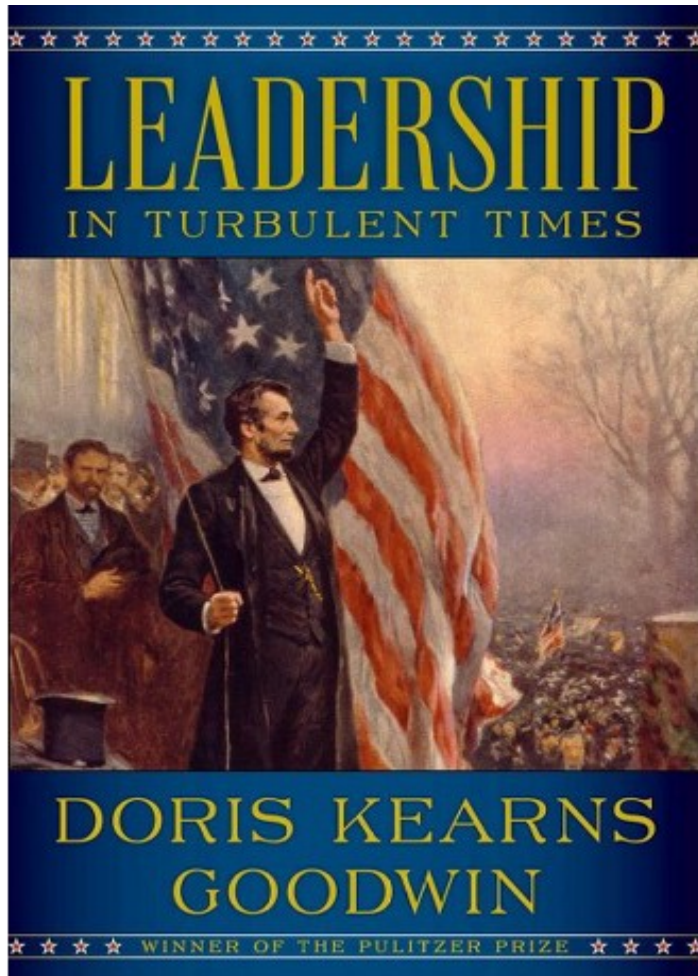


NESE & News from the Northeast

Saul Youssef
Boston University

NESE & News from the Northeast



This is a joke 😊

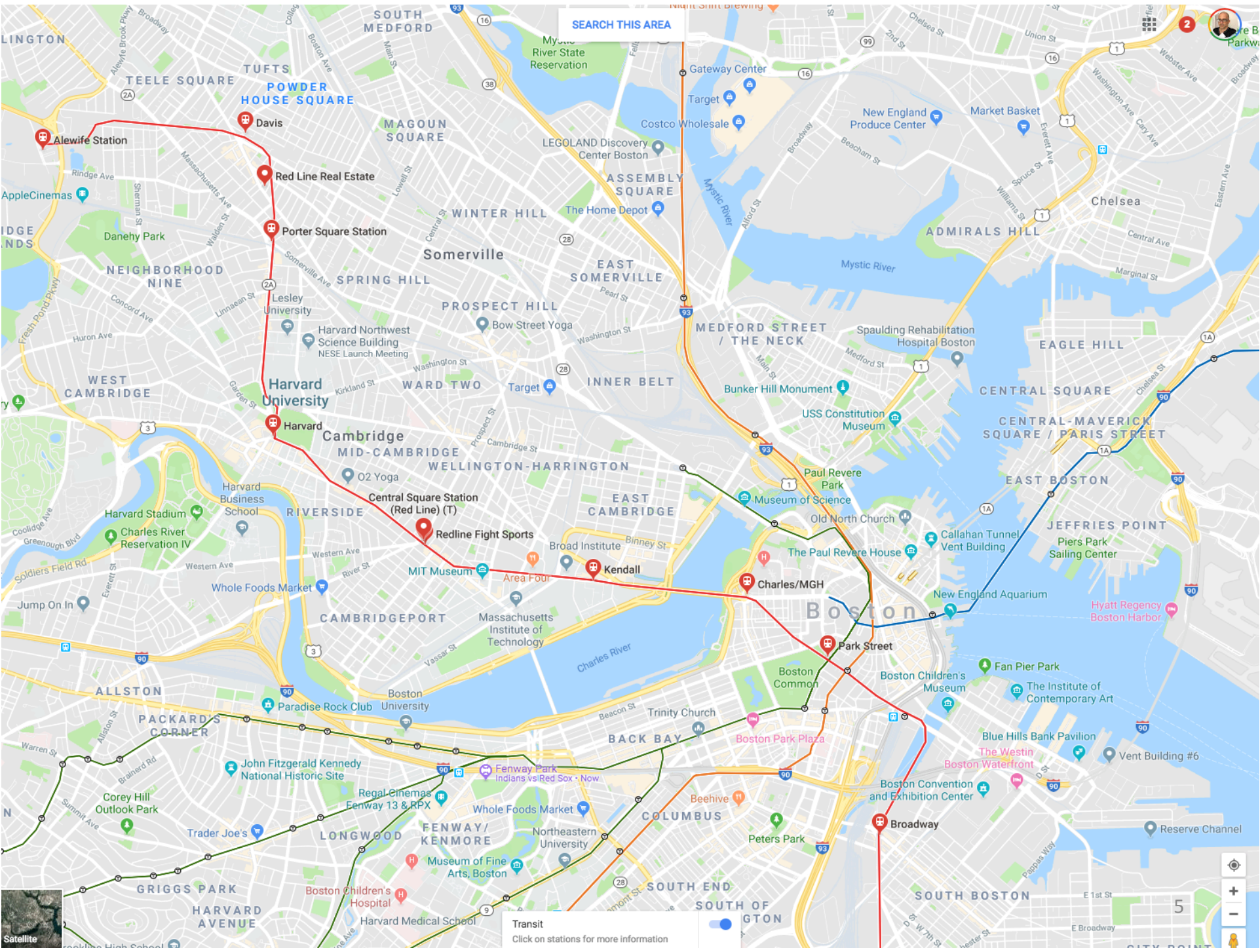
We work in turbulent times of disruptive change

