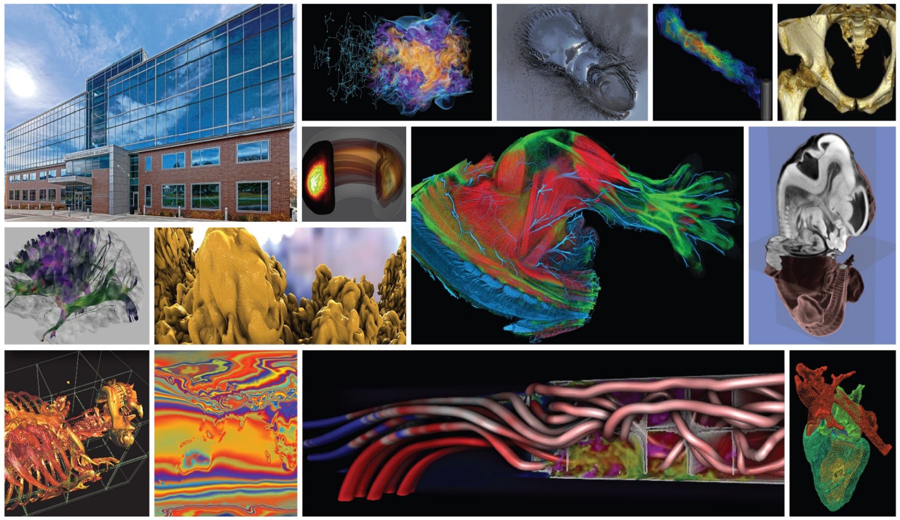
Harnessing the Edge-HPC Continuum for Science

