

HEPCG Work Package 3 Interactive Analysis

Status and Plans

Kilian Schwarz
GSI Darmstadt

Milestones

12/07: PM 27



- Interactive Analysis
 - lokales PROOF Cluster (PM18)
 - PROOF auf dem Grid (PM24)
 - GUI (PM36)
 - Tests in user community (PM36)
 - Meilenstein 3.4.1 (dynamische Generierung von PROOF Clustern (PM24))
- Application Oriented Scheduling
 - Test der Jobumgebung (PM24)
 - Logging des Ressourcenverbrauchs bei Auswahl der Ereignisse (Event Listing) (PM24)
 - Logging des Ressourcenverbrauchs bei Analyse (PM24)
 - Simulation des Ressourcenverbrauchs (PM30)
 - Log Interface (PM30)
 - Meilenstein 3.2.1 - Prototyp (PM24)
- Automatischer Jobmanager
 - Basisfunktionalität (LMU)
 - Tests (LMU)
 - Meilenstein 3.3.1 – Prototyp (LMU)

GSI D-Grid Cluster and GSIAF – present status



150 Mbps



Grid

Grid frontend (gLite/ Globus)

WMS/CE



WMS/CE

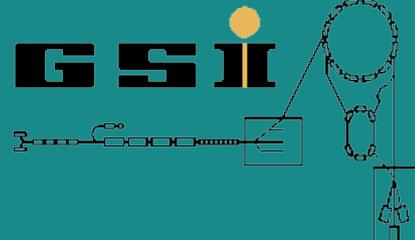
upgrade beginning 2008:
54 compute nodes
35 file servers



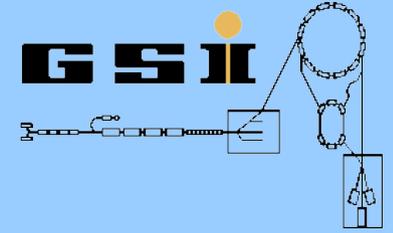
PROOF / Batch



standalone PROOF cluster



Present Status



- Batch Farm/GSI AF and ALICE::GSI::SE_tactical::xrootd nodes dedicated to ALICE:
- 15 D-Grid funded boxes: each
 - 2*2core 2.67 GHz Xeon, 8 GB RAM
 - 2.1 TB local disk space on 3 disks + system disk
- on all machines: Debian Etch 64bit

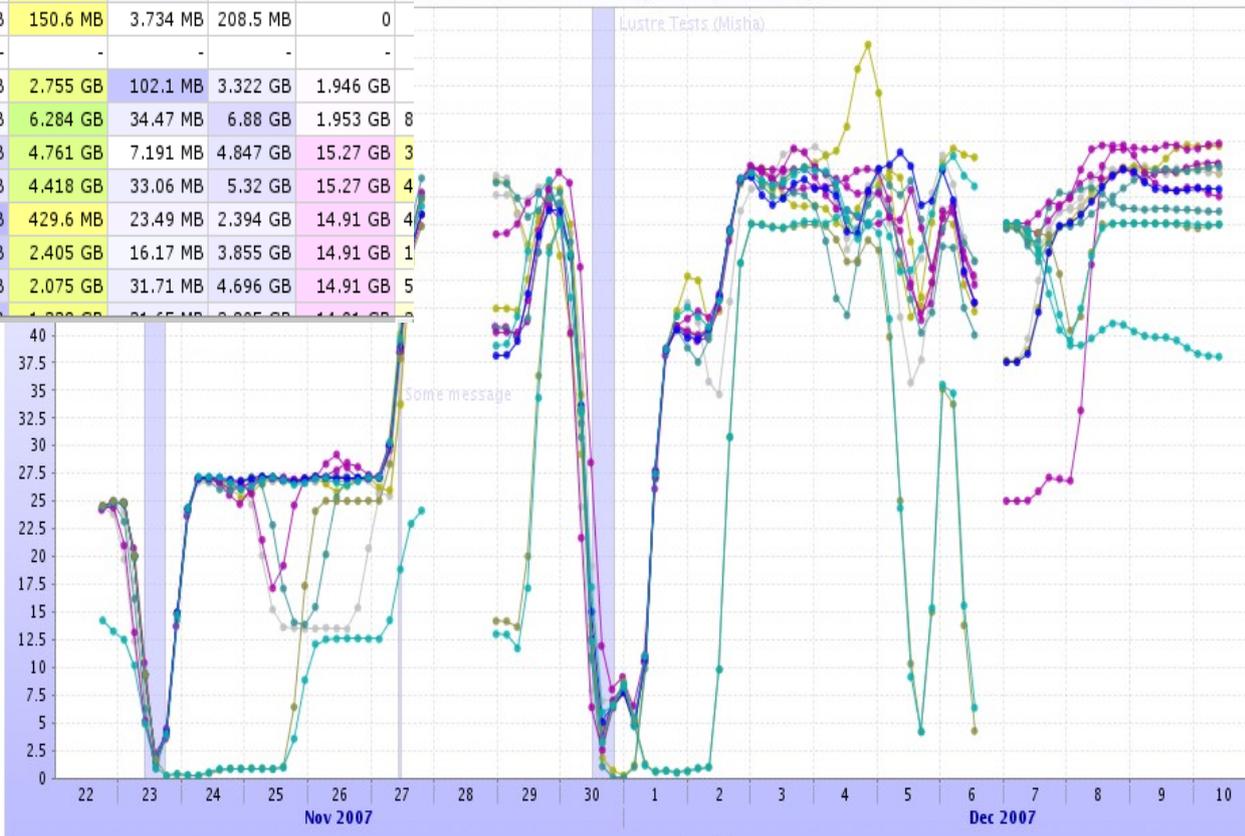
Farm Monitoring (MonaLisa)

