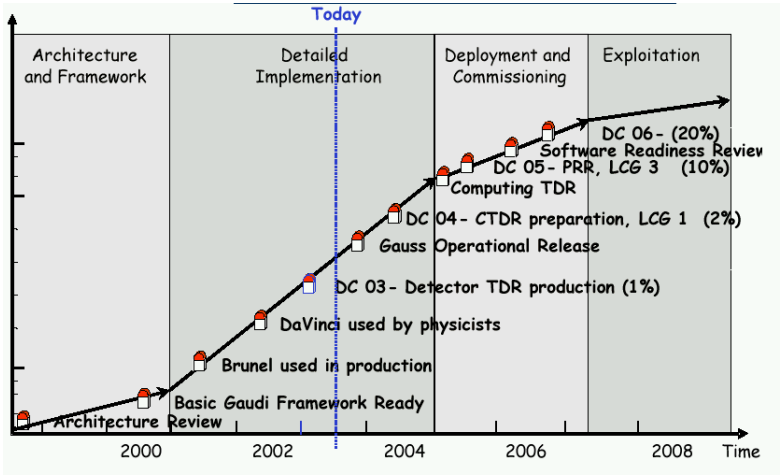
Outline:

- Computing Roadmap
- LHCb Computing Organisation
- Software components
- DC'04 plans
- Conclusions

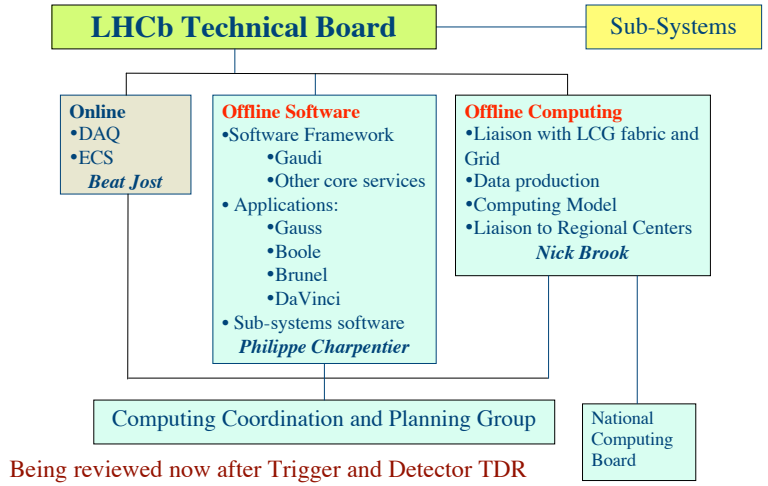
Computing roadmap



LHCb in numbers

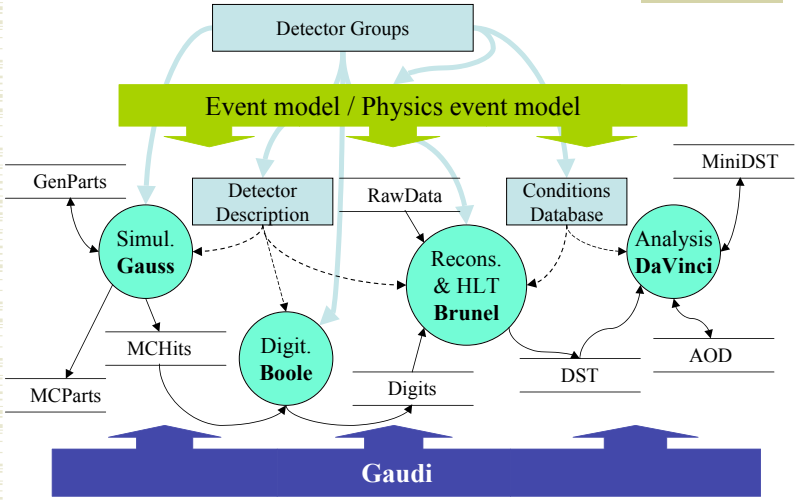
- Number of events per year
  - 2 10^9 events, only a few 10^6 of interest for physics (~25 channels)
  - 3 levels of trigger: L0 (hardware), L1 & HLT (fully software, running on the online selection farm)
  - Independent of the running year (only low luminosity needed)
- Event sizes
  - Raw data: 30-50 kbytes, DST: 100 kbytes
  - Larger for MC data (MC-truth)
- Processing performance
  - Currently
    - Simulation/digitisation: 17.6 (min.bias) / 52 (signal) kSI2k.s
    - Reconstruction: 12.4 kSI2k.s (15s on PIV2.4Ghz)
  - Expected improvement for reconstruction:
    - 1-2s on 2007 machine (factor 8 to 10 compared to now…)

