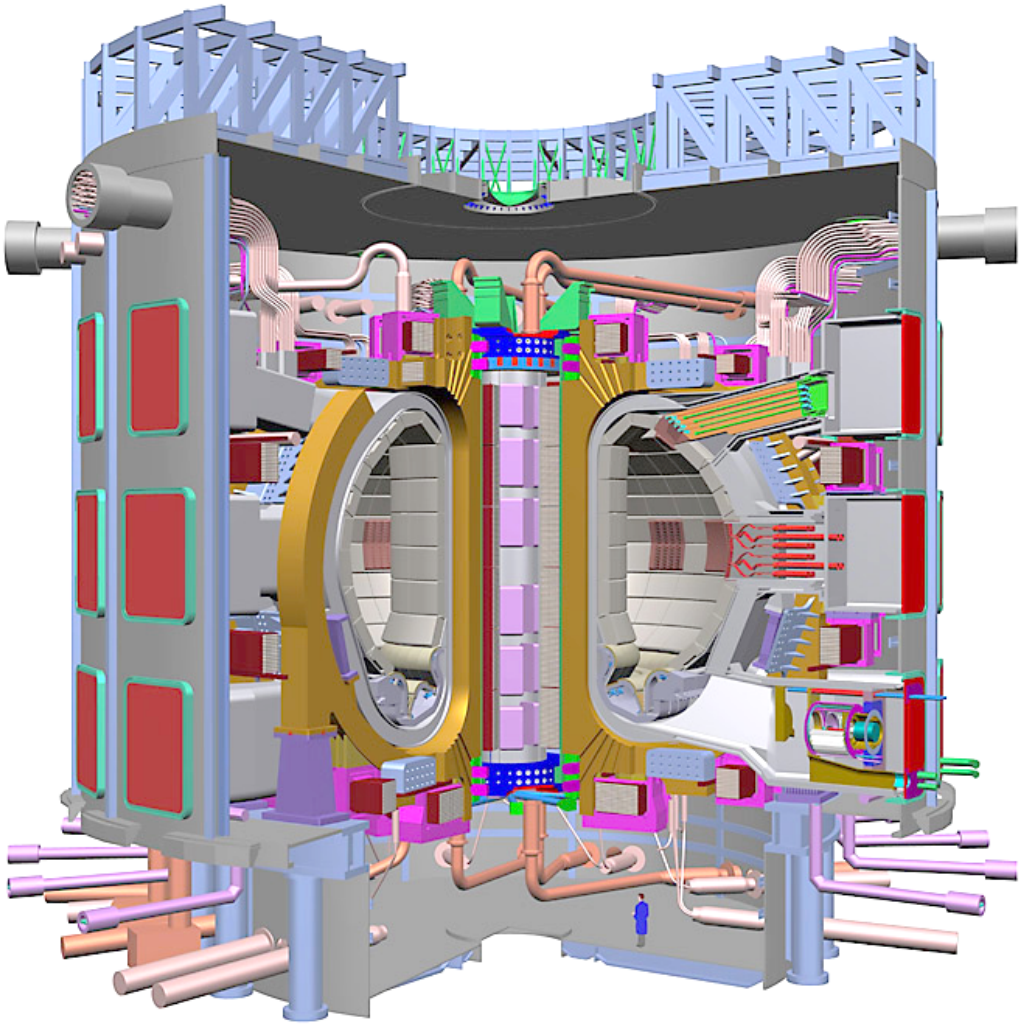
Cloud Computing for Science

Kate Keahey
keahey@mcs.anl.gov
Argonne National Laboratory
Computation Institute, University of Chicago

