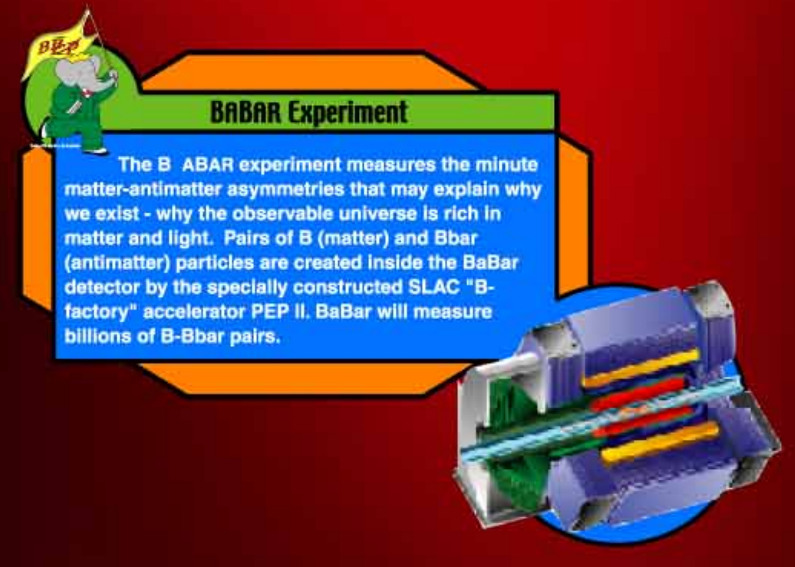
Data Intensive Science

PROCESSING AND MANAGING DATA

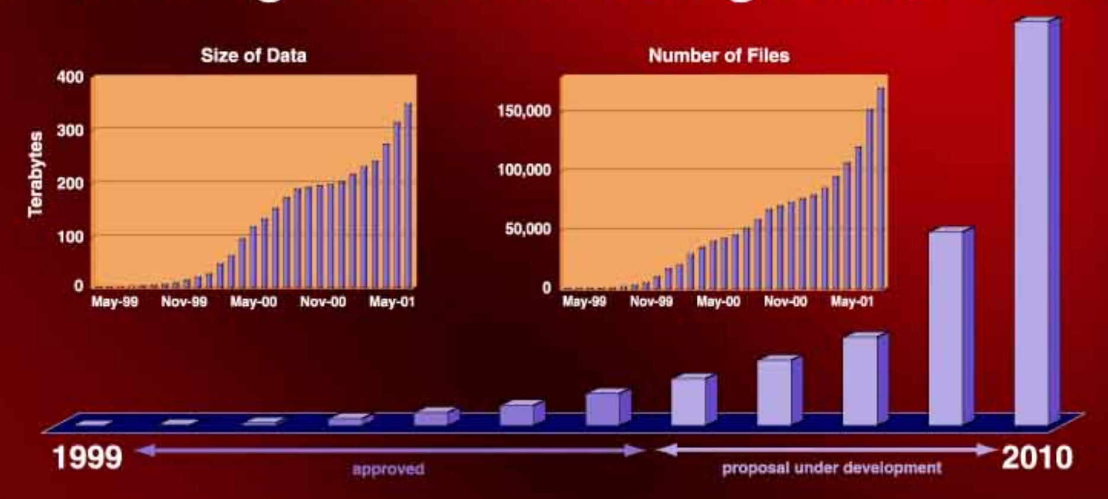
Since May 1999, the BABAR Database System has stored 201,404,275,985,952 bytes of data. In 2009, we expect to store that amount of data every few days.