Manish Parashar

Scientific Computing & Imaging (SCI) Institute

Kahlert School of Computing

University of Utah

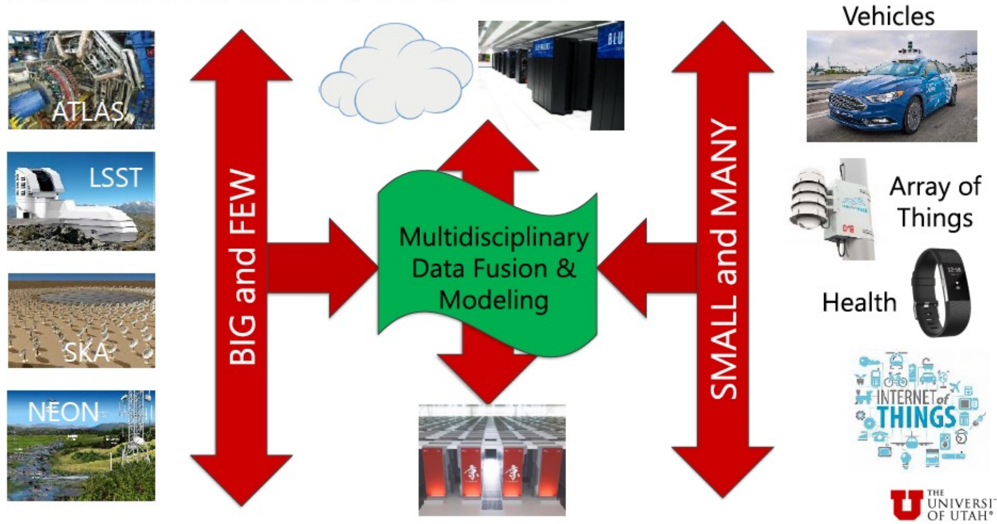


Salishan Conference on High Speed Computing, April 27, 2023

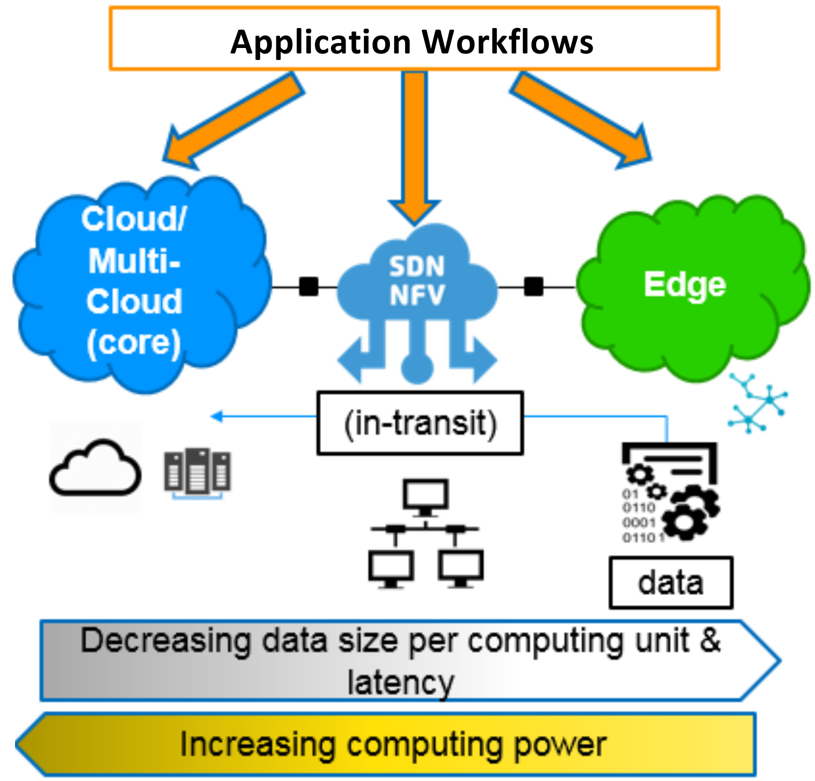
THE UNIVERSITY OF UTAH

www.sci.utah.edu

Emerging Science and the Computing Continuum



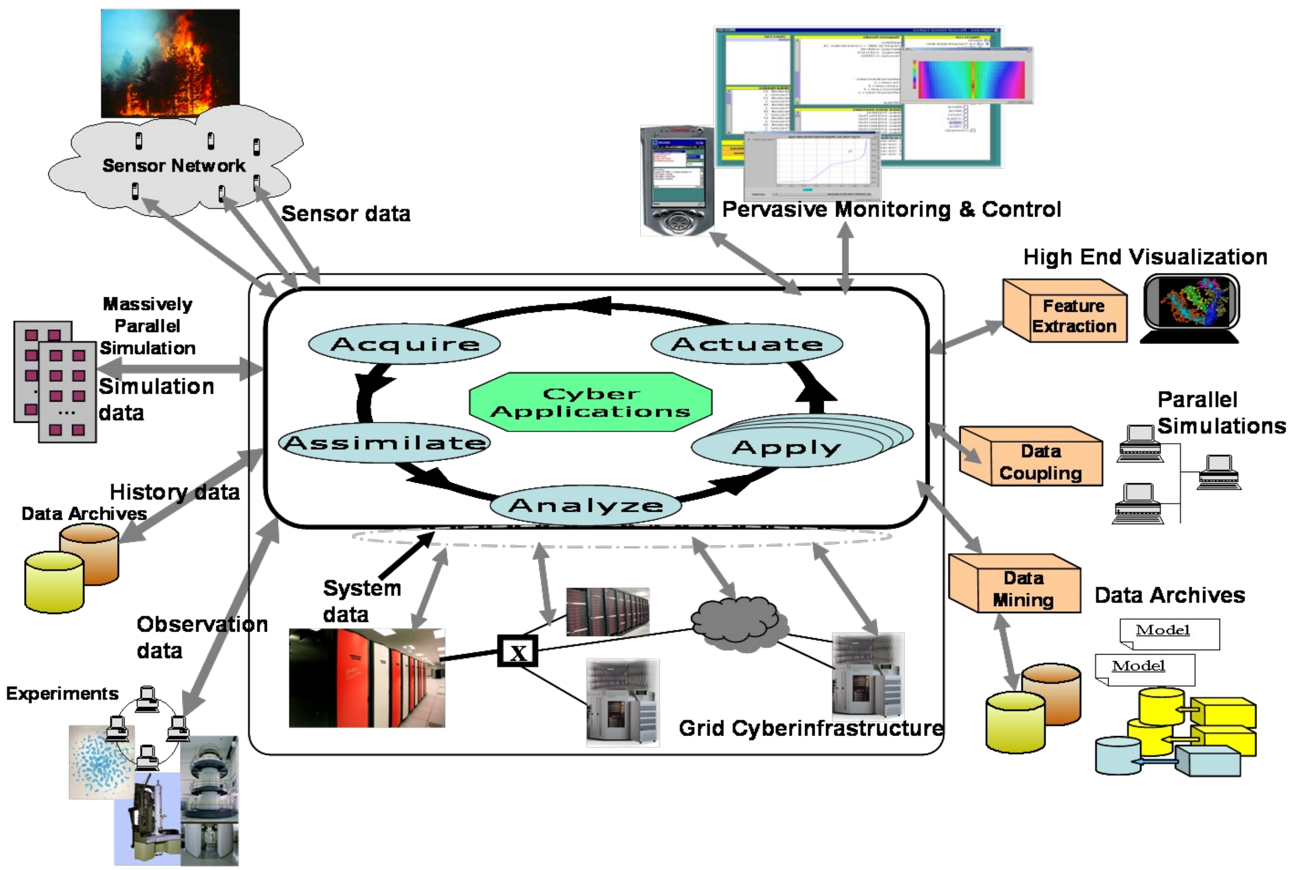
Ack. Dan Reed



Computing across the Continuum [DCC/SIGMETRICS 2020]

- Leverage resources and services at the logical extreme of the network and along the data path to increase the value of the data while potentially reducing costs
- Exploit the rich ecosystem of data and computation resources at the edge so that data is not moved

Emerging Science and the Computing Continuum: Dynamic, Data-driven End-to-end Workflows



