KickStart 2018

David Swanson, PhD
Director, Holland Computing Center

"HCC operates as a core facility and provides services that are critical to the research productivity of faculty and students across the University of Nebraska campuses."

–HCC 5-year Review, 2016

"HCC exists to reduce the time to science."

–David Swanson

HCC is good at …

- **Big Data**: We manage over 10,000 TeraBytes (10PB)

- **Scaled Out Computing**: 35,000 cores, OSG, advanced networking, high bandwidth transfers, Globus Online

- **Scaled Up Computing**: Parallel processing, XSEDE, GPUs

Recent Study of Data Scientists*

- "Available computing power" was the second most common problem for big data analysis.

- Distributed or parallel processing was the least common solution to their big data needs.

- This could be attributed to the difficulty of processing data on distributed resources.

*K Rexer. Data miner survey-2013 survey summary report. Rexer Analytics, Winchester, 2013.

HCC has great people

- Most are computational scientists

- 5 Full time research personnel

- 7 Sys Admins

- 5 Application Specialists

- Favorite collaborative tools: table with a whiteboard

- Collaboration, not outsourcing outfit

HCC has great people

- 118 Schorr Center

- 152 Peter Kiewit Institute

- Buffet Cancer Center 5.12.397

- hcc-support@unl.edu

- http://hcc.unl.edu

HCC is free*

- *Shared usage of resources:* no charge

- *Best effort support:* no charge

- *Dedicated computing, storage or support:* up-front fees

*to many users! Paid for by NRI, NSF, and NU researchers

Storage Options

- Currently 4 types of storage

  - /home — backed up, per machine, 25 GB/user, free

  - /work — purged, per machine, 50 TB / group, free

  - **/common — not purged, mounted on all HCC machines**

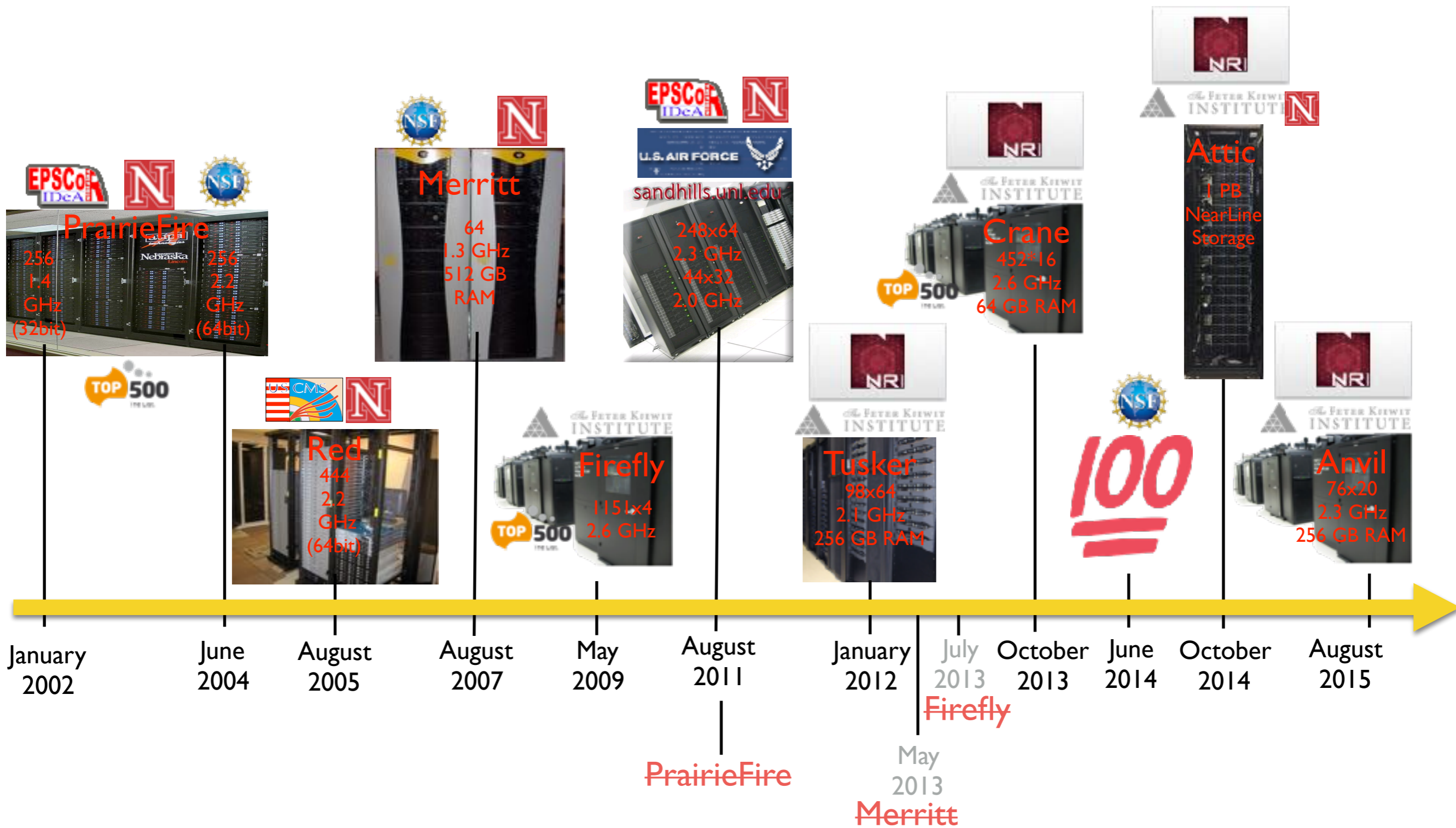 - Attic — near-line archive, not mounted, $25/TB/year