Cloud Computing for Science

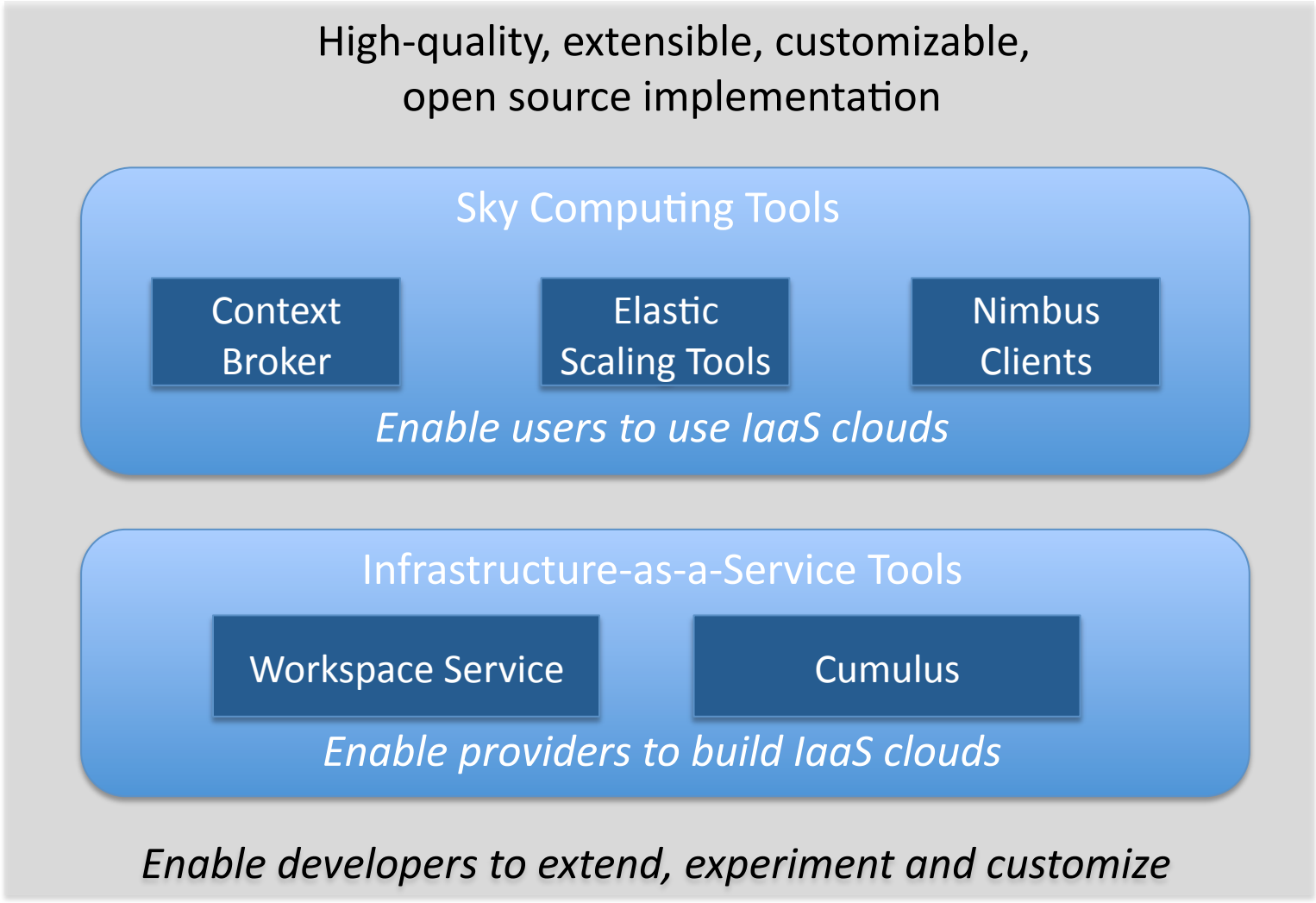


- On-demand computing
- Control over environment

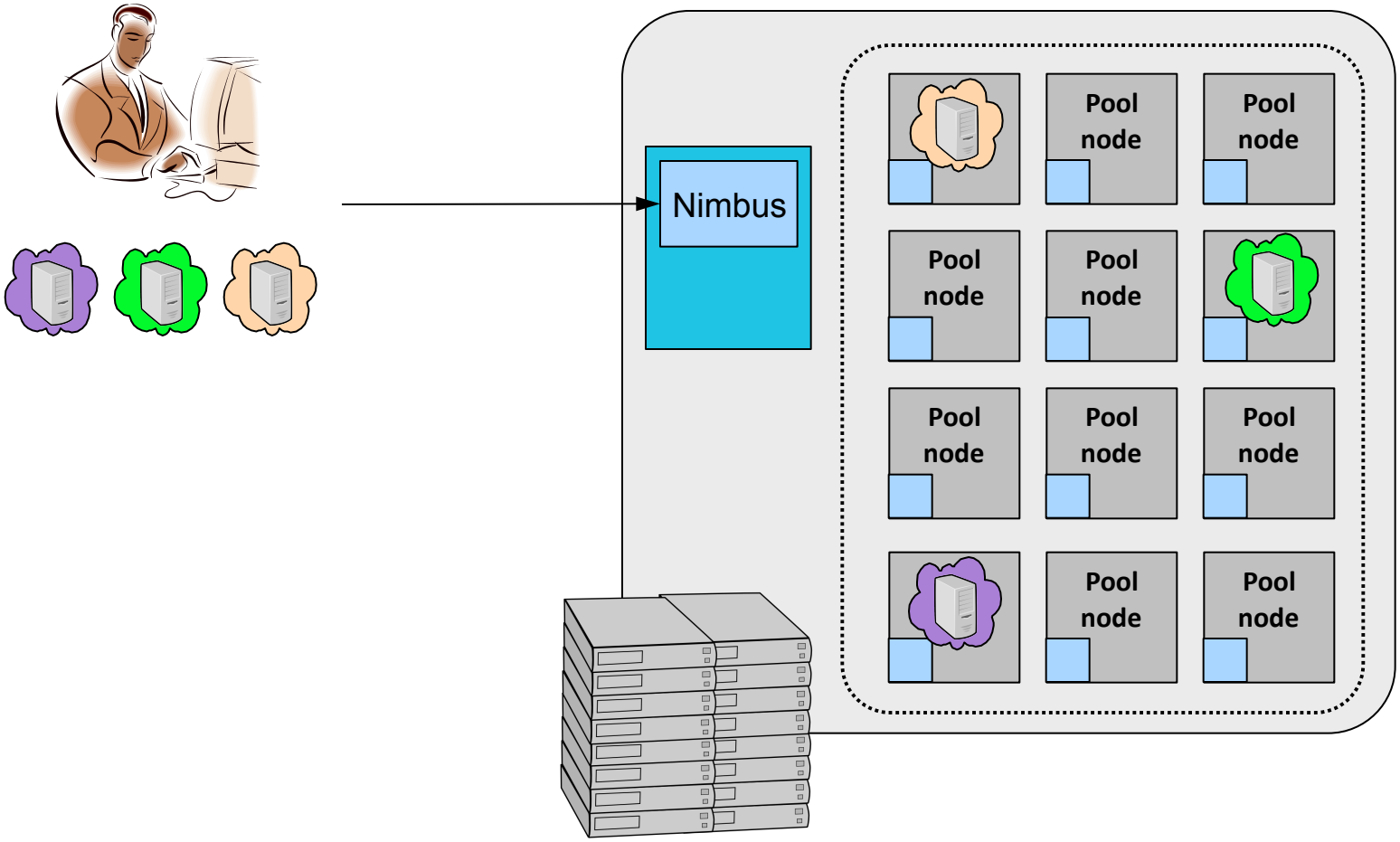
NIMBUS www.nimbusproject.org

Infrastructure-as-a-Service Cloud Computing: the Nimbus Toolkit

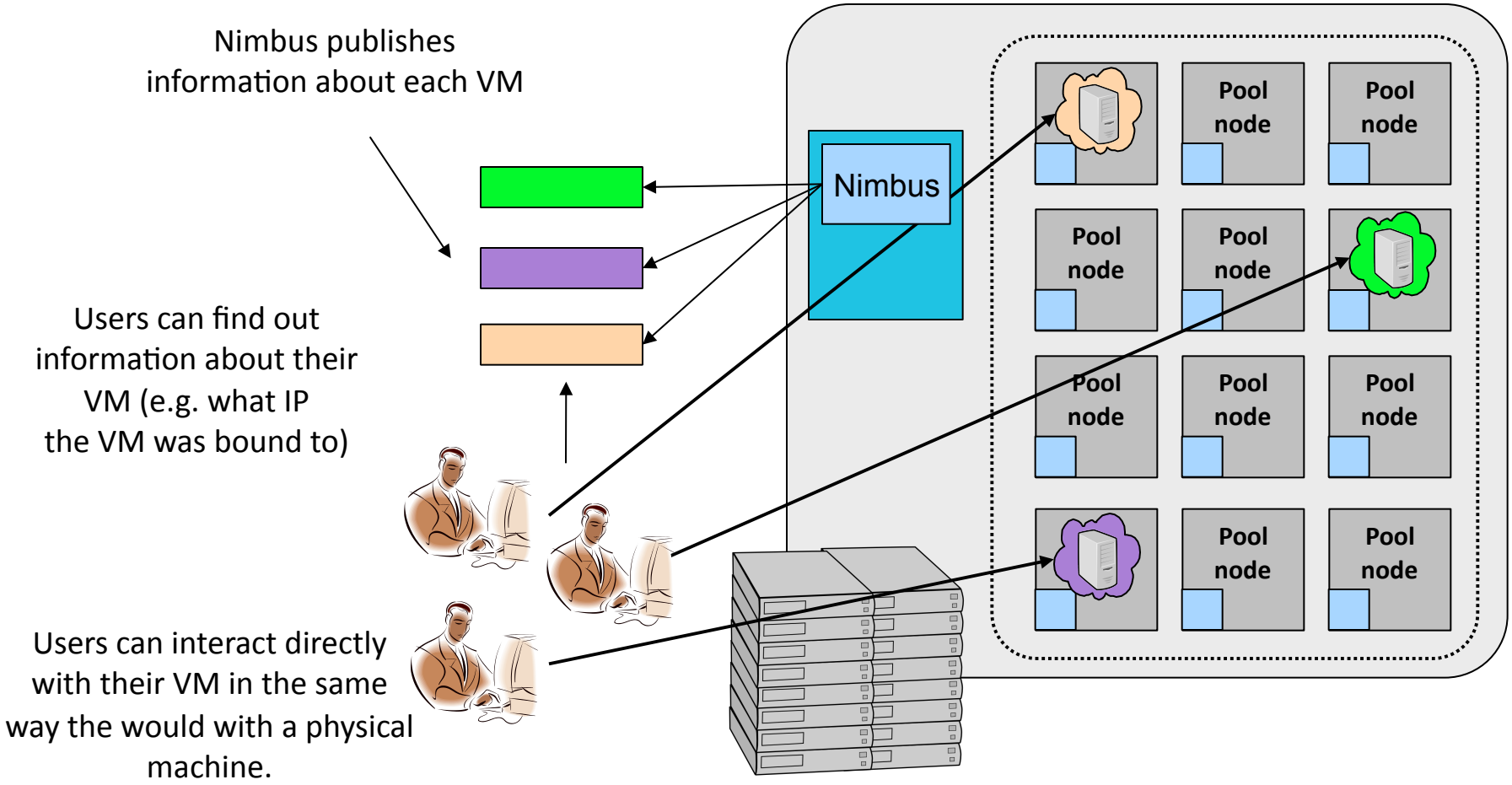
Nimbus Goals



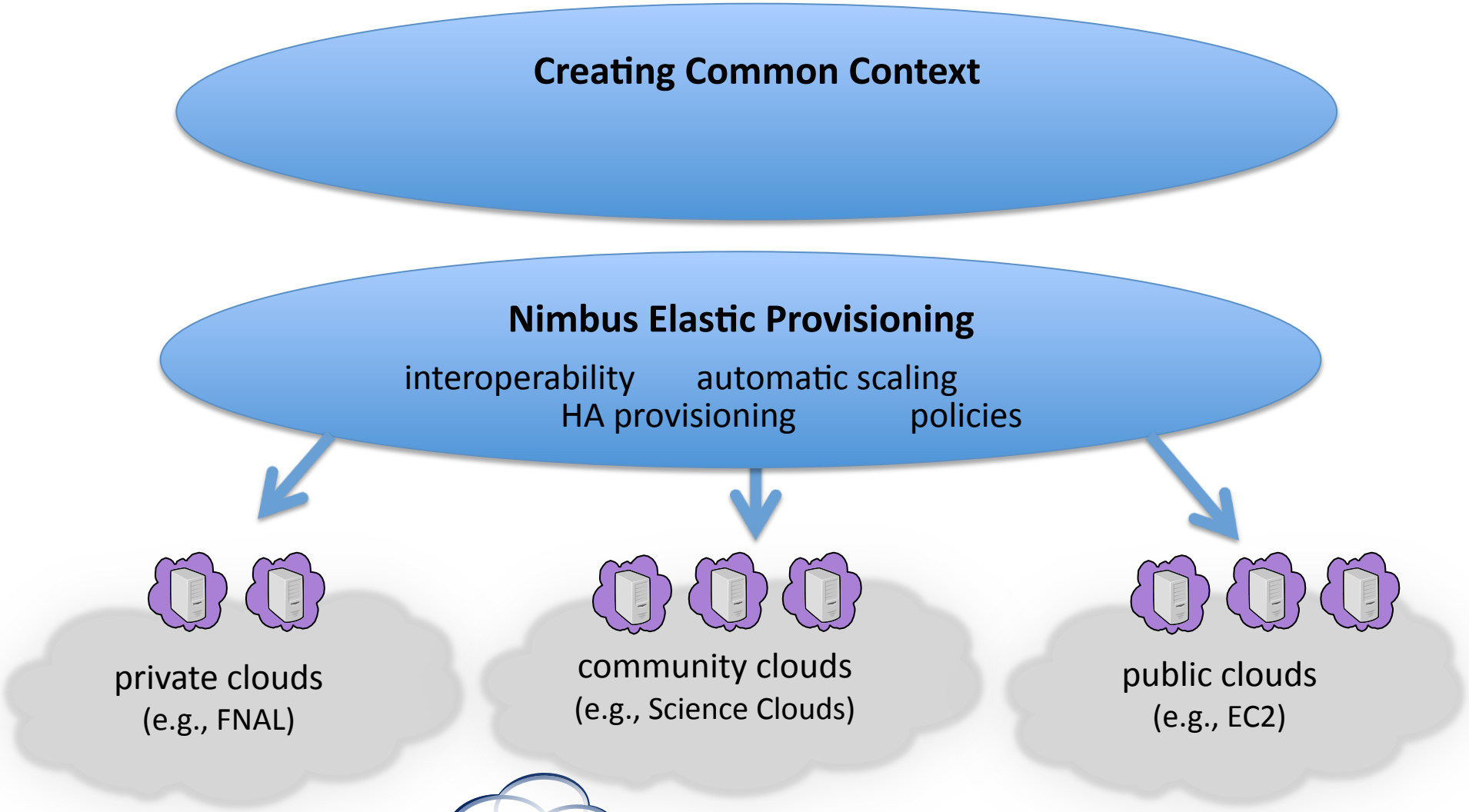
