FACHTAGUNG „FUTURE BUSINESS CLOUDS“

DONNERSTAG, 6. JUNI 2013, 10.00–17.15 UHR – BUNDESMINISTERIUM FÜR WIRTSCHAFT UND TECHNOLOGIE (BMWi), INVALIDENSTRASSE 48, 10115 BERLIN

Beyond the Clouds: Trends in Cloud Computing for Science

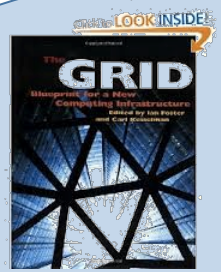
Kate Keahey

keahey@mcs.anl.gov

Argonne National Laboratory

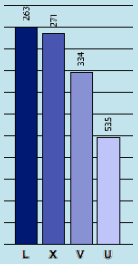
Computation Institute, University of Chicago

Cloud Computing: Looking Back



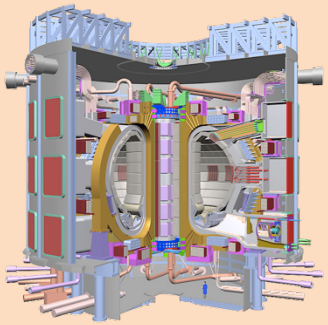
Enabler

Using remote resources



Enabler

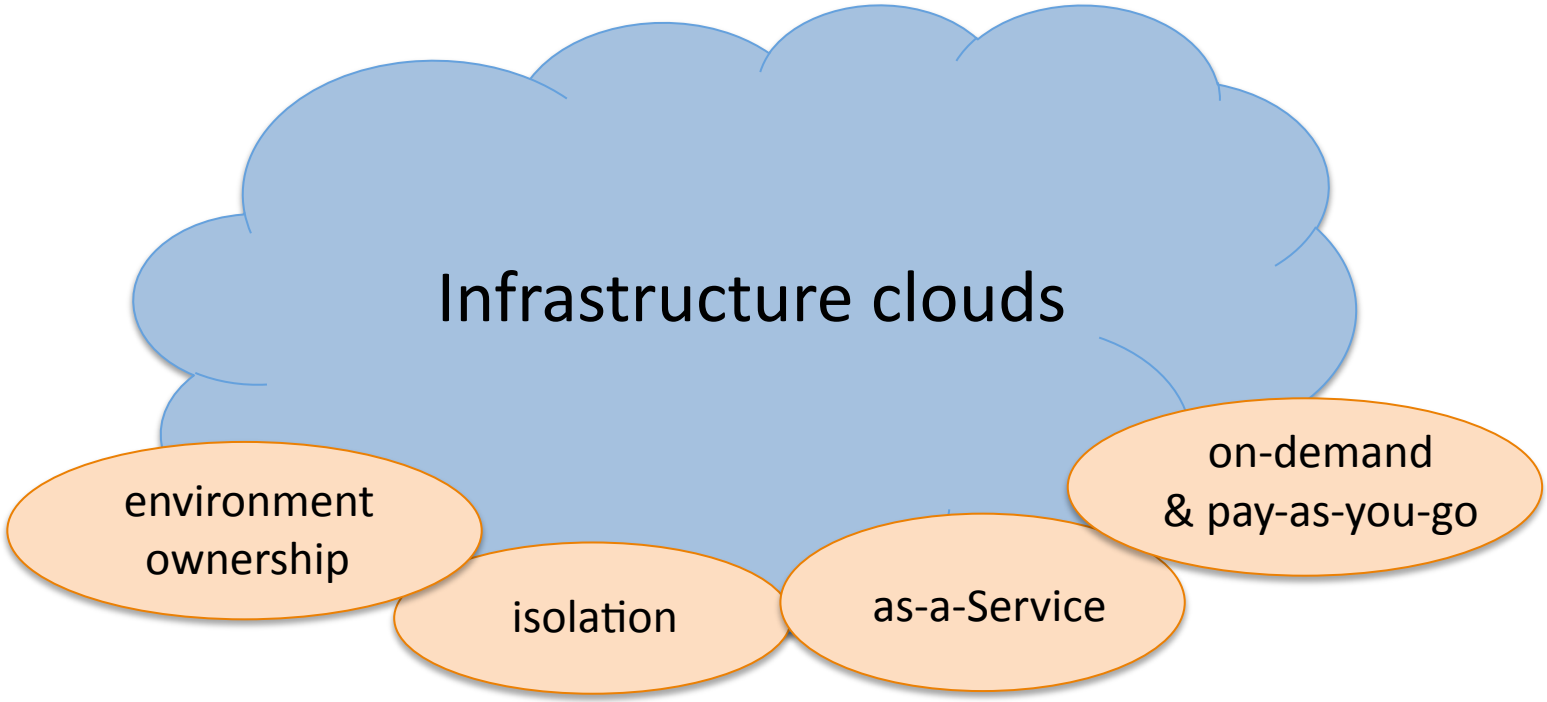
Environment ownership
