RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

Exploring Software Defined Federated Infrastructures for Science

Moustafa Abdelbaky, Javier Diaz-Montes, and Manish Parashar

NSF Cloud and Autonomic Computing Center (CAC)
Rutgers Discovery Informatics Institute (RDI^2)
Rutgers, The State University of New Jersey
http://parashar.rutgers.edu/

CometCloud

Outline

- Federated computing, software defined systems, …. and Science
- Initial explorations with dynamic federation using CometCloud
- Towards a software-defined federated infrastructure for science
- Summary / Conclusion

FEDERATED COMPUTING, SOFTWARE DEFINED SYSTEMS

The Lure of Clouds

- An attractive platform for supporting the computational and data needs of academic and business applications
- The Cloud paradigm:
  - "Rent" resources as cloud services on-demand and pay for what you use
  - Potential for scaling-up/down/out as well as for IT outsourcing
- Landscape of heterogeneous cloud services spans private clouds, public clouds, data centers, etc.
  - Novel dynamic Marketplaces - Heterogeneous offering with different QoS, pricing models, geographical locations, availability, capabilities, and capacities
- Cloud federations extend as-a-service models to virtualized data-centers federations

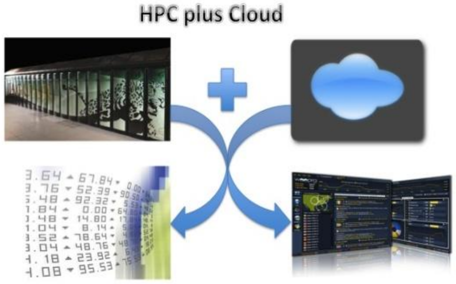
Clouds as Enablers of Science

- Clouds are rapidly joining traditional CI as viable platforms for scientific exploration and discovery
- Possible usage modes:
  - Clouds can simplify the deployment of applications and the management of their execution, improve their efficiency, effectiveness and/or productivity, and provide more attractive cost/performance ratios
  - Cloud support the democratization
  - Cloud abstractions can support new classes of algorithms and enable new applications formulations
  - Application driven by the science, not available resources
- Many challenges
  - Application types and capabilities that can be supported by clouds?
  - Can the addition of clouds enable scientific applications and usage modes that are not possible otherwise?
  - What abstractions and systems are essential to support these advanced applications on different hybrid platforms?

Cloud Usage Modes for Science

- ***HPC in the Cloud*** – outsource entire applications to current public and/or private Cloud platforms

- ***HPC plus Cloud*** – Clouds complement HPC/Grid resources with Cloud services to support science and engineering application workflows, for example, to support heterogeneous requirements, unexpected spikes in demand, etc.



- ***HPC as a Cloud*** – expose HPC/Grid resources using elastic on-demand Cloud abstractions

See Parashar et al, "Cloud Paradigms and Practices for Computational and Data-Enabled Science and Engineering" IEEE CiSE 15, 10 (2013)

Federated Computing for Science (I/II)

- Scientific applications can have large and diverse compute and data requirements

- Federated computing is a viable model for effectively harnessing the power offered by distributed resources
  - Combine capacity, capabilities

- HPC Grid Computing - monolithic access to powerful resources shared by a virtual organization
  - Lacks the flexibility of aggregating resources on demand (without complex infrastructure reconfiguration)

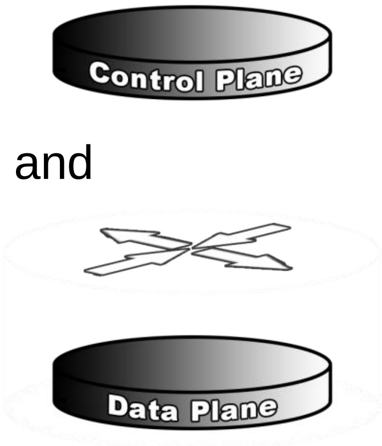
- Volunteer Computing - harvests donated, idle cycles from numerous distributed workstations
  - Best suited for lightweight independent tasks, rather than for traditional parallel computations

RUTGERS

Federated Computing for Science (II/II)

- Current/emerging science and engineering application workflow exhibit heterogeneous and dynamic workloads, and highly dynamic demands for resources
  - Various and dynamic QoS requirements
    - Throughput, budget, time
  - Unprecedented amounts of data
    - Large size, heterogeneous nature, geographic location
- Such workloads are hard to efficiently support using classical federation models
  - Rigid infrastructure with fixed set of resources
- Can we combine the best features of each model to support varying application requirements and resources' dynamicity?
  - Provisioning and federating an appropriate mix of resources on-the-fly is essential and non-trivial

Software Defined ….



- Software Defined Networks
  - An approach to building computer networks that separates and abstracts elements of these systems (Wikipedia)
  - E.g., separation of control and data plane
- Software Defined Systems
  - Based on software defined networking (SDN) concepts
  - Allow business users to describe expectations from their IT in a systematic way to support automation
  - Enable the infrastructure to understand application's needs through defined policies that control the configuration of compute, storage, and networking, and it optimizes application execution
    - Open virtualization, Policy driven optimization and elasticity – autonomics, Application awareness
- See also software defined data centers, ….