IaaS: How it Works



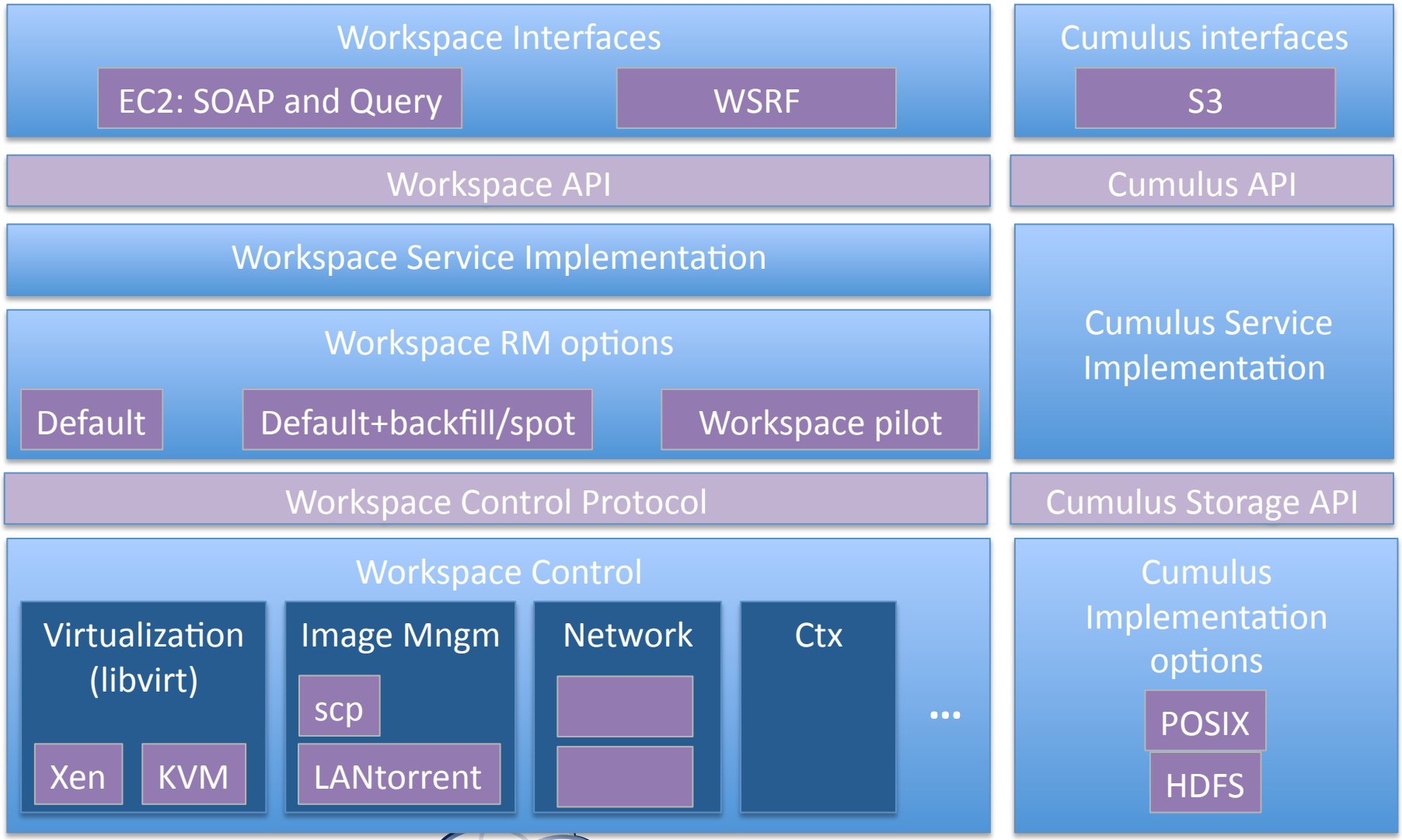
IaaS: How it Works



Sky Computing Tools: Working with Hybrid Clouds



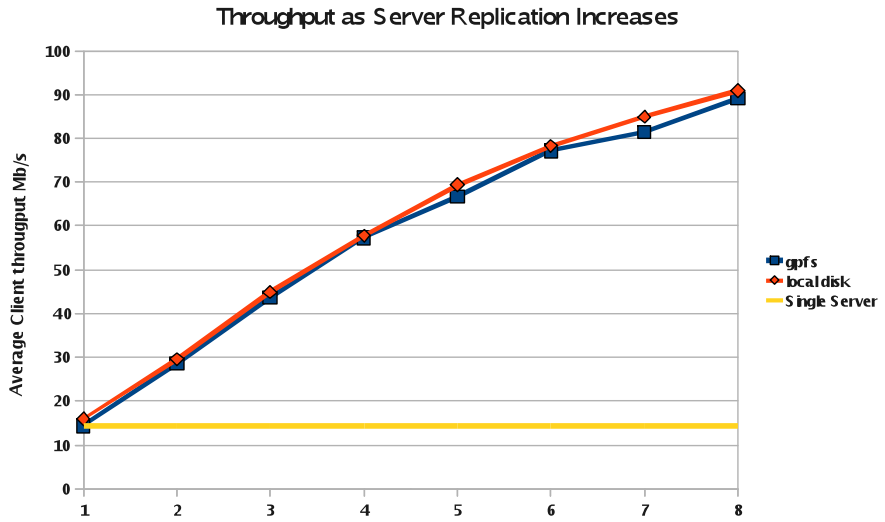
Nimbus: A Highly-Configurable IaaS Architecture

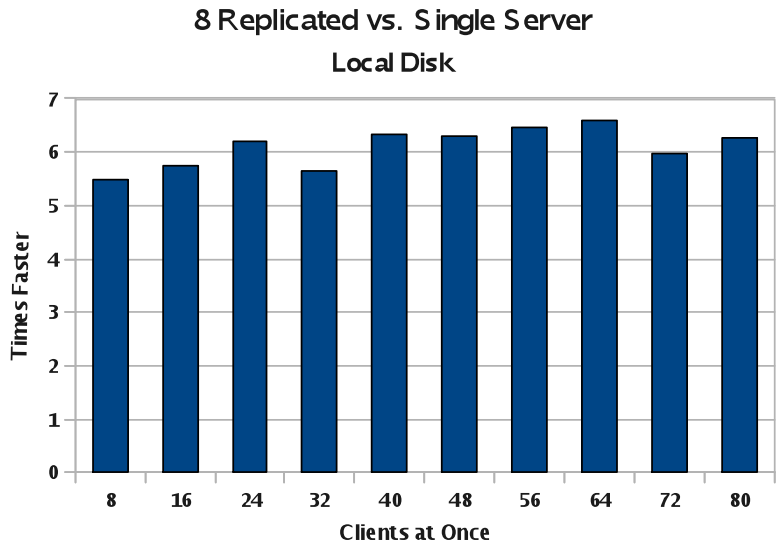


Recent Highlights

Cumulus: a Scalable Storage Cloud

