RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

Computing in the Continuum: Harnessing Pervasive Data Ecosystems

Manish Parashar, Ph.D.

Director, Rutgers Discovery Informatics Institute RDI²

Distinguished Professor, Department of Computer Science

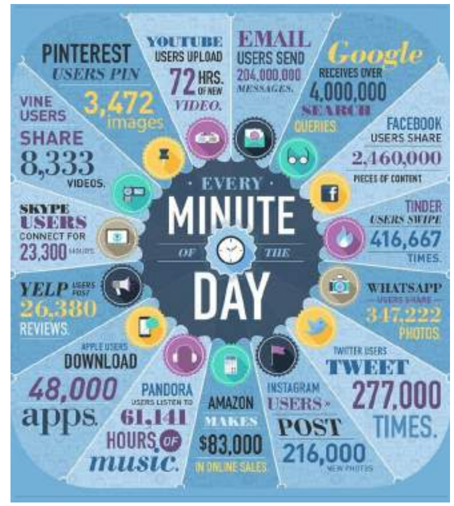
August 07, 2019

Outline

- The emerging data/compute reality
- Computing in the Continuum
- Research challenges (and some ongoing research)
- Conclusion

The Era of Big Data and Extreme Compute

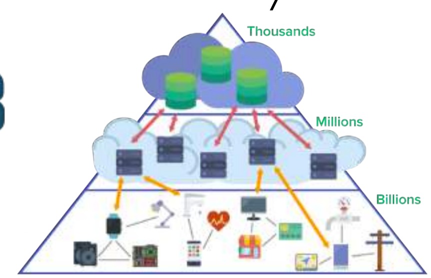
- *Data and computing data are pervasive*
- Extreme scales; extreme data volumes and rates
- Novel paradigms: cloud services everywhere, cloud/fog/edge, in-transit, SDN/NFV, IoT, …
- New technologies: accelerators, storage, communication, …
- New concerns: correctness, energy, fault tolerance, etc.



Entire printed collection of the US library of Congress is 10 Terabytes

1 Exabyte is 100,000 US Libraries of Congress

One BILLION BILLION operations per second by 2021!



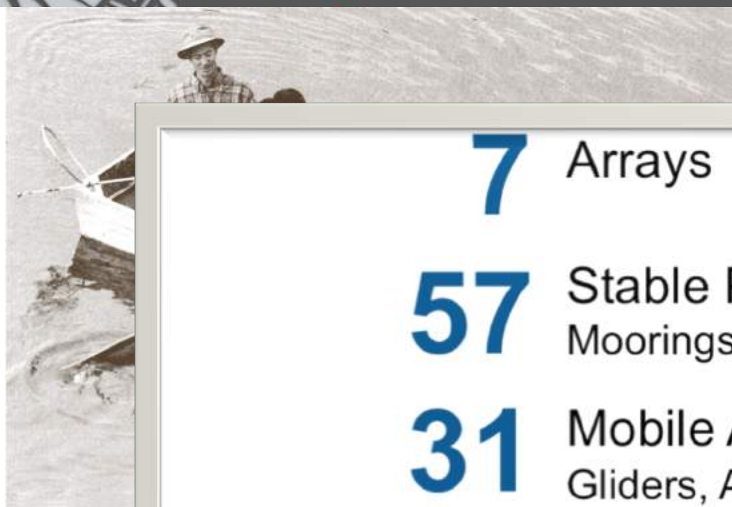
Science & Society Transformed by Data & Computing

- *Nearly every field discovery is transitioning from "data poor" to "**data rich**"*
- *The **scientific process** has evolved to include **computation** & **data***

Science and Engineering in 21st Century

- New paradigms and practices in science and engineering
- Inherently multi-disciplinary
- Extreme scales, data-driven, data and compute-intensive
- Collaborative (university, national, global)

RUTGERS DISCOVERY INFORMATICS INSTITUTE



NSF Ocean Observatories

7 Arrays

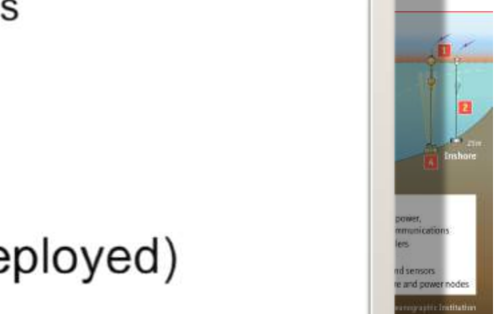
57 Stable Platforms
Moorings, Profilers, Nodes

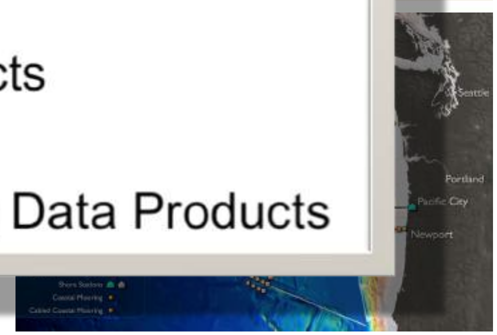
31 Mobile Assets
Gliders, AUVs

1227 Instruments (~850 deployed)

>2500 Science Data Products

>100K Science/Engineering Data Products





ooinet.oceanobservatories.org

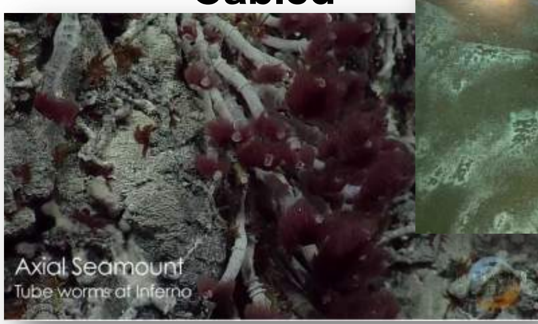
Credit: John Delaney, University of Washington

