



HPC CURRENT USE AND PRACTICE **OPEN SCIENCE HPC CENTER**

NSCI: High-Performance Computing Security Workshop

John West
Director of Strategic Initiatives

This is a program manager's perspective
informed by federal, academic, and industry
experience...

WHAT DO YOU DO, FOR WHOM, AND HOW?

What's the application, user community, system architecture?

ONE TACC SYSTEM

Stampede (NSF-funded)

- #6 in 2013, #12 today
- 9.6 PF peak
- 6,400 nodes
 - 2x Xeon Sandy Bridge and Knights Corner
 - Some NVIDIA K20
- InfiniBand (FDR)
- 270 TB memory, 14 PB disk
- 4.5 MW max power draw
- Weighs .5 million pounds

Metric	Past 12 months	Project to Date
# of users running jobs	5,303	10,458
# of projects	1,795	3,145
# of jobs	1,761,366	7,071,847
Core hours delivered	829,804,026	2,649,762,345
System availability	98.88%	97.71%
Tickets resolved	4,720	14,495



For example, in the past year Stampede...

- Delivered calculations in support of the LIGO team
- Gordon Bell prize for Global Mantle Convection

This is world-class, science-changing computation delivered in a university context.

MODERN SCIENTIFIC COMPUTING

The kinds of things researchers do with computing systems has evolved tremendously over the last 20 years and now includes:

- ▶ Simulation
- ▶ Data Visualization
- ▶ Data Analysis
- ▶ Data Mining/Clustering
- ▶ Storing/Sharing/Curating/Reproducing Digital Objects with Colleagues

TACC IS A SYSTEM OF SYSTEMS



Stampede (Open Science)

Lonestar 5 (ITAR, HIPAA)

- 30k Intel Haswell cores 1.25 PF

Wrangler (HPC + Big Data)

- 0.6 PB usable DSSD flash storage w 1 TB/s read rate + 10 PB Lustre

Maverick (Data Analysis)

- 132 Fat nodes w dual 10 core Ivy Bridge + NVIDIA Kepler K40

Chameleon & Jetstream Cloud

- 1400 nodes OpenStack

Disk and Tape Storage

- 100+ PB storage in HIPAA-aligned data center

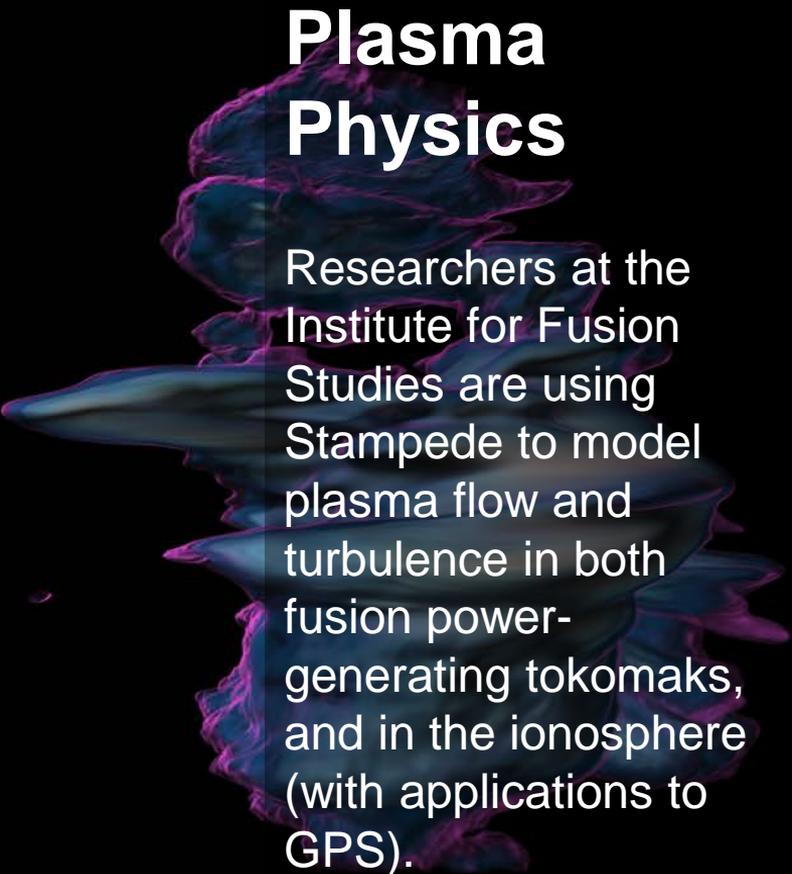


Hikari (Datacenter practices)

- 380V DC Green computing system partnership with NEDO and NTT
- 10k Haswell cores
- HVDC and Solar
- Support for container ecosystem

Plasma Physics

Researchers at the Institute for Fusion Studies are using Stampede to model plasma flow and turbulence in both fusion power-generating tokomaks, and in the ionosphere (with applications to GPS).