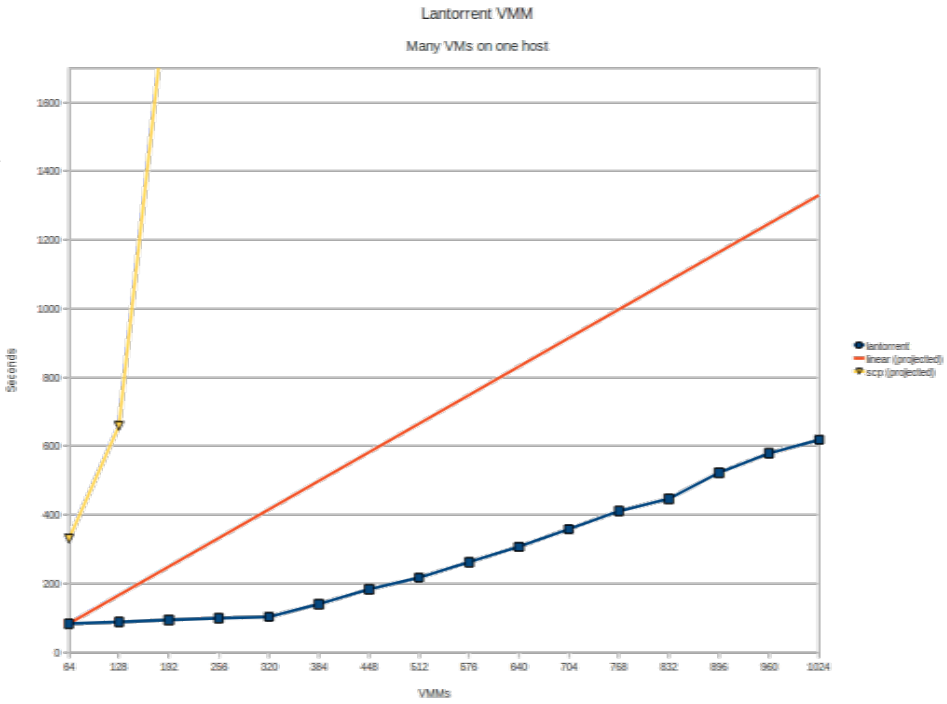
- Challenge: a scalable storage cloud
- S3-compatible open source storage cloud
- Quota support for scientific users
- Pluggable back-end to popular technologies such as POSIX, HDFS, potentially also Sector and BlobSeer
- Configurable to take advantage of multiple servers
- SC10 poster





