Interactive On-Demand HPC at Lincoln Laboratory and Beyond

Durham HPC Days Thursday Keynote

Albert Reuther, Ph.D., MBA

18-June-2026

MIT LINCOLN LABORATORY
SUPERCOMPUTING CENTER

Early Days of Supercomputing

- Very expensive
- Very capable (for the day)
- Lots of demand for precious computing time

- The solution: Develop resource management system with:
  - Queues
  - Priorities
  - Reservations
  - Accounting of CPU hour allocations
  - Metric: System Utilization (>98%)
- Much like scheduling telescope time

LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

MIT Lincoln Laboratory

Department of Defense Federally Funded Research and Development Center

Massachusetts Institute of Technology

MIT Lincoln Laboratory

Mission: Technology in Support of National Security

Key Roles:
- System architecture engineering
- Long-term technology development
- System prototyping and demonstration

FY18 Employees: 3995

FY18 Funding: $1.05B

Facilities: 2.1M sq-ft

Mission Areas:
- Air and Missile Defense
- Homeland Protection
- Air Traffic Control
- Communication Systems
- Cyber Security
- Advanced Technology
- Space Control
- ISR Systems and Technology
- Tactical Systems
- Engineering
- Biotechnology & Human Systems

Lincoln Laboratory Supercomputing Center (LLSC) Role

Mission Areas

- Air and Missile Defense
- Homeland Protection
- Air Traffic Control
- Communication Systems
- Cyber Security
- Advanced Technology
- Space Control
- ISR Systems and Technology
- Tactical Systems
- Engineering
- Biotechnology & Human Systems

Vast Data & Computation

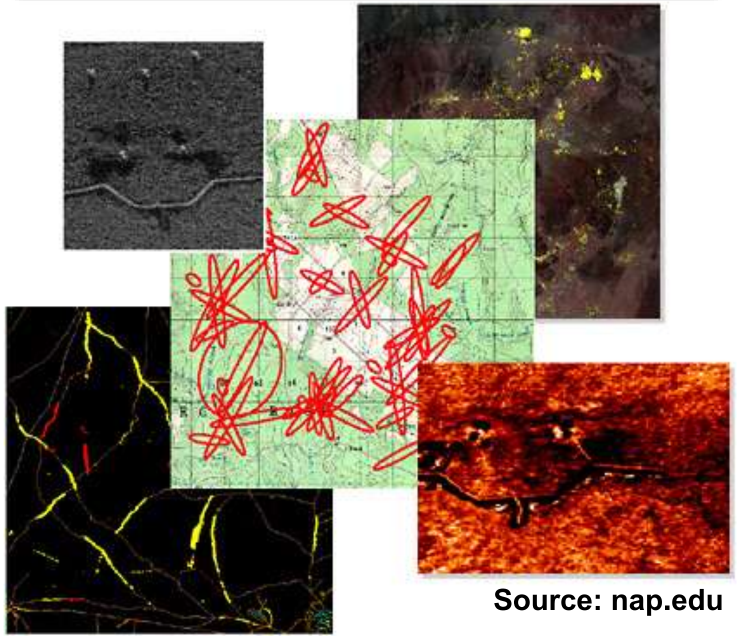
OSINT | Weather | HUMINT | C2 | Ground | Maritime | Air | Space | Cyber

LLSC develops & deploys unique, energy-efficient supercomputing that provides cross-mission

- Data centers, hardware, software, user support, and pioneering research
- 100x more productive than standard supercomputing[1]
- 100x more performance than standard cloud[2]

[1] *Interactive Grid Computing at Lincoln Laboratory*, Bliss et al., LL Journal, 2006
[2] *Scalability! But at what COST?*, McSherry et al., HotOS XV, 2015

Supercomputing Application Types

Physical Modeling and Simulation	Signal and Image Processing	Machine Learning and Data Analysis
		
Source: olcf.ornl.gov	Source: nap.edu	Source: cooperenvironmental.com
• Molecular Dynamics • Finite Element Analysis • Computational Fluid Dynamics • Multi-Physics Modeling • Weather Simulation • Etc.	• Radar (GMTI, SAR, ISAR) • Sonar • Wide Area Motion Tracking • Signal Intelligence (SIGINT) • Electronic Warfare • Etc.	• Target Tracking • Phenomenology Detection • Visualization • Data Clustering • Graph/Network Analysis • Etc.

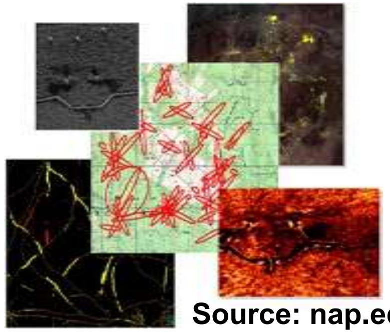
LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Supercomputing Application Types

Physical Modeling and Simulation

Source: olcf.ornl.gov

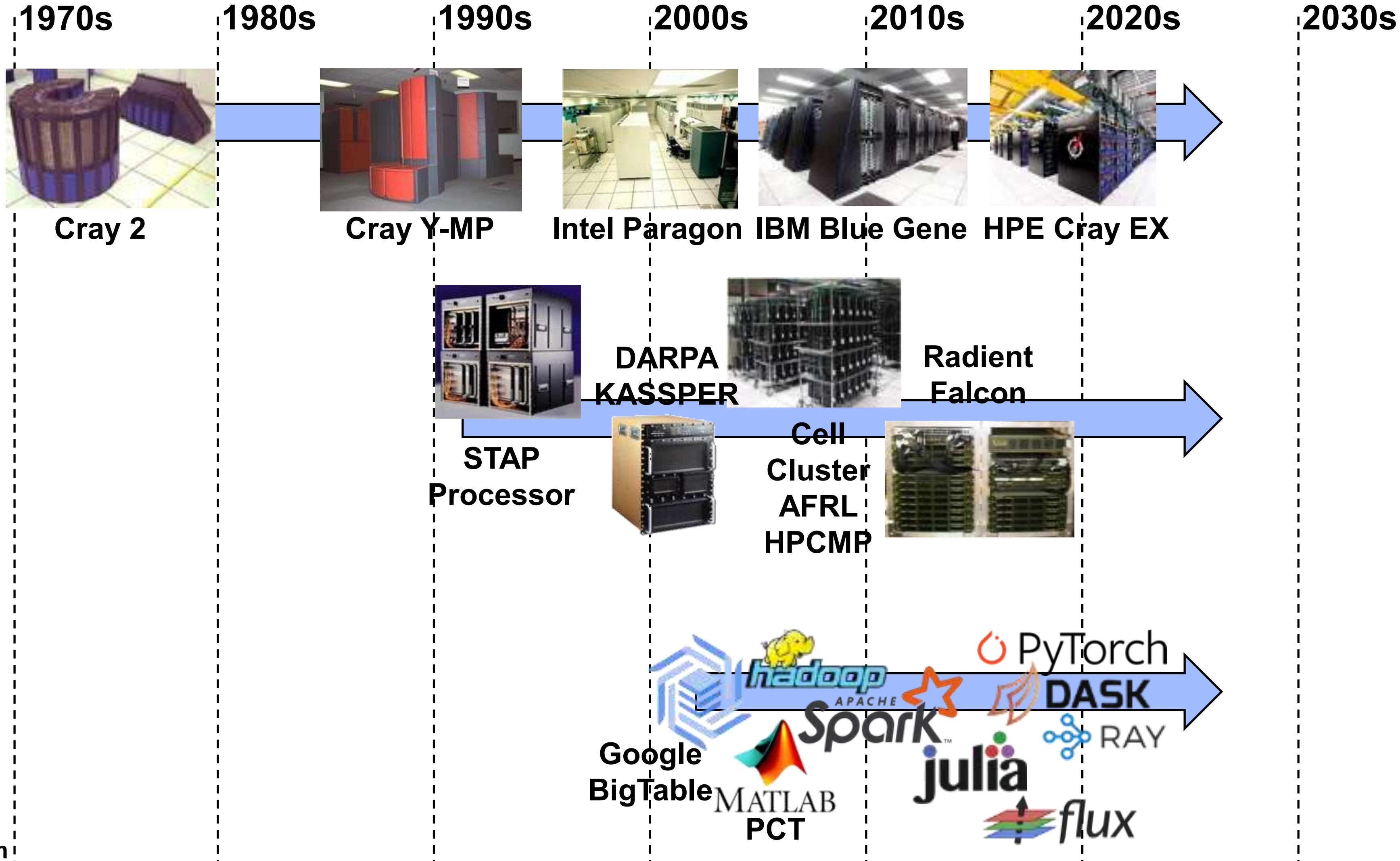
Signal and Image Processing



Source: nap.edu

Machine Learning and Data Analysis

Source: cooperenvironmental.com



STAP = Space-Time Adaptive Processing
PCT = MATLAB Parallel Computing Toolbox

Lincoln Supercomputing Software

- Focused on Enabling Advanced Research -

High Productivity Languages & Environments

python, MATLAB, SIMULINK, jupyterlab, julia, GNU Octave, R

AI Frameworks

PYTORCH, DASK, TensorBoard, TensorFlow, RAY, flux, APACHE Spark, Mamba, PIP, CONDA

3rd Party / Commercial Software

QUANTUMESPRESSO, FEKO, LS-DYNA, Gaussian, Inc., Cart3D, STAR-CCM+, Modtran®, lumerical, NCBI BLAST, WRF Weather Research & Forecasting, SIMULIA ABAQUS, ParaView

High Performance Languages, Compilers, Debuggers, Performance Analysis

Fortran, C, C++, VS Code, intel oneAPI HPC TOOLKIT, NVIDIA HPC, MPI, GCC, oneAPI, NVIDIA CUDA, NVIDIA Nsight Compute

High Performance Databases

OneSparse, PostgreSQL, DB OS, accumulo

Government Simulation Software

CREATE Computational Research and Engineering Acquisition Tools and Environments, AURORA, Kestrel, SENTRi, AFSIM, CAPSTONE

… and much more user installed software

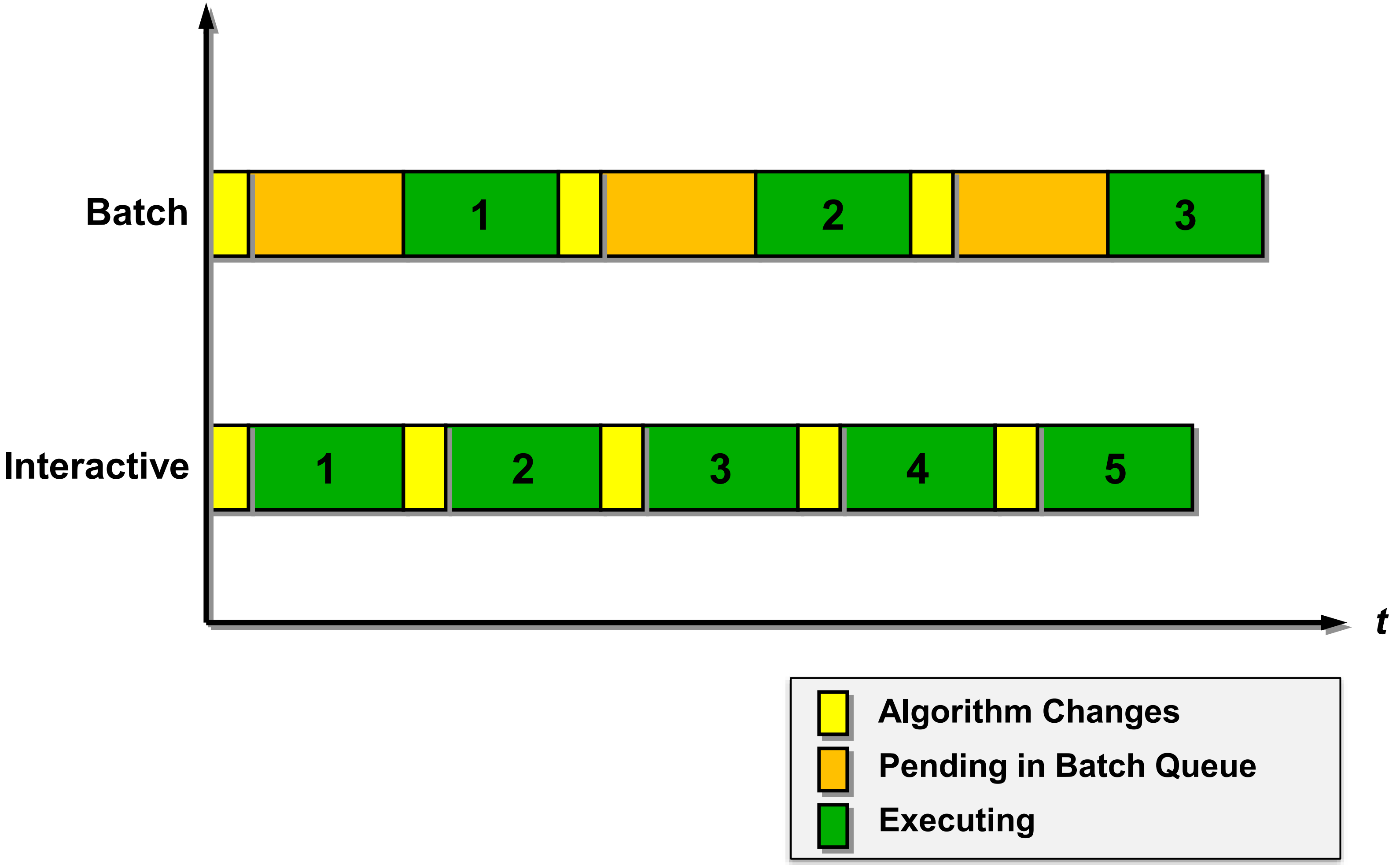
Sketchpad and Interactive Computing

Ivan Sutherland is shown working with Sketchpad on Lincoln TX-2 (1963).
First interactive graphical user computer interface.
https://archive.ll.mit.edu/publications/labnotes/LookingBack_19_1.pdf

We usually use computers in an interactive, on-demand manner

Batch vs. Interactive Development Cycles

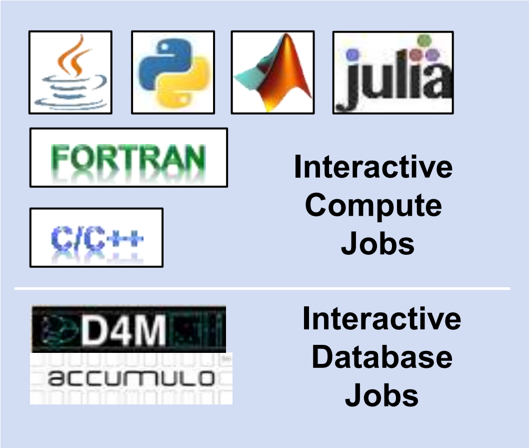
More development turns using interactive