Safety and Severe Storm Prediction

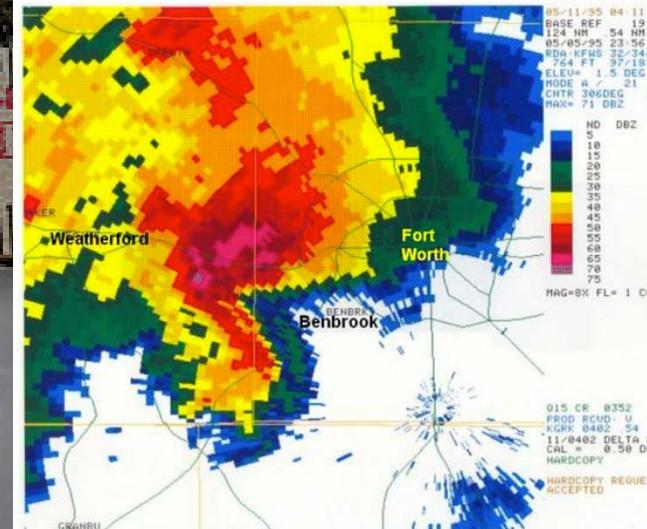
The Oklahoma University CAPS Center is blending ensemble simulations, data assimilation, and machine learning to advance hail forecasts from 2 hours to 24 hours.



PREDICTING SEVERE HAIL STORMS

NSF-supported research at the University of Oklahoma uses supercomputers and simulations to improve storm forecasts

Published on March 22, 2016 by Aaron Dubrow



Radar imagery from 6:56pm shows a close-up of the Mayfest supercell centered west of Benbrook, Texas. The pink and darkest red colors represent radar indicators of large hail with this storm. The storm impacted the Mayfest festival at 7:10pm. Credit: National Weather Service

Finding our Ancestors

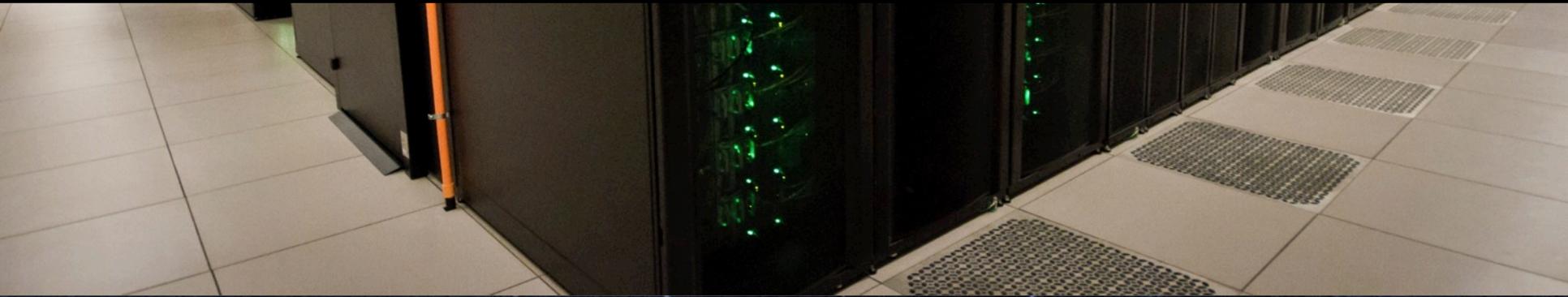
Genomics analysis performed of nine ancient Europeans that bridge the transition to agriculture in Europe between 8,000 and 7,000 years ago reveals that most present-day Europeans derive from at least three highly differentiated populations.



HOW DO WE SUPPORT RESEARCHERS WITH SUCH DIVERSE NEEDS AND BACKGROUNDS?



BUILD A MASSIVE STORAGE CLOUD NEXT TO INNOVATIVE, POWERFUL, USABLE COMPUTERS AT THE END OF FAST INTERNET PIPES



MANY DOMAIN SCIENTISTS ARE NOT EXPERTS AT COMPUTING TECHNOLOGY. CREATE PURPOSE-BUILT, HIGHLY INTUITIVE INTERFACES

ThaleMine

Drill down on the genome. Analyze gene and protein data from TAIR, expression and interaction data from BAR, ortholog data from Panther, and more. Generate, save, and analyze result sets. Use plain text or structured queries to find interactive gene pages and protein reports.