LANTorrent: Fast Image Deployment

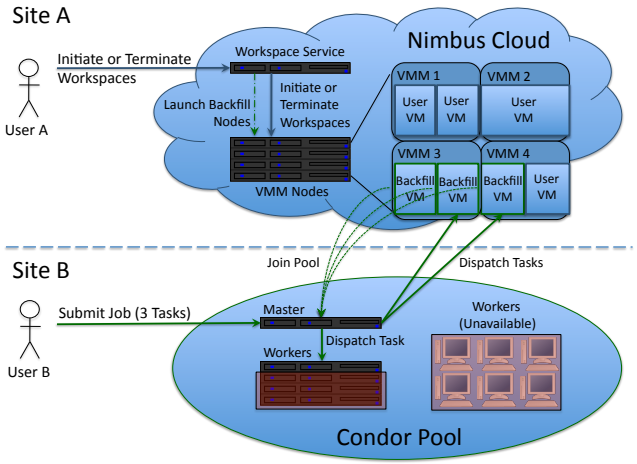
- Challenge: image deployment
- Moving images is the main component of VM deployment
- LANTorrent: the BitTorrent principle on a LAN
- Streaming
- Minimizes congestion at the switch
- Detecting and eliminating duplicate transfers
- Benefit: a thousand VMs in 10 minutes
- Nimbus release 2.6

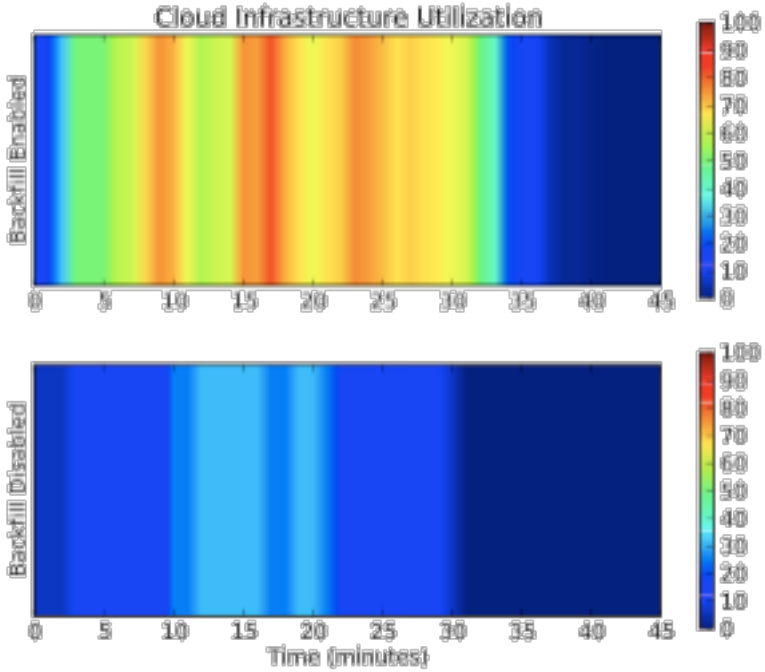


Preliminary data using the Magellan resource
At Argonne National Laboratory

Backfill: Lower the Cost of Your Cloud

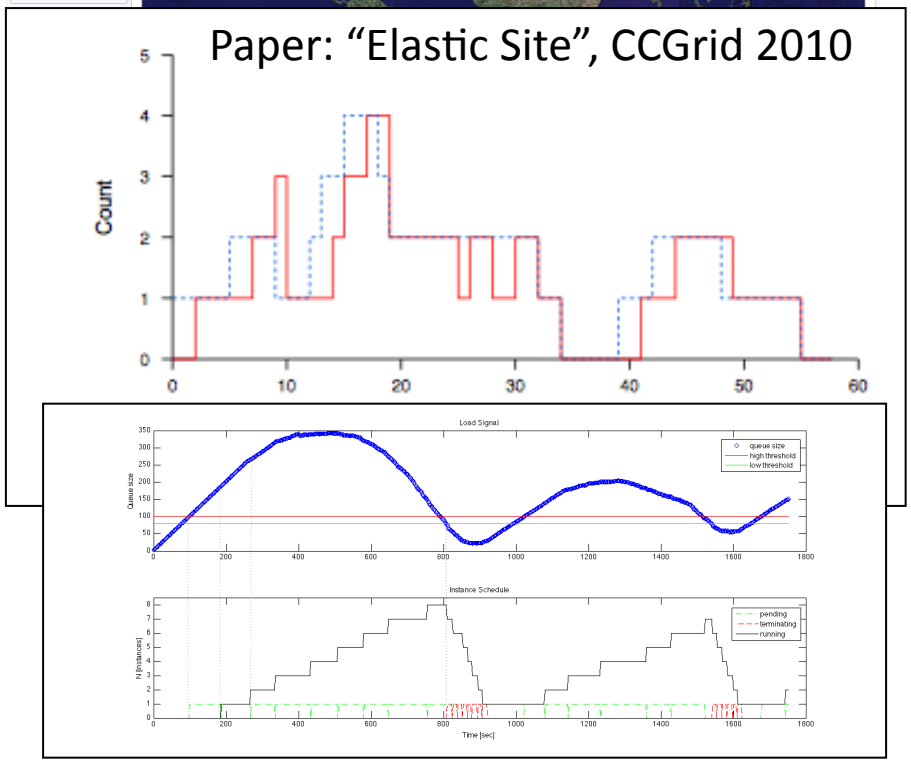
- Challenge: utilization, catch-22 of on-demand computing
- Solutions:
  - Backfill with volunteer VMs: up to 100% utilization
  - Spot pricing
- Open Source community contribution
- Preparing for running of production workloads on FG @ U Chicago
- Extension to Nimbus Workspace Service RM to be available in Nimbus release 2.7