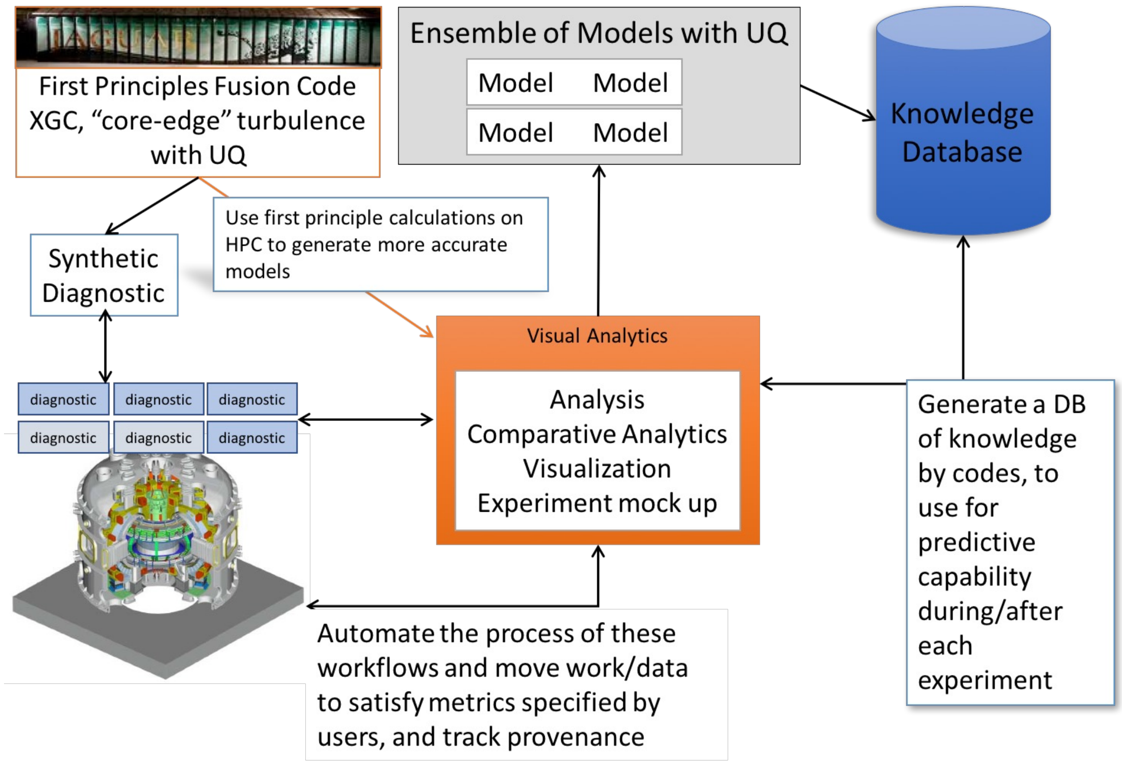
THE UNIVERSITY OF UTAH

SCI www.sci.utah.edu

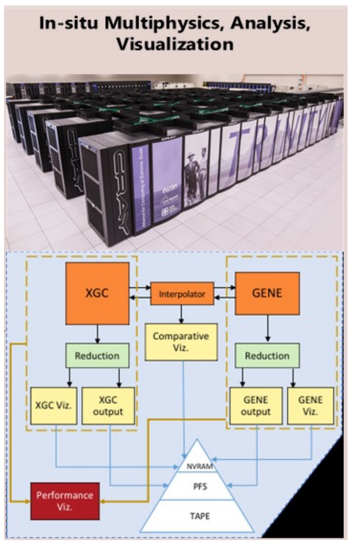
End-to-End Experiment Management

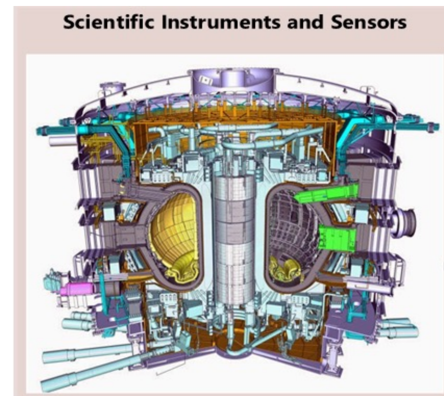
End-to-end Workflows

- High-fidelity instruments, experiments
- Complex coupled simulation workflows composed of coupled simulations and services
- Online data analysis, visualization
- Tight coupling between experiments/observations, simulations and analytics



End-to-end Fusion Workflow



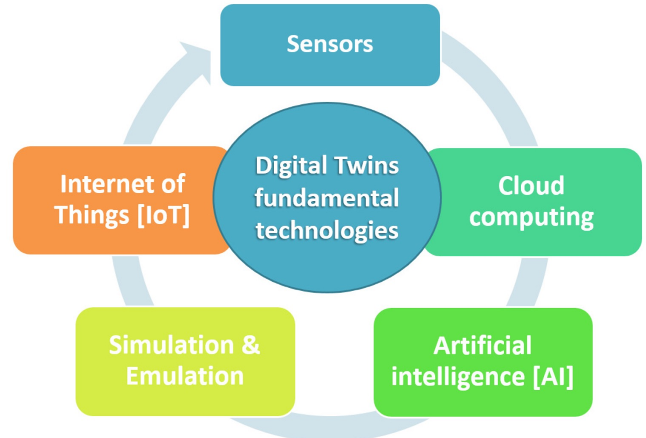


Plasma Disruption Analysis Fusion Workflow

- The goal is to implement a plasma visualization diagnostics system that enables early prediction of anomalies while the tokamak operates
- Data is a time-series of matrix images obtained via direct 2D/3D visualizations of the tokamak plasma
- Analytics include structural analysis and blob detection, filtering & visualization

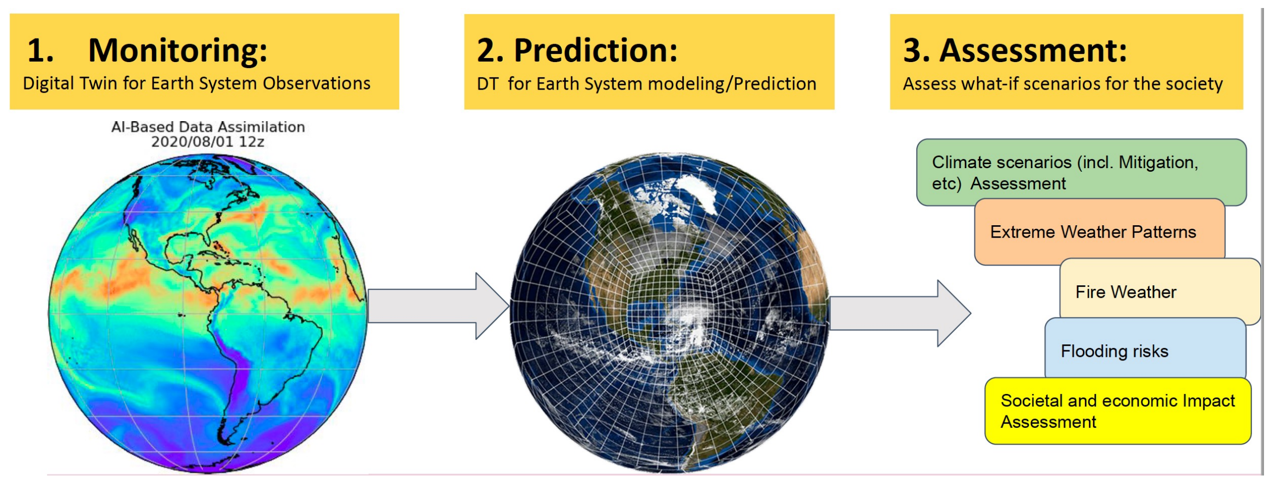
Digital Twins

- A digital representation of an intended or actual real-world physical product, system, or process (a physical twin) that serves as the effectively indistinguishable digital counterpart of it for practical purposes, such as simulation, integration, testing, monitoring, and maintenance.
- NASEM: Foundational Research Gaps and Future Directions for Digital Twins -- https://www.nationalacademies.org/our-work/foundational-research-gaps-and-future-directions-for-digital-twins



https://doi.org/10.3390/asi5040065

Digital Twins for Earth Systems (NOAA)



THE UNIVERSITY OF UTAH®

Urgent Computing: Managing uncertainty under constraints

DEFINITION: Computing under **strict time** and **quality constraints** to support decision making with the desired confidence within a defined time interval

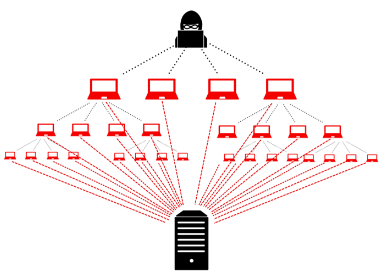
Identify events → Develop response → Trigger appropriate actions

Urgent applications need to react to unforeseen events and to manage complex cost/benefit tradeoffs to meet constraints

- Steer online/on-demand computations based on data/content
- Balance costs of computations versus efficiency and availability of resources
- Support urgency constraints for timely decision-making
- …

Global Challenges and Urgent Science

Earthquake & Tsunamis
Early warning and rapid response requires highly-accurate near-real-time analysis, simulations, and response.