[Mine](#)

JBrowse

Scroll through the genome and zoom in for details. Layer on annotation tracks that display TAIR 10 gene structure data, EPIC-CoGe epigenomics data, and more. Use the latest, fully featured GMOD JBrowse tool (or, for a limited time, the older GBrowse tool).

[Browse](#)

Science Applications

Find other interactive application programs that help you explore plant genomics. AIP hosts applications that integrate data from other web sites using the web services model of on-request data exchange. AIP hosts science apps that were contributed by its members and its users.

[Apps](#)

Find here the latest news and highlights on important natural hazards research discoveries and NHERI program announcements including upcoming meetings. Resources and collaboration tools for the NHERI community will also be provided here.

News & Features

[Register for NHERI Experimental Facilities Workshops](#)
10-19-15
NHERI Experimental Facilities are hosting

<http://araport.org>

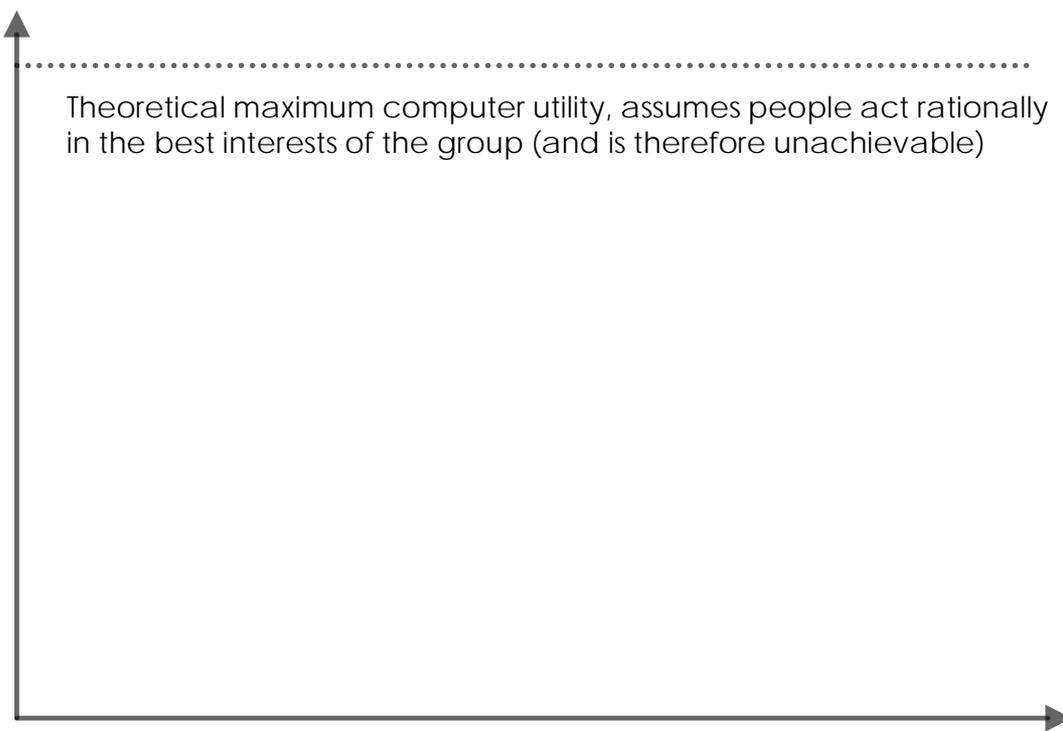
<https://www.designsafe-ci.org>



GIVE EXPERTS ACCESS TO EVERY SINGLE ONE OF YOUR BUILDING BLOCKS.
WEB SERVICE APIs EVERYWHERE. AUGMENT WITH PROFESSIONAL TOOLING.