GSIAF Cluster

Machines status

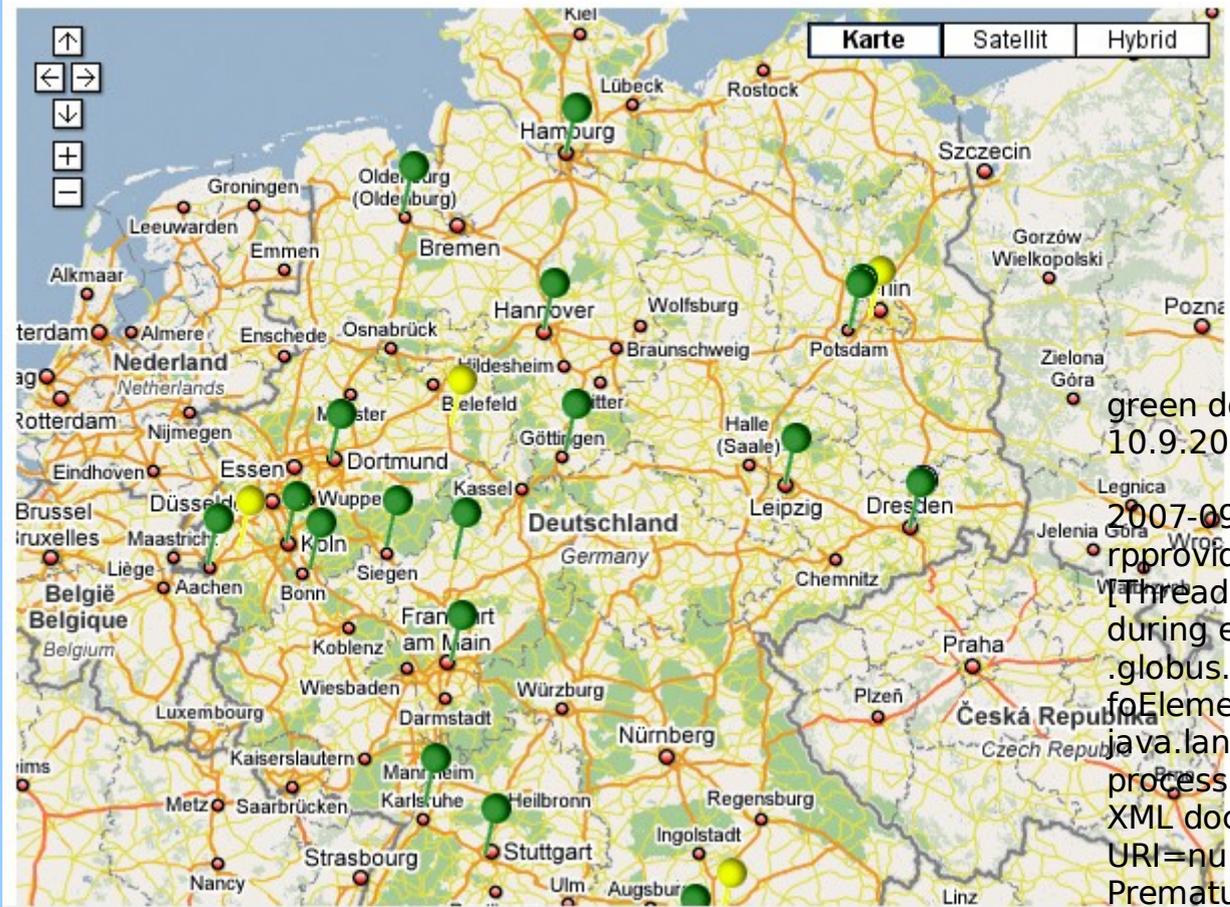
Machine	Machine status		CPU					Memory					Total
	Online	Ping	load	usr	sys	nice	idle	Total	Used	Cached	Buffers	Free	
grid1.gsi.de	Online	0.024ms	0.15	1.51	0.809	0	97.68	4 GB	1.332 GB	956 MB	28.49 MB	2.668 GB	0
grid13.gsi.de	Online	0.371ms	1.24	25.08	27.85	0	47.07	1.194 GB	150.1 MB	592.9 MB	9.875 MB	1.048 GB	0
grid19.gsi.de	Online	0.343ms	0.01	0.35	0.033	0	99.62	1 GB	409.9 MB	563.5 MB	29.99 MB	614.3 MB	0
grid29.gsi.de	Offline	-	-	-	-	-	-	-	-	-	-	-	-
grid32.gsi.de	Online	0.317ms	0	1.05	0.3	0	98.65	4 GB	184 MB	396.6 MB	12.27 MB	3.82 GB	0
grid5.gsi.de	Online	0.911ms	0.19	4.172	1.055	0	94.77	1 GB	403.1 MB	598 MB	14.86 MB	621.1 MB	0
grid8.gsi.de	Online	0.224ms	0.46	10.51	2.218	0	87.27	756.2 MB	547.7 MB	150.6 MB	3.734 MB	208.5 MB	0
gsiaf.gsi.de	Offline	-	-	-	-	-	-	-	-	-	-	-	-
lcg06.gsi.de	Online	0.189ms	0.44	5.402	3.295	0	91.3	3.911 GB	603.3 MB	2.755 GB	102.1 MB	3.322 GB	1.946 GB
lxb255.gsi.de	Online	0.211ms	0	0.008	0.025	0	99.97	7.801 GB	942.8 MB	6.284 GB	34.47 MB	6.88 GB	1.953 GB
lxb256.gsi.de	Online	0.245ms	3.08	31.63	0.234	0	68.13	7.801 GB	2.954 GB	4.761 GB	7.191 MB	4.847 GB	15.27 GB
lxb257.gsi.de	Online	0.181ms	1.17	10.63	0.312	0	89.06	7.801 GB	2.481 GB	4.418 GB	33.06 MB	5.32 GB	15.27 GB
lxb258.gsi.de	Online	0.228ms	6.72	52.19	7.058	0	40.76	7.801 GB	5.407 GB	429.6 MB	23.49 MB	2.394 GB	14.91 GB
lxb259.gsi.de	Online	0.244ms	3.72	42.43	1.726	0	55.85	7.801 GB	3.946 GB	2.405 GB	16.17 MB	3.855 GB	14.91 GB
lxb260.gsi.de	Online	0.324ms	3.8	43.04	2.298	0	54.66	7.801 GB	3.105 GB	2.075 GB	31.71 MB	4.696 GB	14.91 GB

