HEPiX@FNAL and CHEP'98 Personal Impressions

Gunter Trowitzsch: DAQ, Networking, Farming
Wolfgang Friebel: HEPiX, Data Analysis, Data Management, Tools

HEPiX meeting at FNAL

- most sites work on automation of administrative tasks
  - DESY: installation management (sue+cfengine), jumpstart installation of new machines, monitoring tool patrol
  - FNAL: product management (UPS)
- Linux in widespread use, especially for farms
- Increasing role of Linux and NT
  - looking for OS independent tools / methods (AFS, user DB)
- Most sites concerned about security
  - have seen dramatic increase in severity and number of security attacks (see also talk of J.Gamble at CHEP'98)

CHEP'98

- Excerpt of summary talks Sessions A, C, E and plenary talk of R.Mount
- enhanced with some remarks

- About 90 percent of the presentations on the web!
- Only three vendors present
- Poster session left little room for discussions
- Excellent summary talks

What has changed in 11 years since the Asilomar CHEP?

(Toby Burnett, Session A Summary)

- Then: mostly batch-oriented mainframes
  - fortran and common-block memory management
  - graphics only with dedicated processor (and 4010)
  - Just starting to take advantage of interactive computing
- Now: Same problems, but different scale
  - rates, data volume, # channels
- Matched by increased computing power (~100)
- Qualitative difference: new paradigms, tools
  - Networks: distributed processing (CORBA, client/server)
  - OO: C++ & STL, Java; Patterns
  - Commercial Databases: Objectivity, Oracle
  - Graphics: VRML, Open Inventor / OpenGL
  - The Web!

And, how have we responded?

- Creative opportunism: search for novel ways to apply new technology
- Create new HEP-specific tools (GEANT4, ROOT)
  - spawns more innovation
  - some effort is not explicitly shareable
- Must deal with is legacy code (CDF, H1), and people to convert/train
- Some new factors that limit acceptance:
  - use of (expensive) commercial tools
  - non-open design

GEANT4: a new tool made possible by OO

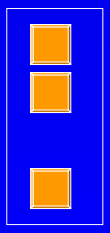
- Not a rewrite of G3; inspires innovation, new creative solutions
- beta release, prod. end 98
- collaboration of 100+ people
- #21 Apostolakis
  - kernel
  - hadronic processes
  - persistency (Objectivity)
  - readout
  - visualization (OpenGL, …)
  - geometry: parametrized volumes, materials, CAD (2..5 times faster than G3)
- #36 Wellish
  - Theory-driven hadronic interactions
- #506/#507 Verderi
  - Flexibility in response: shortcuts (kill particles)
  - Flexibility in descriptions: “ghost volumes”
- #55 Tanaka
  - DAVID: Visualization debugger

ROOT: another new tool!

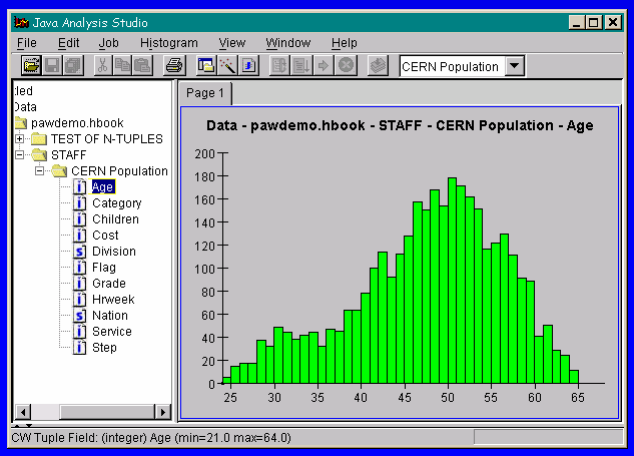
- Framework and tools for analysis, visualization, reconstruction
- Designed to ease transition to C++
- Experiments using it:
  - #176 Carminati: Alice
  - #172 Morrison: PHENIX
  - [Session E #181: STAR]

"Histogram Packages / Analysis frameworks"

