

Using an Object Oriented Database to Store BaBar's Terabytes

Tim Adye
Particle Physics Department

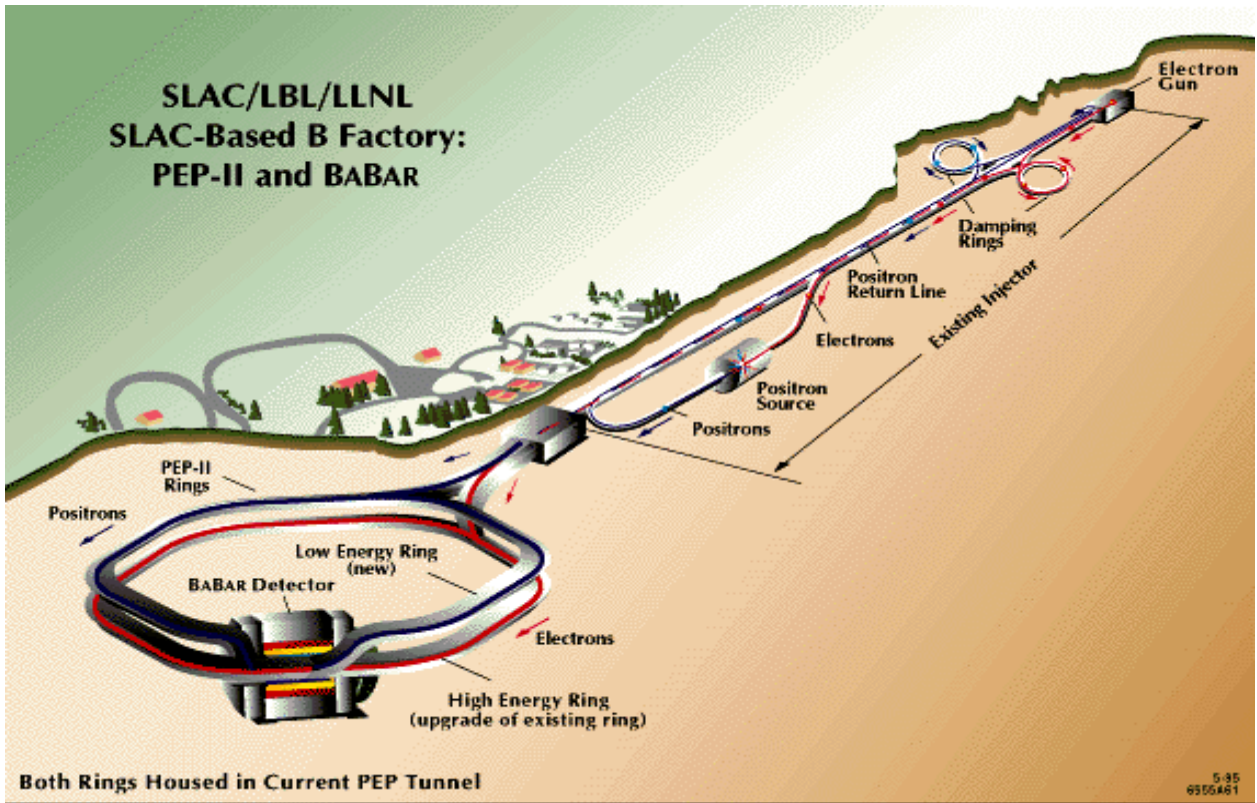
CLRC Workshop on
Advanced Data Storage and
Management Techniques

Outline

- The BaBar experiment at SLAC
- Data storage requirements
- Use of an Object Oriented Database
- Data organisation
- SLAC
- RAL
- Future experiments



- The **BaBar** experiment is based in California at the Stanford Linear Accelerator Center, and was designed and built by more than 500 physicists from 10 countries, including from **9 UK Universities** and **RAL**. —————>
- It is looking for the subtle differences between the decay of the **B^0** meson and its **antiparticle** (\overline{B}^0).
 - If this “CP Violation” is large enough, it could explain the cosmological matter-antimatter asymmetry.
- We are are looking for a **subtle effect** in a **rare** (and difficult to identify) decay, so need to record the results of a **large numbers of events**.