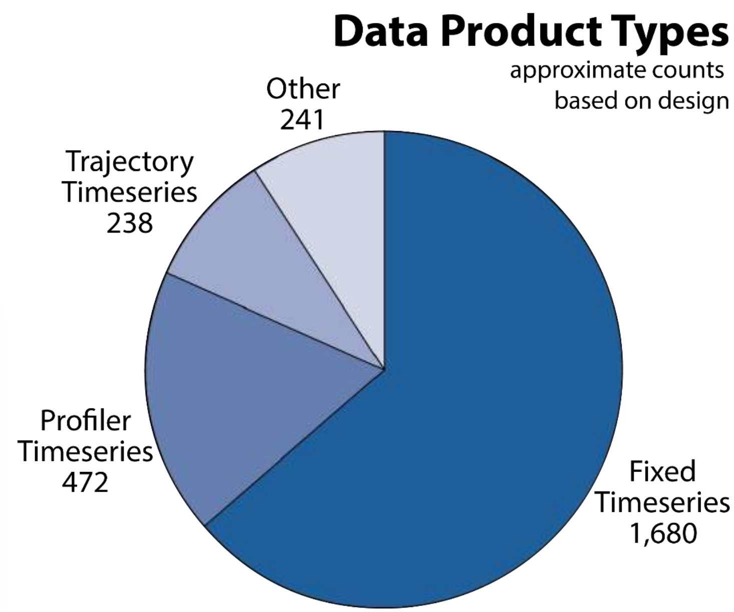
Types of Data

Telemetered

Recovered

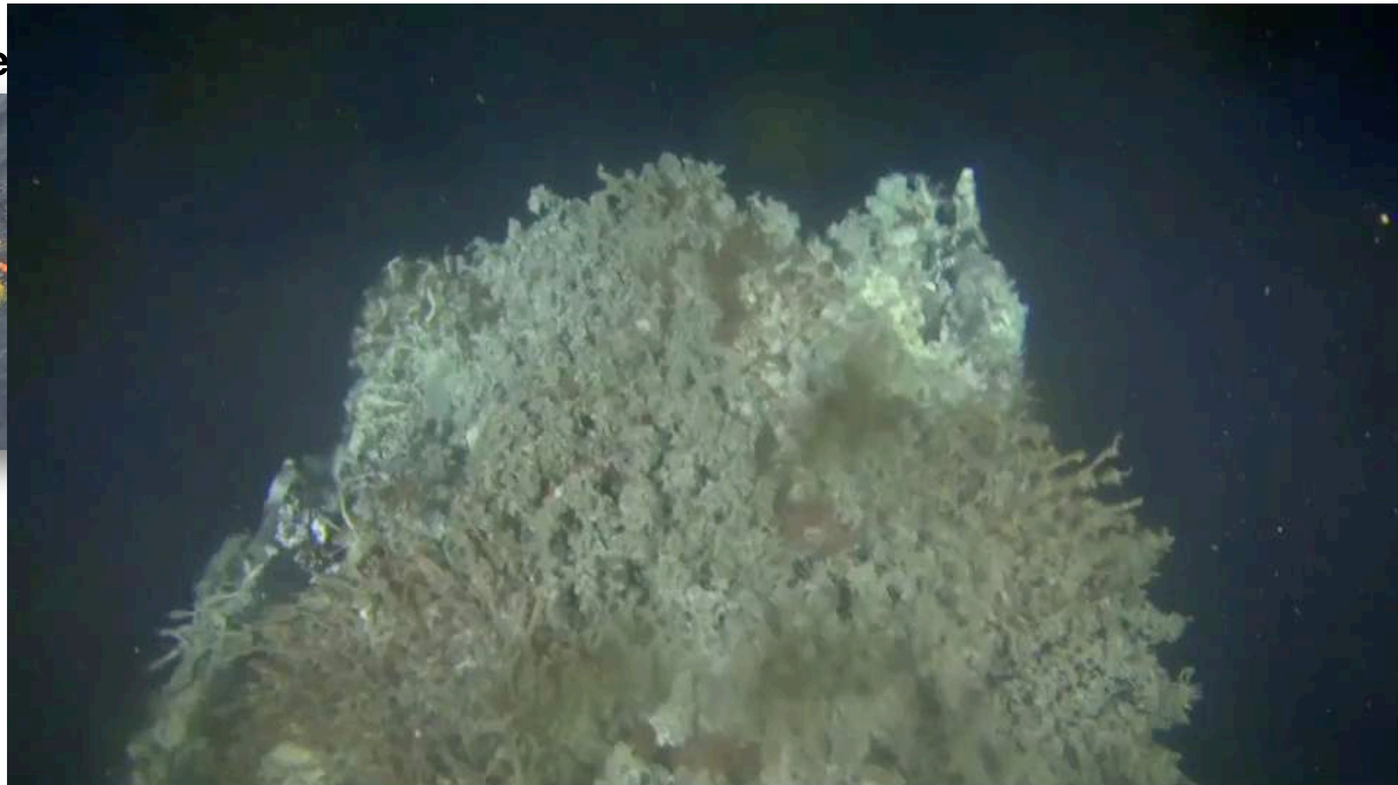
Cabled

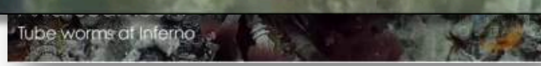




Types of Data

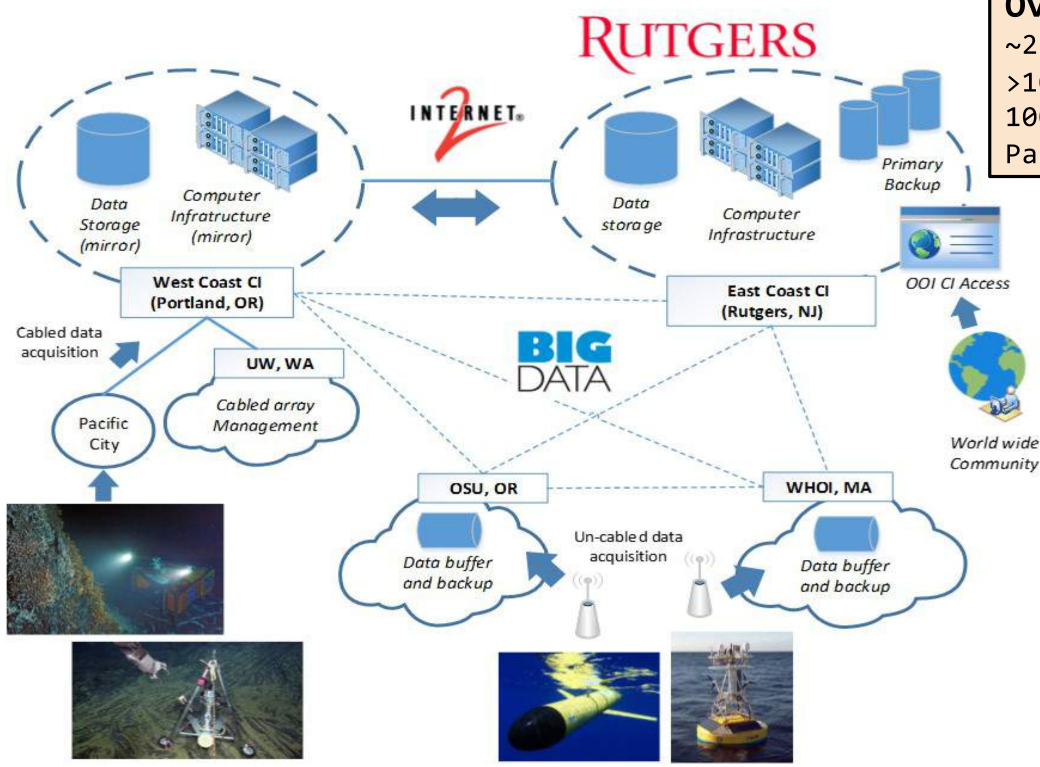
Teleme



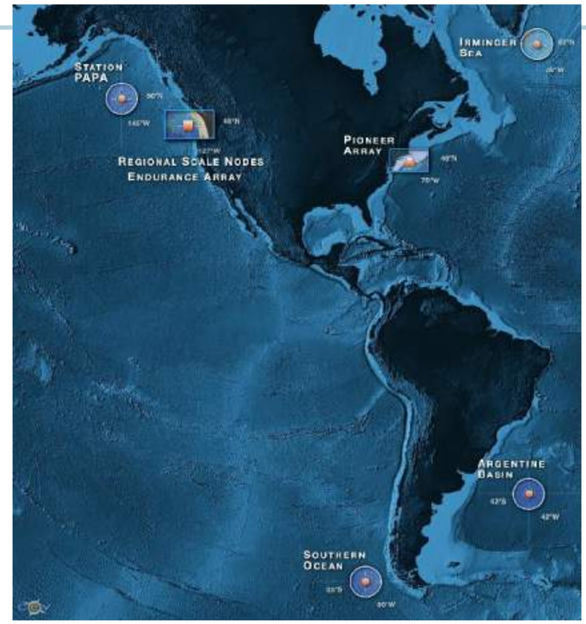


Product Types
approximate counts
based on design

OOI : Robust, Secure and Scalable Production CI



Overall Capacity:
~25PB storage
>100 servers (dual Xeon)
100G network backbone
Palo Alto Firewalls



- **Enterprise-level** system/network
- Industry **best practices**
- 24/7 delivery of data & services
- Integrated software stack
- Mirroring and **fail-over**
- Integrated **cyber-security**

Data-driven Workflows in Science & Engineering

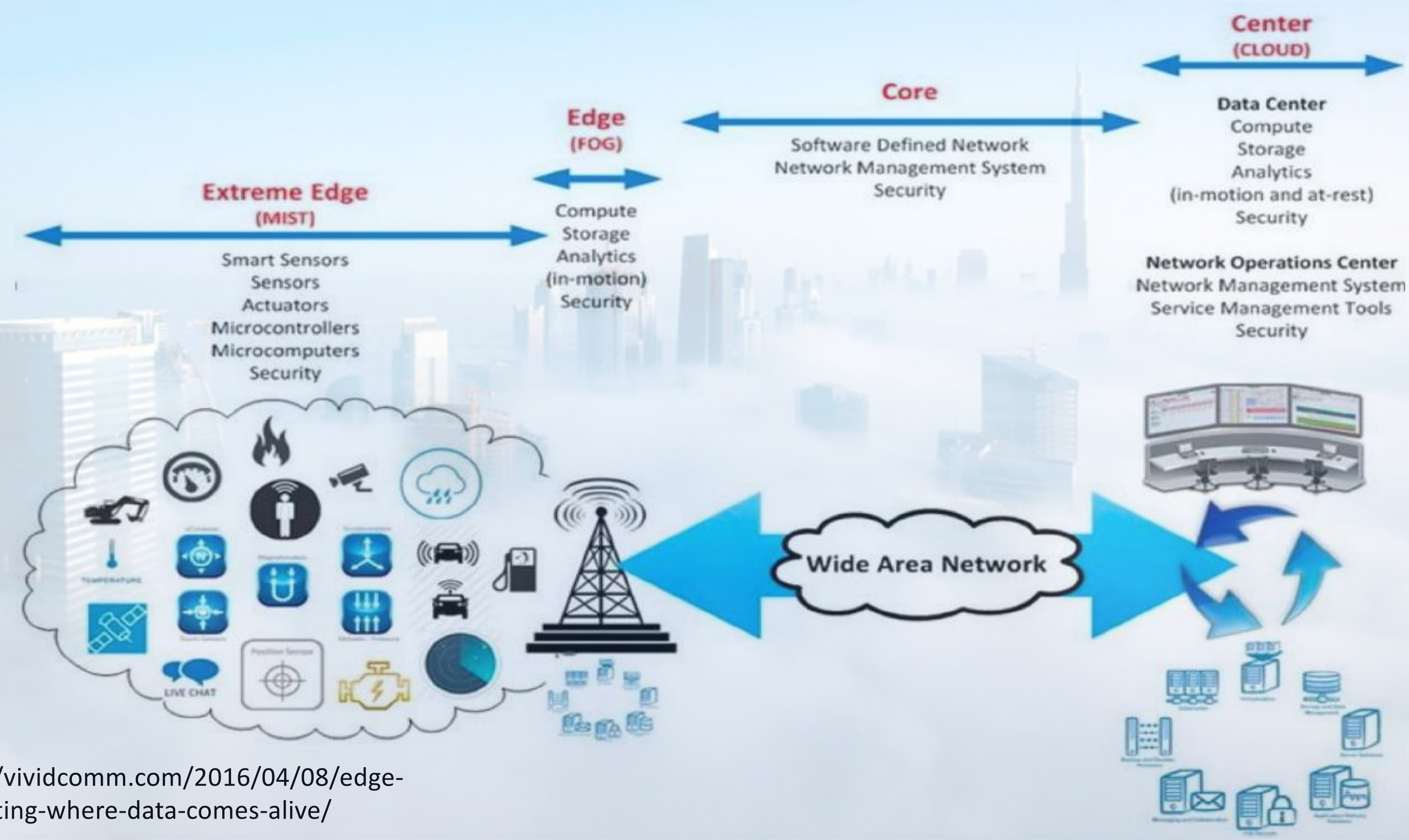
- Large-scale scientific observatories have become an essential part of science and engineering enterprise
  - Provide open, real time access to data from geographically distributed sensors and instruments
- Increasing scale, heterogeneity, and richness of data
  - Data downloads and local processing no longer feasible
- Largely disconnected from each other, from commercial/academic CI services
  - Integrating observatory data into science workflows is a growing challenge
  - New delivery modes for data and data-products essential