HCC On Ramp

1. Nebraska Faculty Connection (HCC Group)

2. Log in (HCC account and Duo set up)

3. Moving and editing your data (Globus)

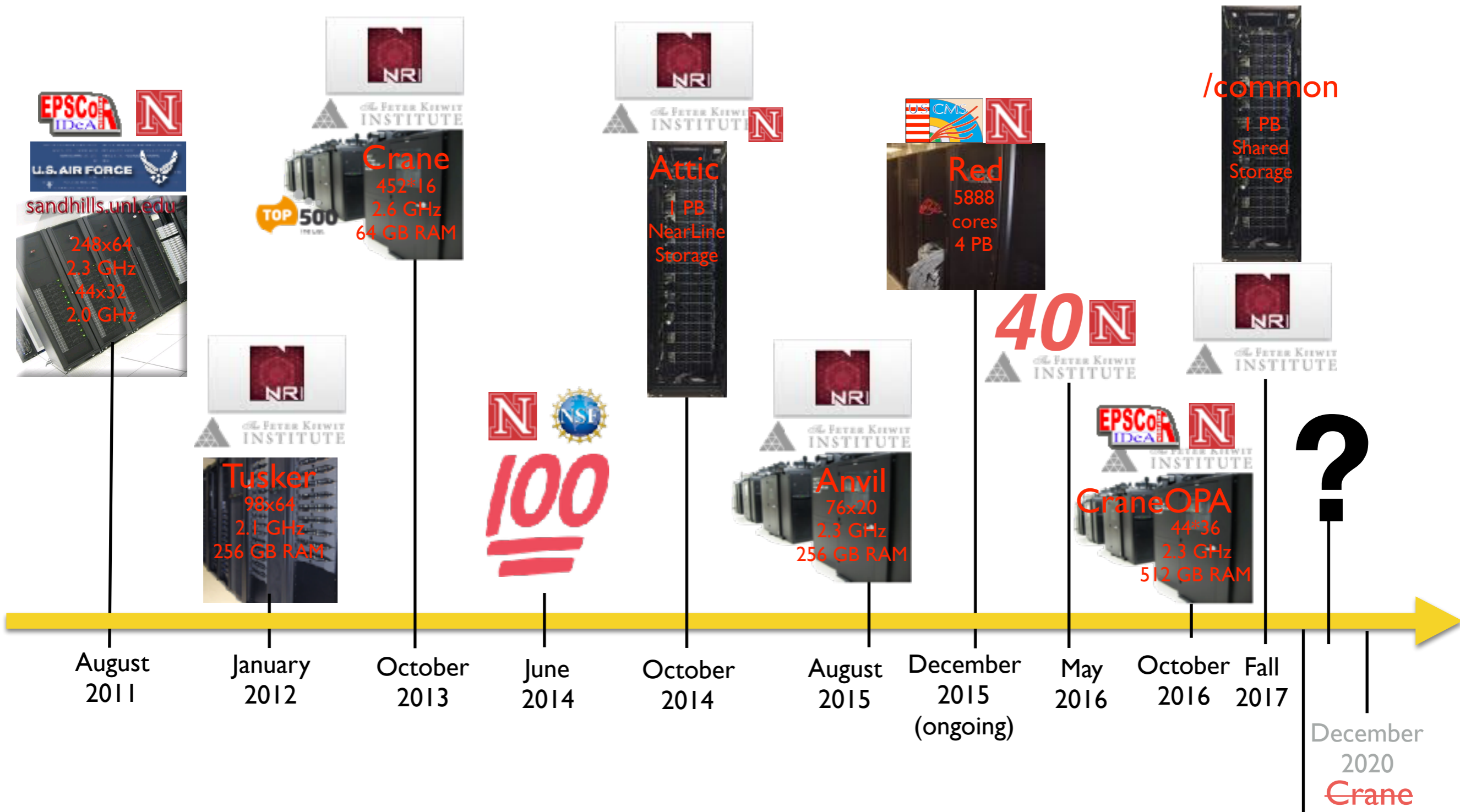4. Launch jobs! (module, SLURM) or (Jupyter Notebooks)

Enough about HCC …

- What is your area of study?