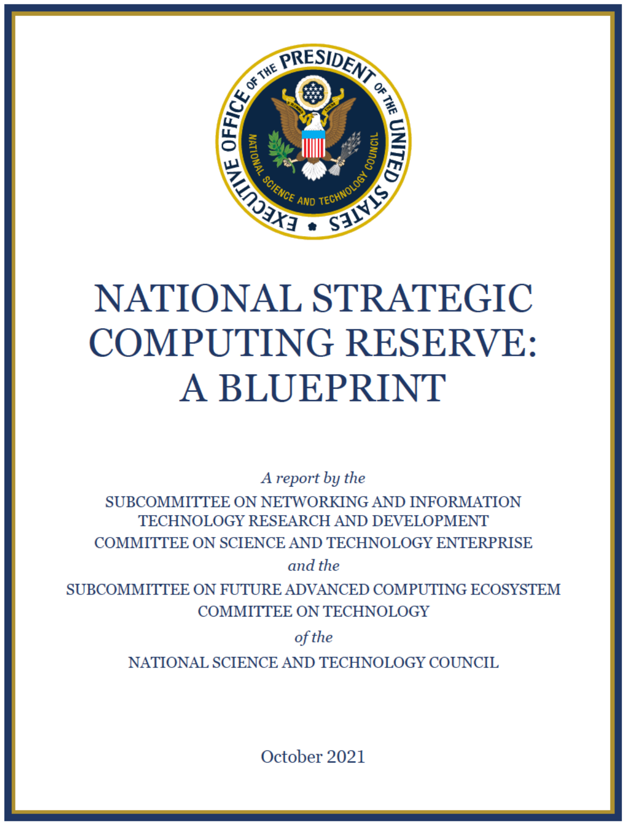


Cyber-Attacks
Amazon Web Services DDoS Attack (Feb 2020) – Peak traffic volume of 2.3 Tbps.

Evacuation traffic management
Traffic jams caused by evacuations from Hurricane Rita in Houston TX, September 23rd, 2005 – real-time situation analysis, response planning and actuation is essential.

Weather and Climate
Hurricane Ida in New Jersey (September 2021) – Sudden flooding resulted due to months of rain falling in a few hours.

National Strategic Computing Reserve (NSCR)


NATIONAL STRATEGIC COMPUTING RESERVE: A BLUEPRINT

A report by the
SUBCOMMITTEE ON NETWORKING AND INFORMATION TECHNOLOGY RESEARCH AND DEVELOPMENT
COMMITTEE ON SCIENCE AND TECHNOLOGY ENTERPRISE
and the
SUBCOMMITTEE ON FUTURE ADVANCED COMPUTING ECOSYSTEM
COMMITTEE ON TECHNOLOGY
of the
NATIONAL SCIENCE AND TECHNOLOGY COUNCIL

October 2021

- Advanced computing cyberinfrastructure can be a strategic National asset in emergency response, if mobilized quickly.
- Goals for an NSCR:
  - Ensure availability of a ready “reserve” of resources (computing, data, software, services) and expertise that can be leveraged nimbly in times of urgent need.
  - Establish policies, processes, and agreements to enable agile, effective, and impactful resource mobilization.
  - Build on continued longer-term strategic investments in resources (computing, data, software, services), expertise.
  - Coordinate across agencies, stakeholder communities, and other national reserves.

Full report available at:
https://www.whitehouse.gov/wp-content/uploads/2021/10/National-Strategic-Computing-Reserve-Blueprint-Oct2021.pdf

Some Research Challenges

- How to drive computation through data
  - Express application behavior based on available data and its content
- How to accommodate uncertainties in data and computation
  - Move away from precise to approximate computing
- How to build applications and manage workflows so that they adapt to increase value
- How to continuously optimize execution in a dynamic data-driven environment
  - How to discover and aggregate services (data, resources, …) that fit the current requirements
- How to develop the system infrastructure and services to support dynamic execution
- How to incorporate utility models, market models, social/trust models, etc.
- How to address security, privacy, ….
- …….

A Motivating Usecase: Wildfire in California leads to Air Pollution in Utah

California's fast-moving Oak Fire burns 14,000 acres and forces thousands to evacuate outside Yosemite National Park

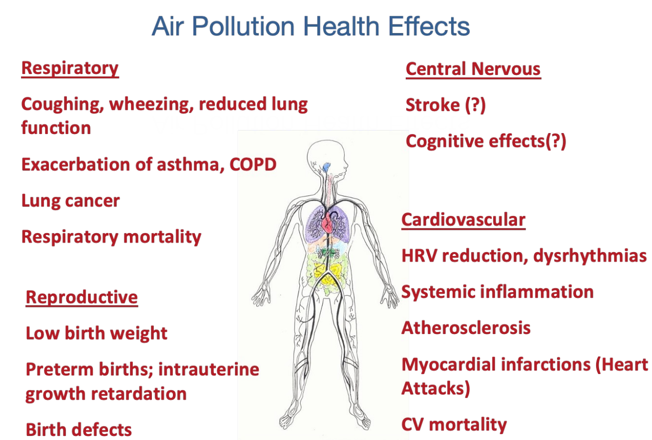
By Jason Hanna, Rebekah Riess, Sara Smart and Andy Rose, CNN

Updated 0645 GMT (1445 HKT) July 25, 2022

Could the exception become the rule? 'Uncontrollable' air pollution events in the US due to wildland fires

Liji M David[1,2], A R Ravishankara[1,2], Steven J Brey[2], Emily V Fischer[2], John Volckens[3] and Sonia Kreidenweis[2]

Published 22 February 2021 • © 2021 The Author(s). Published by IOP Publishing Ltd



Mitigating the negative impacts of wildfires on air quality requires combining knowledge from multiple data sources and integrate it on-demand with distributed computational models.