Emerging Requirements for End-to-End Data-driven Science

- (Near) Real-time delivery
  - Support (near) real-time delivery of raw and processed data to geographically distributed users
- Data processing
  - Provide on-demand computing capabilities that can process data in real-time
- Data transfer
  - Support fast data transfer between large-scale facilities and computing infrastructure
- Workflow description
  - Provide users with mechanism to describe the desired data and computation
- Automated orchestration
  - Orchestrate the entire data-to-discovery pipeline in automated manner

An Evolving Cloud/Services Landscape

An Evolving Cloud/Services Landscape



https://vividcomm.com/2016/04/08/edge-computing-where-data-comes-alive/

An Evolving Cloud/Services Landscape

Datacenter
- HW is the unit of scale

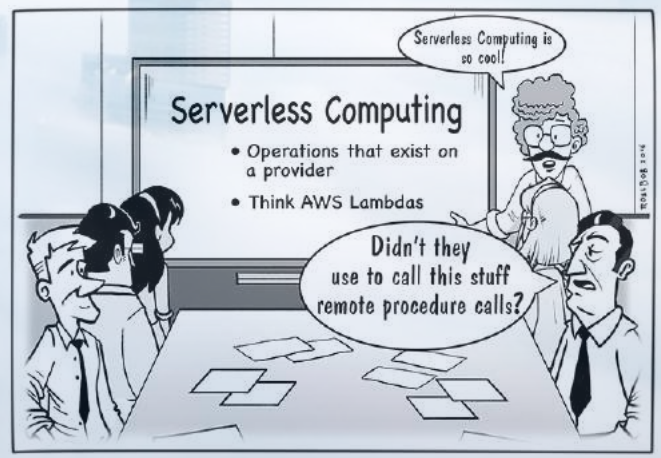
IaaS
- Cloud Datacenters / Software-defined Systems
- OS is the unit of scale

PaaS
- Containers / Microservices
- Application is the unit of scale

FaaS
- Serverless Computing
- Function is the unit of scale
- Event driven

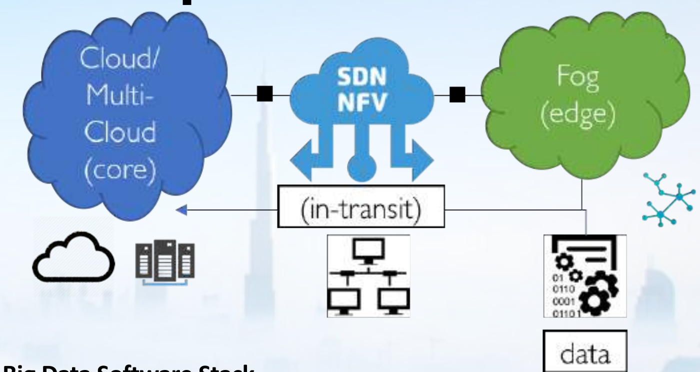
AWS Lambda	Azure Functions	Cloud Functions
Amazon	MicroSoft	Google

APACHE OpenWhisk™

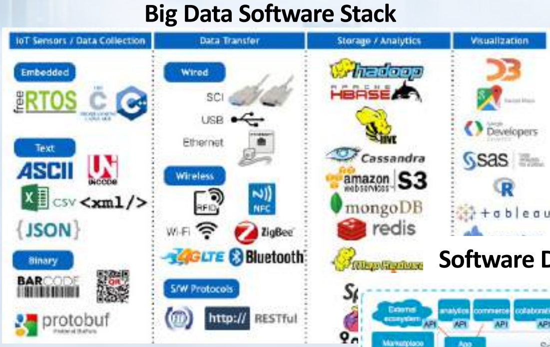


An Evolving Cloud/Services Landscape

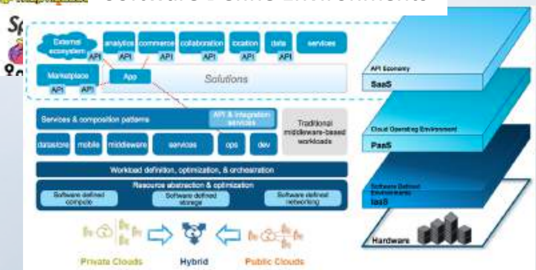
- Cloud
  - Hosted in data centers at the core
  - Relatively inexpensive; seemingly infinite
  - Far from data; data access expensive
- Fog/Edge
  - Computation/storage limited and expensive
  - Closer to the data; lower latencies
  - Limited and unreliable connectivity
- In-Transit
  - Distributed along the data path
  - Limited, but can be effective
  - Intermediate latency
  - Fewer guarantees



