

Simulation of Daily Activities at Regional Centers

MONARC Collaboration

Alexander Nazarenko and Krzysztof Sliwa



K. Sliwa/ Tufts University
DOE/NSF Review January 20, 2000

Baseline Model for Daily Activities

Physics Group Analysis Physics Group Selection Reconstruction ESD Redefinition of AOD+TAG Replication (FTP) Monte-Carlo	200-400 jobs/day 20-40 jobs/day 2 times/year once/month after Reconstruction
--	---

Event processing rate: 1, 000, 000, 000 events/day

Physics Group Selection

Each group reads 100% TAG events and follows:

~10% to AOD

~1% to ESD

~0.01% to RAW

Number of Groups	Follow AOD	Jobs /group
L Groups (L~10)	p% of total TAG (~1%)	1-2
M Groups (M~5)	q% of total TAG (~5%)	1-2
N Groups (N~5)	r% of total TAG (~10%)	1-2

~20 Jobs/Day in total evenly spread among participating RCs

2005-2006 Estimate of Parameters

Parameter	2005	2006
Total CPU	350,000 SI95	520,000 SI95
CPU Unit	400 SI95/box (100 SI95/cpu)	400 SI95/box (100 SI95/cpu)
CPU I/O	40 MBps/box (0.1 MBps/SI95)	40 MBps/box (0.1 MBps/SI95)
AMS I/O for Discs		188 MBps/server
ThroughPut for Tape Storage		2000 MBps
Disk Space	340 TB	540 TB
Tape Space	1 PB	3 PB
LAN	31 GBps	46 GBps

(Les Robertson's estimate of July 99)

Problem Setting: Analysis and Selection

	RAW	ESD	AOD	TAG
Database	1,000,000,000 CERN	1,000,000,000 Tier 1:Locally Tier2:@Tier1	1,000,000,000 Locally @ RC	1,000,000,000 Locally @ RC
Physics Group Analysis 20groups * 10 jobs	0.01%	1%	Folow 100% of the group set	Group set: 1% of total TAG
Physics Group Selection 20groups * 1job	0.01%	1%	10%	100%

CPU (SI95)

250

25

2.5

.25

Totally 220 Independent Jobs: 200 Physics Group Analysis and 20 Group Selection

Problem Setting: Reconstruction and FTP

	Size	ESD	AOD	TAG
FTP	1 DB/Job	Tier 1 Centers	Tier 1 & 2	Tier 1 & 2
Full Reconstruction	6,000,000 events/day	yes	yes	yes
Monthly Reconstruction	100,000,000 events/day	no	yes	yes

CPU (SI95)

250

25

2.5

Participating Regional Centers

5 Tier 1 Regional Centers and one Tier 2 center

RC Name	Data	WAN Connection
CERN (Tier1)	RAW, ESD, AOD, TAG	All RCs
INFN (Tier1)	ESD, AOD, TAG	All Tier 1
KEK (Tier1)	ESD, AOD, TAG	All Tier 1
TUFTS (Tier1)	ESD, AOD, TAG	All Tier 1
CALTECH (Tier1)	ESD, AOD, TAG	All Tier 1 & Caltech-2
CALTECH-2 (Tier2)	AOD, TAG	CERN (RAW) & CALTECH (ESD)

200 Analysis and 20 Selection Jobs are evenly spread among Tier1 RCs

AMS load distribution

One RC (CERN) configured to run **200 concurrent Physics Group Analysis Jobs** and **20 Selection jobs** a day

Participating RC	Data	Jobs
CERN (Tier1)	RAW, ESD, AOD, TAG	200 Physics Group Analysis 20 Physics Group Selection x40

Model1 (optimized AMS distribution)

Model2

—

1 RC vs 5 RC: Group Selection on all data

Model1

One RC (CERN) minimally configured to run **20 concurrent Physics Group Selection Jobs** a day

Participating RC	Data	Jobs
CERN (Tier1)	RAW, ESD, AOD, TAG	20 Physics Group Selection x10

Model2

Five Tier 1 Centers minimally configured to perform **the same task**

Participating RC	Data	Jobs
CERN	RAW, ESD, AOD, TAG	4 Physics Group Analysis x10
INFN	AOD, TAG	4 Physics Group Analysis x10
KEK	AOD, TAG	4 Physics Group Analysis x10
TUFTS	AOD, TAG	4 Physics Group Analysis x10
CALTECH	AOD, TAG	4 Physics Group Analysis x10

Conclusion:

Current configuration provides a possibility to redistribute resources without much increase in the cost; further optimization is needed to increase the efficiency: $1/(Time*Cost)$