History of cpu_usage



D-Grid Ressources

Dynamisch erstellte Karte von Ressourcenanbietern / Wartungszustand



green dot Script was not compatible to Globus 4.0.5
10.9.2007:

```
2007-09-10 17:08:08,243 ERROR  
rpprovider.ResourcePropertyProviderTask  
[Thread-11,timerExpired:159] Unhandled exception  
during execution of org  
.globus.mds.usefulrp.rpprovider.producers.SchedulerIn  
foElementProducer :  
java.lang.Exception: Exception while parsing child  
process stdout into valid  
XML document: org.xml.sax.SAXException: Fatal Error:  
URI=null Line=-1:  
Premature end of file. 2007-09-10 17:08:08,248 WARN  
rpprovider.ResourcePropertyProviderTask [Thread-  
11,timerExpired:164] Timer  
event canceled due to unhandled exception during  
execution of org  
.globus.mds.usefulrp.rpprovider.producers.SchedulerIn  
foElementProducer
```

GSI AF usage experience

- real life analysis work of staged data by GSI ALICE group (1-4 concurrent users)
- 2 user tutorials for GSI ALICE users (10 students each training)
- GSI AF and CAF were used as PROOF clusters during PROOF course at GridKa School 2007
-
- ==> continuous improvement through user requests and debugging

GridKa School 2007

GridKa School 2007 - Agenda - Microsoft Internet Explorer

Adresse <http://gks07.fzk.de/Agenda.html>

	room number 157	Introduction to the Simple API for grid applications SAGA Tutorial	room: Aula
		room number 162	
20:00	"Kneipenabend"		