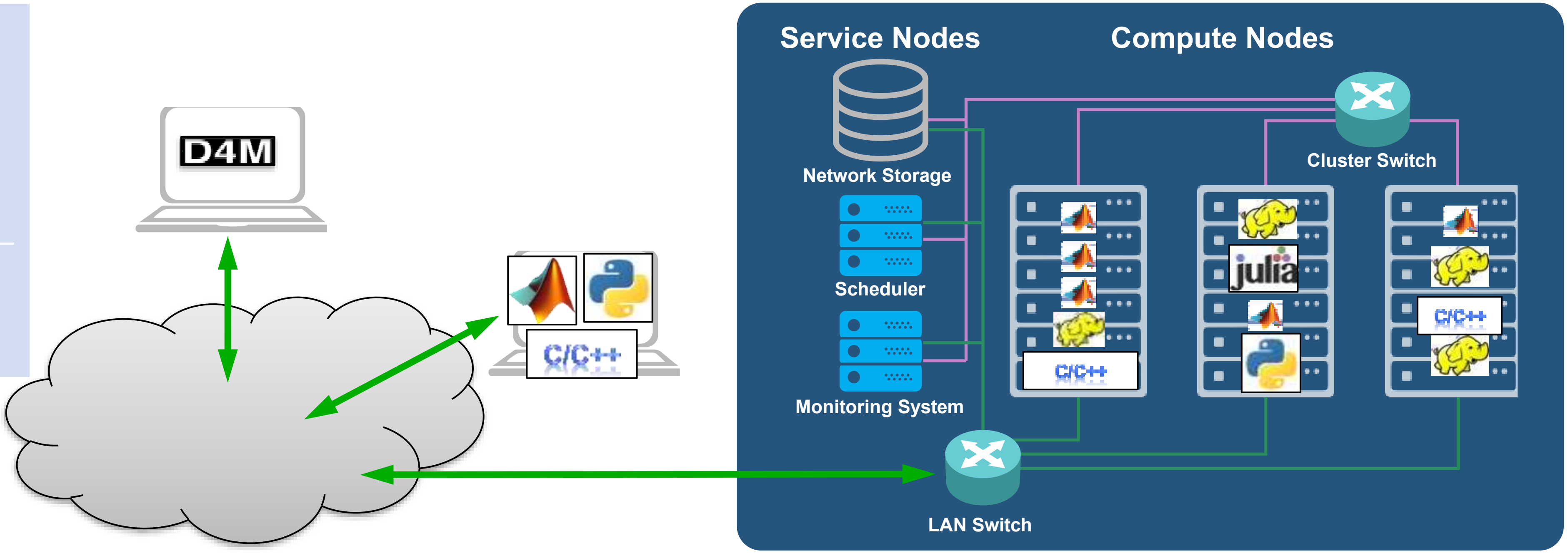


Interactive Best for:

- Rapid Prototyping
  - Algorithm development
  - Data analysis
  - Machine learning training
- Application Steering
  - Real-time / streaming data analytics
  - Debugging/validation
- Visualization

Interactive HPC Enables Broader Userbase



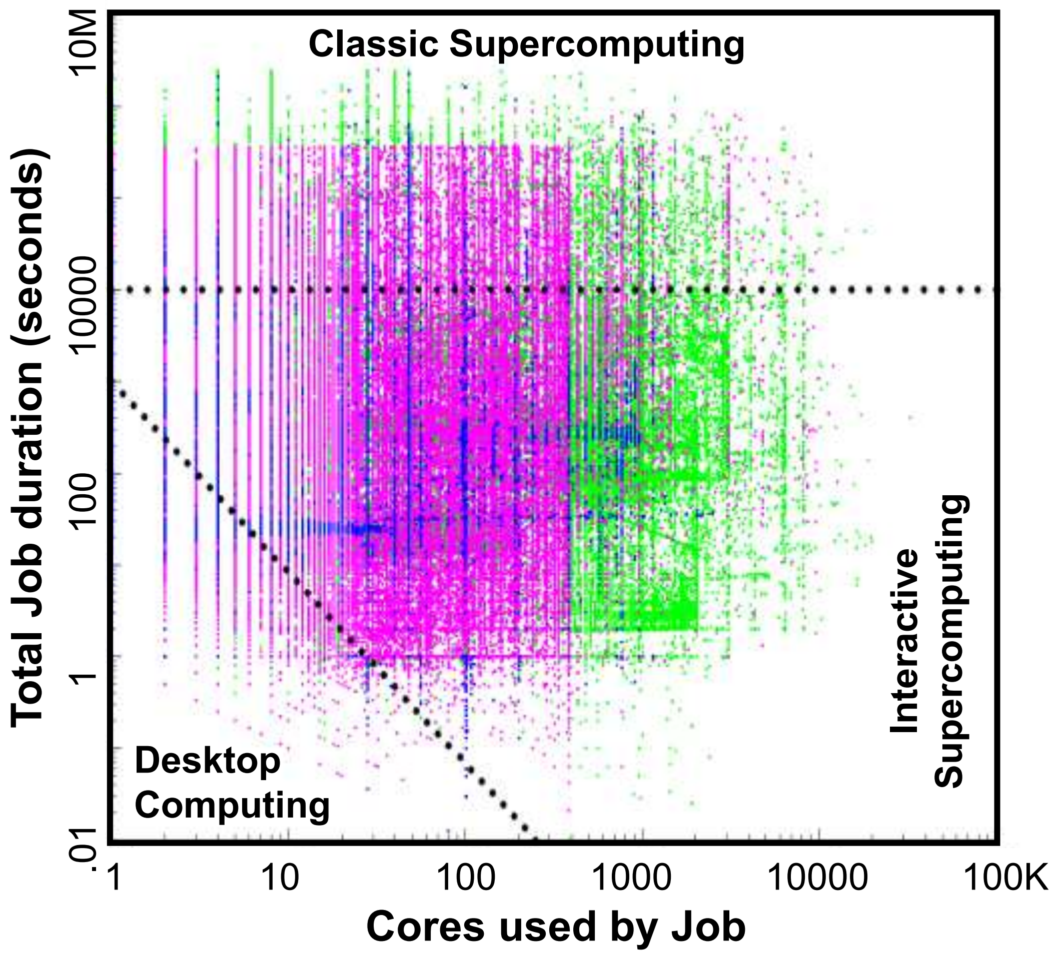


- LLSC provides a software platform that allows users to
  - Launch interactive compute jobs from their desktop
  - Share large volumes of project data
- The LLSC experience provides
  - Reference datasets pre-positioned in databases
  - Software modules and training to reduce user ramp up time
- LLSC responsible for cybersecurity across entire system

- Rapid Prototyping
  - Algorithm development
  - Data analysis
  - Machine learning training
- Application Steering
  - Real-time / streaming data analytics
  - Debugging/validation
- Visualization

N. T. Bliss, R. Bond, J. Kepner, H. Kim, and A. Reuther, “Interactive Grid Computing at Lincoln Laboratory,” *Lincoln Lab. J.*, vol. 16, no. 1, p. 165, 2006.

Userbase Enablement



- **Desktop Computing**
  - **CPU-time <20 minutes**
- **Classic Supercomputing**
  - **Wall-clock time >3 hours**
- **Interactive Supercomputing**
  - **Between desktop and classic supercomputing**
  - **Shortens the "time to insight"**
  - **Ten development turns/day instead of one turn/week**

● TX-Green (89020 Cores) ● TX-E1 (32000 Cores)
● TX-C (6740 Cores)

Note: Profile of jobs executed between April 1st, 2023 and March 31st, 2024.

Interactive HPC enables users to run jobs that don't fit on laptop/desktop

Outline

- Introduction and Motivation
- Scheduler Policy & Organizational Policy
- Tools for Interactive & Urgent Workloads
- Data Management
- User Support and Training
- Summary

Batch vs. Interactive Scheduling Trade-Offs

	All Batch	Batch with Reservations	Interactive with Limits	All Immediate Scheduling
Pro	• Jobs launches optimize system utilization	• Jobs launches maximize system utilization • Reservations specify timeslots to interact with job	• Launch most jobs when submitted • Limits promotes fairness • Accommodates both interactive and batch jobs	• Run all jobs immediately
Con	• Often high latency before execution	• Often high latency before execution	• Lower system utilization	• Frequent scheduler flooding • Lower system utilization

Mullen, J., Reuther, A., *et al.* (2018). Lessons Learned from a Decade of Providing Interactive, On-Demand High Performance Computing to Scientists and Engineers. In: Yokota, R., Weiland, M., Shalf, J., Alam, S. (eds) High Performance Computing. ISC High Performance 2018. Lecture Notes in Computer Science(), vol 11203. Springer, Cham. https://doi.org/10.1007/978-3-030-02465-9_47

LLGrid Productivity Analysis

$$\text{productivity (ROI)} = \frac{\text{time saved by users on system}}{\text{time to parallelize} + \text{time to train} + \text{time to launch} + \text{time to admin.} + \text{system cost}}$$

Production LLGrid model comparisons (2007)

Parameters	*Past 3 Years*	*Past Year (2007)*	*Next Year (2008)*
Lab-wide Users	201	201	251
New Users in Latest Year	81	81	50
Active Users in Latest Year	146	146	175
Simultaneous Jobs	10	16	25
Avg. CPUs per Job	16	32	64
Total Job Launches	137,588	95,525	125,000
Number of System Administrators	4	4	4
Nodes in System	592	592	852
New Nodes in System (Latest Year)	442	442	280
Benefit/Cost (Linear: CPUs)	24.10	43.05	144.62
Benefit/Cost (Logarithmic: $\log_2$(CPUs))	4.17	4.66	9.40

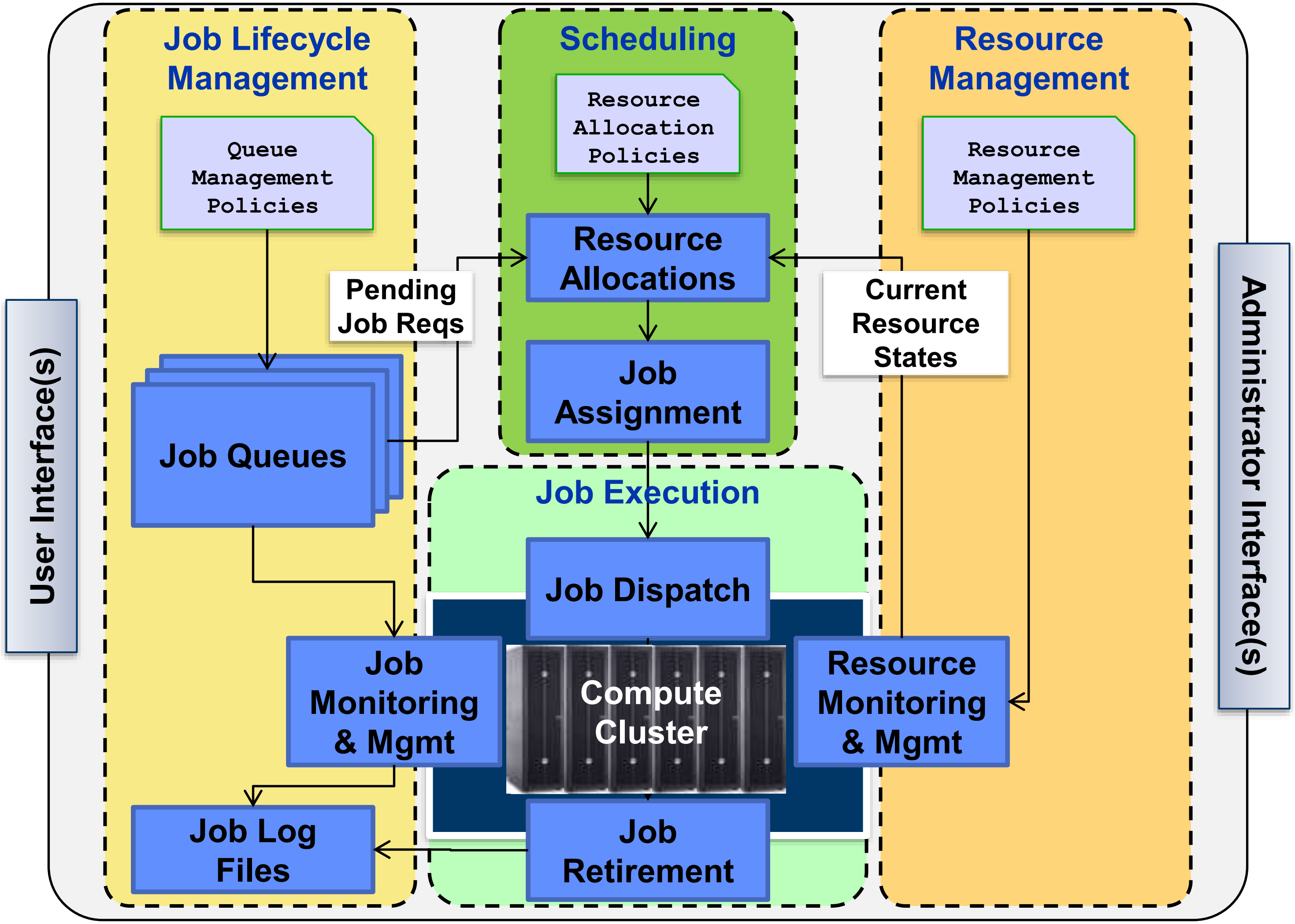
Reuther, A., Kepner, J., McCabe, A., Mullen, J., Bliss, N.T., Kim, H.: Technical challenges of supporting interactive HPC. In: Proceedings of the HPCMO Users Group Conference (UGC) 2007. IEEE, Pittsburgh, PA (June 2007)
Tichenor, S., Reuther, A.: Making the business case for high performance computing: A benefit-cost analysis methodology. CTWatch Quarterly 2(4A), 2–8 (November 2006)

Other Interactive and Urgent Example Scheduling Policies

- **Aggressive Oversubscription – Inst. of Science Tokyo**
  - Single queue for oversubscription
  - Multiple jobs per core/node
  - Memory shared among jobs on node
  - Gaining productivity, some performance slowdown
- **Balanced Interactivity – Tufts**
  - 6-hour execution window
  - Must checkpoint to run longer than 6 hours
  - Resubmission necessary to run from checkpoints
- **DoD HPCMP Scheduling**
  - Aiming for utilization between 60-80%
  - Managing queue expansion < 2
  - QE = (run_time + queue_time) / run_time
- **Perlmutter-on-Demand (NERSC)**
  - Experimented with Slurm vs Kubernetes for resource management
  - Simulated and analyzed policies for various degrees of urgency
  - Demonstrated on-demand urgent data analysis job execution for D-III Program

Resource Manager Architecture

- **Four Primary Tasks**
  - **Job Management**
  - **Resource Management**
  - **Scheduling**
  - **Job Execution**
- **Most activities executed on scheduler server**
- **Job monitoring and resource monitoring daemons run on each compute node**