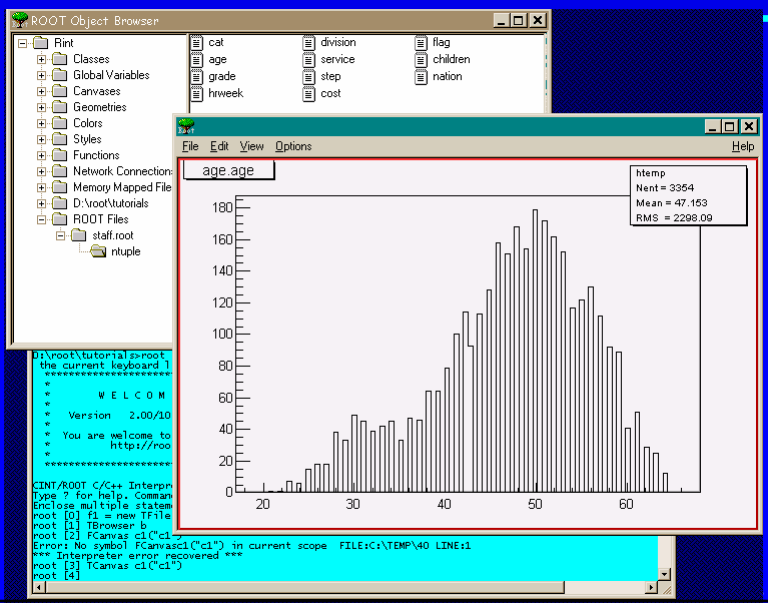
- Introduction:
  - #40 Barrand: OPACS → "Open Scientist"
- The candidates:
  - ROOT (not presented)
  - #221 T. Johnson: Java Analysis Studio
  - #80 Merlino: LHC++
    - "If I didn't have to use this, I would use Java"
  - #163 P. Kunz: HippoDraw
    - update of NextStep app presented at Annecy
    - document paradigm
- Considerations:
  - "openness", or perhaps Cathedral vs. Bazaar
  - dependence on commercial packages



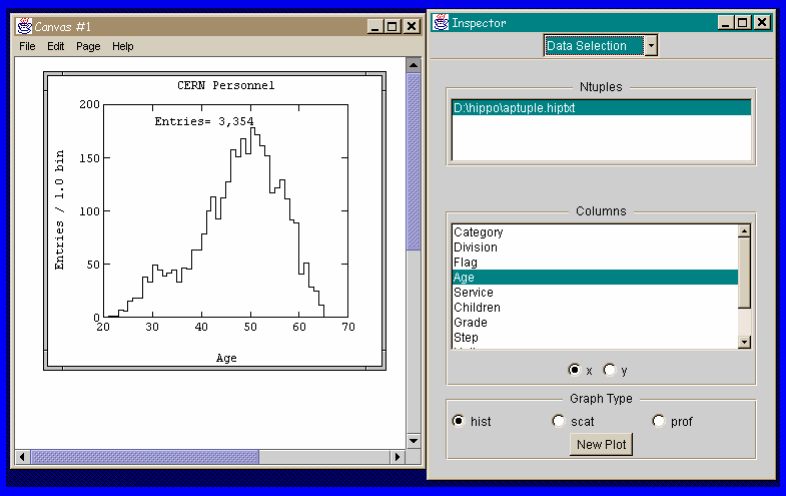
JAS staff example



ROOT staff example



Hippo staff example



Visualization

- #243 Kallenback; HepVis Class Library
  - based on OpenGL/OpenInventor, consistent with LHC, G4
  - provision of a OO class library (Boudreau '94, Barrand '97)
  - Hepvis contents: shapes, widgets, addons, viewers
- #45 D. Brown: BarBar graphics
  - Transition from opacs (based on OpenInventor) to Java-based WIRED client server solution
  - Work as part of "Universal HEP visualization"
- #208 C.Jones: *Spectator* for CLEO II
  - "user-centric design, entity, model, view

Analysis Tools & toolkits

- #63 Jacobsen: Beta
- #160 Ryd: EvtGen
- #51 Soffer: PRD
- #61 Patton: DCHAIN

Concluding remarks

- Physicists are innovative and creative
- some elegant solutions to common problems are not (easily) ported
- Predictions for the next CHEP:
  - Will hear more from ROOT and GEANT4
  - There will be no single “Paw replacement”

Session C Summary: Trends - General (D. Quarrie)

- Widespread adoption of OODBMS
  - Objectivity/DB
  - Event Store & Constants/calibrations/management
    - ATLAS,BABAR,CMS,Delphi,Kloe,STAR,Phenix,Zeus
- Major issues not addressed by existing small-scale installations
  - Scaling, Administration,Roubustness, Security/Protection, Data Distribution, Schema Management
- Major adopters come online in next 12-24 months
  - BABAR, COMPASS, STAR, etc.
  - Need these issues to be understood&resolved

Trends - General (2)

- Several major experiments adopting ROOT as alternative
  - CDF, D0, Phenix
- No clear consensus for a single data management solution
- Different risk analyses with different constraints arriving at different conclusions
- Two alternatives have strengths and weaknesses in different areas
- Exploring these is important and feedback will result in stronger eventual solutions
  - Need collaboration and constructive competition?

