ALICE analysis at GSI (and FZK)

Kilian Schwarz

WLCG @ CHEP 07

ALICE T2 use cases (see computing model)

Three kinds of data analysis

Fast pilot analysis of the data “just collected” to tune the first reconstruction at CERN Analysis Facility (CAF)

Scheduled batch analysis using GRID (Event Summary Data and Analysis Object Data)

End-user interactive analysis using PROOF and GRID (AOD and ESD)

CERN

Does: first pass reconstruction

Stores: one copy of RAW, calibration data and first-pass ESD’s

T1

Does: reconstructions and scheduled batch analysis

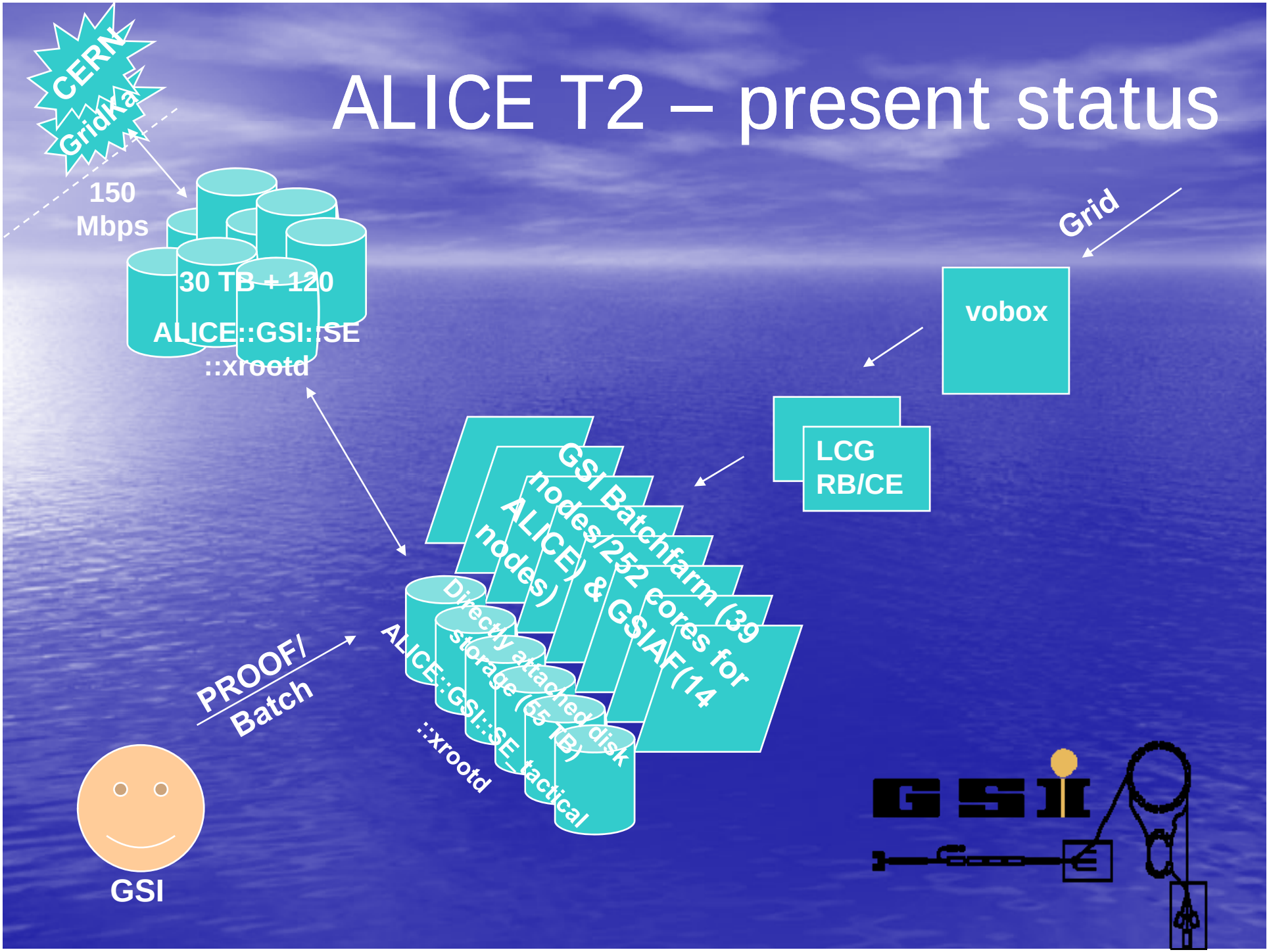
Stores: second collective copy of RAW, one copy of all data to be kept, disk replicas of ESD’s and AOD’s