BABAR Experiment

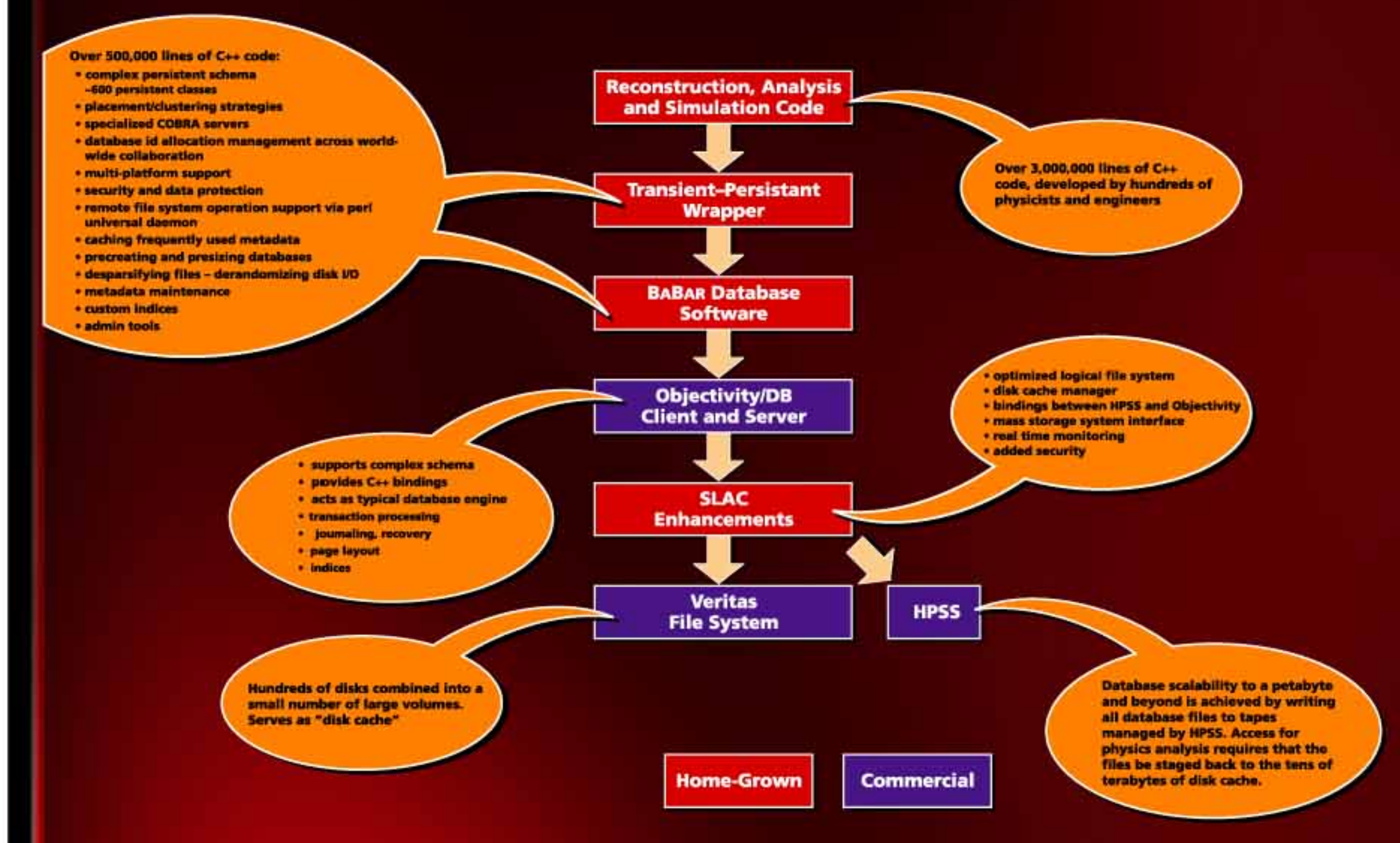
The BABAR experiment measures the minute matter-antimatter asymmetries that may explain why we exist - why the observable universe is rich in matter and light. Pairs of B (matter) and Bbar (antimatter) particles are created inside the BaBar detector by the specially constructed SLAC "B-factory" accelerator PEP II. BaBar will measure billions of B-Bbar pairs.

