ATLAS Computing @ Liverpool

Carl Gwilliam

UNIVERSITY OF LIVERPOOL

3rd July 2008

Liverpool Computing Meeting

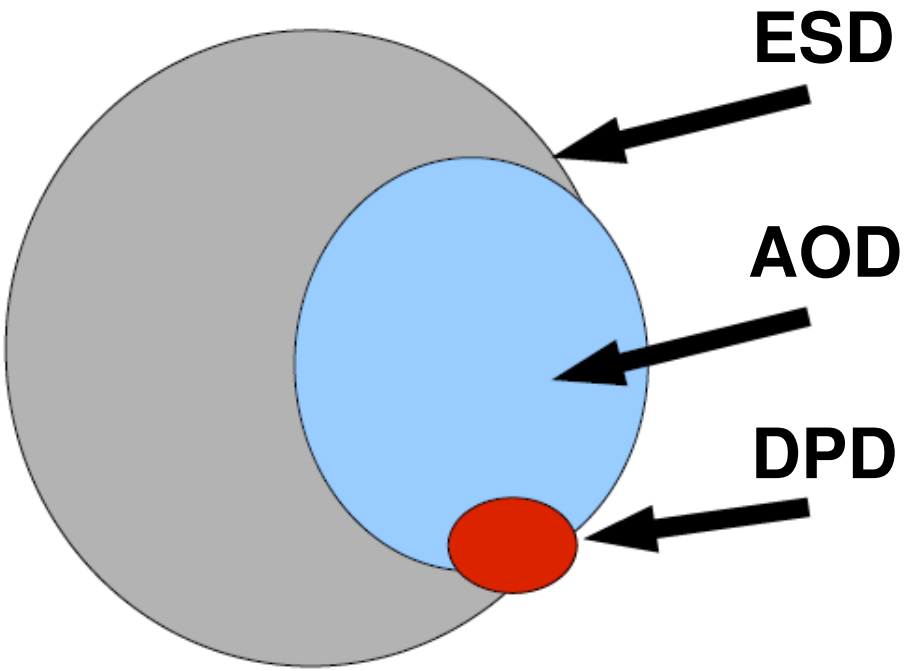
Introduction

- We as the HEP group need to decide on what our computing needs will be in the near (12-18 months) future:
  - Want to make sure we can make the best use of first LHC data.
  - Will also be important for rolling grant proposals.
- We therefore need to think about the likely ATLAS needs over this period.
  - This includes local needs, both interactive and batch, and grid needs
  - It includes storage, CPU and RAM.
- Above all the infrastructure must be easily scalable.

ATLAS Data & Computing Models

Event Sizes (Data)

- RAW
  - ≈1.6 MB/Event
- ESD (Event Summary Data)
  - ≈ 800 kB/Event (± 20%)
  - 500 kB/Event target
- AOD (Analysis Object Data)
  - ≈140 kB/Event (± 20%)
  - 100 kB/Event target
- DPD (Derived Physics Data)
  - ≈ 10 kB/Event, but several flavours.