Thursday, September 13			
9:00-9:45	Grid Access Manager (U. Biewer, IBM) abstract		
9:45-10:30	Research and Education on Microsoft HPC (Dr. Ingo Dahm, Microsoft) A Live Demonstration with Wolfram gridMathematica (Maryam Karbalai, Additive GmbH) MatLabMMA.zip abstract		
10:30-11:00	Break		
11:00-11:30	Grid Applications in D-Grid (W. Gentzsch, D-Grid) abstract		
11:30-12:00	Grid Business Models (J. Altmann, IU Bruchsal) abstract		
12:00-12:30	dCache (O. Synge, DESY) abstract		
12:30-14:00	Break		
14:00-19:00	Unicore course (R. Breu, FZ Juelich, and team) abstract	dCache course (O. Synge, DESY, and team) abstract Installing dCache with dCache Yaim dCache Clients	PROOF (J.F. Grosse-Oetringshaus and Andreas Joachim Peters, CERN) abstract info tar-ball
	room: Aula	room number 157	RGlite demo (Anar Manafov, GSI) abstract room number 162
20:00	School dinner		

Start | Internet Explorer | Microsoft PowerPoint | X-Win32-Konsole | X-Win32-Konsole | Rechner | DE | 12:59

PROOF on the Grid (GUI)

Server

Server's status:

XROOTD <11723>: is running
PROOFAgent <11732>: is running
PROOFAgent's version:
PROOFAgent v1.0.3.1466

application file name: proofagent
protocol version: PProtocol:0.1.0
Report bugs/comments to A.Manafov@gsi.de

PA's pid directory: /tmp/

Start
Stop
Status

Close

Grid

JDL File: /gLitePROOFjdl Browse...

Endpoint: https://grid25.gsi.de:7443/glite_wms_wmproxy_s

Submit 3 worker(s)

100% submitted

Information about the last submitted job:

ID	Status
https://grid25.gsi.de:9000/hRidEstyn...	Running
https://grid25.gsi.de:9000/vvD6p...	Scheduled
https://grid25.gsi.de:9000/kRRqv...	Scheduled
https://grid25.gsi.de:9000/dnnvK...	Scheduled

Close

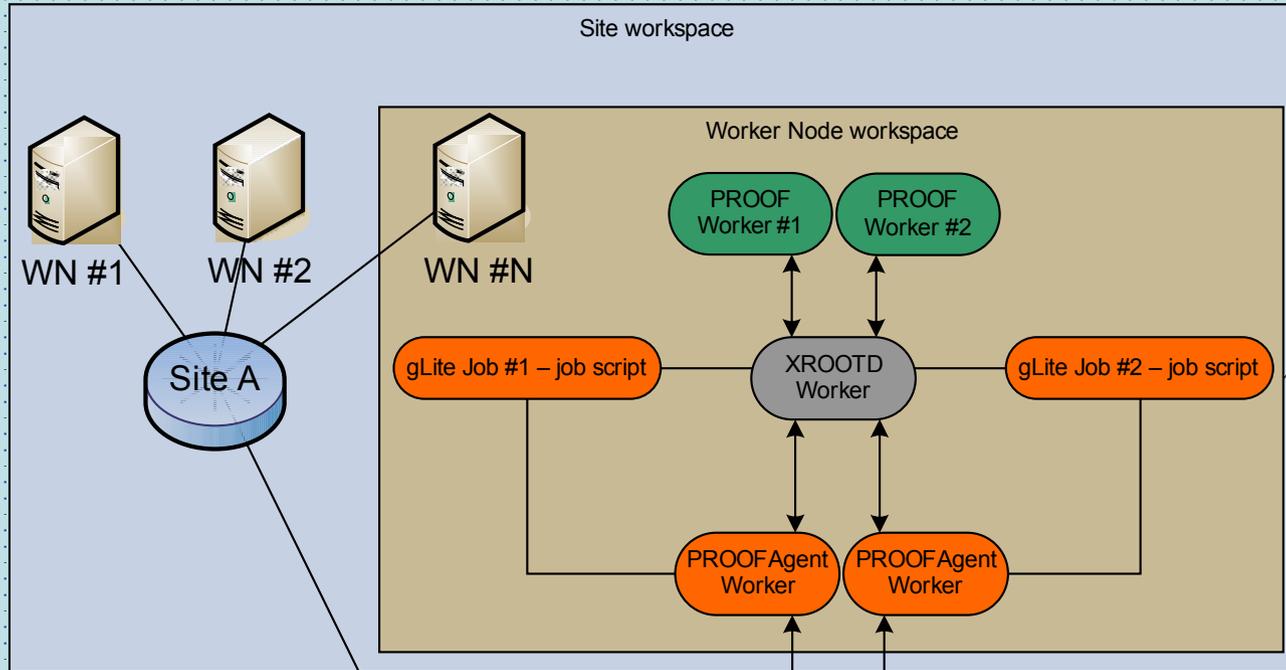
Worker(s)