- Cloud Computing
- Deep Learning and AI
- IoT and Data Science
- Massive urgent scientific challenges
- Uncontrolled security threats

- New institutes
- New national organizations
- New software
- New mandates

Next stop Wonderland





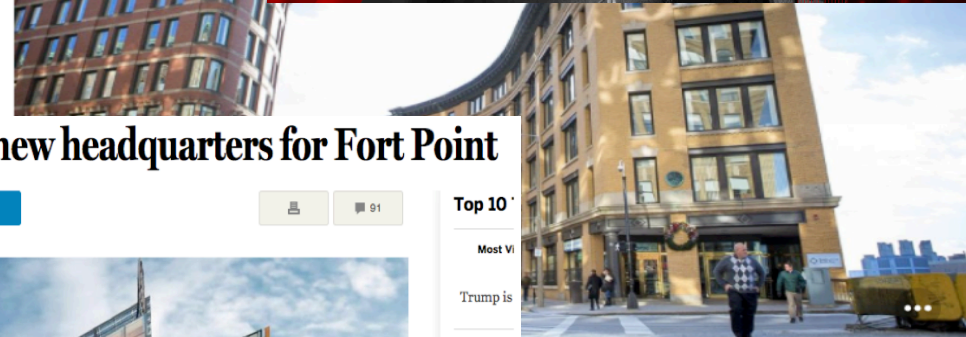
Transit
Click on stations for more information



Technology companies

RedHat
Intel
DELL/EMC
Google
NVIDIA
Amazon
Facebook
IBM
Microsoft
Wolfram Research
Mellanox
Cisco
General Electric
...

Amazon Says 900 Jobs Will Be Added At New Boston Office



GE unveils striking new headquarters for Fort Point



Amazon to add 2,000 jobs in Boston

- Boston is one of the tech giant's largest footholds and is a finalist to win the company's second headquarters.
- Amazon said Tuesday it's invested more than \$400 million in Massachusetts



Boston is No. 1 in the country for startups (again), says U.S. Chamber

[✉ Email](#) [f Share](#) [in Share](#) [🐦 Tweet](#) [★ Save](#) [🖨 Print](#) [📄 Order Reprints](#)



Tufts Health Plan n
monitoring your he
plan as easy as che
your phone.



[Learn More](#)

TRENDING

COMMERCIAL REAL ESTATE

Construction worker killed at Wynn Boston Harbor construction site

7

HEALTH CARE

[Our Mission](#)

[Our Investments](#)

[Our Team](#)

[Board of Directors](#)

[Scientific Advisory Board](#)

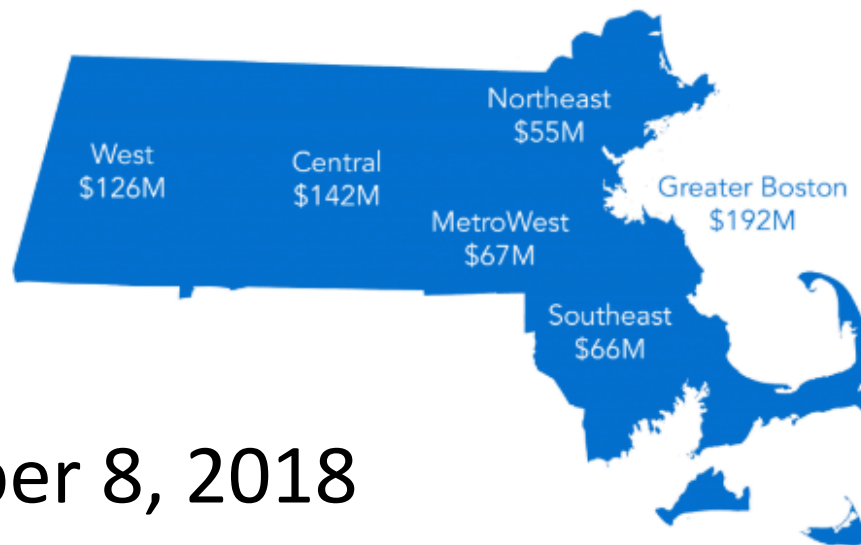
[Biomanufacturing Advisory Panel](#)

[MLSC Careers](#)

Our Investments

The Massachusetts Life Sciences Center has thus far invested or committed more than \$650 million* in state funding, leveraged more than \$2.8 billion in matching investment capital, and made investments thus far projected to create thousands of new jobs in the Commonwealth.

