Full Online Event Reconstruction at HERA-B

Andreas Gellrich
DESY Zeuthen, Germany
HERA-B Collaboration

Monday February 7, 2000

Parallel Session E
Commodity Hardware and Software &
Integration in Farm and Large Systems

Contents

Introduction

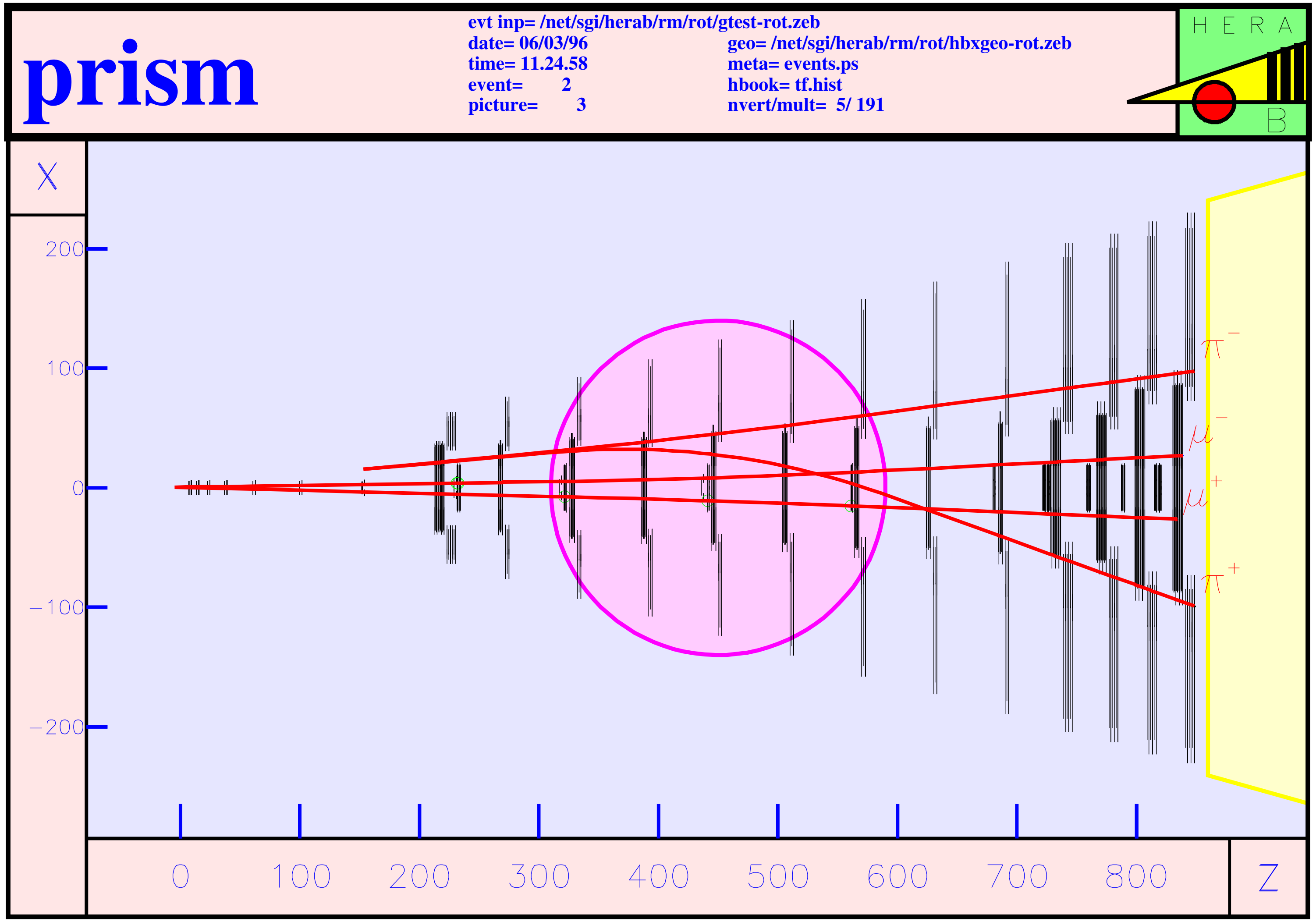
- **Physics:** → **study of B-decays**
  - CP violation in $B^0 \rightarrow J/\psi\ K^0_s$
  - needs O(1000) reconstructed B^0
  - $\sigma_{b\bar{b}} / \sigma_{inel}$ = 12 nb / 13 mb = 10^{-6}

- **Target:** → **wire target in HERA proton beam halo**
  - 10 MHz event rate @ 4 interactions
  - << 1 Hz interesting physics

