# The Linux-PC Farm for Online Event Reconstruction of HERA-B


**Andreas Gellrich**
**DESY Zeuthen, Germany**


**1st Event Filter Farms Workshop at 3rd LHC Computing Workshop**

**Marseille, France**
**September 28 - October 1, 1999**

# Contents

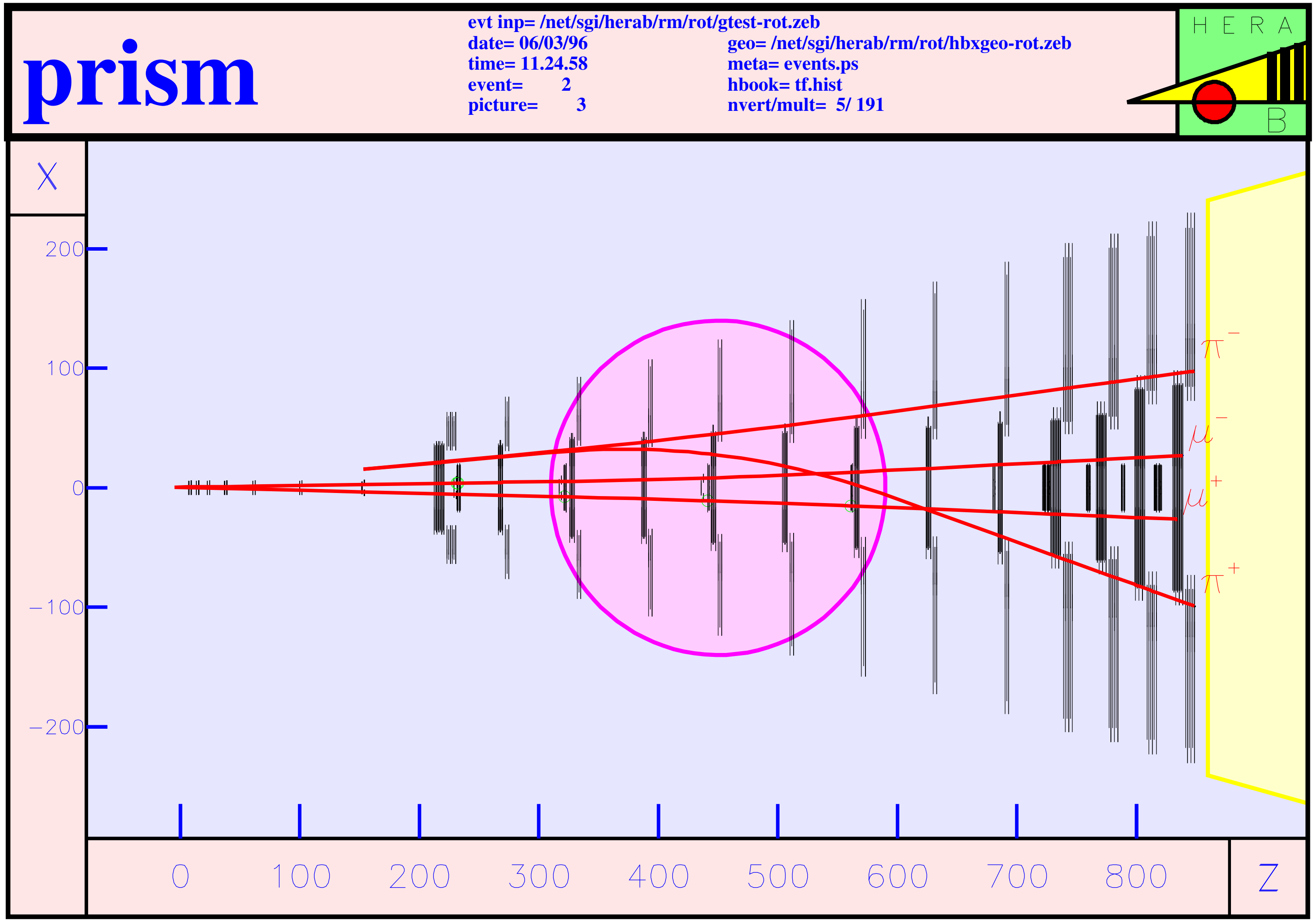
- Introduction
- HERA-B DAQ & Trigger System Overview
- Level-2/3 versus Level-4 Farm
- Details on Level-4 Farm
- Outlook & Summary

