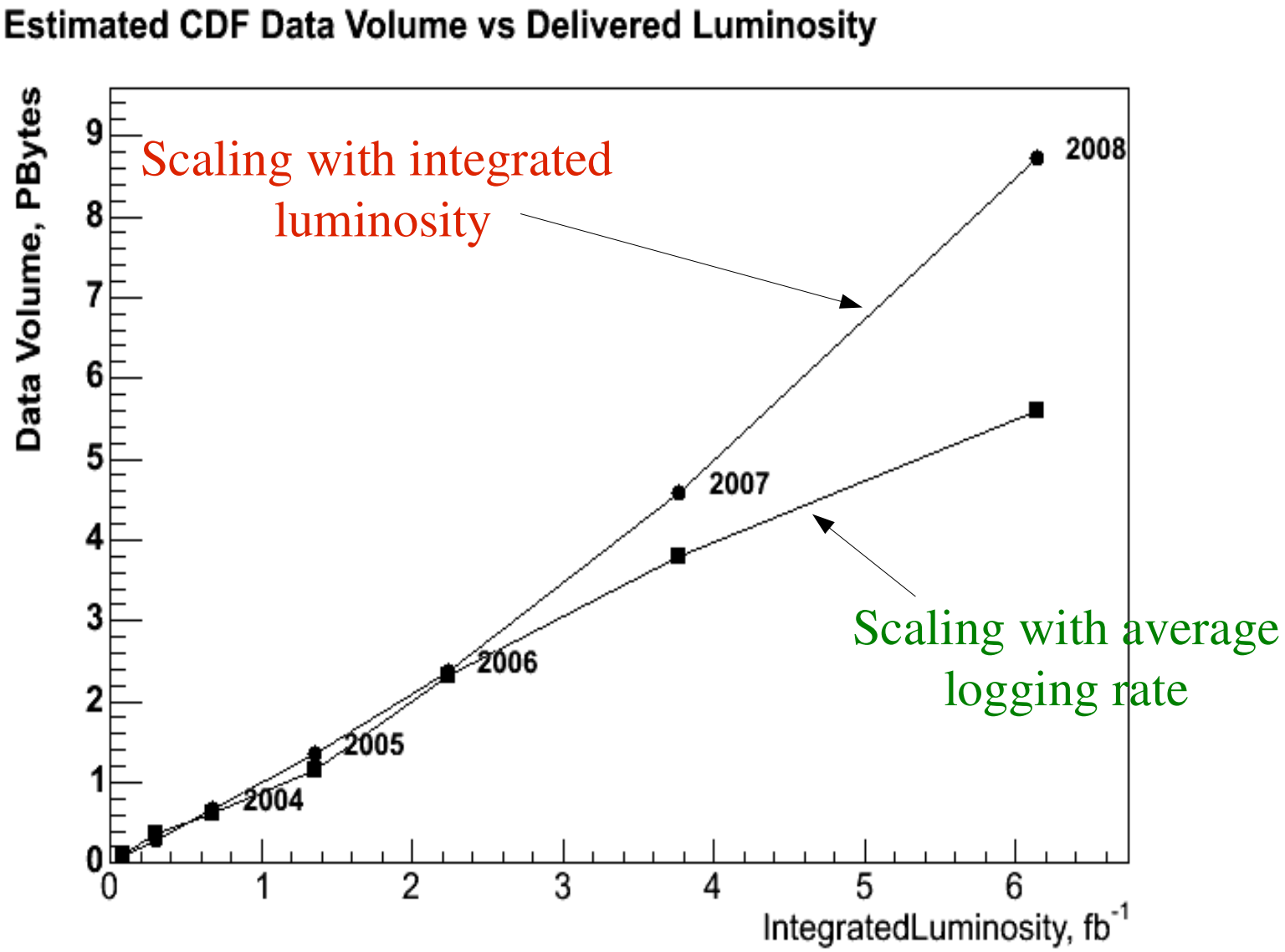
Answers to Review Questions

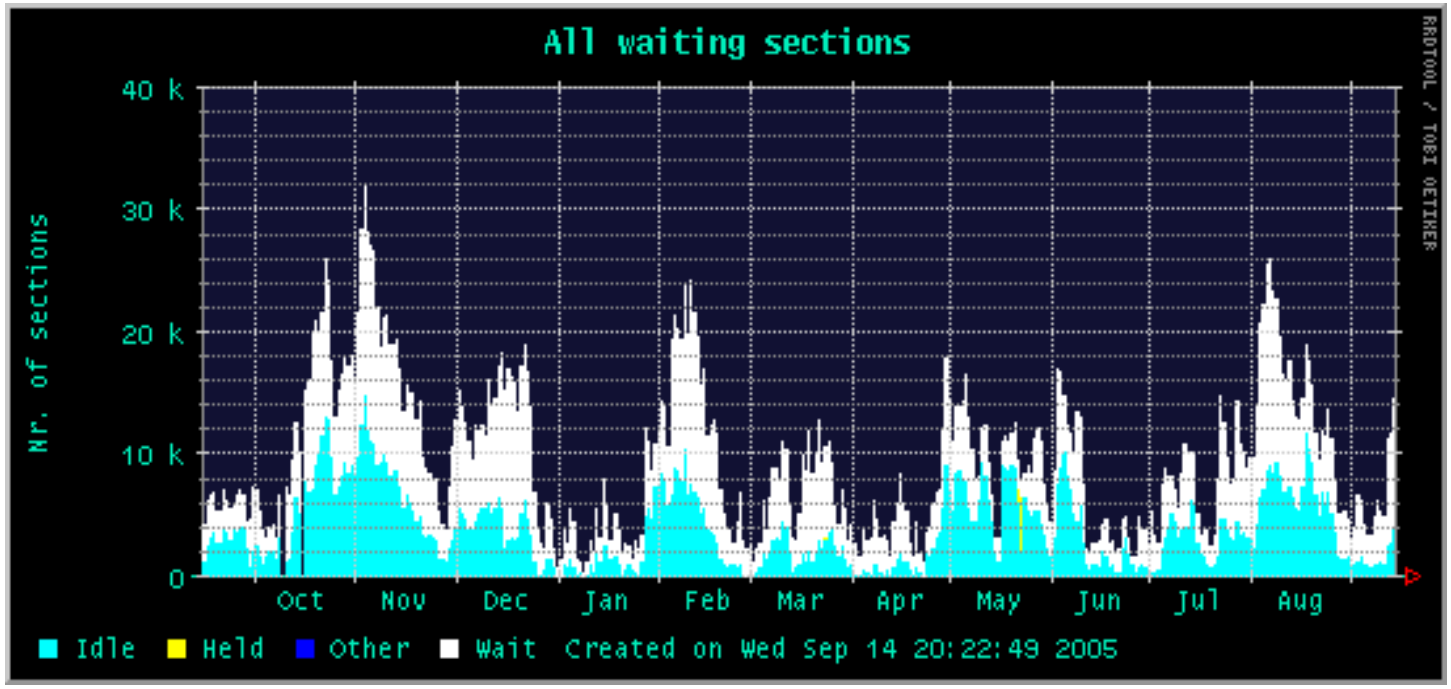
- Question 1: CDF computing model – can you explain the assumptions that go into this plot? If the data rate is determined by trigger bandwidth, we don't understand the plot.
- Answer: We have revised the plot by scaling dataset size with average logging rate -
  - 2004 30 Hz
  - 2005 60 Hz
  - 2006 120 Hz
  - 2007 120 Hz
  - 2008 120 Hz