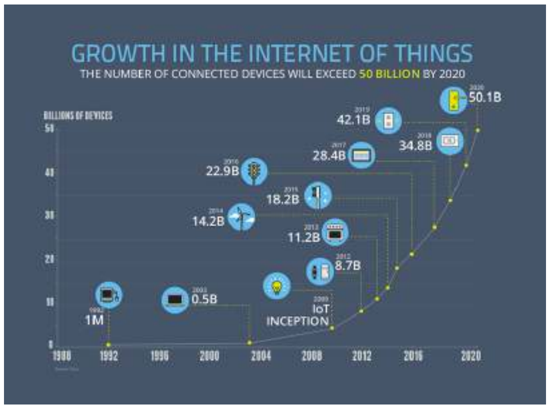
Big Data Software Stack

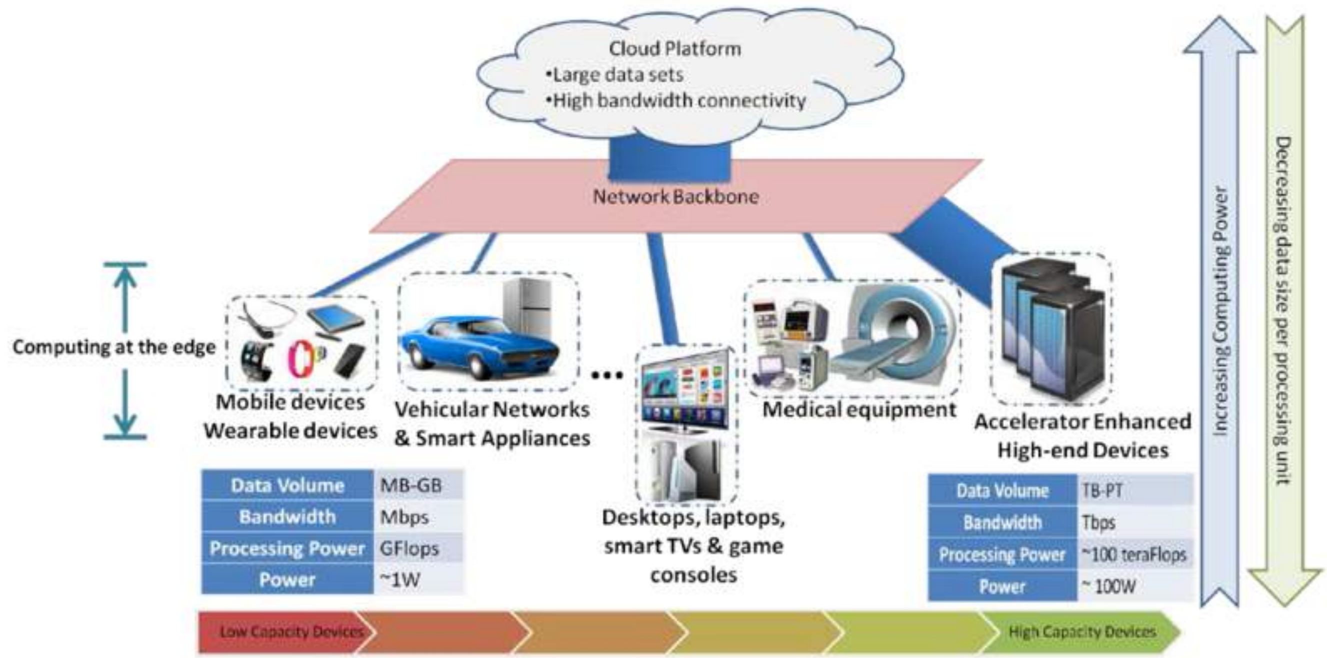


Software Define Environments



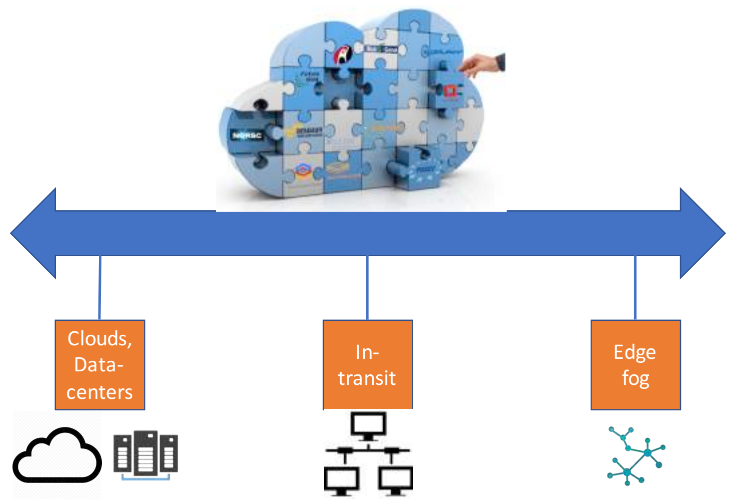
The New Reality – Dynamic, Data Driven!





Computing in the Continuum

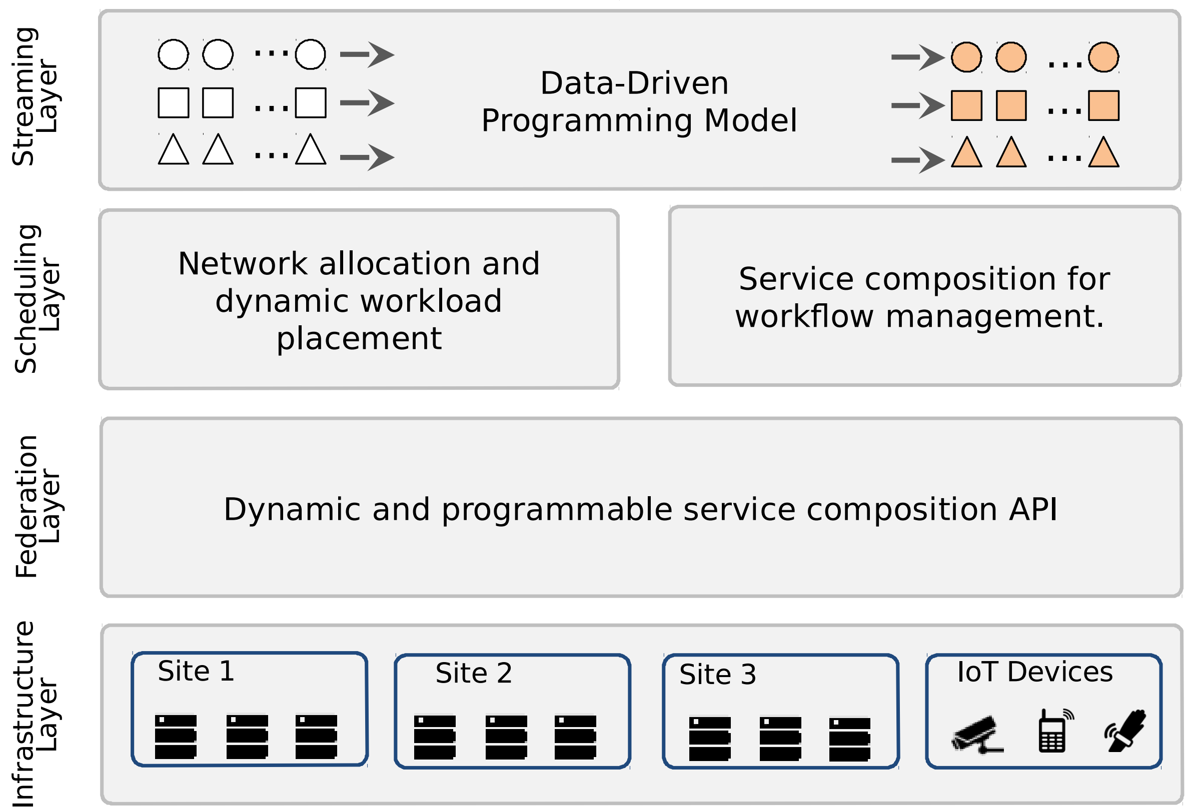
- Leverage resources and services at the logical extreme of the network and along the data path to increase the value of the data while potentially reducing costs
- Exploit the rich ecosystem of data and computation resources at the edge so that data is not moved
- Identify the **high levels of concurrency** that is pervasive throughout the ecosystem as the key to realizing scalable data-centric applications



Research Challenges

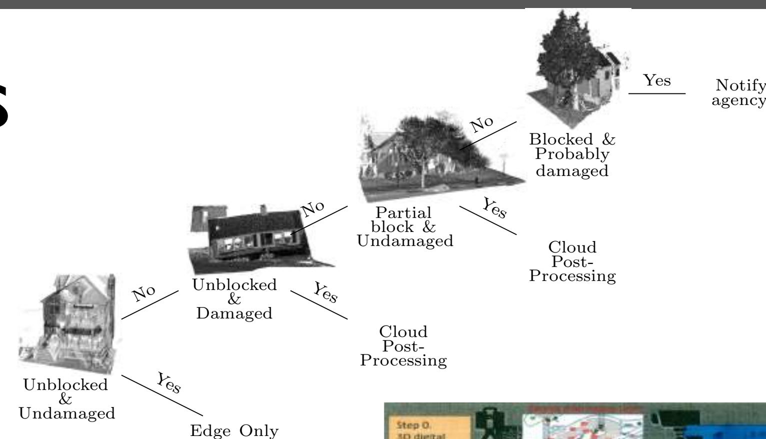
- How to drive computation through data
  - Express application behavior based on available data and its content
- How to accommodate uncertainties in data and computation
  - Move away from precise to approximate computing
- How to build applications and manage workflows so that they adapt to increase their 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 incorporate market models, social/trust models, and utility models
- …….

Computing in the Continuum – Some Research Efforts at RDI2

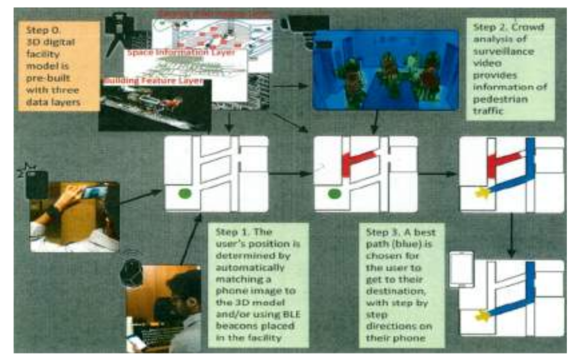


Some Driving Applications

1. Data-driven emergency response to structural damage of civil infrastructure



2. Enhancing mobility for autistic individuals using urban sensing



3. Scientific experiment management

Autonomic Management

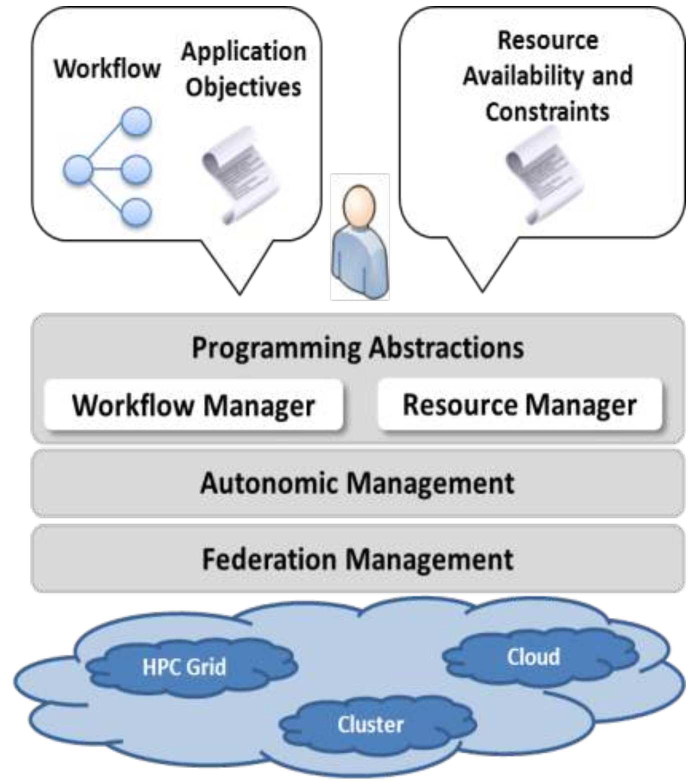
- Optimize resource provisioning and workload allocation to meet objectives and constraints set by users, applications, and/or resource providers
- Create models to translate resource/service capabilities and availabilities into application-level utilities (e.g., throughput, performance, etc.)
- Combine 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?