EXPLORING FEDERATED INFRASTRUCTURE FOR SCIENCE USING COMETCLOUD

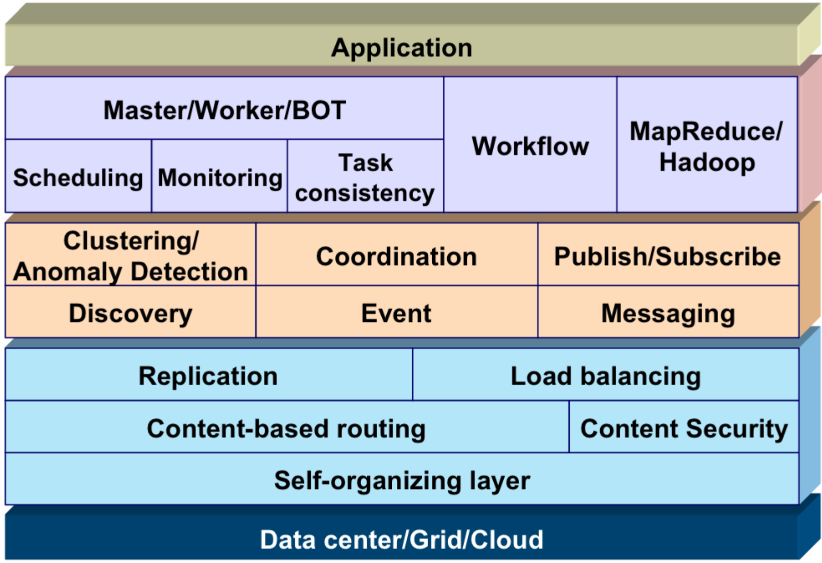
CometCloud

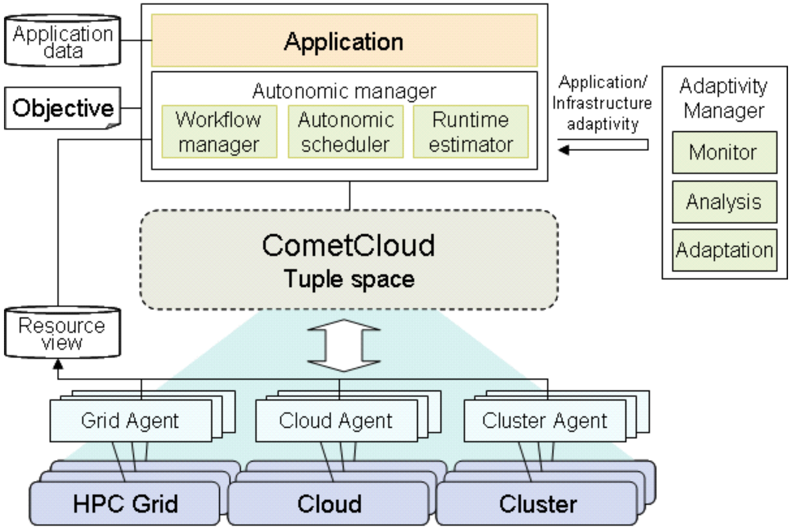
- Enable applications on dynamically federated, hybrid infrastructure exposed using Cloud abstractions
  - **Services:** discovery, associative object store, messaging, coordination
  - **Cloud-bursting:** dynamic application scale-out/up to address dynamic workloads, spikes in demand, and extreme requirements
  - **Cloud-bridging:** on-the-fly integration of different resource classes (public & private clouds, data-centers and HPC Grids)

- High-level programming abstractions & autonomic mechanisms
  - Cross-layer Autonomics: Application layer; Service layer; Infrastructure layer

- Diverse applications
  - Business intelligence, financial analytics, oil reservoir simulations, medical informatics, document management, etc.