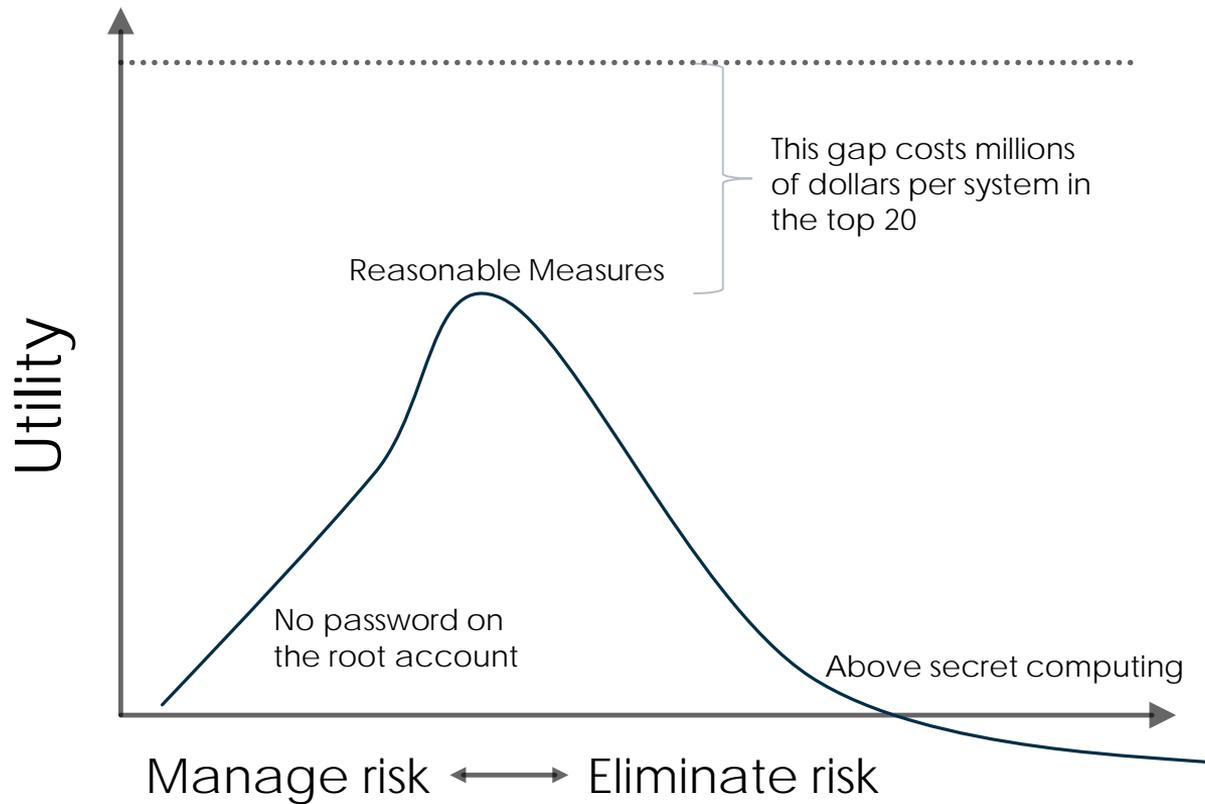


HOW TO TRADE USABILITY AGAINST SECURITY?



Utility

Theoretical maximum computer utility, assumes people act rationally in the best interests of the group (and is therefore unachievable)



BEST OF ENEMIES

- ▶ After establishment of basic order (there is a root password, etc.), computer security practices are a direct and unavoidable drag on user productivity **in the best case**
- ▶ Most mission agencies don't have the best case
 - ▶ Cybersecurity management does not believe in risk management; it believes in risk elimination
 - ▶ (They will tell you otherwise – they are “confused”)
- ▶ The choice is not to tradeoff usability against security, it is to tradeoff scientific and engineering discovery against exposure to risk

TRADING OFF

- ▶ Heavily context dependent
- ▶ Some data are no kidding dangerous, and most risks should be eliminated for these
 - ▶ When you do this, you must KNOW you are willingly sacrificing usefulness / productivity / usability / discovery to protect the data
 - ▶ So don't do it often
- ▶ But this isn't most data
 - ▶ Yet risk elimination pervades the establishment

CYBERSECURITY @ TACC

- ▶ TACC has a “maintain order” model that tries to stay to the left of the peak
 - ▶ Multi-factor authentication, all systems
 - ▶ Escalation for root, no remote login
 - ▶ Intrusion Detection Systems and centralized logging
 - ▶ Support elevated sensitivity data in special enclaves (ITAR, HIPAA, etc.)

TRADING OFF IN THE CONTEXT OF NSCI

- ▶ You (probably) can't change the establishment
- ▶ So build a portfolio for advanced computation that includes NSCI-scale systems deployed in contexts that are culturally capable of treating risk in a way that matches your explicit goals for the system
 - ▶ National security engineering and intelligence
 - ▶ Dual-use computations
 - ▶ Open science

▶ TACC

- ▶ Hosts over 10,000 users who are citizens of scores of countries
- ▶ zero of whom have had a national agency check or background investigation or driver's license check
- ▶ While delivering
 - ▶ 10+ billion CPU hours over 5 years on
 - ▶ 20+ HPC systems
 - ▶ Using open software frameworks and web portals
- ▶ TACC doesn't participate in much security theater
 - ▶ require 15 character passwords you change every 30 days, that must also contain a mix of upper case letters, lower case letters, numbers, and special characters, that must not be in the previous 10 passwords used.

And yet, useful work gets done

Thanks!
john@tacc.utexas.edu