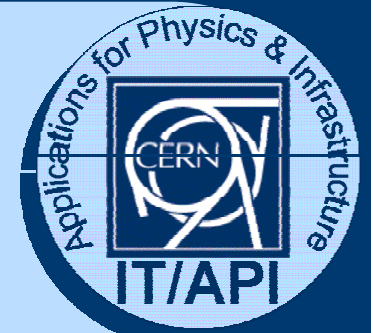


# GEANT4 Parallelism & Grid

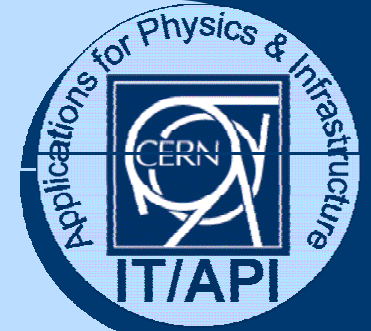
J. Knobloch IT/API  
GEANT4 Workshop Genova  
2 July 2001

# Topics



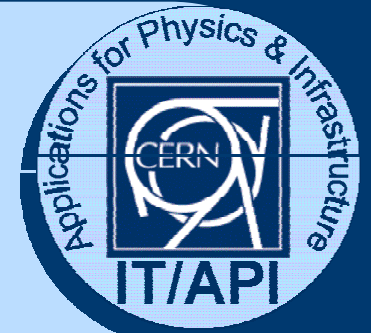
- Why parallel processing?
- What kind of parallel processing?
- Issues to be addressed
  
- To be worked out in the parallel session

# Why parallel processing?



- Single job submission on a multi-processor fabric
- Fast response for interactive development
  - Interesting for rapid histogramming of simple events with “particle-guns”
  - Otherwise it usually requires sub-event parallelism
    - few real-world use cases

# GEANT4 URD:



**UR 19-10.** GEANT4 shall be able to process events in parallel.

**Need** Useful.

**Priority** Not been made part of the LHCC milestones 1994-1998.

**Stability** Evolving.

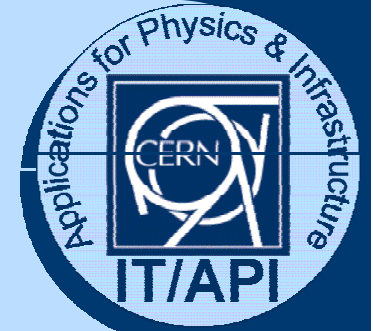
**Source** RD44.

**Clarity** Clear.

**Verifiability** Not applicable.

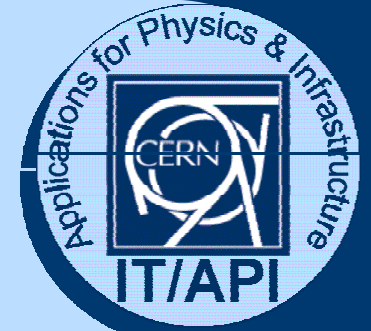
- Assume for now: Event level is sufficient!

# Sub-requirements



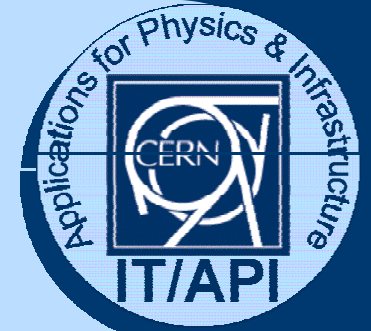
- Repeatability – reproducibility
- Result independence of number of processors
- Exception handling to be dealt with at the processor level and at the control level
- Availability of status information

# Issues



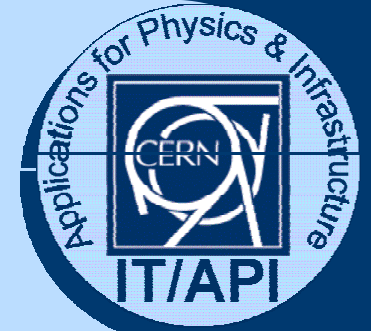
- Initialisation – Geometry & Physics
  - done once and then distributed?
  - > Balance between dataflow and processing time
- Random number distribution
  - seeds per event – central distribution
  - The only way to guarantee reproducibility!
- Data distribution & collection
  - Generator files
  - Event output
  - Summaries, histograms, statistics
  - Database solves many but not all items

# DataGrid



- Work Package 8: High Energy Physics Applications
- Presentation by F. Carminati at FOCUS
- Partners: CNRS, INFN, NIKHEF, PPARC, and CERN
- Task: coordinate the exploitation of the prototype system by the HEP experiments and identify common components to be integrated in a HEP upper middleware layer