T2

Does: simulation and end-user interactive analysis

Stores: disk replicas of AOD’s and ESD’s

ALICE T2 – present status



ALICE T2 – short term plans

- Extend GSIAF to all 39 nodes
- Study coexistence of interactive and batch processes on the same machines. Develop possibility to increase/decrease the number of batch jobs on the fly to give advantage to analysis.
- Add newly bought fileservers (about 120 TB disk space) to ALICE::LCG::SE::xrootd

Plans for the Alice Tier 2&3 at GSI:

Year	2007	2008	2009	2010	2011
ramp-up	0.4	1.0	1.3	1.7	2.2
CPU (kSI2k)	400/260	1000/ 660	1300/ 860	1700/ 1100	2200
Disk (TB)	120/80	300/200	390/260	510/340	660
WAN (Mb/s)	100	1000	1000	1000	...

- Remarks:
- **2/3** of that capacity is for the **tier 2** (ALICE central, fixed via WLCG MoU)
- **1/3** for the **tier 3** (local usage, may be used via Grid)
- according to the Alice computing model no tape for tier2
- tape for tier3 independent of MoU
- hi run in October -> **upgrade operational: 3Q** each year

Computing for Alice at GSI

(Proposal)

(Marian Ivanov)

Priorities (2007-2008)

- Detector calibration and alignment (TPC-ITS-TRD)
  - First test – Cosmic and Laser – October 2007
  - To be ready for first pp collision
- First paper
  - Time scale - Depends on success of October tests
  - Goal : ~ 1 week (statistic about 10^4-10^5 events)
- ==> **The calibration and alignment has the TOP priority (2007-2008)**

Assumptions

- CPU requirements – Relative
  - Simulation ~ 400 a.u
  - Reconstruction ~ 100 a.u
  - Alignment ~ 1 a.u
  - Calibration ~ 1 a.u
- To verify and improve the calibration and alignment several passes through data are necessary
    - The time scale for one iteration ~ minutes, hours ==>
- The calibration and alignment algorithms should be decoupled from the simulation and reconstruction
- The reconstruction algorithm should be repeated after retuning of the calibration

Assumptions

- Type of analysis (requirements)
- First priority
  - Calibration of TPC – 10 ^4 -10^5 pp
  - Validation of the reconstruction - 10^4-10^5 pp
  - Alignment TPC, TPC-ITS – 10^5 pp + 10^4-10^5 cosmic

Assumptions

- Alice test in October – (in one month)
  - Full stress test of system
  - Significant data volume
    - ~20 Tby of raw data from test of 2 sectors (2006)
    - Bottleneck (2006) – The processing time given by time of the data access - CPU time negligible
- We should be prepared for different scenarios
  - We would like to start with the data copied at GSI and reconstruct/calibrate/align locally, later switch to GRID (The same we did in 2006)
  - This approach enables several fast iteration over data

data transfers CERN GSI

- motivation: calibration modell and algorithms need to be tested before October
- test the functionality of current T0/T1 → T2 transfer methods.
- At GSI the CPU and storage resources are available, but how do we bring the data here ?

analysis of TPC test data

preconditon:

- copy to GSI:
  - store at ALICE::GSI::SE::xrootd

out2547.list	Cosmic Scan	A0&1	77
out2548.list	Cosmic Scan	A0&1	67
out2557.list	Cosmic Scan	A0&1	82
out2574.list	Cosmic Stability	A0&1	265
out2660.list	Cosmic Stability	A4&5	313
out2641.list	Cosmic Scan	A4&5	138
out2642.list	Cosmic Scan	A4&5	97
out2643.list	Cosmic Scan	A4&5	224

