2015 CLUSTER PURCHASE: COMMUNITY CLUSTERS TO CLUSTER COMMUNITIES

PURDUE UNIVERSITY

Preston Smith

Director of Research Computing Services

Michael Shuey

Infrastructure Architect

2/27/2015

HPC FACULTY MEETING

COMPUTATION

RESEARCH COMPUTING STRENGTH

Since Steele in 2008, Research Computing has deployed many world-class offerings in computation

PURDUE UNIVERSITY

SIX COMMUNITY CLUSTERS

STEELE

7,216 cores

Installed May 2008

Retired Nov. 2013

COATES

8,032 cores

Installed July 2009

24 departments

61 faculty

Retired Sep. 2014

ROSSMANN

11,088 cores

Installed Sept. 2010

17 departments

37 faculty

HANSEN

9,120 cores

Installed Sept. 2011

13 departments

26 faculty

CARTER

10,368 cores

Installed April 2012

26 departments

60 faculty

#175 on June 2013 Top 500

CONTE

9,280 Xeon cores
(69,600 Xeon Phi cores)

Installed August 2013

20 departments

51 faculty (as of Aug. 2014)

#39 on June 2014 Top 500

TOP TEN CAMPUS SUPERCOMPUTERS IN THE NATION

June 2013 Top 500

U.S. CAMPUS RANKING	UNIVERSITY	NAME	WORLD RANKING
1	PURDUE	CONTE	28
2	INDIANA UNIVERSITY	BIG RED II	46
3	USC	HPCC	53
4	RENSSELAER POLYTECHNIC INSTITUTE	BLUE GENE/Q	76
5	CLEMSON UNIVERSITY	PALMETTO 2	115
6	UNIVERSITY OF ROCHESTER	BLUESTREAK	170
7	PURDUE	CARTER	175
8	UNIVERSITY OF COLORADO	JANUS	239
9	USC	HPC	242
10	UNIVERSITY OF CHICAGO	MIDWAY	301

PURDUE
UNIVERSITY

BID PROCESS

Open bid process:

- Quantity approx. 700 nodes
- Included various interconnects (10 Gbps Ethernet, FDR/EDR Infiniband)
- 20- and 24-core compute nodes
- Memory sizes from 64 GB to 512 GB
- Conventional disk vs. SSD

Prices ranged from $4400/node to $5400/node for base node (64GB)
Vendors included Dell, HP, IBM, etc.

BID RESULTS

HARDWARE SPECS

Base node: HP DL60

- 20-core node, 2.6 GHz Intel “Haswell” processors (E5-2660v3)
    - Larger L3 caches, more cores
    - Double the FLOPS over Carter/Conte processors (FMAC – helps matrix math)
    - Good balance between cost-effectiveness and overall node price
- 64 GB DDR4 memory (minimum)
    - Several memory sizes possible
    - Fastest memory available
- 500 GB local disk
- 10 Gbps Ethernet on all nodes for IP and NFS
- FDR Infiniband interconnect
    - 56 Gbps, 2:1 fat tree – most cost-effective than faster options for most users

Scratch system:

- Approximately 1.4 PB
- 20 GB/sec, IB connectivity to high-performance cluster nodes

IMPLEMENTATION

General implementation schedule:

- Facilities preparation underway
- Board of Trustees meeting April 9th – pending board approval
- Install day May – may need assistance unboxing & racking
- General availability soon after

PURDUE UNIVERSITY

THE BOTTOM LINE

25% more cores at 10% more dollars than Conte,

- **Better floating point**
- **Faster memory**

64G Node	256G Node	512G Node	64G Node + GPU
$4,400.00	$6,500.00	$10,000.00	$11,000.00