Elastic Scaling Tools: Towards Bottomless Resources

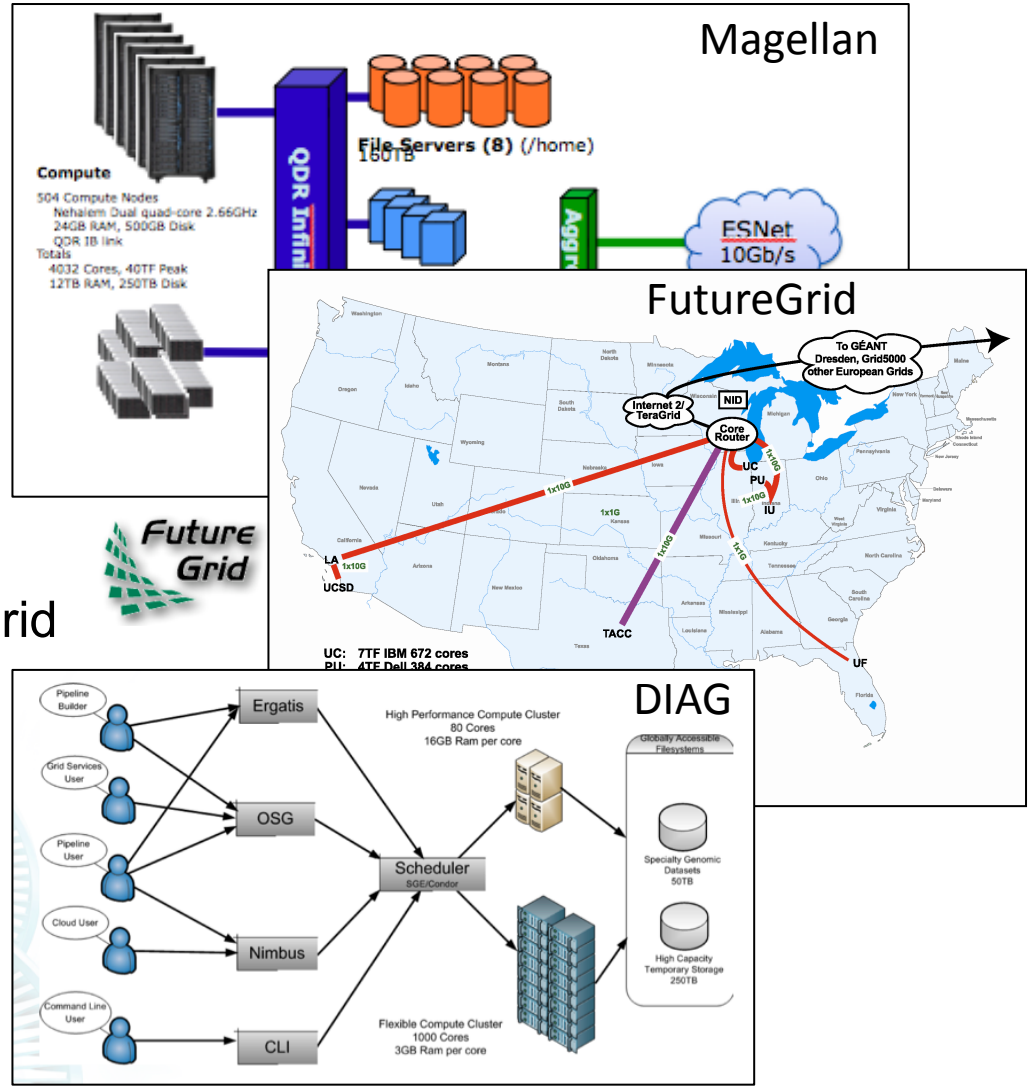
- Early efforts:
  - 2008: The ALICE proof-of-concept
  - 2009: ElasticSite prototype
  - 2009: OOI pilot
- Towards a generic HA Service Model
  - React to sensor information
  - Queue: the workload sensor
  - Scale to demand
  - Across different cloud providers
  - Use contextualization to integrate machines into the network
  - Customizable
  - Latest tests scale to 100s of nodes on EC2
- Release in 2011



Resources, Applications and Ecosystem

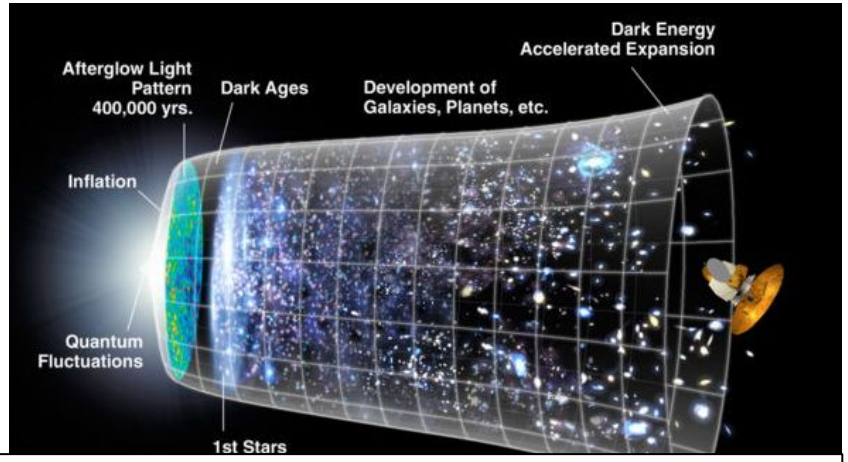
Scientific Cloud Resources

- Science Clouds
  - UC, UFL, Wispy@Purdue
  - ~300 cores
- Magellan
  - DOE cloud @ ANL&LBNL
  - ~4000 cores@ANL
- FutureGrid
  - ~6000 cores
- DIAG =
  - Data Intensive Academic Grid
  - U of Maryland School of Medicine in Baltimore
  - ~1200-1500 cores
- Outside of US:
  - WestGrid, Grid5000



Work by Jerome Lauret (BNL) et al.

- STAR: a nuclear physics experiment at Brookhaven National Laboratory
- Strategy:
  - Nimbus Science Clouds -> EC2 runs
  - Virtual OSG clusters with Nimbus Context Broker
- Impact
  - Production runs on EC2 since 2007
  - The Quark Matter 2009 deadline: producing just-in time results



Priceless?

- Compute costs: $ 5,630.30
  - Fdsf 300+ nodes over ~10 days,
  - Instances, 32-bit, 1.7 GB memory:
    - EC2 default: 1 EC2 CPU unit
    - High-CPU Medium Instances: 5 EC2 CPU units (2 cores)
  - ~36,000 compute hours total
- Data transfer costs: $ 136.38
  - Small I/O needs : moved <1TB of data over duration
- Storage costs: $ 4.69
  - Images only, all data transferred at run-time
- Producing the result before the deadline...

...$ 5,771.37

OCEAN OBSERVATORIES INITIATIVE

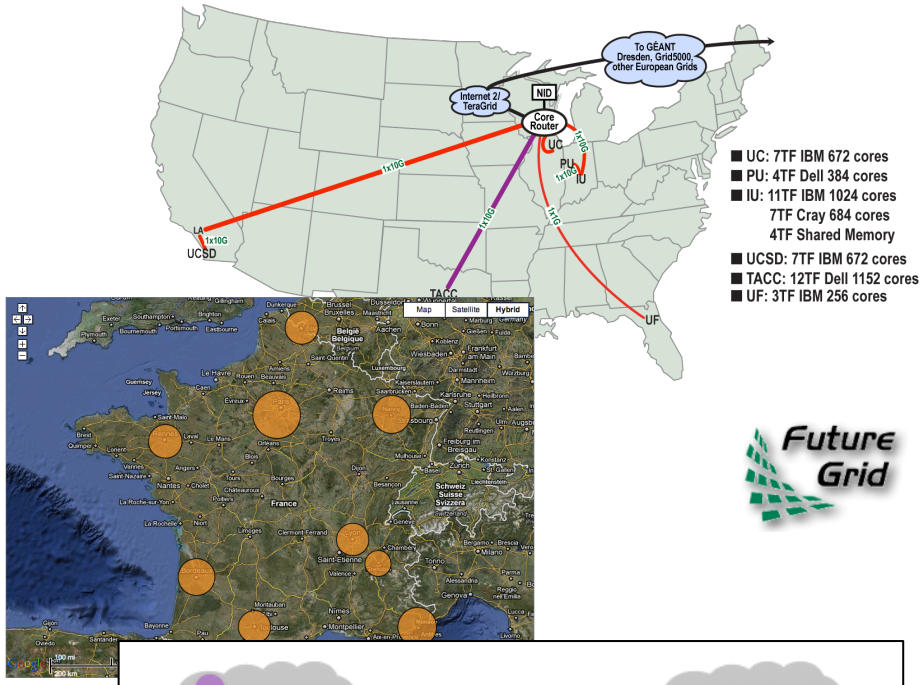
- Large NSF-funded observatory with requirements for adaptive, reliable, elastic computing
- Approach:
  - Private Nimbus regional clouds -> commer… …uds
  - Highly Availab… …ices that provisi… …on many cl… …leed
  - Signif… infra… …d and se… …ent based on th…
- Statu…
  - Scalability and reliability tests on 100s of EC2, FutureGrid and Magellan resources
  - HA elastic services release in Spring 2011

