



# Group Data Processing on the Grid

Pavel Nevski

Brookhaven National Laboratory

# Group Production

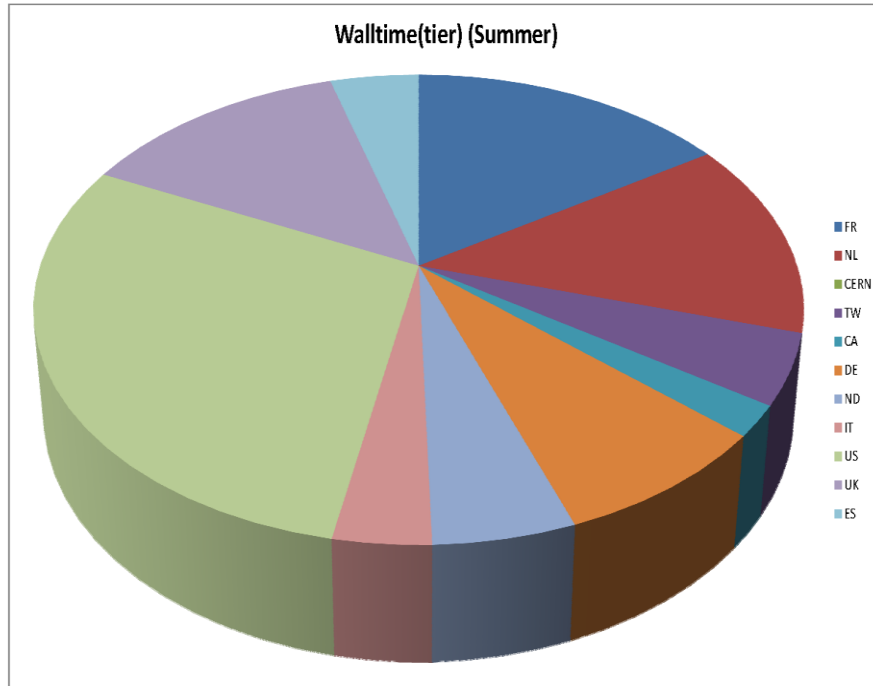


- With the arrival of LHC collisions data, there is a growing need to organize data processing for physics groups in the most efficient way
- Common coordination is in place between physics groups and ATLAS Distributed Computing to support group production
- Our Grid Production system provided the following benefits:
  - Access to computing resources world-wide
  - Guaranteed reproducibility of results
  - Full integration with Data Management system allows automated data delivery to the final destination
  - Organized 24/7 support in case of problems
  - Flexible usage of group resources (mainly disk, CPU is less an issue)
  - Automatic resubmission of failed jobs minimized fraction of events lost
- Group Production is not limited to a common version of the software:
  - Dedicated group production releases are available and centrally supported
    - See Prafulla's talk at this session

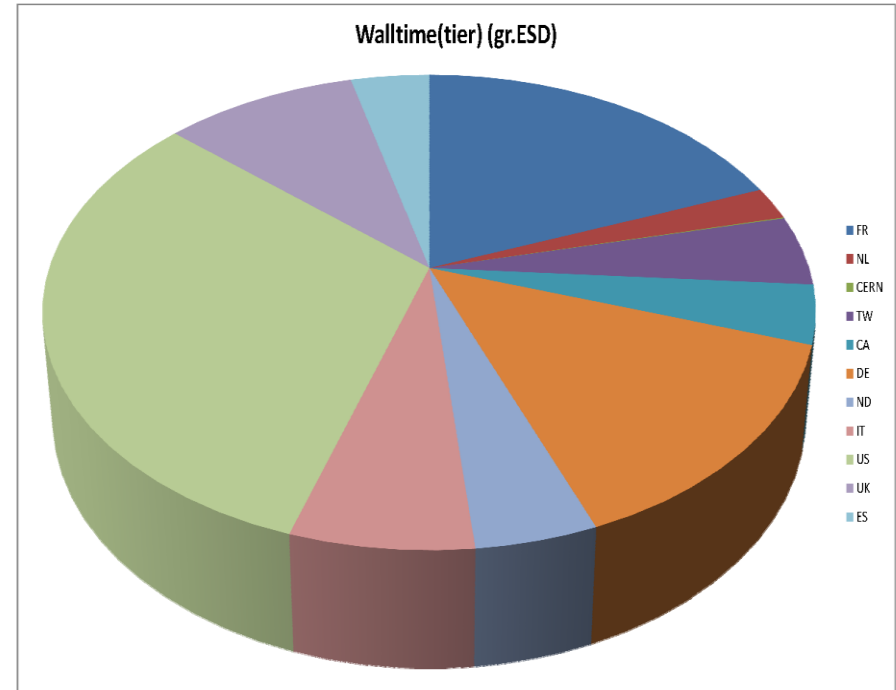
# Resource Sharing



Summer'10 reprocessing  
35,255 CPU-days (wall-time)