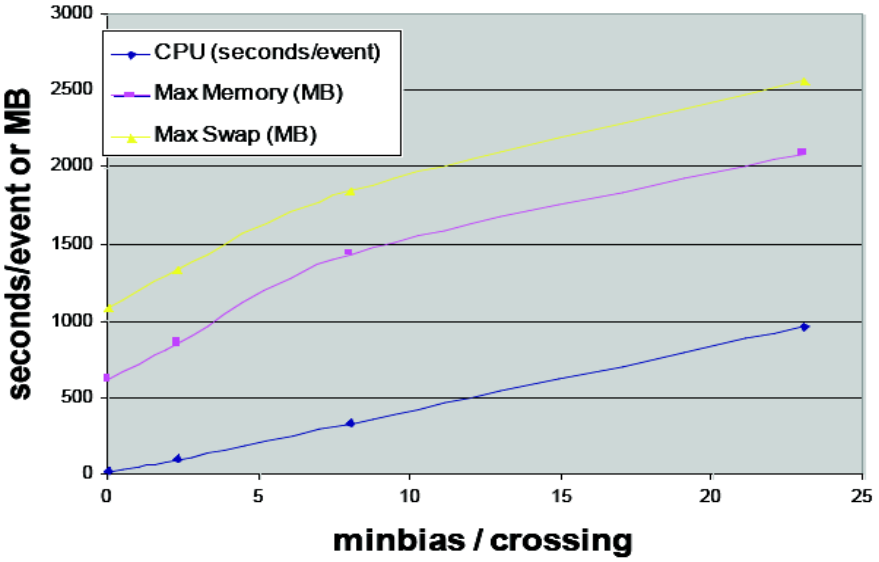


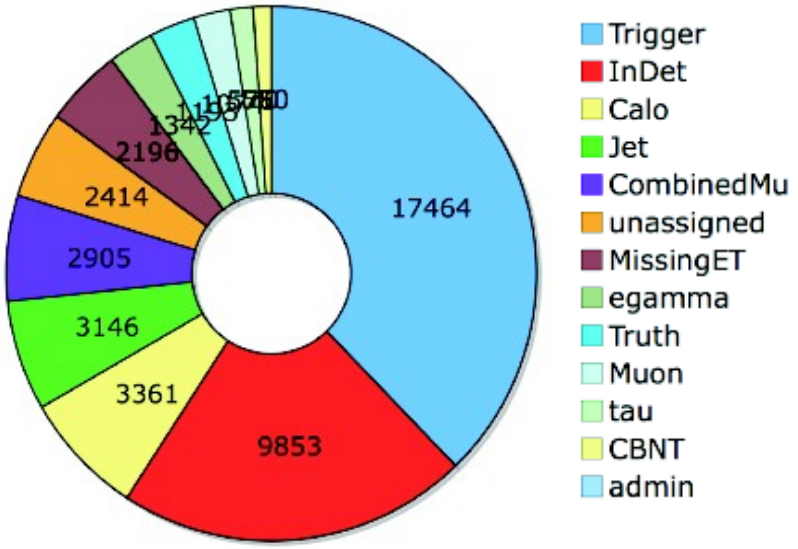
Numbers don't include truth or pile-up!

Timing & Memory Performance

- Digitisation
  - Timing:
    - 27 kSi2K/event (no pile-up)
  - Memory:
    - ≈ 600 MB (no pile-up)
- Reconstruction
  - Timing:
    - 10^{31} (no pile-up): 46 kSi2K/event
    - $2x10^{31}$ (pile-up): 79 kSi2K/event
    - Target: 15 kSi2K/event
  - Memory:
    - Always struggles stay below 2GB

≈ 6s on typical processor





Tiers of ALTAS

- Tier 0
  - Prompt first pass processing on express/calibration & physics streams
  - Restricted for production
- Tier 1 (10)
  - Reprocessing with improved calibration and software
  - Restricted to physics/detector **groups**, not individuals
- Tier 2 (40)
  - Monte Carlo Simulation, producing HITS, ESD, AOD
  - On-demand **user** analysis of AOD datasets, limited access to ESD & RAW
- Tier 3
  - **Local** resources for analysis of **small** DPDs
  - More and more sites are developing and using such systems
  - 20% of UK T2 use prioritised for UK users but mechanisms not yet ready

Jobs **must** go to data!

Initial Data Rates and Volumes

- If we assume:
  - Event sizes: 1.6Mb (RAW), 1 MB (ESD) and 0.2 MB (AOD)
  - Running time: 10 Hours/day for 20 days/month
  - Trigger Rate: 200 Hz
- The resulting data rates are:
  - 11.5 TB/day (RAW), 7.2 TB/day (ESD), 1.4 TB/day (AOD)
- Given we expect about 3 months of running this year this gives:
  - 690 TB (RAW), 432 TB (ESD), 84 TB (AOD) for a single copy
- We also need to factor in about 30M fullsim MC events (=> 6TB AOD), group DPDs and "user" space.

Grid Data Distribution

- T1s contain:
  - Full RAW (10% disk, 90% tape).
  - 2 full copies of ESD.
  - Full AOD at each T1.
  - Full set of group DPDs.
- T2s contain:
  - A total of 30% (10%) of RAW and 150% (30%) of ESD in 2008(9).
  - A further full AOD per T2 cloud (e.g. UK), placed by analysis interest.
- For us, this equates to ≈30TB + MC & group/user space in 2008. For 2009 it's more like 55TB + MC & group/user space.

Requirements for Liverpool

First Guess: Batch

- Short queue
  - 10-20 cores with a 1 hour time limit
  - Not same as interactive since guaranteed all CPU/RAM (but flexible)
- Main queue
  - Currently heavily used so at least equivalent of now (40), probably more.
  - Time limit of 1 day + fair share policy to ensure everyone gets their turn
  - Easy scalability so can add more in future.
- Long jobs (several days)
  - By request on current non-lcg nodes (need to be easier to access).
- High memory
  - Will need high memory queue (at least 2GB/core) for e.g. reconstruction.

First Guess: Interactive

- 20-30 or so cores of which maybe half should be at CERN
  - The ones at CERN would need storage. How much depends on size of data as input to the interactive stage (we have to copy it from Liverpool/grid).
- The amount obviously depends on the split between batch and interactive usage, which we don't really know yet.
  - Given ATLAS data sizes, probably more need for batch (but depends how much we're forced to do on the grid).
- Multi-core machines are best as give more choice on usage.
- Test PROOF as a tool for parallel interactive (point & click) root analysis.

First Guess: Storage

- Our current storage is
  - ≈20 TB local, half of which is already full
  - ≈37 TB grid, of which about 30TB is for ATLAS
- We plan to increase the grid storage to ≈100TB before data-taking, with ATLAS being the main user.
- Local storage will soon be filled even with modest amounts of DPD => likely to need more.
- We can, however, also access the grid storage from local jobs.
  - This still needs to be tested with multiple users.
  - Usefulness depends on what data ATLAS places locally.
- Keep in mind: grid has never really been tested with chaotic user analysis and will also have to cope with data reprocessing.
  - Can we really rely on it from beginning? Do we need more local storage?
- Takes 1-2 months to order+test more so don't need to buy it all now.