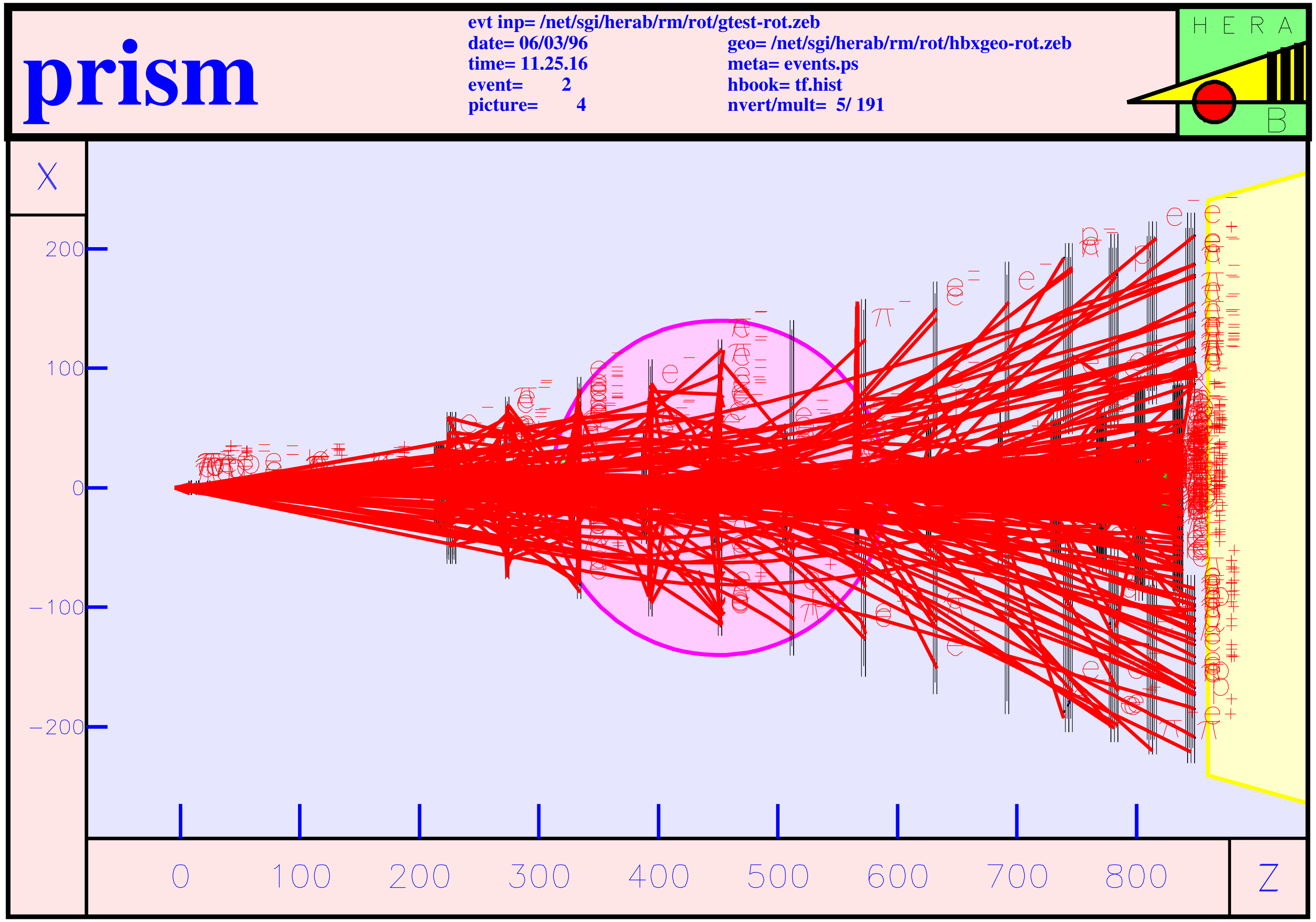
- **Detector:** → **read-out channels: ~600000**
  - tracks / event: ~120
  - occupancy: <=20%

- **Analysis:** → **full online event reconstruction**
  - processing time: 4 sec/event
  - logging rate: 20Hz, 2 MB/sec
  - data volume: 20 TB/year