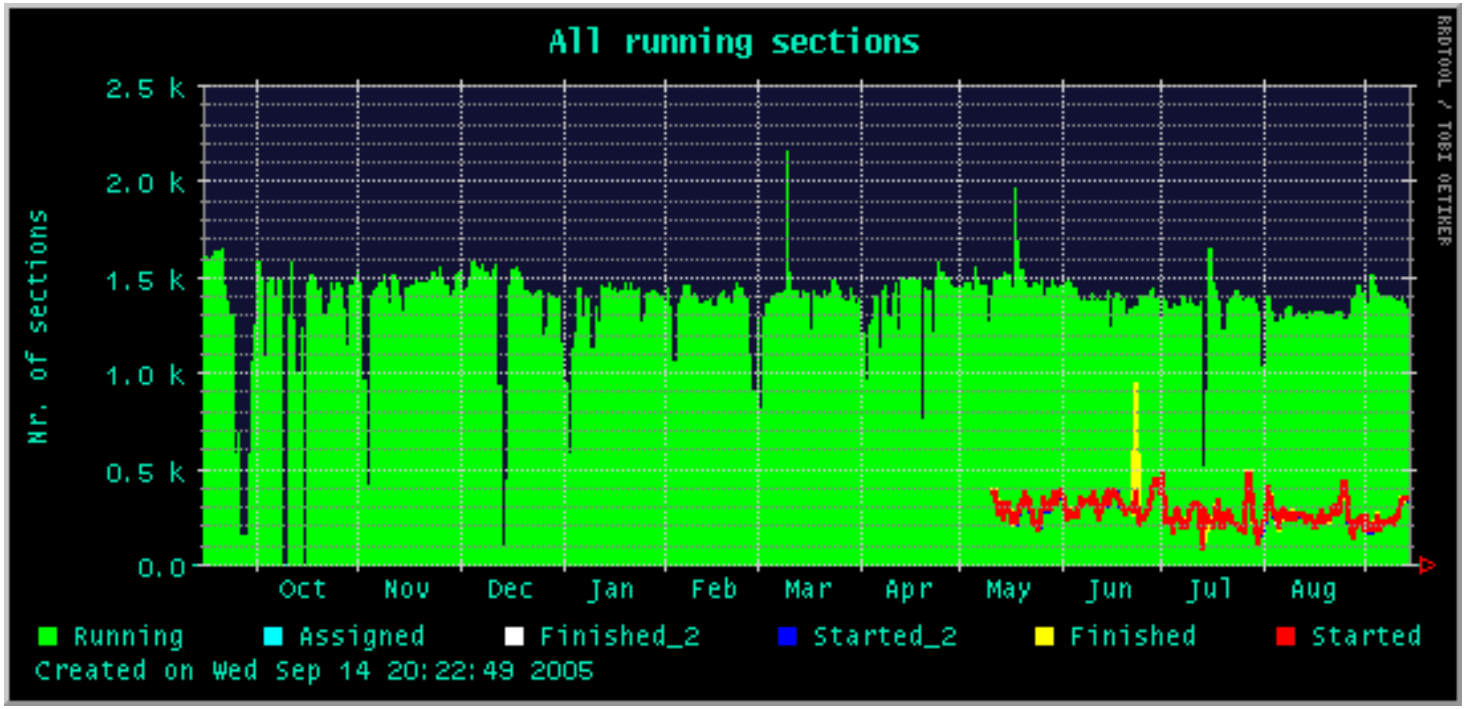

Estimated CDF Data Volume vs Delivered Luminosity
Data Volume, PBytes
IntegratedLuminosity, fb-1
Scaling with integrated luminosity
Scaling with average logging rate
2004
2005
2006
2007
2008
0
1
2
3
4
5
6
7
8
9

- Question 2: Disk space vs tape – There are requests for large disk capacity, eg. 2.2 PB in 2008. How have you done the optimization of tape vs disk?
- Answer: The ratio of disk to tape in our model is maintained at approximately 1:4, which is the ratio CDF has experience with in the last few years of operation. We have found that this ratio allows us to maintain a cache miss rate of ~10%, which we have found to be a good operating point from the physics perspective (latency of data access for analysis).
- We are planning to exploit the caching ability of dCache more fully in the coming months by shrinking the pinned DST datasets and reorganizing the dynamic dCache pools tuned to the physics group access patterns. We expect to learn about the optimization through this process.
- In addition, CDF, D0 and CD are interested in developing a caching model to help predict the optimum disk/tape ratio.

