
Data-Intensive Computing at SDSC

Glenn K. Lockwood, Ph.D.

User Services Group

San Diego Supercomputer Center

glock@sdsc.edu

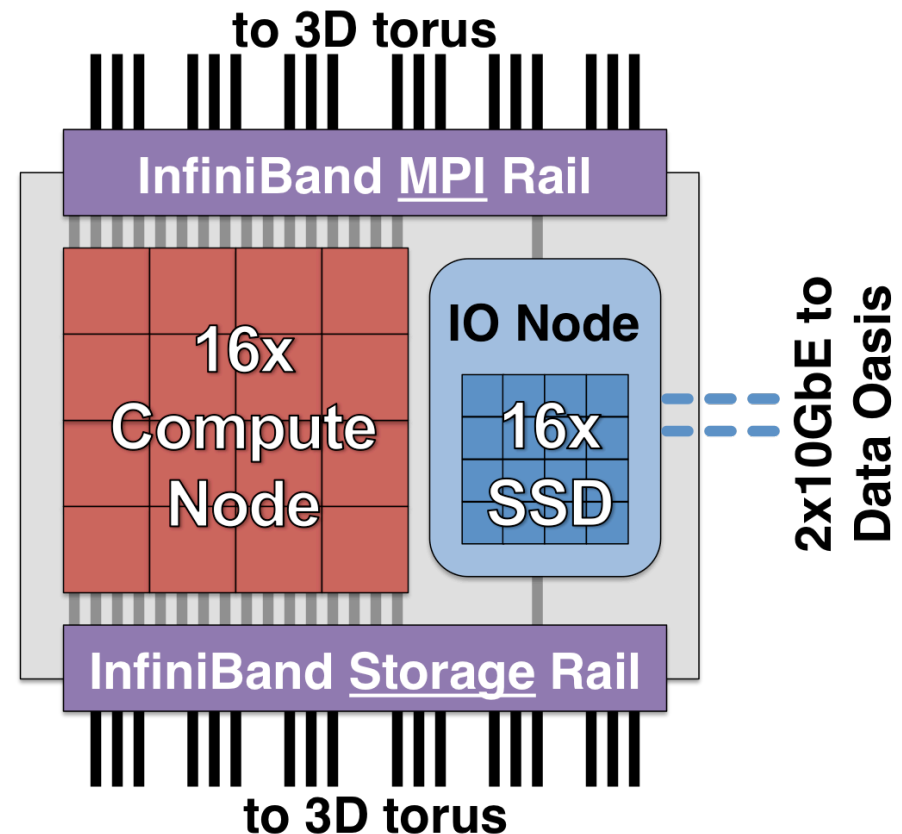
San Diego Supercomputer Center



- **ORU within UC San Diego**
- **XSEDE L1 Service Provider**
 - Gordon (Data-intensive HPC)
 - Data Oasis (4 PB Lustre FS)
 - Comet (2 PF capacity, 2015)
- **Data-intensive computing**
 - LHC/CMS (OSG) - 125 TB
 - Genomics (industry) - 50+ TB
 - Hadoop/YARN/Spark support

Gordon: the Data-Intensive Supercomputer

- **1,024 compute nodes:**
 - 16 cores (Sandy Bridge) and 64 GB DDR3 DRAM
 - 4 GB RAM/core
 - 300 GB SSD
- **Dual-rail 4x4x4 torus of switches**
- **64 I/O nodes**
 - 16x 300 GB SSDs
 - LNet routers
 - IB/10GbE bridging
- **100 GB/sec to Oasis**



SDSC's Data-Intensive Capabilities

Data Oasis Lustre File System

- 4 PB of usable capacity
- 64x Object Storage Servers
- 100 GB/sec to Gordon

Comet (Jan 2015)

- 1.8 to 2.0 PFlops (CPUs)
 - FDR InfiniBand
- Oasis upgrade to 7+6 PB, 40 GbE
- 200 GB/sec from Comet to Oasis



Data-Intensive Computing at SDSC

***ACCELERATING LARGE-
SCALE GENOMIC ANALYSIS***

Case Study: 438 Human Genomes

- **Input data:**

- raw reads from 438 whole human genomes
- 50 TB of compressed data from Janssen R&D



