Special Assignment RunII/SAM

Definition, Status and Prospects:
A Presentation for the Grid and Data Management Meeting

Gerald Guglielmo
(CD/CEPA/OAA/OLA)

April 19th , 2005

For a Limited Time Only

- Informal discussions second week of February
- Informally started on February 14th, 2005
- Formal charge as of February 24th, 2005
- Not longer than September 30th, 2005
  - Earlier termination depends on status
  - End of fiscal year regardless of status
  - Closeout work afterwards as necessary

Goal 1: SAM in Production for CDF

Get the CDF experiment fully operational, in production mode, using SAM as their data handling system for raw event recording through to individual user analyses on-site and off-site. This needs to be achieved in a manner that uses SAM in an appropriate fashion and integrates well with the storage systems, production farm processing for CDF and the CAF software.

Goal 2: Common and Well Documented RunII Data Handling

The second goal is to assure that both CDF and D0 data handling operations are well documented and understood and use common installations and operational approaches where feasible and reasonable. This will help the Run II department to support both CDF and D0 data handling in a consistent and efficient way into the future.

Structure of the Assignment

- Report directly to Computing Division Head
- Work closely with various levels of management
  - RunII department upper management
  - Data Handling group leader
  - SAM project manager
  - CDF and D0 Computing coordinators
  - CCF department management (if necessary)
    - Upper/Lower Storage leaders (if necessary)

Spirit of the Assignment

"The above goals imply a coordinated program of work for the short, medium and long term. The job is therefore one of facilitating the planning, decision-making and program of work necessary to achieve the goals."

---Victoria White

High Level Task Categories

- Carry out a fact-finding mission
  - General overview of projects and interactions done
- Make a plan to achieve these goals
  - Plans and schedules for some subareas now exist
- Work with stakeholders, leaders, etc. for resource allocation and monitor progress
- Carry out tasks at your discretion (potentially becomes legacy code!)

High Level Status of Goal 1

1. SAM in Production for CDF
   - Six subareas identified
     - 1.1 Online data written directly to SAM
     - 1.2 Farms operating smoothly using SAM
     - 1.3 Real user analysis using SAM on Fermilab CAF
     - 1.4 Physics groups prepared to use SAM for summer conference work
     - 1.5 SAM Offsite robust and usable
     - 1.6 SAM-Grid with JIM in Production use

High Level Status of Goal 1 (Cont.)

- Plans and schedules for several subareas exist
  - 1.2 Farms operating smoothly using SAM
  - 1.3 Real user analysis using SAM on Fermilab CAF
- Implementation of first subarea plan in progress
  - 1.3 Real user analysis using SAM on Fermilab CAF
- Overall plan document being written

High Level Status of Goal 2

2. Common and Well Documented RunII Data Handling
   - Pending progress on the first goal

Status for 1.1 Online Data Written Directly into SAM

- Plan/Schedule not yet drafted
- Raw write via SAM since March 10th, 2005
- At or near end of Predator quiet period
  - Need to determine trigger for turning predator off
- Remaining work
  - Ensure robustness, automatic recovery
  - Documentation

Status for 1.2 Farms Operating Smoothly Using SAM

- Plan/Schedule drafted
- Plan driven by hardware purchase
  - Farm Head Node (est. May 10th)
- Some tasks will be staged in later
  - Separate Farm dCache
- Some tasks can be done early
  - Concatenation and dCache write pool

Status for 1.2 Farms Operating Smoothly Using SAM (Cont.)