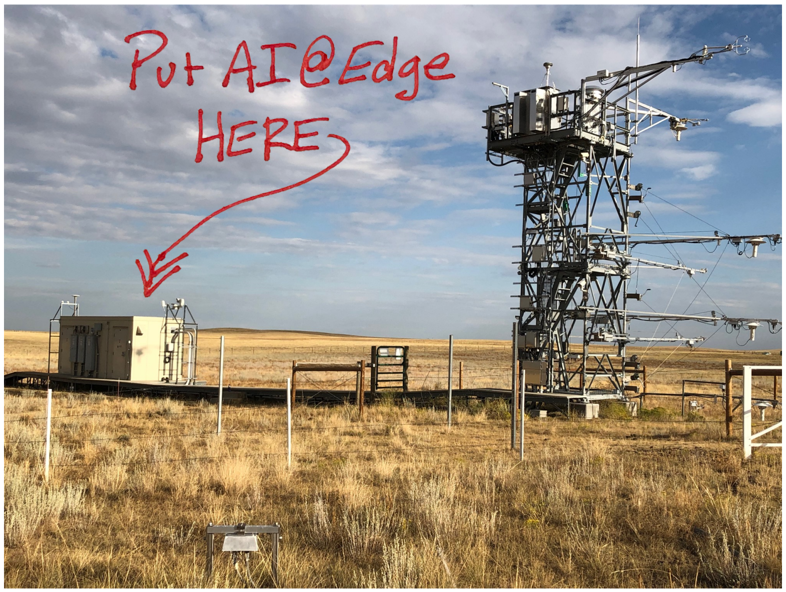
THE UNIVERSITY OF UTAH®

Balouek-Thomert, D., Caron, E., Lefèvre, L. & Parashar, M., Towards a methodology for building dynamic urgent applications on continuum computing platforms, Combined International Workshop on Interactive Urgent Supercomputing (CIW-IUS), 11/2022.

www.sci.utah.edu

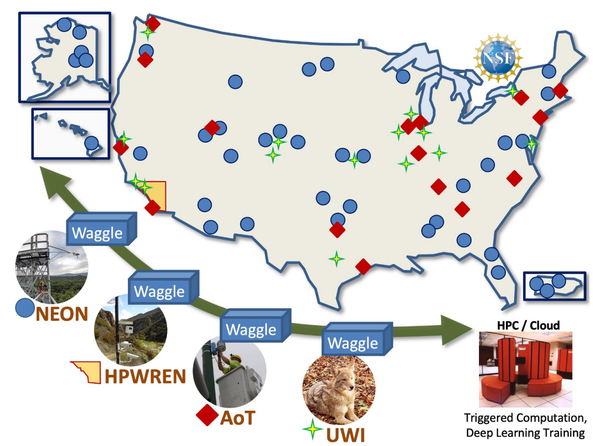
Sensor/Instrument — HPC/Cloud

AI@Edge and the Digital Continuum



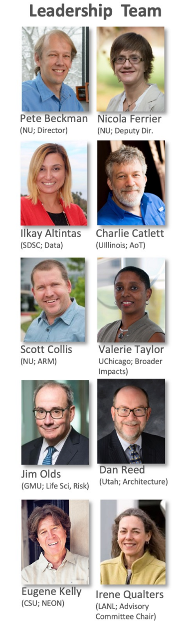
SAGE

Cyberinfrastructure for
AI at the Edge
sagecontinuum.org



Education & Training

Leadership Team

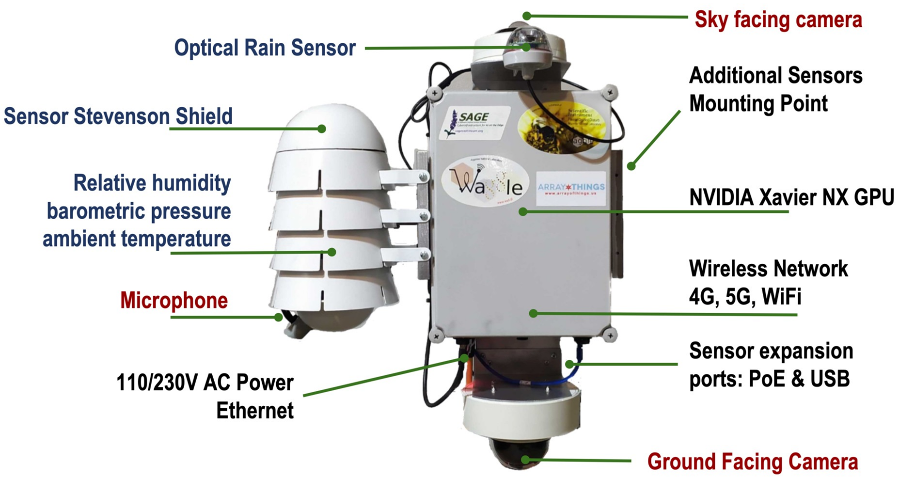


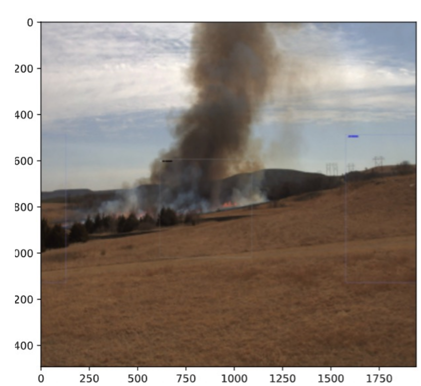
Northwestern University · The University of Chicago · Argonne National Laboratory · Colorado State University · UC San Diego · George Mason University · Array of Things · ARM · neon National Ecological Observatory Network · HPWREN SDSC San Diego Supercomputer Center · Lincoln Park Zoo For Wildlife. For All.