The Challenge of the Increasing Rate of Growth



The BABAR Database System Integrates:

- millions of lines of *custom* C++ code
- a commercial Object Oriented Database System: Objectivity/DB
- all running on over →
  - 2000 CPUs
  - 100 servers



Over 500,000 lines of C++ code:
- complex persistent schema ~600 persistent classes
- placement/clustering strategies
- specialized COBRA servers
- database id allocation management across world-wide collaboration
- multi-platform support
- security and data protection
- remote file system operation support via perl universal daemon
- caching frequently used metadata
- precreating and presizing databases
- desparsifying files – derandomizing disk I/O
- metadata maintenance
- custom indices
- admin tools

Over 3,000,000 lines of C++ code, developed by hundreds of physicists and engineers

- optimized logical file system
- disk cache manager
- bindings between HPSS and Objectivity
- mass storage system interface
- real time monitoring
- added security

- supports complex schema
- provides C++ bindings
- acts as typical database engine
- transaction processing
- journaling, recovery
- page layout
- indices

Hundreds of disks combined into a small number of large volumes. Serves as "disk cache"

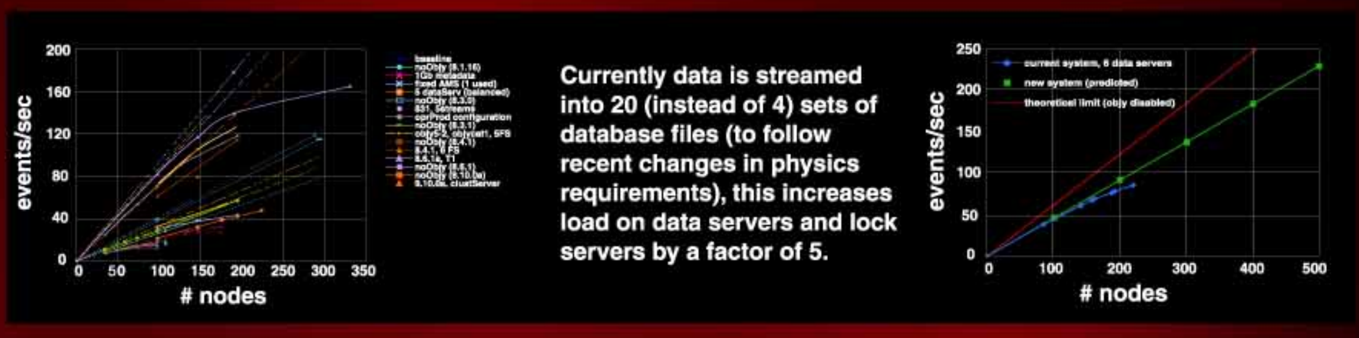
Database scalability to a petabyte and beyond is achieved by writing all database files to tapes managed by HPSS. Access for physics analysis requires that the files be staged back to the tens of terabytes of disk cache.

Recent surveys report the largest database in the world is in the order of 11 terabytes. Hmmmm...

Existing off-the-shelf commercial database software does not scale to hundreds of terabytes.

Optimizing, Tuning....

- redesigning
- optimizing code
- parallelizing, centralizing
- understanding and removing bottlenec ks
- introducing faster hardware



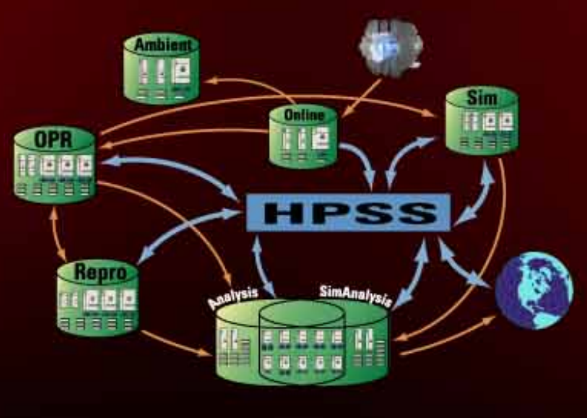
Currently data is streamed into 20 (instead of 4) sets of database files (to follow recent changes in physics requirements), this increases load on data servers and lock servers by a factor of 5.

Look at a Running System...

(...yes, in real time)

The monitor to the left shows load on various servers:
- Online Prompt Reconstruction – 600+ computing nodes are processing data incoming from the detector.
- Analysis –hundreds of physicists are analyzing the data

A Distributed System, with a Non-Trival Data Flow



On average, about 200 GB of data is transferred to SLAC every day. About 180 GB per day is transferred from SLAC to main centers in Europe. From there, data is distributed to local institutes.

These numbers are expected to grow to 1 TB per day in the near future.