- **Computing requirements:**

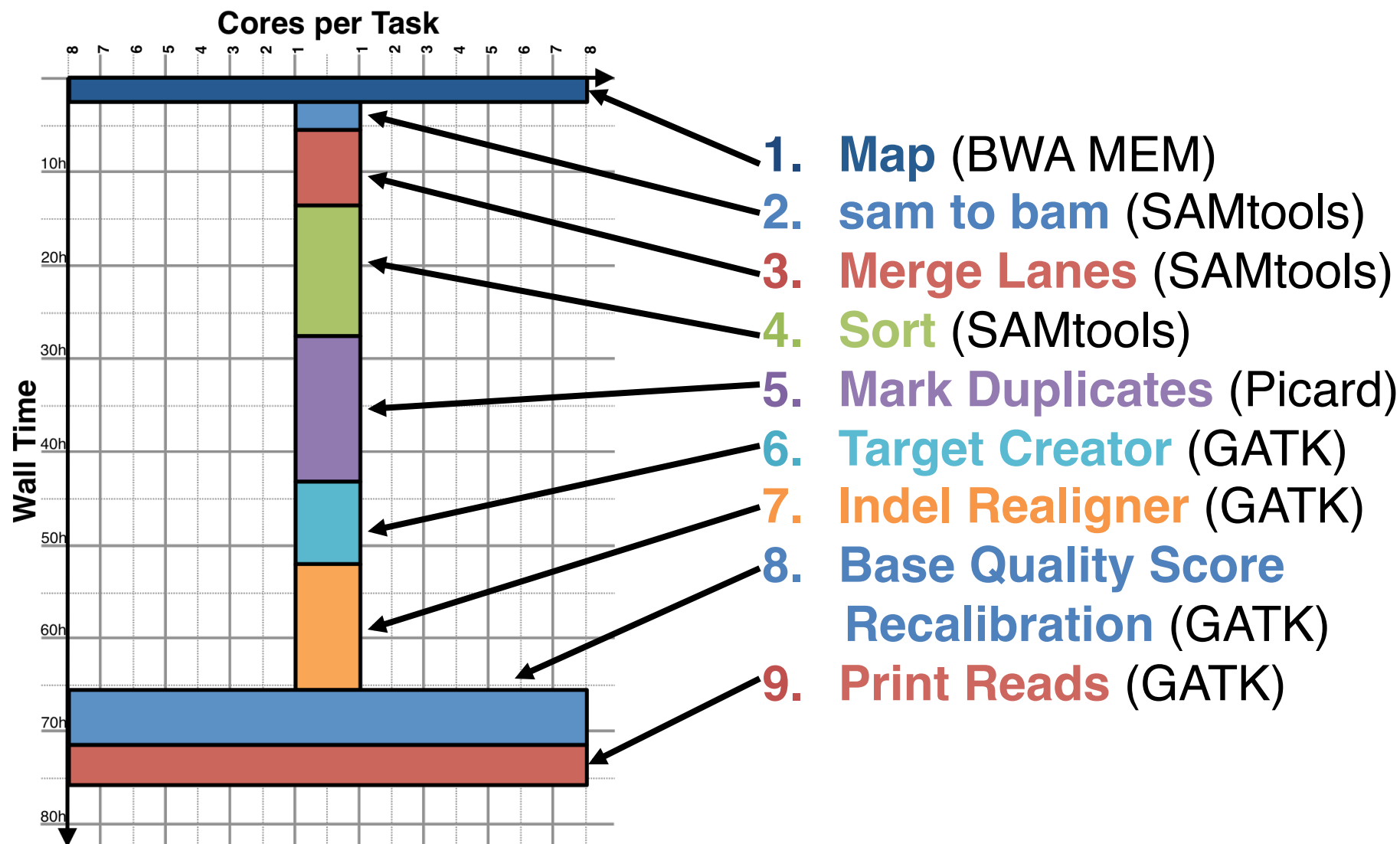
- 9-step read mapping pipeline
- 5-step variant calling pipeline



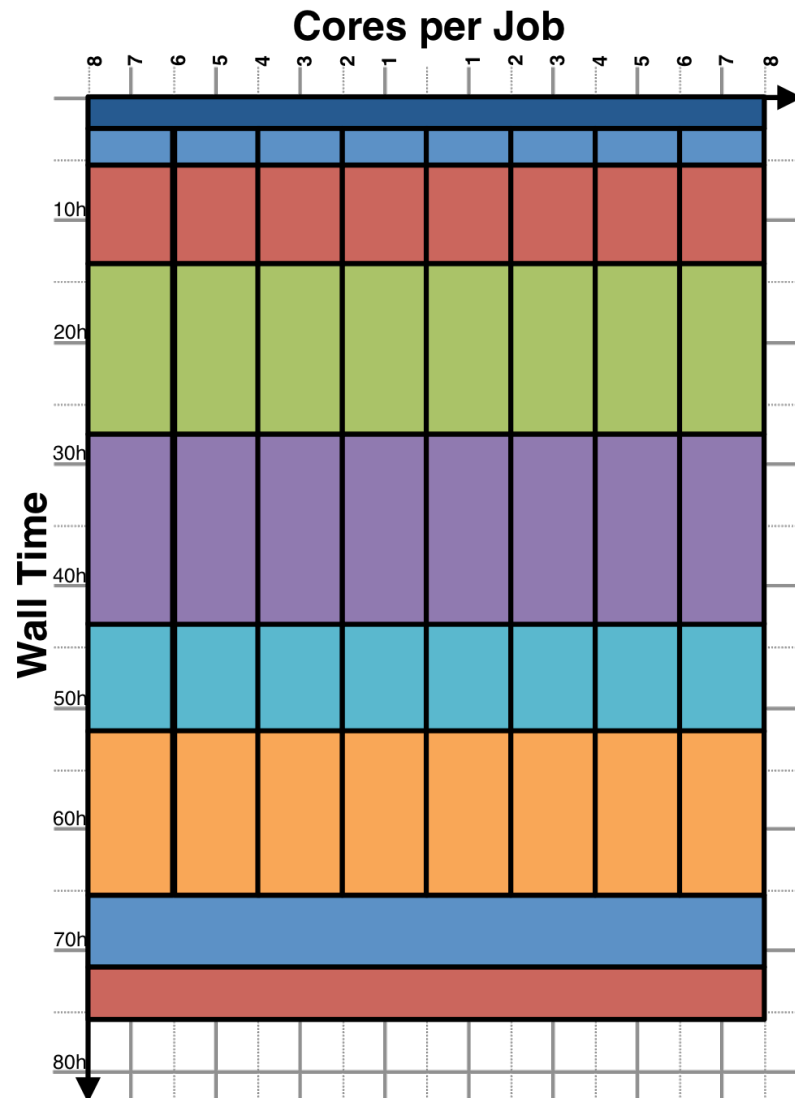
- **Project requirements:**

- FAST turnaround (< 2 months)
- EFFICIENT (batch queue like everyone else)