Group Production on TO ESD  
66,556 CPU-days (wall-time)

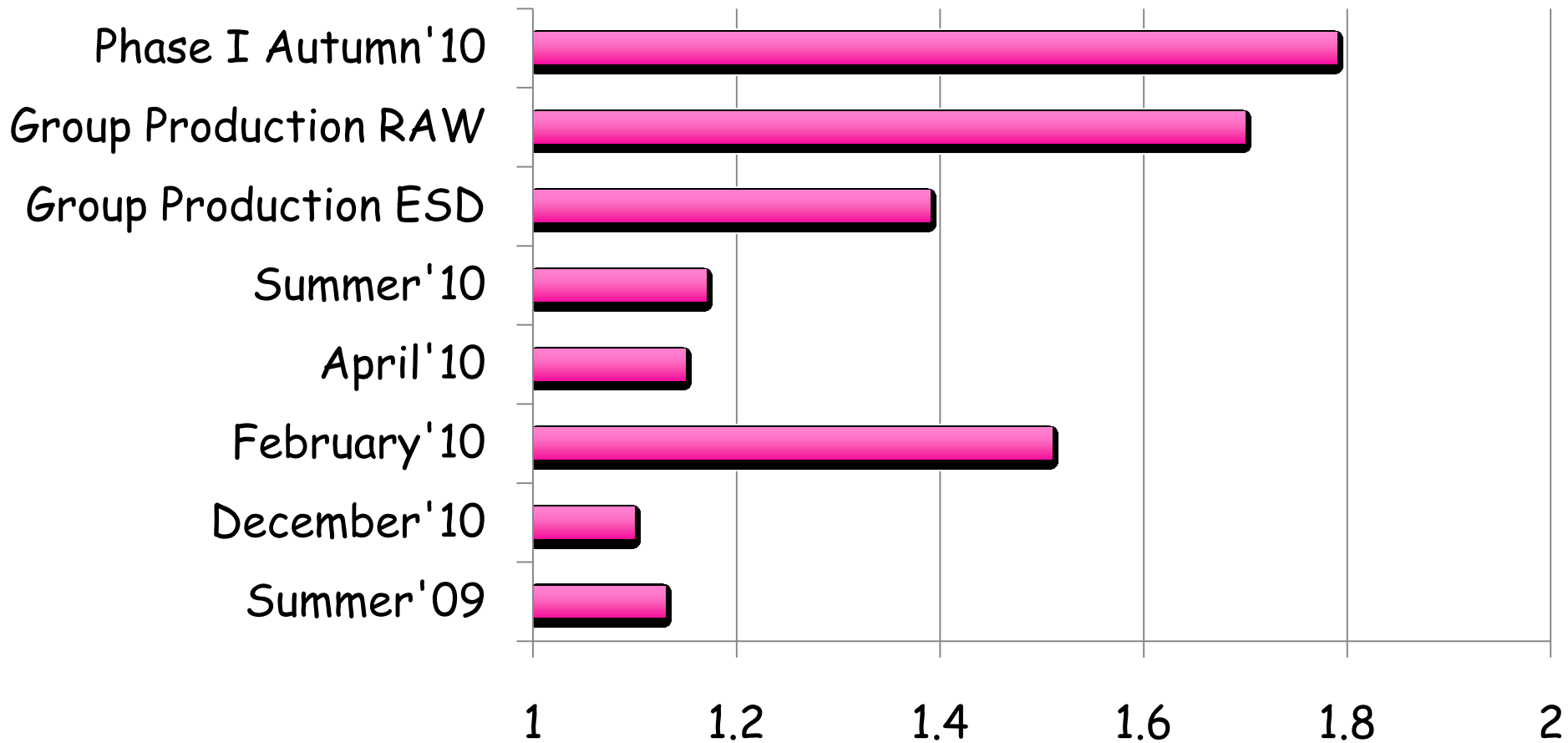


- Already in Summer 2010 Group production used more CPU then the central Data Reprocessing

# Some Pitfalls



- Average number of attempts necessary to minimize event losses is determined by many factors



# Some Guidelines



- Group data processing is running with “no data loss” policy.
  - i.e. Job Autoabortion is switch off
  - We run only on validated sites
- It is the groups responsibility to insure that software is mature enough and no CPU is wasted
- Please monitor your requests closely and stop submission (or terminate tasks) if you see permanent failures
- Fill free to discuss with Junji and me if you are insure what the problem is , do not expect problems to miraculously disappear