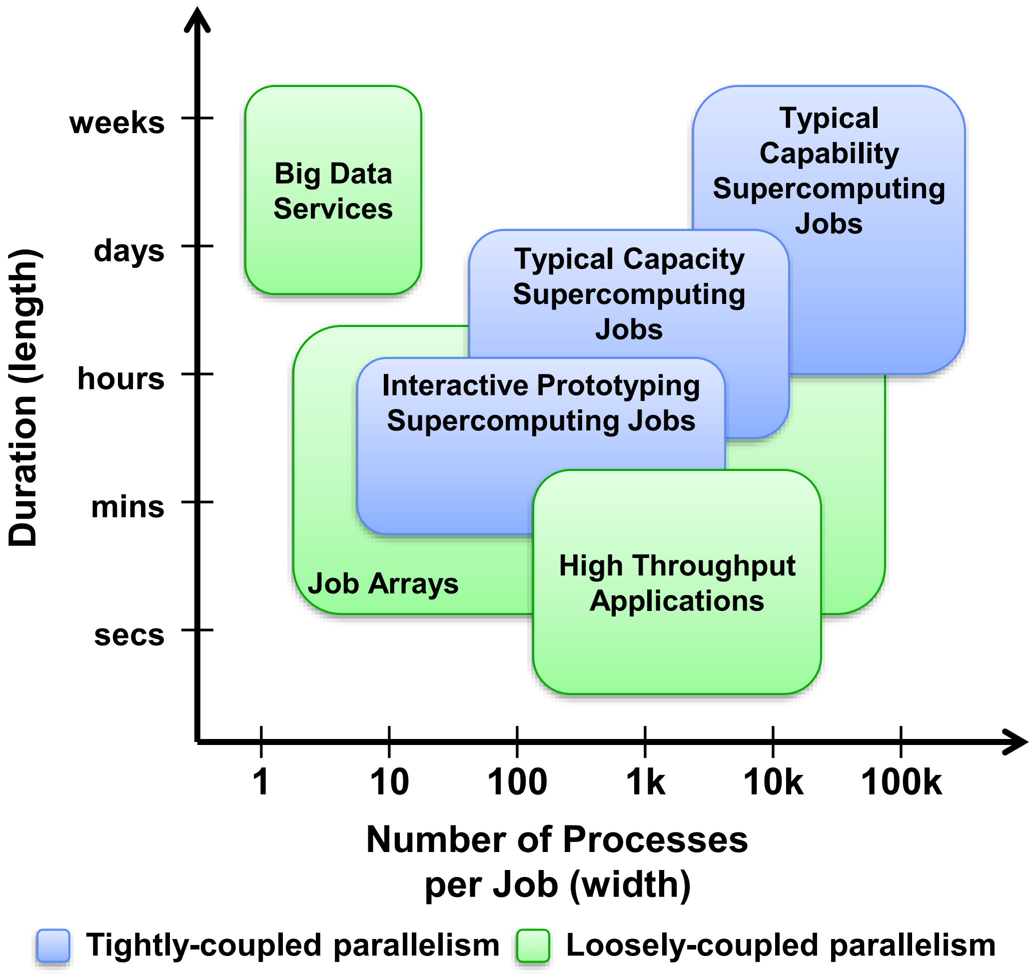


A. Reuther *et al.*, “Scalable system scheduling for HPC and big data,” *J. Parallel Distrib. Comput.*, vol. 111, 2018. DOI: 10.1016/j.jpdc.2017.06.00; arXiv:1705.03102
A. Reuther *et al.*, “Scheduler technologies in support of high performance data analysis,” in *2016 IEEE High Performance Extreme Computing (HPEC) Conference*, 2016. DOI: 10.1109/HPEC.2016.7761604; arXiv:1607.06544

Two Perspectives on Job Types

- **Traditional Supercomputing Paradigm**
  - Job: single process
  - Job Array: series of single (independent) processes as a grouping
  - Parallel Job: set of processes that are all launched simultaneously and communicate synchronously (MPI)

- **Big Data Paradigm**
  - High Throughput Immediate Jobs: Very short-run, time-critical independent jobs (μs to ms execution)
  - High Throughput Batch jobs: Independent, usually short but not so time-critical
  - Service jobs: Independent, persistently executing



A. Reuther *et al.*, “Scalable system scheduling for HPC and big data,” *J. Parallel Distrib. Comput.*, vol. 111, 2018. DOI: 10.1016/j.jpdc.2017.06.00; arXiv:1705.03102
A. Reuther *et al.*, “Scheduler technologies in support of high performance data analysis,” in *2016 IEEE High Performance Extreme Computing (HPEC) Conference*, 2016. DOI: 10.1109/HPEC.2016.7761604; arXiv:1607.06544

Resource Manager Tradeoffs: Schedulers vs Orchestrators

Traditional Supercomputing Schedulers

- Sophisticated priority and accounting mechanisms
- Accommodate batch, reservations, and interactive modes
  - Optimizes for high utilization in batch mode
  - Enables very fast launching in interactive mode
- Multiple user interfaces
  - Command line is primary interface
  - APIs for Python, Java, etc.
- Key Schedulers
  - Slurm
  - IBM LSF
  - PBS Pro, Torque, Moab

IBM Platform LSF

Big Data/Cloud Orchestrators

- Simple queueing system for low job dispatch latency
- Integrated support for containers (Docker, Singularity)
- Lack native support for synchronously parallel jobs (MPI)
- Multiple user interfaces
  - Programming API is primary interface
  - Command line interface
- Key Schedulers
  - Kubernetes
  - Apache Mesos
  - Hadoop YARN

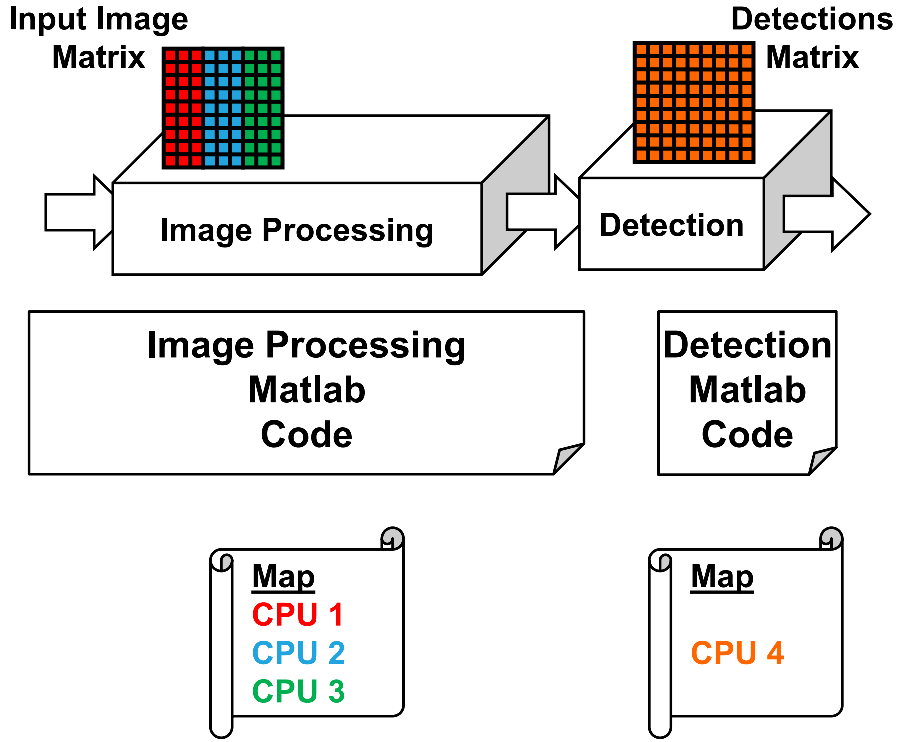
A. Reuther *et al.*, "Scalable system scheduling for HPC and big data," *J. Parallel Distrib. Comput.*, vol. 111, 2018. DOI: 10.1016/j.jpdc.2017.06.00; arXiv:1705.03102
A. Reuther *et al.*, "Scheduler technologies in support of high performance data analysis," in *2016 IEEE High Performance Extreme Computing (HPEC) Conference*, 2016. DOI: 10.1109/HPEC.2016.7761604; arXiv:1607.06544

Outline

- Introduction and Motivation
- Scheduler Policy & Organizational Policy
- Tools for Interactive & Urgent Workloads
- Data Management
- User Support and Training
- Summary

Interactive Parallel Matlab (pMatlab)

Algorithm



Parallel Remappable Software

Storage

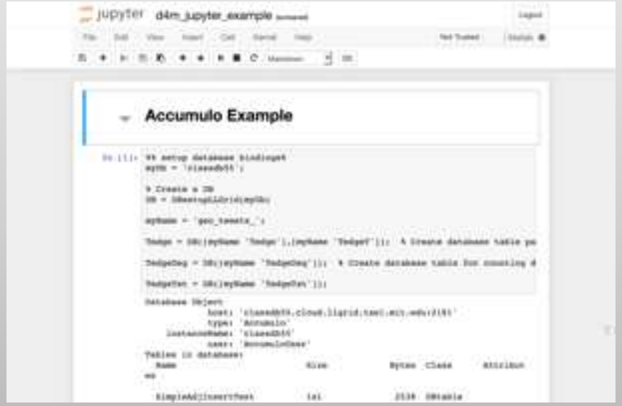
Supercomputer

- **pMatlab software makes Matlab code parallel and remappable**
- **Port Matlab codes in minutes to run (from their desktop) on supercomputer**
- **pMatlab transitioned to open source and industry and textbook (100K users)**

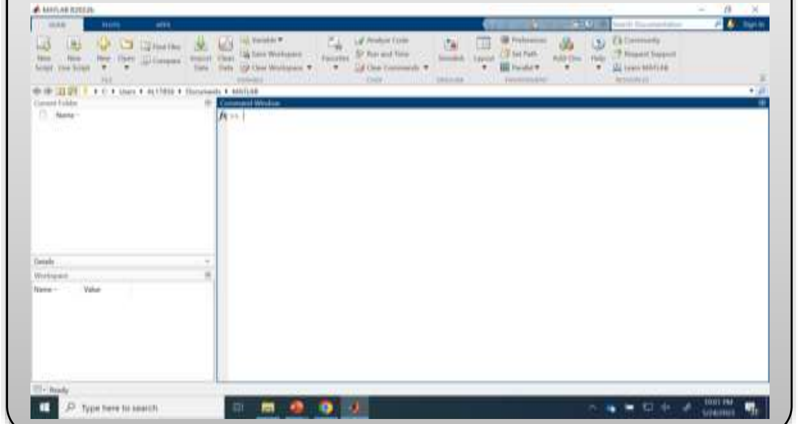
N. T. Bliss, J. Kepner, H. Kim and A. Reuther, "pMATLAB: Parallel MATLAB Library for Signal Processing Applications," ICASSP 2007, pp. IV-1189-IV-1192, doi: 10.1109/ICASSP.2007.367288.

LLSC Advantage: Interactive Supercomputing

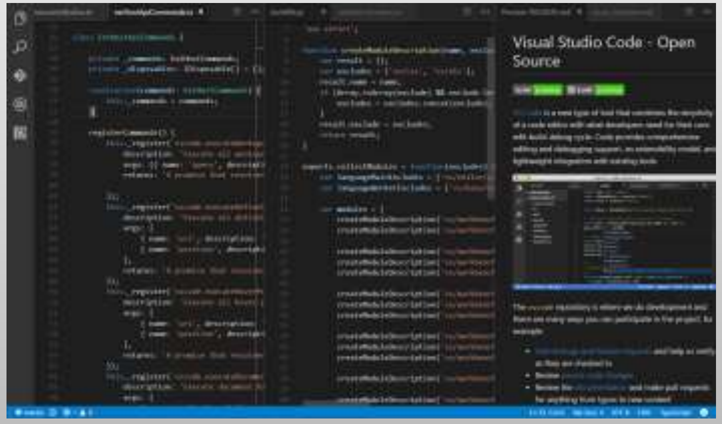
Jupyter Notebook: web-based IDE & more



Interactive MATLAB: N-1 nodes work behind the scenes of IDE



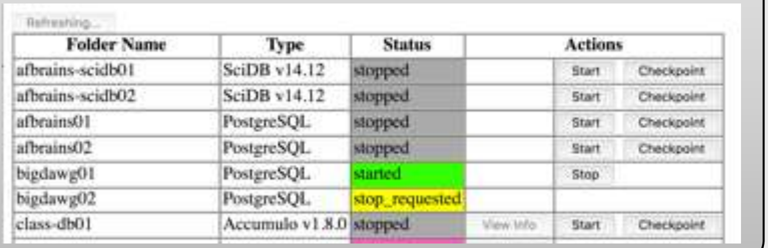
VS Code: Flexible IDE connects via ssh bridge and lightweight server



Interactive Hardware: get a processor core or a whole node

Dynamic Web Dashboards: forward through web gateway/portal

Dynamic Databases: manage world's most powerful databases from a GUI

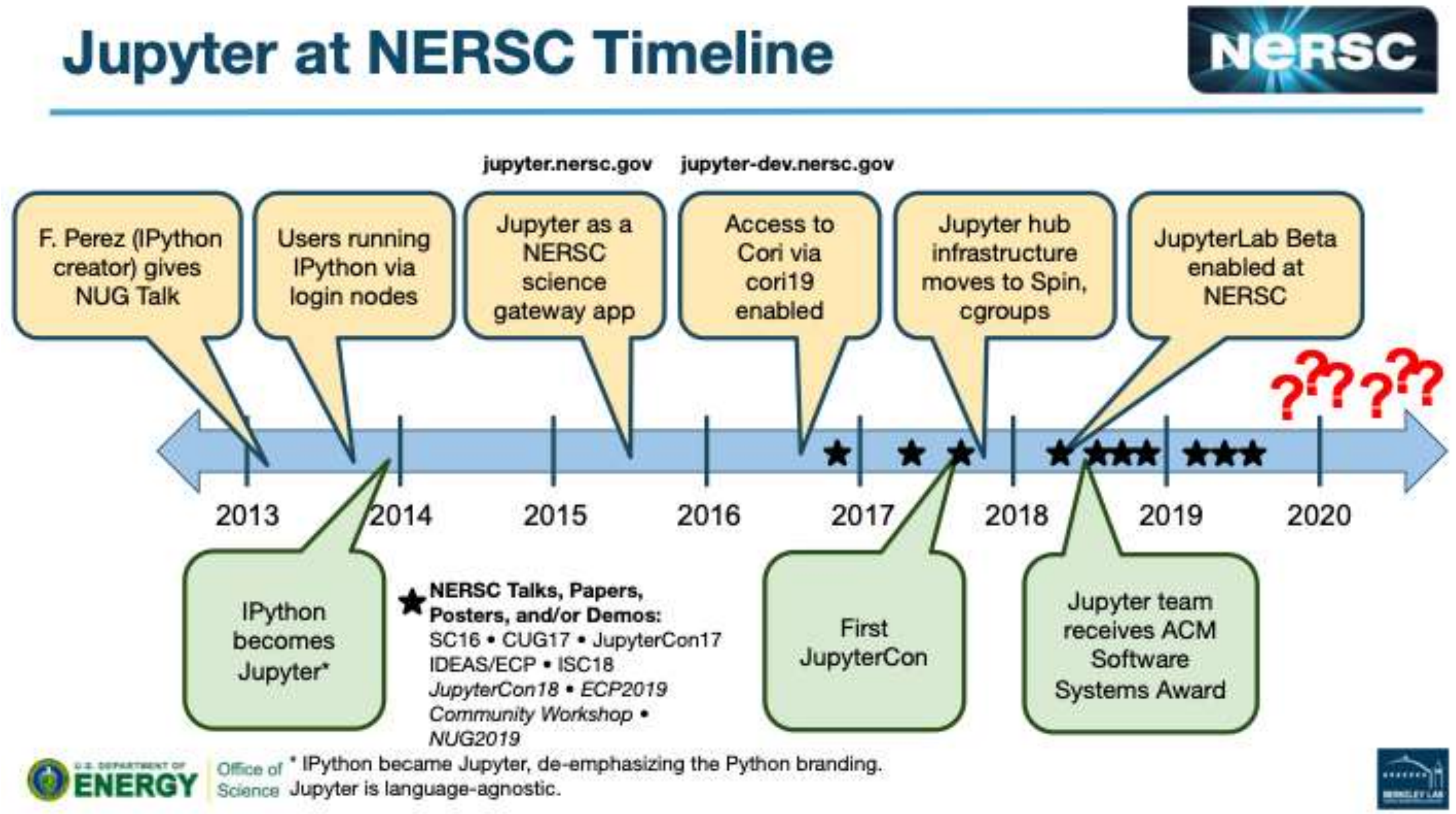


Refreshing...