- Prototype Farm work continues
  - Production processing of Stream A (April 18th)
  - Validate split stream processing (April 25th)
  - Production processing of split stream (~May 2nd)
- Prototype Farm could go production
  - Effort would change focus to a head node upgrade

Schedule for 1.2 Farms Operating Smoothly Using SAM

	Task	1	2	3	4	5	6	7	8
1.2.1	Farm Concat Hardware	Cooper 0.25 Zhang 0.25	Cooper 0.25 Zhang 0.25	<--CAN BE ACCOMPLISHED PRIOR TO START OF PLAN					
1.2.2	Farm dCache Write Pool Deploy	Cooper 0.15 Zhang 0.15 dCache Expert 0.2	Cooper 0.15 Zhang 0.15 dCache Expert 0.2	<--CAN BE ACCOMPLISHED PRIOR TO START OF PLAN					
1.2.3	Farm Head Node Deploy			Syu 0.5					
1.2.4	CAF Software Install				Syu 0.25				
1.2.5	Farm Software Insall					Suen 0.25			
1.2.6	Web Server Install				Syu 0.25				
1.2.7	SAM Software Install				Syu 0.1 Baranovski 0.15				
1.2.8	Farm Test						Suen 0.75 Benjamin 0.25	Suen 0.75 Benjamin 0.25	
1.2.9	Dcache Admin Node Deploy				Cooper 0.15 Zhang 0.15 dCache Expert 0.2	<---WILL BE STAGGED IN AT A LATER TIME			
1.2.10	Dcache Read Pool Deploy				Cooper 0.15 Zhang 0.15 dCache Expert 0.2	<---WILL BE STAGGED IN AT A LATER TIME			

FTE Effort for 1.2 Farms Operating Smoothly Using SAM

	1	2	3	4	5	6	7	
Developer								**Total**
Baranovski	0	0	0	0.15	0	0	0	0.15
Benjamin	0	0	0	0	0	0.25	0.25	0.5
Cooper	0.4	0.4	0	0.3	0	0	0	1.1
Dcache Expert	0.2	0.2	0	0.4	0	0	0	0.8
Suen	0	0	0	0	0.25	0.75	0.75	1.75
Syu	0	0	0.5	0.6	0	0	0	1.1
Zhang	0.4	0.4	0	0.3	0	0	0	1.1
Total	**1**	**1**	**0.5**	**1.75**	**0.25**	**1**	**1**	
Group								
Sam	0	0	0	0.15	0	0	0	0.15
RunII Ops	0.8	0.8	0	0.6	0	0	0	2.2
CD	0.2	0.2	0.5	1	0	0	0	1.9
NonCD Expt	0	0	0	0	0.25	1	1	2.25

Status for 1.3 Real User Analysis Using SAM on Fermilab CAF

- Plan/Schedule written
  - Implementation started April 4th, 2005
- Hardware subplan written
  - Completed April 1st, 2005 (a few changes maybe)
- User Committee Assessment
- Milestone is May 13th, 2005 (end week 6 in plan)

Schedule for 1.3 Real User Analysis Using SAM on Fermilab CAF

Task		0 (4/1)	1 (4/8)	2 (4/15)	3 (4/22)	4 (4/29)	5 (5/6)	6 (5/13)
1.3.1	DBServerFailover	Kumar .25 White .25	White .25					
1.3.2	Hardware config		White .25 Genser .25 Baranovski .25					
1.3.3	Admin Doc		White .2 Genser .2					
1.3.4	DBBrowserTools	Herber .25 UserComm	Herber .25 UserComm	Herber .25 UserComm	Herber .25 UserComm			
1.3.5	MigrateBooks	Herber .25 Benjamin .25	Herber .25 Benjamin .25	Herber .25 Benjamin .25	Herber .25 Benjamin .25			
1.3.6	User Doc AC++...		Bartsch .5	Bartsch .5	Bartsch .5	Bartsch .5	Bartsch .5	Bartsch .5
1.3.7	Job Start Issues		No good solution from batch perspective so use fallback of a timeout. Drop task from schedule at this time.					
1.3.8	Python/NFS	Illingworth .5 Sill .5	Illingworth .5 Sill .5	Illingworth .5 Sill .5				
1.3.9	Verify AC++/SAM		Started Early	Started Early	Genser .25 Bartsch .25			
1.3.10	Load Tests					Genser .5 Benjamin .5	Genser .5 Benjamin .5	<---Need to define metrics
1.3.11	UserDocs/v6 Links			Bartsch .5				
1.3.12	Define Support Proceedures							Genser .5 Bartsch .5
1.3.13	SAMTV/WebServer		Started Early	Started Early	Lyon .25 Carpenter .25 Zhang .25			
1.3.14	SAM V6 DBServer and client performance		Carpenter .5 Illingworth .5	Carpenter .5 Illingworth .5	Carpenter .5 Illingworth .5			
1.3.15	Deploy new hardware	Boyd .2 Cooper .2	<---DONE----					