Trends - Mass Store

- Fairly widespread adoption of HPSS
  - Not without problems
  - Also OSM
- Several sites looking at alternatives/adjuncts
  - Home grown (ENSTORE) & commercial (ACSLS)
- Viable PC-based Mass Stores & commodity hardware
  - FNAL, KEK
- Good collaboration between Laboratories
  - e.g. DESY, FNAL, Jefferson Lab
- Coming to grips with Objectivity integration issues
  - Good collaboration between laboratories & vendor

Trends - Constants/Calibrations

- OODBMS
  - In many cases common technology with event store
- Use of CORBA to isolate implementation
  - distributed access
- Integration of detector production data with other calibration data
- Use of OODBMS for management of pseudo-realtime processes
  - Sometimes in conjunction with persistent object manager

Trends - Data Model & Access

- Hiding of persistence behind a transient layer
- Access optimizations
  - clustering/placement
    - Predictive
      - Within databases, tapes & filesystems
    - Heuristic
      - Monitoring of access patterns
  - Reclustering/cloning
  - Indexing of datasets/collections
  - Parallelism
  - Request batching

Concluding remarks for session C

- Vibrant field, very full session
- Pushing the envelope
- Major adoption of OODBMS technology
  - not the only choice however (ROOT)
  - More collaboration/constructive competition
- Not without problems and unresolved issues
- Transition to major production installations in next 12-24 months
  - Testbeds for LHC?
- Next CHEP should be interesting!

Session E Summary (F. Ruggieri)

- CLASS Libraries
- Communication & ORB
- Data Access Models
- Middleware Software
- Information & Process Management
- Software Management, Test & Distribution
- Others

Class Libraries

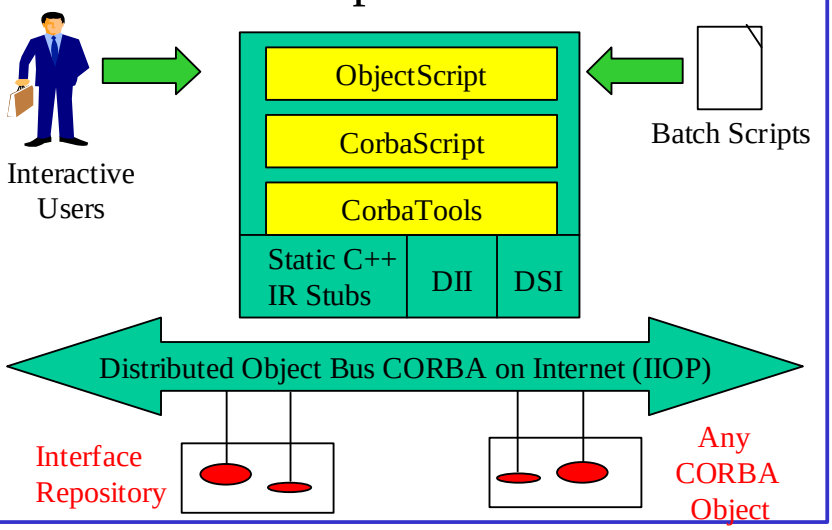
- Development of classes & methods for clusters, track finding, Particle Id & Calibration (Babar).
- Specific Experiment Classes:
  - C++ & OO approach with ROOT (STAR)
  - ZOOM - Fermilab (CDF, D0)
- SW Qualification and Metrics (Babar).

C++ Compiler (KAI, ANSI, Proprietary,…)

Communication & ORB

- CORBA implementations for non demanding applications:
  - Slow Control & Run Control (CLEO III)
  - Graphics Analysis (Babar)
- OO communication for On-Line (CMS)
- CorbaScript to ease the testing.
- Publish & Subscribe with SmartSockets (CLAS).

CorbaScript Architecture



LIFL - GOODE

Middleware

- Distributed Computing:
  - Computing Grid (CLIPPER)
  - Java “glue” for HEP (Winner)
  - Nile
- User interfaces: SUEZ (CLEO)

Conclusions 1/2

- OO software and Class Libraries are growing fast, but it’s probably time for:
  - a new step into “Cooperative Work”
  - some metrics tools for Quality Control
- Java, ORB & Middleware software are close to make our life easier and can be acquired/bought.