http://cometcloud.org

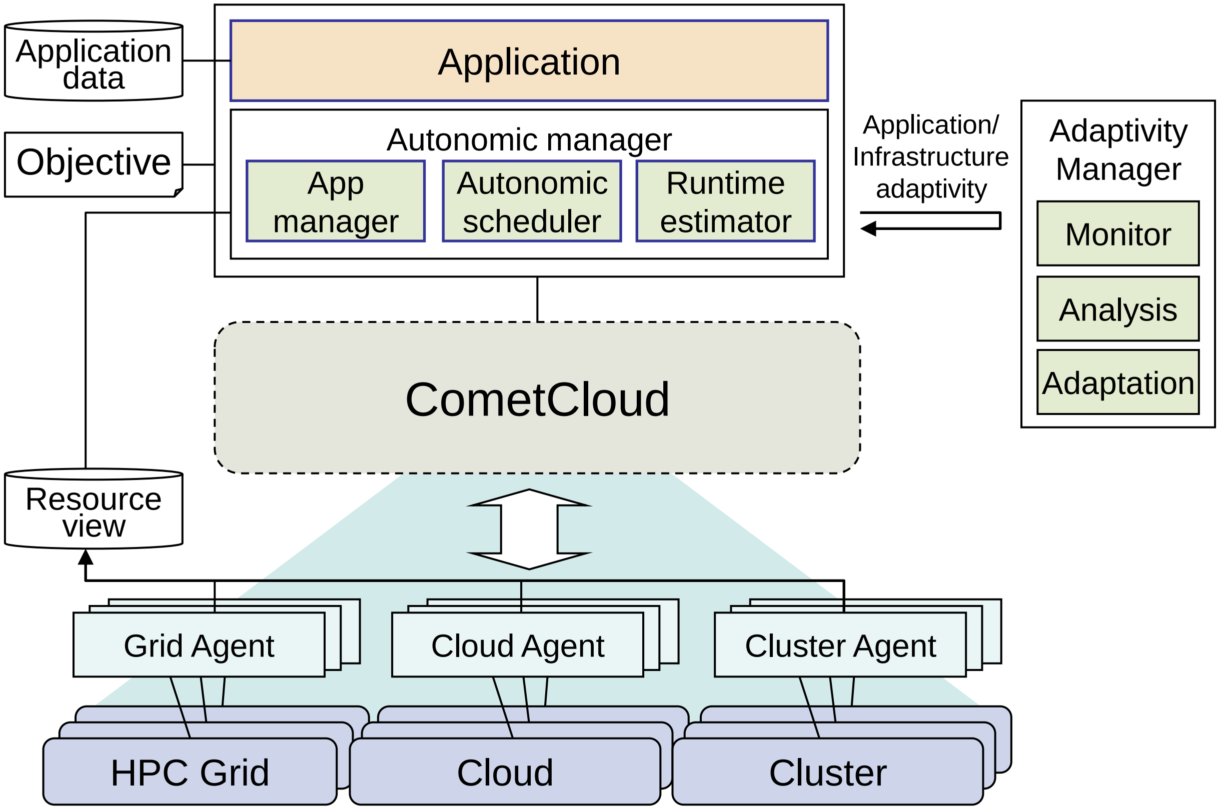




Federated (hybrid) computing infrastructure

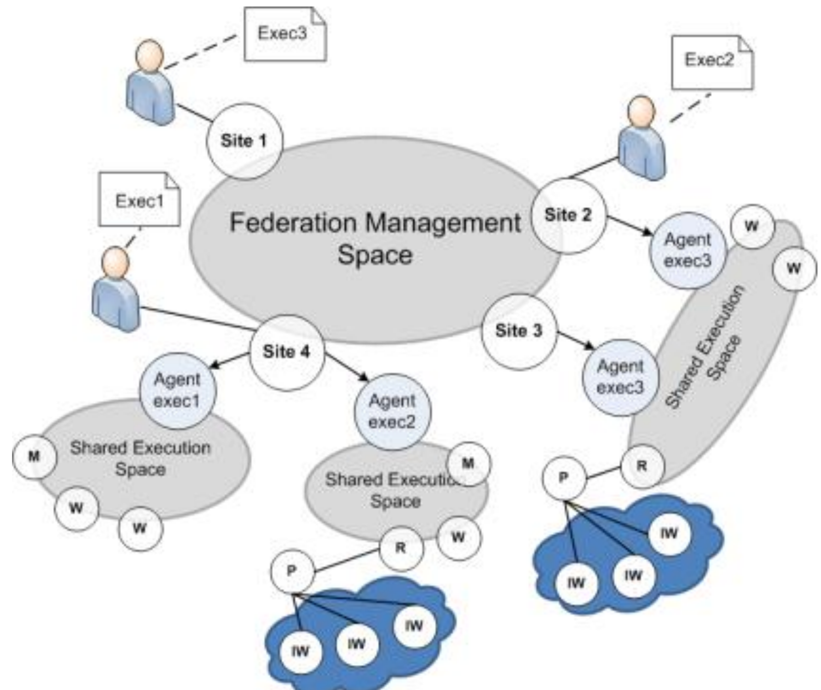
Autonomics in CometCloud

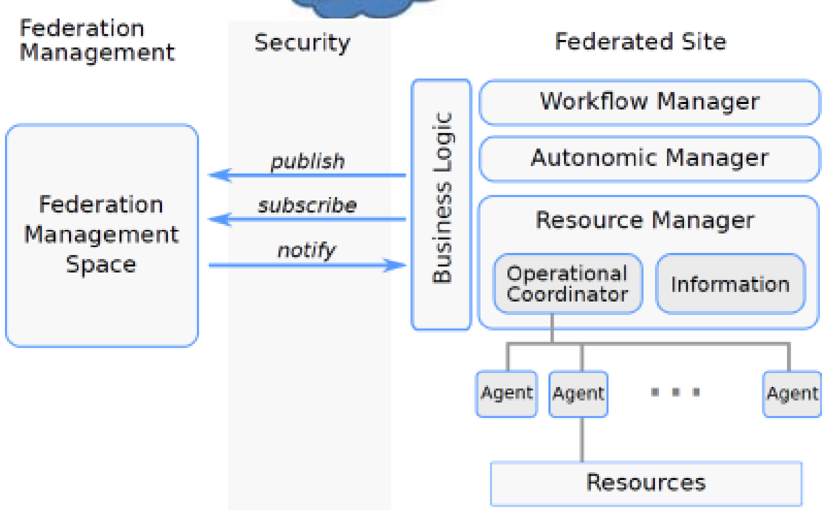
- **Autonomic manager** manages workflows, benchmarks application and provision resources.
- **Adaptivity manager** monitors application performance and adjusts resource provisioning.
- **Resource agent** manages local cloud resources, accesses task tuples from CometCloud and gathers results from local workers so as to send them to the workflow (or application) manager.



On-Demand Elastic Federation using CometCloud

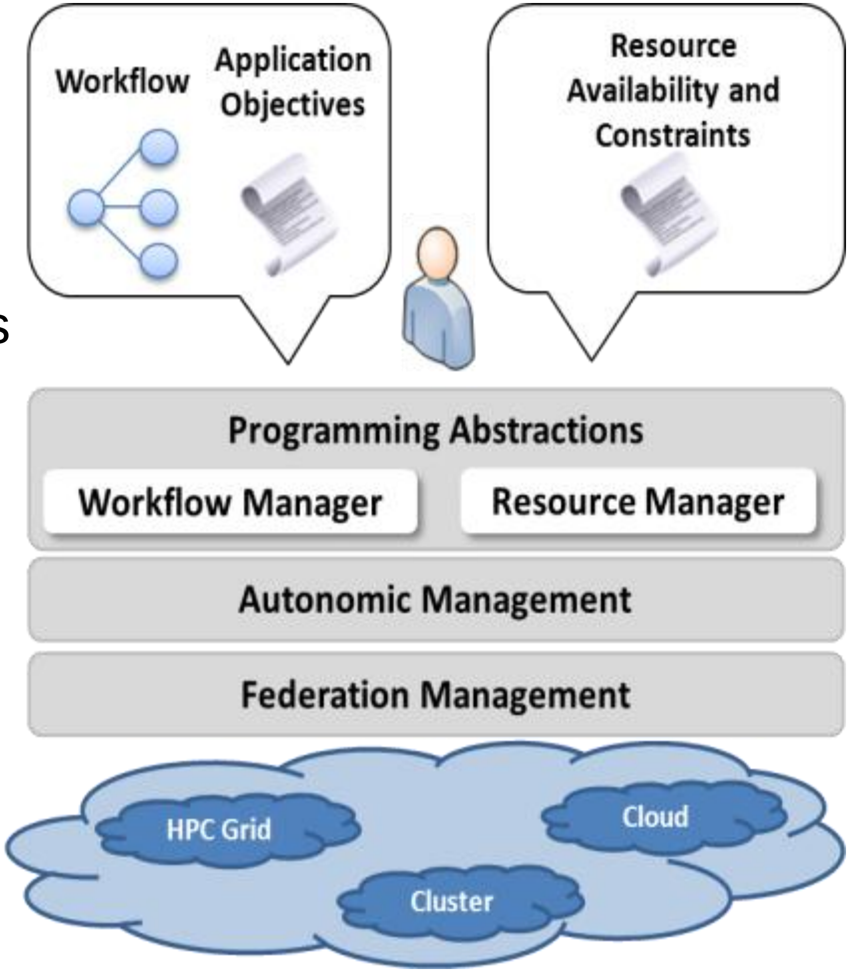
- Autonomic cross-layer federation management
  - Resources specified based on availability, capabilities, cost/performance constraints, etc.
  - Dynamically assimilated (or removed)
  - Resources coordinate to:
    - Identify themselves / verify identity
    - Advertise their resources capabilities, availabilities, constraints
    - Discover available resources
- Federation coordinated using Comet Spaces
- Autonomic resource provisioning, scheduling and runtime adaptations
- Business/social models for resource sharing





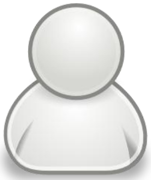
Software Defined Cyberinfrastructure Federations for Business and Science?

- Combine cloud abstractions with ideas from software-defined environments
- Create a nimble and programmable environment that autonomously evolves over time, adapting to:
  - Changes in the infrastructure
  - Application requirements
- Enable efficient data processing by
  - Allocating computing close to data sources
  - Process data in-situ and/or in-transit
- Independent control over application and resources



Software-defined Ecosystem

User/Provider



Define federation programmatically using rules and constraints

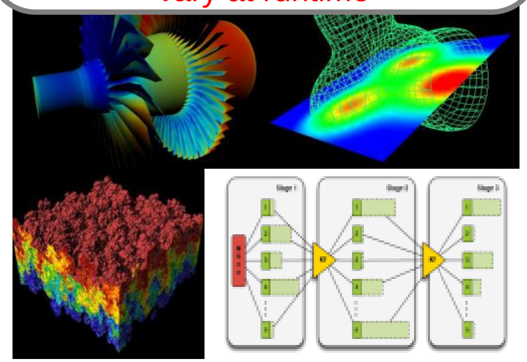
- Availability
- Capacity & Capability
- Cost
- Location
- Access policy

- vary at runtime -

Scientific Applications & Workflows

- Workflow definition
- Objectives (deadline, budget)
- Requirements (throughput, memory, I/O rate)
- Defined in terms of science (e.g., precision, resolution)