*approximate distribution of investments the MLSC has made across the state, rounded to the nearest million, as of June 2017



Due November 8, 2018

We're swimming in new Data Science Institutes

For Data Science, location matters!

Lichtman Lab at Harvard
Home Research People Papers Press Resources Contact Us

Saturated Reconstruction of a Volume of Neocortex
Our July 2015 report in Cell →

CATALYST FEATURES, UNCATEGORIZED
INTERVIEW: BU physicist on creating new, math-driven field of medicine
September 29, 2015 12:29 am | by Ross Hu

Imagine a world in which people can monitor their health the same way they would check the weather on their phones.

This is the world envisioned by Plamen Ivanov, a professor of physics at Boston University, and the team he leads at the Laboratory for Network Physiology, a project that over the summer was awarded \$1 million for medical research from the W.M. Keck Foundation.

"We proposed to Boston University and then to the Keck Foundation to develop an atlas," Ivanov said, "an atlas of dynamic interactions among organ systems."

Ivanov and his team call their new field "network."

Boston University physicist Plamen Ivanov is coordinating organ system functions on a network. PHOTO BY MAE DAVTCHIAN

Rajen Kilachand Center for Integrated Life Sciences & Engineering

Professor Andrew McCallum, Priscilla Chan and Mark Zuckerberg

- [Boston University Data Science Initiative](#)
- [Harvard Data Science Initiative](#)
- [MIT Institute for Data, Systems and Society](#)
- [Data Science at Northeastern University](#)
- [Center for Data Science, UMASS/Amherst](#)

• • •

Professor Andrew McCallum, Priscilla Chan and Mark Zuckerberg

January 16, 2018

UMass Amherst Center for Data Science partners with Chan Zuckerberg Initiative

BY: ED BLAGUSZEWSKI



Boston University  @BU_Tweets · 14m

President Brown today announced a proposed project to give data science a home on Comm Ave — a 17-floor building meant to resemble a stack of books that will be home to @BU_Computing, mathematics and statistics, and computer science.



BU to Build Data Sciences Center

Boston University announced its plans to build the BU Data Sciences Center. By bringing mathematics and statistics and computer science dep...

bu.edu

October 1, 2018

M.I.T. Plans College for Artificial Intelligence, Backed by \$1 Billion



The Massachusetts Institute of Technology is taking a particularly ambitious step in preparing students to develop, and consider the implications of, artificial intelligence. It is creating a new college, backed by a planned investment of \$1 billion.

Cody O'Loughlin for The New York Times

New York Times...*Yesterday!*¹¹





Lack of Perspective



Logan
Airport





MGHPCC

Holyoke, Massachusetts

Near zero carbon footprint

>350,000 x86 cores now

Space/power/cooling for 785 racks

90,000 square feet of computer floor

Space for 2d building or pods

Consortium: BU, Harvard, MIT, NEU, UMASS,
Commonwealth of Massachusetts, DELL/EMC and Cisco



National compute & storage resources

San Diego Supercomputer Center:

- Comet: 46752 CPUs
- Comet GPU: 1728 processors
- Oasis (storage): 4000 TB



Pittsburgh Supercomputing Center

- Bridges: 21056 CPUs
- Bridges GPU: 1344 processors
- Bridges large memory: 160 CPUs
- Bridges Pylon (storage): 10000 TB

● Compute Resources
● Storage Resources

MGHPCC



Texas Advanced Computing Center:

- Stampede2: 368280 CPUs
- Wrangler: 2304 CPUs
- Ranch (storage): 61440 TB
- Wrangler (storage): 10000 TB

350,000 cores
~50000 TB storage

NORTHEAST CYBERTEAM



The Northeast Cyberteam is an NSF-funded initiative to build a team of research computing facilitators (RCFs) to support researchers at small and mid-sized institutions in the region.

We are looking for exceptional individuals who have some research computing experience, and are interested in learning more! If matched to a project, you will be assigned a mentor, and together you will provide assistance to a researcher working on a computationally intensive project. Stipends are available and current projects are listed below.

NSF Funded...., John Goodhue PI & MGHPCC Executive Director

- + University of Maine
- + University of Vermont
- + University of New Hampshire

NORTHEAST CYBERTEAM



About Us



computing power in the northeast.