Folder Name	Type	Status	Actions		
afbrains-scidb01	SciDB v14.12	stopped		Start	Checkpoint
afbrains-scidb02	SciDB v14.12	stopped		Start	Checkpoint
afbrains01	PostgreSQL	stopped		Start	Checkpoint
afbrains02	PostgreSQL	stopped		Start	Checkpoint
bigdawg01	PostgreSQL	started		Stop	
bigdawg02	PostgreSQL	stop_requested			
class-db01	Accumulo v1.8.0	stopped	View Info	Start	Checkpoint

A. Prout *et al.*, "MIT SuperCloud portal workspace: Enabling HPC web application deployment," *2017 IEEE High Performance Extreme Computing Conference (HPEC)*, Waltham, MA, USA, 2017, pp. 1-6, doi: 10.1109/HPEC.2017.8091097.

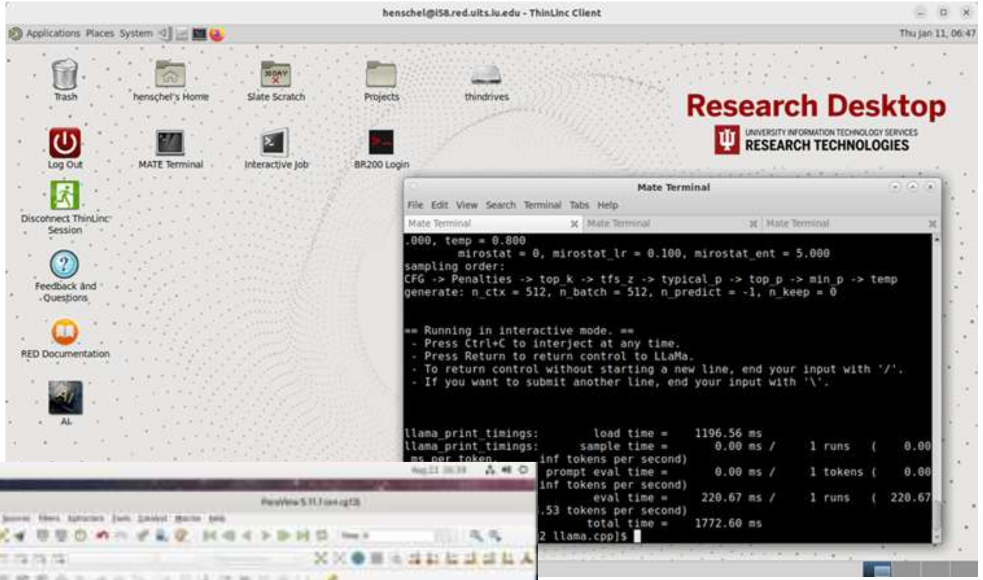
Jupyter on HPC: Developed at NERSC



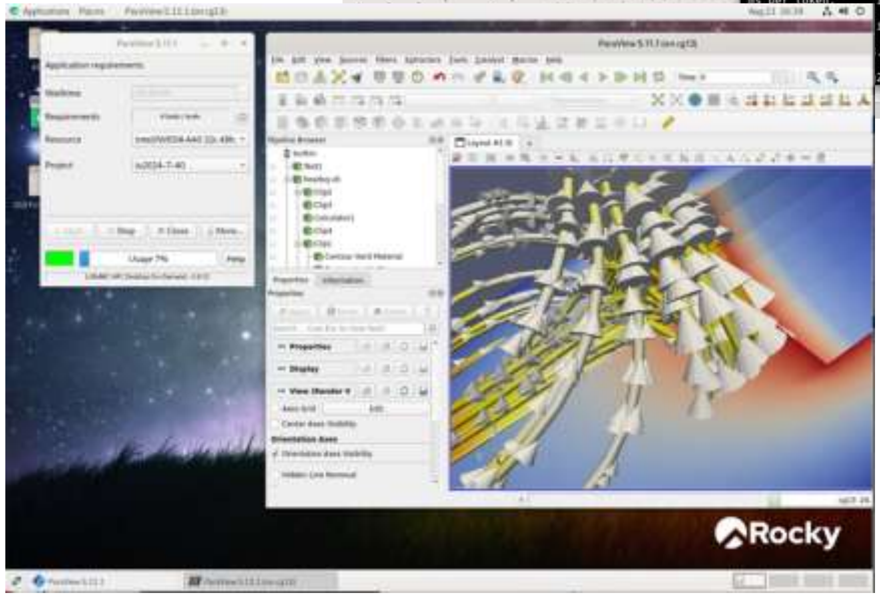
From: Rollin Thomas, “Jupyter: Redefining the User Interface to HPC,” NERSC User Group (NUG), July 2019.

Research Desktops

- Familiar desktop environment for students and researchers
- Many integrated tools
  - Graphic-heavy applications
  - Familiar UIs
  - Scheduler interactions
- Interface to scheduler or orchestrator



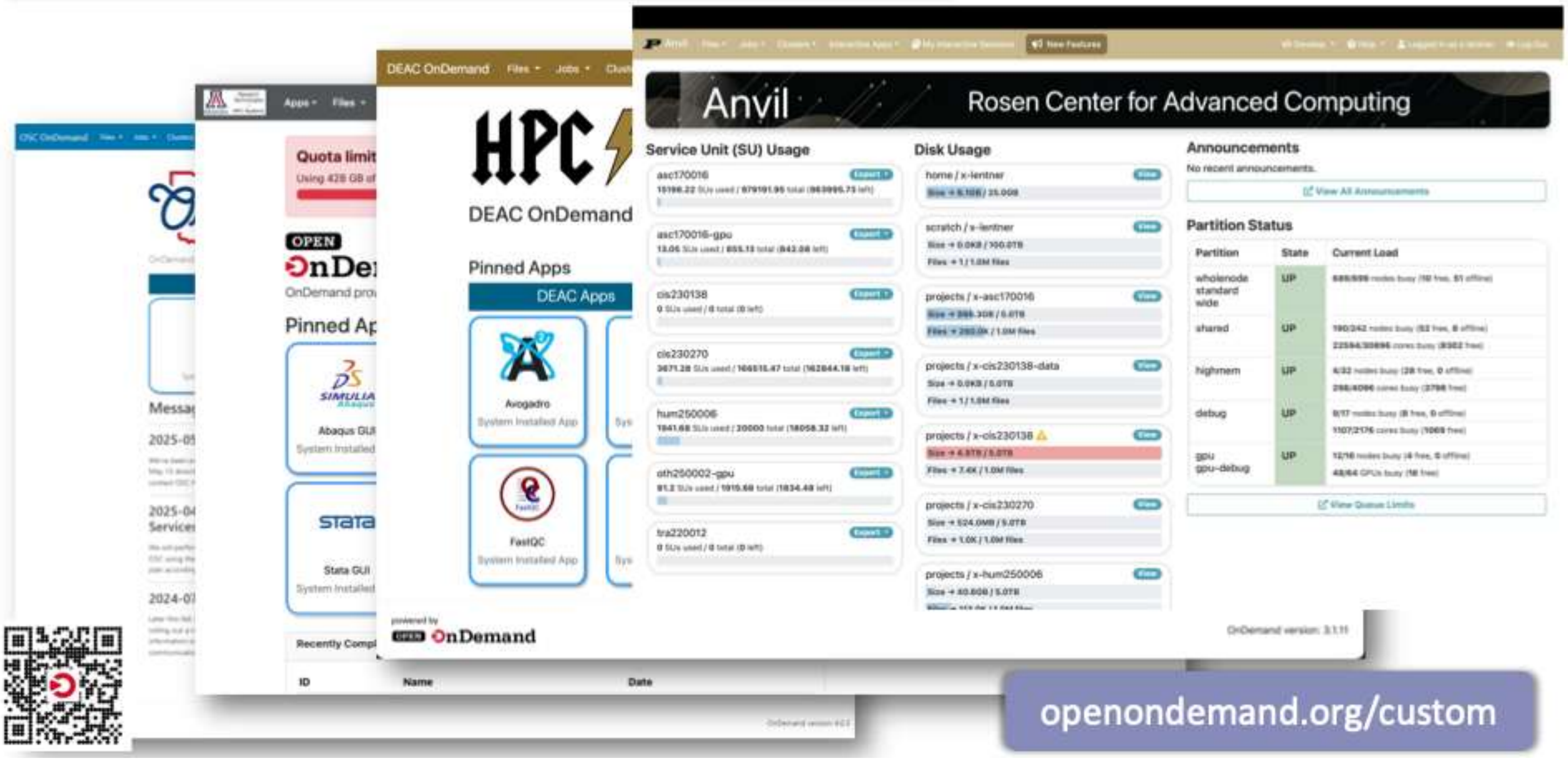
Research Desktop – Indiana Univ.



Lunarc Research Desktop – Lund Univ.

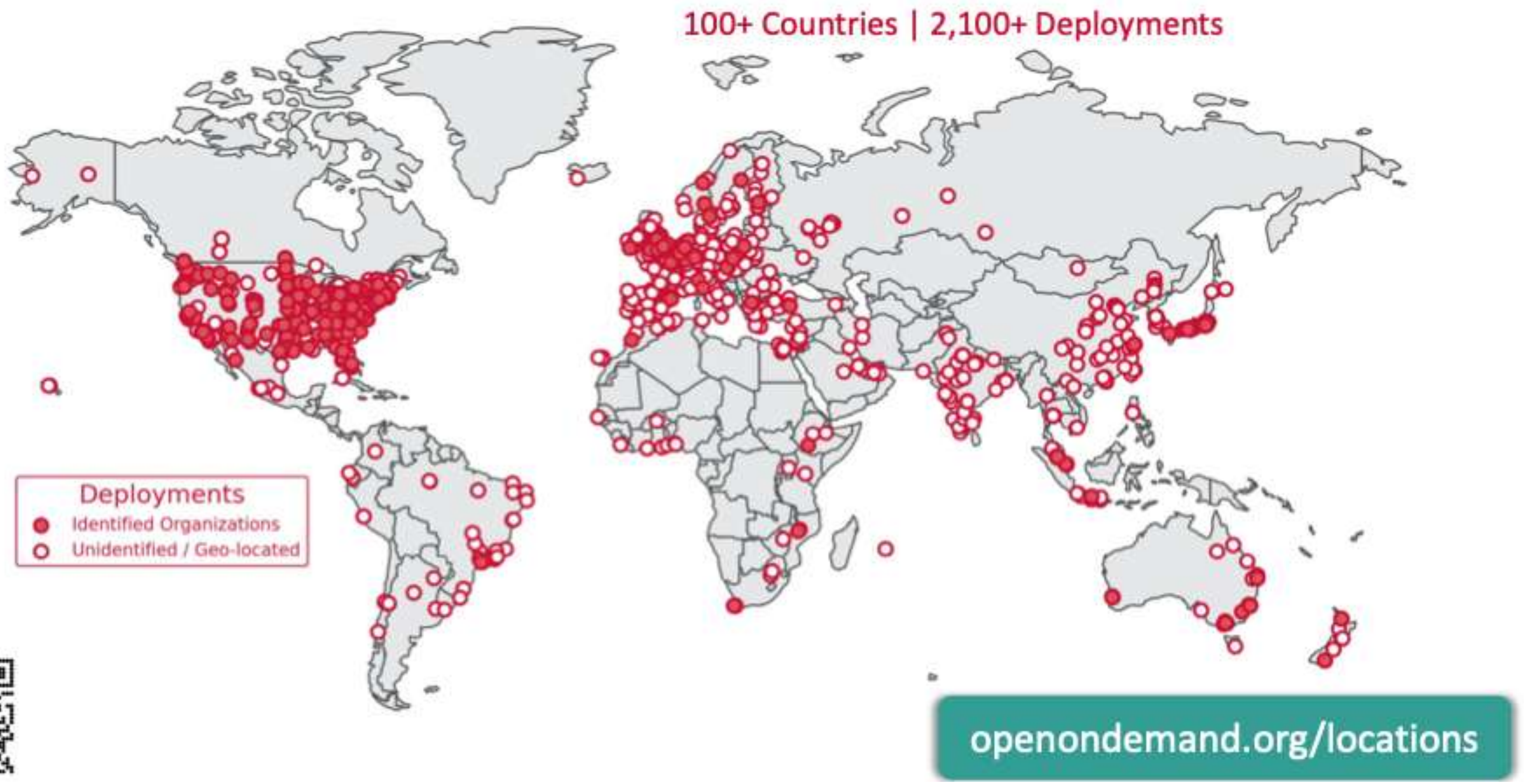
LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Web-based Portal: Open OnDemand



From: Alan Chalker, “Open OnDemand: Connecting Computing Power with Powerful Minds”, WIUHPC Featured Talk at SC25. St. Louis, Nov. 2025.

Open OnDemand Portal: Worldwide Adoption



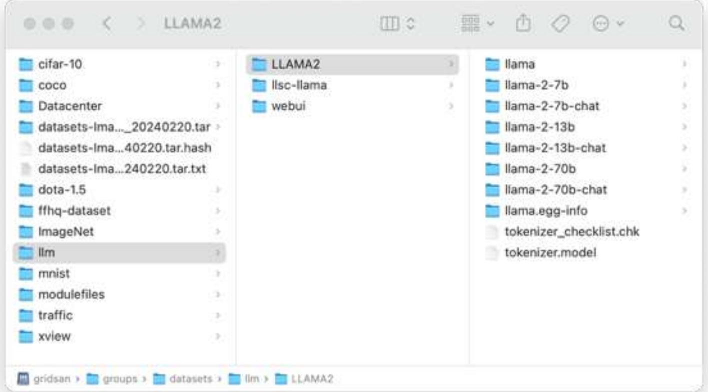
From: Alan Chalker, “Open OnDemand: Connecting Computing Power with Powerful Minds”, WIUHPC Featured Talk at SC25. St. Louis, Nov. 2025.

