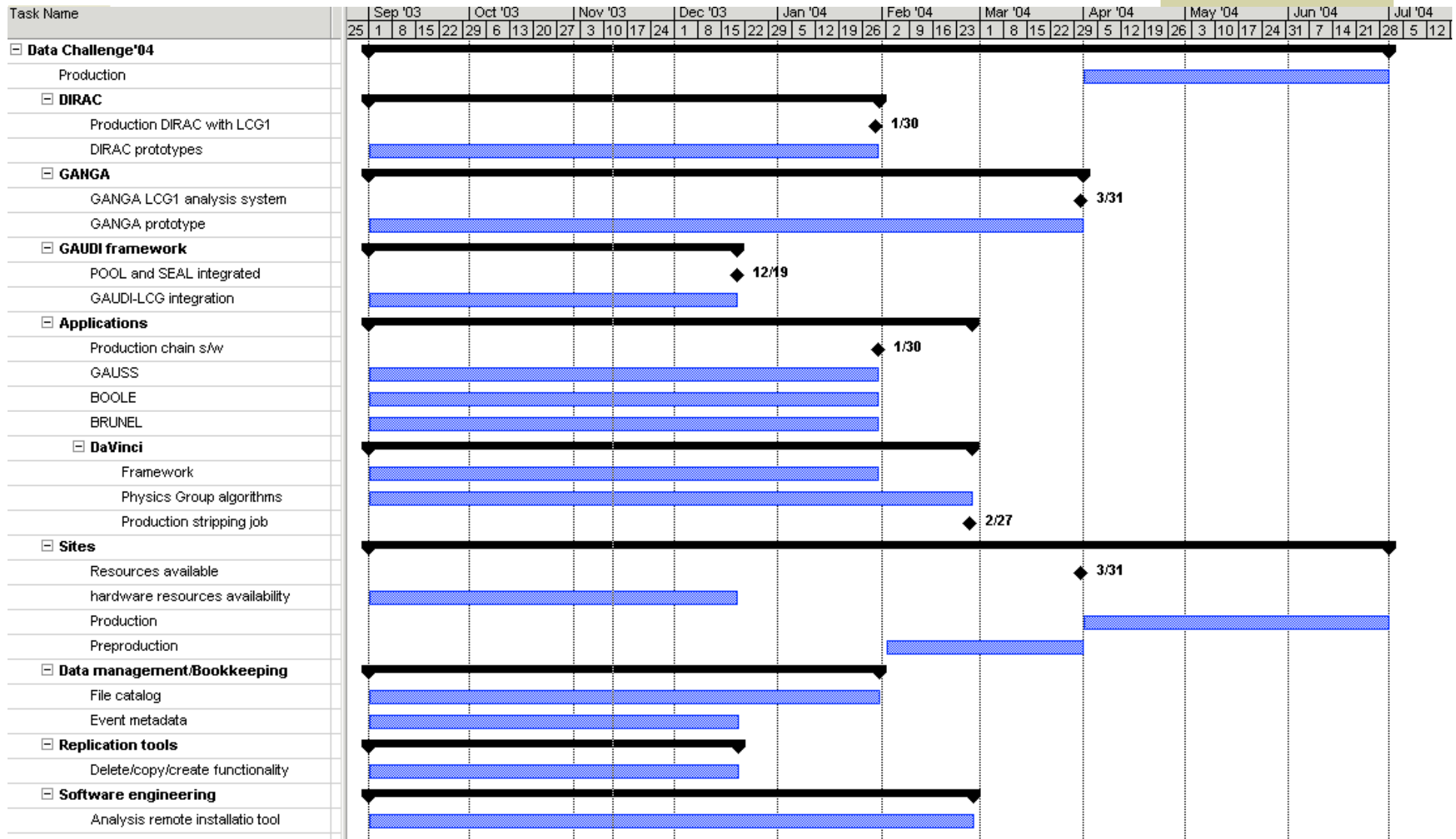**LHCb plans fr DC'04**
Meeting with LHCC referees (26/01/2004)
Ph.Charpentier