Contact Us

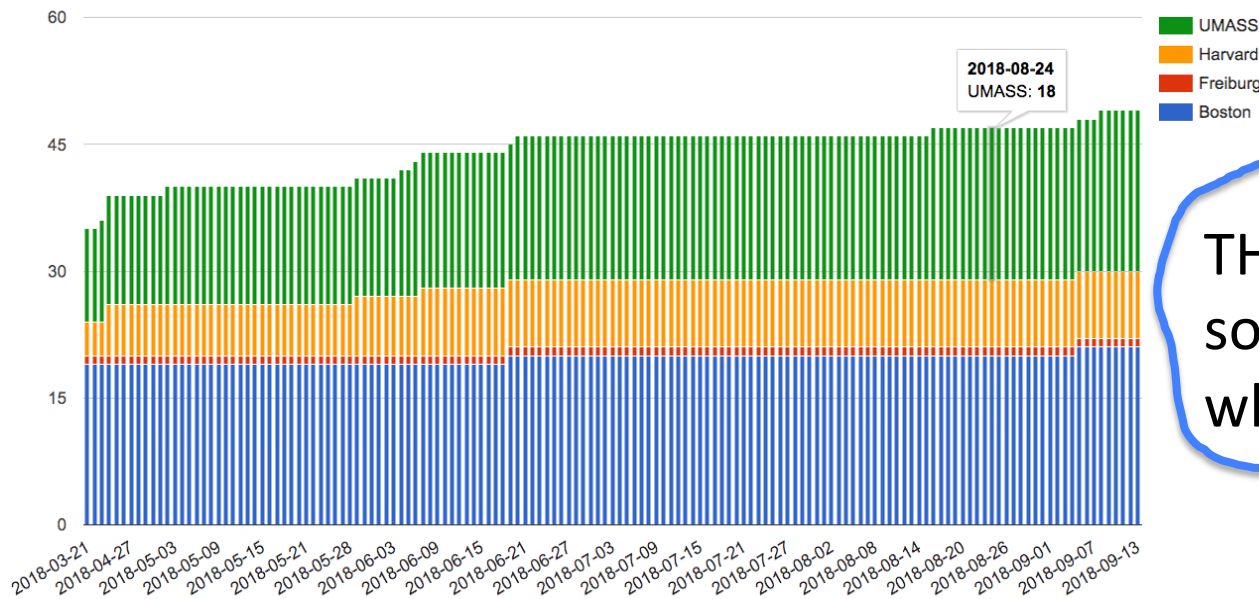
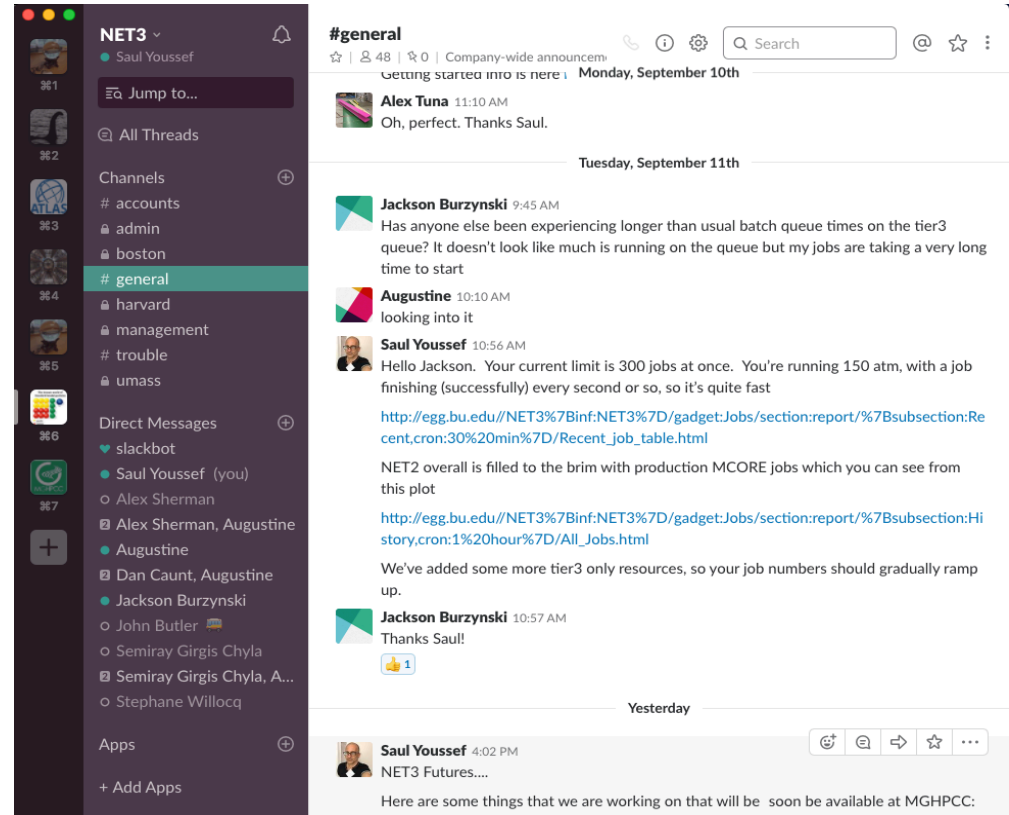
Northeast Tier 3

Boston University

Harvard University

UMASS/Amherst

- Operated by Wayne's group as part of NET2
- 49 Users
- Gradually growing
- Almost as many UMASS users as BU users
- Essentially no extra work for Augustine
- Use of Slack is extremely helpful
- UMASS bought in with three nodes



THEME: If you're good at something, do it for the whole consortium.