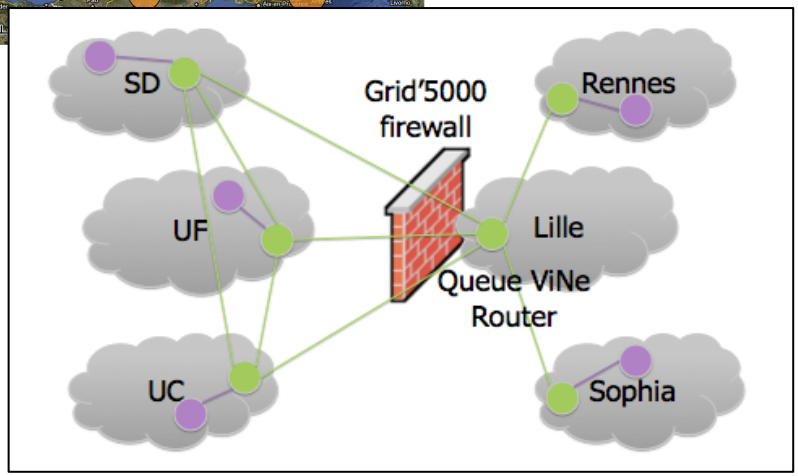
Trail-blazing project

Sky Computing @ Scale

Work by Pierre Riteau et al, University of Rennes 1

- Approach:
  - Combine resources obtained in multiple Nimbus clouds in FutureGrid and Grid' 5000
  - Deployed a virtual cluster of over 1000 cores on Grid5000 and FutureGrid – largest ever of this type
  - Combine Context Broker, ViNe, fast image deployment
- Grid'5000 Large Scale Deployment Challenge award
- Demonstrated at OGF 29 06/10
- TeraGrid '10 poster