Monitor connections (available 3 out of 3 worker(s)):

master depc218.gsi.de
worker dech001@grid4.gsi.de:53240 (redirect through local
worker dech001@grid44.gsi.de:36829 (redirect through local
worker dech001@grid17.gsi.de:59676 (redirect through local

Close

and ROOT gLite interface



Workspace prerequisites:

- gLite WN
- ROOT
- XROOTD
- Outgoing connection

PROOF on the Grid /
dynamic creation of
PROOF Clusters

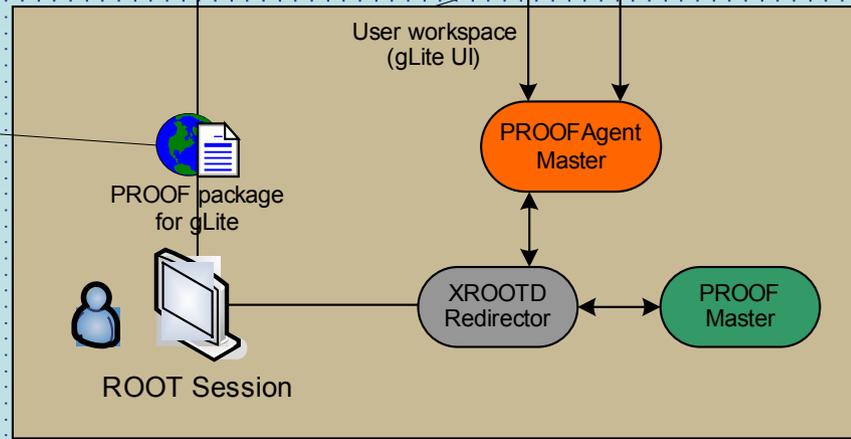


Workspace content:

- gLite UI
- ROOT
- XROOTD (with GSI authentication)
- xpd.cfg (generic XROOTD config)
- PROOFAgent (master mode)
- proofagent .cfg.xml
- Server_gLitePROOF.sh
- PAConsole (optional)

Content of gLite job:

- gLitePROOF .jdl
- gLitePROOF .sh
- xpd.cfg (generic XROOTD config)
- PROOFAgent (worker mode)
- proofagent .cfg.xml

