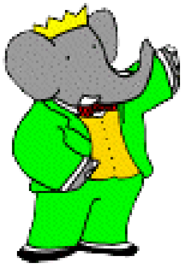
Computing and analysis experience in BaBar



Francesco Fabozzi

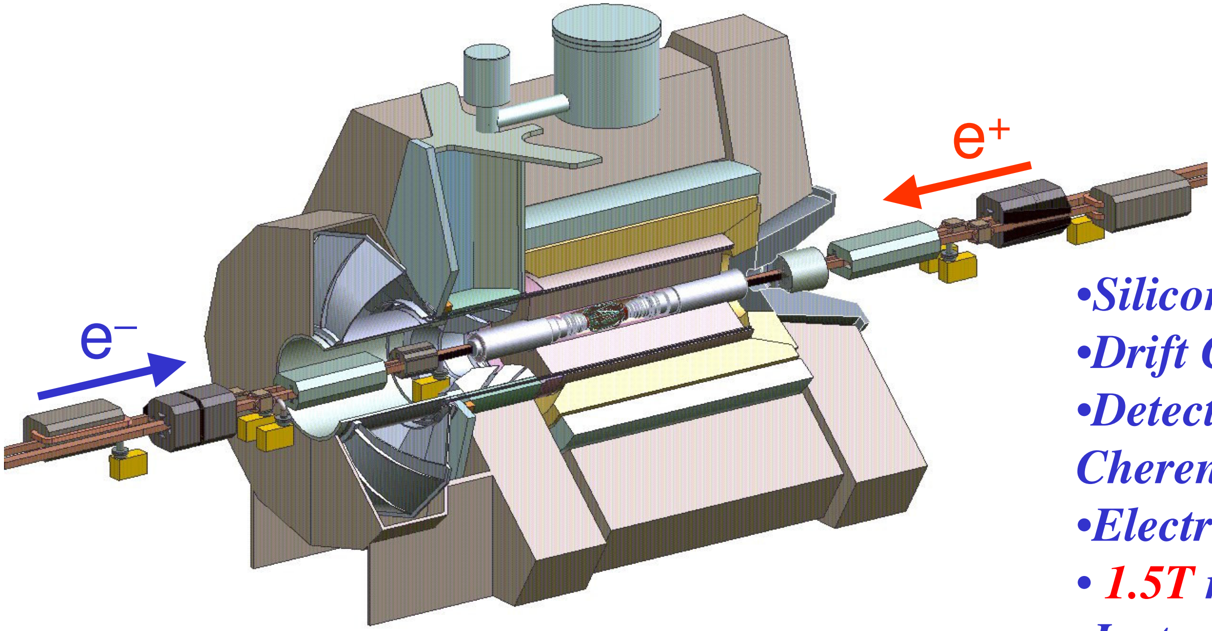
INFN-Napoli

III Workshop italiano sulla Fisica di Atlas e CMS

Bari, 20 – 22 Ottobre 2005

The BaBar experiment @ SLAC

- Study of CP violation in B meson decays
  - B mesons produced at PEP-II B-factory
    - $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\overline{B}$ decays
    - About 1M $B\overline{B}$ / fb^{-1}

