FY13-14 Hardware Projection Inputs

Collecting up the inputs from the various projects:

Fermi:

- 200 cores DC for Level 1 data processing; 500 TB/yr storage
- Full reprocessing of Pass8 in CY13: ~6 months with 3000 cores (current allocation)
- no new compute nodes
- ~1 PB disk, few PB tape
- 2 new Oracle servers
- no plan yet for next full reprocessing

CDMS:

FYI: SuperCDMS currently has 164 TB of disk space spread out over
6 disk servers. Right now we use about half of it. We also have a reserved
allocation of 115 cores.

The SuperCDMS cluster needs model has several components:

1/ Analysis has traditionally not relied very much on MC (for various
reasons, some good). There is work going on to change that but it's
slow progress. However, I do have some estimates for needs:

FYI:
Single job i.e. single event:
Memory: ~500MB
Average time: ~20 mins (yes, that is correct)

Major production:
Events: 600k
Size on disk: 1 TB
Total time: ~6 weeks at current load
(400 - 600 jobs parallel on avg)
Total expected in 1yr: ~ 2
Total expected to be stored on disk in steady state: ~3 - 4

Specialist study:
events: 100k
size on disk: 15 GB
Total time: ~1 Week@ 500 cores avg
Total expected in 1 yr: ~4
Total expected to be stored on disk in steady state: ~8 - 10

Summary:
Currently, the avg number of cores available is fine (although at peak
times we get ~3000 cores and need to throttle throughput because the disk
can't keep up). In the steady state I expect to use no more than 6TB of
data. If we find that we are running many more SNOLAB simulations, I
expect to go through a cycle of deprecating old simulations, where we
through away all intermediate raw data which reduces the size of the
simulation on disk by ~ factor 10.

When we switch to a monolithic Geant4 solution, average time per job
should be reduced slightly and per-job memory requirements should be
reduced significantly (no promises yet).

2/ MC support for SNOLAB R&D: Difficult to estimate both number of events
and frequency since it's R&D driven.

3/ Possible test facility data processing: No estimates, but should be
small (cores and data volume). Possible start is late this summer and
until SNOLAB starts up. This processing would be time critical, but
should be below the current 115 core allocation.

For 1/ and 2/ we don't really have a required DC level, but rely on
getting a lot of cores for a limited amount of time. 1/ is not so time
critical so if we got less cores we would just run longer (up to a certain
point).

4/ Data processing for SNOLAB: We don't know how much data we will have
(readout and trigger schemes are being discussed), and no decision about
where to do it (SLAC or Fermilab). I am gathering information about
current processing speed and how different algorithms are expected to
scale when we get x10 more detectors.

DarkSide-50:

We don't have a great "model" for our needs yet.

We will need 10TB to hold the Veto data for analysis this year - this might
grow by 10TB/year over the next two years.
We will also need 5-10 TByte of space to support the data acquisition test
stand that we are going to build. For this, I propose to repurpose parts of
the 15TB SuperB allocation (which can also be initially used to support the
veto analysis) - the remaining part of the SuperB allocation is still being
used for the FDIRC detector R&D.

CPU needs are expected to be extremely modest. The DAQ test stand will be
self-sufficient in terms of CPU. For the veto analysis I would expect an
average of less than 10 core-years over the next year, utilized in frequent
few-hour peaks of 20-30 cores. We can provide an update as soon as we have
reconstruction software in hand and better estimates.

EXO:

draft report on EXO historical usage and future
predictions of SLAC computing resource usage here:

https://docs.google.com/document/d/1jPNwYQZHM20HIkd08PJCZ-j0GbF_Nd_MHmVVmc9TNRg/pub

ATLAS:

ATLAS Tier-2.

Currently funded at $600k flat-flat per year including labor, recharge and
hardware. Historically (this has left about $225k/year (net of SLAC
purchasing charge) to go to vendors.
Baseline guidance/agreement with US ATLAS is to spend 40% on CPU and 60%
on disk. This is often changed dramatically at the last moment depending
on where ATLAS is feeling the resource crunch. Taking the 40:60 guideline
and using the planning spreadsheet developed two years ago, we expect
acquisitions of:

2013 2014 2015 2016
CPU HS06 7806 10019 13259 17252
Disk TB 772 991 1311 1706

Dis TB are always net delivered to physics.

ATLAS Group

Currently has about $100k worth of Ariel Schwartzman's early career award
invested in a Proof Cluster. This is proving very valuable for
data-intensive analysis jobs. No serious ongoing planning has been done,
but keeping this facility viable in the face of increasing ATLAS data
looks like costing ~$35k/year on average.

KIPAC:

I did a very cursory stab at a model for kipac.

I calculated the number of "power-user-equivalent" (PUE) users and propose to say each PUE needs 32 cores and 3 TB/year.

We have about 50 PUE's at kipac yielding 1600 cores and 150 TB/year.

DES storage needs add 200-300 TB/year which I think has no budget.

DES computing is partly represented above and really should be separate. I'll try to do a better job next week when I can talk to Risa.

In the end we are using 1500 cores and burning 150TB/year of space. As cores get better and storage costs go down I expect people to get better at using them and filling up disks so that overall usage will increase at something like 15%/year.

HEP Theory:

for theory we could start from the HPC white paper, it applies to SLAC
computing as well, since 50% of the authors reside in Bldg 48 😉

We estimated about 500k CPU hours in 'pleasingly parallel' mode per one
of Tom/JoAnne's pMSSM analyses. There are likely to be at least two per
year, plus debugging, testing and training new students, say three.
I know from Ahmed that data handling can be a problem, as they
copy/unpack/pack lots of small gzip'ed files before/after analysis.

For QCD we estimated 150k CPU hours per project, with typically ~10 such projects per year. Right now I am processing most of this on other machines, because the NFS at SLAC is often very slow and we also have problems writing/reading many small files (order 1000-10000). Having MPI available would be perfect, we can make use of as few or as many nodes as the system provides.

In general I think the computing model for theory is flexible, we can mostly adjust to what is available. If we have the chance to shape some of the new system, then my vote would be for improved I/O first and good MPI support second. I guess this overlaps with your needs and the needs of KIPAC.