



## LHCb thinking on Regional Centres and Related activities (GRIDs)

F Harris



## Overview of presentation

- Country situations and a possible LHCb model for RCs
- Status of EU Grid proposal and LHCb involvement
- Comments on LHCb attitude to Tapes vs. Disks (and some related points)



## Overview of current situation

- **DISCLAIMER** Nothing is 'agreed' in the MOU sense (requires negotiations in collaboration and with funding agencies), but we have the following viewpoint
- We are trying to apply (1/3, 2/3) rule overall
  - Good candidates for regional centres are
    - **Tier1** Lyon, INFN, RAL, Nikhef
    - **Tier2** Liverpool, Glasgow/Edinburgh
  - Discussions going on
    - Russia (?Tier1 for all expts ? Networking)
    - Switzerland (? Tier2 centre for LHCb)
    - Germany (? LHCb use of a national centre)
  - Discussions just beginning
    - Spain
    - Poland
    - Brazil



## Strategy for LHCb country computing planning

- Make case to funding agencies based on
  - Detector etc. studies 2001-2
  - Physics +trigger studies up to startup
  - By startup have facilities in place to match pro-rata requirement for whole expt (see experiment model )
  - Each country has its own constraints (financial, existing infrastructure,etc.) leading to different possibilities for Tier-1/2)
  - Get involved in GRID related activities as appropriate(?manpower)



## For example - planning in the UK

- Computing requirements for 2001-3 for UK/LHCb dominated by detector (RICH+VELO) construction + some trigger optimisation (physics background studies in general start late 2003 but some now)

–	CPU(PC99)	STORAGE (TB)
– 2001	200-400	5-10
– 2002	200-400	5-10
– 2003	400-600	10-20

- Satisfied(?) by MAP(Liverpool) + JIF (all 4 LHC expts)

– JIF proposal (know result late 2000) for all 4 experiments

–	CPU(PC99)	STORAGE (TB)	+ networking enhancement
– 2001	830	25	
– 2002	1670	50	
– 2003	3100	125	



**REAL**

Generates RAW 100 kB  
 reconstructs ESD 100 kB  
 AOD 20 kB  
 TAG ~100+ B  
 stores RAW+ESD+AOD+TAG