➔ Antonio Amorim: Databases for Calibration & Alignment at HERA-B

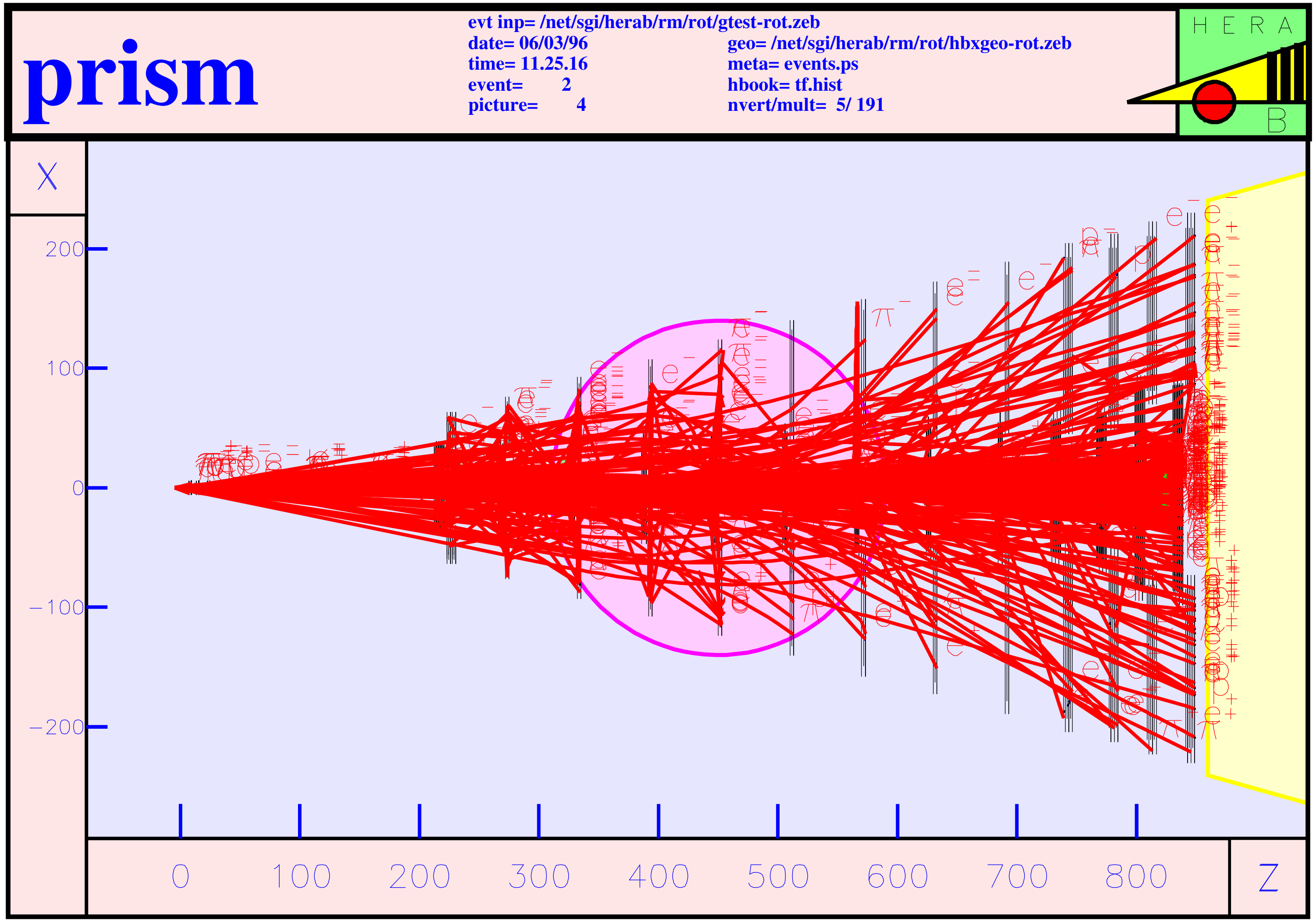
# Introduction

- **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