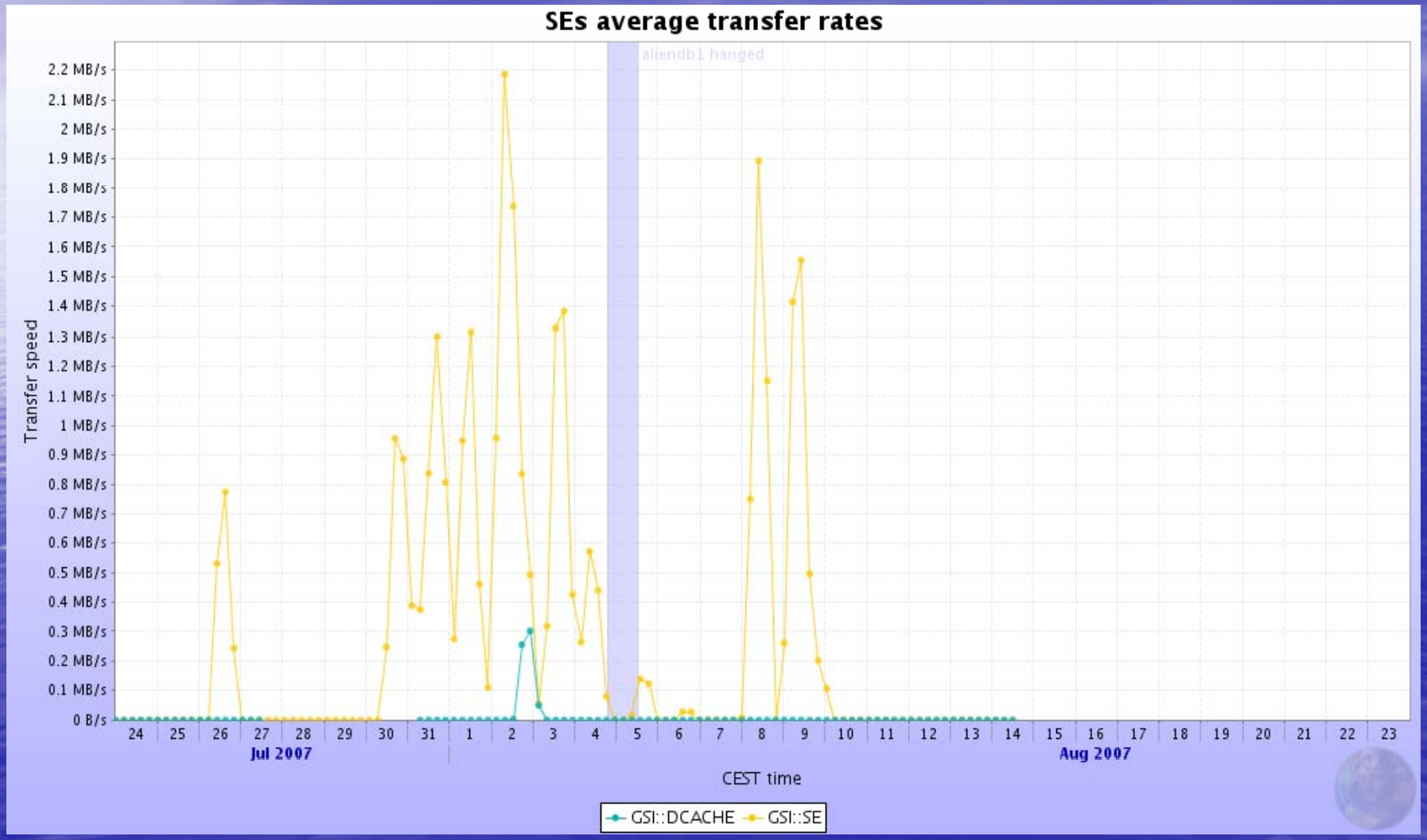
...