1 RC vs 6 RC: Reconstruction+Analysis+Selection

Model1

One RC (CERN) minimally configured to run all the Jobs a day

Participating RC	Data	Jobs
CERN (Tier1)	RAW, ESD, AOD, TAG	20 Physics Group Selection x10 200 Physics Group Analysis Full Reconstruction and FTP

Model2

Five Tier 1 Centers optimized to perform the same task with 30 MBps WAN

Participating RC	Data	Jobs
CERN	RAW, ESD, AOD, TAG	4 Physics Group Selectionx10 40 Physics Group Analysis Full Reconstruction and FTP
INFN	ESD, AOD, TAG	4 Physics Group Selectionx10 40 Physics Group Analysis
KEK	ESD, AOD, TAG	4 Physics Group Selectionx10 40 Physics Group Analysis
TUFTS	ESD, AOD, TAG	4 Physics Group Selectionx10 40 Physics Group Analysis
CALTECH	ESD, AOD, TAG	4 Physics Group Selectionx10 40 Physics Group Analysis
CALTECH-2 (Tier2)	AOD, TAG	40 Physics Group Analysis 20 Physics Group Analysis

Conclusion:

Current configuration provides a possibility to optimize the CPU power and reduce the cost; further optimization is possible to reduce WAN bandwidth to 30 MBps

6 RC: Two types of Reconstruction+Analysis+Selection

Five Tier 1 and one Tier 2 Centers optimized to perform the complete set with 30 MBps WAN and optimized LAN

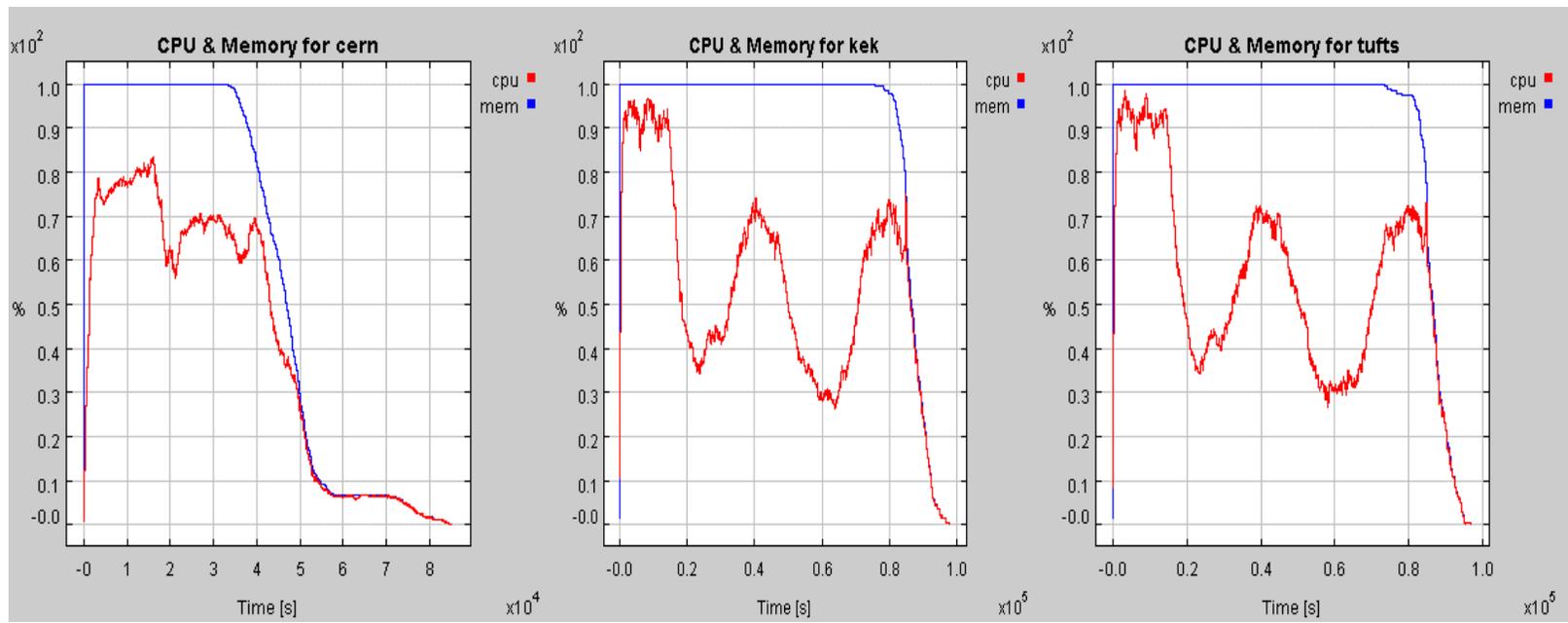
Participating RC	Data	Jobs
CERN	RAW, ESD, AOD, TAG	4 Physics Group Selectionx10 40 Physics Group Analysis Full Reconstruction and FTP Monthly Reconstruction and FTP (10days)
INFN	ESD, AOD, TAG	4 Physics Group Selectionx10 40 Physics Group Analysis
KEK	ESD, AOD, TAG	4 Physics Group Selectionx10 40 Physics Group Analysis
TUFTS	ESD, AOD, TAG	4 Physics Group Selectionx10 40 Physics Group Analysis
CALTECH	ESD, AOD, TAG	4 Physics Group Selectionx10 40 Physics Group Analysis
CALTECH-2	AOD, TAG	4 Physics Group Selectionx10 40 Physics Group Analysis 20 Physics Group Analysis