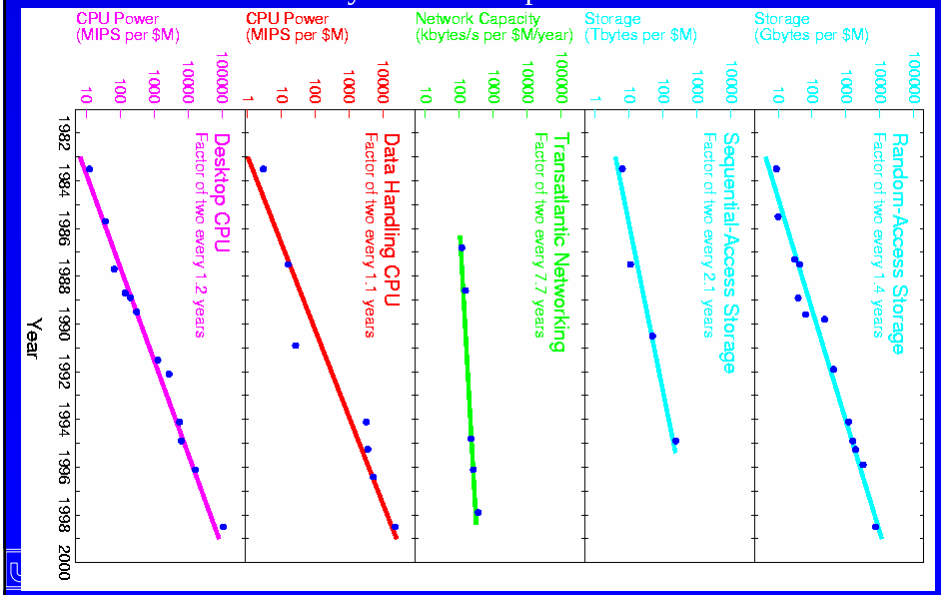
Conclusions 2/2

- Software Tools have become essential for the HEP developers community:
  - Development
  - Testing
  - Packaging
  - Distribution
- Data Modeling is important and many lessons can be learned from the (almost) running experiments.

Excerpt from plenary talk by R.Mount

Future Directions for HENP Computing

Technology, Cost and Tariff Evolution Since 1983: My Personal Experience



Parallel Computing in HENP (3)

Will we get into parallel processing?

No and Yes:

No: Today's arguments against parallel processing will remain valid;

Yes: Invisible parallelism will creep upwards

- Striped disks;
- Multithreaded database servers;
- Automatic parallel decomposition of database queries

Storage versus Computation (2)

Computation is becoming cheaper faster than mass storage (I think).

Could we eliminate explicit storage of processed data?

Now? No

In 2005? Possibly

⇒ Need DBMS that will transparently retrieve or recompute objects. DBMS should decide whether to cache objects that have been recomputed or retrieved via WAN.

Who is working on this?

Mass Storage (1)

- Disk technology and price is driven by a vigorous commodity market.
- Robust tape technology is specialized and small-market.
- Mass-storage management software is at the lunatic fringe.

Will all this continue?

Networks

HENP (intercontinental data-intensive collaborative science) is network hungry and chronically undernourished.

Is any relief in sight?

Why not try wishful thinking?

Networks: Wishful Thinking (1)

- There is a near-term market for T1–DS3 per household at $50/month;
- There are no major technological barriers;

∴ It will happen;

∴ Intercontinental infrastructure to support it will appear.

Planning and Management of HENP Software Development

- Difficult (at least for me);
- Encourage new ideas;
- Encourage collaboration;
- Encourage optimal use of commercial software;
- Build on the GEANT4 collaborative model;
- Balance healthy diversity against damaging duplication;
- Trust more valuable than management.

Relationship with Scientific Computing and Computing Science

- Easy to see why we have absolutely nothing to gain from working with Biology, Chemistry, Climate, Fusion .....
- Equally easy for Climate etc. to see why they have nothing to gain from working with HENP;

e.g. "Data + Visualization Corridors" (ASCI)

= "Data + Data Management + Data Analysis + Visualization Corridors" (HENP)

Perceptions of HENP (e.g. "Every Politician Knows")

- Scientists who will not collaborate with
  - their colleagues at other universities/labs,
  - other "real" sciences,
  - computing science,
  - hardware and software industry,

  are wasting resources.
- Arguments that "we cannot use commodity computing" are not worth listening to.

Perceptions are substantially correct.