Viewpoint from a University Group

Bill Lockman

SCIPP

Santa Cruz Institute for Particle Physics

with input from: Jim Cochran, Jason Nielsen, Terry Schalk

WT2 workshop

April 6-7, 2009

Outline

Introduction

Personnel and activities

Hardware projection

Sites

Configurations

Conclusions

Introduction

Up until a couple of months ago, SCIPP had not given serious consideration to building a T3

Our plan was to:

- Use lxplus at CERN
- Use the grid
- Use SCIPP computing hardware at CERN
- Use available cycles at SLAC T2 (20% of 1200 cores)

ATLAS computing model anticipates T2 resources may not be sufficient to handle production and D^3PD-based analysis demands

- T3 needed to offload demand (a more "*flexible*" and "*nimble*" model)

⇨ Revisiting the T3 option in light of new estimates from forthcoming T3 report

UCSC ATLAS Physicists

Facility (5):

Alan Litke[1], Jason Nielsen, Bruce Schumm, Abe Seiden, Hartmut Sadrozinski

Staff (2):

Alex Grillo, Bill Lockman

Postdocs (2):

Sofia Chouridou[1], Jovan Mitrevski[1]

Graduate Students (5):

Andrea Bangert, Daniel Damiani, Ken Fowler, Gabe Hare[1], Peter Manning

[1] currently at CERN

Physics Directions

Standard Model (SM) physics:

- Underlying event: Hare, Nielsen
- W($\mu\nu$) + jets: Chouridou, Nielsen
- Z(ee) + jets: Fowler, Nielsen, Seiden

New Physics (NP):

- GMSB ($\gamma\gamma$ + missE_t): Baggert, Damiani, Litke, Mitrevski, Nielsen, Schumm
- Universal Extra Dimensions: Manning, Seiden

UCSC input to US ATLAS model

Input: # of analyses associated with the specified stream started in the specified year

performance ESD/pDPD at T2	# (2009)
e-gamma	(1)*
muon	
track	
W/Z(e)	2
W/Z(μ)	2
W/Z(τ)/missE_t	
gam-jet	
minbias	(1)*

physics stream (AOD/D^{1}PD) at T2	# (2009)	# (2010)
e-gamma	1	1
muon	1	
jet/missE_t	1	

*not included in spreadsheet calculation

estimated DISK/CPU resources at T3

Resources required to perform a single 1 hour pass on the D^3PDs at the T3:

2009	2010
29 cores	169 cores
10 TB	60 TB

#cores (kSpecInt2K-s) = [# of events]•[MC factor]•[1/(transform rate)]

#TB = [# of events]•[MC factor]•[D^3PD size/event (TB)]

#events includes a stream /perfDPD reduction factor

transform (D3PD plots) rate = 10kHz on 1 KSpecInt2K CPU

MC factor = 5 = data + 4•data MC

D3PD size/event (TB) = 5KB/event•10^{-9} TB/KB

This represents 5×10^9 events per hour in 2010

T3 Site

The SCIPP T3 site is not obvious. At least 3 choices:

Site:	**Advantages:**	**Disadvantages:**
SCIPP	•cooling, power, space is provided •cost of management/support	•limited cooling, power, space •probably can't scale past 2010 size •limited and shared connectivity (1Gb/s) •support probably not 24/7
SLAC	•proven track record •24/7 support •Direct connection to SLAC T2 •load sharing possible	•cost of power, cooling, management
NERSC	•existing hardware, infrastructure, management	•we have little experience with NERSC

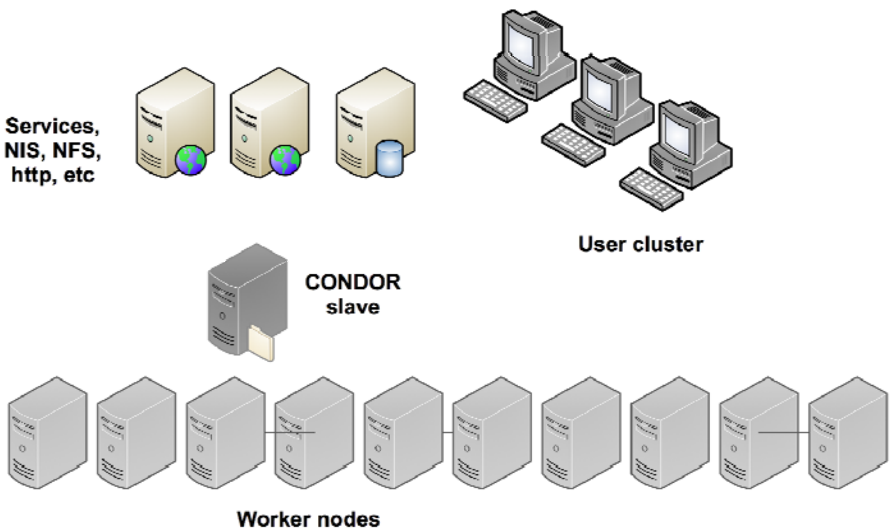
This maps into the type of T3 center (T3w, T3g, T3gs, T3af) we envision for UCSC