How much data?

- Since BaBar started operation in May, we have recorded 7 million events.
 - 4 more years' running and continually improving luminosity.
 - Eventually record data at ~ 100 Hz; $\sim 10^9$ events/year.
 - Each event uses 100-300kb.
 - Also need to generate 1-10 times that number of simulated events.
- Recorded 5 Tb
 - Expect to reach ~ 300 Tb/year
 - I.e. 1-2 Pb in the lifetime of the experiment.

Why an OODBMS?

- BaBar has adopted **C++** and **OO** techniques
 - The first large HEP experiment to do so wholesale.
- An **OO Database** has a more natural interface for **C++** (and **Java**).
- Require **distributed** database
 - Event processing and analysis takes place on many processors
 - 200 node farm at SLAC
- Data structures will **change** over time
 - Cannot afford to reprocess everything
 - Schema evolution
- **Objectivity** chosen
 - Front runner also at CERN

How do we organise the data?

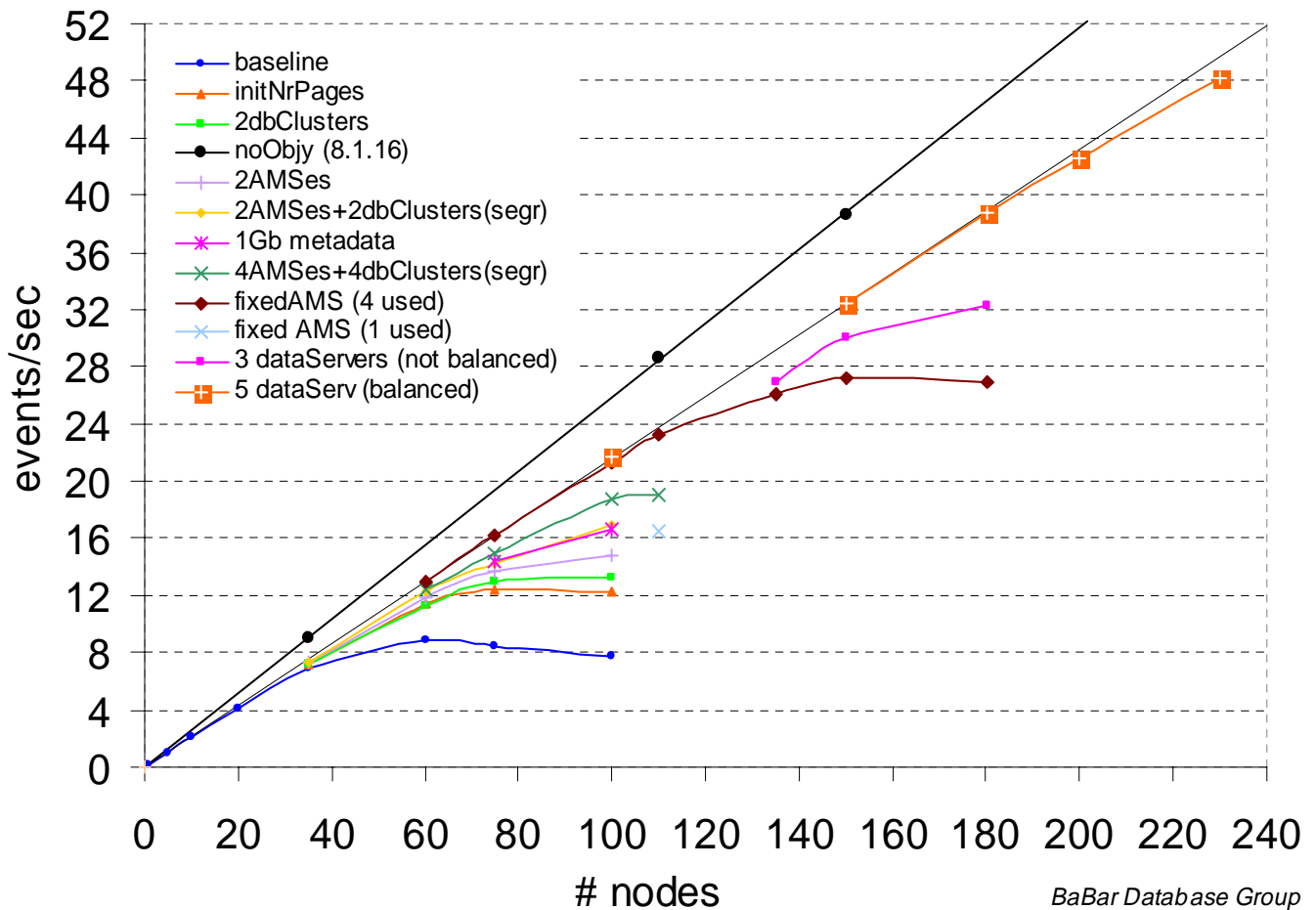
- **Traditional** HEP analyses read each event and **select** relevant events, for which additional processing is done.
 - Can be done with sequential file
 - Many different analyses performed by BaBar physicists.
- In BaBar there is **too much data**.
 - Won't work if all the people to read all the data all of the time.
 - Even if all of it were on disk.
- Organise data into different **levels of detail**
 - Stored in separate files
 - tag, "microDST", full reconstruction, raw data
 - Objectivity keeps track of **cross-references**
- Only read more detailed information for **selected** events.
 - But different selections for different analyses