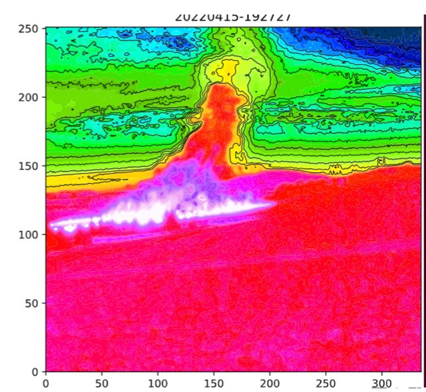
Ack: Pete Beckman, ANL

Ack: Pete Beckman, ANL

NEON Mobile Deployment Platform (MPD) with SAGE Konza Prairie for controlled burn: April 2022.



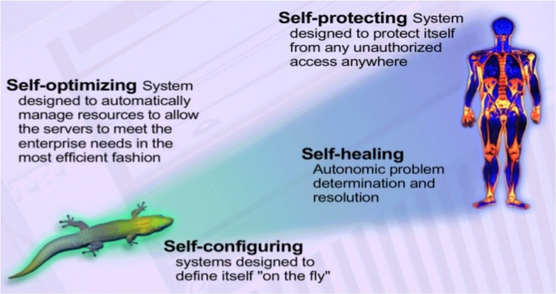




THE UNIVERSITY OF UTAH

www.sci.utah.edu

Autonomic Management



- Optimize resource provisioning and workload allocation to meet objectives and constraints set by users, applications, and/or resource providers
- Create models to translate resource/service capabilities and availabilities into application-level utilities (e.g., throughput, performance, etc.)
- Combine predictive and reactive approaches to improve decisions
- Quantify errors and uncertainties to offer confidence levels
  – How much error can I tolerate to maintain certain QoS?

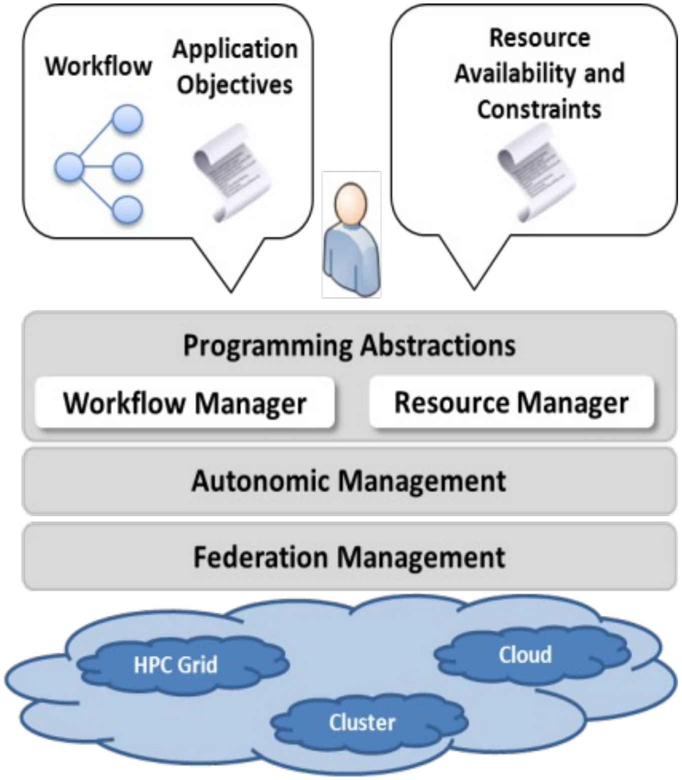
I. Petri, O. F. Rana, L. F. Bittencourt, D. Balouek-Thomert and M. Parashar, "Autonomics at the Edge: Resource Orchestration for Edge Native Applications," *IEEE Internet Computing*, vol. 25, no. 4, pp. 21-29, 1 July-Aug. 2021, doi: 10.1109/MIC.2020.3039551.

Parashar, M., Hariri, S. (2005). Autonomic Computing: An Overview. In: Banâtre, JP., Fradet, P., Giavitto, JL., Michel, O. (eds) Unconventional Programming Paradigms. UPP 2004. https://doi.org/10.1007/11527800_20

Distributed Software-defined Environments

M. AbdelBaky

- Create a nimble and programmable ecosystem that autonomically evolves over time, adapting to:
  - Define composition programmatically using constraints
  - Respond to changes in the infrastructure, application requirements
  - Allocate computational resources close to digital data sources
  - Process data in-situ and/or in-transit



- AbdelBaky M, Diaz-Montes J, Parashar M. Software-defined environments for science and engineering. *The International Journal of High Performance Computing Applications*. 2018;32(1):104-122. doi:10.1177/1094342017710706
- M. Abdelbaky and M. Parashar, "A General Performance and QoS Model for Distributed Software-Defined Environments," in *IEEE Transactions on Services Computing*, vol. 15, no. 1, pp. 228-240, 1 Jan.-Feb. 2022, doi: 10.1109/TSC.2019.2928300.

R-Pulsar: Enabling Data-driven Workflows across the Continuum (rpulsar.org)

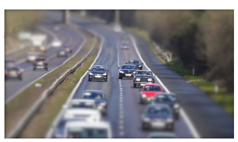
Daniel Balouek-Thomert

A programming system for building data-to-discovery pipelines

- **Programming support** for data-driven workflows and application models
- **System management** for discovery and federation of sensors and computational resources / execution of workflows

Producers

Traffic



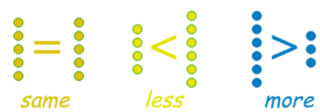
Smoke

Any physical/virtual sensors

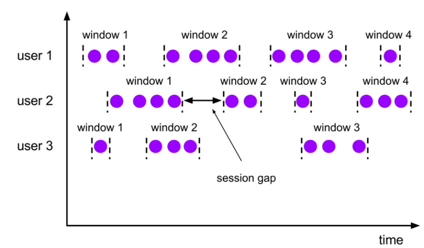
DATA COLLECTION

Consumers / Filters

Individual values



Windows



FILTERING / RULE MATCHING

Actions

Send a message

Approximate computations

Change system parameters or execution mode

Launching computation (local, cloud, hybrid, etc.)