THEME: If you're good at something, do it for the whole consortium.

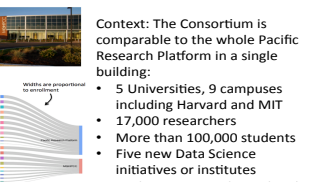
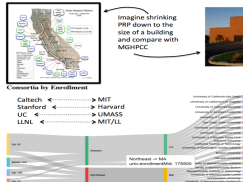
1st Deployment being cabled now. 12PB starter; 40+% buy-in already RedHat Partnership Going!

NESE: The Northeast Storage Exchange

Saul Youssef,¹ Scott Yockel,² Chris Hill,³ John Goodhue,⁴ Devesh Tiwari,⁵ and Mike Zink⁶
 Boston University,¹ Harvard University,² MIT,³ MGHPCPC,⁴ Northeastern University,⁵ University of Massachusetts⁶



Massachusetts Green High Performance Computing Center (MGHPCC)
 • 15 megawatts, 785 racks, approx. 2/3 full now
 • >350,000 x86 cores now
 • Redundant 100Gb fiber ring to national research networks
 • Secure, single use site
 • Room for second building and/or expansion via pods



PROJECT ORGANIZATION

- Management, planning, outreach, science coordination, sustainability: *All PIs*
- Operations: *Scott Yockel*, Harvard FAS Research Computing, BU, RedHat
- Networking: *Jefferson Burson*, Harvard University IT
- Technology planning: *Devesh Tiwari*, Northeastern University
- Ocean apps and iRods: *Christopher Hill*, MIT
- Block storage for clouds, CloudLab: *Mike Zink*, UMass/Amherst
- NET2, LHC applications, federated LHC storage: *Saul Youssef*, Boston University
- Collaboration with RedHat, Mass. Open Cloud: *Orran Krieger*, Boston University
- Authentication and Globus: *Jim Culbert*, MGHPCPC
- Education, Outreach, Open Storage Platform: *John Goodhue*, MGHPCPC

For Data Science, location matters!



DATA SCIENCE
 Every one of the five Consortium member universities has a new data science institute, initiative or program.

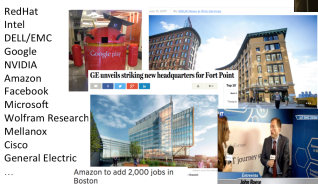
- Boston University Data Science Initiative
- Harvard Data Science Initiative
- MIT Institute for Data, Systems and Society
- Data Science at Northeastern University
- Center for Data Science, UMass/Amherst

Every Consortium university will be able to create their own control room showing Northeast U.S. cyberinfrastructure operations including NESE. This will be great for:
 • Community building
 • Student projects and training
 • Inspiring new generations of operations software based on Data Science and AI

Amazon Says 900 Jobs Will Be Added At New Boston Office

Partnership with RedHat

- Via BU/RedHat "Collaboratory"
- Design consultation...including with Sage Weil
- Installation help
- Embedded RedHat personnel
- Use of RedHat facilities in Boston
- Premium software Subscriptions



GOALS

- To meet the storage needs of the data revolution for science, engineering, education and technology.
- To be the start of a national cyberinfrastructure in the Northeast U.S.

STRATEGY

- Take advantage of the basic economics of storage.
- Use our unique consortium and the MGHPCPC facility as a starting point.
- Organize the project for long term growth, with institutional responsibilities, community building, long term planning, integration with education, and long term technology tracking.
- Use our unique environment for partnerships with universities, technology companies, biotechs, hospitals, institutes, and new data science centers.



University buy-in, Project buy-in, NESE project funds, Inherited Equipment

Item	Value
2 x 100Gb/s uplinks to 100Gb/s Fabric	
Basic Storage Unit	
1U	As dense as 60 bay J80D
	12 x 10TB 7200 rpm spinning
	Mass Storage
	4 x 280GB Micron high endurance SSD
	Bluestore Ceph + CephFS
	1 x 32GB NVMe
	Ceph write-ahead log
	2 x 100Gb/s
	Matches drives I/O
	1 x Single Socket Intel Xeon SkyLake 2.2 GHz (3.0 GHz Turbo), 35C/20T with AVX-512 SMT0 coprocessor

3.9 PB raw per rack when fitted
 Space for more buy-ins
 Start of UMass/Amherst buy-in

First Deployment

- Six racks at two sites
- 12 PB raw storage
- Small, flexible 120TB OSD
- Designed for all storage types: CephFS, Block storage, S3, Swift, Globus endpoint
- SSDs for Ceph Bluestore
- 100Gb/s networking

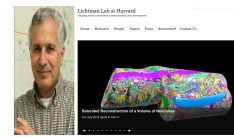
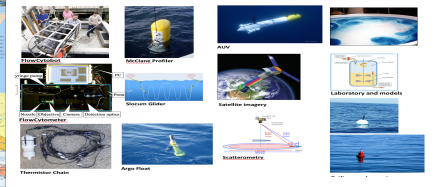
Buy-ins done or planned from...