What happens at SLAC?

- Cannot store everything on disk
 - Maybe 5 Tb, but not 1 Pb.
 - Already buying ~1 Tb disk per month.
- Analysis requires frequent access to summary information.
 - Keep tag and “microDST” on disk
 - Rest in mass store (HPSS at SLAC)
- Main challenge is getting this to scale to hundreds of processes/ors reading and writing at the same time.
 - The vendor seems to believe we can do it.
 - “The Terabyte Wars are over
While other vendors quarrel about who can store 1 Terabyte in a database, the *BaBar* physics experiment at the *Stanford Linear Accelerator Center (SLAC)* has demonstrated putting 1 Terabyte of data PER DAY into an Objectivity Database.”
 - Top news item on *Objectivity* web site
 - But it took a lot of work...

Performance Scaling

- A lot of effort has gone into improving speed of **recording** events



- Ongoing work on obtain similar improvements in data **access**.

RAL as a Regional Centre

- Cannot do everything at SLAC
 - Even with all the measures to improve analysis efficiency at SLAC, it cannot support **entire collaboration**.
 - Network connection from UK is **slow**, sometimes very slow, occasionally **unreliable**.
- Therefore need to allow analysis outside SLAC.
 - “Regional Centres” in UK, France, and Italy.
 - **RAL** is the UK Regional Centre.
- Major challenge to **transfer data** from SLAC, and to **reproduce** databases and analysis environment at RAL.

RAL Setup

- At RAL, have just installed Sun analysis and data server machines with **5 Tb disk**
 - UK Universities have 0.5-1 Tb locally
 - All part of £800k JREI award
- Import microDST using **DLT-IV**
 - ~70 Gb/tape with compression
- Interfaced to **Atlas Datastore** (see John Gordon's talk).
 - Less-used parts of the federation can be **archived**
 - Can be brought back to disk on demand
 - needs further automation
 - Also acts as a local **backup**.

Other Experiments

- BaBar's requirements are **modest** with respect to what is to come.
 - 2001 Tevatron Run II: **~1 Pb/year**.
 - 2005 LHC: **many Pb/year**.
- Choice of HSM.
 - **HPSS** is expensive. Maybe we don't need all the bells and whistles.
 - But already in use at SLAC/CERN/...
 - **EuroStore** (EU/CERN/DESY/...)
 - **ENSTORE** (Fermilab)
 - **CASTOR** (CERN)
 - Which way should RAL go?
- Is **Objectivity** well-suited to our use?
 - Develop our own?
 - Espresso (CERN)
 - BaBar is being watched closely...