ACTION TRIGGERS

THE UNIVERSITY OF UTAH

SCI www.sci.utah.edu

A Motivating Use case: Wildfire in California leads to Air Pollution in Utah

California's fast-moving Oak Fire burns 14,000 acres and forces thousands to evacuate outside Yosemite National Park

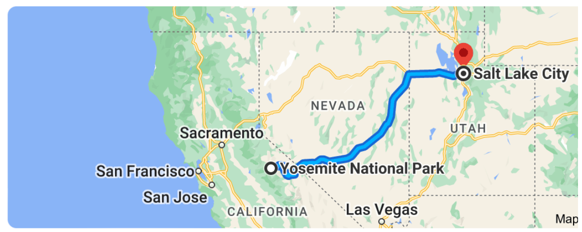
By Jason Hanna, Rebekah Riess, Sara Smart and Andy Rose, CNN

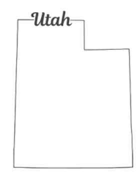
Updated 0645 GMT (1445 HKT) July 25, 2022

Could the exception become the rule? 'Uncontrollable' air pollution events in the US due to wildland fires

Liji M David[1,2], A R Ravishankara[1,2], Steven J Brey[2], Emily V Fischer[2], John Volckens[3] and Sonia Kreidenweis[2]

Published 22 February 2021 • © 2021 The Author(s). Published by IOP Publishing Ltd





Balouek-Thomert, D., Caron, E., Lefèvre, L. & Parashar, M., Towards a methodology for building dynamic urgent applications on continuum computing platforms, Combined International Workshop on Interactive Urgent Supercomputing (CIW-IUS), 11/2022.

Sage (beta)

Edge (Field cameras)

1. Smoke detection at the edge of the network using Camera Imagery

WIFIRE

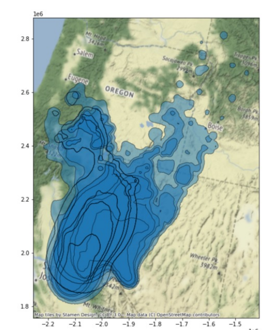
Cloud (SDSC)

2. Wildfire simulations to determine the severity and direction of fires

THE UNIVERSITY OF UTAH

HPC (Utah)

3. Pollution Concentration maps to support decision-making



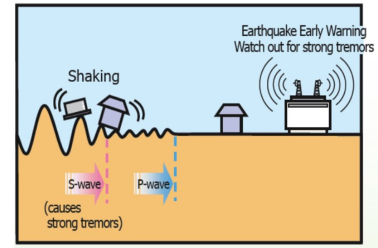
www.sci.utah.edu

THE UNIVERSITY OF UTAH

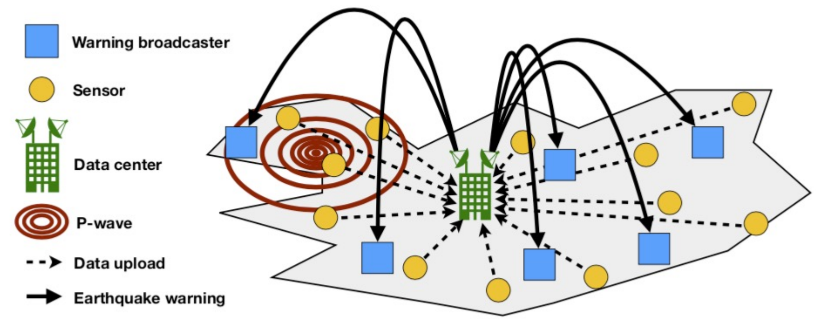
Earthquake Early Warning

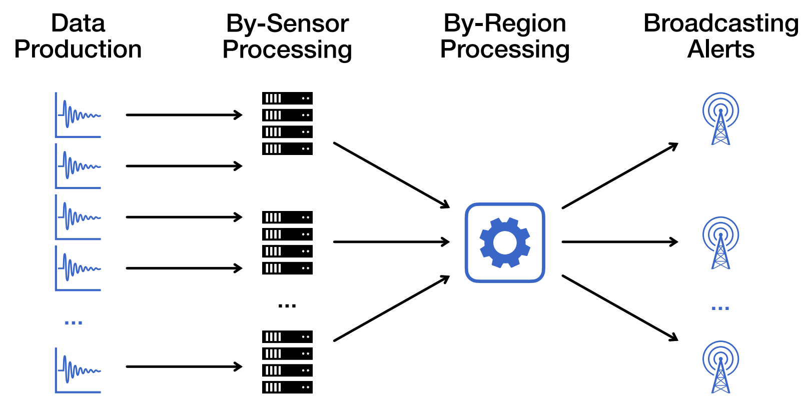
- **Earthquake Early Warning (EEW)** requires earthquakes to first be characterized (magnitude, location, speed of displacement, etc.)
  - A ***single data source*** doesn't able to cover a whole spectrum of events:
    - Seismometers are good for the smaller earthquakes (< 6.5); High-precision GPS are good for larger earthquakes.
  - ***Centralized data processing*** cannot support real-time, high volume data processing

- **Goal:** Combine ***multiple data sources*** to cover the whole spectrum of events; leverage the ***CI continuum*** for preserving latency and resiliency



Credit: Japan Meteorological Agency





Fauvel, K, et al. "A distributed multi-sensor machine learning approach to earthquake early warning," Proceedings of the AAAI Conference on Artificial Intelligence, **Outstanding Paper Award in Artificial Intelligence for Social Impact,** 2020.