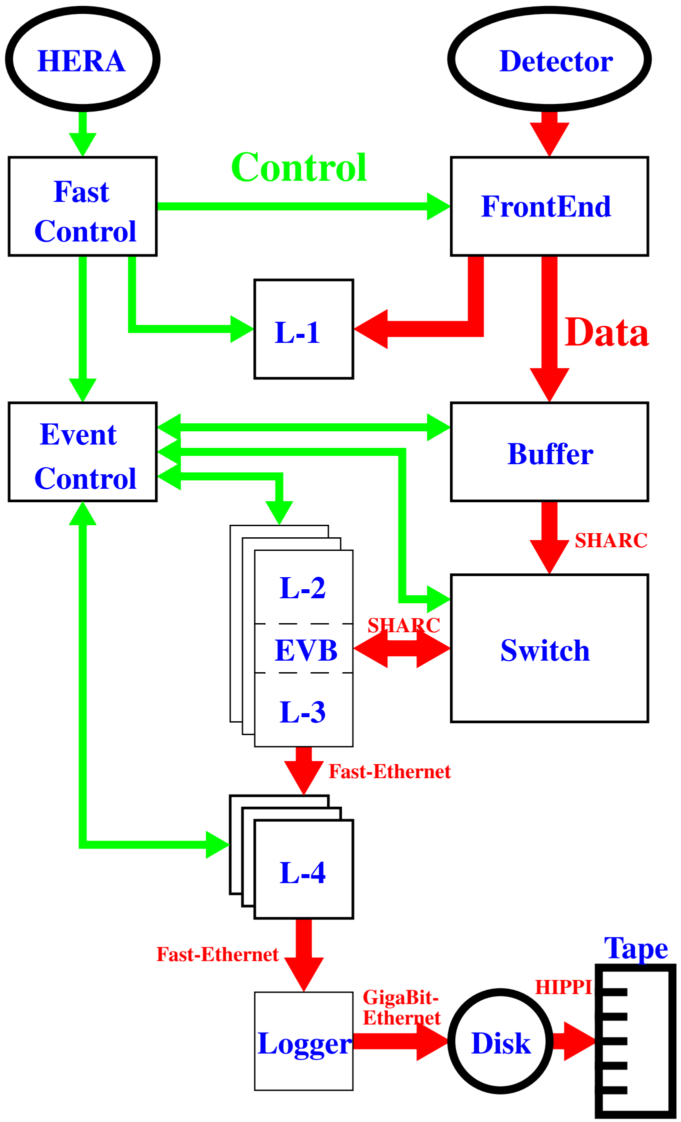
“Golden” Decay



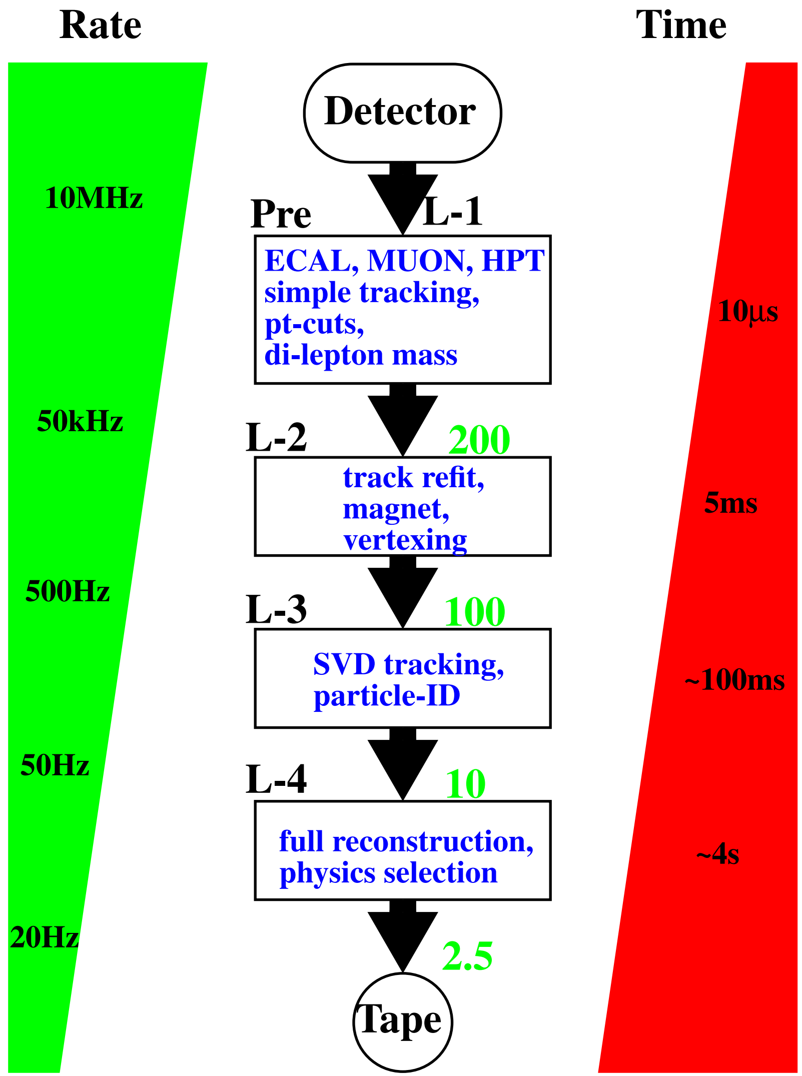
“Golden” Decay (cont’d)



DAQ & Trigger



HERA
Detector
Fast Control
Control
FrontEnd
L-1
Data
Event Control
Buffer
SHARC
L-2
EVB
L-3
SHARC
Switch
Fast-Ethernet
L-4
Fast-Ethernet
Tape
GigaBit-Ethernet
HIPPI
Logger
Disk



Rate
Time
10MHz
50kHz
500Hz
50Hz
20Hz
Detector
Pre
L-1
ECAL, MUON, HPT
simple tracking,
pt-cuts,
di-lepton mass
10µs
L-2
200
track refit,
magnet,
vertexing
5ms
L-3
100
SVD tracking,
particle-ID
~100ms
L-4
10
full reconstruction,
physics selection
~4s
2.5
Tape

Purpose and Tasks

- **Full Online Event Reconstruction:**
  - 50 Hz * 4 sec = 200 nodes
  - → multi-processor farm

  - run offline developed software online
  - → provide appropriate software environment
  - → make offline developments online-compliant (I/O)
- **Event Classification:**
  - mark events due to there physical contents
  - to be used in event directories
- **Final Event Selection:**
  - L-4 trigger step
- **Data Logging:**
  - add reconstruction information to event
  - send events to logger

Purpose and Tasks (cont'd)

- **Data Quality Monitoring:**
  - use availability of data
  - use high statistics
  - → central collection (gathering) of histograms

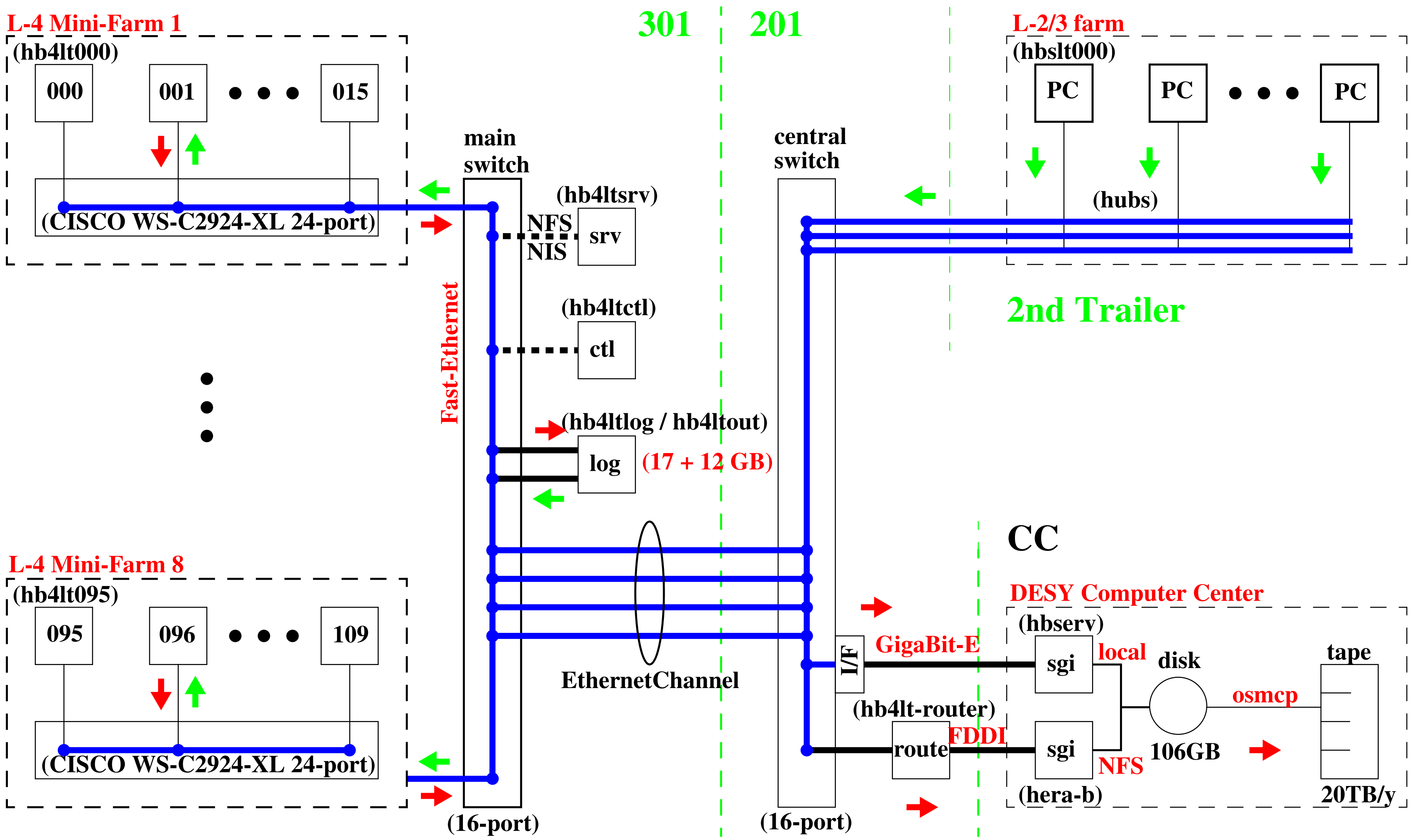
- **Preparation of Data for Calibration and Alignment:**
  - use availability of data
  - use high statistics
  - → central collection (gathering) of data
  - → feedback system for database constants

