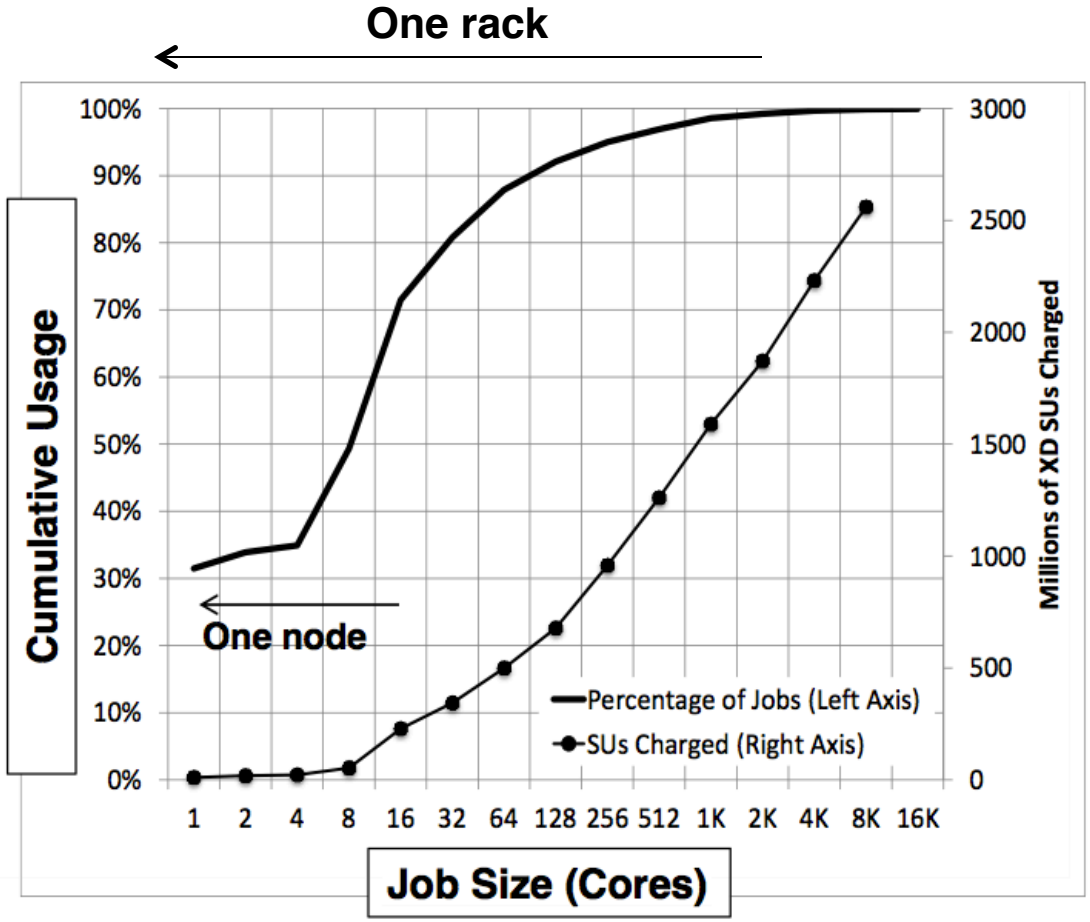

Michael L Norman
PI Comet Project
Director, SDSC
COMET
IS HERE
SDSC

HPC for the 99%

- 99% of jobs run on NSF's HPC resources in 2012 used <2,048 cores
- And consumed >50% of the total core-hours across NSF resources