Some Recent Research

- Programmable/dynamic distributed infrastructure composition
- Workflow management and service composition across distributed infrastructure
- In-transit computing and data-optimized scheduling
- Data-driven programming models for streaming applications

Distributed Software-defined Environments

- Combine cloud service abstractions with ideas from software-defined environments
- Create a nimble and programmable ecosystem that autonomically evolves over time, adapting to:
  - Changes in the infrastructure
  - Application requirements
- Independent control over application and resources
- Enable efficient processing to transform data into actionable knowledge that drive critical decision making
  - Allocate computational power close to digital data sources
  - Process data in-situ and/or in-transit



M. Abdelbaky, J. Diaz and M. Parashar, "Enabling Distributed Software-Defined Environments Using Dynamic Infrastructure Service Composition," CCGrid'2017.
M. Abdelbaky and M. Parashar, "A General Performance and QoS Model for Distributed Software-Defined Environments", IEEE TSC 2019.

Distributed Software-Defined Environments

Applications & Workflows

- Description and dependency
- Workload behavior
- QoS objectives (e.g., deadline, budget)

- varies at runtime -

Autonomic Management

- Identify utility of composition
- Ensure objectives are always met
- Adapt & reconfigure services

- on the fly -

User/Resource Provider

Define composition programmatically using rules and/or constraints

- Availability
- Application requirements (e.g., type of resources, memory, I/O rate)
- Cost
- Location
- Access policy

- varies at runtime -

Exposed as a service to applications & workflows

Synthesize a space-time ecosystem

For Example….

- Containerized a synthesized cancer bioinformatics workflow
- Ran across 5 different clouds in 7 different regions using 15 different types of resource classes, 110 VMs
- Deployed up to 7000 containers
- Varying workloads
- Varying resource availabilities, properties, and constraints

Site Name & VM Type	# Cores	Max. VMs†	Speedup	Cost ⇑
AWS east t2.micro	1	10	2.39	0.013
AWS east t2.small	1	10	2.39	0.026
AWS east t2.medium	2	10	3.35	0.052
AWS east t2.large	2	10	3.47	0.104
AWS west t2.micro	1	10	2.52	0.013
AWS west t2.small	1	10	2.33	0.026
AWS west t2.medium	2	10	3.45	0.052
AWS west t2.large	2	10	3.47	0.104
Chameleon m1.small	1	8	2.49	0.026
Chameleon m1.medium	2	6	3.99	0.052
Chameleon m1.large	4	4	5.87	0.209
Azure east Standard-A1	1	3	1.00	0.044
Azure west Standard-D1	1	3	1.70	0.077
Google east n1-standard-1	1	3	2.40	0.05
IBM Bluemix	N/A	3∗	N/A	0.028
Dell cluster	8	12	N/A	N/A

Note: The # of containers per instance = # of cores per instance.
∗ Max number of containers for Bluemix.
† – Maximum number of available VMs per type.
⇑ – Real cost ($) per hour for all cloud providers except Chameleon, which was estimated base on AWS pricing.

RUTGERS DISCOVERY INFORMATICS INSTITUTE

Experiment: Dynamic Workload & Infrastructure

Triggering Event	Time (min)

=0.09

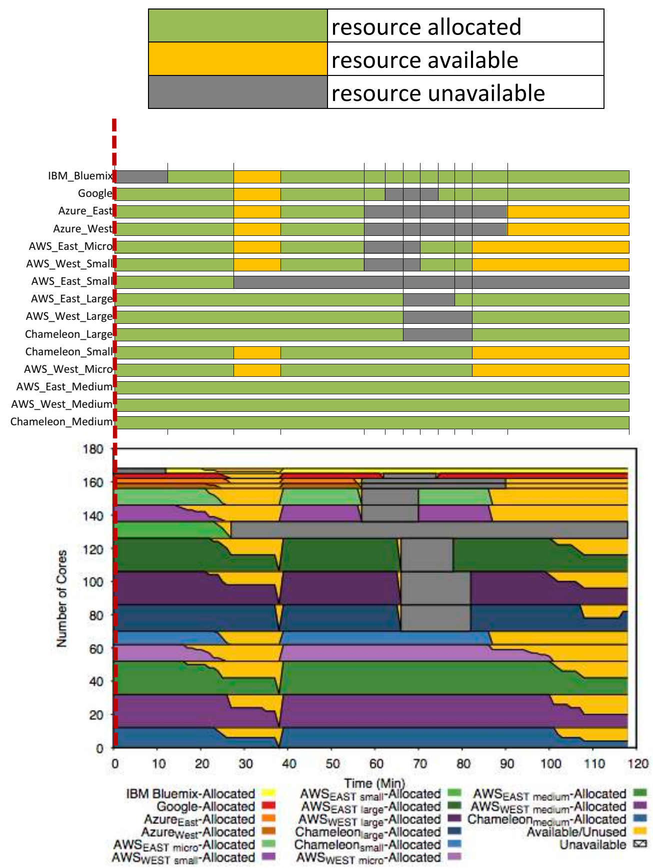
Available Resource Classes

Selected Resource Classes

Current Total Number of Cores

Details

workload size



End-to-End Data Processing and Delivery in Science and Engineering

- We cannot process data within the constraints in the core of the network due to the data movement costs.
- There are capabilities in the network and at the edge that can be leveraged.
- We want to develop solutions to leverage these resources.

- Q1) How to exploit and provision resources that exist in the continuum but are not being used or leveraged?
- Q2) How to leverage application's characteristics to meet end-to-end QoS? How to increase quality of solution using edge and in-transit nodes?
- Q3) How to schedule workflows on the evolving computing landscape to end-to-end QoS? How to manage QoS at runtime using edge and in-transit resources?

Key Research Ideas (Ali Reza Zamani's Thesis)

- Using SDN technology to provision in-transit resources
- Partially process data using edge and in-transit resources

[C1]: CCGrid 2015
[C2]: SOSE 2016
[J1]: IEEE TSC 2017
[C3]: ICDCS 2017

- Increase accuracy of the approximate computing techniques using edge and in-transit resources

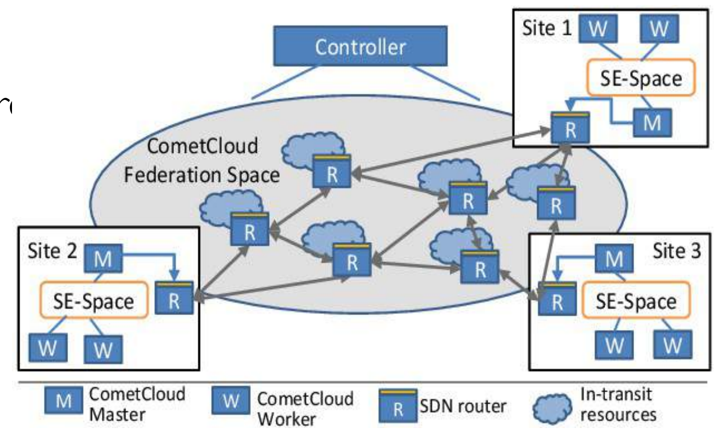
[C4]: e-Science 2017
[C5]: FiCloud 2017
[J2]: FGCS 2018
[C6]: UCC 2018

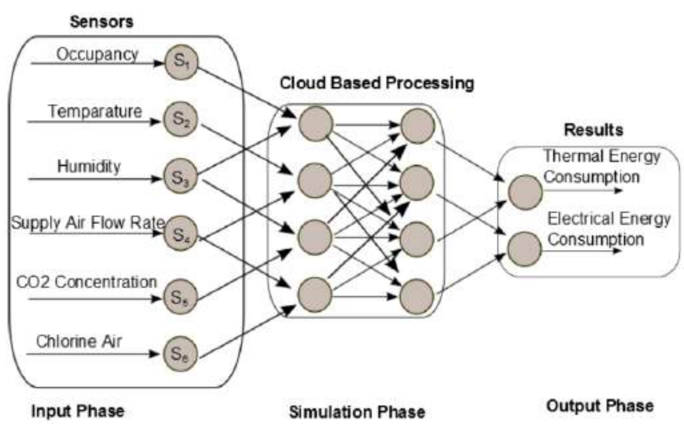
- Scheduling workflow stages over edge and in-transit nodes
- QoS monitoring at edge and in-transit nodes