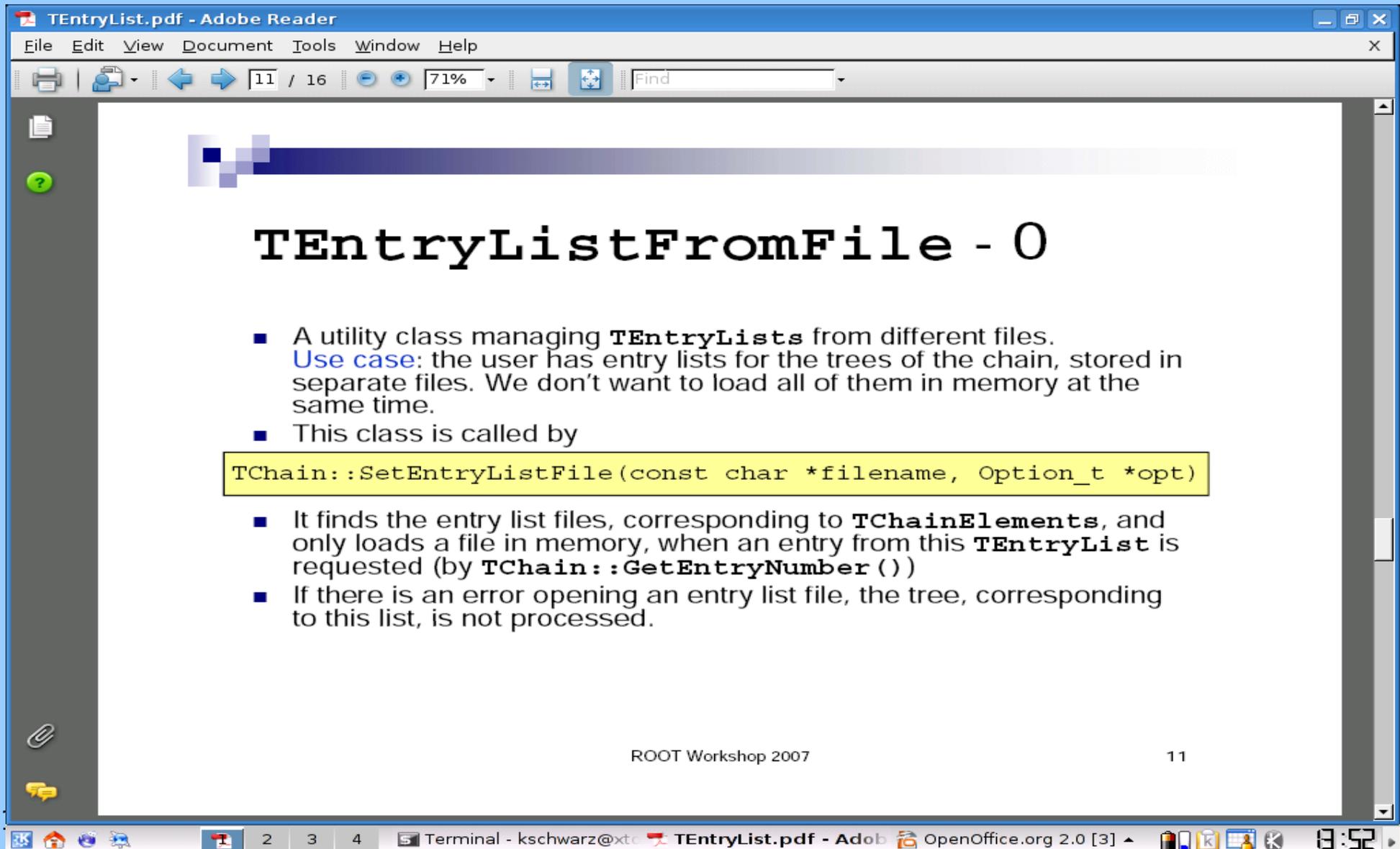


Legend:

- Newly developed components
- ROOT components
- XROOTD

13-Dec-2007

application oriented scheduling



TEntryListFromFile - 0

- A utility class managing **TEntryLists** from different files.
Use case: the user has entry lists for the trees of the chain, stored in separate files. We don't want to load all of them in memory at the same time.
- This class is called by

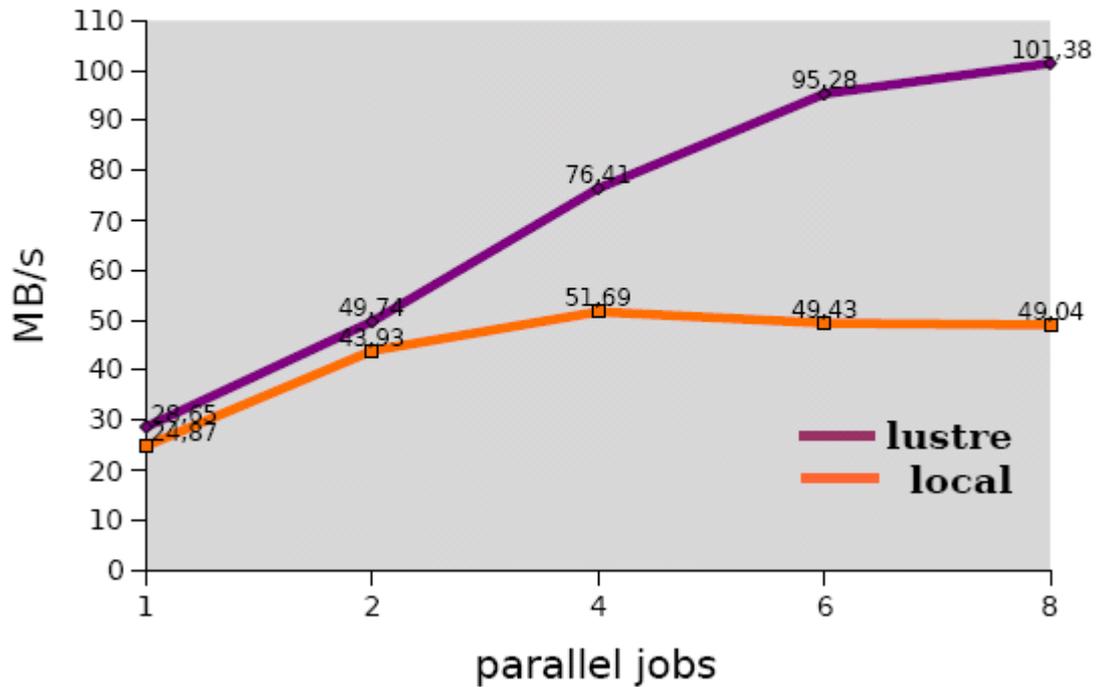
```
TChain::SetEntryListFile(const char *filename, Option_t *opt)
```

- It finds the entry list files, corresponding to **TChainElements**, and only loads a file in memory, when an entry from this **TEntryList** is requested (by **TChain::GetEntryNumber()**)
- If there is an error opening an entry list file, the tree, corresponding to this list, is not processed.