Model1 (fixed values)

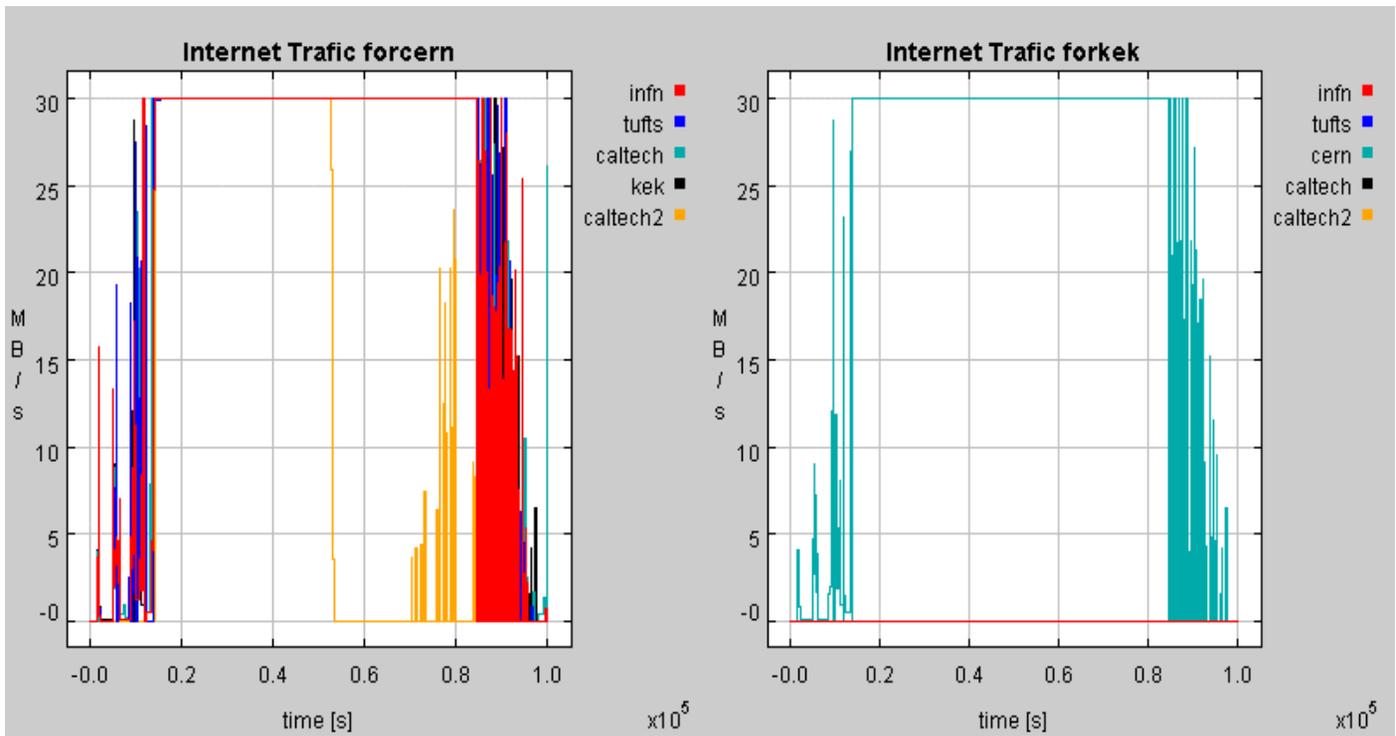
Model2 (randomized data processing times and sizes)

Conclusion: Current configuration provides a possibility to run daily the complete set of jobs at 6 centers with the WAN bandwidth 30 MBps and the network parameters not exceeding the estimate of 2005