- Harvard FAS RC
- MIT
- UMASS/Amherst
- Boston University
- NET2 project
- We already have 40% buy-in**

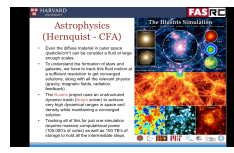
FIRST SCIENCE



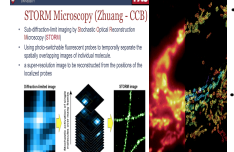
- NESE storage for the U.S. ATLAS Northeast Tier 2 center (NET2, BU/ Harvard) is one of the first production applications. 6PB and growing.
- Federating NESE storage with OSIRIS/Also NSF-DIBBS and with tape storage at Brookhaven National Laboratory to produce a prototype national federation which can be scaled to the 1 EB needed in 2026 for the high luminosity LHC
- Ocean Modeling by Chris Hill's group at MIT uses data from ships, fixed stations, from space and from simulations.



- Jeffrey Lichtman's lab at Harvard produces ~10 PB per year of electron microscopy images of mouse brains with the goal of mapping neuronal circuits and understanding brain function
- Mark Friedl of BU needs to analyze 10s of PB of multi-spectral time-series satellite images of the Earth.



- Lars Hernquist's group at Harvard studies galaxy formation and cosmology with simulations which need 100s of TB for intermediate steps.
- A next-gen gene sequencing facility at Harvard needs to store and process tens of TB per day.

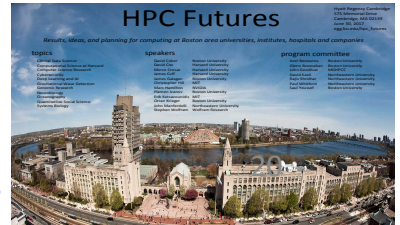


- Advanced microscopy at Harvard images mitochondria in single cells and produces approximately 1 PB of data per year.
- The CASA project at UMass/Amherst aims to revolutionize rapid response to hazardous weather. Data currently at DFW metplex.

INTERVIEW: BU physicist on creating new, math-driven field of medicine



- Network physiology is a new field invented by Plamen Ivanov at BU where unique medically important insights can be extracted from multi-source human physiology data with fast enough access for clinical response.
- These are only some of the first expected NESE applications to be discussed at meetings like HPC Futures.



MIT Supercloud

Education and Outreach

PILOT project including BU

THEME: If you're good at something, do it for the whole consortium.

- Home
- Requesting An Account
- Getting Started
- How To Use
- Jupyter Portal
- Online Courses
- Contact

Login using Touchstone

Systems and Software

This page lists information about the system and available [software](#), [languages](#), [compilers](#), [modules](#), etc. This is only a partial list, so if there is anything you are interested in that isn't listed here, please [contact us](#).

MGHPCC TX-E1 Specifications

Summary						
Number of Nodes		56				
Total CPU Cores		1348				
Total GPUs		10				
Distributed Storage		873 TB				
CPU-Only Nodes						
Processor	Nodes	Cores	RAM	Local Disk		
Intel Xeon	25	16	64 GB	16 TB		
Intel Xeon	7	2 x 14	256 GB	12 TB		
AMD Opteron*	20	2 x 16	192 GB	8 TB		
GPU Nodes (Intel Xeon CPU)						
GPU Type	GPU Interconnect	Nodes	GPUs	CPU Cores	RAM	Local Disk
Volta V100	NVLink	1	4	2 x 14	500 GB	2 TB
Volta V100	PCIe	3	2	2 x 14	500 GB	2 TB