Outline

- Introduction and Motivation
- Scheduler Policy & Organizational Policy
- Tools for Interactive & Urgent Workloads
- Data Management
- User Support and Training
- Summary

Frequently Used Curated Datasets and Models available on LLSC Systems

LLM Models



ImageNet

14M images, 231 GB

Microsoft COCO

330K images, 18GB

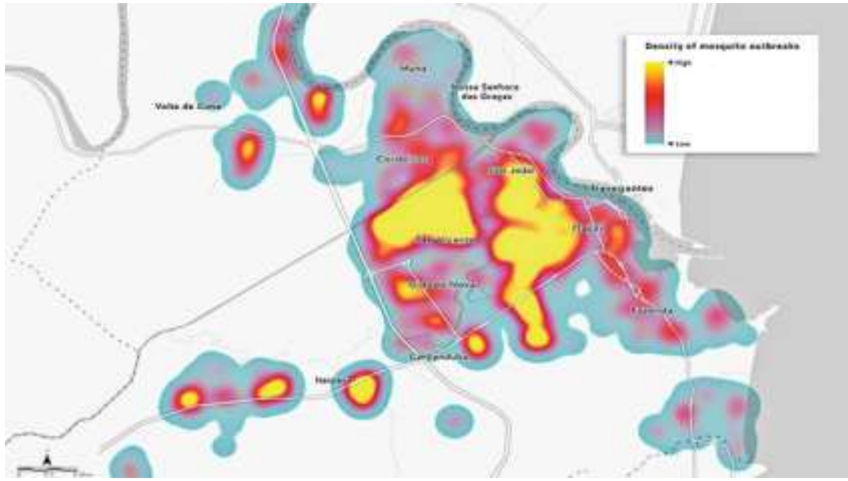
Moments in Time

1M videos, 10T

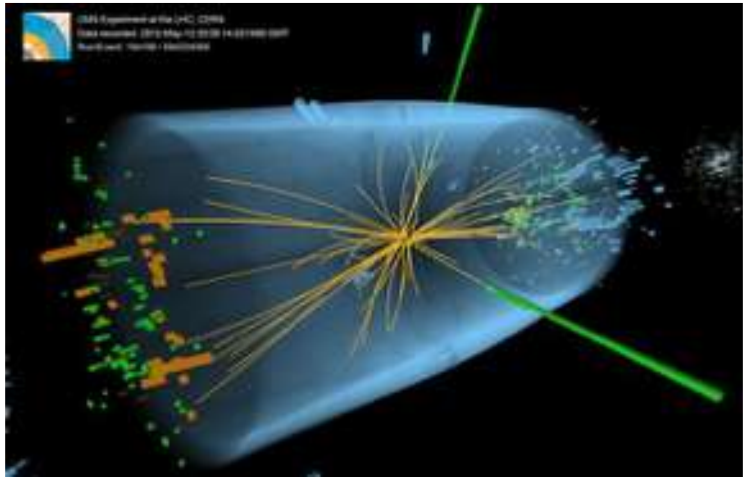
MIT Places Dataset

8M images, 600 GB

AI requires many large, diverse datasets, and many AI models are very large

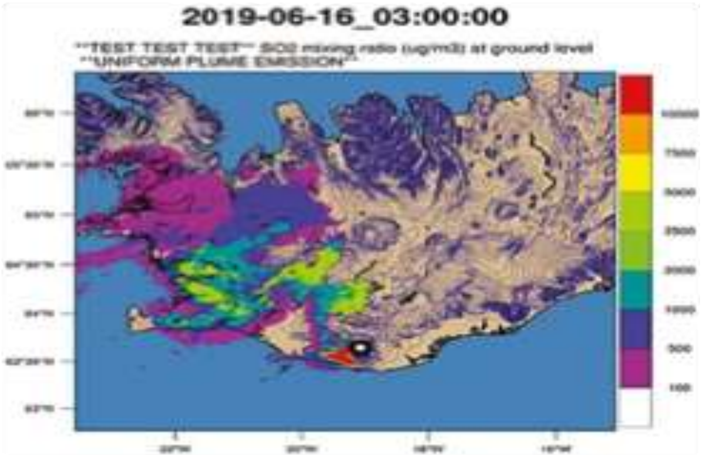


Mosquito Borne Illness Tracking – EPCC

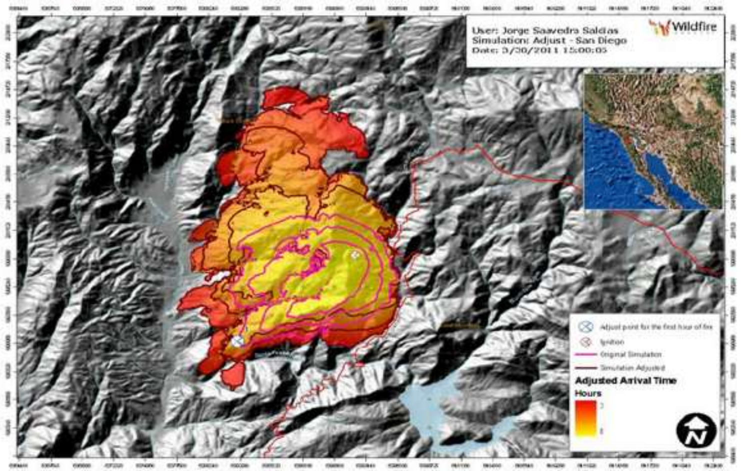


LHC Data Analysis – NERSC

Tsunami Prediction – Tohoku Univ.



Volcanic Gas Dispursion – Univ. of Leeds



Wildfire Risk Assessment – Technosylva

Gamma Ray Burst Localization – WUSTL

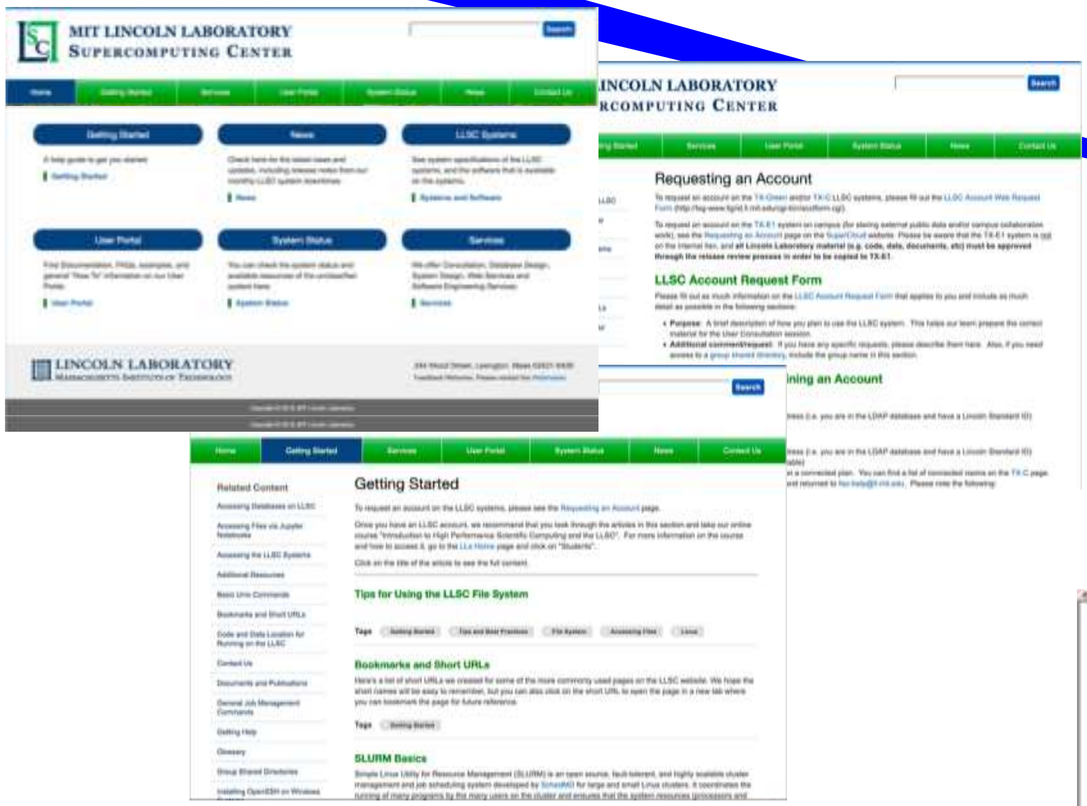
LHC = Large Hadron Collider
WUSTL = Washington Univ. at St. Louis

Outline

- Introduction and Motivation
- Scheduler Policy & Organizational Policy
- Tools for Interactive & Urgent Workloads
- Data Management
- User Support and Training
- Summary

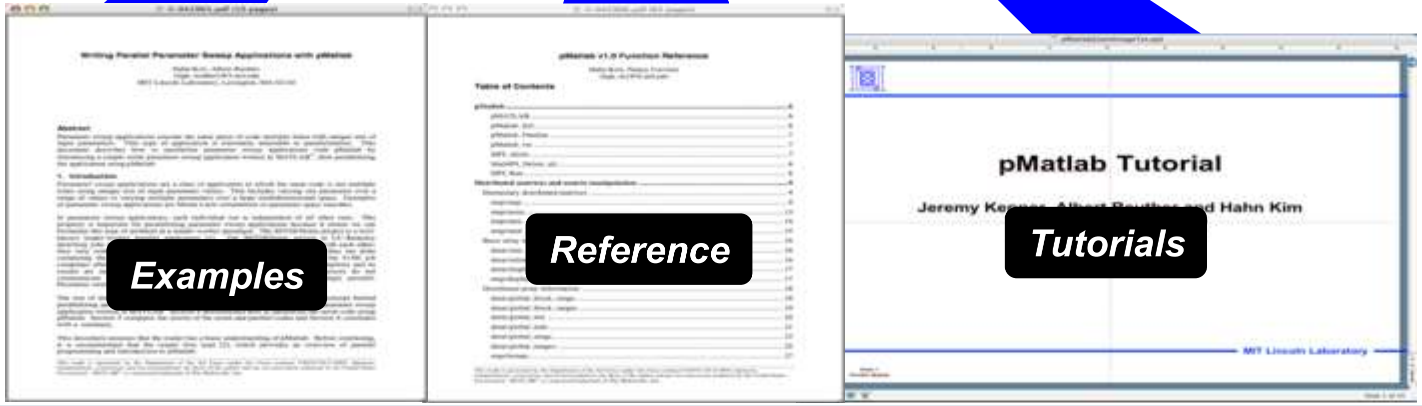
LLSC Team – Expert Consulting

http://short/LLSC



Wiki Site

Expert Consulting



Examples ***Reference*** ***Tutorials***

Mullen, J., Reuther, A., *et al.* (2018). Lessons Learned from a Decade of Providing Interactive, On-Demand High Performance Computing to Scientists and Engineers. In: Yokota, R., Weiland, M., Shalf, J., Alam, S. (eds) High Performance Computing. ISC High Performance 2018. Lecture Notes in Computer Science(), vol 11203. Springer, Cham. https://doi.org/10.1007/978-3-030-02465-9_47

Educational Resources

http://short/LLSC

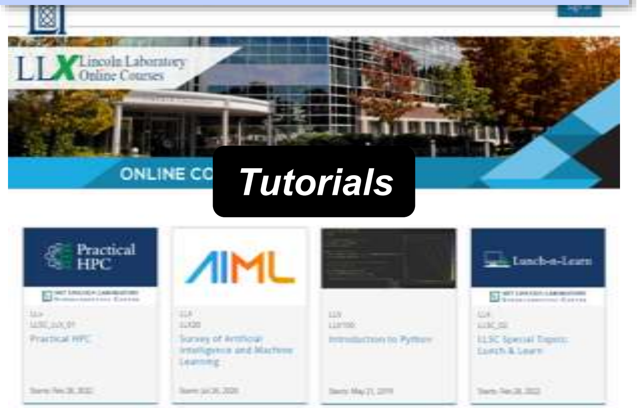
http://short/LLSC-consults
http://short/LLSC-OH-LnL

Consult Meetings

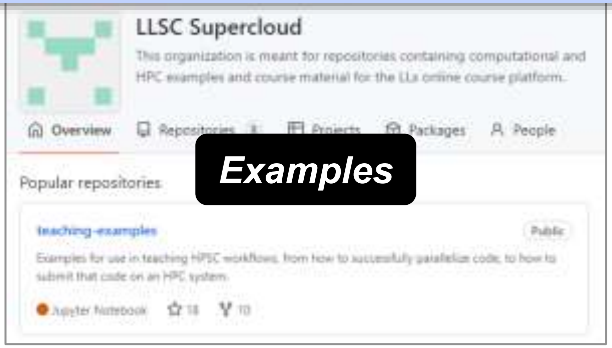
Lunch & Learn Sessions

Drop In Office Hours

http://short/LLX



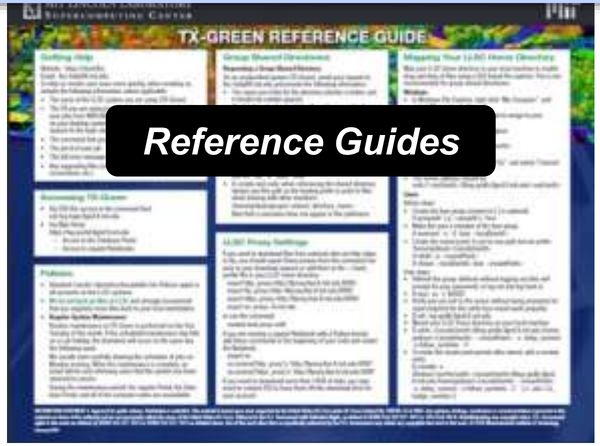
https://github.com/llsc-supercloud
/home/gridsan/groups/courses_shared
~/examples



LLSC-Help@ll.mit.edu (Lincoln)
SuperCloud@mit.edu (MIT campus)