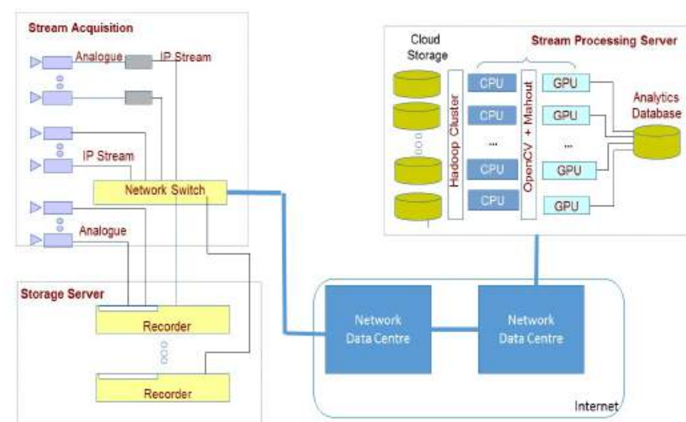
[C7]: SBAC-PAD 2018
[J3]: CCPE 2019
[J4]: FGCS 2019

In-transit Computing (with O. Rana, I. Petri, Cardiff)

- Use computational capabilities along the path
  - Leverage excess (idle) resources within the network infrastructur
  - Enable general purpose computation in the network data center
- Extend SDN controllers to manage resources at network data centers
  - Resource provisioning/allocation for in-transit processing
  - Explore approximations in-transit



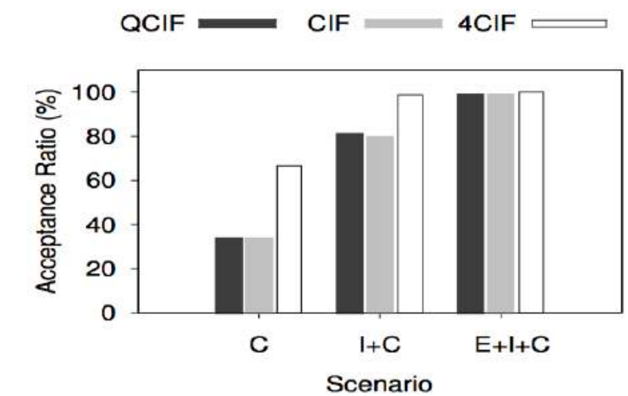


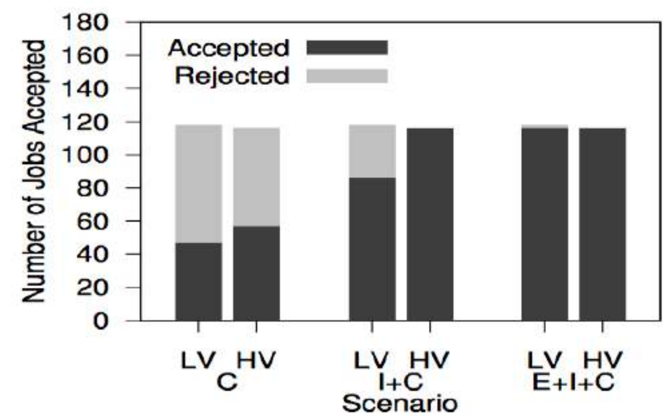


A. Zamani, M. Zou, J. Diaz-Montes, I. Petri, O. F. Rana, A. Anjum and M. Parashar, "Deadline Constrained Video Analysis via In-Transit Computational Environments", IEEE Transactions on Services Computing. 2017.

Near real-time processing at the Edge and In-transit

- Objective:
  - Leverage in-transit services to improve QoS.
- Workload: Video surveillance
- Experiment:
  - 3 cameras capturing video with 3 different qualities: QCIF, CIF and 4CIF
  - 3 scenarios: C (core), I+C (in-transit + core) , E+I+C (edge, in-transit, core)
  - Video chunks may high or low values
  - Edge resources sample video frames based on value
- Results:
  - Using in-transit resources located along the data path from source to destination, the infrastructure was able to significantly increase the number of accepted jobs.
  - High value jobs receive premium service





Edge-based Approximation

- "Performing exact computation or operating at peak-level service demand require a high amount of resources, allowing selective approximation or occasional violation of the specification can provide disproportionate gains in efficiency."

- Combine capability in the Data Centre with "approximate" algorithms in-transit or at the edge
  - EnergyPlus + a trained neural network (as a function approximator for EnergyPlus behaviour)

- But why?
  - EnergyPlus ~ Execution time (Minutes)
  - Neural Network Training ~ Execution time (Minutes)
  - Trained (FF) Neural Network ~ Execution time (Seconds)

- Trigger re-training when input parameters change significantly
  - Each EnergyPlus execution provides potential training data for the neural network

Edge processing and Approximation

Scenario	Edge Resources	Site Resources
Base	c4.2xlarge	c4.2xlarge
Higher	c4.2xlarge	c4.4xlarge
Highest	c4.2xlarge	c4.8xlarge

Objectives:

- Execute jobs within deadline and budget
- Minimize waiting times at the cloud
- Use approximation when resource constrained: loop reduction, parameter interval reduction. parameter value skipping, ANN

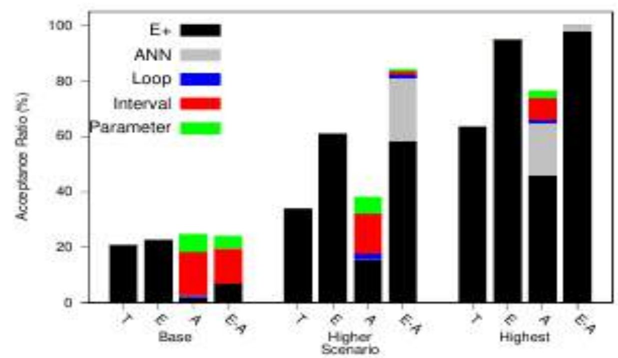
Workload: EnergyPlus workflow

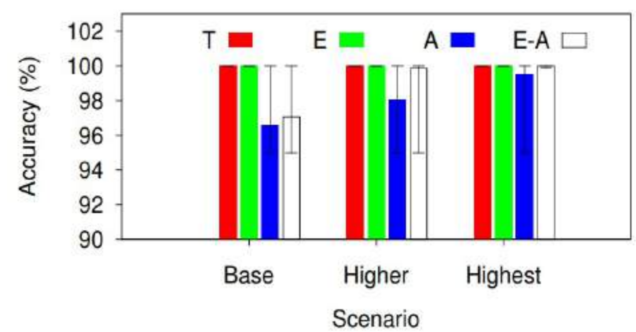
Experiment:

- Four strategies: Traditional (process at core), Edge-enabled, Approximation, Edge-enabled + approximation

Results:

- Job acceptance ratio increases with availability of edge and approximation.
- Edge resource contribute to increase the accuracy of the approximation techniques.



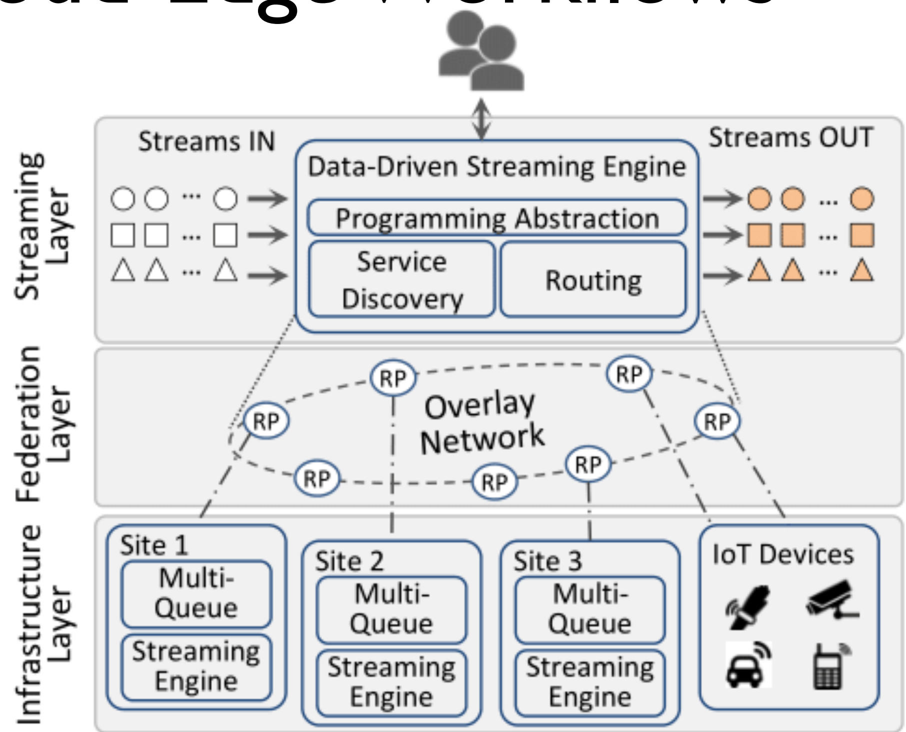


A. Zamani, I. Petri, J. Diaz-Montes, O. F. Rana and M. Parashar, "Edge-supported Approximate Analysis for Long Running Computations", To appear in FiCloud 2017.

A. Zamani, M. Zou, J. Diaz-Montes, I. Petri, O. Rana, and M. Parashar. "A computational model to support in-network data analysis in federated ecosystems." FGCS 80 (2018): 342-354.

