

# DØ Requests for SAMGrid

Adam Lyon

Fermilab / Computing Division / DØ

Grid Activity Coordination Meeting  
27 February 2007

## DØ uses SAMGrid (JIM) for all offsite production

- Monte Carlo production
- p20 Reprocessing
- Primary processing (future?)

## DØ uses SAMGrid (JIM) for all offsite production

- Monte Carlo production
- p20 Reprocessing
- Primary processing (future?)

## The SAMGrid project phase transition to maintenance

- Freeze features
- But developers still available for crises, bug fixes, consulting
- Day-to-day support transitions to REX OPS team
- Support for running SAMGrid will not diminish!

## DØ uses SAMGrid (JIM) for all offsite production

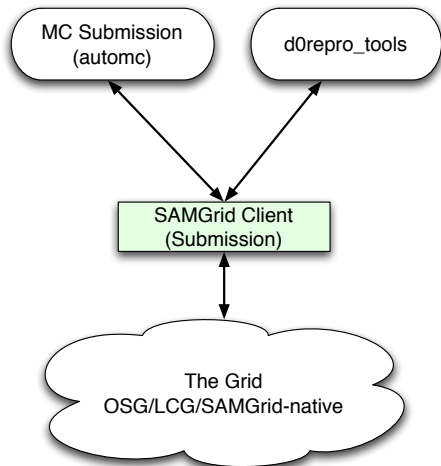
- Monte Carlo production
- p20 Reprocessing
- Primary processing (future?)

## The SAMGrid project phase transition to maintenance

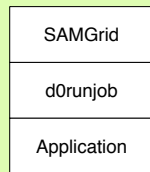
- Freeze features
- But developers still available for crises, bug fixes, consulting
- Day-to-day support transitions to REX OPS team
- Support for running SAMGrid will not diminish!

**Determine remaining DØ requests for SAMGrid now!**

# SAMGrid in Simple Pictures



## Job on worker node



- SAMGrid extensively sets up d0runjob
- d0runjob really a collection of scripts; SAMGrid writes macro to specify what to run

## Complete migration to SAM v7

- Complete MC Production chain works with `pythia`
- Need to integrate other generators (e.g. `alpgen`)  
Requires `d0runjob` work and SAMGrid integration.
- Status: `d0runjob` development in progress  
TODO: Finish development, integrate with SAMGrid and test

## Complete migration to SAM v7

- Complete MC Production chain works with *pythia*
- Need to integrate other generators (e.g. *alpgen*)  
Requires *d0runjob* work and SAMGrid integration.
- Status: *d0runjob* development in progress  
TODO: Finish development, integrate with SAMGrid and test

## Write any stage of MC output

- Current implementation only writes out final MC files
- Store intermediate MC files
- Status: *d0runjob* writes meta-data and issues SAM store commands for intermediate files  
TODO: Integrate with SAMGrid and test

### Start production from intermediate stage

- **Current implementation must start from generation with pythia**
- Desire to start production with a dataset of generated files (or any other list of intermediate files)
- Status: No changes to `d0runjob` required. Just needs new macro  
TODO: Determine how SAMGrid can write these macros



## Start production from intermediate stage

- **Current implementation must start from generation with pythia**
- Desire to start production with a dataset of generated files (or any other list of intermediate files)
- Status: No changes to `d0runjob` required. Just needs new macro  
TODO: Determine how SAMGrid can write these macros

## Generate CAF trees

- **Currently only thumbnails are generated**
- Make CAF trees from the thumbnails in the job
- Status: Development required in `d0runjob`. Easier if CAF tree maker is RTE'fied (not sure, but doubt it)  
TODO: Finish development, integrate with SAMGrid and test

## Run trigsim in MC production

- Currently we cannot run trigsim in MC production
- Desire to run trigsim as a production step
- Status: d0runjob development complete  
TODO: Integrate with SAMGrid and test

## Run trigsim in MC production

- **Currently we cannot run trigsim in MC production**
- Desire to run trigsim as a production step
- Status: d0runjob development complete  
TODO: Integrate with SAMGrid and test

## Run Recocert

- **Currently implemented**
- Make certification plots
- Status: Working

## Skimming

- Read in merged, unskimmed thumbnails. Run skim application (writes out many files of events separated by physics). Store unmerged skimmed files somewhere (durable cache), declare to SAM
- Recovery from failures **crucial**
- Currently skimming jobs are run by hand on the analysis farm by experts; recovery done by hand
- Have undated requirements document from Daniel Wicke
- TODO:
  - `d0repro_tools` needs development to launch jobs and handle recovery scenarios
  - `d0runjob` needs development to wrap skim application
  - SAMGrid then integrates and tests