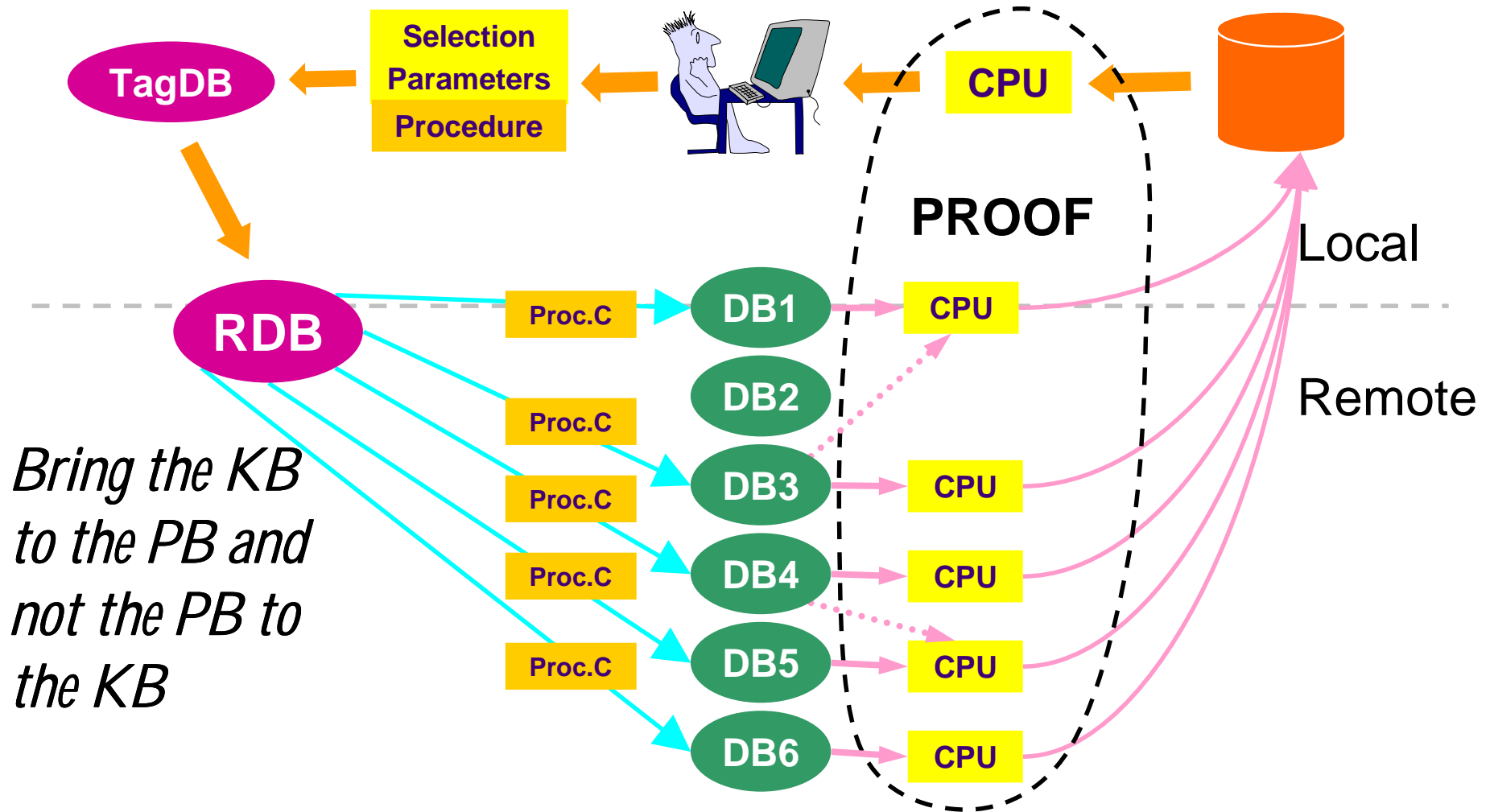
# WP8 – Experiments activities



- ALICE: distributed analysis model using AliRoot DataGRID Services
- ATLAS: Requirements; Tests with GEANT3; Tests with Objectivity planned for this autumn
- CMS: Definition of requirements; Planning first tests; European and US labs involved
- LHCb: Distributed MC production at CERN, RAL, Liverpool and Lyon; Current Testbed-0 tests using CERN and RAL (Globus problems encountered)



# ALICE distributed analysis model



# LHCb Short term Use Case for GRID

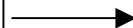
## 1. Production started by filling out a Web form:

- Version of software
- Acceptance cuts
- Database
- Channel
- Number of events to generate
- Number of jobs to run
- Centre where the jobs should run



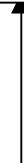
## 2. Web form calls a java servlet that:

- Creates a job script (one per job)
- Creates a cards file (one-three per job) with random number seeds and job options (The cards files need to be accessible by the running job)
- Issues a job submit command to run script in batch-> WP1



## 4. Java program calls servlet at CERN to:

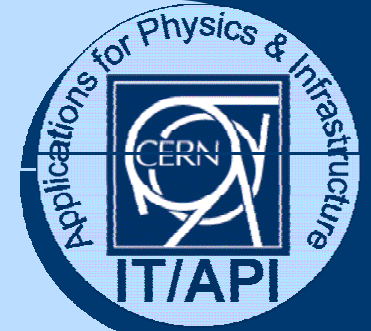
- Transfer data back to CERN -> WP2
- Update meta-database at CERN -> WP2



## 3. Script does the following:

- Copies executable, detector database and cards files
- Executes executable
- Executable creates output dataset
- Output copied to local mass store -> WP5
- Log file copied to web browsable area
- Script calls java program (see 4)

# Summary



- Event level parallelism is required for GEANT4
- A number of issues have to be addressed
- Some of them will be solved in the context of GRID middleware
- Third-party components (persistency and analysis tools) need to be “Grid-enabled”, too.
- The effort required to implement and test this in GEANT4 must be provided – by GRID resources?
- ... see Parallelism in Parallel Sessions ...