HPC for the 99%

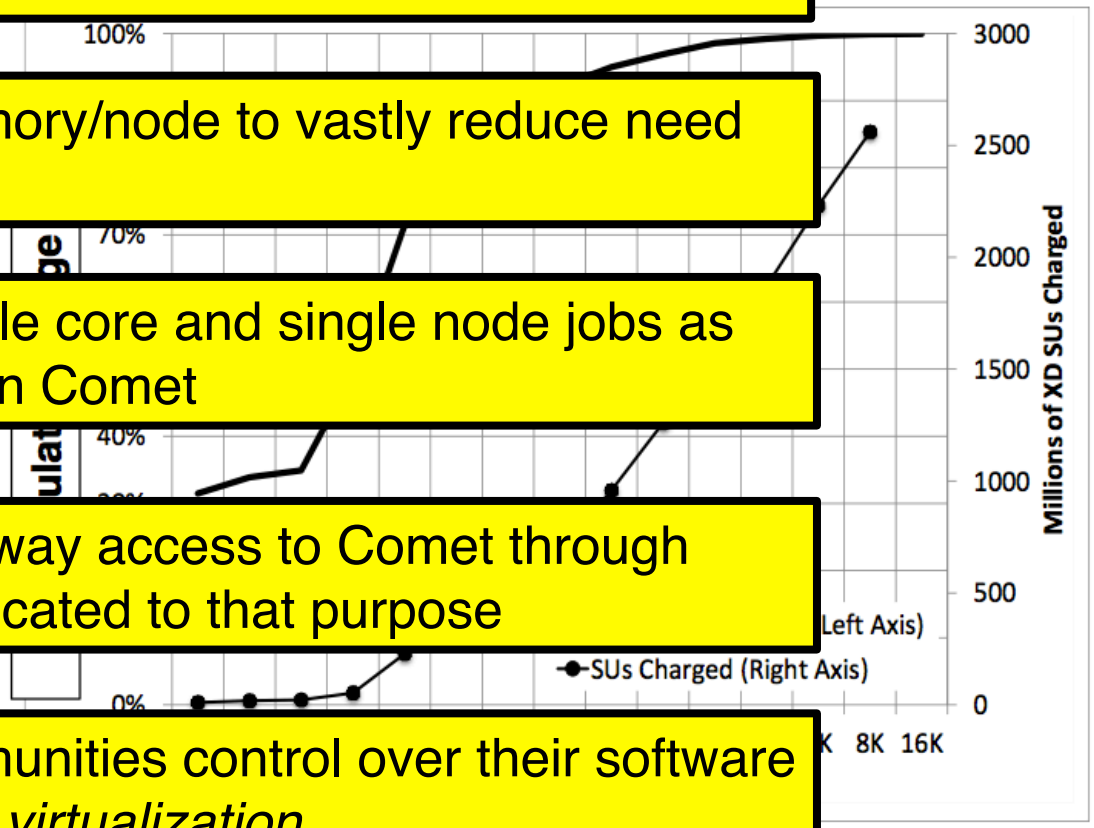
Idea 1: *architect* an HPC system where 99% of the jobs run inside a single rack with full bisection BW

- **99% of jobs run on NSF resources in 2012 use**
- **And >50% of the total core NSF resources**