- **Event Data Reprocessing in Shutdown Periods:**
  - use vast processing power of the farm
  - → copy entire data files to nodes for processing

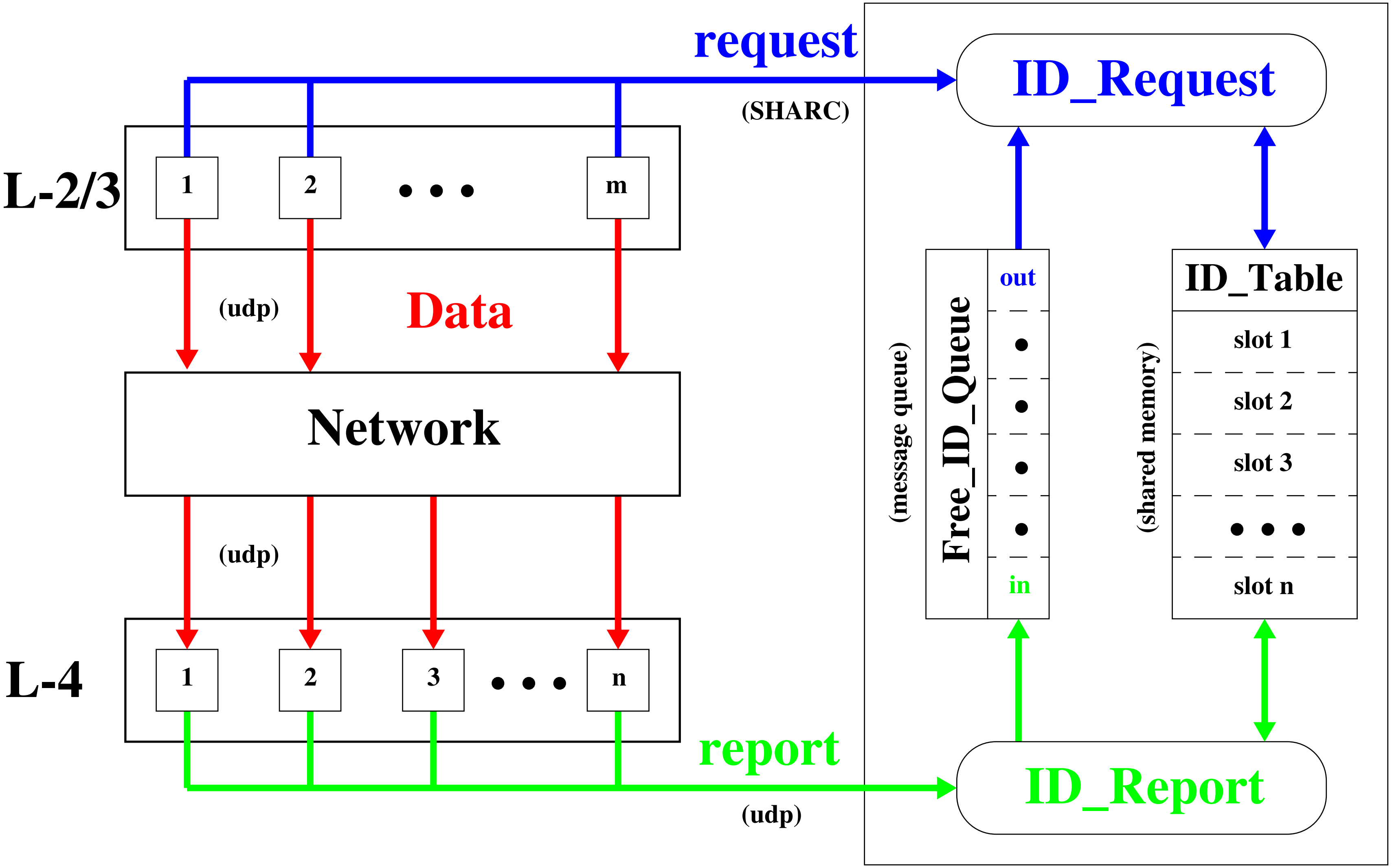
Implementation

- **Processor Nodes:**
  - 80 dual-PIII/500MHz
  - 20 single-PII/450 MHz
  - 256 MB SDRAM, 13 GB disks
- **Network:**
  - Fast-Ethernet
  - Gigabit-Ethernet uplink
  - CISCO-switches
  - 8 switched “mini-farms”
- **Services:**
  - NFS/NIS service (executables, files)
  - slow control (http)
  - local logging (36 GB, DLT7000)
- **Operating System:**
  - Linux (S.u.S.E.)