R-Pulsar: Programming Cloud-Edge Workflows

- Data-drive Stream Processing: *Location, Content, Quality, Uncertainty* aware
- Content-based, event-driven specification of which topology to execute; where to execute
- Builds on the *Associative Rendezvous* paradigm for content-based decoupled interactions with programmable reactive behaviors
- Implementation based on Storm and Kafka



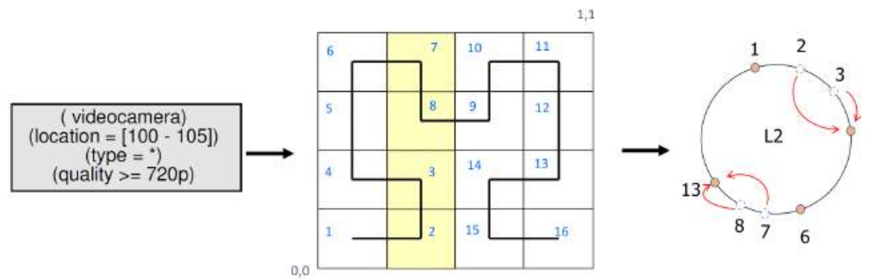
Software for Raspberry Pi and Android
https://github.com/egibert/Rutgers-Pulsar

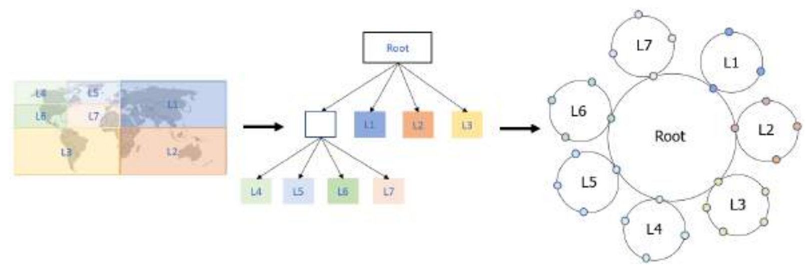
R-Pulsar – Key Ideas

- Data/Resources described using locally defined profiles
- Associative (in-memory) DHT / Pub/Sub messaging / Structured overlays
- Reactive behaviors leveraging serverless computing
  - Time/Energy/Accuracy tradeoffs

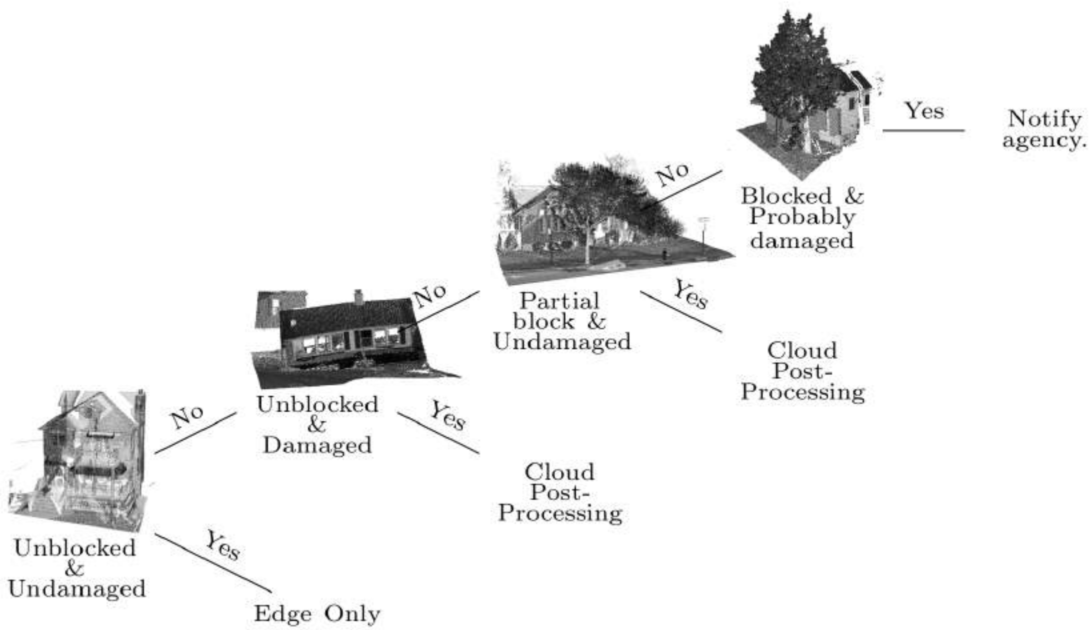
(videocamera)
(location = [100, 101, 102])
(type = Public)
(quality = 720p)

(videocamera)
(location = [100 - 105])
(type = *)
(quality >= 720p)





Use case: Disaster Response



Content-driven disaster response workflow decision stages.

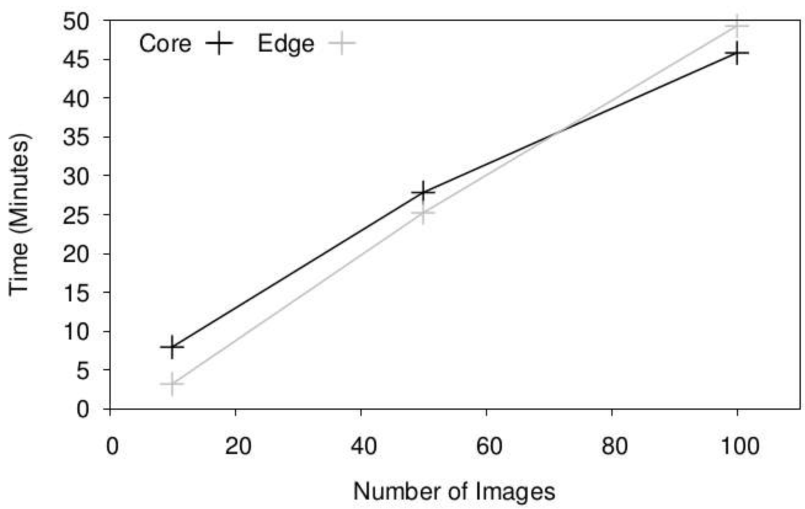
(With Xuan Hu, Jie Gong, Dept. of Civil Engr., Rutgers)

Need to quickly and efficiently determine whether the building conditions are safe or not for evacuees to return.

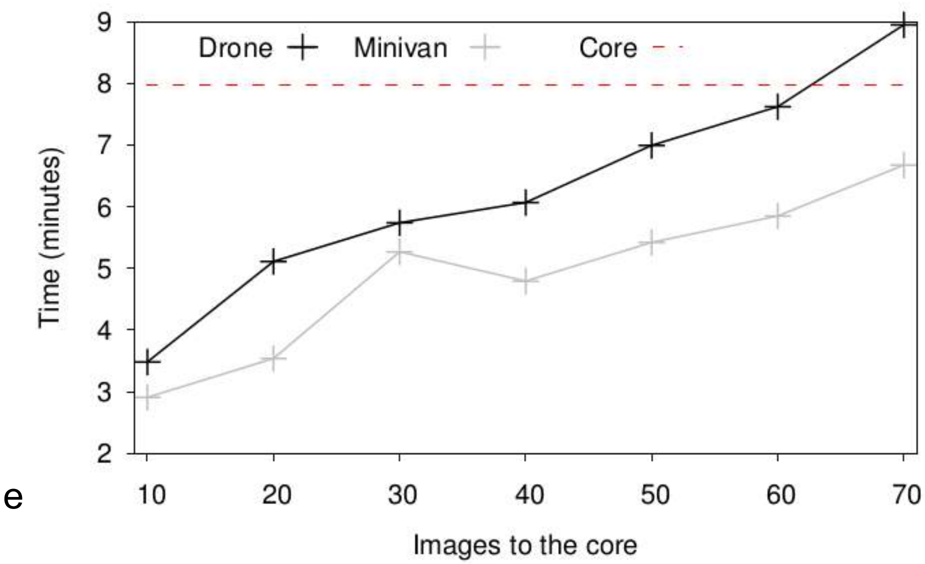
Sending images to the core is expensive, also not all images contain useful information.

Sample workflow consisted of 3.7 GB of LiDAR images (largest = 33.8 MB, smallest = 1.8 KB)

Experimental Results

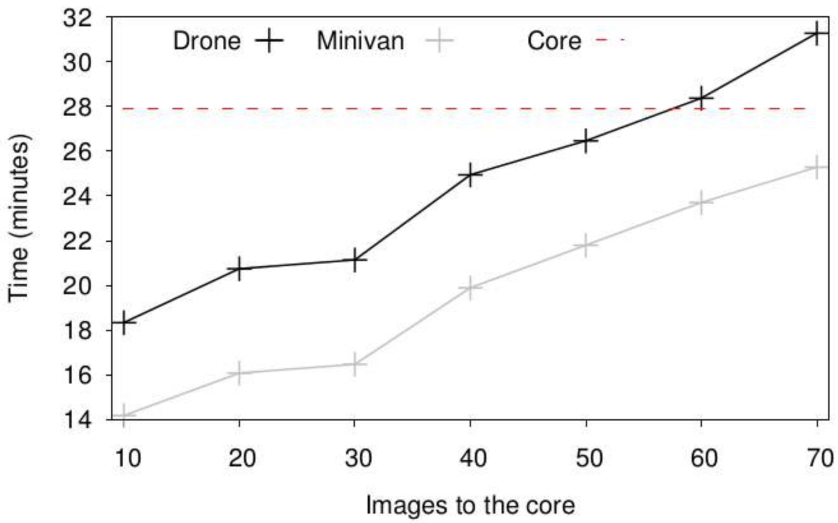


- Naïve implementation without any data "quality" trade-off we see that as the computation grows (the affected area is large) the core performs better than the edge due to limited resources at the edge of the network.

