Database Futures Workshop

Rapid Summary

Tony Cass

- Oracle
- Other SQL
- NoSQL
- Heresy

Oracle

- **Many mission-critical applications**
  - Relatively well understood and relatively stable
- **Administrative & Engineering**
  - "Trivial" data volumes, but many users
- **Experiments**
  - Large data volumes, but growth expected to be ~linear with physics data volume
    - In some cases, hardware capability growth outstrips system requirement growth.
- **Accelerator**
  - Very large data volume, pressure to store more (e.g. Beam Loss Monitor data)
  - O(10PB) by 2020
  - O(EB) for CLIC, however (and LHC initial estimates were off by factor of 50…

- **Live long and prosper!**

- Other SQL is mostly MySQL; some SQLite.

- Systems run fine now (c.f. low support requirements for ALICE DAQ environment)…

- … but if someone were to offer support, there would be takers.
  - Slippery slope? (SQL Server; can there be a "canned" MySQL environment or would there be many?)

- The key issue seems to be the difficulty of providing efficient read performance for essentially random queries—databases have been optimised for inserts and "production queries".

  - Even an issue in BE where people have been well used to tight constraints on read access as this continues the practice of the LEP era.

- The time required to develop an application delivering reasonable performance seems to be less with NoSQL systems than with Oracle (the target of all comparisons).

- But, for me at least, the requirements are still unclear.

  - How many "random read" applications are there? Will each need a dedicated NoSQL database?

  - Can a single NoSQL database support all requirements?

    - Unlikely: both ATLAS and CMS have chosen different platforms for different applications…

  - Is ease-of-setup at the cost of future maintenance woes?

- Is today's MySQL support model adequate for the NoSQL world?

Simon's Heretical Thought

- NoSQL is only interesting as a disruptive technology.

- The disruptive technology is physics data in a database
  - i.e. this is what analysis programs run against…