Read Mapping from the HPC Perspective

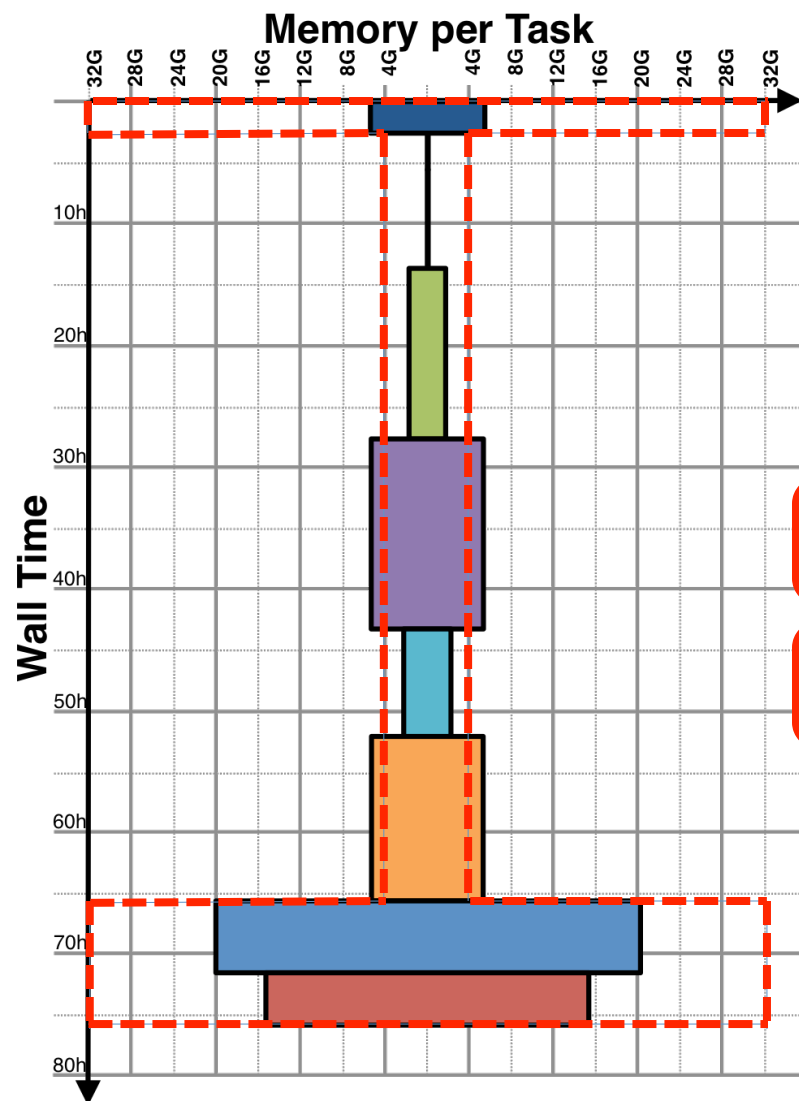


Maximize Efficiency by Keeping Cores Busy, Right?



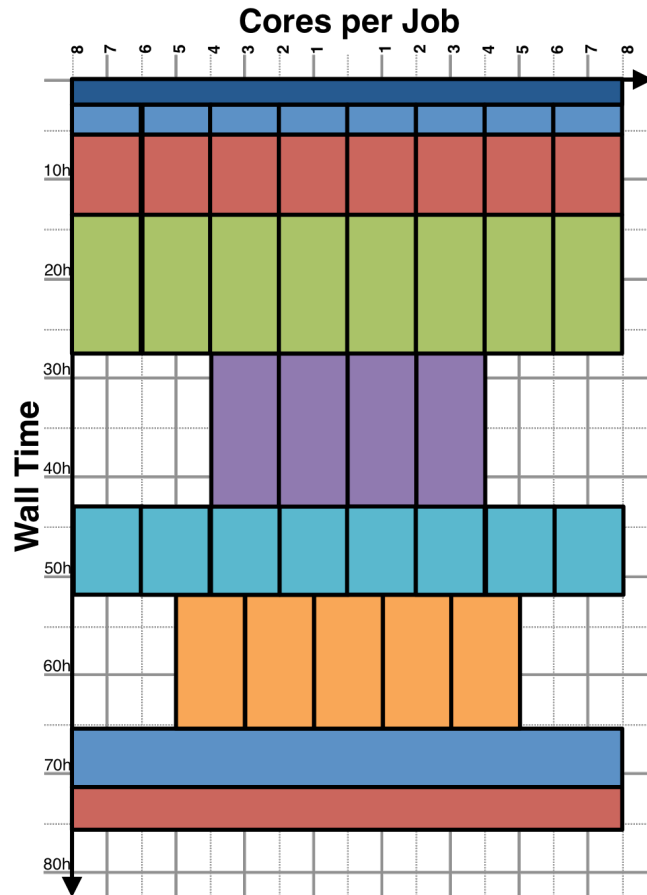
1. Map 8 threads/task
2. sam to bam 1 thread/task
3. Merge Lanes 1 thread/task
4. Sort 1 thread/task
5. Mark Duplicates 1 thread/task
6. Target Creator 1 thread/task
7. Indel Realigner 1 thread/task
8. BQSR 8 thread/task
9. Print Reads 8 thread/task

Read Mapping from the Data Perspective



1. Map 6.5 GB/task
2. sam to bam 0.0056 GB/task
3. Merge Lanes 0.0040 GB/task
4. Sort 1.7 GB/task
5. Mark Duplicates 5.3 GB/task
6. Target Creator 2.2 GB/task
7. Indel Realigner 5.3 GB/task
8. BQSR 20 GB/task
9. Print Reads 15 GB/task

Limiting Resources



1. Map
2. sam to bam
3. Merge Lanes
4. Sort
5. Mark Duplicates
6. Target Creator
7. Indel Realigner
8. BQSR
9. Print Reads

CPU
CPU
CPU

IO

Memory

CPU

Memory

CPU

CPU

When you can't feed
memory fast enough

Out-of-core Calculations: How to Murder Lustre

16 threads generate...