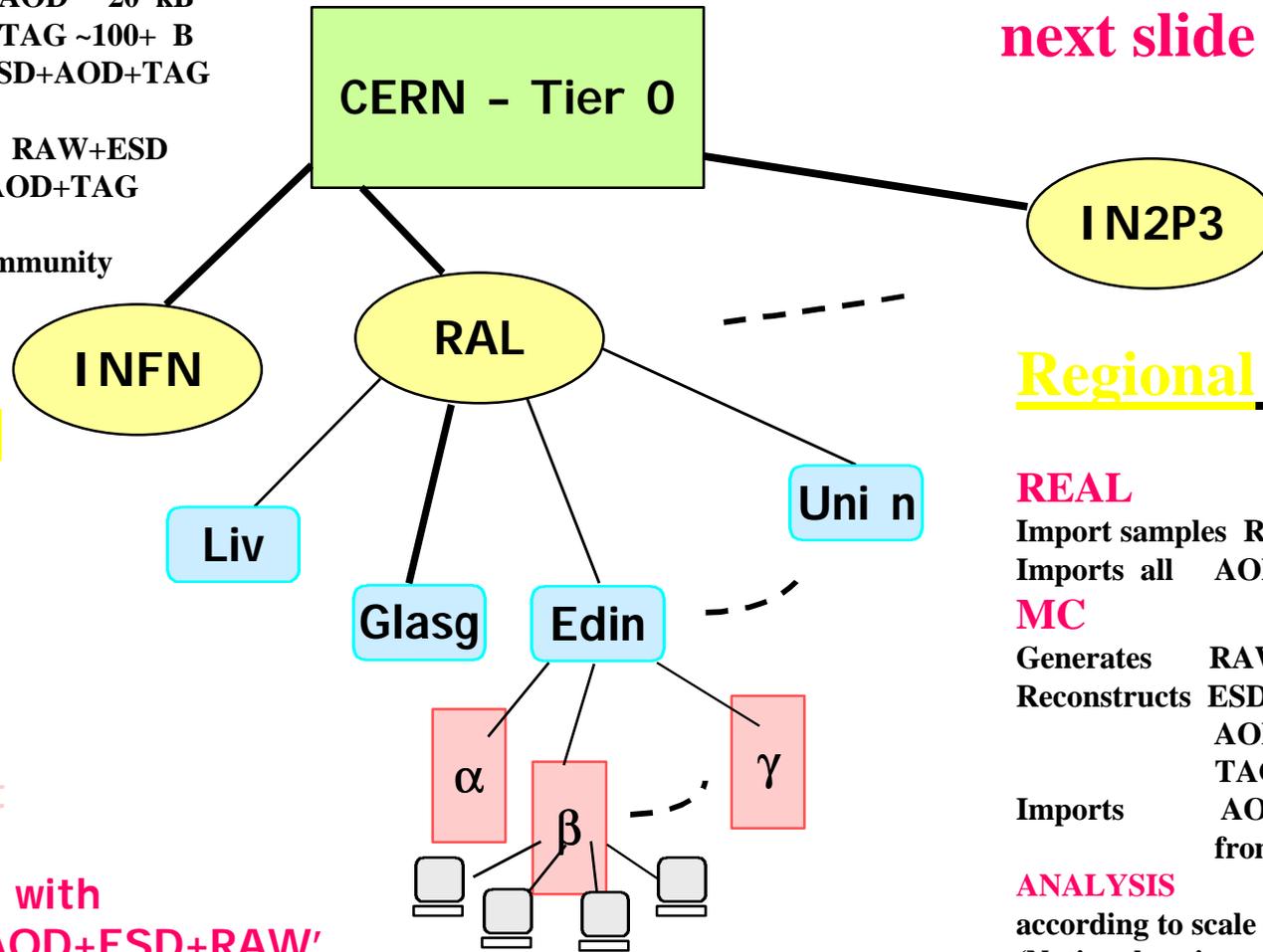
**MC**

Import samples RAW+ESD  
 Imports all AOD+TAG

**ANALYSIS**

For 'CERN' community

**But we want a GRID  
 not a hierachy, see  
 next slide -----**



**Tier 1**

**Tier 2**

**Regional Centres**

**REAL**

Import samples RAW+ESD  
 Imports all AOD+TAG

**MC**

Generates RAW 200 kB  
 Reconstructs ESD 100 kB  
 AOD 30 kB  
 TAG ~100+ B

Imports AOD+TAG  
 from other centres

**ANALYSIS**

according to scale of centre  
 (National,region,university)

Department  
 Desktop

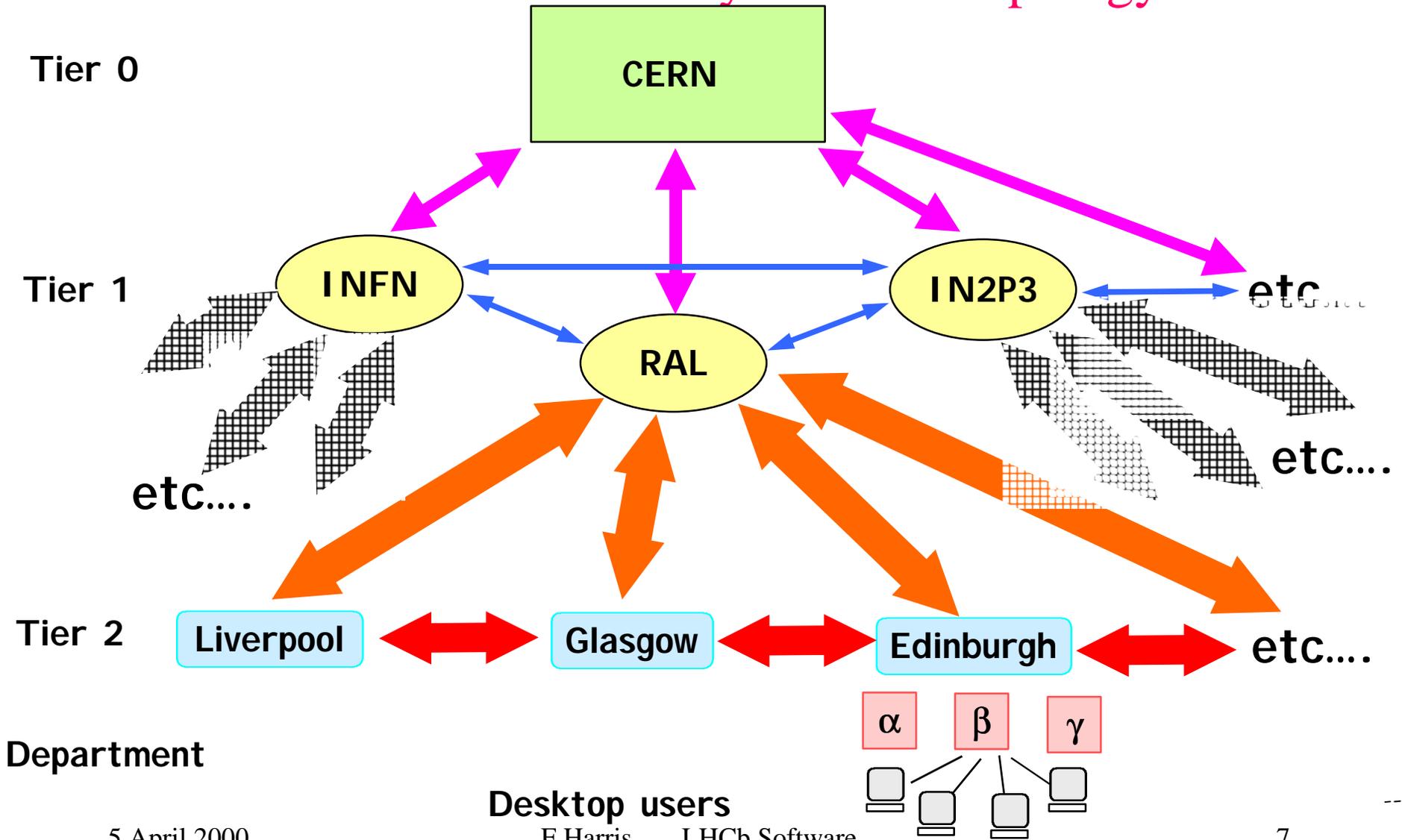
**ANALYSIS with  
 'Ntuples +AOD+ESD+RAW'  
 (10\*\*5 ev take ~ 100 GB)**