- vary at runtime -



Autonomic Manager

- Identify utility of federation
- Negotiate with application
- Ensure applications' objectives and constraints
- Adapt and reconfigure resources and network on the fly

Synthesize a space-time federated ACI

Exposed as a cloud to the application/workflow

Elastic Cyber-infrastructure

Software-defined ACI: ACI-as-a-Cloud

- Software defined ACI federations exposed using elastic on-demand Cloud abstractions
- Declaratively specified to define availability as well as policies and constraints to regulate their use
  - Use of a resources may only be allowed at certain times of the day, or when they are lightly loaded, or when they have sufficient connectivity, etc.
  - Prefer certain type of resources over others (e.g., HPC versus clouds or “free" HPC systems versus the allocation-based ones)
  - Specify how to react to unexpected changes in the resource availability or performance
  - Use resources only within the US or Europe due to the laws regulating data movement across borders
- Evolve in time and space -- the evaluation of these constraints provides a set of available resources at evaluation time
- Leverage software-defined networks to customize and optimize the communication channels or software-defined storage to improve data access

Software-defined ACI: Platform as a Service

- Platform as a Service to decouple applications from the underlying ACI Cloud

- Key components
  1. An API for building new applications or application workflows
  2. Mechanisms for specifying and synthesizing a customized views of the ACI federation that satisfies users' preferences and resource constraints
  3. Scalable middleware services that expose resources using Cloud abstractions
  4. Elasticity exposed in a semantically meaningful way
  5. Autonomics management is critical

- CometCloud provides some of these - currently focusing on 2

Many technical issues

- **Deployability**: Must be easy to deploy by a regular user without special privileges
- **Standardization/Interoperability**: Interact with heterogeneous resources
- **Self-discovery**: Discovery mechanisms to provide a realistic view of the federation
- **Scalability and extended capacity**: Scale across geographically distributed resources
- **Elasticity**: Ability to scale up, down or out on-demand
- **Security, Authentication, Authorization, Accounting**……
- ….

Related Work - Cloud Federation

- Cloud Bursting (scaling out to a cloud when needed)
  - Extending local cluster to a cloud with different scheduling policies (M. D. de Assuncao et. al)
  - Extending Austrian Grid with a private cloud (S. Ostermann et. al)
  - Extending grid resources to a Nimbus cloud (C. Vazquez et. Al)
- Hybrid Grid and Cloud
  - Creating a large-scale distributed virtual clusters using federated resources from FutureGrid and Grid'5000 (P. Riteau et. al)
  - Infrastructure to manage the execution of service workflows in a union of a grid and a cloud (L. F. Bittencourt et. al)
- Cloud of Clouds
  - Federation of Amazon EC2 and NERSC's Magellan cloud (I. Gorton et. al)
  - Using Pegasus and Condor to federate FutureGrid, NERSC's Magellan cloud and Amazon EC2 (J.-S. Vockler et. al)
- Federation Models
  - Composing cloud federation using a layered service model (D. Villegas et. al)
  - Cross-federation model using customized cloud managers (A. Celesti et. al)
  - A reservoir model that aims at contributing to best practices (B. Rochwerger et. al)

Relevant Related Projects

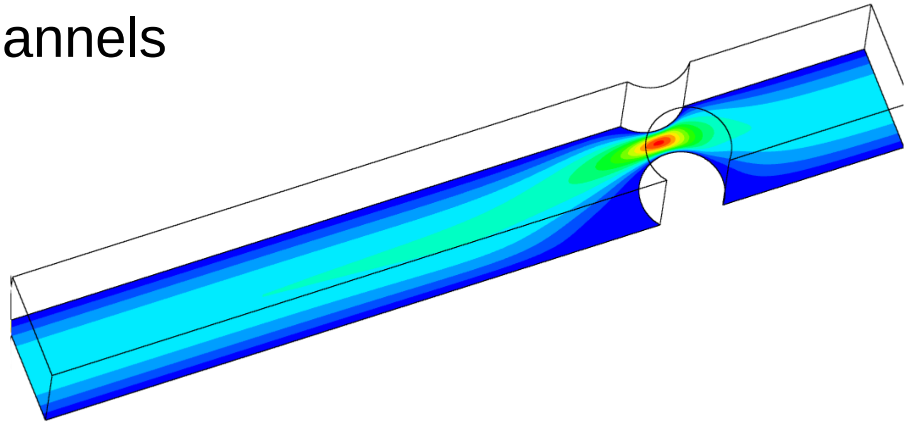
- FED4FIRE (European Union FP7)
  - A common federation framework for developing, adapting or adopting tools that support experiment lifecycle management, monitoring and trustworthiness
- InterCloud (Univ. of Melbourne, Australia)
  - Utility-oriented federation of cloud computing environments for scaling of application services
- Business Oriented Cloud Federation (Univ. of South Hampton, UK)
  - Cloud federation model via computation migration for real time applications; targets real-time online interactive applications, online games
- ….

Autonomics in Multi-Cloud Environments

- Links with Control theory From Chenyang Lu (Washington Univ. in St Louis)
  - Provide QoS and related guarantees in open, unpredictable environments
- PACMan Alan Roytman, Aman Kansal, Jie Liu and Suman Nath, “PACMan: Performance Aware Virtual Machine Consolidation”, Proceedings of ICAC 2013, San Jose, USA (USENIX/ACM)
  - VM Consolidation and dynamic VM allocation
- AGILE H. Nguyen, Z. Shen, X. Gu, S. Subbiah, J. Wilkes, “AGILE: Elastic distributed resource scaling for Infrastructure-as-a-Service”, Proceedings of ICAC 2013, San Jose, USA (USENIX/ACM)
  - Medium term predictions using Wavelets
  - Use of an “adaptive” copy rate
- TIRAMOLA “Automated, Elastic Resource Provisioning for NoSQL Clusters Using TIRAMOLA” Dimitrios Tsoumakos, Ioannis Konstantinou, Christina Boumpouka, Spyros Sioutas, Nectarios Koziris, CCGrid 2013, Delft, The Netherlands
  - Modelling decisions as a Markov Decision Process to support elastic behaviour
- Autoflex: Service Agnostic Auto-scaling Framework for IaaS Deployment Models” Fabio Morais, Francisco Brasileiro, Raquel Lopes, Ricardo Araujo, Wade Satterfield, Leandro RosaIEEE/ACM CCGrid 2013, Delft, The Netherlands
  - Reactive and proactive auto scaling mechanisms based on monitoring

An Initial Experiment: Fluid Flow in Microchannel

- Controlling fluid streams at microscale is of great importance for biological processing, creating structured materials, etc.
- Placing pillars of different dimensions, and at different offsets, allows "sculpting" the fluid flow in microchannels
- Four parameters affect the flow:
  - Microchannel height
  - Pillar location
  - Pillar diameter
  - Reynolds number