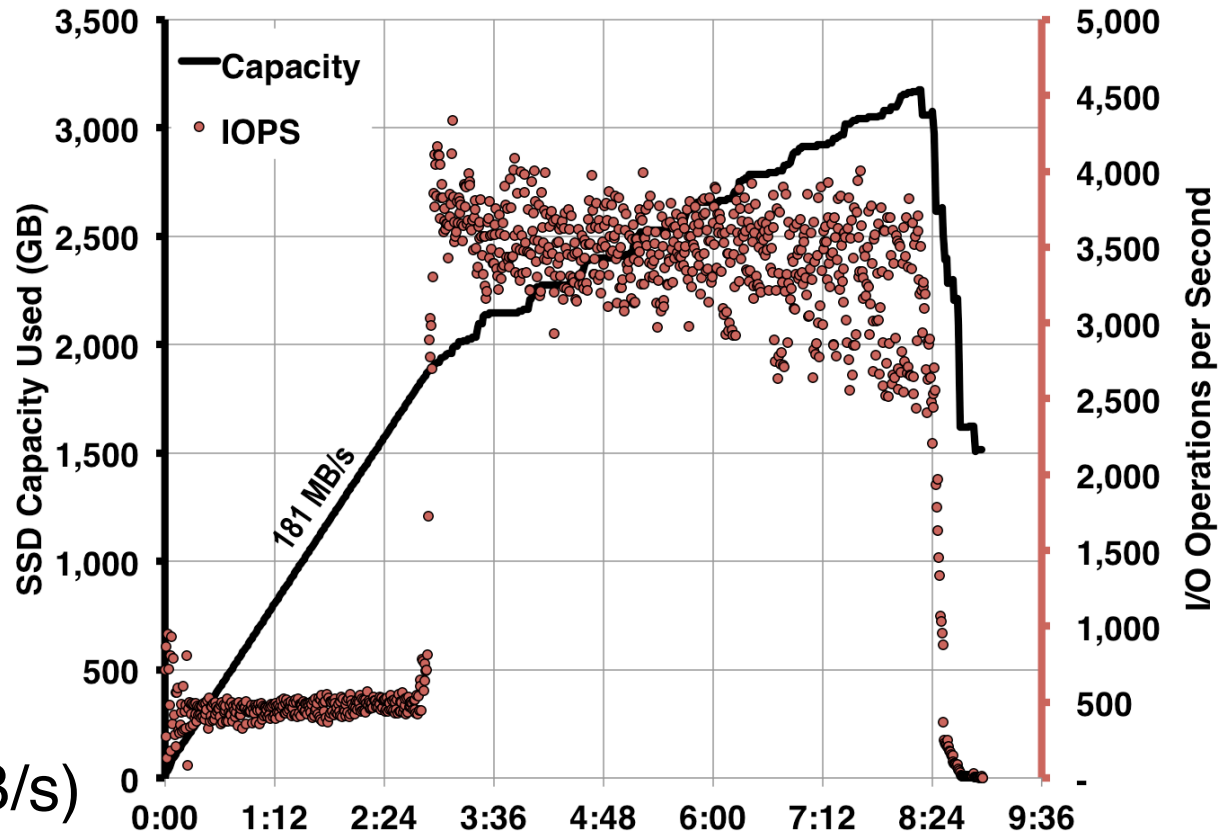
- 10,000 temp files
- 3,500 IOPS
- 1 GB/s read rate

and require

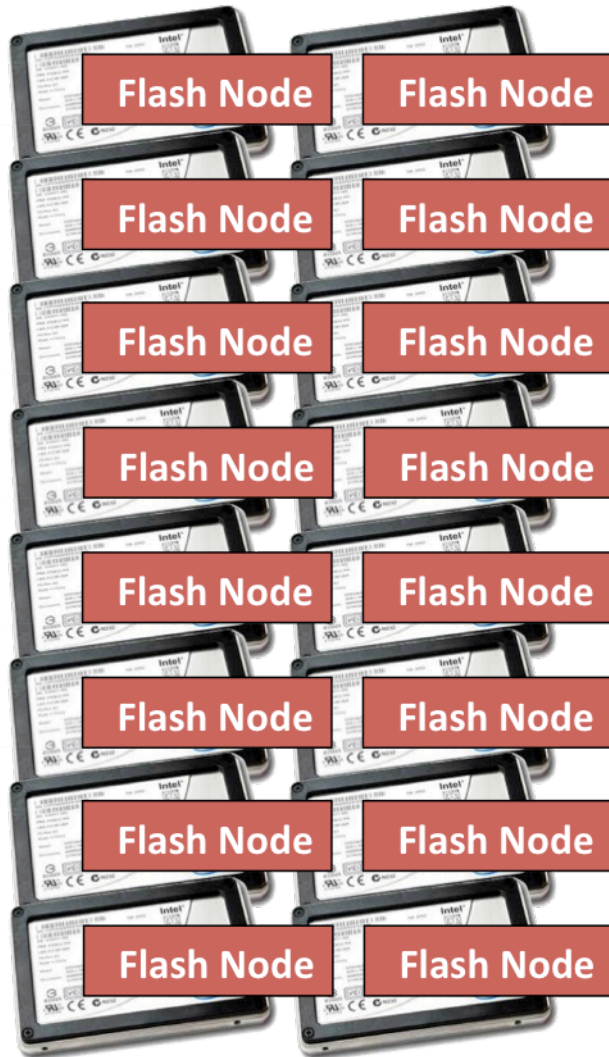
- 25 GB DRAM
- 1.6 TB input data
- 3.5 TB scratch