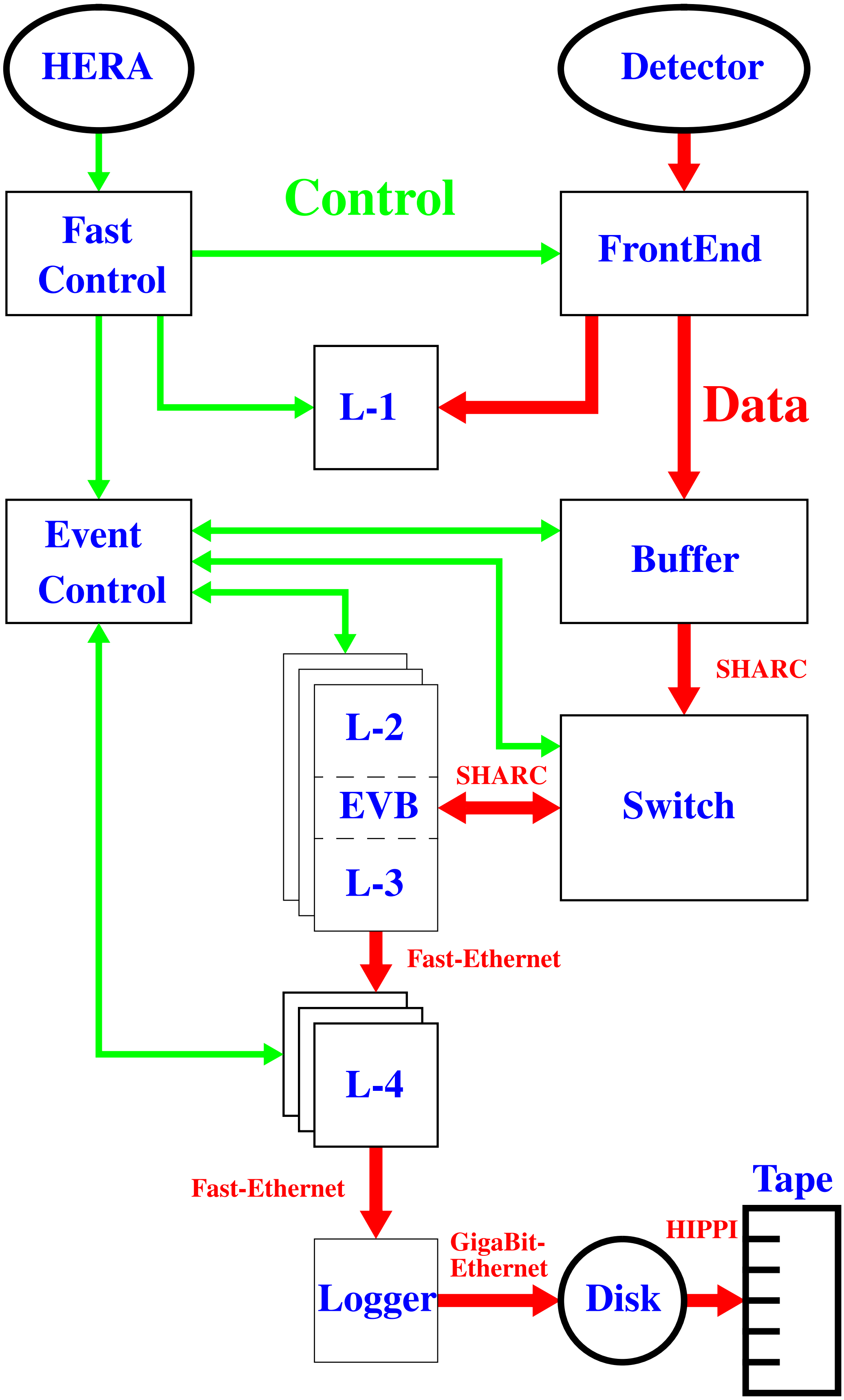
# Golden $B^0$-Decay



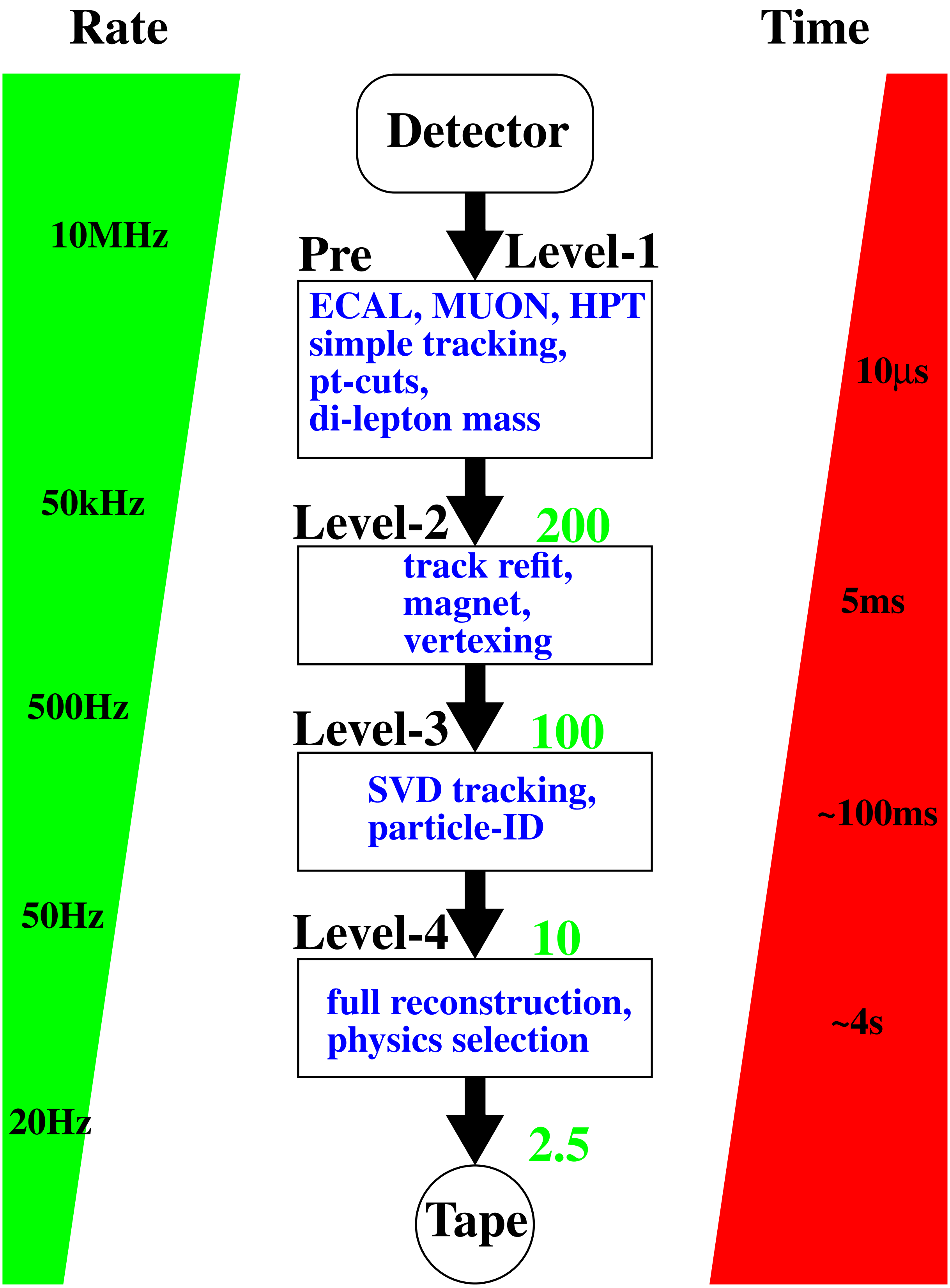
# Golden $B^0$-Decay (cont'd)