- Each point in the parameter space represents simulation using the Navier-Stokes equation (MPI-based software)
- Highly heterogeneous and computational cost is hard to predict a priori
- Global view of the parameter space requires 12,400 simulations (three categories)

Fluid Flow in Microchannel Experiment Setup

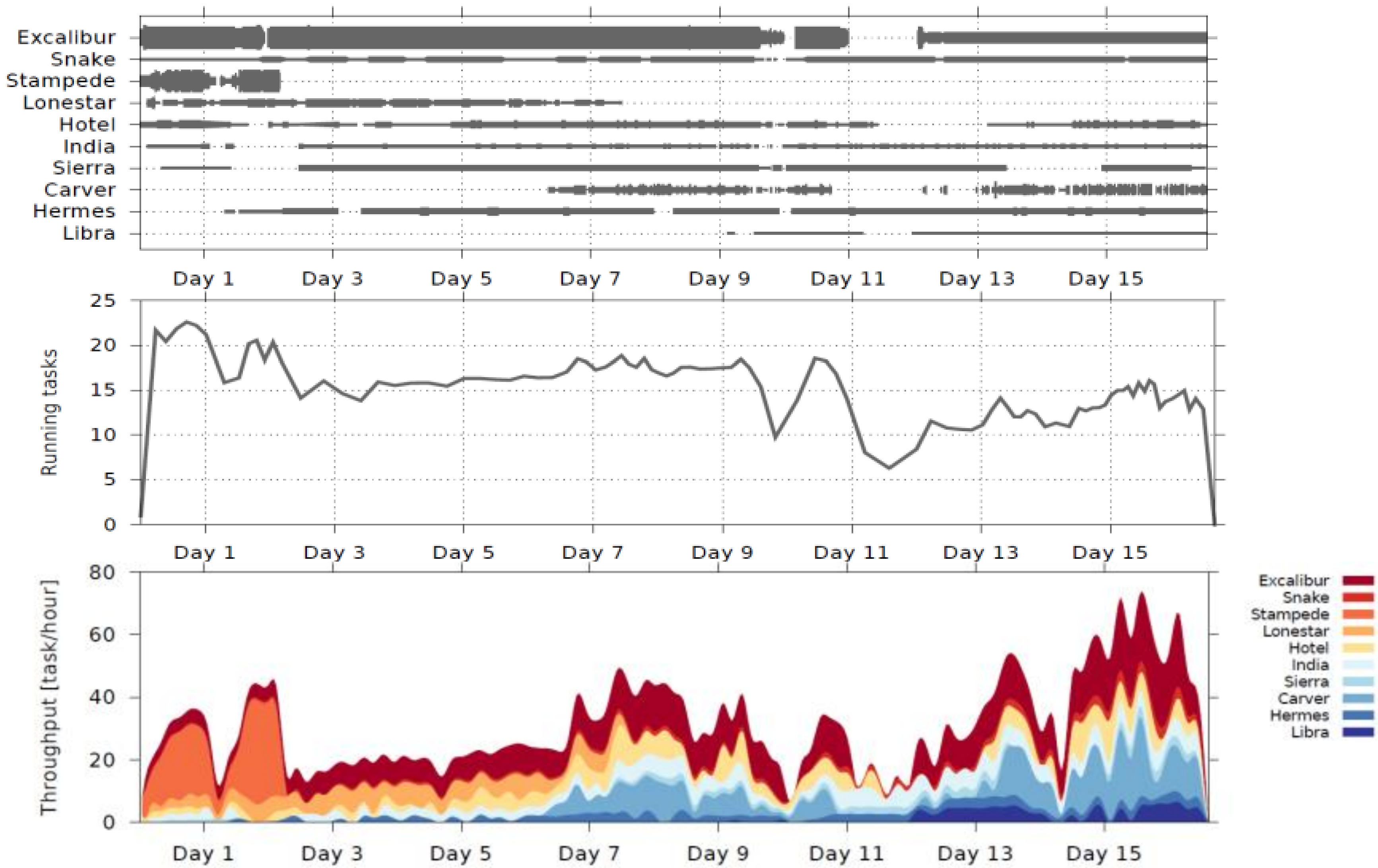
- Minimum Time of Completion - Elastically and opportunistically federate resources
- Global view of the parameter space requires 12,400 simulations (three categories)
- Experiment completely performed within user space (SSH)
- 10 different HPC resources from 3 countries

Name	Type	Cores†	Network	Scheduler
Excalibur	IBM BG/P	8,192	BG/P	LoadLeveler
Snake	Linux SMP	64	N/A	N/A
Stampede	iDataPlex	1,024	IB	SLURM
Lonestar	iDataPlex	480	IB	SGE
Hotel	iDataPlex	256	IB	Torque
India	iDataPlex	256	IB	Torque
Sierra	iDataPlex	256	IB	Torque
Carver	iDataPlex	512	IB	Torque
Hermes	Beowulf	256	10 GbE	SGE
Libra	Beowulf	128	1 GbE	N/A

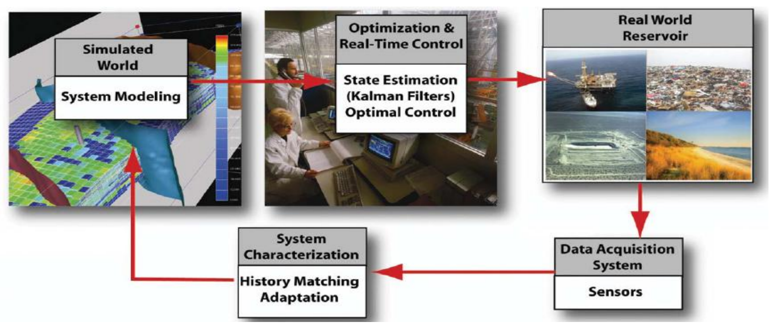
Note: † – peak number of cores available to the experiment.