Outline:

o Applications

o Data management

o Production tools

o Analysis tools

o Resources

High level schedule

Task Name	Sep '03	Oct '03	Nov '03	Dec '03	Jan '04	Feb '04	Mar '04	Apr '04	May '04	Jun '04	Jul '04
Data Challenge'04											
Production											
DIRAC											
Production DIRAC with LCG1						1/30					
DIRAC prototypes											
GANGA											
GANGA LCG1 analysis system							3/31				
GANGA prototype											
GAUDI framework											
POOL and SEAL integrated				12/19							
GAUDI-LCG integration											
Applications											
Production chain s/w						1/30					
GAUSS											
BOOLE											
BRUNEL											
DaVinci											
Framework											
Physics Group algorithms											
Production stripping job						2/27					
Sites											
Resources available							3/31				
hardware resources availability											
Production											
Preproduction											
Data management/Bookkeeping											
File catalog											
Event metadata											
Replication tools											
Delete/copy/create functionality											
Software engineering											
Analysis remote installatio tool											

Application software schedule

- ○ Gaudi: 19th December '03 (Deadline met)
  - ❑ Dictionary available for the full event model
  - ❑ POOL tested with all persistent classes
  - ❑ Still performance problems (but improving)
  - ❑ Windows version missing
  - ❑ XML file catalogue for production
- ○ Applications: 30th January '04
  - ❑ Framework using new Gaudi, first use of GEANT 4 in production
    - ✰ Still a few issues to fix with Gauss (MCTruth, decay tree…)
  - ❑ Sub-system software: all algorithmic/event model changes ready
- ○ Pre-selection/stripping algorithm: 27th February '04
  - ❑ DaVinci single stripping job, selection results saved

Data management schedule

○ File catalogue technology choice

  ❑ AliEn file catalog as master, RLS used in LCG, XML in WNs

  ❑ Main priority from LHCb for ARDA

○ XML file catalog conversion: 30th January '04

  ❑ Allows to wait for the POOL implementation (needed for analysis)

○ Replication tools: 27th February '04

  ❑ Copy, delete, move files

  ❑ Based on AliEn data management, but must incorporate LCG2 SE (RLS-based)

○ Metadata schema: 30th January '04

  ❑ Based on the existing BKDB (replica table removed, use FC instead)

  ❑ Additions for tag collections, new workflow

Production environment schedule

○ Dirac v2: 30th January '04

 ❏ Workflow definition

 ❏ Workload Management System (Service-oriented, candidate for ARDA sevice)

 ❏ Production management tools, Monitoring

○ Production tools deployment

 ❏ Pilot sites (CERN, Lyon, LCG…): 30th January '04

 ❏ Limited number of sites for pre-production (February)