...which demand

- high capacity (TBs)
- high bandwidth (GB/s)
- high IOPS (kIOPS/node)



Gordon BigFlash



Noflash

Noflash

Noflash

Noflash

Noflash

Noflash

Noflash

Noflash

Noflash

Noflash

Noflash

Noflash

Noflash

Noflash

Noflash



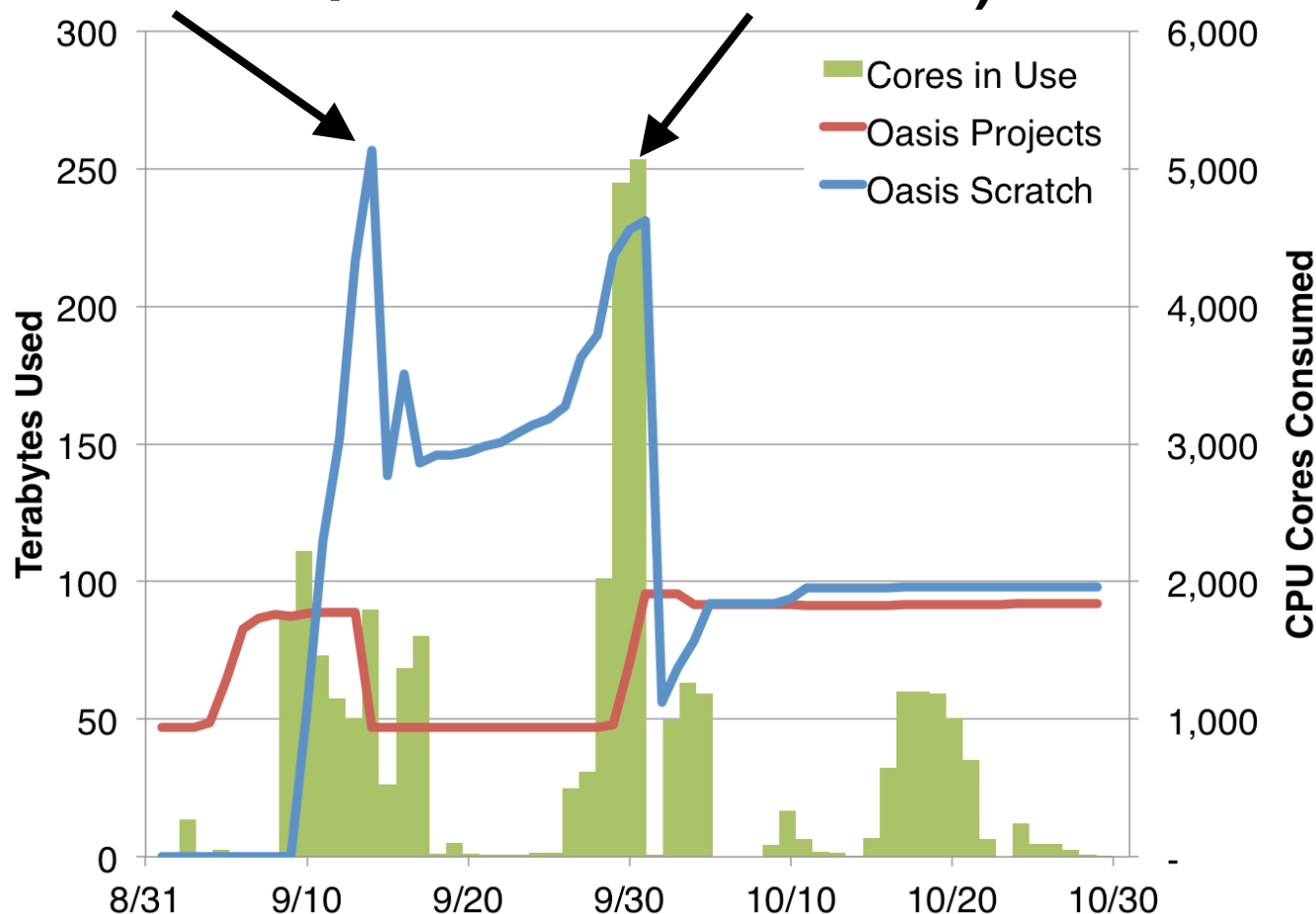
Aggregate 16x SSDs
on a single compute
node to get