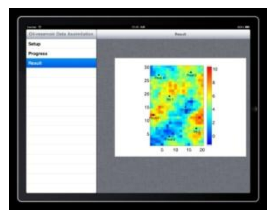
Summary of the Experiment

~16 days of continuous execution	12,845 tasks processed (445 extra)
2,897,390 CPU-hours consumed	400 GB of data generated



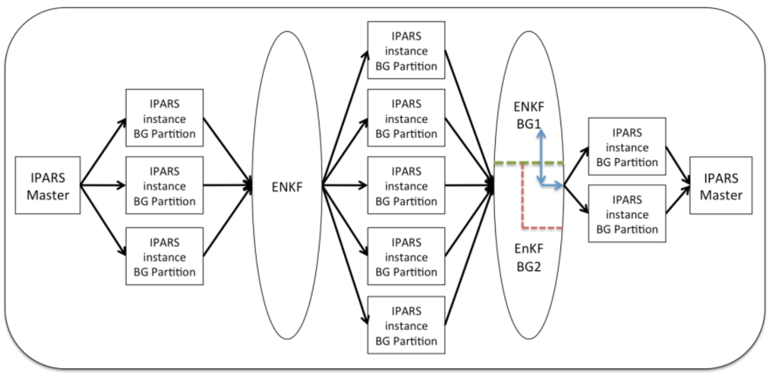
HPC as a Service (Winner SCALE'11)





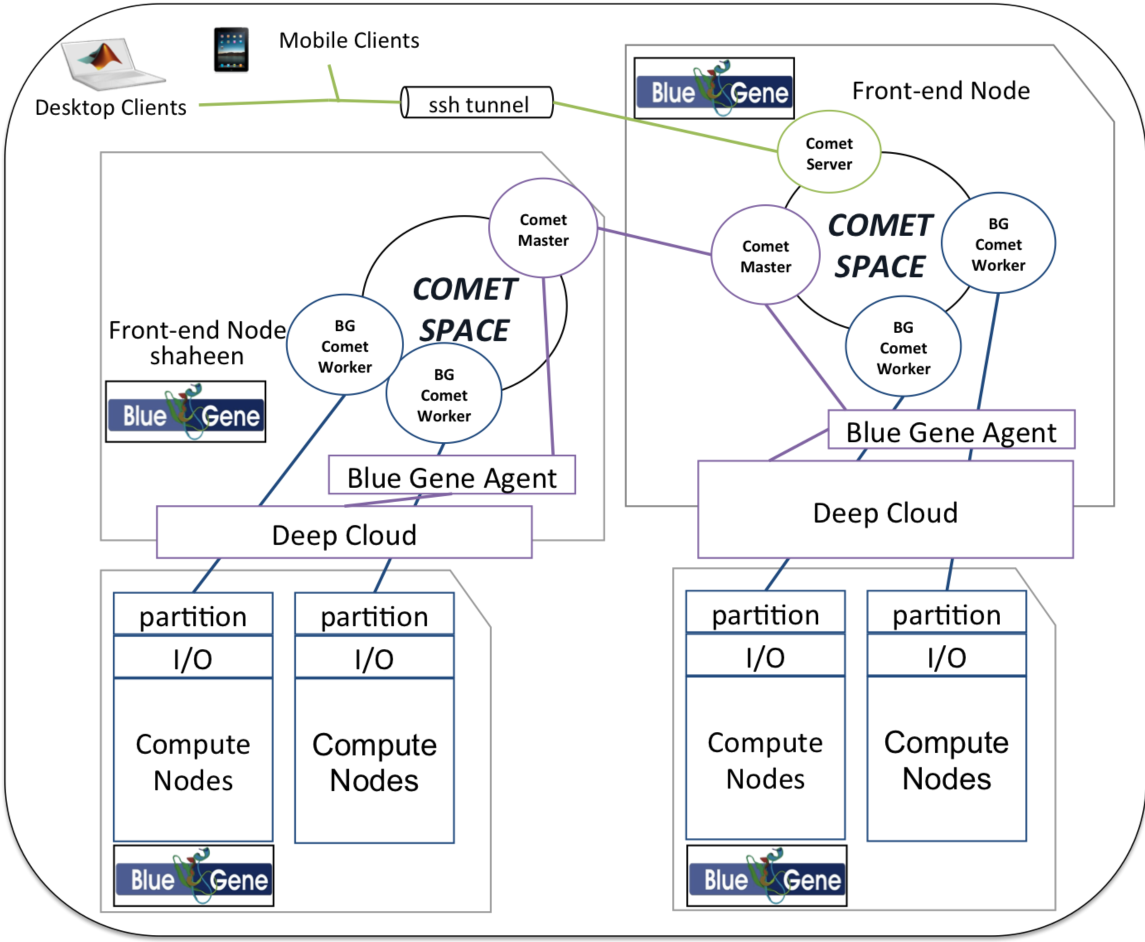
Demonstrated how the cloud abstraction can be effectively used to support ensemble geo-system management applications on a geographically distributed federation of supercomputing systems using a pervasive portal running on an iPad

http://nsfcac.rutgers.edu/icode/scale



HPC as a Service (IEEE Computer 10/12)

- HPC as a Service using federation of IBM Blue Gene/P systems
- Elastically scale up to 22K processors



Accelerating Protein Folding using Advanced Computational Infrastructure (Rutgers + BMS)

Individual trajectories

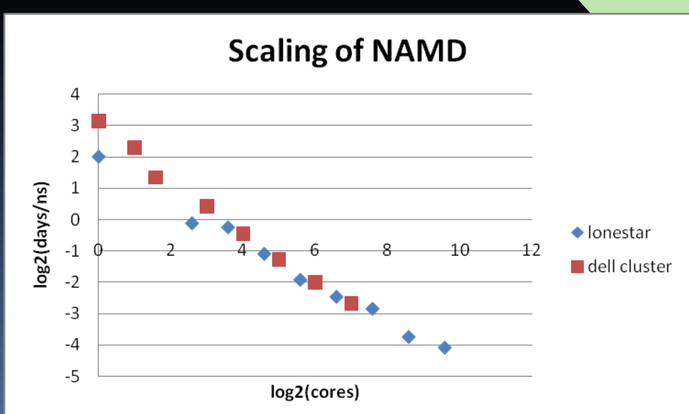
- Parallel NAMD trajectories
- Asynchronous communication in cometCloud

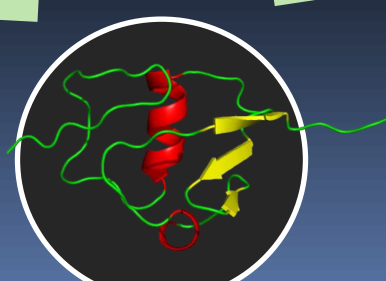
Science

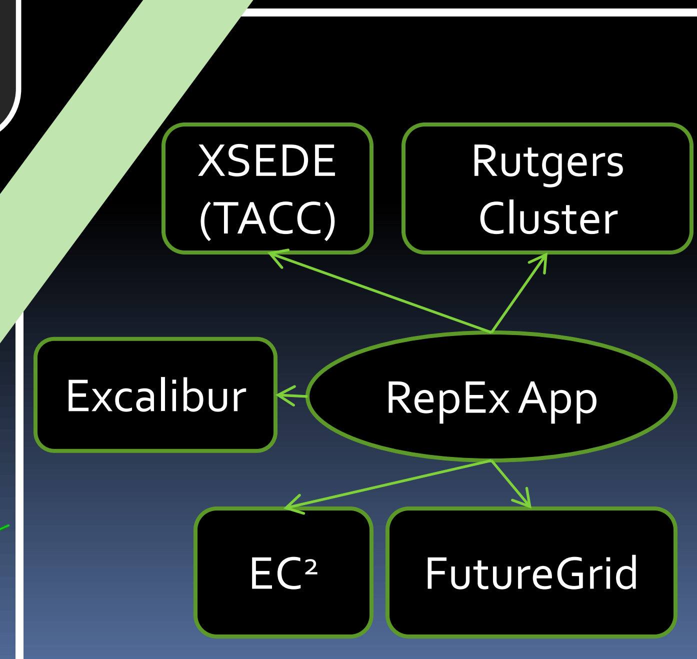
- Be smart about using resources
- Commodity hardware versus high end resources
- Terminate or restart resources