ROOT Workshop 2007 11

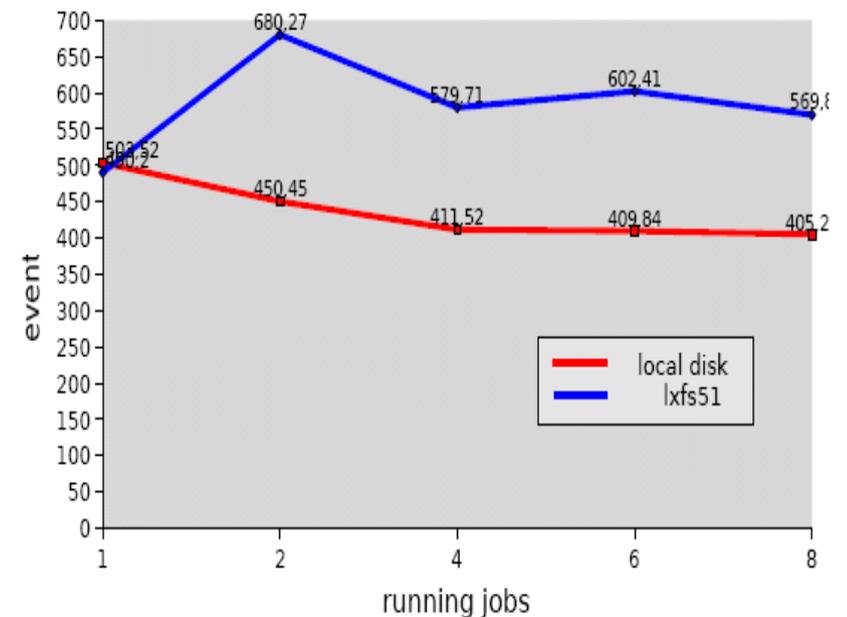
optimised PROOF I/O

Parallel Jobs Reading from Local Disk / from Lustre Cluster



ROOT Files 1.2 GB

Speed for parallel jobs (simple task) at lxb281



13-Dec-2007

AliESDs: 3.5 MB

Identifying differences between local ROOT and PROOF jobs. Simplifying debugging and resource monitoring

PROOF Processing Logs: localhost

```
// ----- Start of element log -----  
// Ordinal: 0.0 (role: worker)  
// Path: Anna@localhost:1093//tmp/proof/Anna/session-denbp035-1197467813-14327/worker-0.0-denbp035  
// # of retrieved lines: 2  
// -----  
14:56:57 14336 Wrk-0.0 | Error in <TTree::SetBranchAddr>: The pointer type given "Short_t" (2)  
14:56:57 14336 Wrk-0.0 | Error in <TTree::SetBranchAddr>: The pointer type given "Short_t" (2)  
// ----- End of element log -----
```

PROOF Query Progress: Anna@localhost

Executing on PROOF cluster "localhost" with 2 parallel workers:
Selector: Analyzer.C
2 files, number of events 20, starting event 0

Initialization time: 0.5 secs
Processed: 20 events (0.00 MBs) in 0.4 sec
Processing rate: 47.4 evts/sec (0.0 MBs/sec)

Close dialog when processing is complete

Max. memory usage Memory Plot

Stop Cancel Close Show Logs Rate plot

Choose workers: all master 0.0 Lines: from -10 to 0 Grep for: Error Display

Remote Grep in logs

Memory monitoring

Simple GUI for log handling

```
*****  
* VERSION 5.17.07 5 DECEMBER 2007 *  
* You are welcome to visit our Web site *  
* http://root.cern.ch *  
*****  
ROOT 5.17/07 (branches/dev/proof@21318, Dec 12 2007, 14:56:43 on linux)  
CINT/ROOT C/C++ Interpreter version 5.16.27, Oct 25, 2007  
Type ? for help. Commands must be C++ statements.  
Enclose multiple statements between { }.  
root [0] .L ~/script/prooftest2.C+  
root [1] prooftest2()  
Starting master: opening connection ...  
Starting master: OK  
Opening connections to workers: OK (2 workers)  
Setting up worker servers: OK (2 workers)  
PROOF set to parallel mode (2 workers)
```