- 4.4 TB RAID0 array
- 3.8 GB/s bandwidth
- 320,000 IOPS

Footprint on Gordon: CPUs and Storage Used

**257 TB + 80 TB
Lustre used at peak**

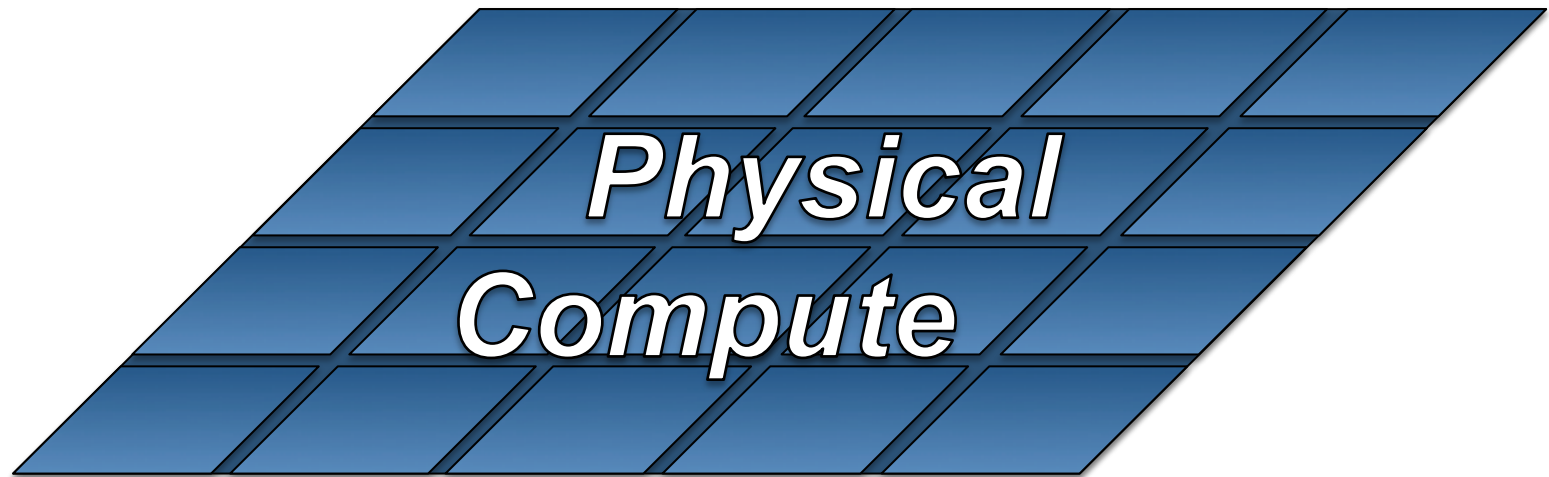
**5,000 cores (30% of
Gordon) in use at once**