test data transfer to T2 and test SE

Laser:

out2572.list	31
out2657.list	171
out2728.list	195
out2798.list	215
out2906.list	90
out3094.list	6 directories
out3189.list	4 directories
out2612.list	114
out2686.list	177
out2746.list	41
out2851.list	Job <167677> - 345

data transfer CERN GSI



Software development

- Write component
- Software validation - Sequence:
  1) Local environment (first filter)
     1) Stability – debugger
     2) Memory consumption – valgrind, memstat (root)
     3) CPU profiling - callgrind, vtune
     4) Output – rough, quantitative – if possible
  2) PROOF
     1) For rapid development – fast user feedback
     2) Iterative improvement of algorithms, selection criterias ...
     3) Improve statistic
  3) production using GRID/ALIEN
     1) Improve statistic
  4) alternative scenario: local batch system
     1) Memory consumption – valgrind, memstat
     2) CPU profiling
     3) Output - better statistic

analysis of test TPC data

- using various analysis techniques
  - local batch farm at GSI (read from ALICE::GSI::SE)
  - PROOF@GSI (GSIAF) – copy data to ALICE::GSI::SE_tactical (PROOF cluster – directly attached disks)
  - Grid@GSI (submit to AliEn – jobs should arrive at GSI – since this is where the data are)

Proposal

- Algorithmic part of our analysis, calibration software should be independent of the running environment
  - TPC calibration classes (components) as example (running, tuning OFFLINE, used in HLT, DAQ and Offline)
- Analysis and calibration code should be written following a **component** based model
  - TSelector (for PROOF) and AliAnalysisTask (at GRID/ALIEN) – just simple wrapper

analysis of test TPC data

- using Grid methods
  - analysis partition: GSI should be included
  - JDL: specify that CERN CE should not be used since data of interest are stored at CERN and GSI. Job should then take the other alternative.

analysis of TPC test data

- Executable="tpcRecAlienLocal.sh";
- InputFiles={"LF:/afs/cern.ch/alice/tpctest/AliRoot/HEAD/TPC/recTPC.C","/afs/cern.ch/alice/tpctest/AliRoot/HEAD/TPC/AnalyzeESDtracks.C"};
- InputDataCollection="LF:/alice/cern.ch/user/h/haavard/jdl/runs/run$1.xml";
- InputDataList="tpcRec.xml";
- InputDataListFormat="xml-single";
- OutputArchive={"log_archive:stdout,stderr,*.log@Alice::CERN::castor2",
- "root_archive.zip:AliESD*.root,TPC.Rec*.root@Alice::CERN::castor2",
- "tpc_archive.zip:FitSignal.root,TimeRoot.root,TPCsignal.root,TPCtracks.root,TPCdebug.root@Alice::CERN::castor2"};
- OutputDir="/alice/cern.ch/user/h/haavard/$2/$1/#alienfirstfilename#_dir";
- Split="file";
- SplitArguments = {"#alienfirstfilename#"};
- Arguments = " $1 $2 $3 ";
- Workdirectorysize={"4000MB"};
- Packages={"VO_ALICE@APISCONFIG::V2.1"};
- Requirements = (!other.CE=="ALICE::CERN::LSF");

analyse TPC test data

tpcrecAlienLocal.sh

…

- command aliroot -q -b "$ALICE_ROOT/TPC/recTPC.C($runtype)"
- command aliroot -q -b "$ALICE_ROOT/TPC/AnalyzeESDtracks.C+($run)"

…

recTPC.C

…

- AliReconstruction rec;
- rec.SetDefaultStorage("local://$ALICE_ROOT");
- rec.SetRunReconstruction("TPC");
- rec.SetFillESD("TPC");
- rec.Run(0);

…

Analysis of ESDs

Ana Marin, Sylwester Radomski

28.08.2007

ALICE Analysis Meeting, GSI

A few numbers – pp LHC running year

- Total number of events = 10^9

- Number of events per second = **100**
- Number of seconds per day = 60*60*24 = **86 400**
- Number of events per day = **8.6M**
- Events size (ESD) = **0.2 MB**
- Data volume per day = **1.7 TB**

- Number of days = **150**
- Total number of events = **1.3 G**
- Total data size (ESD) = **250 TB**