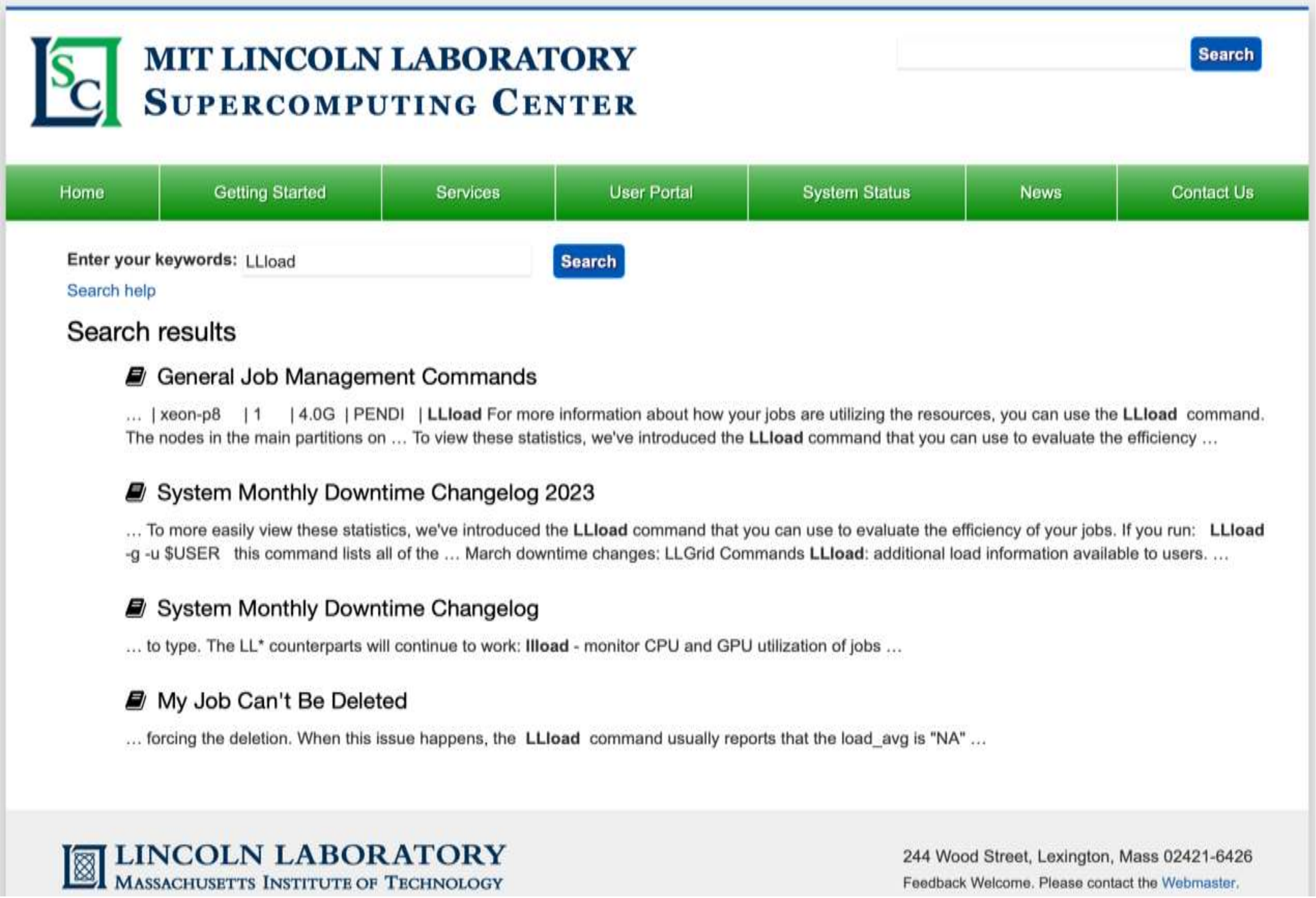
Email

http://short/LLSC-Ref-Guides

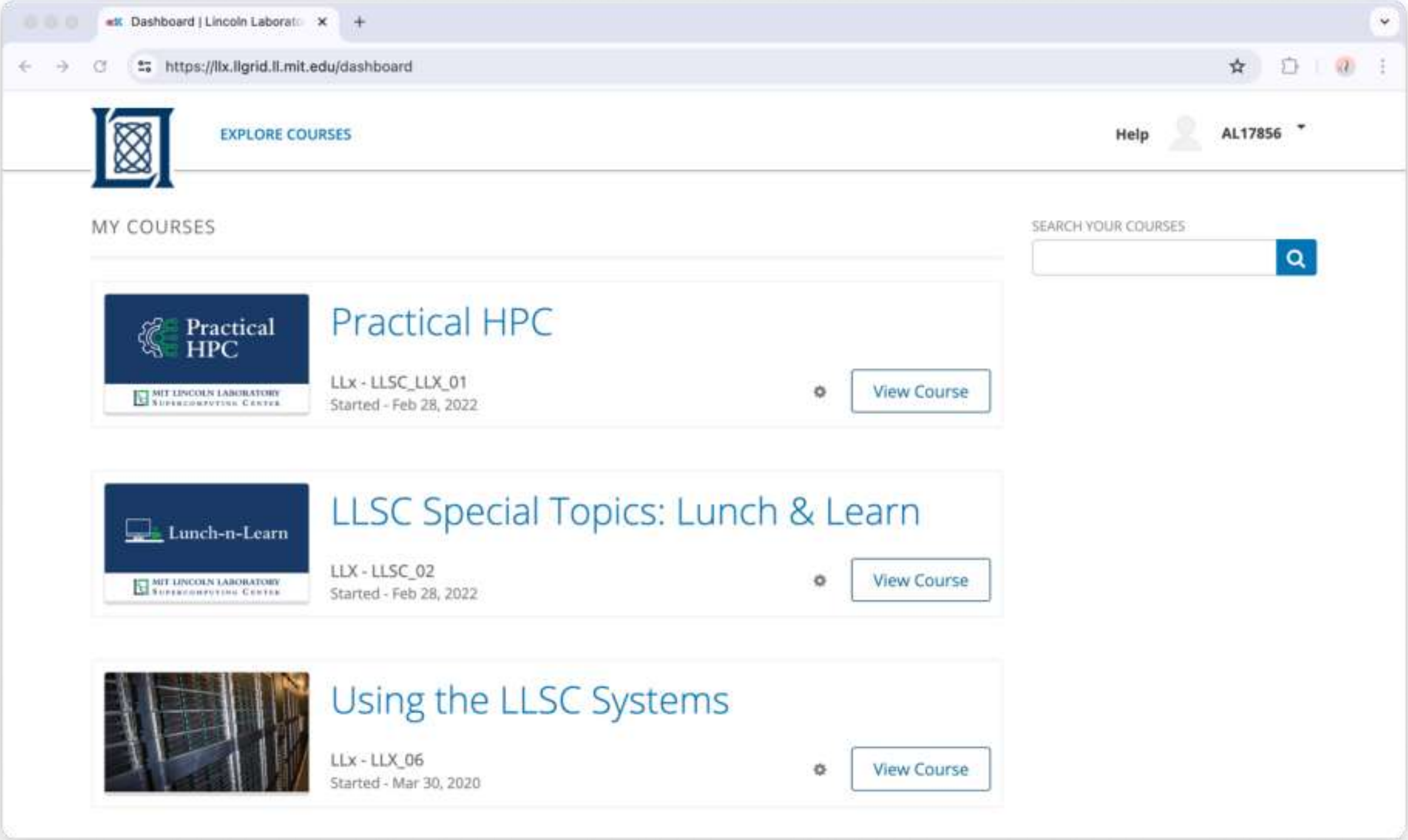


Educational Resource: Wiki Site (https://short/llsc)

- **Instructions on getting account and getting started**
- **Many pages on using LLSC systems**
- **Links to User Portal**
- **News about systems**
- **Very capable search functionality**



Educational Resources: Tutorials (https://short/LLX)



Outline

- Introduction and Motivation
- Scheduler Policy & Organizational Policy
- Tools for Interactive & Urgent Workloads
- Data Management
- User Support and Training
- Summary

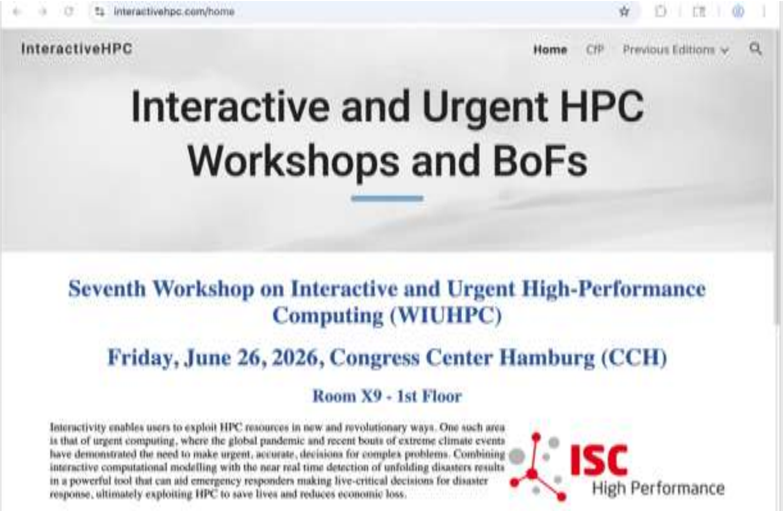
Summary

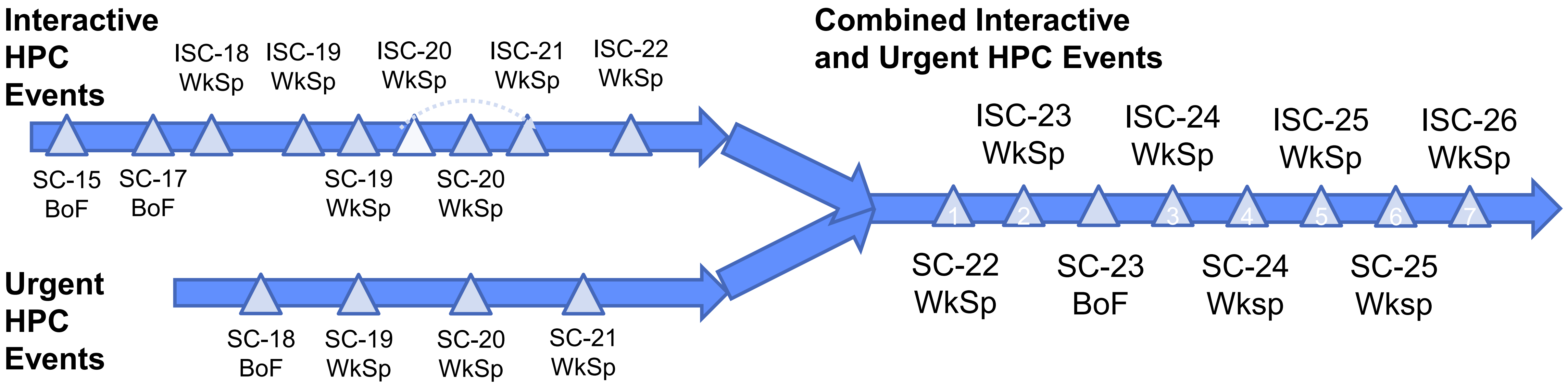
- Productivity in HPC requires greater interactivity with HPC systems
- Different approach to resource management and scheduling
  - Lower utilization
  - Agile engineering insight turns
- Many tools for interactive and urgent HPC
  - Integrated applications
  - Notebooks
  - Research desktops
  - Portals
- Many challenges with data management and urgent job orchestration
- Higher percentage of novice users
  - Need basic training
  - Early wins encourage further use and exploration

Workshop on Interactive and Urgent HPC (WIUHPC)

Combined Totals:

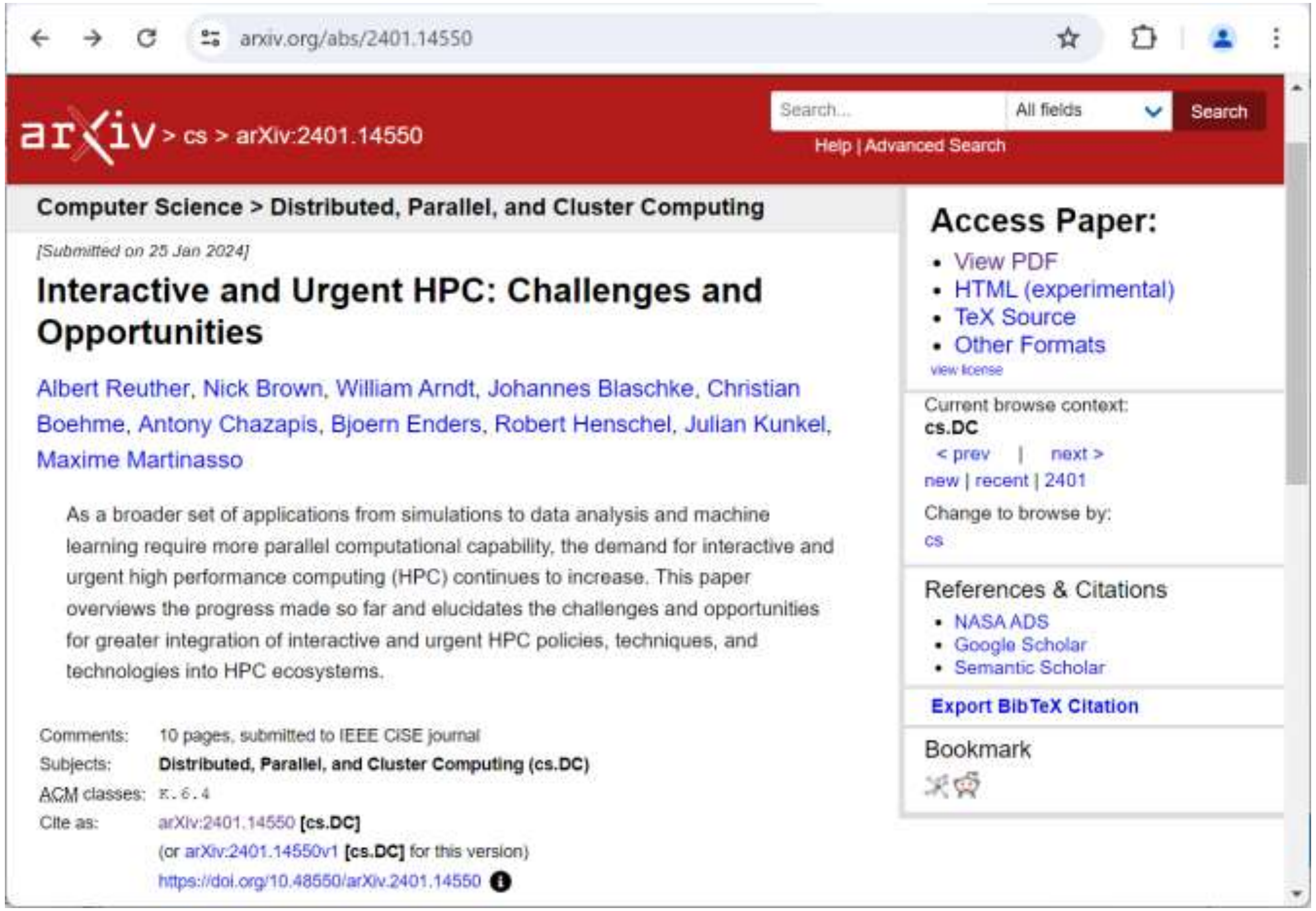
- 11 Keynote/Featured Talks
- 41 Research Paper Talks
- 25 Lightning Talks
- 7 BoF Talks
- 13 Panels
- 1 Position Paper
- 1 Book Chapter





Interactive and Urgent HPC - Position Paper

- **Written to inspire research, experimentation, and papers**
  - Introduced at SC23 BoF
  - Posted on arXiv
- **Addresses State of Practice & Research Challenges and Opportunities**
  - Organizational and System Policy
  - Scheduling Jobs
  - Technologies
  - Data Management
  - System User Support



arxiv.org/abs/2401.14550

arXiv > cs > arXiv:2401.14550

Search... All fields Search

Help | Advanced Search

Computer Science > Distributed, Parallel, and Cluster Computing

[Submitted on 25 Jan 2024]

Interactive and Urgent HPC: Challenges and Opportunities

Albert Reuther, Nick Brown, William Arndt, Johannes Blaschke, Christian Boehme, Antony Chazapis, Bjoern Enders, Robert Henschel, Julian Kunkel, Maxime Martinasso

As a broader set of applications from simulations to data analysis and machine learning require more parallel computational capability, the demand for interactive and urgent high performance computing (HPC) continues to increase. This paper overviews the progress made so far and elucidates the challenges and opportunities for greater integration of interactive and urgent HPC policies, techniques, and technologies into HPC ecosystems.

Comments: 10 pages, submitted to IEEE CiSE journal
Subjects: Distributed, Parallel, and Cluster Computing (cs.DC)
ACM classes: K.6.4
Cite as: arXiv:2401.14550 [cs.DC]
(or arXiv:2401.14550v1 [cs.DC] for this version)
https://doi.org/10.48550/arXiv.2401.14550

Access Paper:
- View PDF
- HTML (experimental)
- TeX Source
- Other Formats

view license

Current browse context:
cs.DC
< prev | next >
new | recent | 2401
Change to browse by:
cs

References & Citations
- NASA ADS
- Google Scholar
- Semantic Scholar

Export BibTeX Citation

Bookmark

https://doi.org/10.48550/arXiv.2401.14550

Interactive and Urgent HPC – Book Chapter

- To appear in Springer Nature book
  - *Title:* Urgent Computing across the Device-Edge-Cloud Continuum: Concepts, Applications, and Future Directions
  - *Editors:* Daniel Balouek, Manish Parashar
  - Posted on arXiv
- Covers State of the Research
  - Organizational and System Policy
  - Scheduling Jobs
  - Technologies, Environments, Tools
  - Data Management
  - Benchmarking
  - User Training
  - User Case Studies
  - Path Forward

arxiv.org/abs/2603.22542

Cornell University

Learn about arXiv becoming an independent nonprofit.