Idea 2: *increase* memory/node to vastly reduce need for multiple rack runs

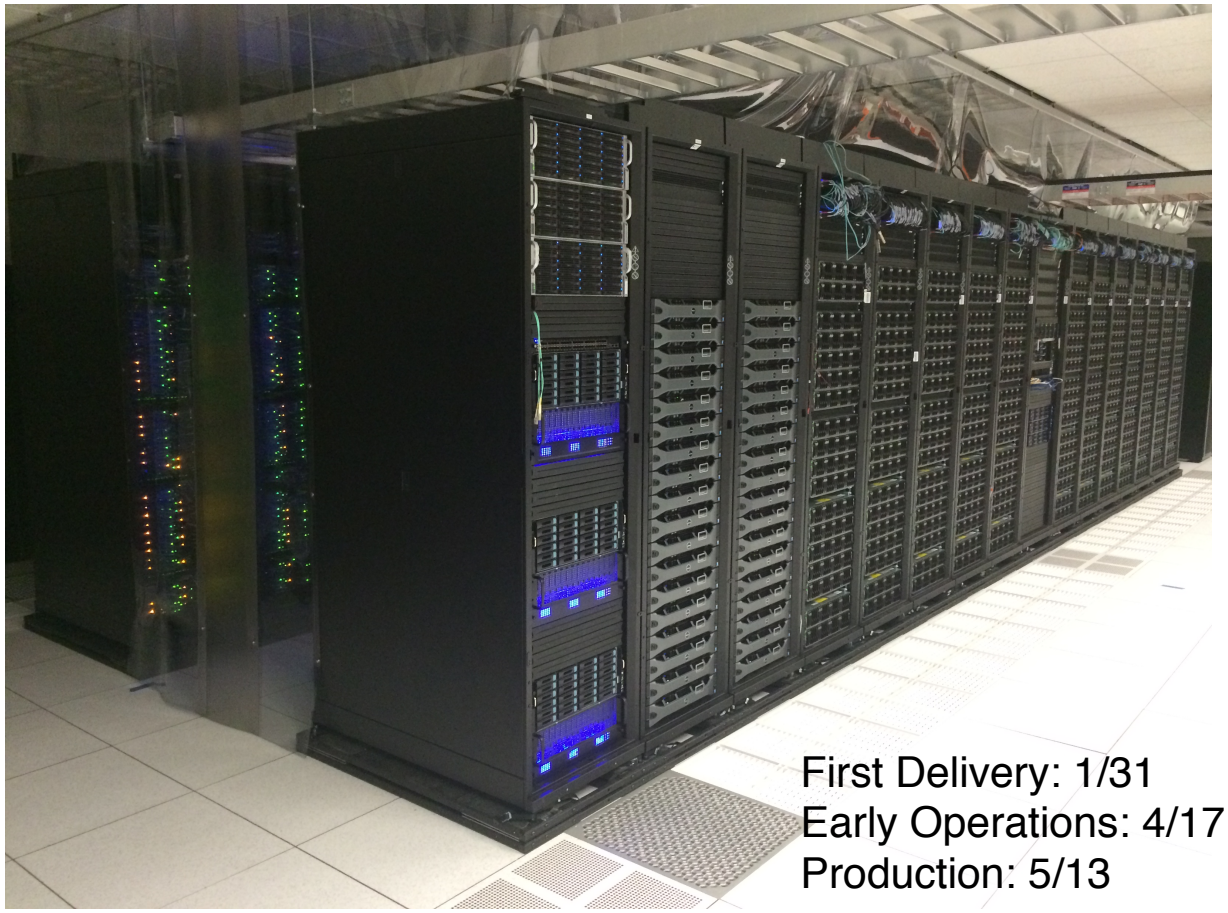
Idea 3: *embrace* single core and single node jobs as "first class citizens" on Comet

Idea 4: *support* Gateway access to Comet through ancillary servers dedicated to that purpose



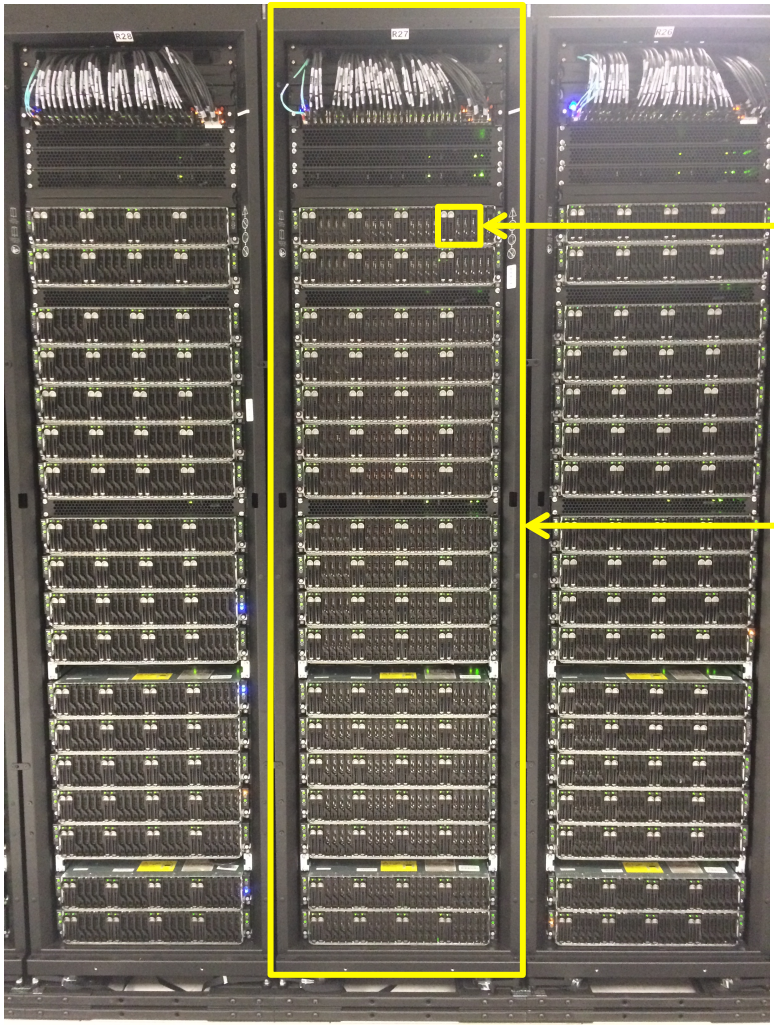
Idea 5: *provide* communities control over their software environment through *virtualization*