# DAQ & Trigger Scheme



# Rates & Times



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

# Triggering at HERA-B

- **Requirements:**
  - select << 1 Hz physics from 10 MHz multi-interaction events
- **Strategy:**
  - high B-mass: high-$p_T$ decay particles
  - long B-lifetime: detached vertices
- **Detectors:**
  - pad chambers: high-$p_T$ tracks
  - calorimeter: electrons
  - muon system: muons
  - tracking chambers: tracks, vertices
- **Levels:**
  - pre-trigger: Regions-of-Interest building (ECAL, MUON)
  - Level-1: simple (hardware) track finding (di-lepton mass)
  - Level-2: track re-fit, magnet tracking, secondary vertex finding in SVD
  - Level-3: SVD tracking outside RoI, particle-ID
  - Level-4: full online event reconstruction, event classification & selection

# Level-2/3 versus Level-4

- **Level-2:** → **latency and bandwidth dominated**
  - ◆ **RoI based (event fractions $O(1\%)$)**
  - ◆ **reduction factor ~100**
  - ◆ **high input rate of 50kHz**
  - ◆ **very low processing time $O$(1 ms)**
- **Level-3:** → **still latency and bandwidth dominated**
  - ◆ **full event unit available (after event building)**
  - ◆ **reduction factor ~10**
  - ◆ **still high input rate of ~500 Hz**
  - ◆ **low processing time $O$(100 ms)**
- **Level-4:** → **processing time dominated**
  - ◆ **full event reconstruction**
  - ◆ **reduction factor ~1**
  - ◆ **moderate inpute rate of 50 Hz**
  - ◆ **long processing time $O$(1 s)**

# Level-2/3 Farm Concept

- **Hardware:** → 100 + 140 nodes
  - PII / 300 MHz + PIII / 400 MHz, 64 MB RAM
  - floppy drive
  - NO hard disk (diskless)
  - cheap graphics card
  - Fast-Ethernet
- **Data link:** → bandwidth: 40 MB/sec, latency: 1 μsec
  - SHARC-to-PCI i/f cards including low level i/f driver
- **Slow control:**
  - CAN-bus card
- **Software:**
  - RedHat Linux
  - bootp to load set-up
  - Linux boot/root from floppy
  - one process environment

# Level-4 Farm Motivation

- **Online event reconstruction:**
  - derive quantities needed for calibration & alignment
  - avoid re-processing (processing time: ~4 sec/event)
    (data volume: 20 TB/year)
  - allow for immediate physics analysis

- **Further Level-4 tasks:**
  - event classification
  - final event selection (Level-4 trigger step)
  - data quality monitoring
  - calibration & alignment monitoring
  - event data logging to archive

# Level-4 Farm Concept

- **Basic requirements:**
  - performance (O(1 sec/event))
  - scalability (10-200 nodes)
  - flexibility (<100 kB/event>, 5 MB/sec)
  - costs (~1.5 kDM/CPU)
- **Hardware aspects:**
  - long term availability & support
  - space consumption, power supply, cooling
  - reliability
  - maintenance
- **Software aspects:**
  - operating system & development platform, compiler
  - long term availability & support
  - boot procedure
  - system control & monitoring
  - maintenance

# Level-4 Software Environment

- *"Physics"* software:
  - event reconstruction
  - event classification
  - Level-4 trigger step
  - (Monte Carlo)
  - (detector simulation (GEANT))

- Frame program: (Arte)
  - program control (interactive)
  - data I/O: (offline (file) vs. online (memory))
  - table structure
  - Fortran, C/C++
  - Unix platforms: IRIX_mips4, Linux_intel, (Solaris_sparc, HP-UX)

→ Give up separation between online and offline.

→ Use offline developments online.

# Level-4 Farm Studies

- **VME (embedded):**
  - e.g. PowerPC 604 / 100 MHz
  - e.g. Data-Strobe links (10 MB/s)
  - standard online system +
  - housing +
  - real-time capabilities +
  - expensive -
  - not used in offline environment -
  - link interface -

- **PC (WS):**
  - commodity solution +
  - off-the-shelf components +
  - high-end processors available +
  - well supported operating systems +
  - standard network solutions +
  - limited real-time capabilities (-)
  - housing (-/+)

# Level-4 Farm Implementation

- **Hardware:**
  - commodity / off-the-shelf products
  - high-end machines
  - Fast-Ethernet / GigaBit-Ethernet
  - modular network