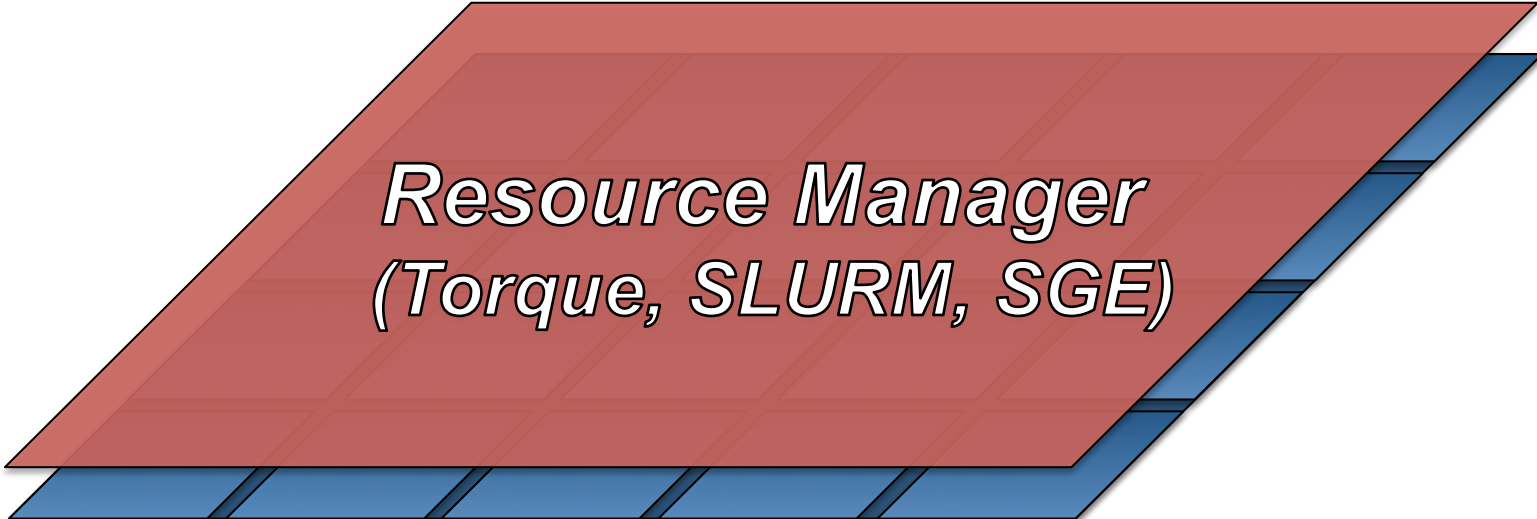
Data-Intensive Computing at SDSC

HADOOP IN TRADITIONAL HPC ENVIRONMENTS

myHadoop: Run Hadoop on Traditional HPC Infrastructure



myHadoop: Run Hadoop on Traditional HPC Infrastructure

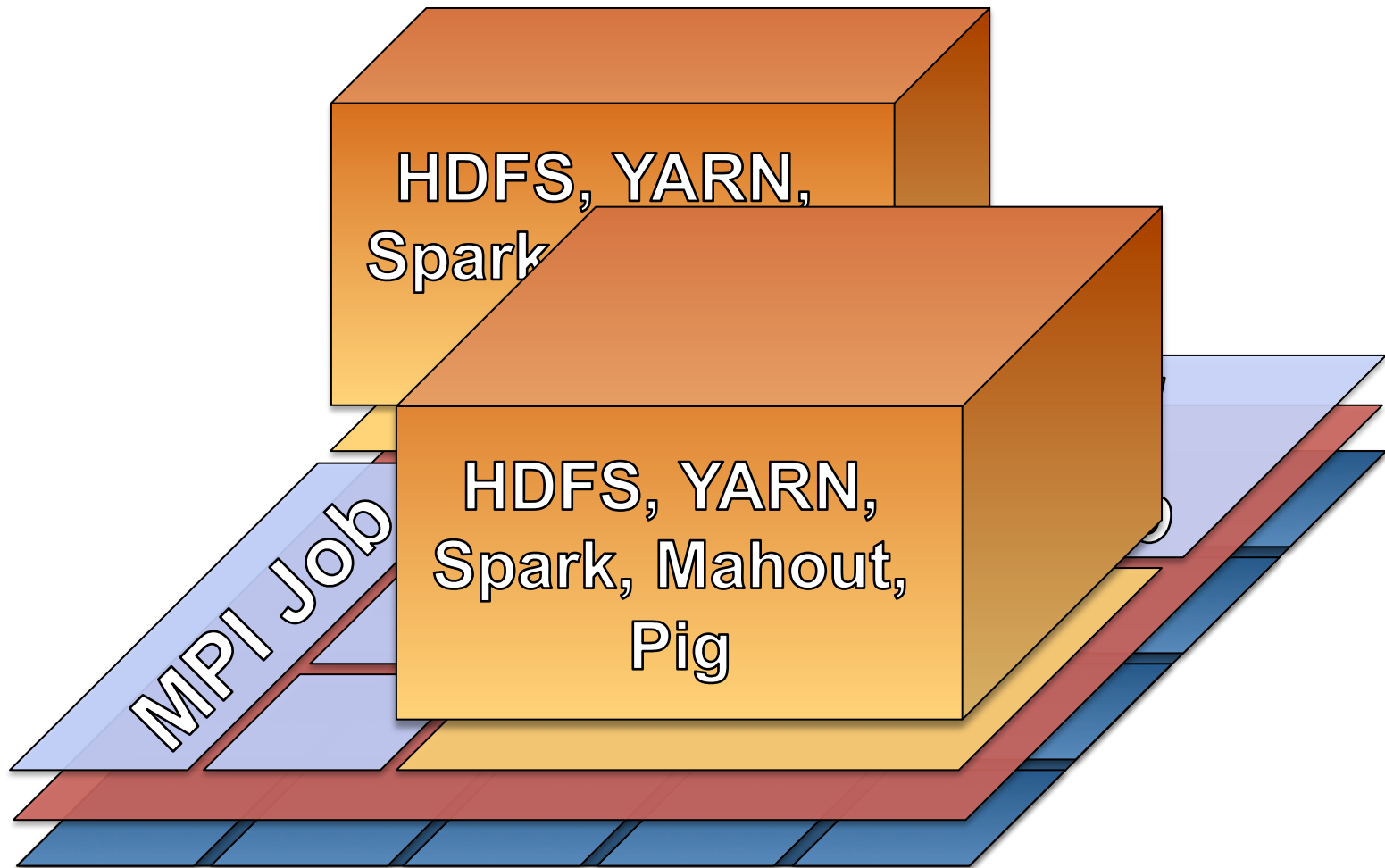


*Resource Manager
(Torque, SLURM, SGE)*

myHadoop: Run Hadoop on Traditional HPC Infrastructure



myHadoop: Run Hadoop on Traditional HPC Infrastructure



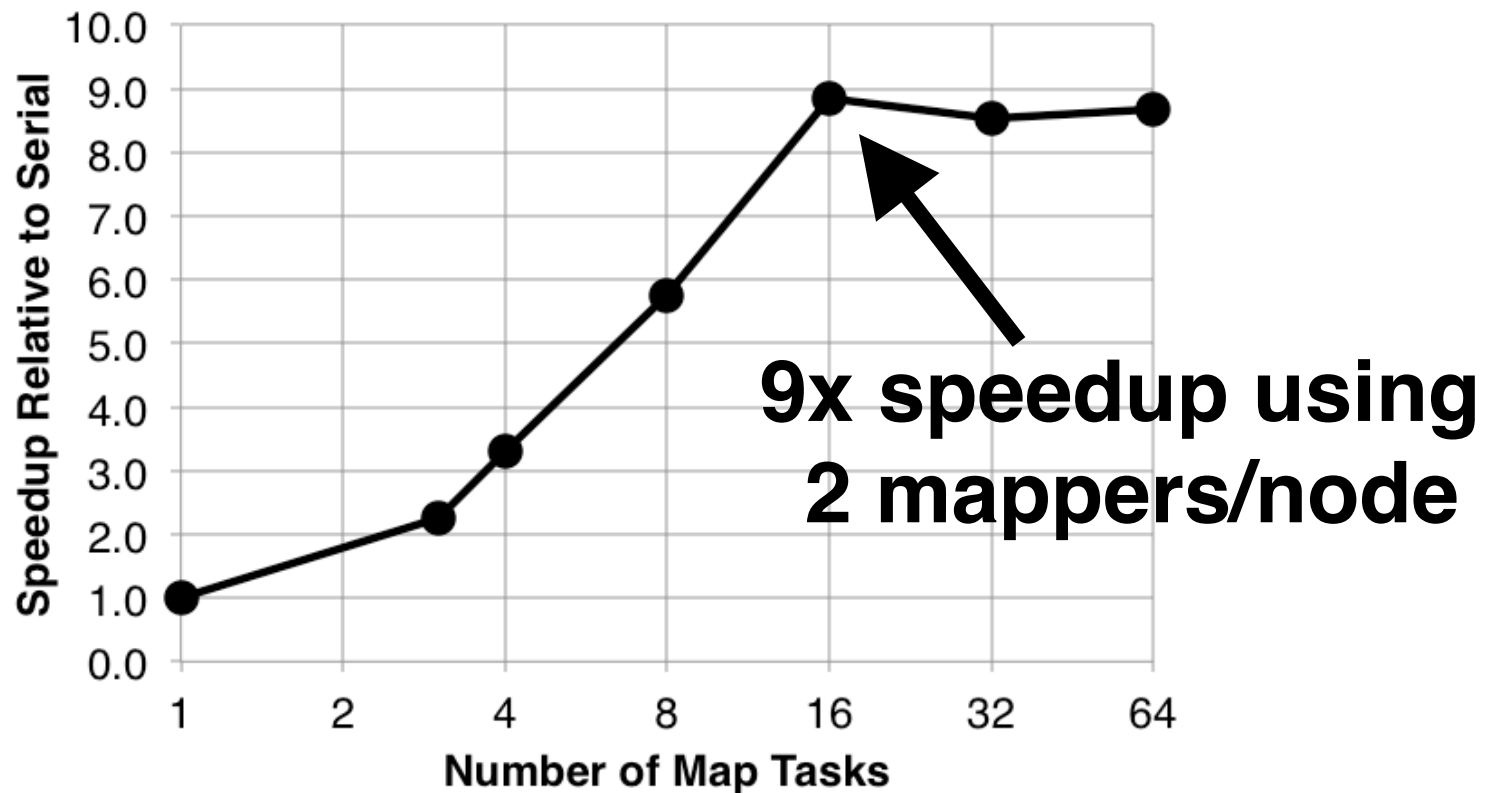
Hadoop on HPC: Experiences

- **myHadoop adds a capability but is imperfect**
- **HDFS is ultimately infrastructure**
 - Users don't like transferring to HDFS for each job
 - Use HPC technologies (Lustre, GPFS) instead
- **Limited adoption *so far***
 - Pre/post-processing (e.g., ETL)