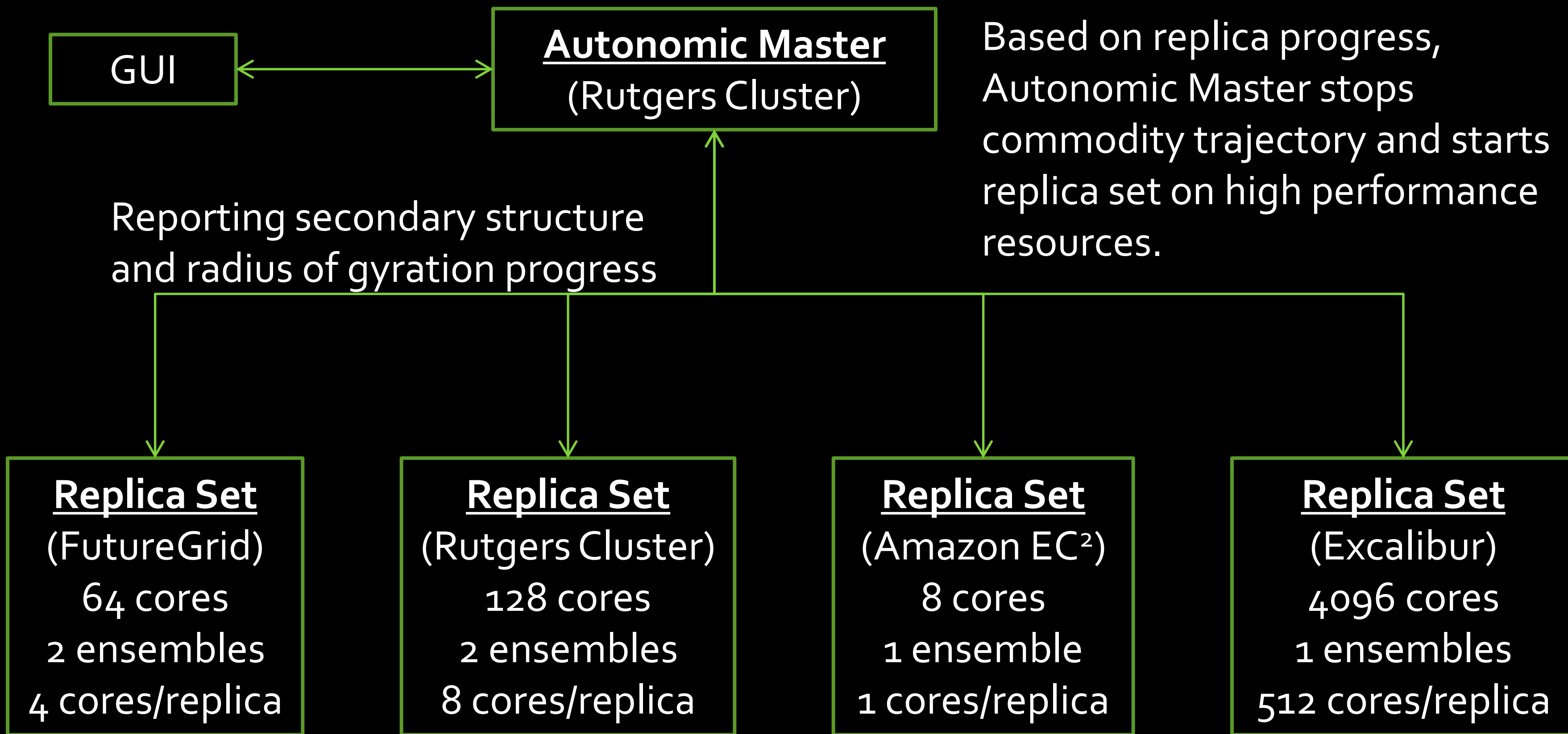
Infrastructure

- Federated clouds









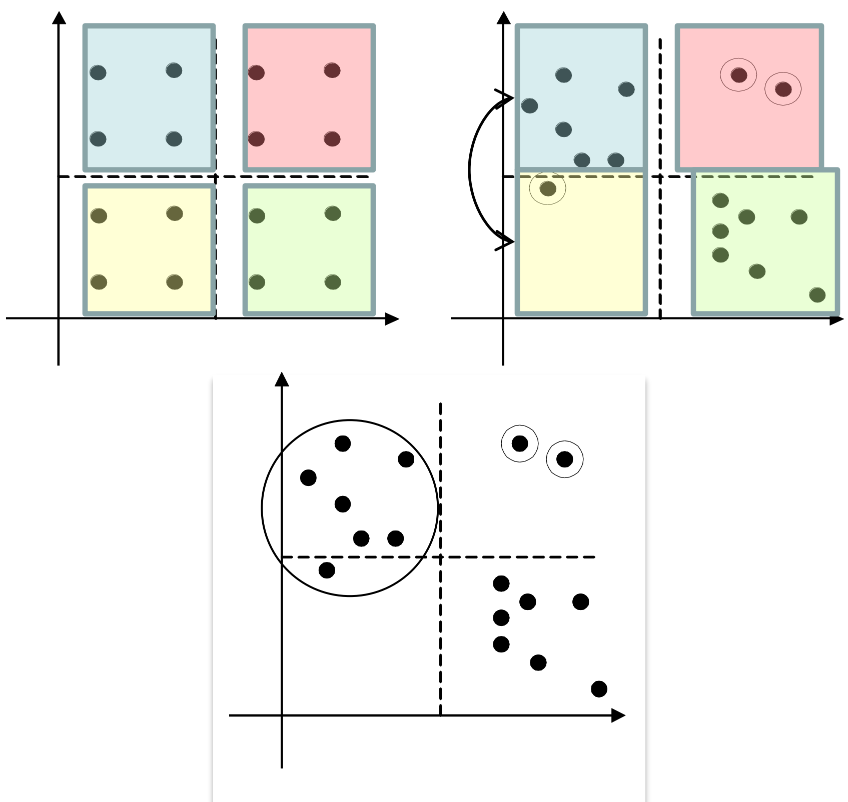
*Could run multiple replicas per temperature to improve likelihood of asynchronous exchange on heterogeneous hardware.