 ❏ Full scale deployment: 15 March '04

 ❏ Requires outbound connectivity

○ Risks

 ❏ Delays in readiness: many new components, in particular LCG

DIRAC Workload Management & LCG
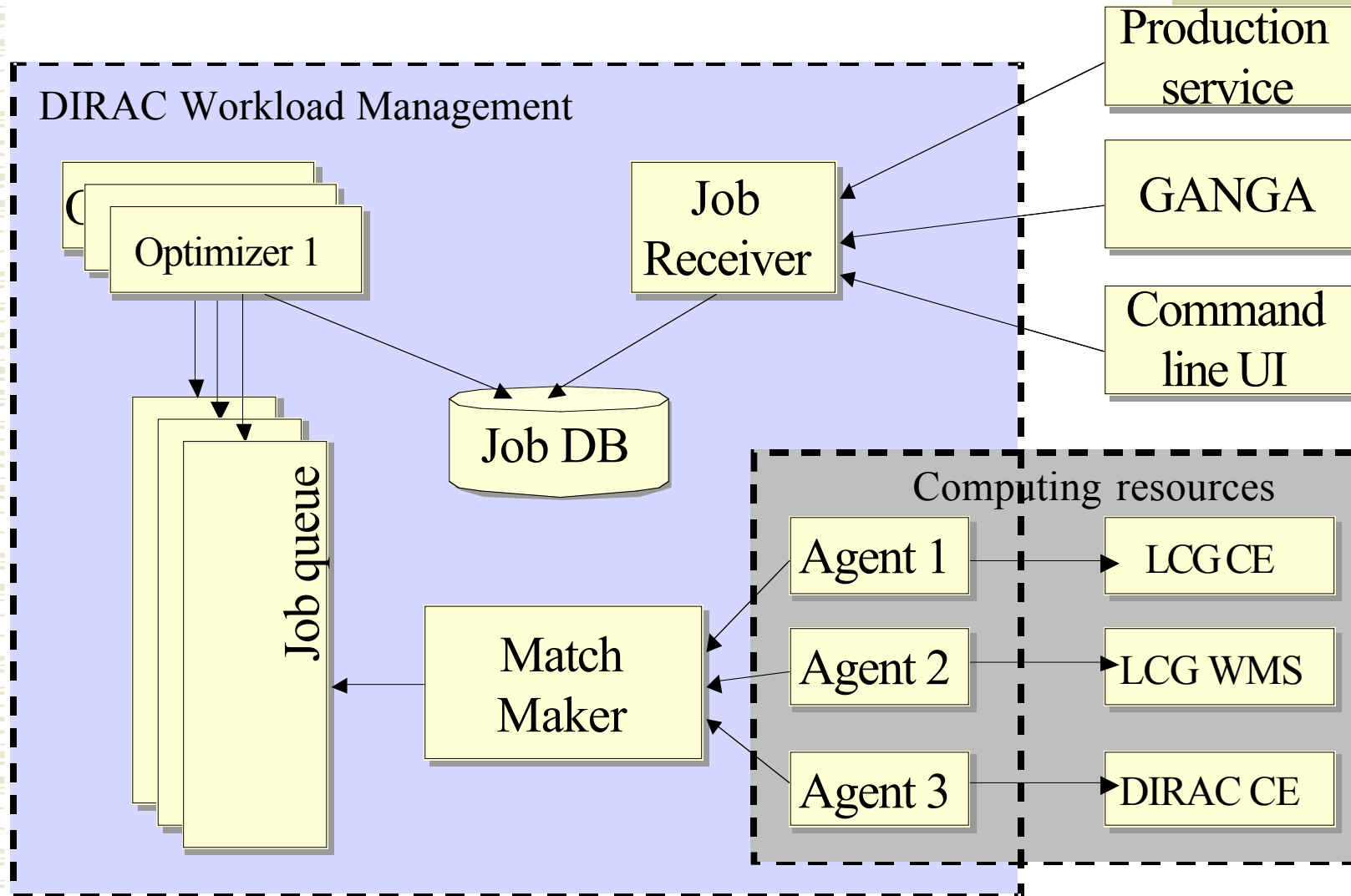
LCG-2 usage

- ○ All production jobs submitted via Dirac

- ○ LCG-2 is a specific Dirac CE (possibly 2: through RB and directly CE)

- ○ As for ALICE, if LCG works well it will take many jobs

- ○ If not, we foresee to refrain non-LCG sites and delay our production in order to properly test and use LCG-2

- ○ Files stored on LCG-SE will be made accessible for analysis also outside LCG

  - ❑ Alien file catalog + RLS

Analysis environment schedule

○ Biggest challenge: based on **GANGA** (ATLAS/LHCb)

○ Must provide improved functionality and user-friendliness compared to "standard" LXBATCH processing

○ Many issues to be settled

❑ Software deployment (user code fast evolving)

❑ How to reproduce the user's environment?