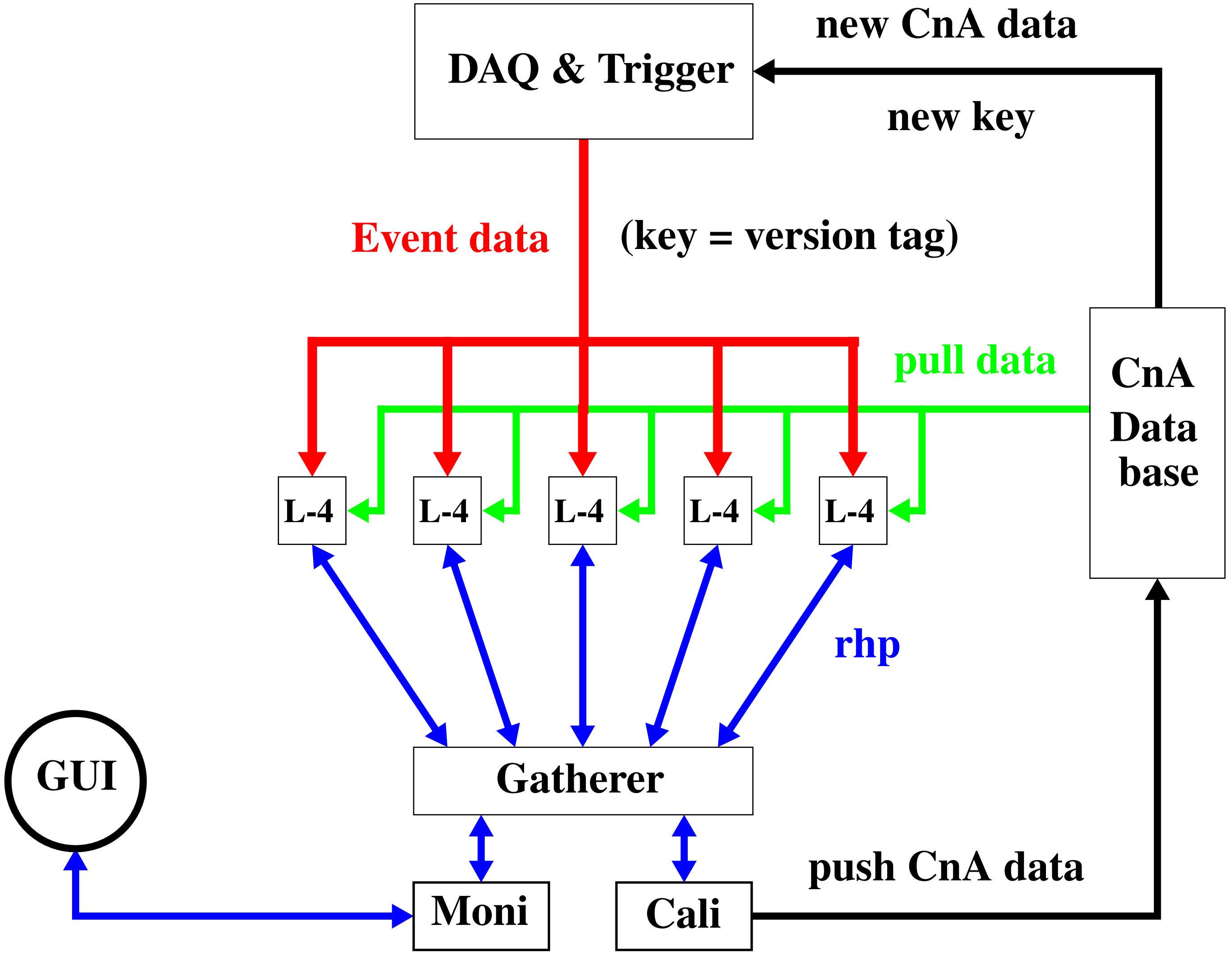
Farm Network



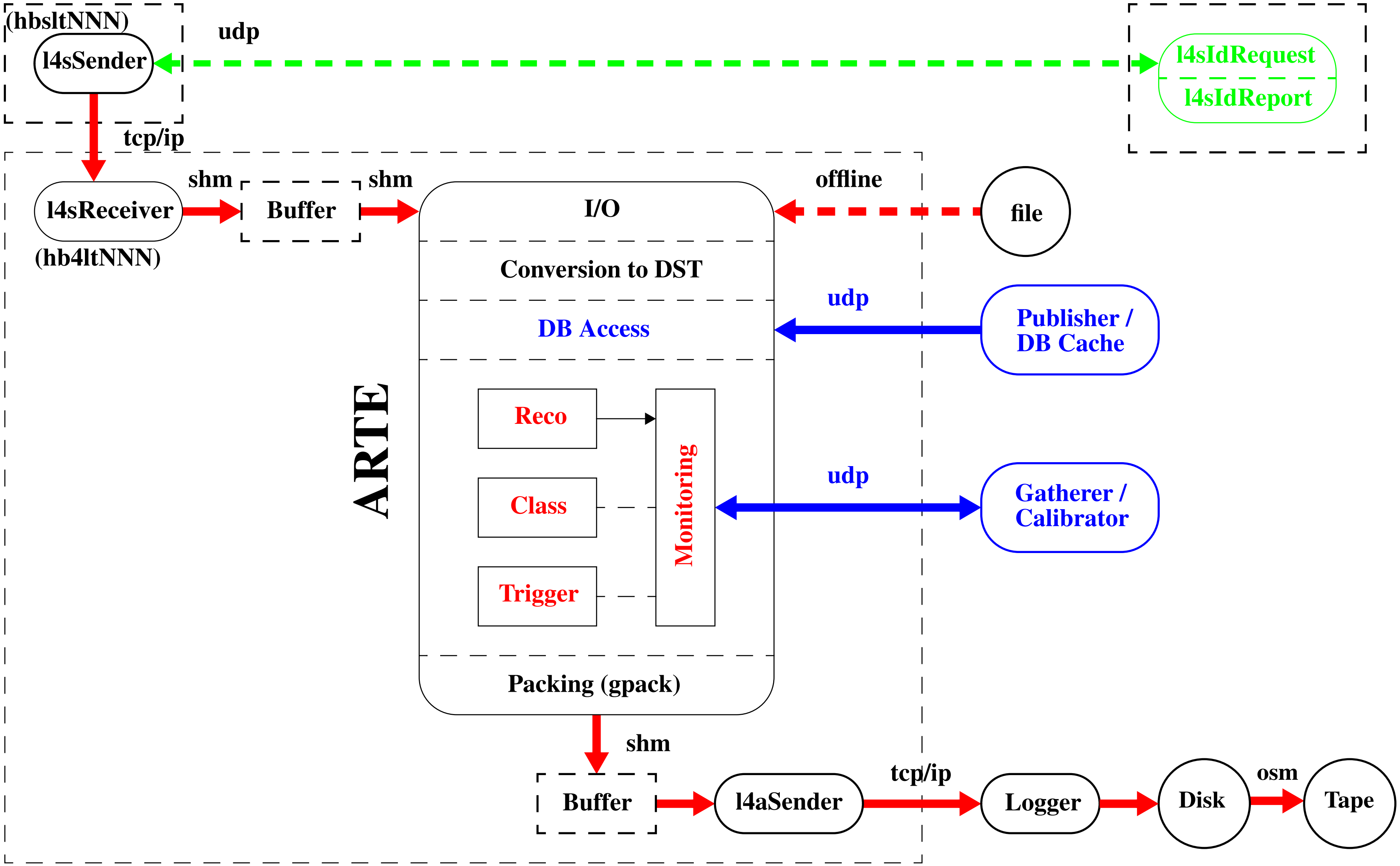
Farm Event (Fast) Control



Calibration & Alignment

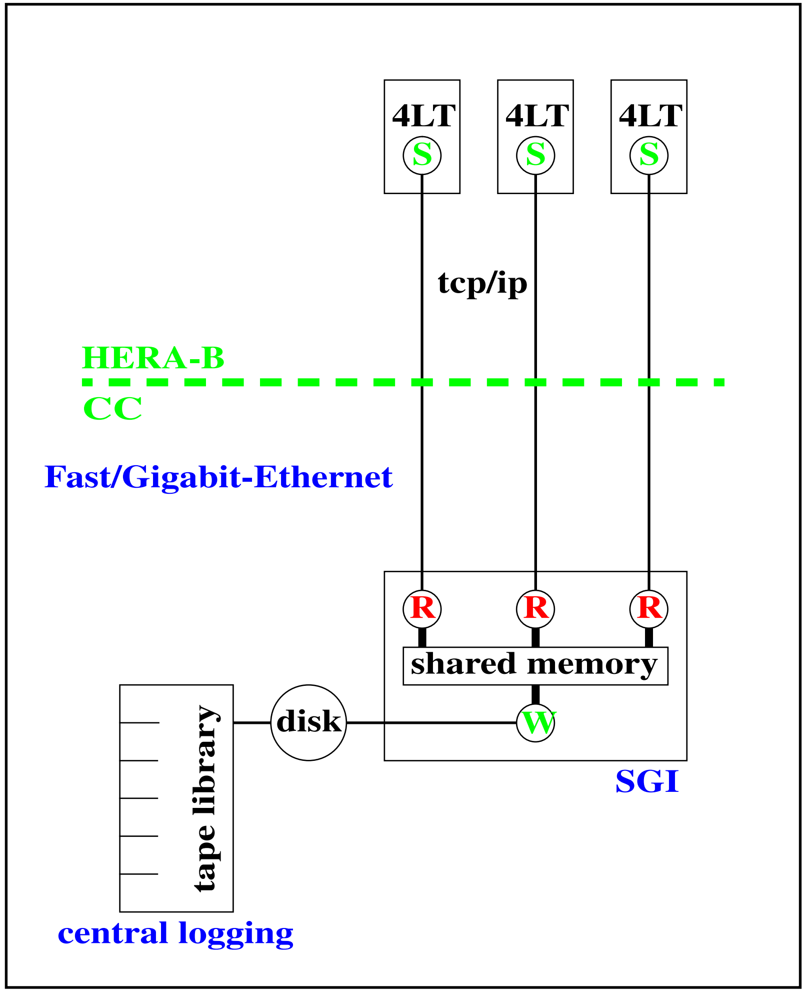


Farm Node Processes

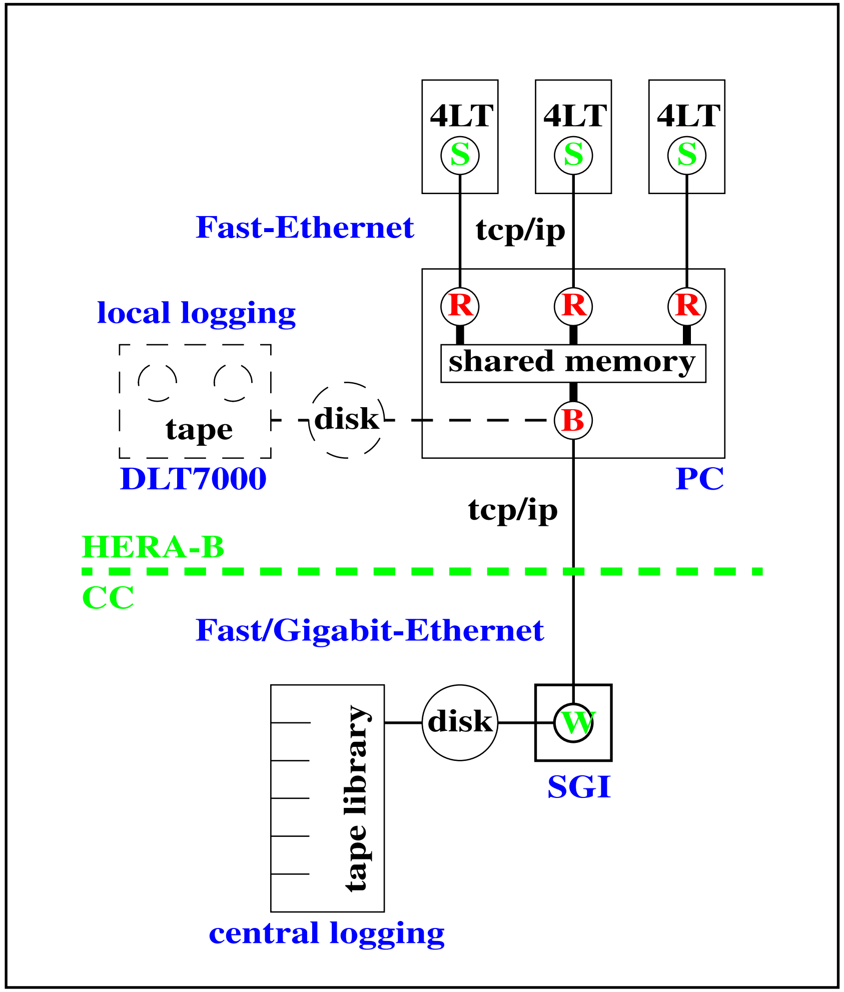


Event Data Logging

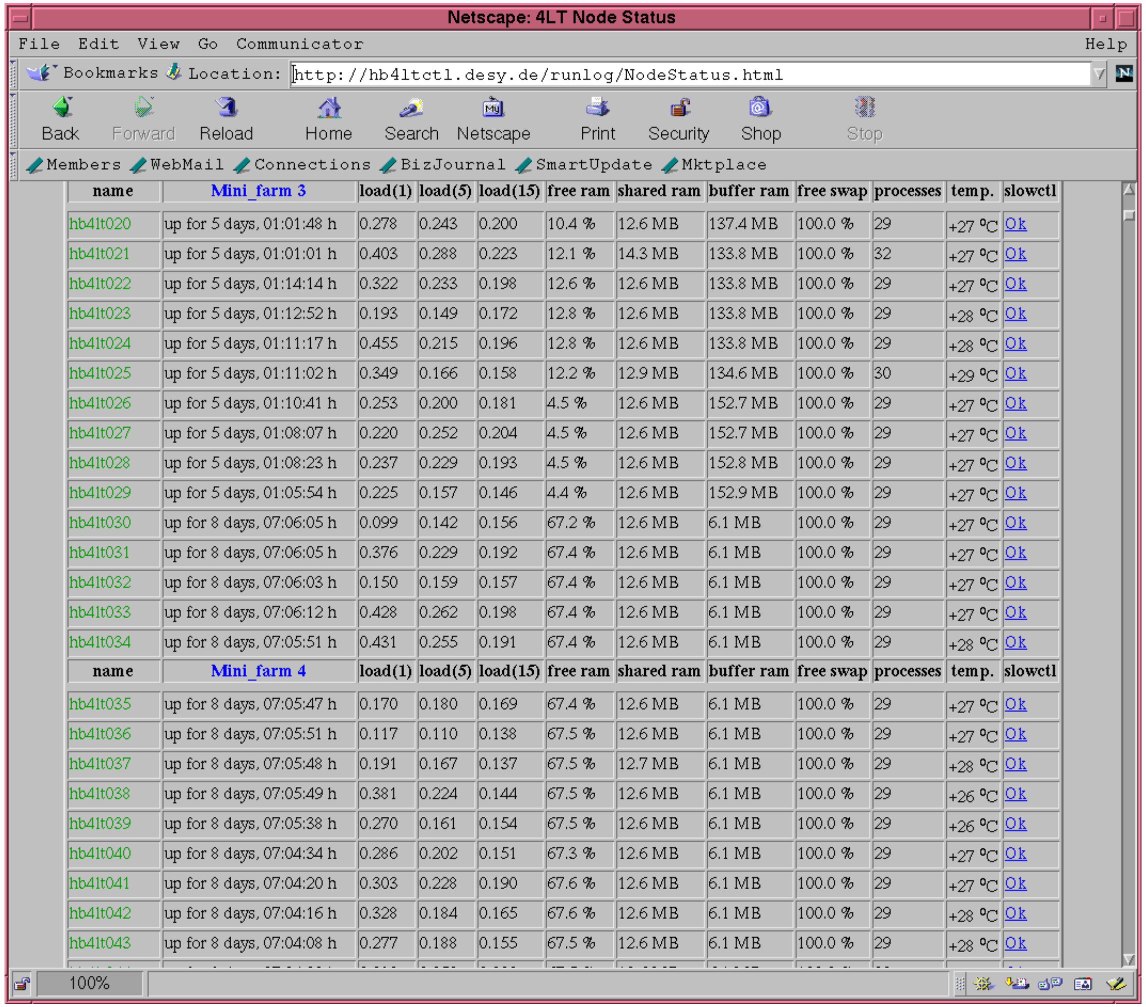
Central Logging & Archiving



Local Logging & Central Archiving



Farm Slow Control



name	Mini_farm 3	load(1)	load(5)	load(15)	free ram	shared ram	buffer ram	free swap	processes	temp.	slowctl
hb4lt020	up for 5 days, 01:01:48 h	0.278	0.243	0.200	10.4 %	12.6 MB	137.4 MB	100.0 %	29	+27 °C	Ok
hb4lt021	up for 5 days, 01:01:01 h	0.403	0.288	0.223	12.1 %	14.3 MB	133.8 MB	100.0 %	32	+27 °C	Ok
hb4lt022	up for 5 days, 01:14:14 h	0.322	0.233	0.198	12.6 %	12.6 MB	133.8 MB	100.0 %	29	+27 °C	Ok
hb4lt023	up for 5 days, 01:12:52 h	0.193	0.149	0.172	12.8 %	12.6 MB	133.8 MB	100.0 %	29	+28 °C	Ok
hb4lt024	up for 5 days, 01:11:17 h	0.455	0.215	0.196	12.8 %	12.6 MB	133.8 MB	100.0 %	29	+28 °C	Ok
hb4lt025	up for 5 days, 01:11:02 h	0.349	0.166	0.158	12.2 %	12.9 MB	134.6 MB	100.0 %	30	+29 °C	Ok