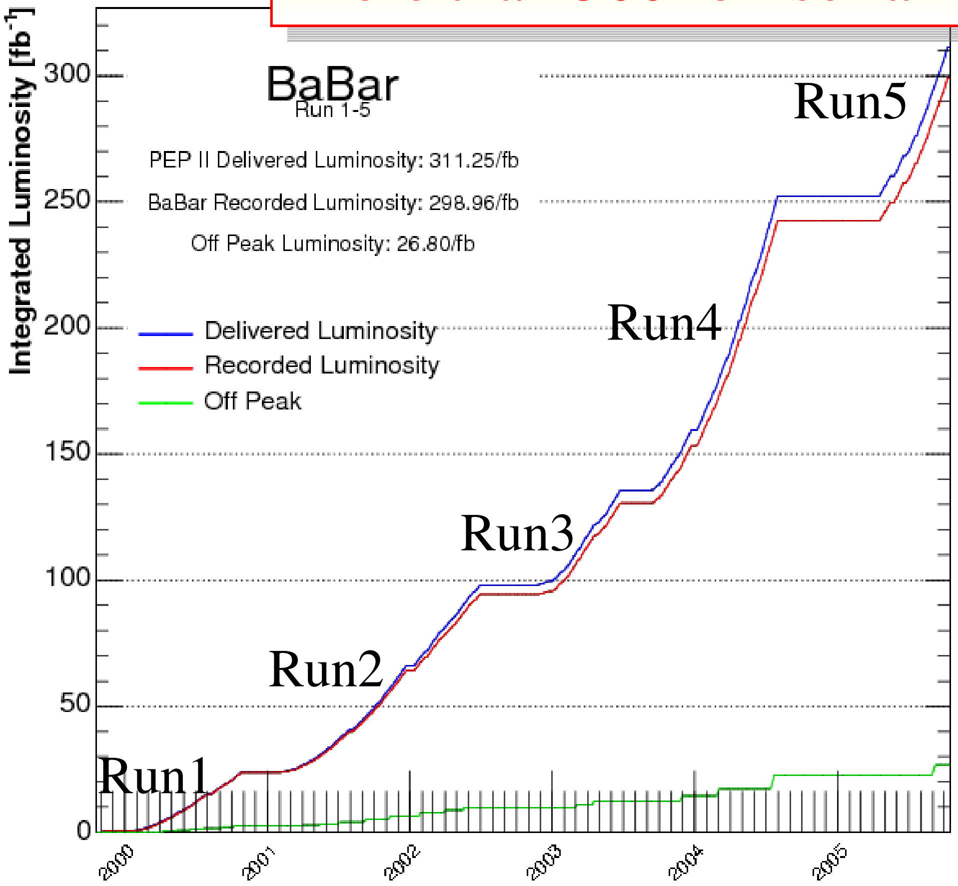


•Silicon Vertex Tracker (SVT)
•Drift Chamber (DCH)
•Detector of Internally Reflected Cherenkov Light (DIRC)
•Electromagnetic Calorimeter (EMC)
• 1.5T magnet
•Instrumented Flux Return (IFR)

The BaBar experiment @ SLAC

- The physics goals of BaBar require high statistics
  - High luminosity of the machine
  - High efficiency in data acquisition
- The computing model (CM) has to be flexible enough to scale with the huge amount of available data