- What question do you wish you could answer?

- Do you currently use HCC resources?

- What type of computing (storage/networking/visualization) do you do now?

- Are you experiencing any bottlenecks?

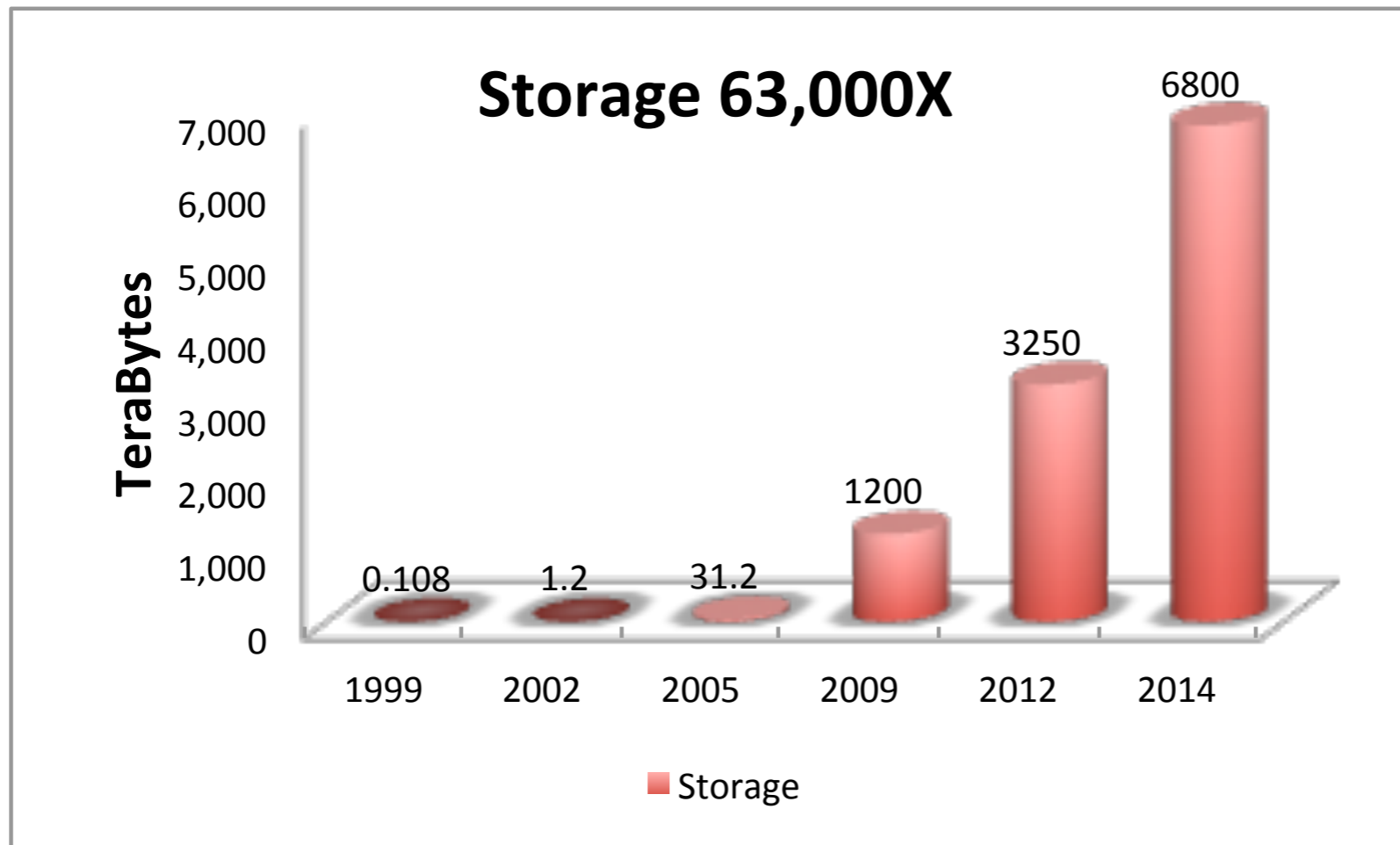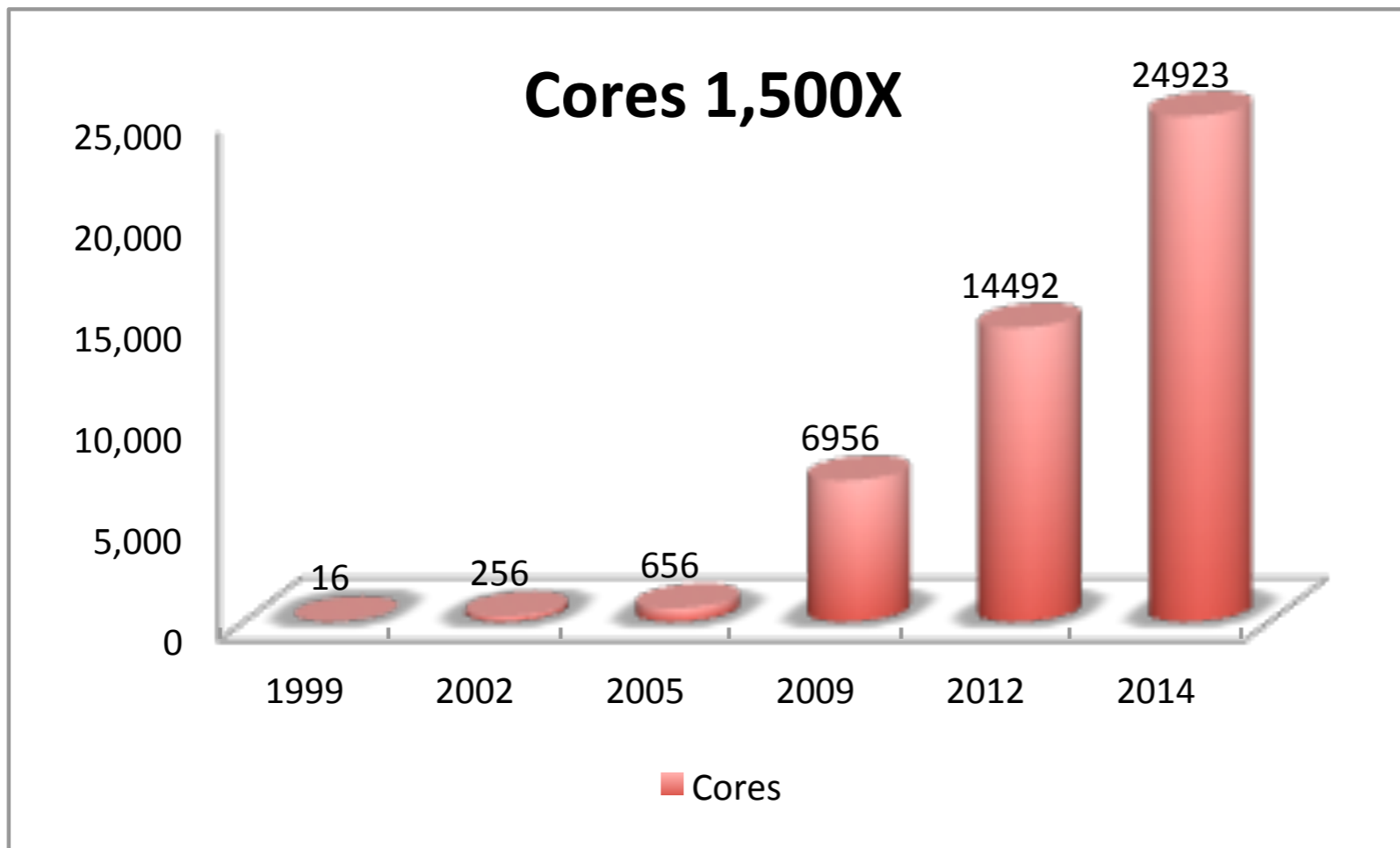# Parallelism Rules