IN THE PAST

ONE COMMUNITY

PURDUE
UNIVERSITY

TODAY

NEIGHBORHOODS

PURDUE
UNIVERSITY

TODAY

NEIGHBORHOODS

PURDUE
UNIVERSITY

TODAY

NEIGHBORHOODS

PURDUE UNIVERSITY

TODAY

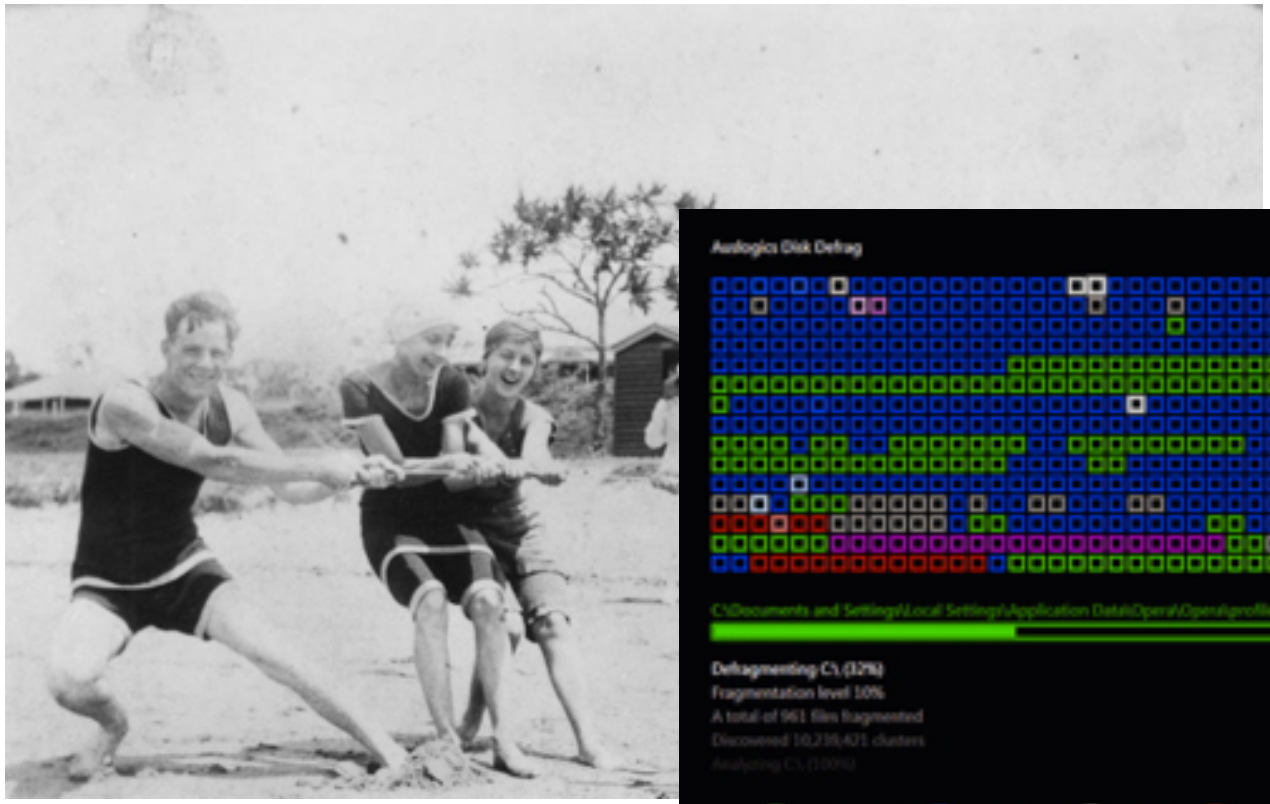
NEIGHBORHOODS

PURDUE UNIVERSITY

ONE SYSTEM

MANY DIFFERENT (CONFLICTING) NEEDS

- Parameter sweeps vs CFD
- Monte Carlo simulation vs climate modeling
- R vs molecular dynamics



PURDUE UNIVERSITY

Community Clusters to Cluster Communities

What neighborhoods are in our community?

By domain? Chemistry, mechanical engineering, statistics, structural biology

By Hardware Configuration? Accelerator cluster, big memory cluster, data-intensive cluster

By Profile of Work: single core, multi-core, multi-node, inter-node communication needs, I/O characteristics, memory requirements

COMPUTATION

NEW MODEL – ORGANIZED BY COMMON PROFILES

Community Clusters to Cluster Communities

What neighborhoods are in our community?