More than 300 fb^{-1} so far



Requirements of the CM

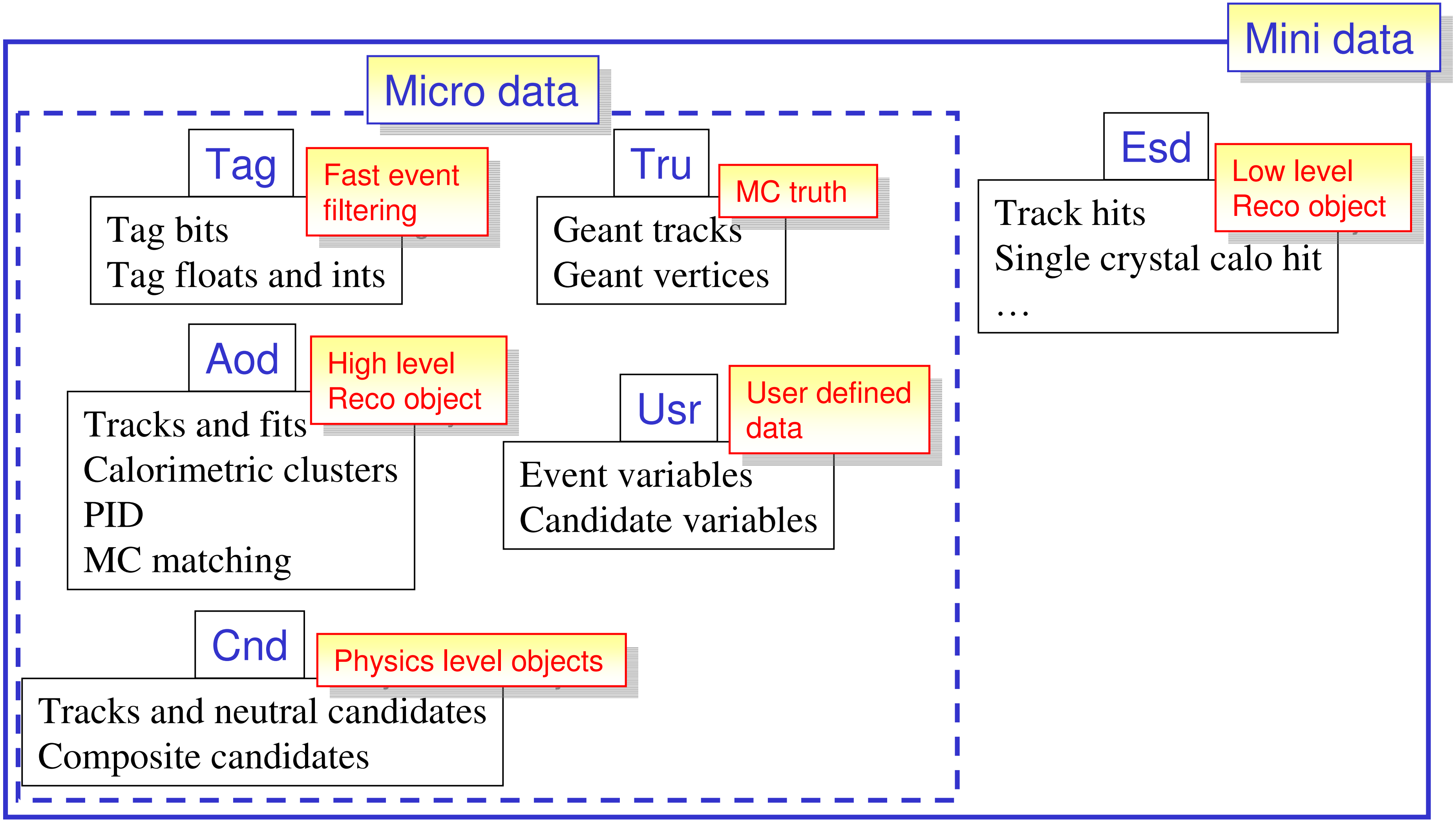
- Increase physics productivity
  - Data available in a short amount of time in a format suitable for physics analysis
    - New/updated physics results soon after the data are collected
  - Example: analysis model
- Avoid overload of CPU and disks resources
  - Optimized use of resources
  - Distributed computing
    - Use all the possible resources available in the collaboration
  - Example: simulation production

