Thoughts on APEL cloud accounting

It would be nice if we could answer a question like:

How many HS06 did ATLAS use at CERN between 1.1.2015 and today including:
- Batch resources
- Cloud cpu resources
- Dedicated resources (voboxes etc)

?

Thoughts on APEL cloud accounting

See John Gordons proposal on extensions of the records.
Personal view:

- StartTime and EndTime should refer to the VM and used to find out if a VM was running (thus counted) in the period in question or not

- Add SampleStartTime and SampleEndTime for the record. Cputime etc should report the usage within this interval

- To reduce the effect missing records add
    - Cumulative cpu time, cumulative network usage etc

Benchmarking

- Benchmarking is tricky but can be made to work if certain conditions are met:
    - No cpu overcommitment on the hypervisor
    - No cpu emulation
- The idea is to gather hw information on the VM itself, and get the rating from a lookup table
- This is done on the CERN batch farm for both physical and virtual worker nodes

Classification of VMs

Combine information relevant for performance:

- Vendor (AMD, Intel)

- OS version

- CPU type

- Number of cores

- Memory speed

Getting the reference values

- Benchmarking itself is done at CERN like this:
  - For new batches of hardware we redo the benchmarking
  - New worker nodes are created in a spare partition
  - All are benchmarked **at the same time** with large statistics
  - The most pessimistic values are taken as reference values