Comet
"HPC for the long tail of science"



First Delivery: 1/31
Early Operations: 4/17
Production: 5/13

~70 TF supercomputer in a rack



1 CN = 24 core Haswell @ 2.5 GHz
+ 128 GB DDR4 DRAM
+ 320 GB SSD

1 rack = 72 nodes
= 1728 cores
= 9.2 TB DRAM
= 23 TB SSD
= FDR InfiniBand
FatTree (non-blocking)

IT'S ABOUT THE SOFTWARE STUPID!

HPC

BIG DATA

Gateways

MONDAY, August 10	
8:00 – 8:30	Registration, Coffee
8:30 – 8:45	**Welcome** Mike Norman, SDSC Director
8:45 – 9:30	**Introduction, Orientation** Bob Sinkovits, Director for Scientific Computing Applications, SDSC
9:30 – 10:15	**How do I launch and manage jobs on the system?** Mahidhar Tatineni, User Services Manager, SDSC
10:15 – 10:45	Break
10:45 – 12:15	**Launching and Managing Jobs** Mahidhar Tatineni, User Services Manager, SDSC
12:15 – 1:30	Lunch at Café Ventanas
1:30 – 3:00	**How do I manage my data on the file system?** Amit Majumdar, Division Director, Data Enabled Scientific Computing
3:00 – 3:30	Break
3:30 – 5:00	**How do I know I'm making effective use of the machine?** Bob Sinkovits, Interim Director for Scientific Computing Applications, SDSC
5:30 – 8:30	**Reception at Wayne Pfeiffer's home overlooking the Pacific,** *Sweater or jacket recommended* Shuttle provided from SDSC driveway

TUESDAY, August 11	
8:00 – 8:30	Coffee
8:30 – 10:00	**How do I automate my job pipeline to ensure reproducibility?** Ilkay Altintas, SDSC's Chief Data Science Officer, Director, Workflows for Data Science (WorDS) Center of Excellence SDSC
10:00 – 10:15	Break
10:15 – 12:15	**How do I manage my software?** Andrea Zonca, HPC Applications Support Specialist, SDSC
12:15 – 1:30	Lunch at Café Ventanas
1:30 – 3:30	**How do I mine and get insight from my data?** Natasha Balac, Director, Predictive Analytics Center of Excellence (PACE), SDSC Amit Chourasia, Senior Visualization Scientist, SDSC
3:30 – 3:45	Break
3:45 – 4:15	SDSC Data Center Tour
4:15 – 5:00	Hands-on practice continues with mentors available for questions