The Computing Model 2

- The present computing model (CM2) has proven enough flexibility to scale with the luminosity ramp-up
  - Adopted after the experience of the first years of data taking
    - The previous computing model (CM1) was not able to scale with luminosity
- The CM2 was the results of a huge effort within the collaboration
  - Developed in ~ 1 year (2003)

The analysis model

- The event store is organized in different components

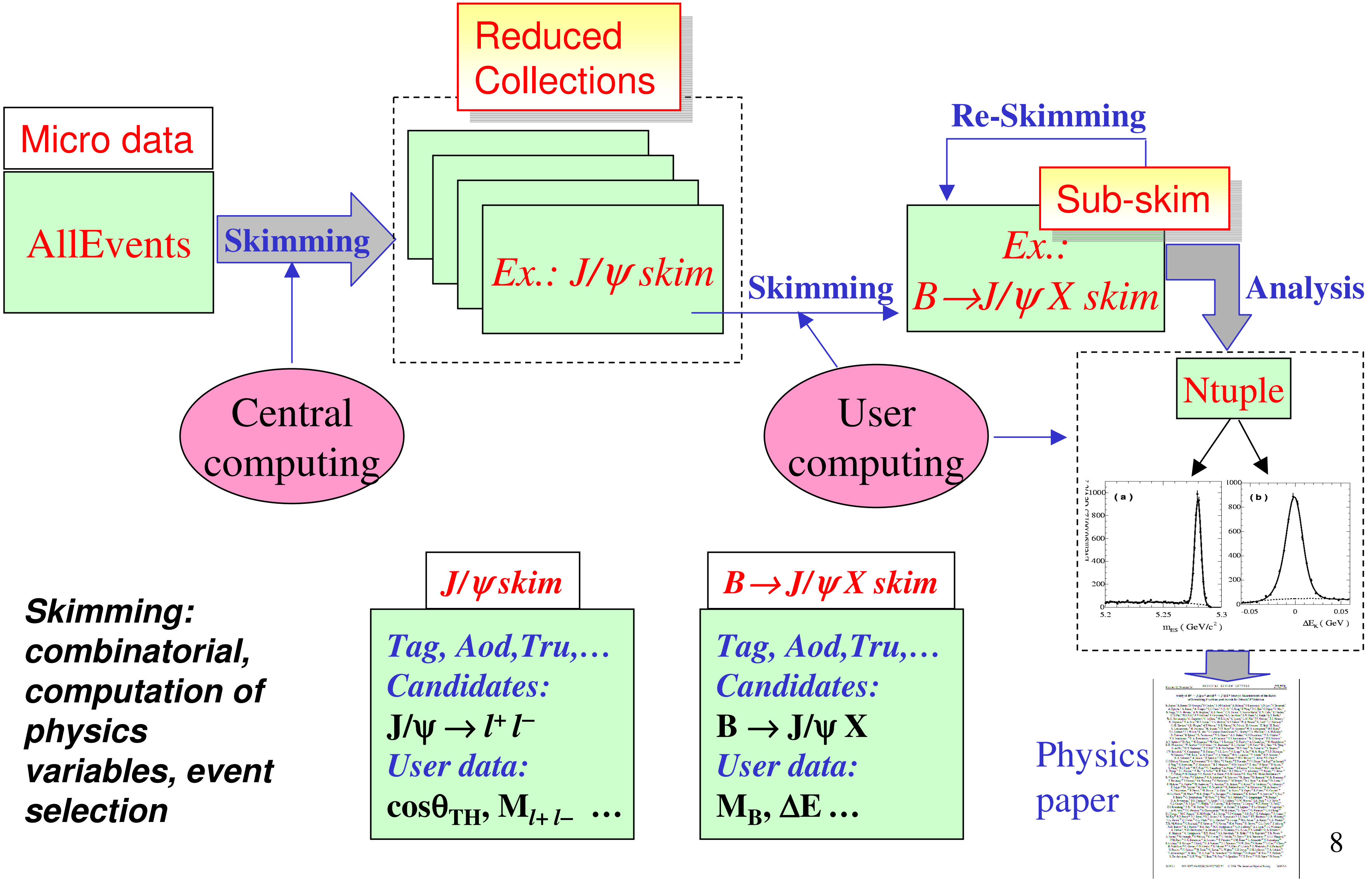


The analysis model

- “Micro” are a subset of “Mini”
  – No separation between data for physics analysis and reco data for detector studies
    - Changes in reconstruction reflect directly into analysis
    - User can specify different levels of data access when running on the same collection
- Persistency of composite candidates
  – Save CPU time when doing analysis iteration
- Addition of user-defined data
  – Store physics variables needed in the analysis

The analysis model

- From reconstructed data to physics results

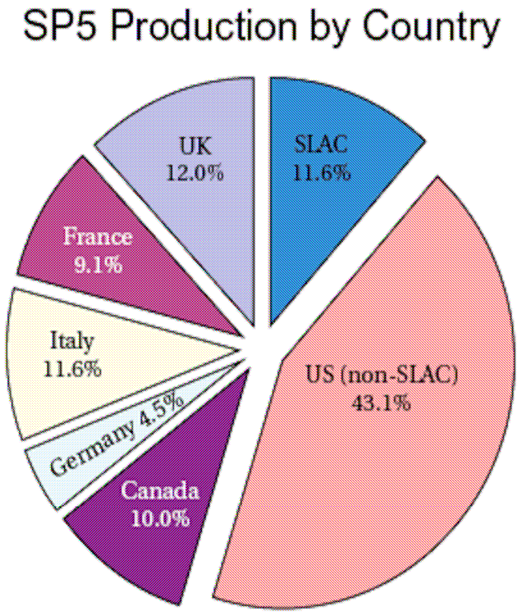


The analysis model

- Centralized data reduction (“skimming”)
  - Reduced collections to run user analysis code
  - Possibility to store particle candidates and analysis variables in the output collections
    - Save computing time when re-running
- No need of massive ntuples productions
  - Users can produce their own sub-skims
    - Access to the experiment software
    - Interactive access from ROOT of the collections
      - ROOT I/O based event store
    - Can be exported to small sites for the final analysis

Simulation Production (SP)

- Distributed computing effort
  - More than 20 centers contribute to MC production
    - Almost 2000 CPU
    - Equally distributed between USA and Europe
- Typical production cycle
  - few billion events in ~1 year
  - almost 100 TB of data
  - management of ~1 M jobs on Linux nodes



Simulation Production (SP)