Tier-3 types

	T3w	T3g	T3gs	T3af
stands for:	workstation	grid access	grid services	analysis facility
approx. number of cores	$\sim 8 - 32$	> 80	> 168	limited by agreement
format	towers	towers [ANL model, see E.2.1] or rack [Duke model, see E.2.2]	rack	rack
storage capacity	$\sim$ few TB	$>$ 20TB	$>$ 30TB	limited by agreement
clustered? batch?	no or minimal	yes, headless workers	yes, headless workers	yes headless workers
interactive ROOTtuple analysis?	yes	no	no	no
MC, e.g. $t\bar{t}$	few hundreds ATLFAST in hours; millions generator in hours	few thousands ATLFAST in days; millions generator in hours	few millions ATLFAST in days; many millions generator in hours	few millions ATLFAST in days; many millions generator in hours
data production capability?	no	no	yes	yes
support level	owner/ group	group	group/dept professional	lab professional
network rating	100Mbps	$\geq$ 1 Gbps	10Gbps	10Gbps
software, services	ROOT Athena	ROOT, OSG Athena Local Resource Manager (e.g. Condor, PBS, ArCond[see E.2.1]) DQ2 endpoint "outsourced" catalog, subscription	ROOT, OSG Local Resource Manager (e.g. Condor, PBS) robust network file system (e.g. dCache, xRootd) DQ2 site services	same as T3gs
specialized cooling/ power	none	none (towers) CRAC (rack)	CRAC 10's kW	facility
costs	$\geq$ 20k	$\geq$ 30k	$\geq$ 80k	negotiated

Strawman T3G system

Figure 23: Generic tower-based T3g system.



top: towers: 20TB 120kSI2K
bottom: rack: 24TB 160kSI2K (5KW)

Table 24: Minimal strawman T3g system. Such a system would provide approximately 120kSI2k processing.

component	typical model	quantity	unit cost, k$
switch	Cisco 1GB	1	2.5
worker towers	Intel-based E5410 2.33GHz, 2 TB storage 8GB RAM	10	2.0
server elements	DELL PE1950 E5440 processor, 2.83MHz, 16GB RAM, 250GB drive	4	0.5
total cost			$24.5k

Table 25: Strawman T3g system designed to fit into an already existing rack. Other systems are certainly possible. At added expense and slightly reduced capability, but with considerable simplification in cabling, etc., a blade-based system would fit in a rack as well. Such a system would provide approximately 160kSI2k processing.

component	typical model	quantity	unit cost, k$
UPS	DELL	1	1.0
switch	DELL PowerConnect 48GbE, portmanaged	1	1.5
servers	DELL PE2950 E5440 processor, 2.83GHz, 32GB RAM, 250GB drive	1	4.2
compute elements	DELL PE1950 E5440 processor, 2.83GHz, 16GB RAM, 250GB drive	10	2.4
storage elements	DELL MD1000	2 (24TB, usable)	5.4
total cost			$41.5k

Strawman T3GS system

Figure 21: Generic single-rack T3gs system.