- Question 2: Are you folding in disk management costs?
- Answer: No, however
  - The number of servers is not expected to increase linearly with dataset sizes, due to increasing disk capacity, and due to retirements
  - The new file servers are expected to be more reliable than the existing 3Ware servers, based on the experience of CMS. This information is factored into the recent decision to purchase ATABeast servers.

- Question 3a): Analysis model – do you really need to make ntuples in one month? What is the impact if you stretch this out to level the load?
- Answer: We have used large scale ntuple production as an estimator of the total analysis needs. Our experience has been that ntuple production occurs in bursts, triggered by
  - Conferences
  - Updates to user analysis code and high level algorithms such as b-tagging, tau identification, or improved jet reconstruction and energy scale
- This analysis pattern is supplemented more steady state operations
  - Ntuple reading and analysis
  - Miscellaneous activities like trigger studies, top mass fitting etc.
- The combination of the 'burst mode' activities and other, more steady analysis patterns has historically led to full DC utilization of the CAF.


All waiting sections
Nr. of sections
40 k
30 k
20 k
10 k
0
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Idle
Held
Other
Wait
Created on Wed Sep 14 20:22:49 2005
RRDTOOL / TOBI OETIKER


All running sections
Nr. of sections
2.5 k
2.0 k
1.5 k
1.0 k
0.5 k
0.0
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Running
Assigned
Finished_2
Started_2
Finished
Started
Created on Wed Sep 14 20:22:49 2005
RRDTOOL / TOBI OETIKER

- CDF Computing Usage Task Force did a detailed study of analysis CPU usage and concluded that
  - 50% of CAF usage is reading DST's to make ntuples
  - 20% was user MC
  - 30% miscellaneous
- This is consistent with our assumption that large-scale ntuple production provides a way of modelling the total average analysis needs with the appropriate scaling properties.
  - The model incorporates the different CPU needs of B physics ntupling and high p_T physics ntupling
- For 2005, ntupling-based model predicts need of 3.3 THz as CPU of analysis CAF, to be compared with actual 3.2 Thz

- Question 3b): Is the current usage of the analysis farm dominated by ntuple production?
- Answer: yes, as explained on the previous slides.

- Question 4) Analysis off site. How much disk space at remote sites (like LCG) will be required to support this model? Are the sites going to be able to support this, especially as LHC ramps up?
- Answer: Our initial plan is to pin specific datasets, eg. Specific B physics datasets, at remote sites and use those sites as B physics analysis centers
- As an example, the hadronic B dataset for 1/fb is ~35 TB. Our Italian collaborators have indicated that Italian dCAF has 40 TB available for CDF now.
- Similarly, Karlsruhe and Japan have 30 TB and 10 TB for CDF.
- For the longer term, we will discuss at the International Finance Committee meeting next month the commitment of disk and CPU for CDF.

- Question 5) Reconstruction vs analysis off-site. Is it really more difficult to do reco off-site than analysis? (It is easier to get CPU than disk at off-site locations)
- Answer: Yes, for a number of reasons
  - Expert manpower (1-2 FTE per site) required to preserve integrity of primary reconstruction (CDF also does splitting)
  - Reconstruction needs massive and continuous shipment of data both to and from the remote site. Analysis in the mode of pre-staged, pinned datasets for specific physics topics is a simpler problem.
  - Some local disk needed for primary reconstruction as well, for staging and concatenating data and making efficient use of remote CPU
  - Opportunistic use of CPU can be exploited more easily for analysis compared to primary reconstruction. The latter requires a predictable schedule
  - It is more important for CDF to find additional CPU for analysis compared to reconstruction (because analysis needs are factor of ~3 bigger)
  - Problem of moving analyses off-site is factorizable into moving "one" analysis at a time, can clone simple solution many times
  - Bologna, Karlsruhe, Rutgers and Wisconsin have expressed interest in hosting physics datasets