Equipment Timeline

Equipment Timeline

Crane
452*16
2.6 GHz
64 GB RAM

Attic
1 PB
NearLine
Storage

Red
5888
cores
4 PB

/common
1 PB
Shared
Storage

sandhills.unl.edu
248x64
2.3 GHz
44x32
2.0 GHz

Tusker
98x64
2.1 GHz
256 GB RAM

Anvil
76x20
2.3 GHz
256 GB RAM

40

CraneOPA
44*36
2.3 GHz
512 GB RAM

?

August
2011

January
2012

October
2013

June
2014

October
2014

August
2015

December
2015
(ongoing)

May
2016

October
2016

Fall
2017

December
2020
Crane

January
2019
Tusker
Sandhills

More to come...

- HCC is buying incrementally, annually

  - MRI or other grant funding could enhance this

  - What should we target?

- Establishing an annual process

  - Gather input in fall

  - purchase in spring

 - add incrementally until next spring

Cores 1,500X

Year	Cores
1999	16
2002	256
2005	656
2009	6956
2012	14492
2014	24923

Storage 63,000X

TeraBytes

Year	Storage
1999	0.108
2002	1.2
2005	31.2
2009	1200
2012	3250
2014	6800

It takes ca. 6 days to transfer 6,800 TB at 100 Gbps

Science is a Team Sport
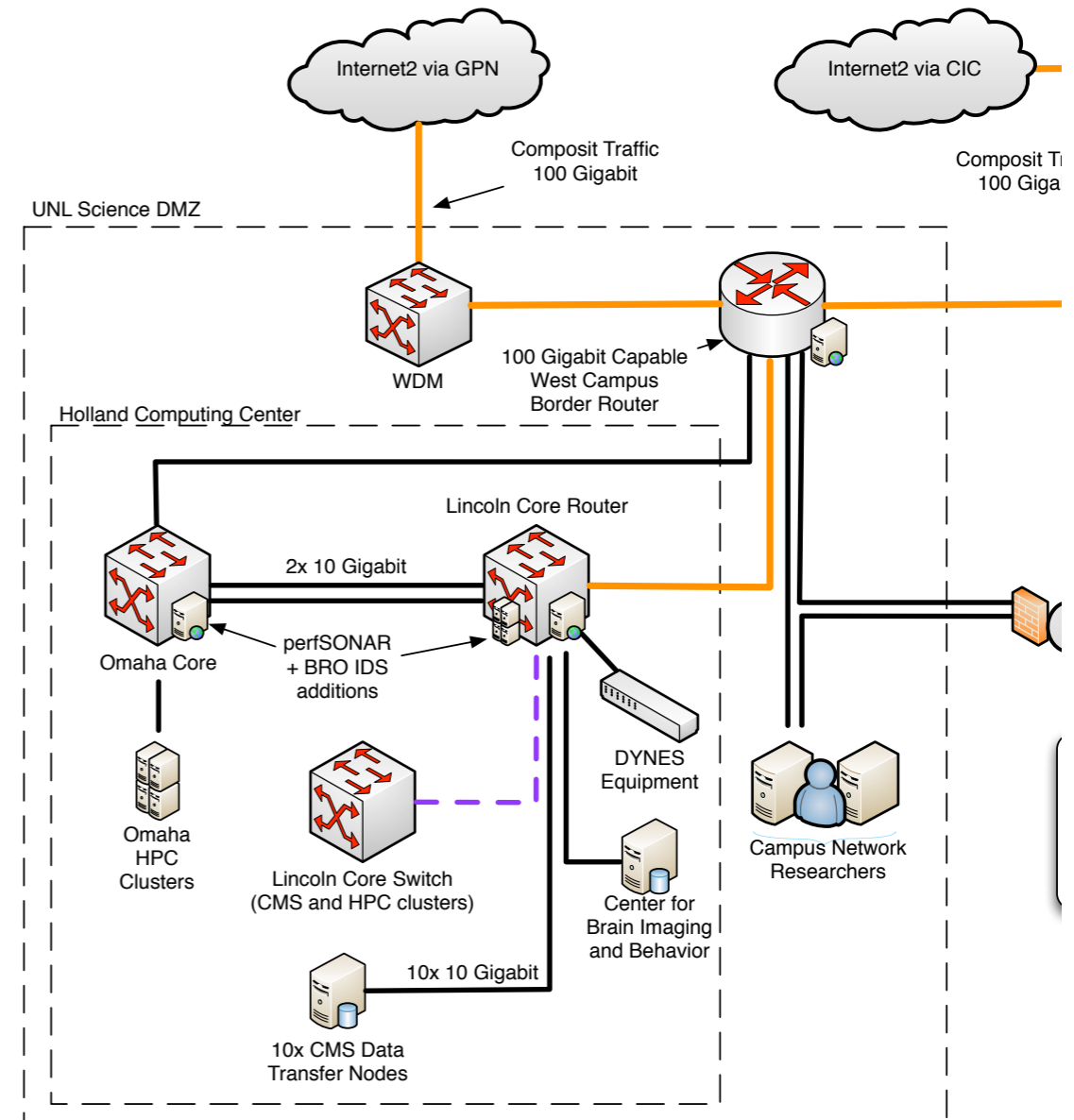
LHC and CMS

- In Run 1 the LHC created 15 PB of data/year

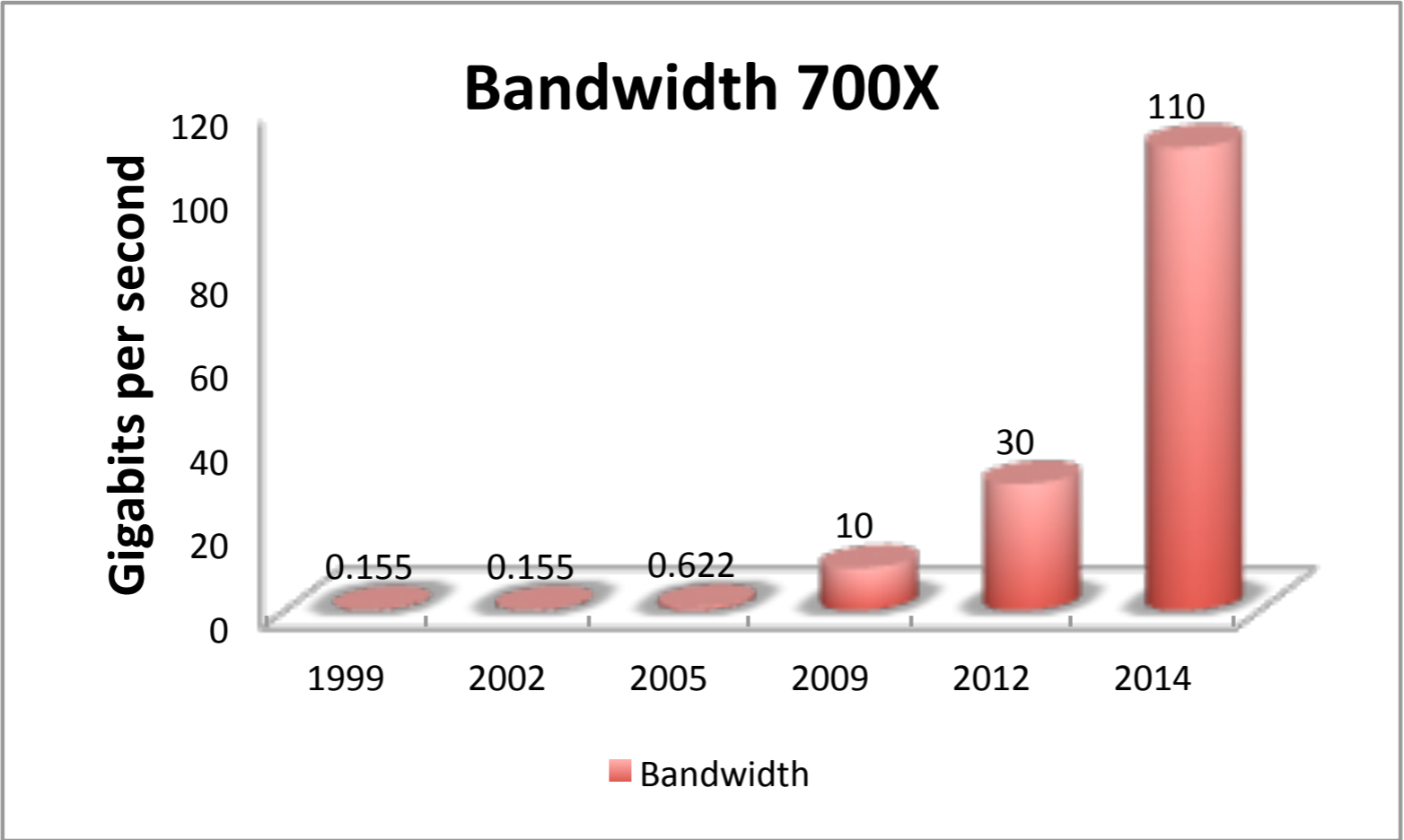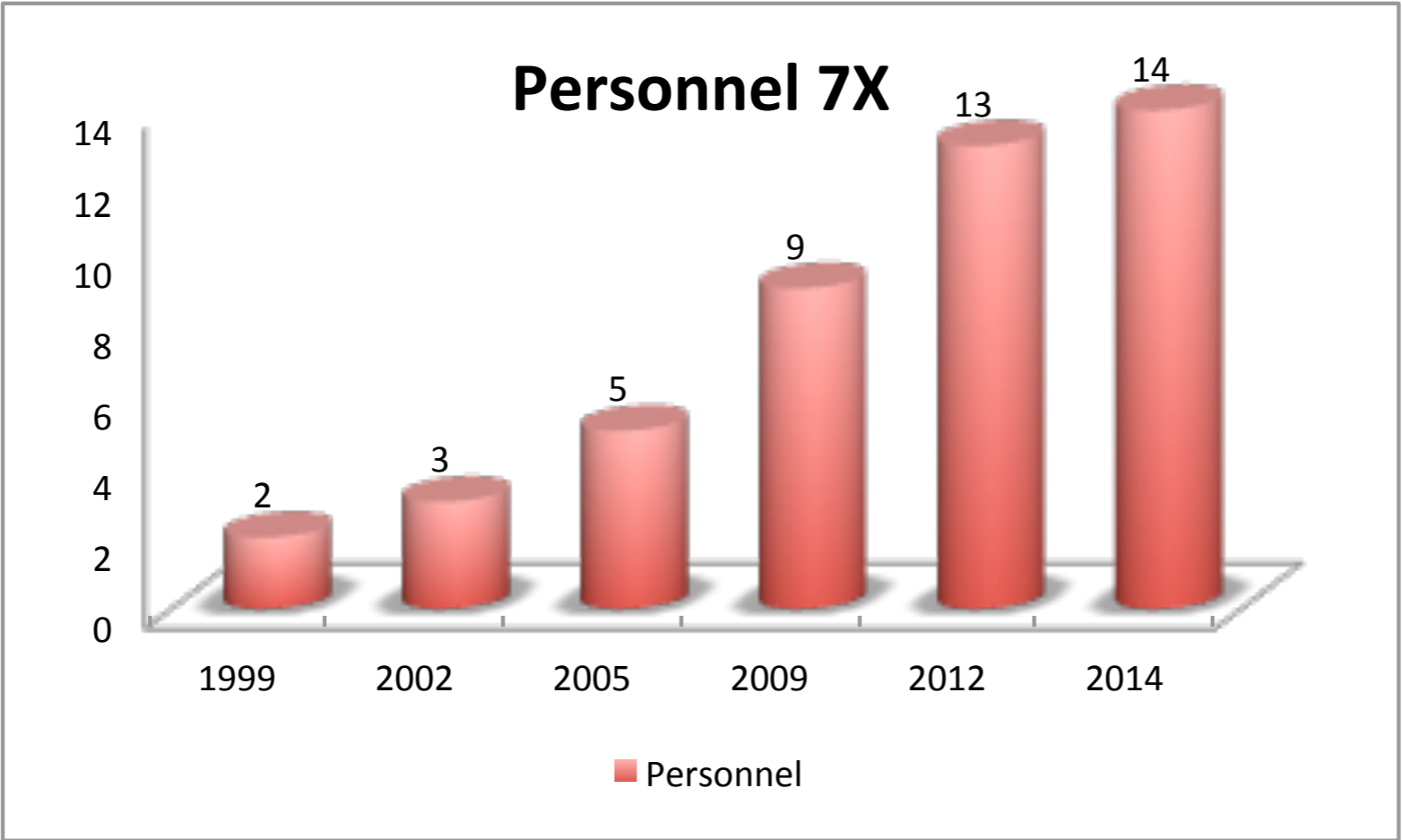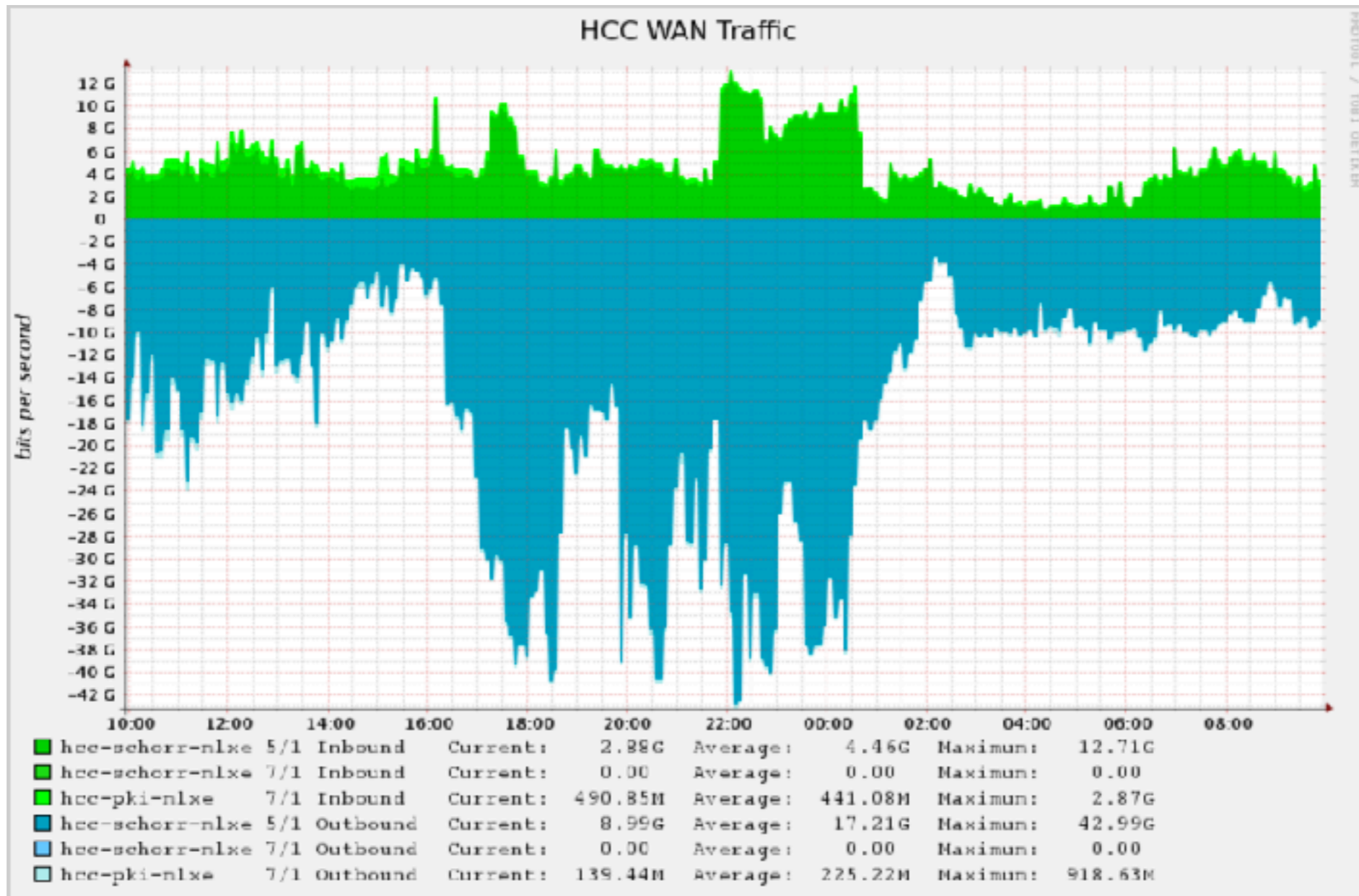- In Run 2 expected output is 30 PB/year (1 GB/sec)
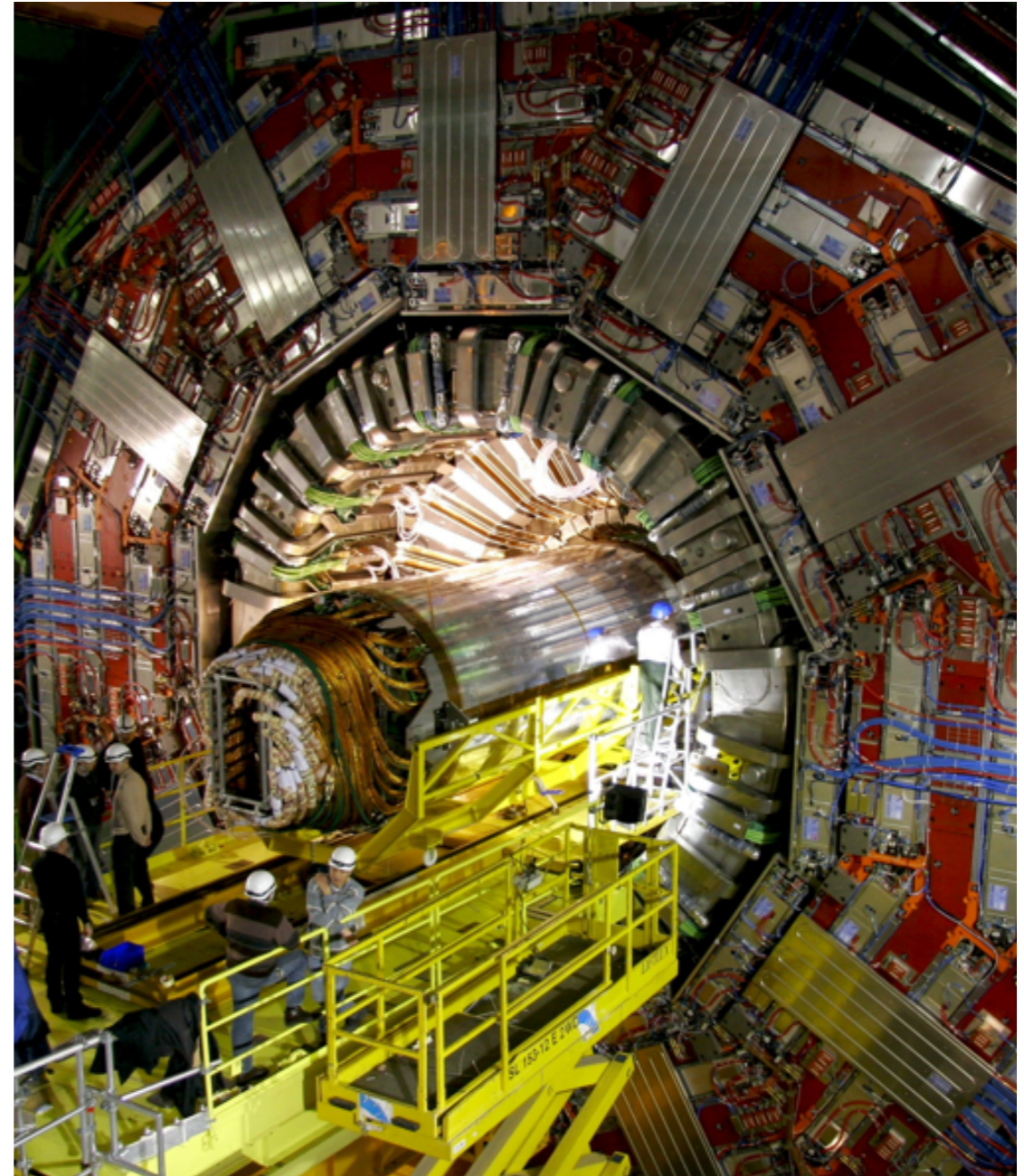
Collaboration Infrastructure

- Network Nebraska

- 100 Gbps between HCC and Internet2 via GPN

- DMZ to/from HCC

- OSG access to HCC resources

- DTNs, Globus Connect

Personnel 7X

Year	Personnel
1999	2
2002	3
2005	5
2009	9
2012	13
2014	14

Bandwidth 700X

Gigabits per second

Year	Bandwidth
1999	0.155
2002	0.155
2005	0.622
2009	10
2012	30
2014	110

Local speed record to date
42 (+10!) Gbps

- Software Defined Networks
- PerfSONAR
- Performance Testing

Getting Started