WEDNESDAY, August 12 PARALLEL SESSIONS		
8:00 – 8:30	Coffee	
	Track 1 *Auditorium*	**Track 2** *Synthesis Center E-B143*
Session 1 8:30 – 12:00	**GPU Computing and Programming** **Andreas Goetz, Co-Director, CUDA Teaching Center, Co-Principal Investigator, Intel Parallel Computing Center** This session provides an introduction to massively parallel computing with graphics processing units (GPUs). The use of GPUs is becoming increasingly popular across all scientific domains since GPUs can significantly accelerate time to solution for many problems. Participants will be introduced to essential background of the GPU chip architecture and will learn how to program GPUs via the use of Libraries, OpenACC compiler directives, and CUDA programming. The session will incorporate hands-on exercises for participants to acquire the skills to use and develop GPU aware applications.	**Predictive Analytics** Natasha Balac, Director, Predictive Analytics Center of Excellence (PACE), SDSC This session is designed as an introduction for attendees seeking to extract meaningful predictive information from within massive volumes of data. The session will provide an introduction to the field of predictive analytics and a variety of data analysis tools to discover patterns and relationships in data that can contribute to building valid predictions.
12:00 – 1:30	Lunch at Café Ventanas	
Session 2 1:30 – 5:00	**Performance Optimization** Bob Sinkovits, Director for Scientific Computing Applications, SDSC This session is targeted at attendees who both do their own code development and need their calculations to finish as quickly as possible. We'll	**Spark for Scientific Computing** Andrea Zonca, HPC Applications Specialist Mahidhar Tatineni, User Services Manager, SDSC Apache Spark is a cluster computing framework extensively used in Industry to process large

PARALLEL SESSIONS			
8:00 – 8:30	**Coffee**		
	Track 1	**Track 2**	
	12:00 – 1:30	Lunch at Café Ventanas	
Session 3 8:30 – 12:00	**Session 4** 1:30 – 5:00	**Parallel Computing using MPI & Open MP** Mahidhar Tatineni, User Services Manager, SDSC This session is targeted at attendees who are looking for a hands-on introduction to parallel computing using MPI and Open MP programming. The session will start with an introduction and basic information for getting started with MPI. An overview of the common MPI routines that are useful for beginner MPI programmers, including MPI environment set up, point-to-point communications, and collective communications routines will be provided. Simple examples illustrating distributed memory computing, with the use of common MPI routines, will be covered. The OpenMP section will provide an overview of constructs and directives for specifying parallel regions, work sharing, synchronization and data scope. Simple examples will be used to illustrate the use of OpenMP shared-memory programming model, and important run time environment variables Hands on exercises for both MPI and OpenMP will be done in C and FORTRAN.	**Workflow Management** Ilkay Altintas, SDSC's Chief Data Science Officer, Director, Workflows for Data Science (WorDS) Center of Excellence SDSC This session will start with a crash course on workflow management basics. We will then explore common computing platforms including Sun Grid Engine, NSF XSEDE high performance computing resources, the Amazon Cloud and Hadoop with an emphasis on how workflow systems can help with rapid development of distributed and parallel applications on top of any combination of these platforms. We will then discuss how to track data flow and process executions within these workflows (i.e. provenance tracking) including the intermediate results as a way to make workflow results reproducible. We will end with a lab session on using Kepler to build, package and share workflows interacting with various computing systems.
	5:30 – 9:00	**Beach BBQ Dinner at La Jolla Shores Hotel**, *sweater or jacket recommended* 8110 Camino Del Oro, La Jolla, CA 92037 Shuttle provided from SDSC driveway	
	Pure Python prototype code can be gradually optimized by pushing the most computationally intensive functions to the GPU without the need to implement code in CUDA or OpenCL.		

FRIDAY, August 14	
8:00 – 8:30	Coffee
8:30 – 9:30	**Emerging Technologies in HPC** Shawn Strande, Deputy Director, SDSC
9:30 – 11:00	**Lightning Rounds**
11:00 – 11:30	**Wrap up**
11:30AM	**Adjourn** Thank you for attending we hope you enjoyed the week! *(To-go box lunches will be available)*

THANKS TO THE ORGANIZERS AND LECTURES

THANKS TO YOU FOR ATTENDING!