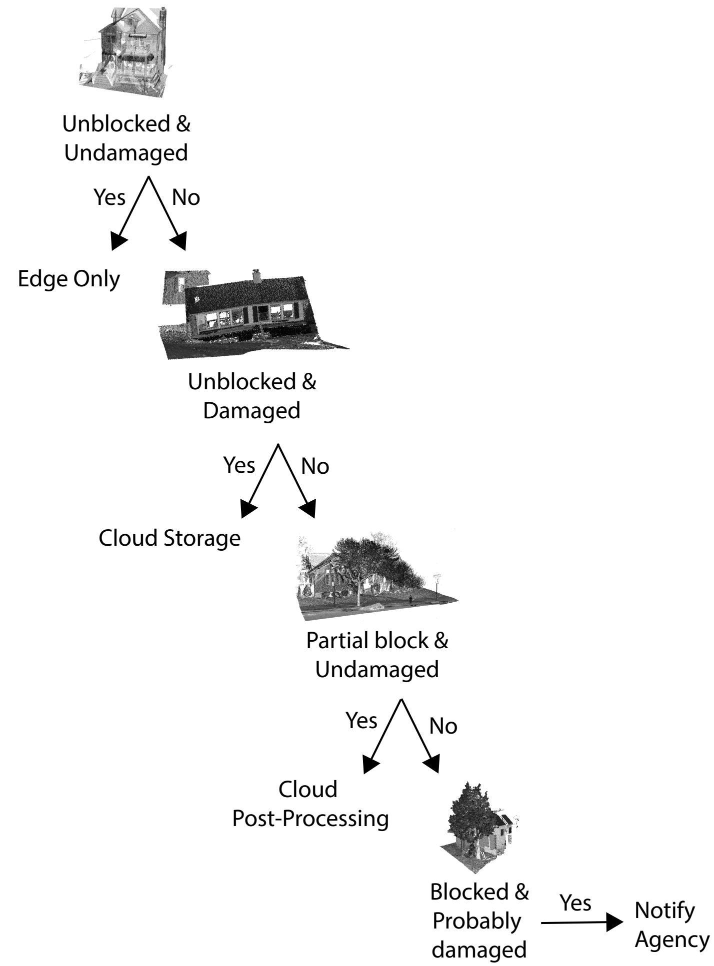
Hurricane Sandy -- Disaster Response

Goal: Quickly and efficiently determine whether the civil infrastructure (buildings, bridges) conditions are safe or not for evacuees to return

Edge Approximation: Canny-edge detection algorithm

- Multi-stage algorithm to detect a wide range of edges in images
- Ease to apply approximation techniques (substitution, discarding) at function and input parameters level

E. Renart, D .Balouek-Thomert, X.Hu, J. Gong and M. Parashar. Online Decision-Making Using Edge Resources for Content-Driven Stream Processing, eScience'17, October 2017



Content driven disaster response workflow decision stages.

THE UNIVERSITY OF UTAH®

SCI www.sci.utah.edu

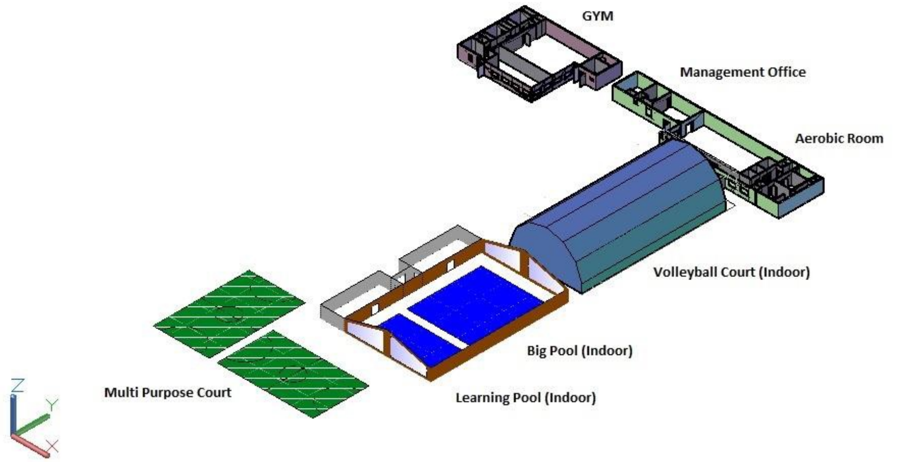
Smart Building Management

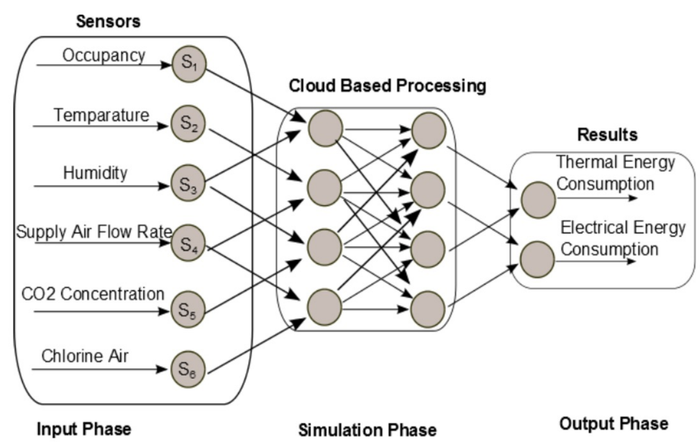
Goal: Improve safety and energy efficiency of the instrumented FIDIA SportE2 Facility

HPC/Cloud Application: EnergyPlus building energy simulation (https://energyplus.net)

Edge Approximation: Trained neural network as a function approximator for EnergyPlus

- Retraining triggered by changes in inputs/errors





A. Zamani, I. Petri, J. Diaz-Montes, O. F. Rana and M. Parashar, "Edge-supported Approximate Analysis for Long Running Computations", FiCloud 2017.

Summary

Extreme and pervasive compute and data can enable urgent science	Large volumes of geographically distributed data needs to be processed by complex application workflows in a timely manner
Computing in the continuum	Leverage resources and services at the logical extremes, along the data path, in the core to process data in support of urgent science
Many challenges across multiple layers	Application formulation, programming systems, middleware services, standardization & interoperability, autonomic engines, etc.

Stay Tuned!

QuickPar 2023 – First International Workshop on Urgent Analytics for Distributed Computing

Co-located with Euro-Par 2023, Limassol, Cyprus

IMPORTANT DATES:

- **Paper Deadline:** May 5, 2023
- **Author Notification:** June 19, 2023
- **Camera ready papers:** July 2, 2023
- **Workshop (Full Day):** August 28 or 29, 2023

https://2023.euro-par.org/

EURO-PAR CONFERENCE 2023

THE UNIVERSITY OF UTAH

ORGANIZERS:

Daniel Balouek-Thomert, Inria
Manish Parashar, University of Utah

SCI 35 www.sci.utah.edu

Thank you

www.sci.utah.edu

Manish Parashar

manish.parashar@utah.edu