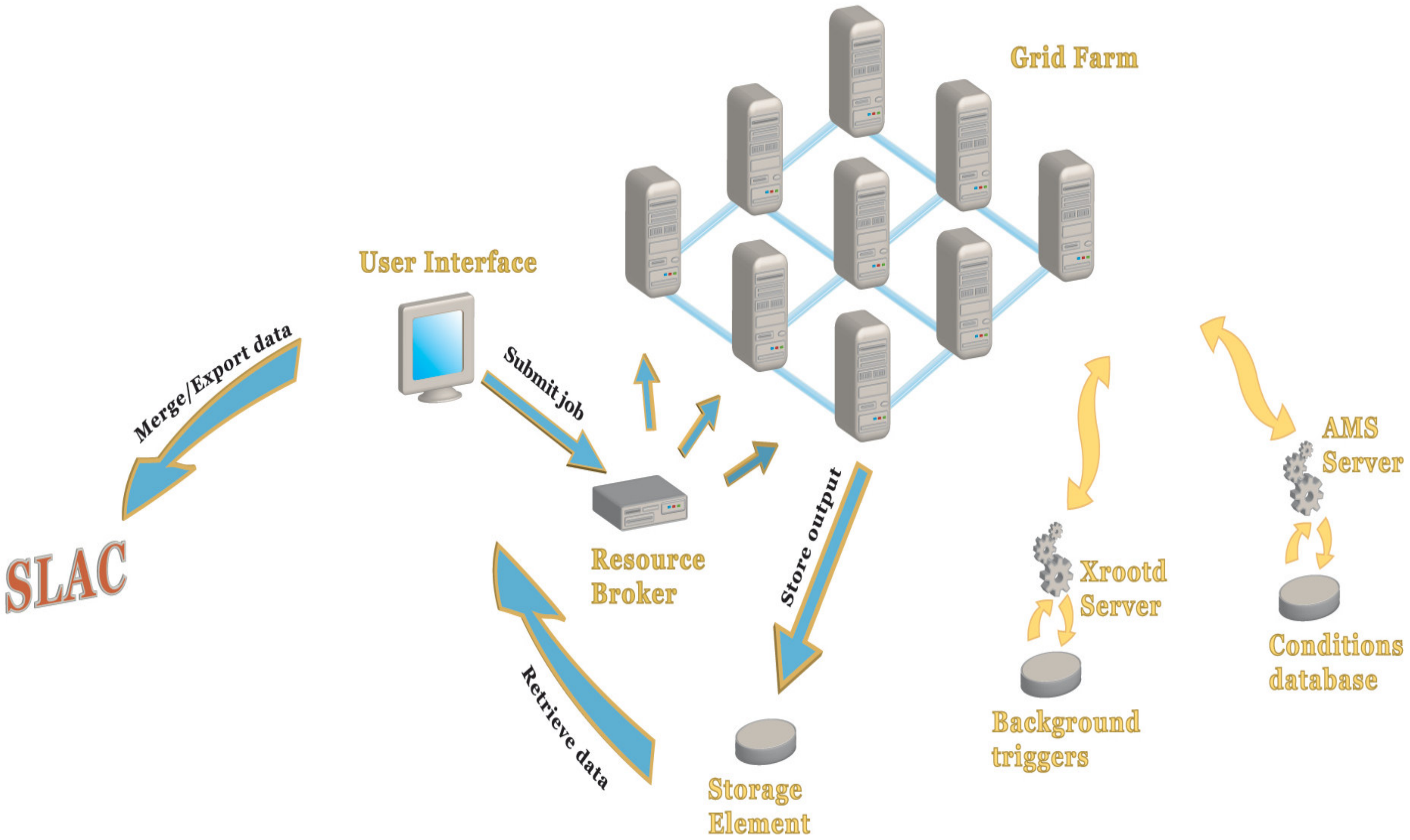
- Centralized management of production (requests, job assignement and archiving)
  - Central production database at SLAC
- Jobs require in each site:
  - a specific release of BaBar software
  - Configuration and Condition database (Objectivity)
  - Detector noise info (xrootd server)
- Set of tools (= perl scripts) for jobs management (ProdTools)
  - Interface between central production database and external resources
    - Flexibile (heterogeneous resources)
    - Robust and self-recovering system (reduce human effort)
      - Low failure rate (below 1% with new event store)
- Transfer back to SLAC of produced data for archiving
  - On average 200GB/day transfer
    - Dedicated file server

SPGrid-Italia

- Simulation Production is a natural candidate for a Grid-based setup
  - Big effort in Italy to gridify BaBar software using LCG middleware
- Resources:
  - ~ 400 shared CPU
    - Virtually all Grid-IT resources
  - UI in Ferrara with ProdTools installed
  - BaBar RB in Ferrara
  - AMS server (Objy DB) and xrootd server installed in Ferrara, Napoli, Padova
    - Non-gridified, accessed via WAN

SPGrid-Italia

- SPGrid-Italia scheme



SP-Grid status

- Starting from June, official SP8 and SP6 Italian production is performed also on the Grid
  - Main limitation in resource usage due to Objectivity condition database (max 70 concurrent access to DB for optimal performances) and use of WAN for access
  - Event production rate ~ 2M events / week
  - Gridifications of BaBar resources at CNAF T1 is planned in the next few weeks ⇒ relevant increase of available resources for productions
- SPGrid efforts also in UK
  - Part of UK SP resources have been gridified
    - Common approach with Italian SPGrid
- SPGrid will start soon in Germany
- Also plannings in US and Canada
  - Planned use of GridX1 resources in Canada for production
  - Planned use of Open Grid Science for US production