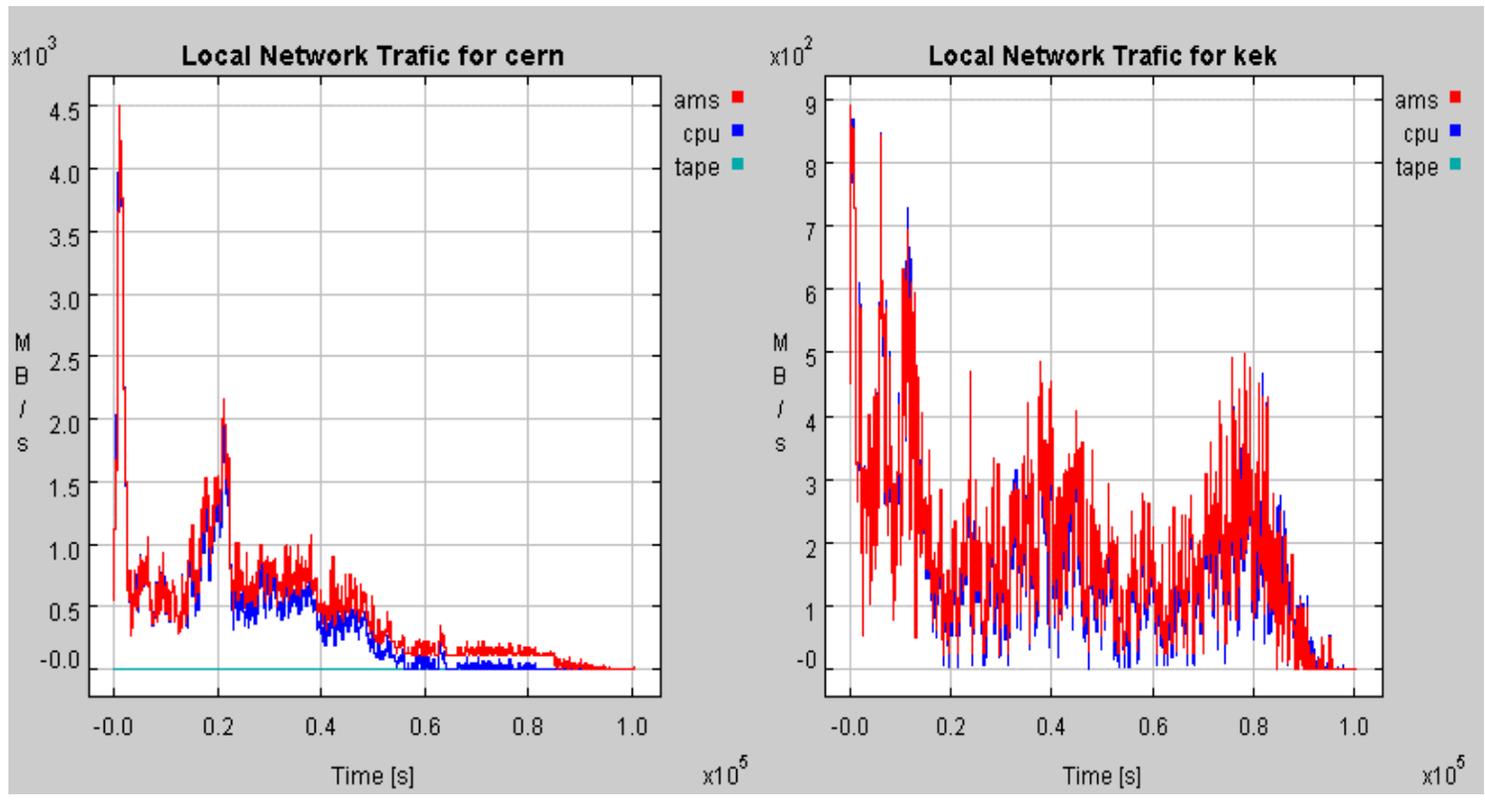
K. Sliwa/ Tufts University
DOE/NSF Review January 20, 2000



K. Sliwa/ Tufts University
 DOE/NSF Review January 20, 2000



K. Sliwa/ Tufts University
 DOE/NSF Review January 20, 2000



K. Sliwa/ Tufts University
DOE/NSF Review January 20, 2000

Future Work:

- **Replication job:** partial dataset replication from CERN to Regional Centers
- **Flexible data access:** data exchange between Regional Centers without getting data directly from CERN (depending on load, availability,...)
- **Imposing coherence on the concurrent jobs:** if Reconstruction and/or Replication is taking place, Analysis/Selection jobs should be able to monitor new data availability if requested
- **Improving Cost function:** adding cost of WAN, adding other hidden costs currently not accounted for
- **Optimization with respect to the parameter *Time*Cost* for a given task run on different architectures**