Caltech HPC Cluster:
Initial planning thoughts to facilitate discussion

Kaushik Bhattacharya

24 February 2017

Caltech

Background

- Need an institute-wide high performance computing
  - Essential tool across disciplines and divisions
  - Individual closet-sized clusters are inefficient and unsustainable
  - Enables critical research that is not otherwise possible
- Cloud is complementary, not yet a substitute
- National facilities are a complement, not a substitute
- Caltech has spent about $1.6M/year over the last 10 years on computing clusters through startups, gifts and grants

Proposed model

- Staggered model
  - Use $3-4M initial investment to seed and institute wide heterogeneous cluster
  - Add nodes in years 2 and 3 if sufficient new funds (startups, gifts, grants) become available; pool funds to subsequent year if funds are insufficient
  - Freeze the cluster after year three, run for 5-7 years
  - Start new cluster/evaluate transition to the cloud every 4-6 years
- Housed in Powell Booth/South Mudd
- Management
  - Managed by the Vice Provost
  - Run by IMSS
  - Faculty advisory committee: Dan Meiron (chair), Mitch Guttman, Jonathan Katz, Tom Miller, Mark Simons, Maria Spiropulu
- Faculty can buy nodes (fair share) through startups, grants, gifts
  - Fair share usage
  - Lower user fees

Faculty survey: December 2016

- 147 unique responses
- Snapshot (CPU use)

Typical Core	BBE	CCE	EAS	GPS	HSS	PMA	U	Total	Heavy GPU
500+	1	2	6	2		8	1	20	1
100-500	5	1	8	4		4		22	
50-100	5	3	5	4	1	4		22	5
15-50	1	3	5	2	2	2	2	17	
1-14	6	2	7	4	11	11	3	44	

- Suggests we need 12,000 – 15,000 CPUs (consistent with installed capacity)
- GPU use is not that high, but growing
- Some users need small core count, but lots of memory
- Not all heavy users need high speed interconnect

$4M strawman based on Dell preliminary quote

- Strawman
  - 8000 cores, 1 petaflop
  - 250 CPU nodes,
  - 40 GPU nodes,
  - 2:1 infiniband connections,
  - 500 CPUs with extra memory
- Caveat: Skylake, NVLINK pricing is not finalized
- Likely to meet immediate campus demand
  - perhaps needs growth of GPU as FRAM (GPS cluster) shuts down
  - Assume that some of the new clusters continue to operate for 3-5 years
  - Some users are better served by the cloud
- HPC committee members have the spread-sheet and have been asked to configure the machine as they see fit

User fees

- Very early strawman

Usage (K hours)	Owner		Non owner	
	¢/core hour	Annual ($)	¢/core hour	Annual ($)
CPU rates				
0-50	0	0	0	0
50-500	0.75	3375	1.5	6750
500-3000	0.375	12750	1.125	34875
3000-10000	0.1875	25875	0.9375	100500
GPU rates				
0-10	0	0	0	0
10-100	7.5	6750	17.5	15750
100-500	3.75	21750	13.75	70750
500-3000	1.875	68625	11.875	367625

- Comparison

	Pod	AWS	Spot
CPU-hour	7-9	9	2
GPU-hour	197	90	20

- Estimated fees based on strawman cluster: $1.13M

Caltech

Timeline and communication

- March 2017
  - Confirmation of the availability of core funding $2M
  - Exploration of additional funds – faculty, divisions, startups …
  - Discussion with vendors
- April – May 2017
  - Negotiation with vendors
  - Order
- September 2017
  - Delivery

- Will post regular updates on imss.caltech.edu/hpc
- Please contact Bhattacharya, Meiron, Guttman, Katz, Miller, Simons, Spiropulu if you have questions/comments/suggestions

Caltech