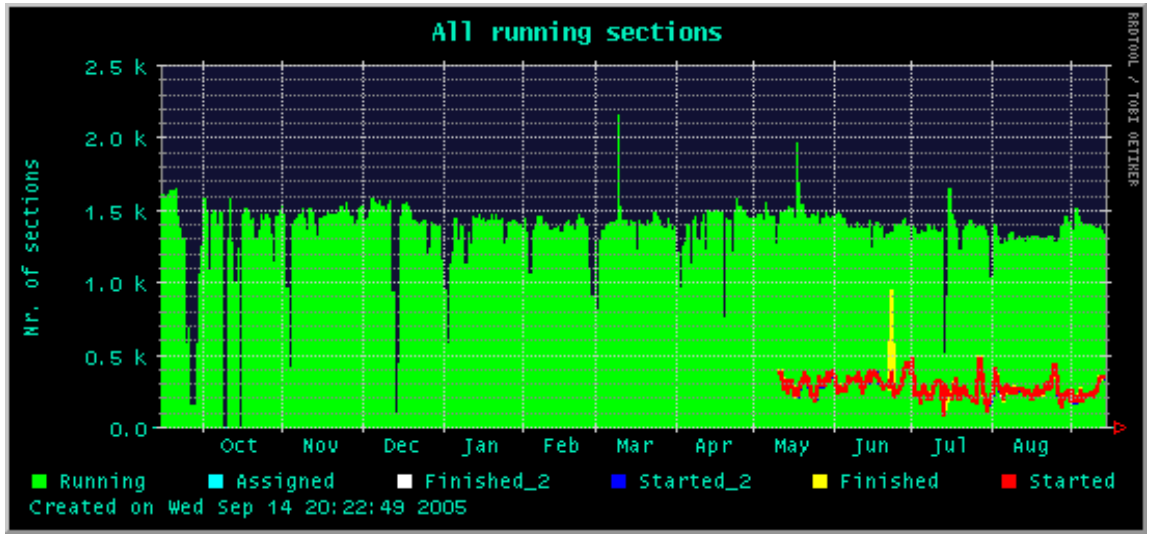
- Question 6) What are the requirements for the Condor Glide-in?
- (i) Does it require privileged installation?
- (ii) Do worker nodes need external network connectivity?
- Answer: We need a Globus Gatekeeper to enter the computing element. No other requirements on the GRID site.
- We install and maintain our headnode, like for any other CondorCAF. This requires root access during installation (Kerberos) but can be operated without root privileges afterwards. This is a limitation of the CondorCAF and is not Glide-in specific.
- Worker nodes require outgoing connectivity. In the current release, the headnode must be able to talk to the worker nodes, but this requirement should be removed soon.

- Question 7) Frontier requirements:
- (i) What are the requirements on the resource provider to install this? Will all sites agree to this install?
- Answer: Resource provider (eg. DCAF) is required to a “FronTier Squid” (a CPU running a web-based application) which provides the user with read-only access.
- We have asked all remote sites to install squids – many did, some did not yet.
- Even if there is no Squid installed at a remote site, the CDF code will work, and the jobs will be accessing the Fermilab Squid and not Oracle.

- Question 8) SAM and CDF. Adam's slide 18: There are large factors in file storage requirements between CDF and D0. It sounded like the suggested solution is to have a more gradual ramp up to this scale. Does CDF agree that this is a workable solution?
- Answer: yes, CDF agrees.

- Question 9) Looking at the SCS Ganglia link for the last year: http://fnpca.fnal.gov/ganglia/?r=year&s=descending&c there is a cluster called 'CDF-Farms'. What resource does this describe – analysis cluster, reconstruction, other?
- Answer: It describes the reconstruction farm + test farm

- Question 10) What is the average utilization of the reconstruction and analysis compute resources for the last year?
- Answer: for analysis farm, >90%



- for reconstruction farm, ~90% (big gaps correspond to reconfigurations)