We gratefully acknowledge support from the Simons Foundation, Schmidt Sciences, Massachusetts Institute of Technology, and all contributors.

Donate

arXiv > cs > arXiv:2603.22542

Search... All fields Search

Help | Advanced Search

Computer Science > Distributed, Parallel, and Cluster Computing

[Submitted on 23 Mar 2026]

Interactive and Urgent HPC: State of the Research

Albert Reuther, William Arndt, Johannes Blaschke, Christian Boehme, Nick Brown, Antony Chazapis, Bjoern Enders, Jens Henrik Goebbert, Robert Henschel, Julian Kunkel, Maxime Martinasso, Michael Ringenburg, Rollin Thomas

When we think of how we use smartphones, e-commerce, collaboration platforms, LLMs, etc., most of our interactions with computers are interactive and often urgent. Similar trends of interactivity and urgency are coming to HPC, with applications from simulations to data analysis and machine learning requiring more parallel computational capability and more interactivity. This chapter overviews the progress made so far along with some vectors of what the path forward will bring for greater integration of interactive and urgent HPC policies, techniques, and technologies into our HPC ecosystems.

Comments: 32 pages, 3 figures

Subjects: Distributed, Parallel, and Cluster Computing (cs.DC)

ACM classes: A.1

Cite as: arXiv:2603.22542 [cs.DC]

(or arXiv:2603.22542v1 [cs.DC] for this version)

https://doi.org/10.48550/arXiv.2603.22542

Access Paper:

View PDF

HTML (experimental)

TeX Source

view license

Current browse context: cs.DC

< prev | next >

new | recent | 2026-03

Change to browse by: cs

References & Citations

NASA ADS

Google Scholar

Semantic Scholar

Export BibTeX Citation

Bookmark

https://doi.org/10.48550/arXiv.2603.22542

30th IEEE HPEC Virtual Conference September 14-18, 2026 (ieee-hpec.org)

- Premiere conference on High Performance Extreme Computing (850+ participants in 2025)
- 150+ presentations; 2026 Distinguished Speakers:
  - Prof. Saman Amarasinghe (Perkins Prof EECS MIT; ACM Fellow)
  - Prof. Yaneer Bar-Yam (President NECSI)
  - Craig Connors (VP & CTO Cisco Infrastructure & Security)
  - Prof. Danna Freedman (Quantum@MIT Director; MacArthur Fellow)
  - Prof. Sigal Gottlieb (UMass Dartmouth; SIAM & AWM Fellow)
  - Gregory Kurtzer (CEO CIQ & Founder of CentOS & Singularity)
  - Dave Martinez (MIT Lincoln Lab Fellow; IEEE Fellow)
  - Dr. CJ Newburn (Nvidia Distinguished Engineer)
  - Prof. Katsuhisa Ozaki (Shibaura Institute of Technology)
  - Prof. Daniela Rus (CSAIL Director; AAAS, ACM, AAAI, IEEE Fellow)
- 2026 Special Sessions
  - Age of Mixed-Precision
  - Bridging Quantum and HPC
  - Large Language Models: Opportunities and Challenges
  - MIT/Amazon/IEEE Graph Challenge
  - GraphBLAS Forum
  - BRAINS: Building Resilience through Artificial Intelligence for Networked Systems
  - Scaling Research Computing Education

IEEE HPEC
High Performance Extreme Computing

IEEE Boston Section

Sponsors	Thank You	Technical Organizer
NVIDIA MITRE	New England Section of SIAM Boston Chapter of the IEEE Computer Society MIT Student Chapter of SIAM	

- **HPEC focuses on hardware, software, systems, and applications where performance matters**
- **HPEC welcomes experts and people who are new to the field**

LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Questions?

Albert Reuther

reuther@ll.mit.edu

https://www.ll.mit.edu/biographies/albert-i-reuther

MIT LINCOLN LABORATORY
SUPERCOMPUTING CENTER

Backups

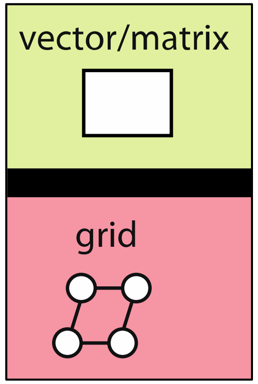
MIT LINCOLN LABORATORY
SUPERCOMPUTING CENTER

Large and Growing Need for Supercomputing

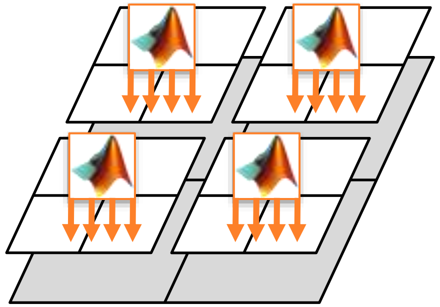
- requires world-class research to optimize diverse applications on complex hardware -

- **Challenge**: maximizing the efficiency of thousands of users distinct software
- **Approach**: LLSC developed innovative programming, *allocation*, and *scheduling* technologies

pPython Parallel Python[1]

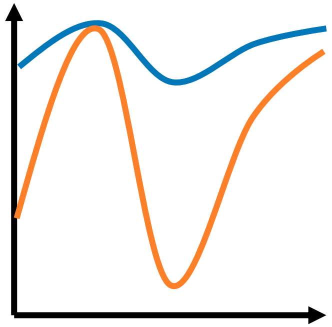


Hierarchical Allocation[2]

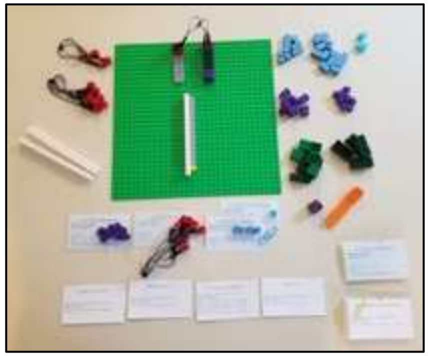


3D Monitoring & Management[3]

Spot Scheduling[4]



HPC Training via Serious Games[5]



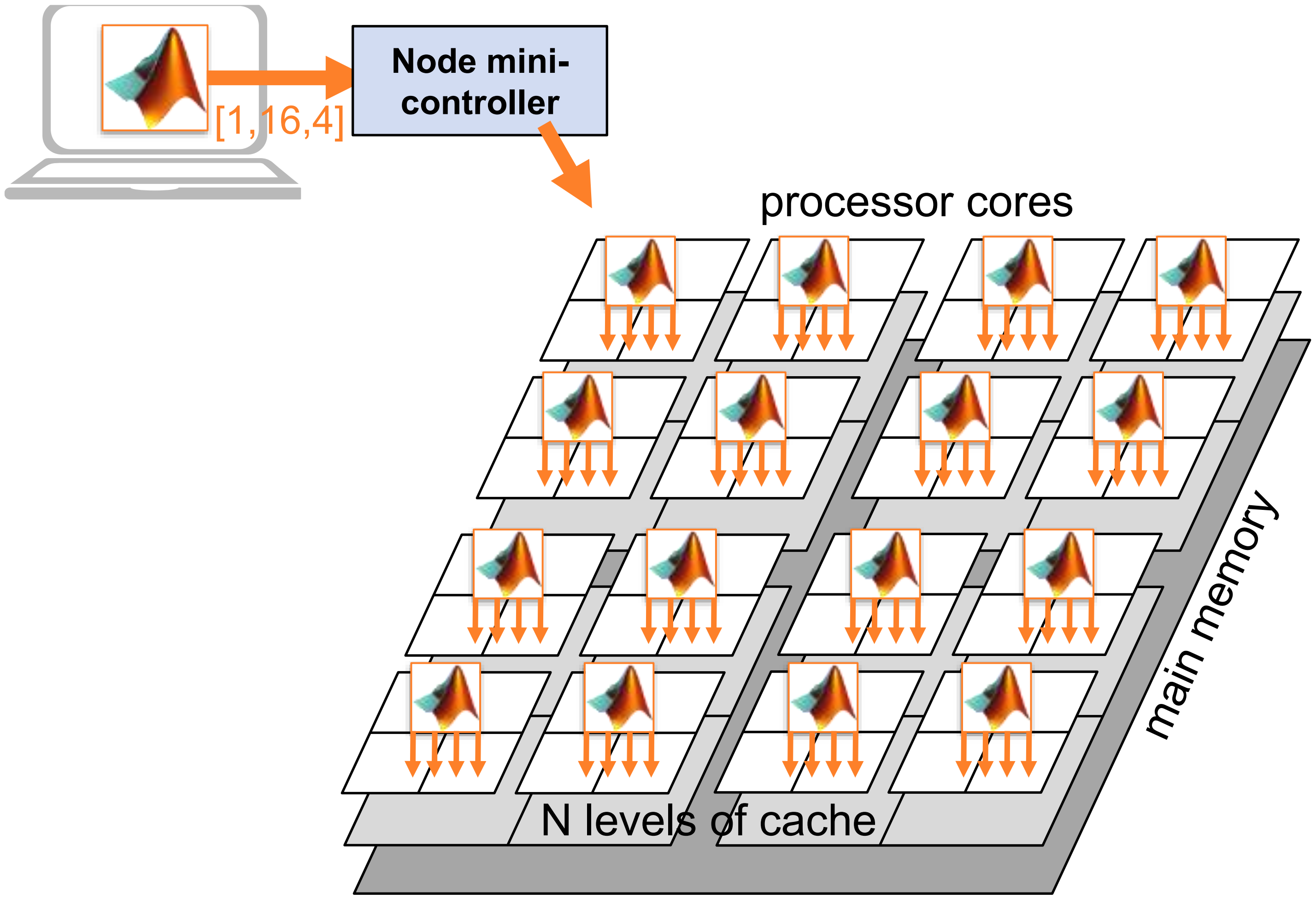
[1] *pPython Performance Study*, Byun et al, HPEC 2023
[2] *Node-Based Job Scheduling for Large Scale Simulations of Short Running Jobs*, Byun et al., HPEC 2021
[3] *Supercomputer 3D Digital Twin for User Focused Real-Time Monitoring*, Bergeron et al, HPEC 2024
[4] *Best of Both Worlds: High Performance Interactive and Batch Launching*, Byun et al., HPEC 2020
[5] *Teaching and learning HPC through serious games*, Mullen et al., JPDC 2021

Innovative *Allocation* Technologies

- **Traditional allocation**
  - **Cores are slots to be filled with jobs**
  - **10,000 cores = 10,000 jobs to be allocated**
    - **No ability to optimize for memory/core architecture**

- **Hierarchical allocation**
  - **User specifies: nodes, cores, and threads**
  - **Allocator divides up jobs among nodes**
  - **Dynamically writes hardware aware mini-controller on each node to independently start, monitor, and stop user processes on node**

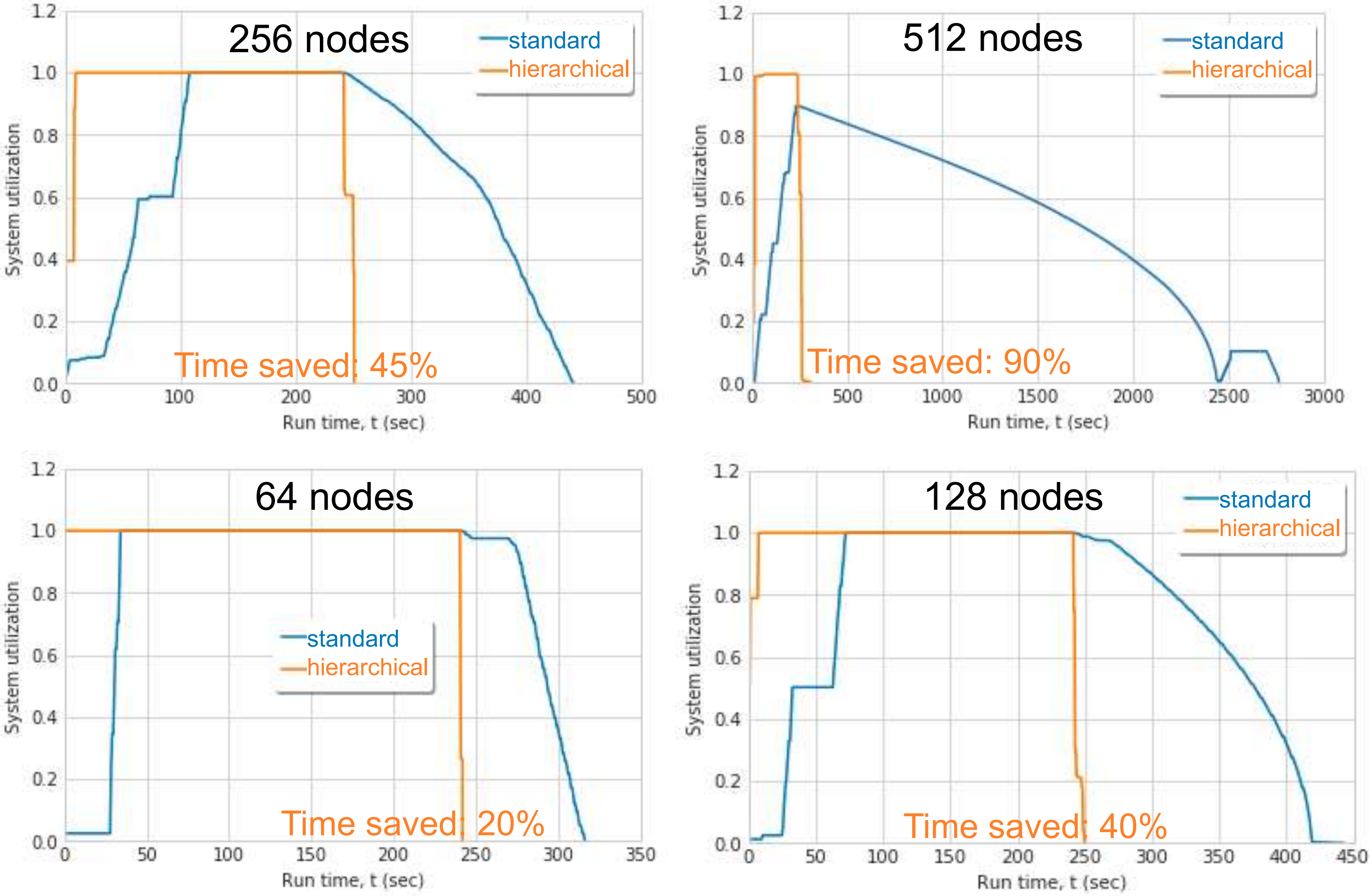
User requests: 1 node with 16 processes with 4 threads per process



Byun et al., *Node-Based Job Scheduling for Large Scale Simulations of Short Running Jobs*, IEEE HPEC 2021

