Month on All Stations  
Year ending 15-Dec-2005  
(CDF Production)





# Conclusions

- SAM usage at CDF is increasing
- AC++/SAM/CAF interfaces are maturing
- help from the collaboration especially with the ntuple use case