## Run raw2sim on zerobias files

- MC production uses zerobias data in .sim format to overlay with MC events. Need to convert raw zerobias files into sim files with raw2sim application. Store into SAM
- Recovery from failures is not critical (just run more)
- zerobias raw files come from the detector (according to Todd)
- Raw files are big, and sim files are big too - no concatenation required
- Currently run by hand on analysis farm
- No requirements document
- TODO:
  - d0repro\_tools needs development to launch jobs
  - d0runjob needs development to wrap application
  - SAMGrid then integrates and tests

## Production of CAF Trees

- Run merged, skimmed thumbnails through the CAF tree maker, write out CAF Root Tree files. Store into SAM.
- Recovery from failures crucial
- Currently run by hand on analysis farm by Common Samples Group
- No requirements document
- Will concatenation be necessary? Probably not - but note no standard concat'er for Root files
- TODO:
  - `d0repro_tools` needs development to launch jobs and handle failure scenarios
  - `d0runjob` needs development to wrap application
  - SAMGrid then integrates and tests

## Standalone Recocert

- Run merged (skimmed?) thumbnails through recocert application for certification plots. Concatenate output? Store on tape?
- Recovery from failures desirable?
- Currently run by hand on analysis farm by SAM shifters
- Undated requirements document written by Daniel Wicke
- TODO:
  - `d0repro_tools` needs development to launch jobs and handle failure scenarios
  - `d0runjob` needs development to wrap application
  - SAMGrid then integrates and tests

# Summary

Task	Reqs?	repro?	runjob?	SAMGrid?
MC Complete v7	y	-	N	N
MC Write any stage	y	-	Y	N
MC Start any stage	y	-	-	N
MC Generate CAF	y	-	N	N
MC Trigsim	y	-	Y	N
MC Recocert	y	-	Y	Y
P Skimming	Y?	N?	N	N
P raw2sim	N	N	N	N
P Generate CAF	N	N	N	N
P Recocert	Y?	N?	N	N



# Merging (concatenation)

- Want to store big files into enstore, not lots of small ones
- Sources of small files:
  - Reco'ed thumbnails (from multiple streams)
  - MC thumbnails
  - Skimmed thumbnails (from multiple skims)
  - Recocert root files (if stored to tape)
  - CAF Trees from production & MC (maybe big enough)
- Merging began as a one-of exercise on one-stream Reco'ed thumbnails
- Now a two-of exercise with merging of MC thumbnails
- Problem is the Meta-data. Since MC and Production meta-data are so different, we have two merging applications
- De-hard-coding meta-data assumptions is the work!

## Merge (concatenate) multiple streams/skims?

- SAMGrid is a thick layer here
- **We need to negotiate requirements**
  - Do we launch one concatenation job that concatenates different skims?
    - Does SAMGrid determine what to merge? (hard)
    - Does `d0repro_tools` determine what to merge? (hard)
  - Do we launch one job per skim/stream to concatenate (easier for tools, but harder to manage)?
  - Who determines the meta-data?
  - Can `d0runjob` get more involved?

- DØ Responsibility
  - Peter Love (Imperial)
  - Michel Jaffre & Laurent Duflot (in2p3)
- Integration (macro writing, environment setup) performed by SAMGrid. Integration takes  $\sim 1$  day to write, several days to debug
- This above step must be performed for each new feature of d0runjob

- DØ Responsibility
  - Peter Love (Imperial)
  - Michel Jaffre & Laurent Duflot (in2p3)
- Integration (macro writing, environment setup) performed by SAMGrid. Integration takes  $\sim 1$  day to write, several days to debug
- This above step must be performed for each new feature of d0runjob

## Future plan?

- SAMgrid is only responsible for moving files to the worker node and passing JDL info; d0runjob does the rest.
- Unclear what effort it would take to make that work

# Road map (what I think is should be)

- 1 Recently finished identifying tasks
- 2 Prioritization of tasks ← We are here
- 3 **Acquire firm requirements**
- 4 Get time estimates from SAMGrid/d0runjob/d0repro\_tools
- 5 iterate
- 6 Determine path for d0runjob/SAMGrid integration; Determine path for concatenation
- 7 Negotiate task list
- 8 Do it

- SAMGrid Integration: Parag
  - THE SAMGrid Integration Expert
  - Currently busy with support of p20 reprocessing
  - Dehong picking up the load
  - Help from OSG troubleshooting task force
- SAMGrid Concatenation: Andrew
  - Wrote previous merging infrastructure
  - Slowly transitioning to SRM, but still majority of time on SAMGrid
- d0runjob: DØ (Peter Love, Michel Jaffre)
- d0repro\_tools: DØ (Daniel Wicke)
- DØ CPB is helping to focus DØ requests
- Will this be enough people? We don't know yet.