The LHCb Computing Projects



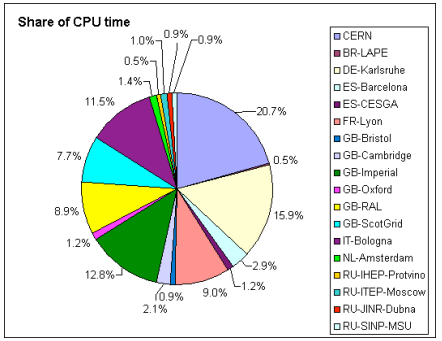
Being reviewed now after Trigger and Detector TDR

Software components

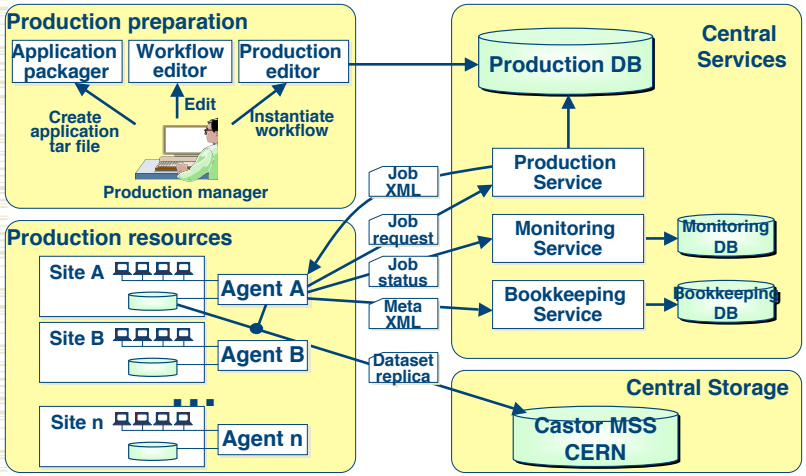


Production and Analysis tools

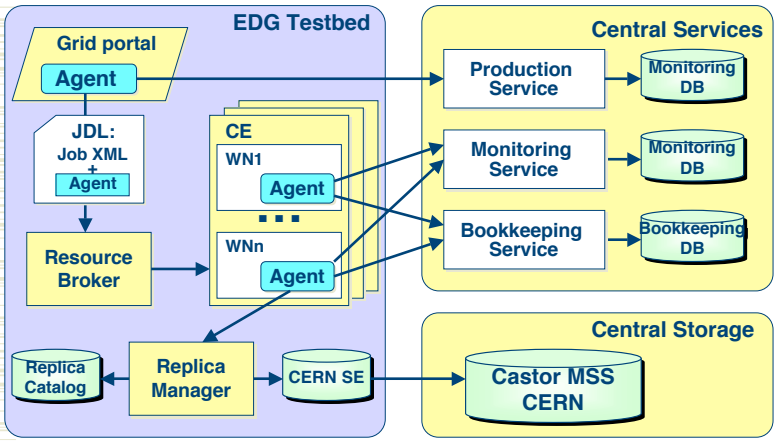
- Production system: Dirac
- Used successfully for DC'03
  - 47 million events
  - 59 days on 17 centres
  - 36600 jobs with 92% success rate
  - 85% produced outside CERN
  - Only very small fraction on EDG (71% failure rate)
- Easy and fast installation
  - Typically 1 day for local support with help from CERN manager
- Production run by 0.5 FTE



Dirac architecture



Dirac and EDG/LCG



Plans for 2004

- DC'04 first aim: provide information for the Computing TP/TDR (June 2005)
  - Emphasis put on Analysis (production has been demonstrated already)
  - Needed to trigger Distributed Analysis: provide more/better quality data to physicists

Event type	Generation	Simulation	L0/L1 reduction	Reprocessing	Time (CPU hours)
Minimum bias	150+11 M	150 + 11 M	0,27% + 100%	0.4 + 11 M	2003333 + 238333
B generic	50 + 4 M	50 + 4 M	100,0%	50 + 4 M	2333334 + 171111
B signal	30 + 0.5 M	30 + 0.5 M	100,0%	30 + 0.5 M	2133334 + 35556
Totals	245.5 M	245.5 M		95.9 M	
				Total time	6 915 001
				Total CPU (SPECint2k * hours)	2 766 000 400