Shell No. 4 Shell Shell No. 3 Shell No. 6 Shell No. 2 Shell No. 5

03:00 pm 12.12.2007

Memory consumption against time for a given job

additional job information via ALICE MonaLisa repository. Output is stored in ROOT trees which can be browsed offline

The screenshot shows a Java IDE with two windows. The left window displays a scatter plot titled "job.fMem_usage.fValue:job.fMem_usage.fTimestamp (fJobID==8681205)". The y-axis is labeled "job.fMem_usage.fValue" and ranges from 9 to 12. The x-axis is labeled "job.fMem_usage.fTimestamp" and ranges from 1197.456 to 1197.462, with a multiplier of $\times 10^9$. The plot shows a clear upward trend in memory usage over time. The right window shows a text log file titled "JStore_2007_12_12.log" with columns for timestamp, host, user, and various system metrics.

Timestamp	Host	User	Metric	Value
1197456471977	GSI	ALICE::GSI::LCG_Jobs	disk_free	205584.098
1197456471977	GSI	ALICE::GSI::LCG_Jobs	disk_total	205904.727
1197456471977	GSI	ALICE::GSI::LCG_Jobs	disk_usage	1.0
1197456471977	GSI	ALICE::GSI::LCG_Jobs	disk_used	320.629
1197456471977	GSI	ALICE::GSI::LCG_Jobs	mem_usage	10.7
1197456471977	GSI	ALICE::GSI::LCG_Jobs	open_files	148.0
1197456471977	GSI	ALICE::GSI::LCG_Jobs	page_faults_maj	0.0
1197456471977	GSI	ALICE::GSI::LCG_Jobs	page_faults_min	1491422.0
1197456471977	GSI	ALICE::GSI::LCG_Jobs	rss	1799372.0
1197456471977	GSI	ALICE::GSI::LCG_Jobs	run_ksi2k	42559.2
1197456471977	GSI	ALICE::GSI::LCG_Jobs	run_time	15420.0
1197456471977	GSI	ALICE::GSI::LCG_Jobs	virtualmem	2864576.0
1197456471977	GSI	ALICE::GSI::LCG_Jobs	workdir_size	14.195
1197456471978	GSI	ALICE::GSI::LCG_Jobs	host	lxb283.gsi.de
1197456471978	GSI	ALICE::GSI::LCG_Jobs	job_user	aliproduct
1197456471978	GSI	ALICE::GSI::LCG_Jobs	status	10.0
1197456472037	GSI	ALICE::GSI::LCG_Jobs	host	lxb283.gsi.de
1197456472037	GSI	ALICE::GSI::LCG_Jobs	job_user	aliproduct
1197456472037	GSI	ALICE::GSI::LCG_Jobs	status	10.0
1197456472051	GSI	ALICE::GSI::LCG_Jobs	cpu_ksi2k	36495.48
1197456472051	GSI	ALICE::GSI::LCG_Jobs	cpu_time	13223.0

```
end with ';', '@':abort > @
root [6] logtree->Draw("job.fMem_usage.fValue", "fJobID==8681205")
<TCanvas::MakeDefCanvas>: created default TCanvas with name c1
(Long64_t)76
root [7] logtree->Draw("job.fMem_usage.fValue:job.fMem_usage.fTimestamp", "fJobID==8681205")
<TCanvas::MakeDefCanvas>: created default TCanvas with name c1
(Long64_t)76
root [8] logtree->Draw("job.fMem_usage.fValue:job.fMem_usage.fTimestamp", "fUser.fData==\"aliproduct\"")
<TCanvas::MakeDefCanvas>: created default TCanvas with name c1
(Long64_t)7588
root [9] logtree->Draw("job.fMem_usage.fValue", "fUser.fData==\"aliproduct\"")
<TCanvas::MakeDefCanvas>: created default TCanvas with name c1
(Long64_t)7588
root [10] logtree->Draw("job.fMem_usage.fValue:job.fMem_usage.fTimestamp", "fJobID==8681205")
<TCanvas::MakeDefCanvas>: created default TCanvas with name c1
(Long64_t)76
root [11] □
```

Text log file, collected by the Java client of MonaLisa

Examples of TTree::Draw() commands for plotting

outview

- distribute RGLite via ROOT
- more discussion and testing with community concerning usage and setup of local and distributed PROOF clusters
- implementation of new features after user feedback up to Release version
- proceed with PROOF log and debug interface until Release version will be published