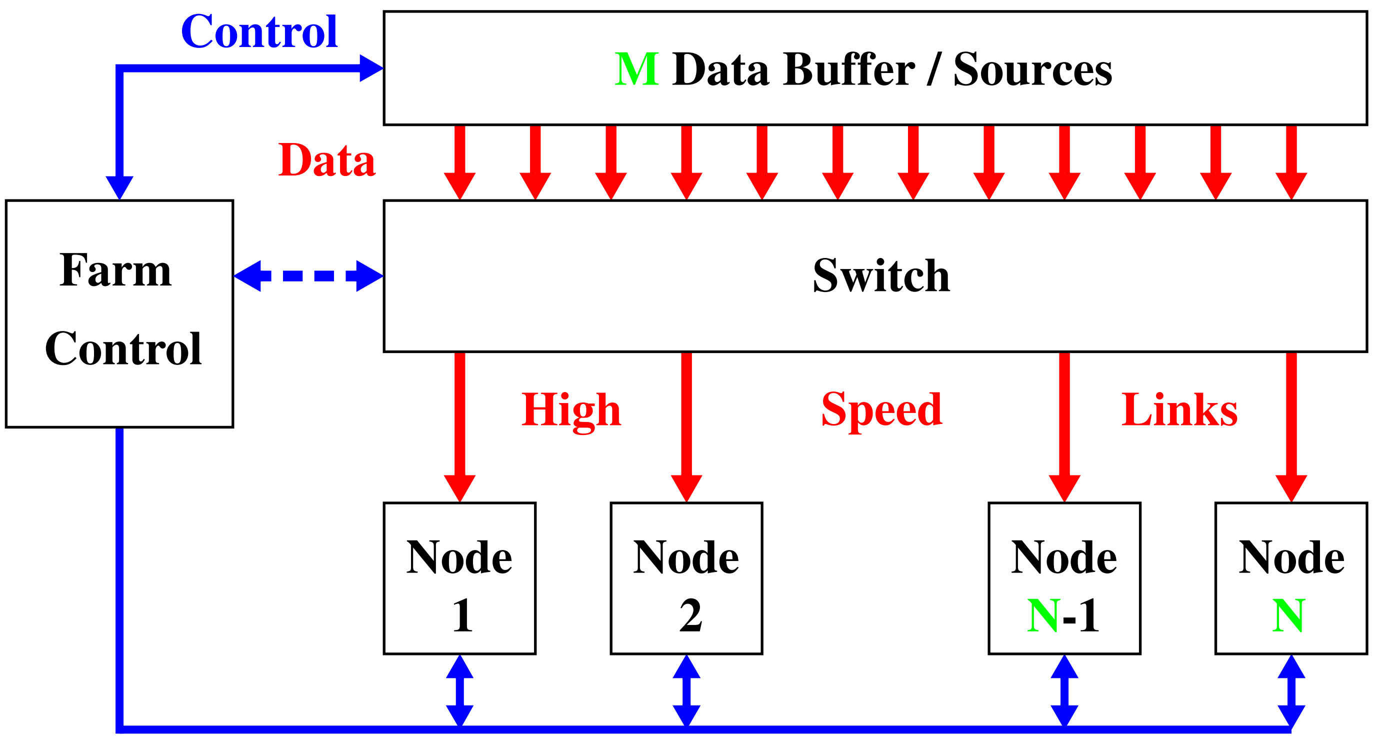
- **Farm control:**
  - event data from m Level-2/3 to n Level-4 nodes
  - monitoring

- **Node processes:**
  - multi-process environment
  - embedded "*physics*" software

- **Monitoring:**
  - collect data for alignment & calibration
  - monitor data flow
  - *remote* histogramming package

# General Farm Scheme

(Dynamic point-to-point connections between M sources and N destinations.)



# Level-4 Farm Hardware

- **Processor nodes:**
  - **Motherboard: Asus P2B(-D) (100 MHz)**
  - **CPU: Pentium II/III / 400/450 MHz**
  - **L2-cache: 512 kB**
  - **RAM: 64 / 2 x 128MB SDRAM**
  - **Hard disk: 2.5 / 4.3 GB**
  - **Graphics: Elsa Winner 1000 S3 Trio**
  - **Fan: High quality**
  - **Housing: Midi-Tower (ATX standard)**