HPC (Rice): Multiple cores or nodes, probably MPI. Benefit from high-performance network and parallel filesystem. The vast majority of campus - 80% of all work!

HTC (Hammer): Primarily single core. CPU-bound. No need for high-performance network.

Life Science (Big memory): Use entire node to get large amounts of memory. Less need for high-performance network. Needs large, fast storage.

CHALLENGES

CHALLENGES TO SMALLER COMMUNITIES

- HPC and HTC communities prefer different points to optimize the scheduler.
- Small but key communities (like large memory) lose benefits of standby queues when fewer nodes are spread between several clusters.
- HTC or large memory communities often have little need for HPC-specific optimizations
    - Accelerators
    - High-speed, low-latency networks

Emerging communities often don't fit in existing model at all!

Big Data Analytics
Graphics Rendering
Nontraditional platforms (Windows, cloud)

FOR THE MAJORITY:

NO CHANGE FROM TODAY!

Rice: A traditional HPC system just like Carter or Conte

The same, familiar model:

- New cluster acquisition every 12-18 months
- Each a distinct, non-heterogeneous system.

Nothing different for you!

PURDUE UNIVERSITY

UNDERSERVED COMMUNITIES

A BETTER EXPERIENCE

Hammer – HTC
Big Memory

HTC or big memory clusters expanded annually with each purchase.

Better Community Cluster Experience

PURDUE UNIVERSITY

Pay for resources on-demand!

- Burst capacity for a deadline coming up?
- Access to a different type of node for a short period?

Pay for on-demand node instances in one-month increments.

PURDUE
UNIVERSITY

RESEARCH DATA STORAGE

UPDATES SINCE THE FALL MEETING

RESEARCH DATA DEPOT

SERVICE DETAILS

At $150/TB per year:

- Storage oriented around your research lab, with
  - Snapshots
  - Multi-site copies of your data
  - Disaster protection
  - A scalable, expandable storage resource optimized for HPC
- Access to Globus data transfer service, and endpoint sharing

PURDUE UNIVERSITY

SCRATCH IMPROVEMENTS

CARTER

New scratch filesystem on order for Carter!

1.5 PB filesystem
Expect Conte-like 100TB, multi-million file user quotas

Existing Carter scratch will remain only on Hansen

PURDUE UNIVERSITY

Well received!

- Since a fall go-live, over 105 research groups are participating.
  - ***Many are not HPC users!***
- Over .7 PB provisioned to date
- A research group purchasing space has purchased, on average, 8.6TB.

GET ACCESS

GIVE IT A TRY!

To buy 1 or more TB of space,
Or to set up a trial for your lab

Order online:

https://www.rcac.purdue.edu/purchase/depot/

PURDUE UNIVERSITY

GLOBUS

STATISTICS

Data moved in 2014:

13 TB in, 19TB out

200k files both directions

55 unique users

Coming soon:

Globus interface to Fortress

https://transfer.rcac.purdue.edu

PURDUE UNIVERSITY

HADOOP

- “hathi” Hadoop cluster for prototyping big data applications
- Spark, Hbase, Hive, Pig

36 research groups,
411 users on the system already

https://www.rcac.purdue.edu/compute/hathi/

PURDUE
UNIVERSITY

RESEARCH SERVICES

OTHER SERVICES YOU MIGHT BE INTERESTED IN

VERSION CONTROL

NEED GIT OR SVN?

Repositories for your lab, managed by your queue management tool.

PURDUE UNIVERSITY

NEED HELP?

COFFEE BREAK CONSULTATIONS

- Hard to solve problems with HPC?
- Need help building your software or optimizing your workflow?
- Need to learn what resources are available?

PURDUE UNIVERSITY

SCHOLAR

HPC FOR INSTRUCTION

- Need to teach students to use HPC in a course?
- Scholar cluster is available to any instructor at no cost.

Spring 2015:	EAPS
CS	AGRY
STAT	ANSC
CHEM	ChemE

Just send a CRN

PURDUE UNIVERSITY

THE END

QUESTIONS?

Questions?

PURDUE UNIVERSITY

SPARE SLIDES

EVERYTHING PAST HERE IS SPARE

PURDUE
UNIVERSITY