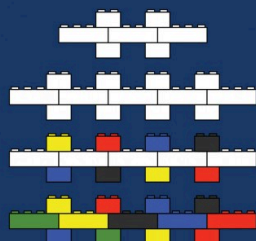


Jeremy Kepner

Mathematics of Big Data

Spreadsheets, Databases, Matrices, and Graphs

Jeremy Kepner and Hayden Jananthan



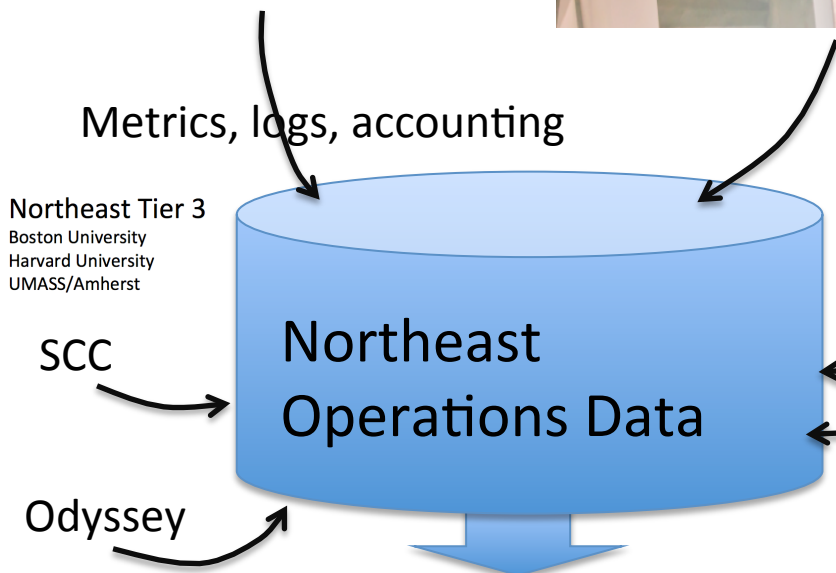
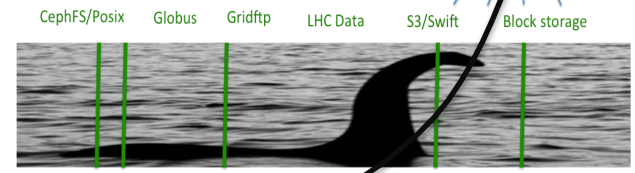
Foreword by Charles E. Leiserson

MIT LINCOLN LABORATORY SERIES

* Coming soon!

MASS OPEN CLOUD

An Open Cloud Exchange Public Cloud



Northeast Tier 3
Boston University
Harvard University
UMASS/Amherst

Metrics, logs,
accounting

- Save accounting and analytics for everything together
- Generate “control rooms”
- Starting point for Data Science and AI projects
- Cybersecurity hub
- Helps to establish trust, excitement, student involvement



BU, HU, MIT, NEU,
UMASS, MGHPCC

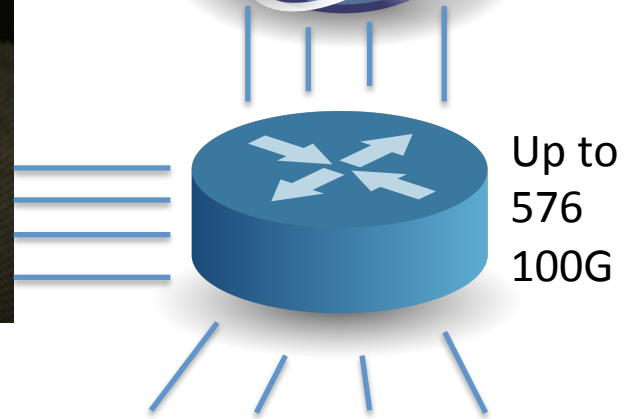


All collaborators get to make and use a “control room”!

Inspiration for Students, Visitors, Data Science, Curation, histories, accounting, data shared ²³



Internet2, Internet1
ESNet, LHCONE,...