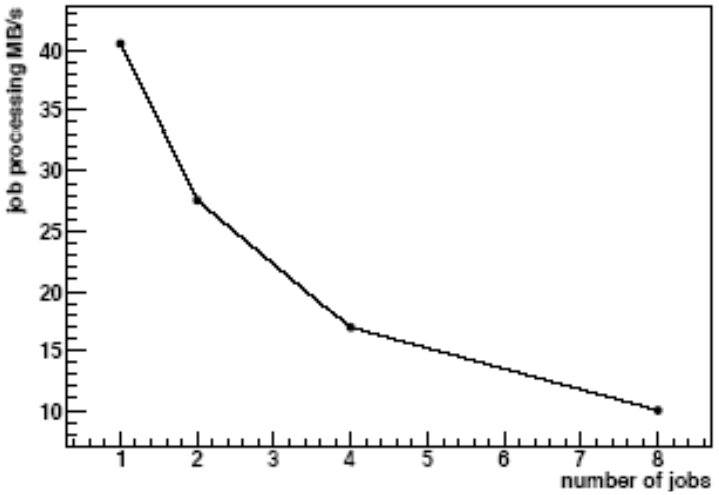
Computing TDR, p. 62

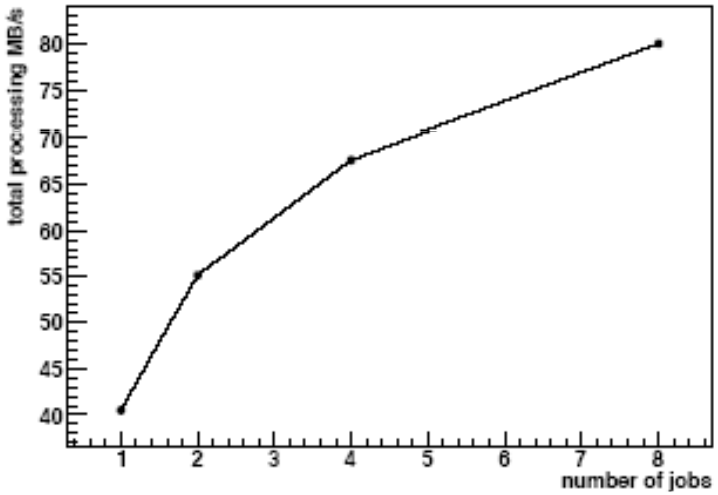
Physics Data Challenge 2006 (pp-minbias)

- Focus on central barrel events
- Events per file (job) = 100 (1 s)
- Files per unit = 1000
- Number of events (pdc06) $\sim$ **52 M** events
- Number of ESD files $\sim$ 520 k
- Maximum number of analyzed events:
  - Physics Week, Feb. 2007: **4.6 M (1.8 local + 2.8 grid)**
  - Currently: **9.05 M grid** (improvements in framework stability)
- PDC06 corresponds to 6 days of LHC running
- Number of **analyzed ESDs** corresponds to 1 day

Local Processing Architecture

- AMD Opteron processor, 2x2, 1.5 TB disk space, raid 5
- Data are read only from the local disk – no network traffic





- Total speed $\sim$ 400 ev/s, CPU utilization $\sim$ 30%
- Analyzing 1.5 TB set takes around 6 hours.
- 4 AliRoot simulations in addition do not change the numbers
- **Analysis dominated by data input stream**

Experience with GRID

Staging is central part of the analysis
but is not transparent in AliEn framework

- one not staged file triggers "job kill"
- **Grid performance**
  - job yield – 97%-99.7% (a respectable number for distributed computing)
  - total job time – 0.8 sec/ev (one local machine equivalent of 300 grid jobs)
  - cpu job time – 0.02 sec/ev (CPU is idle for 97% of the time)
  - local: total time 0,003 sec/ev, cpu 0.01 sec/ev
- Possible reasons for the grid performance
  - analysis speed dominated by reading data from a disk server ?
  - inefficient network connection ?
  - **detailed understanding needed**

Analysis at GridKa T1

- centrally steered batch analysis
- preconditions:
  - functioning SE with xrootd interface
    - technically feasible with dCache
  - GridKa hesitates so far to open xrootd ports to the world on all xrd doors
    - security concept should be rediscussed

summary

- at GSI ALICE analysis will be done using Grid, PROOF, and local batch.
- batch cluster and GSIAF are on the same machines. We will be able to increase/decrease batch/Grid jobs dynamically to give advantage to GSIAF/PROOF analysis if needed
- data transfer to GSI has still to be exercised
- analysis performance on the Grid has to improve with respect to data I/O.
- xrootd access to T1 centres has to be discussed