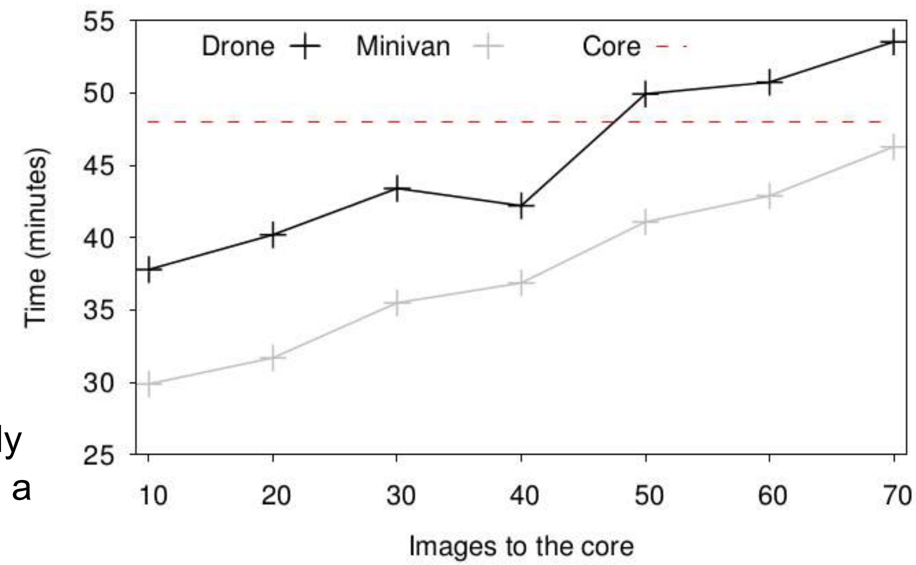


- Disaster response workflow with 10 images. When only 10% of the images need further processing we observe a speed up of 71% compared to sending all 10 images to the core of the network

Experimental Results



- Disaster response workflow with 50 images. When only 10% of the images need further processing we observe a speed up of 50%.



- Disaster response workflow with 100 images. When only 10% of the images need further processing we observe a speed up of 34%.

Towards Intelligent Data Delivery

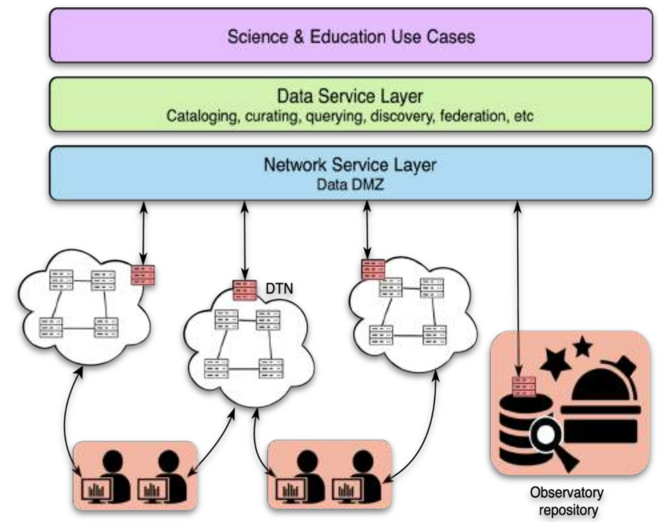
- Performance degradation at repository
  - While processing requests: Many concurrent requests
  - During data transfers: Network bandwidth limitation / contention

- Performance degradation experienced by users
  - Request response latency
  - Slow data retrieval speed; data transfer
  - Data sizes are large

- Data discovery is hard
  - Data heterogeneity / siloed repositories
  - Lack of data recommendation tool

Proposed approach (work in progress)

- Leverage in-network processing/storage (e.g., DTNs)
- Leverage locality
  - Spatial and temporal
  - Domain affinities
  - Cross-domain / cross-facility affinities
- Develop knowledge networks, recommendation engines



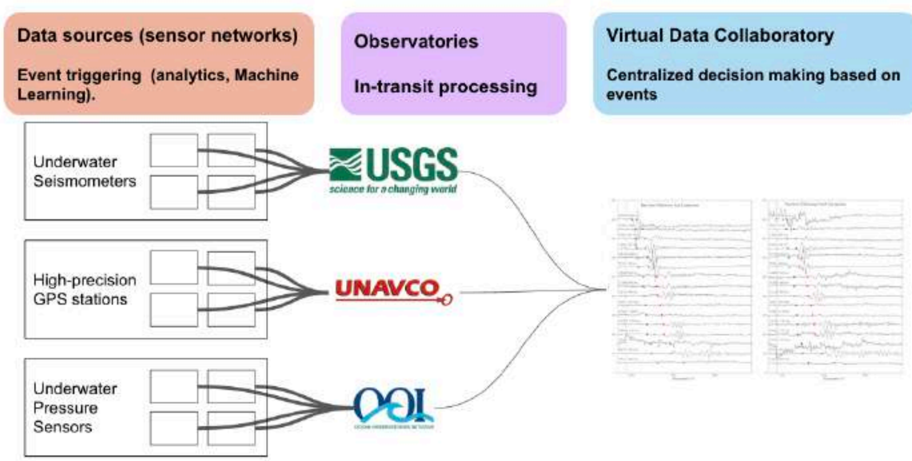
Virtual Data Collaboratory Platform

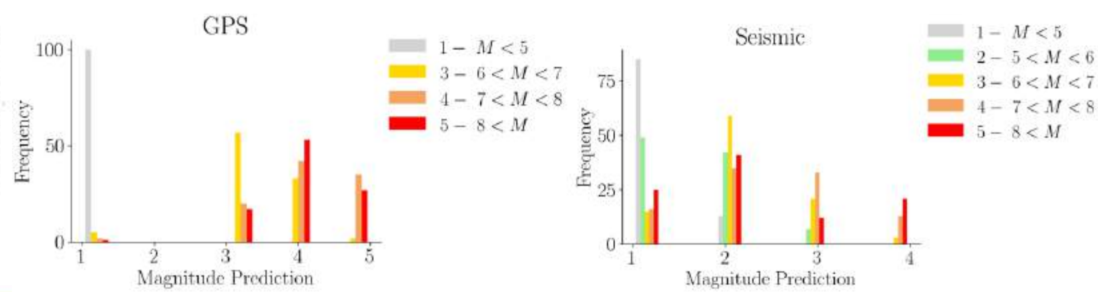
Use-case: Tsunami Early Warning

Increase precision and delay for Tsunami warning by analyzing multiple geo-graphically-distributed data sources simultaneously, in collaboration with UNAVCO.

To issue Tsunami Early Warnings, earthquakes must first be characterized (magnitude, location, speed of displacement, etc.).
Seismometers are good for the ***smaller* earthquakes** (< 6.5), **high-precision GPS** are good for ***larger* earthquakes**.
Goal: combining multiple data sources to improve the precision and delay to issue warnings by covering the **whole spectrum** of events.

# 60s MTS	GPS (# Events)	Seismic (# Events)
Magnitude < 5	7,718 (170)	1,038 (349)
$5 \leq$ Magnitude < 6	3,859 (85)	None
$6 \leq$ Magnitude < 7	991 (4)	266 (4)
$7 \leq$ Magnitude < 8	432 (6)	249 (6)
Magnitude > 8	265 (4)	133 (4)
Total	13,265 (269)	1,686 (363)





Some Related Projects

- ExoGENI
  - A framework that orchestrates a federation of independent cloud sites located across the US and circuit providers (e.g., NLR and Internet2) through their native IaaS API interfaces, and links them to other GENI tools and resources
- FELIX
  - A federation framework that allow users to build their own virtual slices using resources of remote Internet facilities
- CloudLab/Chameleon
  - Experimental meta-cloud environments that provide bare-metal/vm access and control across multiple university sites
- Atos
  - Multi-cloud service integration
- DITAS
  - Simplified logistics/data-management for cloud and edge environments
- Fluid Computing
  - Replication and real-time synchronization of application states on several devices

Summary

- Data-driven/data-intensive reality
  - Large scale, heterogeneous in nature and geographic location
  - Large volumes of data that need 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…
- Many challenges accross multiple layers
  - Application formulation, programming systems, middleware services, standardization & interoperability, autonomic engines, etc.

Thank You!

M. AbdelBaky, D. Balouek-Thomert, E. Renart, A. Zamani, M. Zou, M. Parashar

Rutgers Discovery Informatics Institute (RDI[2])

Rutgers, The State University of New Jersey

Email: parashar@rutgers.edu

WWW: http://parashar.rutgers.edu/

CometCloud: http://cometcloud.org