FTE Effort for 1.3 Real User Analysis Using SAM on Fermilab CAF

	0	1	2	3	4	5	6	
Developer								**Total**
Baranovski	0	0.25	0	0	0	0	0	0.25
Bartsch	0	0.5	1	0.5	0.5	0.5	1	4
Benjamin	0	0.25	0.25	0.25	0.5	0.5	0	1.75
Boyd	0.2	0	0	0	0	0	0	0.2
Carpenter	0	0.5	0.5	0.75	0	0	0	1.75
Cooper	0.2	0	0	0	0	0	0	0.2
Genser	0	0.45	0.25	0.25	0.5	0.5	0.5	2.45
Heavy	0	0	0	0	0	0	0	0
Herber	0	0.5	0.5	0.5	0	0	0	1.5
Illingworth	0.5	1	1	0.5	0	0	0	3
Kumar	0.25	0	0	0	0	0	0	0.25
Lyon	0	0	0	0.25	0	0	0	0.25
Sill	0.5	0.5	0.5	0	0	0	0	1.5
White	0.25	0.7	0	0	0	0	0	0.95
Zhang	0	0	0	0.25	0	0	0	0.25
Total	**1.9**	**4.65**	**4**	**3.25**	**1.5**	**1.5**	**1.5**	
Group								
Sam	0.75	2.7	1.75	1.75	0	0	0	6.95
RunII Ops	0.9	1.7	2	1.25	1	1	1.5	9.35
CD	0.25	0	0	0	0	0	0	0.25
NonCD Expt	0	0.25	0.25	0.25	0.5	0.5	0	1.75

Status for 1.4 Physics Groups Prepared to Use SAM for Summer Conference Work

- Plan/Schedule not yet drafted
- May fall naturally, in part, out of subarea 1.3
  - Validation an ongoing process
- Ashutosh has a group of super users identified
  - 5 users on committee to try analysis with SAM
  - Waiting on response from DH group
  - Vanguard of user adoption

Status for 1.4 Physics Groups Prepared to Use SAM for Summer Conference Work (Cont.)

- Need SAM on CAF for user analysis first
  - Cannot sell groups on something not yet there
- Re-evaluate once CAF user analysis milestone achieved.

Status for 1.5 SAM Offsite Robust and Usable

- In use now offsite
  - Rely on centrally located Naming Service and SAM database
- Plan/Schedule not yet drafted
- Ideas actively being discussed
- Need to evaluate robustness
- Configuration and management
  - Python freezing may help here too

Status for 1.6 SAM-Grid with JIM in Production Use

- Plan/Schedule not yet drafted (by me)
  - Caveat JIM team does have plans for work
- Not following closely at this time (lower priority)
  - Progress is being made and no red flags so far
- Focus more attention when time permits
  - Need to review before June if possible

Summary of Special Assignment Goal 1 Status

- SAM in Production for CDF
  - Work in Progress
  - Some plans/schedules written
  - Many things left to do
    - Planing, implementing, debugging, etc.
  - At least one mini milestone achieved
  - Focus on amassing successes

Summary of Special Assignment Goal 1 Status (Cont.)

- Action Items Short List
  - Need defined metrics and scope for load tests
  - More advanced planning instead of on the fly
  - Formal response to shift definition recommendation
  - Delegation of tasks to prevent over commitment
  - Activity completion criteria
  - Desire keep on schedule
  - Momentum for last 5% of an activity

Summary of Special Assignment Goal 2 Status

- Common and well documented RunII data handling
  - Pending successes on goal 1