CephFS/Posix

Globus

Gridftp

LHC Data

S3/Swift

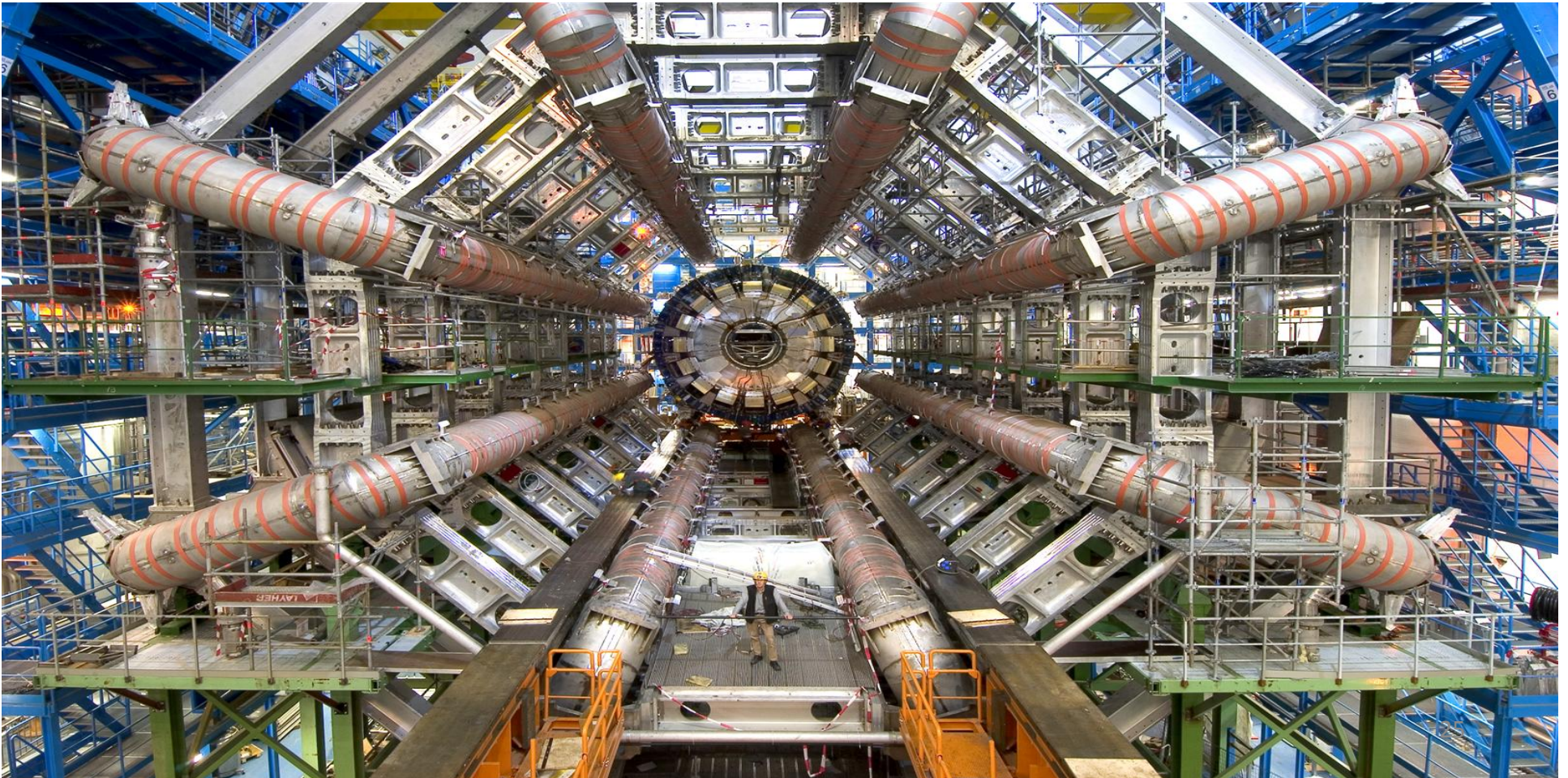
Block storage



University buy-in, Project buy-in, NESE project funds, Inherited Equipment

Why this is going to work

- NET2 – the U.S. ATLAS Northeast Tier 2 Center, one of four Tier 2 centers in the U.S. for the ATLAS LHC Experiment at CERN
 - About 6PB of GPFS storage and 10,000 cores, 100G to internet2 & ESnet. Joint BU/Harvard project.
 - Spends about \$150K per year on storage hardware.
 - [Storage in AWS would cost more than \\$1M per year instead.](#)
 - It wouldn't work anyway because we need ~200Gb/s between the storage and the computing fabric. Easy on the MGHPCC floor, very expensive otherwise.



Starter NESE OSD deployment

1 U

As dense as a 60 bay JBOD

12 x 10TB 7200 rpm spinning drives

Main storage

4 x 480GB Micron high endurance SSD

Bluestore Ceph database + CephFS metadata

1 x 32GB NVMe

Ceph write-ahead log

2 x 10Gb/s

Matched to the drives

1 x Single Socket Intel Xeon Skylake 2.2 Ghz (3.0 Ghz Turbo), 10C/20T

...with AVX-512 SIMD coprocessor

Plenty of CPU for 12 OSD unit

12 PB Raw, 40% buy-in

Summary:

Northeast Projects that Might Interest You

1. MGHPCC
2. Northeast CyberTeam (NSF funded)
3. Northeast perfSonar Mesh (Shawn helped with this)
4. Massachusetts Life Sciences Center
5. Northeast Tier 2 Center (NSF, analogous to Shawn's AGLT2)
6. Northeast Storage Exchange [NESE] (NSF, analogous to Shawn's OSiRiS)
7. Open Storage Network (NSF, John Goodhue co-PI)
8. Massachusetts Open Cloud
9. MIT Supercloud
10. HPC Futures

HPC Futures

Hyatt Regency Cambridge
575 Memorial Drive
Cambridge, MA 02139
June 30, 2017

Results, ideas, and planning for computing at Boston area universities, institutes, hospitals and companies

topics

Clinical Data Science
Computational Science at Harvard
Computer Science Research
Cybersecurity
Deep Learning and AI
Gravitational Wave Detection
Habitable Zone Exoplanets
Genomic Research
Juno Mission to Jupiter
Neurobiology
Quantitative Social Science

speakers

David Coker	Boston University
David Cox	Harvard University
Merce Crosas	Harvard University
James Cuff	Harvard University
Jason Dittmann	MIT
Mark Hamilton	NVIDIA
Erik Katsavounidis	MIT
Orran Krieger	Boston University
John Manfredelli	Northeastern University
Mark Michalski	Massachusetts General Hospital
Stephen Wolfram	Wolfram Research

program committee

Azer Bestavros	Boston University
Glenn Bresnahan	Boston University
John Goodhue	MGHPCC
Christopher Hill	MIT
Plamen Ivanov	Boston University
David Kaeli	Northeastern University
Rajiv Shridhar	Northeastern University
Paul Whitford	Northeastern University
Saul Youssef	Boston University



Thanks for listening.
Let's stay in touch 😊