 - Training and experimentation
 - Streaming (Python) and pipes (C++) popular

HDFS on SSDs: ETL of Called Variants

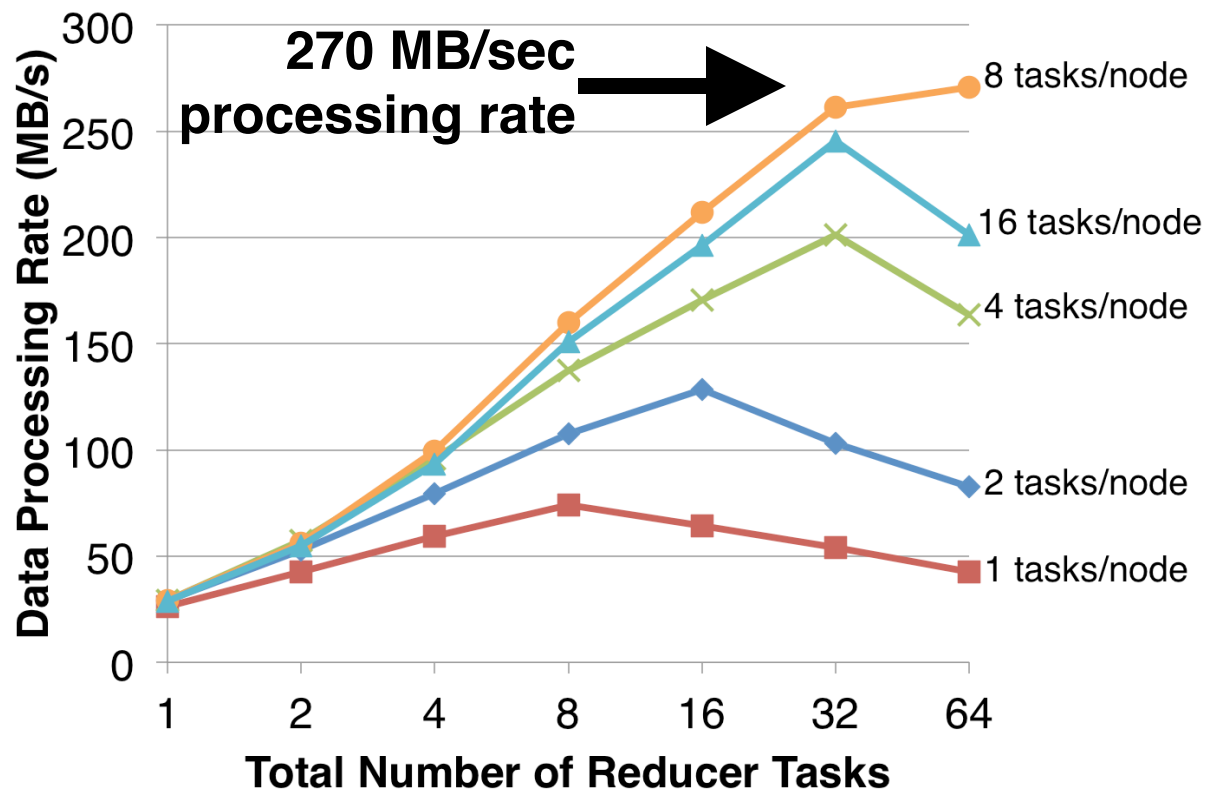
8-node Hadoop cluster on Gordon

8 GB VCF (Variant Calling Format) files



HDFS on Lustre: Performance Scaling

8-node Hadoop cluster on Gordon
7.7 GB of chemical evolution data



Acknowledgments

Janssen R&D

- Chris Huang
- Sarah Lamberth
- Zhenya Cherkas
- Carrie Brodmerkel
- Ed Jaeger
- Martin Dellwo
- Lance Smith
- Mark Curran
- Sandor Szalma
- Guna Rajagopal

SDSC

- Wayne Pfeiffer
- Haisong Cai
- Mahidhar Tatineni
- Rick Wagner
- Bill West

Scripps Translational Science Institute

- Kristopher Standish
- Tristan Carland
- Nicholas J. Schork
(now with J. Craig Venter Institute)

Questions?

glock@sdsc.edu