DC'04 storage needs

(Preliminary numbers, fresh from last week…)

Storage (GB)		
Event type	oodigi	oodst
Minimum bias	232 + 6 380	99 + 2 712
B generic	25 000 + 1 833	10 000 + 733
B signal	15 000 + 250	6 000 + 100
Total	48 695	19 644

OOSim (Gauss - MCHits+MCParticles)		
Version	Min. bias	Generic B
	500	1100

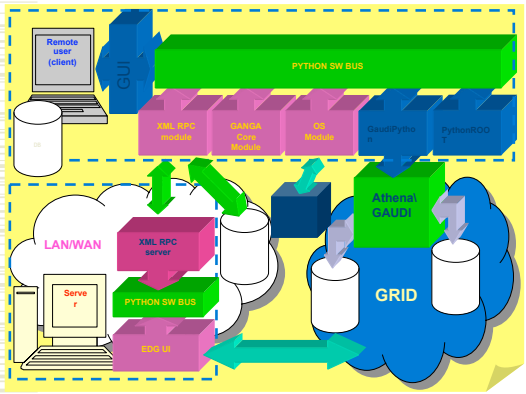
OODIGI (Boole - Digi+MCHits+MCParticles)			
Version	Min. bias	Triggered min. bias	B decay
	160	580	500

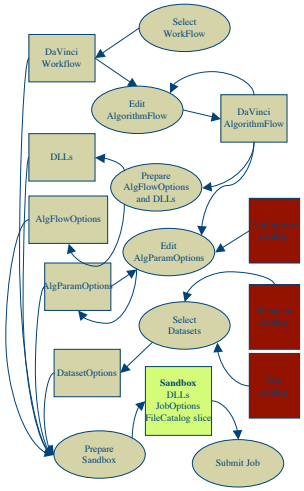
oodst(Brunel - MCParticles, Reco, MCTruth relations)			
Version	Min. bias	Triggered min bias	B decay
	68	247	200

Plans for DC'04

- Most urgent need:
  - File catalogue
    - ARDA file catalogue considered for analysis
    - For data produced on LCG, files would be duplicated (or replicated) on ARDA file catalogue
    - Decision to be taken in December (backup AliEn)
- Current developments
  - Dirac new functionalities
    - Adapted to re-processing, more flexible workflow definition
    - Grid service compliant
  - Analysis: Ganga
    - User wizard for distributed analysis job submission
    - Interactive analysis being investigated (besides classic Ntuple/ROOT)

GANGA principles





Analysis model

- Tried out for the 2003 analysis (Trigger and re-optimisation TDRs), to be improved for DC'04
- Production of DSTs for min. bias, signal samples, bb inclusive background
- Physics group provide (released) software for event preselection (rejection factor ~500)
- Event selection performed as a central "production-like" task
  - Selected streams produced
  - Stripped DSTs produced contain < 100 kevents, hence manageable even on a laptop!
  - Selection eventually replaced by HLT selection

Analysis model (cont'd)

- Full DSTs distributed to all Tier1 centres
  - Re-running stripping in a distributed way
- Stripped DSTs distributed to only selected centres (including Tier2, served by their Tier1)
  - To enforce "distributed" analysis (avoid users take the "easy" route!)
  - Need a file catalogue and data management tools
- Use GANGA as a user analysis wizard
  - Used by selected users in May '04
  - Available to the collaboration in July '04
  - Analysis going on until 2005

Computing in 2007 and beyond

- First pass processing on the Online selection farm
  - Should only use ~20% of the farm capacity (2s @ 200 Hz for 2000 CPUs)
- LCHb plans to use the Online farm for re-processing outside data taking periods (as well as for MC simulation at low priority)
- Needs in terms of MC: to be defined, not clear picture now
- Analysis
  - Plan to use Online streaming of data
  - Small datasets for each analysis team, distributed to Tier2/3
  - Full datasets (real and MC) on Tier1's

Conclusions

- LHCb is certainly less demanding in 2007, nevertheless quite demanding now (for production) in order to prepare the Computing TP/TDR
  - HLT performance assessment (strategy and CPU)
- Analysis model much lighter (multiple exclusive channels with "easy" pre-selection

- LHCb very actively preparing for DC'04
  - First priority on Computing
  - Aim at Distributed Analysis (2nd half '04)
  - No point using software not on the roadmap towards 2007
    - If not possible, time-scale for TP/TDR should be re-evaluated