Hierarchical Allocation Performance

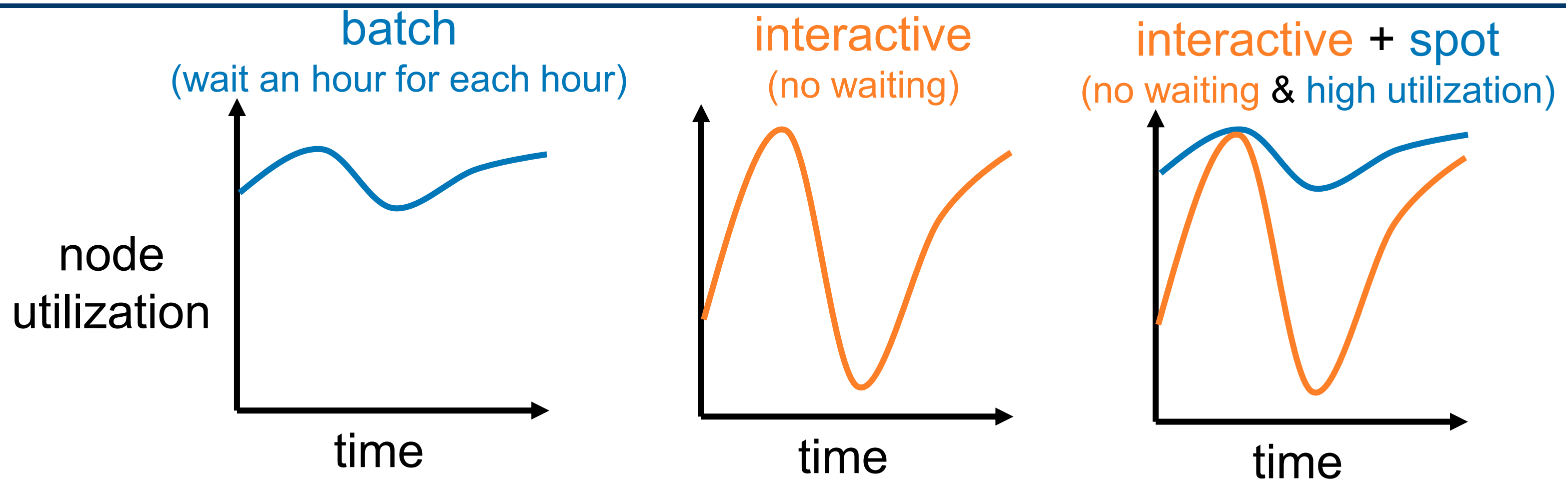
- **Dramatically accelerates large interactive jobs**
- **Transitioned to default LLSC approach**



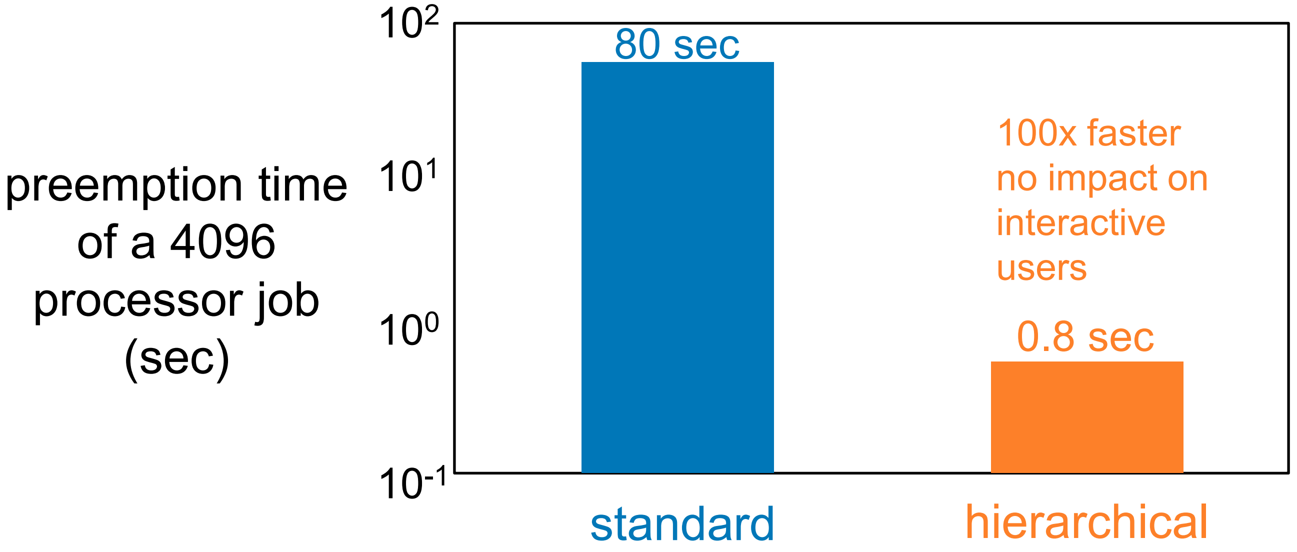
Byun et al., *Node-Based Job Scheduling for Large Scale Simulations of Short Running Jobs*, IEEE HPEC 2021

Innovative *Scheduling*

- **Traditional dilemma**
  - **Batch: 90% of nodes used, but user waits an hour per hour of run time**
  - **Interactive: launches immediately, but 50% of nodes used**
  - **Same wall-clock performance but interactive is preferred and more efficient**



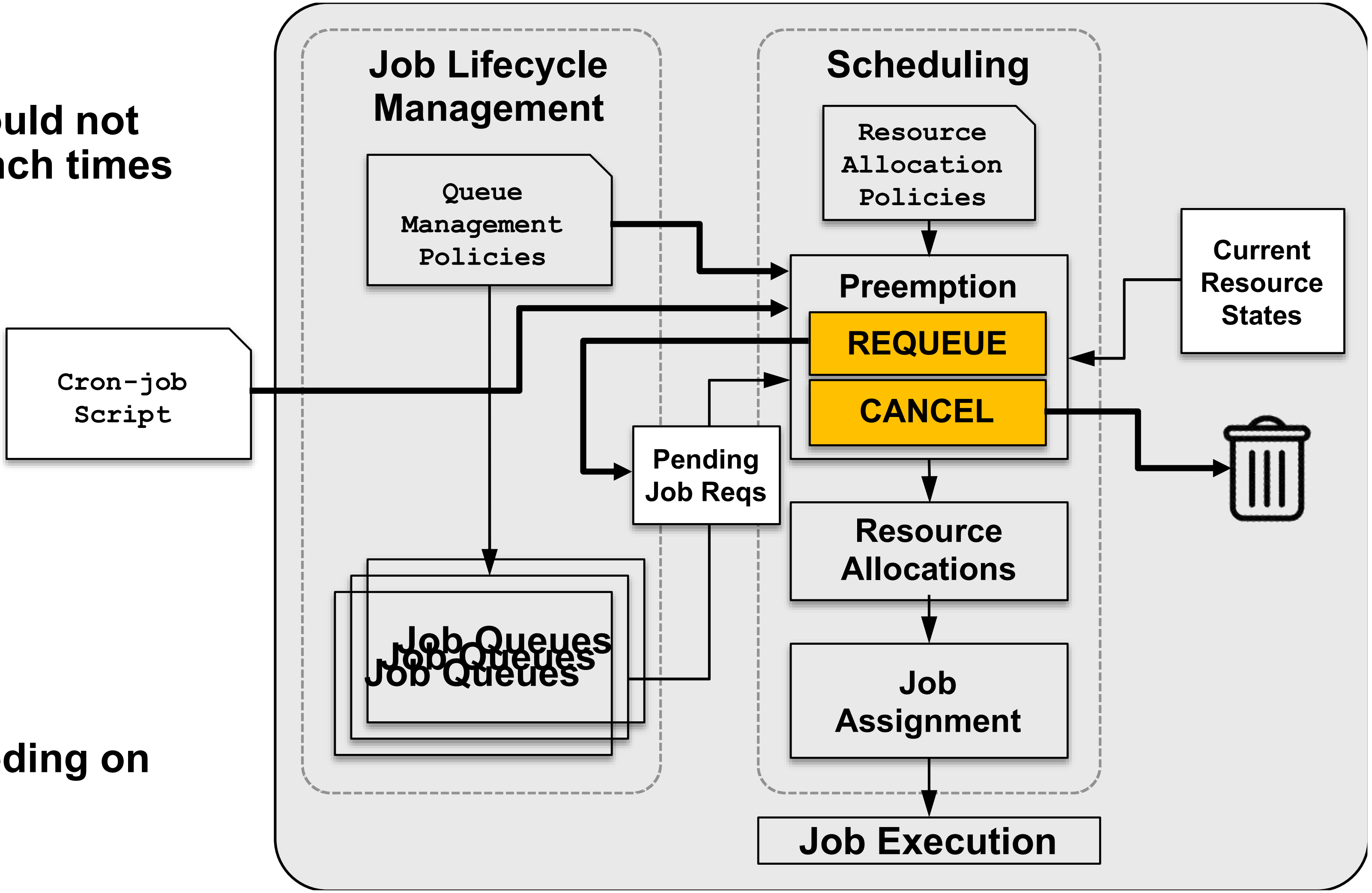
- **Best of both worlds**
  - **Spot jobs that can be instantly preempted by interactive jobs**
  - **Achievable with Hierarchical Allocation fast preemption**



C. Byun *et al*., "Best of Both Worlds: High Performance Interactive and Batch Launching," *2020 IEEE High Performance Extreme Computing Conference (HPEC)*, Waltham, MA, USA, 2020, pp. 1-7, doi: 10.1109/HPEC43674.2020.9286142.

Various Approaches of Spot Job Implementation

- **Goal**
  - Low priority spot jobs would not adversely impact the launch times of the normal jobs



- **Reserve N nodes for next interactive launch**
- **User cron job to**
  - Adjust spot queue limits
  - Kill spot jobs (LIFO) impeding on reserve nodes

C. Byun *et al*., "Best of Both Worlds: High Performance Interactive and Batch Launching," *2020 IEEE High Performance Extreme Computing Conference (HPEC)*, Waltham, MA, USA, 2020, pp. 1-7, doi: 10.1109/HPEC43674.2020.9286142.

History of Supercomputing at Lincoln Laboratory


1951 Whirlwind


1953 Magnetic-Core Memory Array


1956 TX-0

1958
- AN/FSQ-7 (Whirlwind II)
- Average Response Computer (ARC)
- CG-24
- TX-2

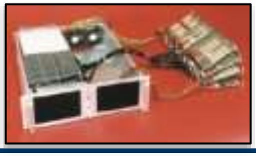
1962 Lincoln Instrument Computer (LINC)


1963 Sketchpad


1970 Fast Digital Processor (FDP)

1970 GENESYS

1974 Lincoln Digital Voice Terminal (LDVT)



1977 : Lincoln Digital Signal Processor (LDSP)

Late 1970s High-speed FFT pipelined processor

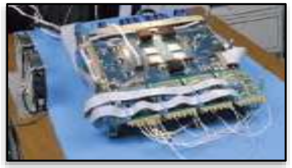
1978 Micro-Processor Based LPC Vocoder (LPCM)


1982 Compact LPC Vocoder

Early 1990s
- APT processor
- RAPTOR processor
- Space-Time Adaptive Processing Library (STAPL)

1992 Radar Surveillance Technology Experimental Radar (RSTER) processor

1999 Parallel Vector Library (PVL)



2002
- ISR Processing and Array Technology (IPAT) processor
- MatlabMPI

2003 pMatlab

2004
- pMapper gridMatlab
- LLGrid TX-2500

2004 Knowledge-Aided Sensor Signal Processing and expert Reasoning (KASSPER) processor

2007
- Parallel Vector Tile Optimizing Library (PVTOL)
- Real-Time Communication Layer (RTCL)

2012 D4M

2014 LLGrid TX-Green

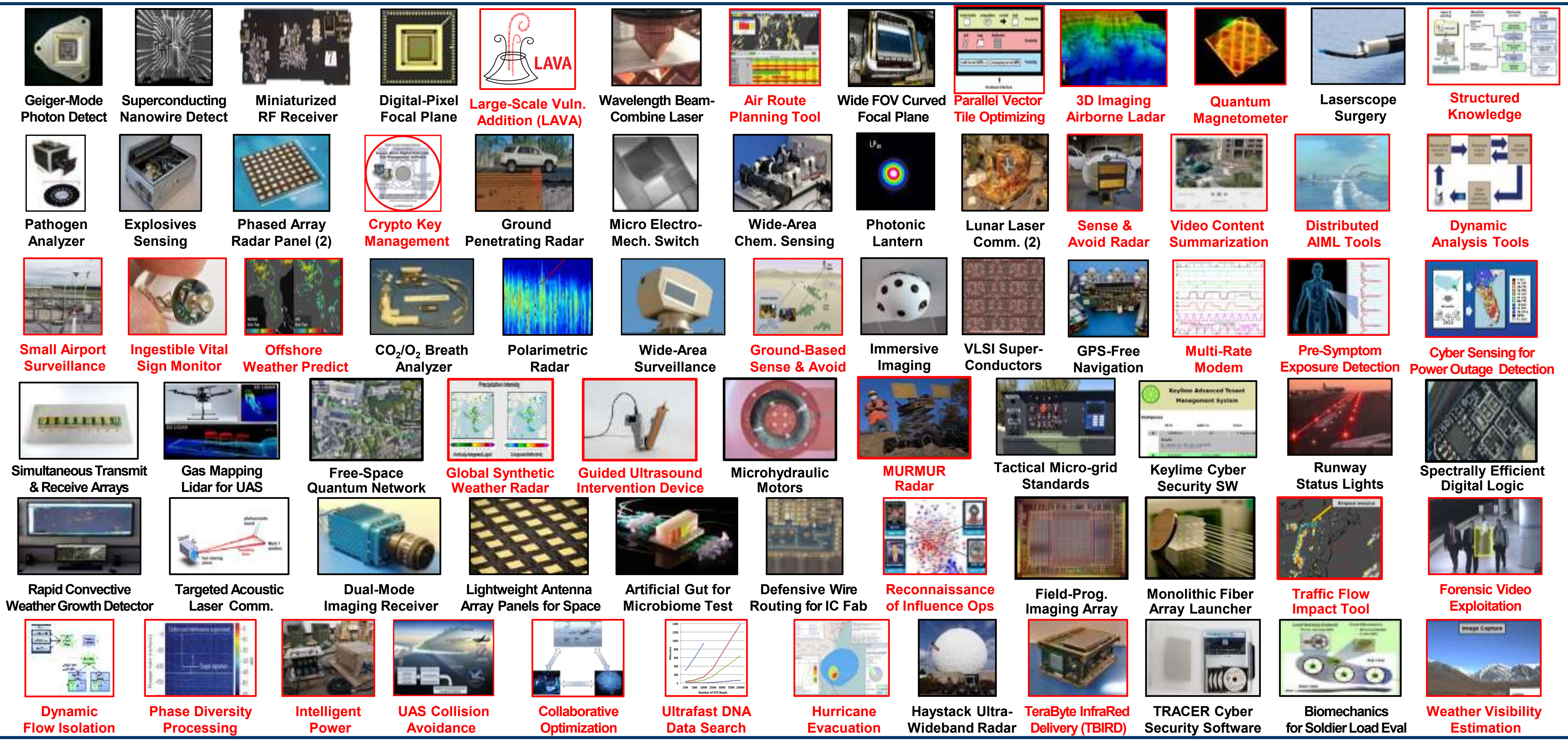
2015 BigDAWG

2016 Lincoln Laboratory Supercomputing Center (LLSC)

1950 1960 1970 1980 1990 2000 2010

https://www.ll.mit.edu/r-d/cyber-security-and-information-sciences/lincoln-laboratory-supercomputing-center

MIT Lincoln Laboratory R&D 100 Awards



LLSC enabled

https://www.ll.mit.edu/sites/default/files/page/doc/2024-01/RD100_Awards_booklet_86Winners_2023_V2_2.pdf

LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

HPC Carpentry, Software Carpentry, RSE

WIUHPC Participating Organizations