hb4lt026	up for 5 days, 01:10:41 h	0.253	0.200	0.181	4.5 %	12.6 MB	152.7 MB	100.0 %	29	+27 °C	Ok
hb4lt027	up for 5 days, 01:08:07 h	0.220	0.252	0.204	4.5 %	12.6 MB	152.7 MB	100.0 %	29	+27 °C	Ok
hb4lt028	up for 5 days, 01:08:23 h	0.237	0.229	0.193	4.5 %	12.6 MB	152.8 MB	100.0 %	29	+27 °C	Ok
hb4lt029	up for 5 days, 01:05:54 h	0.225	0.157	0.146	4.4 %	12.6 MB	152.9 MB	100.0 %	29	+27 °C	Ok
hb4lt030	up for 8 days, 07:06:05 h	0.099	0.142	0.156	67.2 %	12.6 MB	6.1 MB	100.0 %	29	+27 °C	Ok
hb4lt031	up for 8 days, 07:06:05 h	0.376	0.229	0.192	67.4 %	12.6 MB	6.1 MB	100.0 %	29	+27 °C	Ok
hb4lt032	up for 8 days, 07:06:03 h	0.150	0.159	0.157	67.4 %	12.6 MB	6.1 MB	100.0 %	29	+27 °C	Ok
hb4lt033	up for 8 days, 07:06:12 h	0.428	0.262	0.198	67.4 %	12.6 MB	6.1 MB	100.0 %	29	+27 °C	Ok
hb4lt034	up for 8 days, 07:05:51 h	0.431	0.255	0.191	67.4 %	12.6 MB	6.1 MB	100.0 %	29	+28 °C	Ok
name	Mini_farm 4	load(1)	load(5)	load(15)	free ram	shared ram	buffer ram	free swap	processes	temp.	slowctl