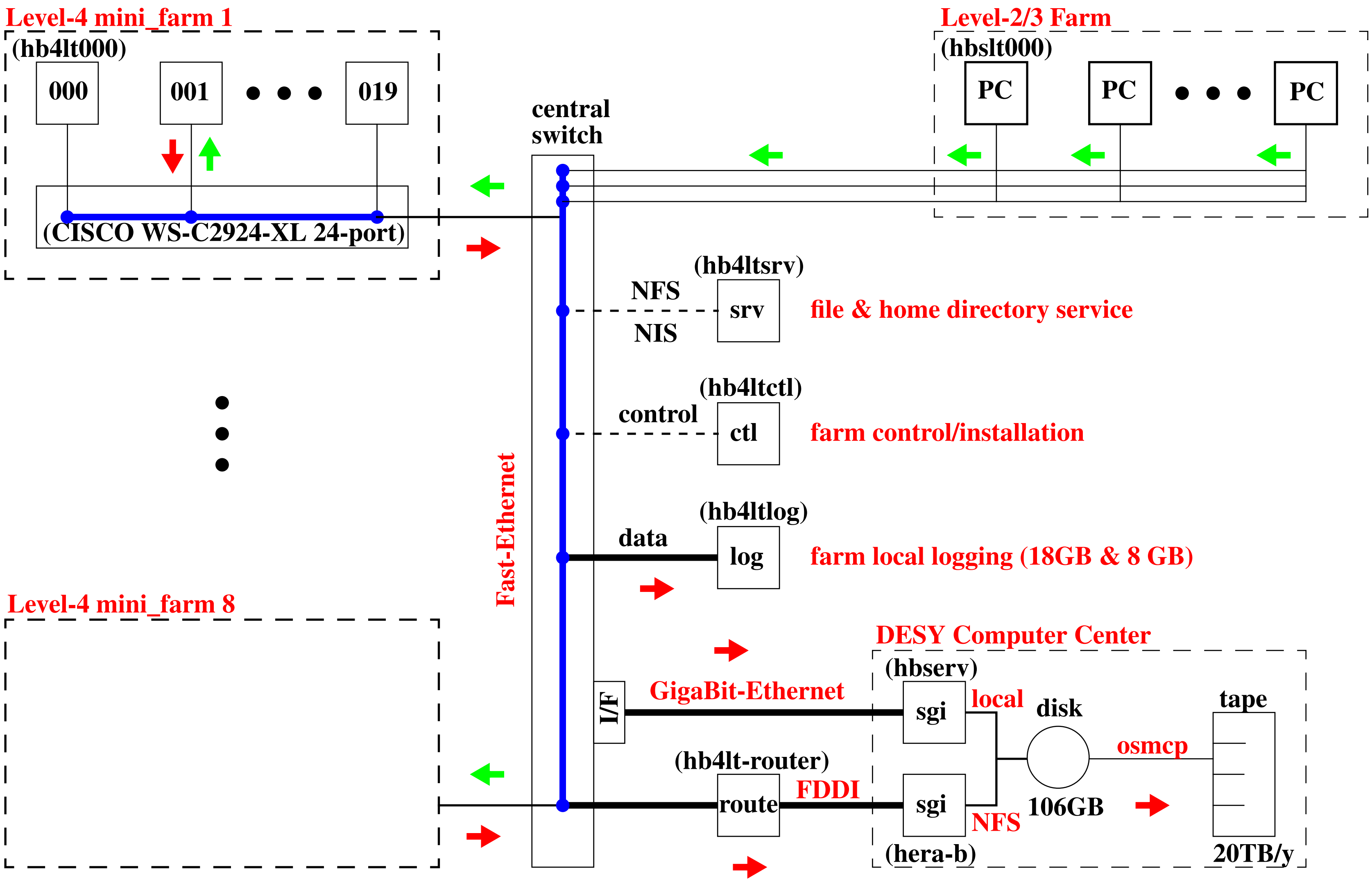
- **Operating system:**
  - **Linux Distribution: S.u.S.E 5.2 / 6.2**
  - **Kernel: 2.0.xx / 2.2.x**

- **Network:**
  - **Standard: Fast/GigaBit-Ethernet**
  - **Net card: SMC Ultra 9432 / Intel EtherPro 100+**
  - **Connections: Fast-Ethernet switches, GigaBit-Ethernet uplink**