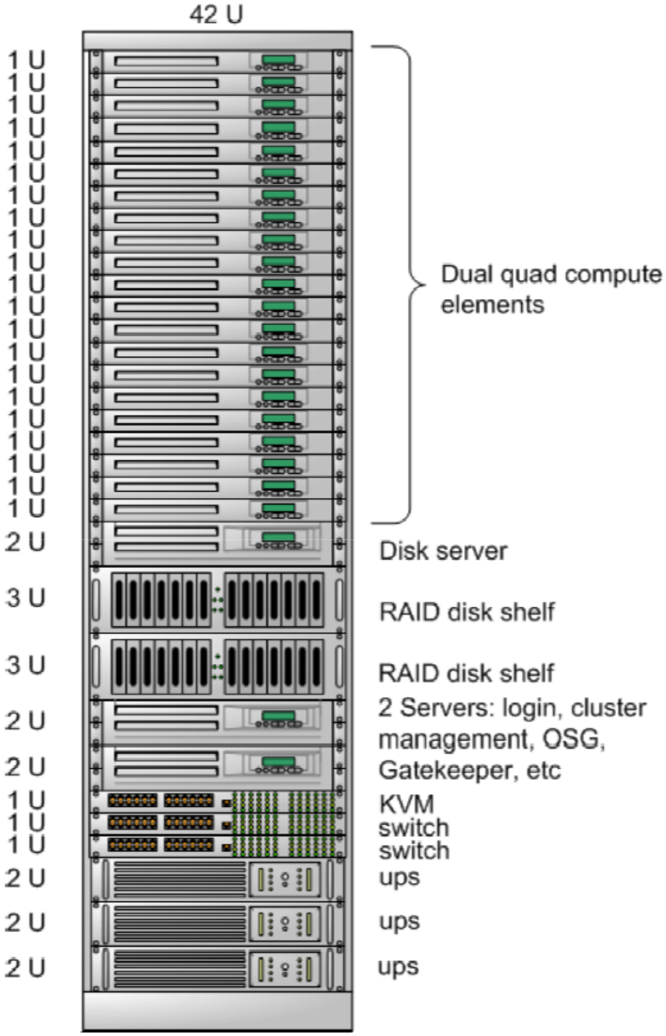


Table 23: Strawman T3gs system designed to fit in one, 42U rack with maximum processing and storage possible. Other systems are certainly possible. At added expense and slightly reduced capability, but with considerable simplification in cabling, etc., a blade-based system would fit in a rack as well. Such a system would provide approximately 320kSI2k processing.

component	typical model	quantity	unit cost, k$
UPS	DELL	3	1.0
switch	DELL PowerConnect 48GbE, portmanaged	2	1.5
servers	DELL PE2950 E5440 processor, 2.83GHz, 32GB RAM, 250GB drive	3	4.2
compute elements	DELL PE1950 E5440 processor, 2.83GHz, 16GB RAM, 250GB drive	21	2.4
storage elements	DELL MD1000	2 (24TB, usable)	5.4
KVM	Belkin	1	1.3
rack			1
total cost			$82.1k

What's missing from $$$ AFAICT

- Display heads/workstations
- Additional cache disks to access D^{3}PD efficiently using PROOF
  - solid state drives a possibility
- $alaries

Summary

Like many universities, we find ourselves catching up to the new T3 estimates

- We need guidance to help optimize costs, system reliability, etc.

All possible options for sites are currently on the table

Extra

SI2K values for various CPUs

Table 22: Estimates of SI2k values collected from various sources for popular processors. From [15].

processor	nickname	Padova	HEP	HEPIX	OSG	BNL
Intel X5355	clovertown	2755	1322	1413	2178	
Intel E5345	clovertown	1190	1267	1889		
Intel E5335	clovertown	2123			1678	
Intel 5160	woodcrest	3161	1505	1602	2420	
Intel 5440	harpertown					2264
Opteron 270		1282	941	1056	1452	1270
Opteron 2214		1352	965	1097	1518	
Opteron 2216						1625
Opteron 2218		1648	1193	1347	1827	1625
Opteron 285		1692	1225	1383	1787	
Opteron 280		1549	1121	1266	1683	
Xeon 3.2 Hz		1516	855			1290
Xeon 3.06 Hz		1427	1166	1402	1169	945
Xeon 2.8 GHz					1123	
Xeon 2.4 GHz			1055	1264	911	747
PIII 1.25 GHz		611	299	319	501	
Opteron 275		1389	1005	1135	1521	1341

An immediate WT2/3 use case

- Simulate different GMSB scenarios in WT2 batch farm using ATLAS Fast Simulation (1 minute/ev):
  - 2 GMSB scenarios (pointing & non-pointing γ)
  - 10 neutralino mass points/scenario
  - 10K events/point ➔ 200K events