*8 temperatures = 1 ensemble

http://youtu.be/sg2C7N7g5CU

Enterprise Business Data Analytics

- Decentralized Clustering Analysis
- Algorithm to study large multi-dimensional information space
- Search and correlate different attributes with known data sources, and allow visualizing and interpreting the results interactively



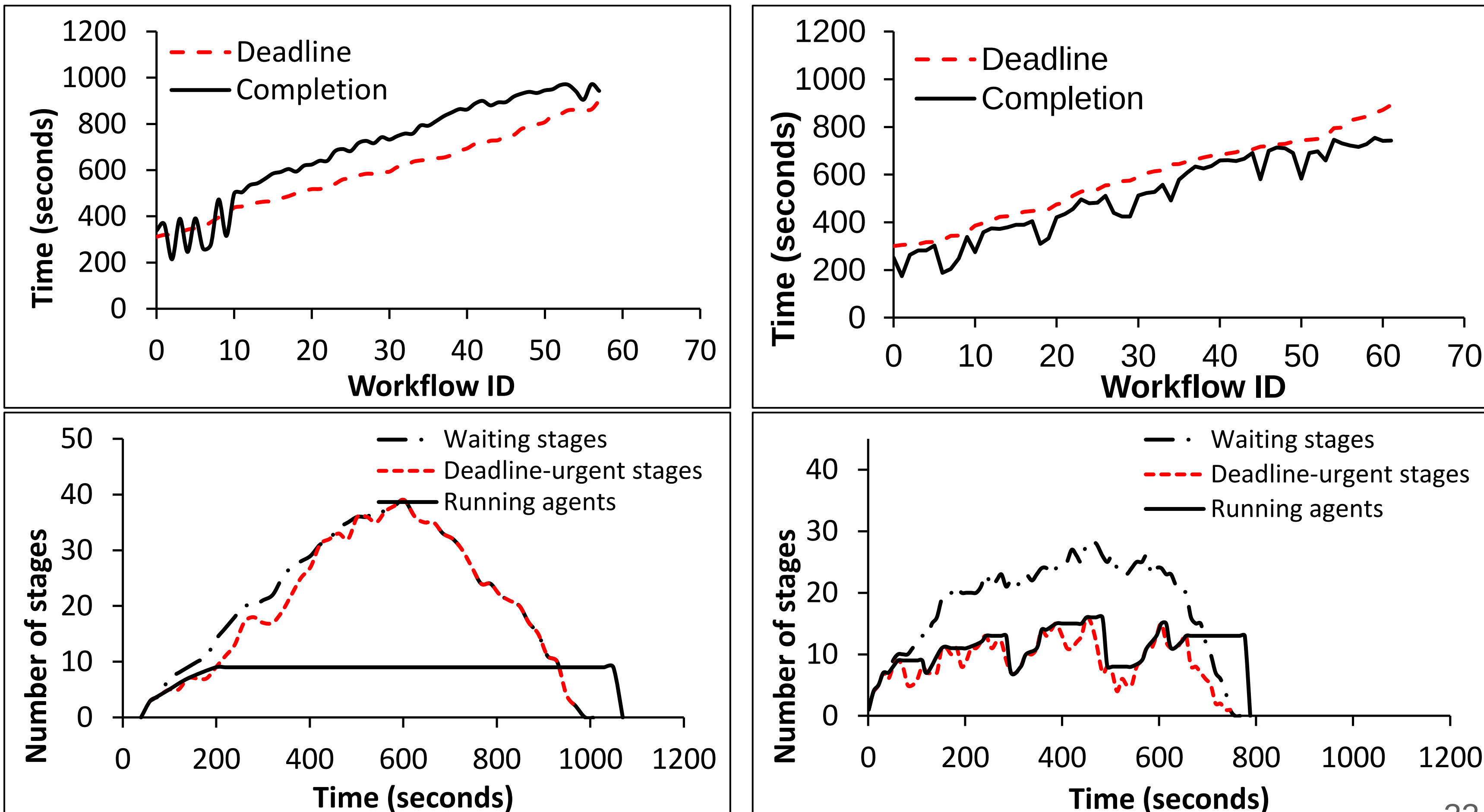
- The space is divided into regions and each region is assigned to a processing node
- Clusters are recognized by evaluating the relative density of points in a given region
- Nodes must communicate with neighbors to account for clusters that occur across region boundaries

Experiment

- Deadline-driven workflows
  - Each workflow has 3 different stages of the DOC application
    - Each stage of the workflow has a different execution time
    - Each stage is a task which is completed by 1 agent and 2 workers
  - Deadline for a workflow is set to average 300 seconds (100 seconds per stage)
  - Submitting workflows every 10 seconds during initial 600 seconds of experiment
  - CloudBurst – No CloudBurst
- Resources
  - Rutgers cluster has 27 machines
  - Amazon EC2 - c1.medium instance type

Deadline-Driven Results

No CloudBurst | **CloudBurst**



Other experiments

- Data-Driven Workflows on Federated Clouds [Cloud'14]
- Federating Resources using Social Models [IC2E'14]
- Elastic Federations for Large-scale Scientific Workflows [MTAGS'13]
- HPC plus Cloud Federations [e-Science'10]
- …. [See cometcloud.org]

- Testbed using resource in US (RU, FutureGrid, XSEDE, IBM), UK (Cardiff), Amazon EC2

- Experiments successful…. but can the model be generalized?

Summary

- Emerging CDS&E workflows have dynamic and non-trivial computational/data requirements
  - Necessitate dynamically federated platforms that integrate heterogeneous resources / services
  - Provisioning and federating an appropriate mix of resources on-the-fly is essential and non-trivial
- Software-defined Advanced Cyber-Infrastructure for Science
  - Software defined ACI federations exposed using elastic on-demand Cloud abstractions
  - Application access using established programming abstraction/platforms for science
  - Autonomic management is critical
- Many challenges at multiple layers
  - Application formulation, programming systems, middleware services, standardization & interoperability, autonomic engines, etc.

The CometCloud Team

- Ph.D. Students
  - Moustafa AbdelBaky, Dept. of Electrical & Computer Engr.
  - Mengsong Zou, Dept. of Computer Science
  - Ali Reza Zamani, Dept. of Computer Science
  - Shivaramakrishnan Vaidyanathan, Dept. of Computer Science
- Faculty
  - Manish Parashar, Ph.D. - Dept. of Computer Science, Rutgers Discovery Informatics Institute (RDI2)
  - Javier Diaz-Montes, Ph.D. - Rutgers Discovery Informatics Institute
  - Esma Yildirim, Ph.D. - Rutgers Discovery Informatics Institute

And many collaborators….

CometCloud: http://cometcloud.org

Thank You!

Manish Parashar, Ph.D.
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