❑ JobOptions edition

❑ Security: user authentication vs group login

○ GANGA functional prototype: 31 March '04

❑ Available to a few test users only, limited number of sites (could bypass security issue providing individual accounts)

❑ First: submission to local batch system, then LCG (and/or Dirac)

Analysis risks

- Schedule is extremely tight!
- GANGA is our baseline as analysis tool
  - Although DC'04 could be analysis using classic tools (CERN LXBATCH processing), this is not the goal!
  - It is mandatory that physicists are committed to test and give feedback to the tool developers
    - It will take many months to get to a fully reliable system
  - Foresee a fast evolution cycle, starting in April '04
  - At the beginning, the turnover will be worse than normal
  - We are not under pressure to provide analysis results
    - If quick results are desperately needed, standard analysis still exists

Resources for DC'04

Computing data challenge 2004
All times normalised to 1GHz PIII processors

Start date:	2004/04/01
End Date:	2004/06/30
Number of days	90

Last update: 26/01/04

Events requested		Simulation		Digitization		Reconstruction		
Type	Total	Time (CPU hours)	oosim size (MB)	Time (CPU hours)	oodigi size (MB)	Time (CPU hours)	Stripping factor	oodst size (MB)
Min bias standard	75 000 000	69 444	650 000	22 917	22 200 000	110 417	0,001	4 800
Min bias special	11 000 000	189 444	1 540 000	3 361	3 256 000	16 194	0,001	704
B generic	50 000 000	3 444 444	27 500 000	26 389	25 200 000	222 222	1,000	8 100 000
B signal	20 950 000	1 443 222	11 522 500	11 057	10 558 800	93 111	1,000	3 393 900
Totals	156 950 000	5 146 556	41 212 500	63 724	61 214 800	441 944		11 499 404

Average CPUs used (Totals/Number of days)	
1GHz PIII processors	2617
SPECint2k	1 046 708

Total time	5 652 224
Total CPU (SPECint2k * hours)	2 260 889 444

CPU time by site		Share of available		% of needed,	Storage required
Site	SPECint2k*hours	by site	by country	by site	(Gbytes)
CERN	851450800	20,7%	20,7%	37,7%	20 051
BR	0	0,0%	0,0%	0,0%	0
CH	19400000	0,5%	0,5%	0,9%	249
DE-Karlsruhe	650 000 000	15,8%	15,8%	28,7%	8 350
ES	205 200 000	5,0%	5,0%	9,1%	2 636
FR-Lyon	367 200 000	8,9%	8,9%	16,2%	4 717
GB-Imperial	400 000 000	9,7%		17,7%	5 138
GB-Liverpool	398 000 000	9,7%	53,6%	17,6%	5 113
GB-RAL	366 000 000	8,9%		16,2%	4 702
GB-ScotGrid	47 000 000	1,1%		2,1%	604
IT-Bologna	432 000 000	10,5%	10,5%	19,1%	5 549
NL-NIKHEF	162 000 000	3,9%	3,9%	7,2%	2 081
PL	75 600 000	1,8%	1,8%	3,3%	971
RU	129 600 000	3,2%	3,2%	5,7%	1 665
	4 103 450 800	100,0%	100,0%	181,5%	61 825

Storage (GB)		
Event type	oosim	oodst
Minimum bias	2 190	6
B generic	27 500	8 100
B signal	11 523	3 394
Total	41 213	11 499

Conclusions

○ Main goals for DC'04:

  ❑ Computing Model definition

  ❑ Experience to be used for the Computing TDR

  ❑ Provide data for HLT studies, Physics performance improvement

○ Time is critical

  ❑ Computing TP/TDR: June 2005

  ❑ Computing Model chapter to LCG: December 2004

  ❑ Analysis experience: from July 2004

  ❑ Production has to start in April 2004

○ Risks

  ❑ Not enough experience with realistic system

  ❑ Implication: revise schedule