5 April 2000

F Harris LHCb Software  
 Workshop



# More realistically - a Grid Topology



5 April 2000

Desktop users  
F Harris LHCb Software  
Workshop



## EU GRID proposal status (<http://grid.web.cern.ch/grid/>)

- EU Reaction to pre-proposal of 30 M Euro - come back with a proposal of 10 M Euro maximum!
- Scaled down proposal being worked on - to be submitted early May
  - Main signatories (CERN,France,Italy,UK,Netherlands,ESA) + associate signatories (Spain,Czechoslovakia,Hungary,Spain,Portugal,Scandinavia..)
  - Project composed of Work Packages (to which countries provide effort)
- LHCb involvement
  - Depends on country
  - Essentially comes via ‘Testbeds’ and ‘HEP applications’



## EU Grid Work Packages

- Middleware
  - Grid work scheduling C Vistoli(INFN)
  - Grid Data Management B Segal(IT)
  - Grid Application Monitoring R Middleton(RAL)
  - Fabric Management T Smith(IT)
  - Mass Storage Management O Barring(IT)
- Infrastructure
  - Testbed and Demonstrators F Etienne(Marseille)
  - Network Services C Michau(CNRS)
- Applications
  - HEP (LHCb involved) H Hoffmann(CERN)
  - Earth Observation L Fusco(ESA)
  - Biology C Michau(CNRS)
- Management
  - Project Management F Gagliardi(IT)



## GRID LHCb WP Physics Study(DRAFT)

- The total sample of  $B > J\Psi/K_s$  simulated events needed is  $\sim 10$  times the number produced in the real data.
- In one year of datataking we expect to collect and fully reconstruct  $10^5$  events, therefore need  $10^6$  simulated events.
- The number of events that have to be generated, stored and reconstructed to produce this sample is  $10^7$ .
- 10% of the ESD data copied for systematic studies ( $\sim 100$  GB).
- The total amount of data generated in this production would be :

RAW data	200 kB/event	$\times 10^7$	= 2.0 TB
Generator data	12 kB/event	$\times 10^7$	= 0.12 TB
ESD data	100 kB/event	$\times 10^7$	= 1.0 TB
AOD data	20 kB/event	$\times 10^7$	= 0.2 TB
TAG data	1 kB/event	$\times 10^7$	= 0.01 TB



## *Grid LHCb WP - Grid Testbed (DRAFT)*

- MAP farm at Liverpool has 300 processors would take 4 months to generate the full sample of events
- All data generated (~3TB) would be transferred to RAL for archive (UK regional facility).
- All AOD and TAG datasets dispatched from RAL to other regional centres, such as Lyon and CERN.
- Physicists run jobs at the regional centre or ship AOD and TAG data to local institute and run jobs there. Also copy ESD for a fraction (~10%) of events for systematic studies (~100 GB).
- The resulting data volumes to be shipped between facilities over 4 months would be as follows :

Liverpool to RAL	3 TB (RAW ESD AOD and TAG)
RAL to LYON/CERN/...	0.3 TB (AOD and TAG)
LYON to LHCb institute	0.3 TB (AOD and TAG)
RAL to LHCb institute	100 GB (ESD for systematic studies)



## Thoughts on mass storage usage (see our note)

- We would like as much active data online on disk as possible
- Use tape for archiving 'old' data (? Some have suggested all disk systems- but how do you decide when/what to throw away)
- R/D - try strategy of moving job to the data (Liverpool COMPASS)
- ? If 2.5 Gb/s networks prove not to be affordable then we may need to move data by tape. Don't want to do that if possible!