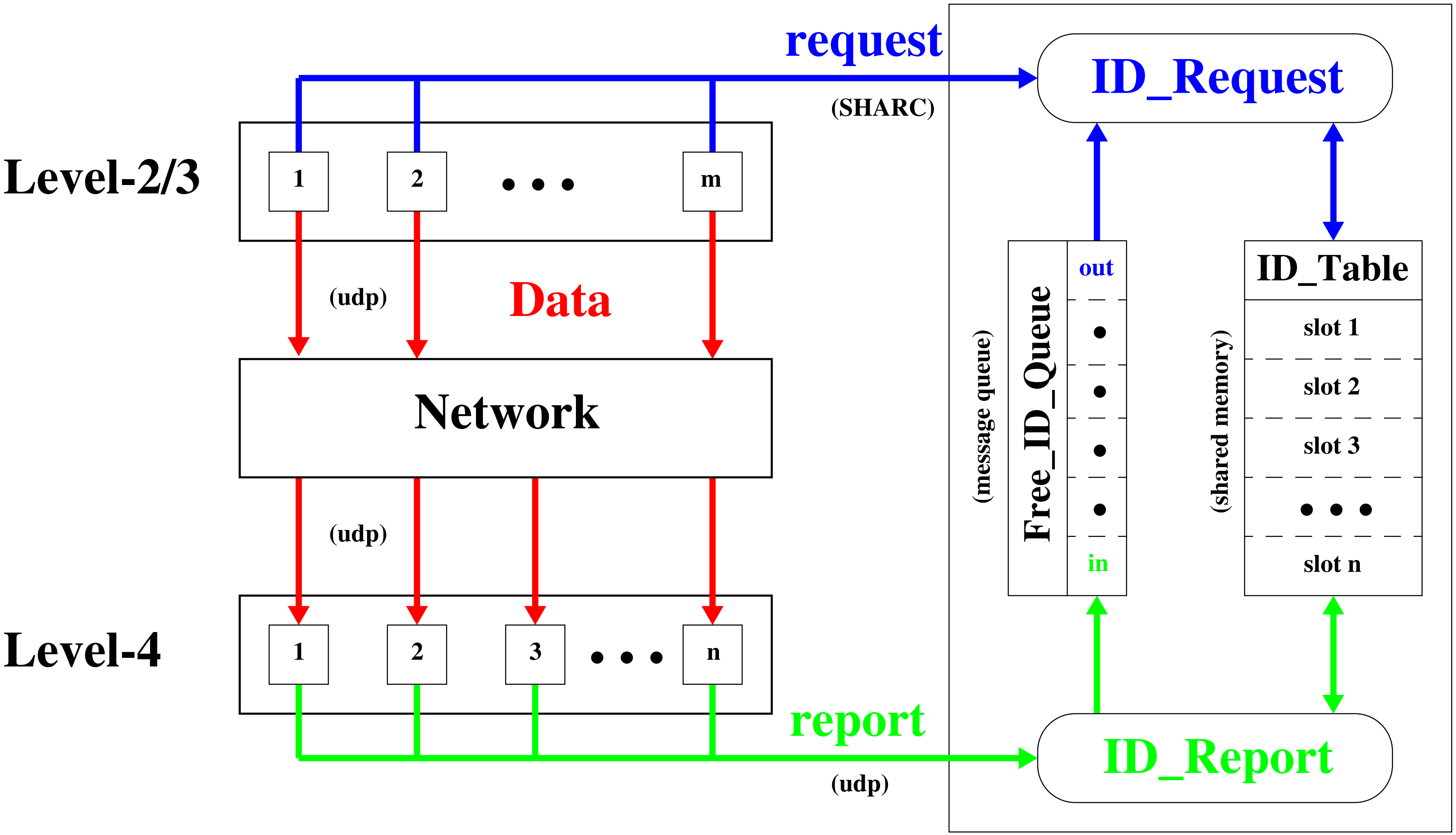
# Level-4 Farm

## (HERA-B Level-4 Farm: 23-Sep-1999)

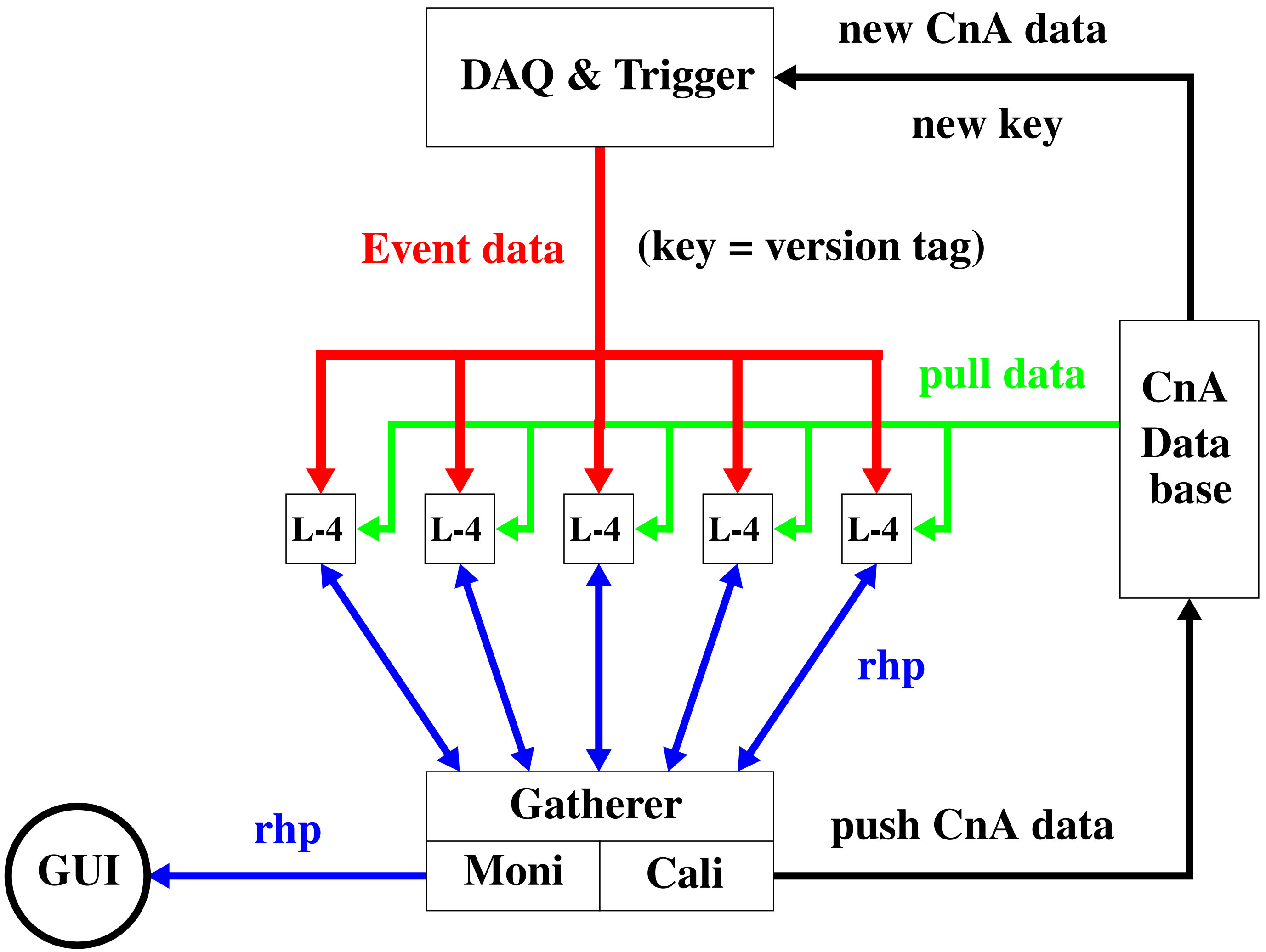
# Level-4 Farm Network



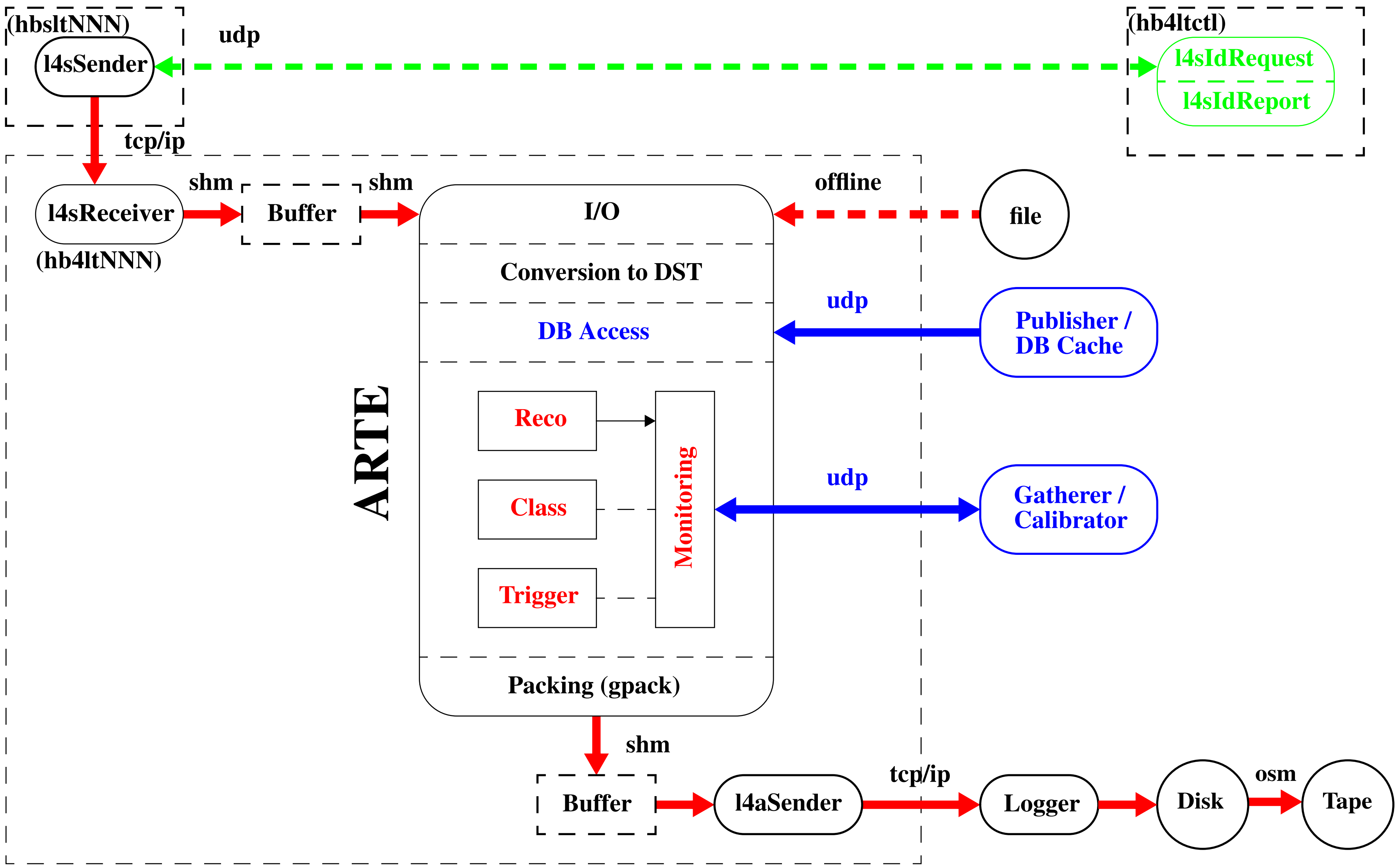
# Level-4 Farm Event Control



# Level-4 Farm Calibration & Alignment



# Level-4 Farm Node Processes



# Status

- **Reliability:**
    - 5 nodes permanently used
    - no Linux crashes
    - >200 days uptime
- **Performance:**
    - minimal processing so far
    - 5-120 kB/event
    - sustained rate of 2.5 MB/sec @ >50 kB/event
    - peak rate 3.5 MB/sec
    - total volume of ~5 TB in 1999 so far

    - new: GigaBit-Ethernet uplink to CC (logging rate: > 5 MB/sec)
- **Problems/Bottlenecks:**
    - Fast-Ethernet-to-FDDI
    - version 5 versus version 6 (→ static linking)

# Outlook & Summary

- **1999:**
  - detector commissioning
  - di-electron ECAL trigger
  - $J/\psi$ signal
  - cross section measurement
  - Level-4: online reconstruction (partly) on <50 CPUs
- **2000:**
  - CP violation measurement
  - full detector
  - $J/\psi$ trigger
  - Level-4: 50 Hz * 2-4sec = 100-200 CPUs
- **2001+:**
  - full physics program
  - full detector
  - final trigger
  - full online reconstruction
  - Level-4: 50 Hz * 4 sec = 200 CPUs