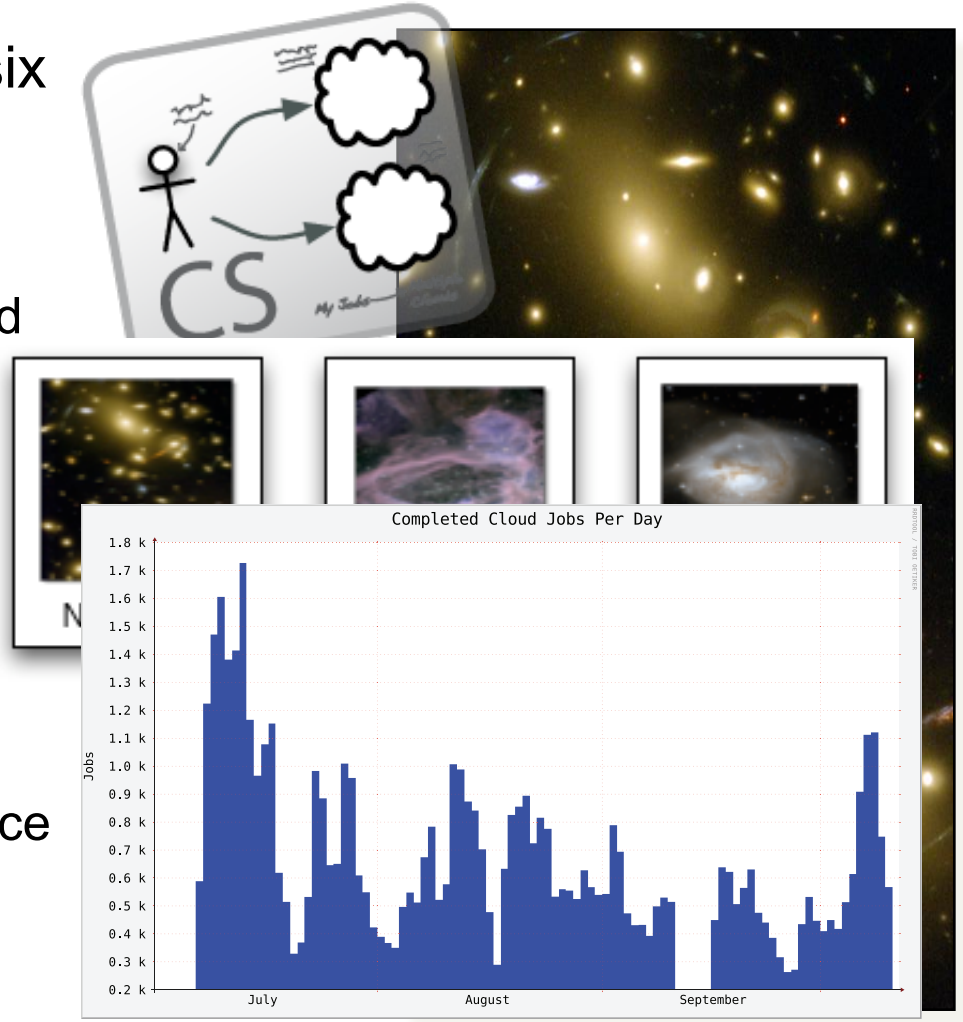


CANFAR
Canadian Advanced Network for Astronomical Research

University of Victoria NRC·CNRC

Work by the UVIC team

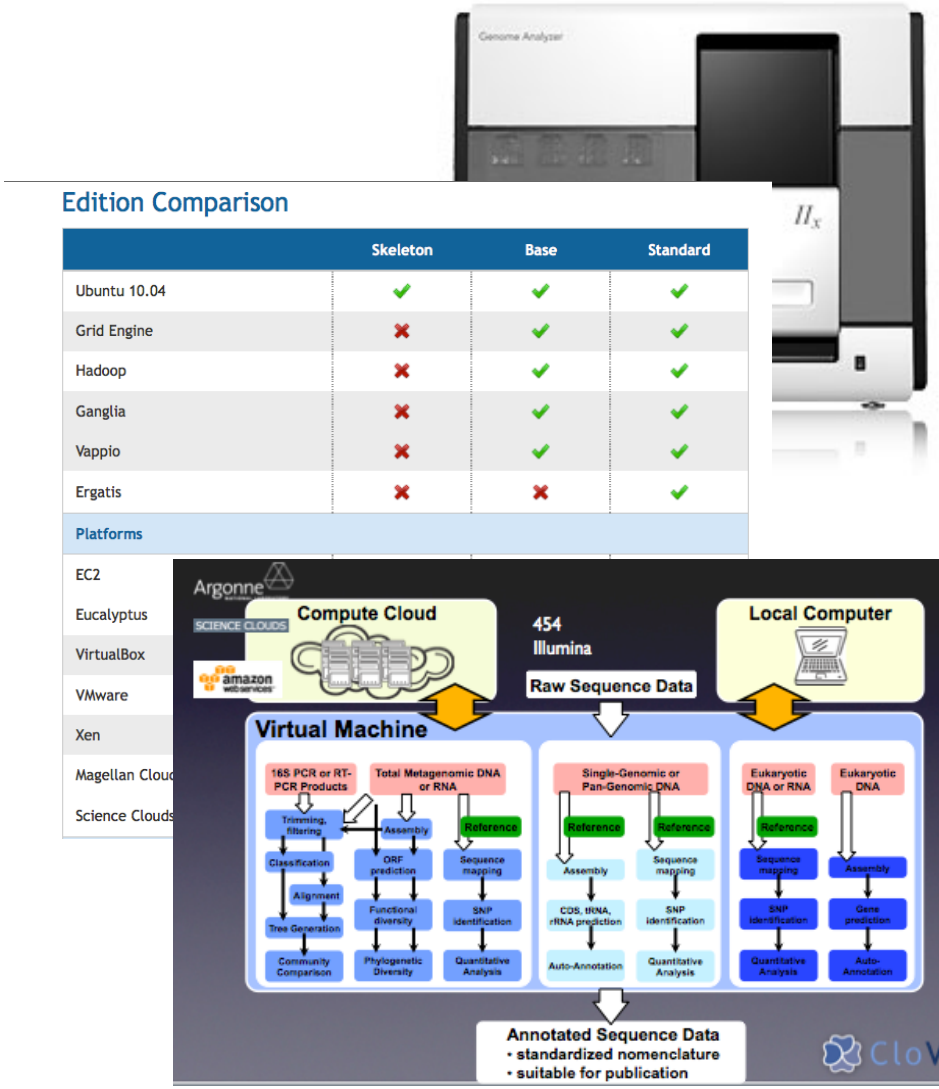
- Provide infrastructure for six observational astronomy survey projects
- Strategy:
  - Running on a Nimbus cloud on WestGrid
  - Dynamic Condor pool for astronomy
  - Appliance creation and management
- Status:
  - MACHO experiment Dark Matter search
  - In production operation since July 2010



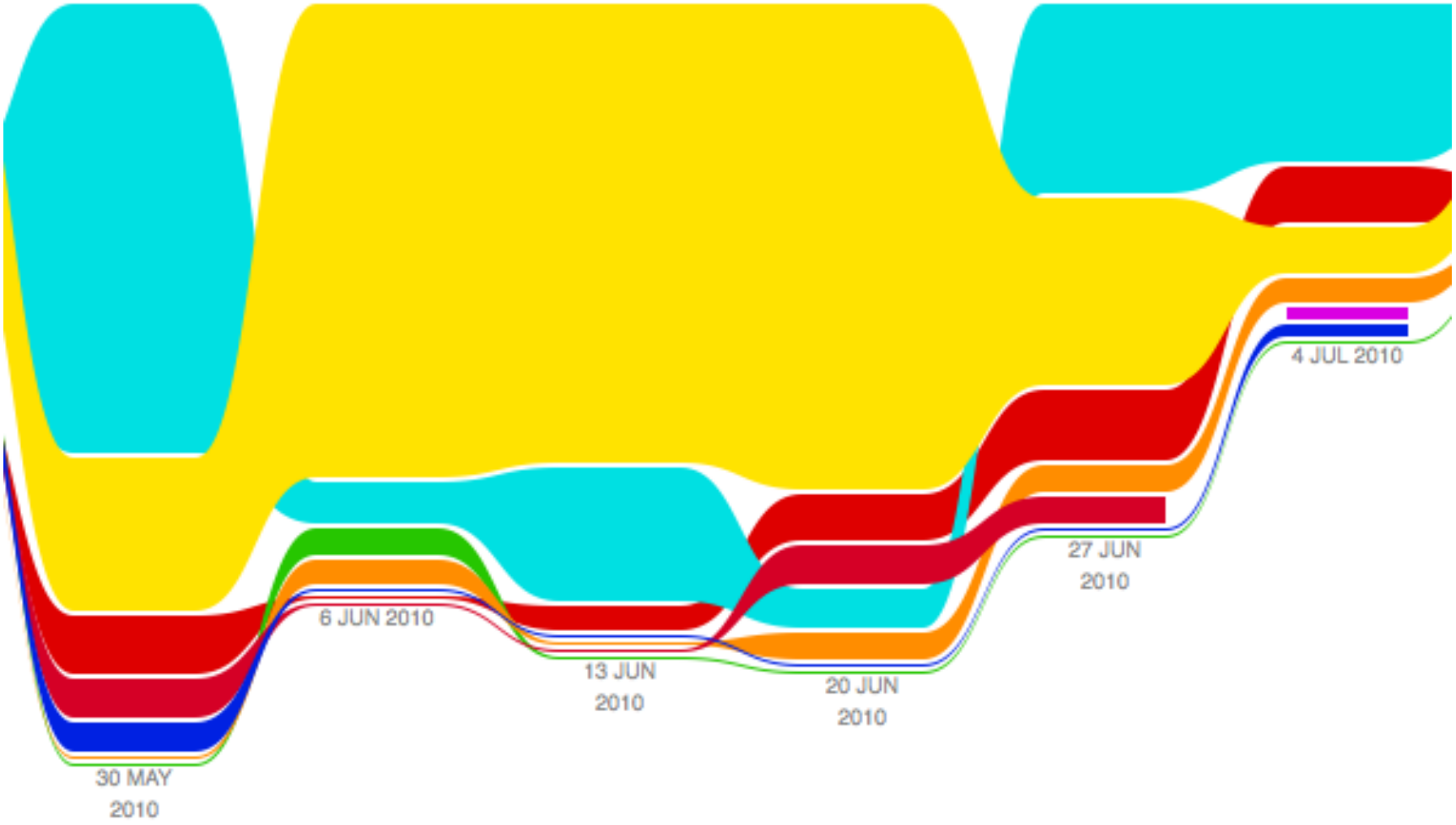
CloVR

Sam Angiuoli
Institute for Genome Sciences
University of Maryland School of Medicine

- The emergent need for processing
- A virtual appliance for automated and portable sequence analysis
- Strategy:
  - Running on Nimbus Science Clouds, Magellan and EC2
  - A platform for building appliances representing push-button pipelines
- Impact
  - From desktop to cloud
  - http://clovr.org



The Nimbus Team



The Nimbus Team

- Project lead: Kate Keahey, ANL&UC
- Committers:
  - Tim Freeman - University of Chicago
  - Ian Gable - University of Victoria
  - David LaBissoniere - University of Chicago
  - John Bresnahan - Argonne National Laboratory
  - Patrick Armstrong - University of Victoria
  - Pierre Riteau - University of Rennes 1, IRISA
- Github Contributors:
  - *Tim Freeman, David LaBissoniere, John Bresnahan, Pierre Riteau, Alex Clemesha, Paulo Gomez, Patrick Armstrong, Matt Vliet, Ian Gable, Paul Marshall, Adam Bishop*
- *And many others*
  - *See http://www.nimbusproject.org/about/people/*

Parting Thoughts

- Cloud computing is here to stay
- A change of paradigm -> a change of pattern
  - New technology requirements
    - Cost comparisons, scaling, data management, appliance management, etc.
  - New work patterns and new opportunities
- Better together: open source collaboration!