hb4lt035	up for 8 days, 07:05:47 h	0.170	0.180	0.169	67.4 %	12.6 MB	6.1 MB	100.0 %	29	+27 °C	Ok
hb4lt036	up for 8 days, 07:05:51 h	0.117	0.110	0.138	67.5 %	12.6 MB	6.1 MB	100.0 %	29	+27 °C	Ok
hb4lt037	up for 8 days, 07:05:48 h	0.191	0.167	0.137	67.5 %	12.7 MB	6.1 MB	100.0 %	29	+28 °C	Ok
hb4lt038	up for 8 days, 07:05:49 h	0.381	0.224	0.144	67.5 %	12.6 MB	6.1 MB	100.0 %	29	+26 °C	Ok
hb4lt039	up for 8 days, 07:05:38 h	0.270	0.161	0.154	67.5 %	12.6 MB	6.1 MB	100.0 %	29	+26 °C	Ok
hb4lt040	up for 8 days, 07:04:34 h	0.286	0.202	0.151	67.3 %	12.6 MB	6.1 MB	100.0 %	29	+27 °C	Ok
hb4lt041	up for 8 days, 07:04:20 h	0.303	0.228	0.190	67.6 %	12.6 MB	6.1 MB	100.0 %	29	+27 °C	Ok
hb4lt042	up for 8 days, 07:04:16 h	0.328	0.184	0.165	67.6 %	12.6 MB	6.1 MB	100.0 %	29	+28 °C	Ok
hb4lt043	up for 8 days, 07:04:08 h	0.277	0.188	0.155	67.5 %	12.6 MB	6.1 MB	100.0 %	29	+28 °C	Ok

- **Requirements:**
  - shift crew usage
  - remote access
  - status control
  - temperature control
  - monitoring
- **Implementation:**
  - one process per node
  - one file per node
  - sysinfo
  - lm_sensors.o
  - /proc/sensors
  - http-service
- **Alternatives:**
  - CAN-bus
  - central slow control

Performance

- **Online Reconstruction:**
  - not all detector components in yet
  - 4 sec / event seems to be feasible

- **Event Data Transfer:**
  - tcp/ip-based protocol
  - → one socket connection per node
  - → 1.3 kHz * 4 kB = 5 MB/sec to L-4
  - → 200 Hz * 4 kB = 1 MB/sec to disk
  - → 60 Hz * 130 kB = 8 MB/sec to disk

- **Logging & Archiving:**
  - Open Storage Manager osm
  - → <= 5 MB/sec net rate to tape
  - → 7 TB in 1999

Summary

- Full online event reconstruction at HERA-B is the goal.
- High statistics for monitoring, calibration, and alignment is needed.
    - → Bring applications to the data.

- 200 CPUs are installed.
- A cheap, scalable, flexible, powerful system was built.
- Intel-PCs with dual-Pentium boards and local disks are used.
- Linux is standard HEP platform for online and offline.

- Design parameters for data transfer were clearly exceeded.
- Processing time of 4 sec/ event is feasible.

→ Standard online event reconstruction has started at HERA-B.