Isolation



The Need

Environment ownership
On-demand processing

Infrastructure Clouds



Cloud Computing: Looking Forward

Cloud computing

On-demand, elastic outsourcing

Continuous Data Acquisition

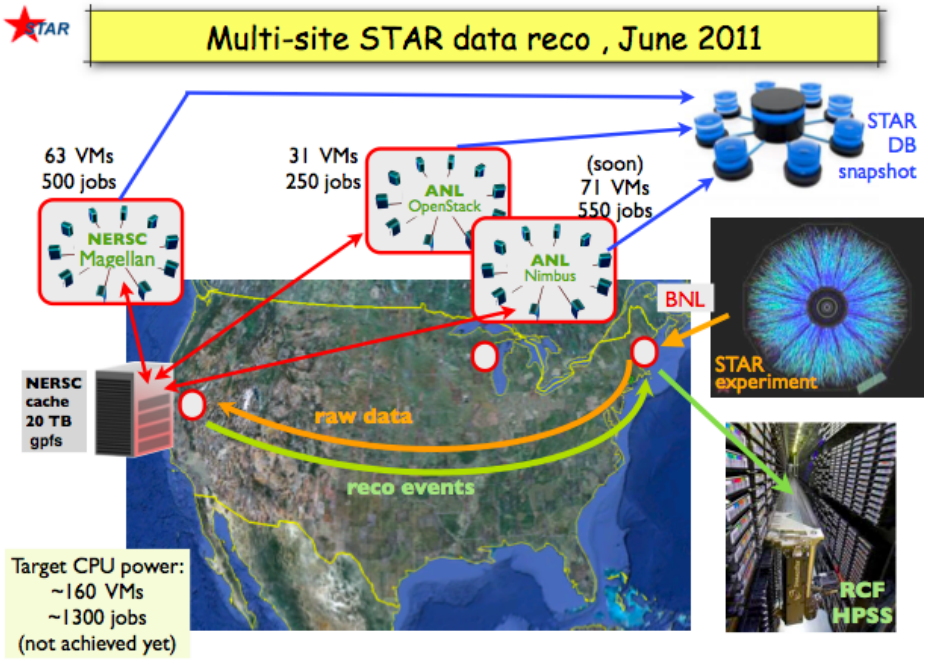
Cheaper and reliable sensors

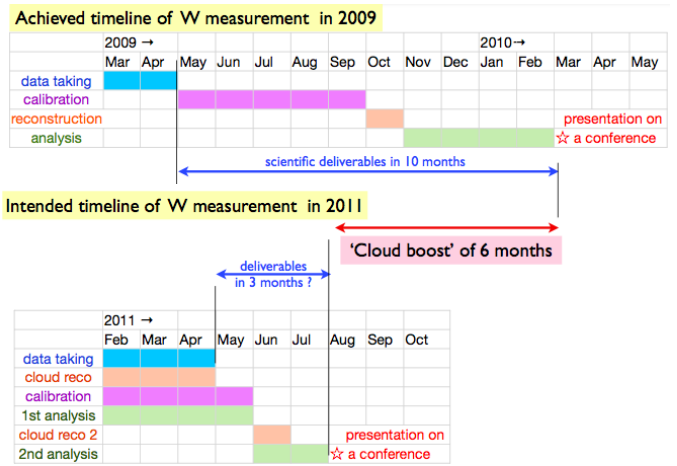
Event-based data processing

Observatories,
Fine-grained management

STAR ☆ The STAR Experiment

- Use case: cloud processing for W-boson reconstruction
- Overall achievement
  - "10 months became 3 months"
- Benefits:
  - Reduce "time to science"
  - Near real-time processing

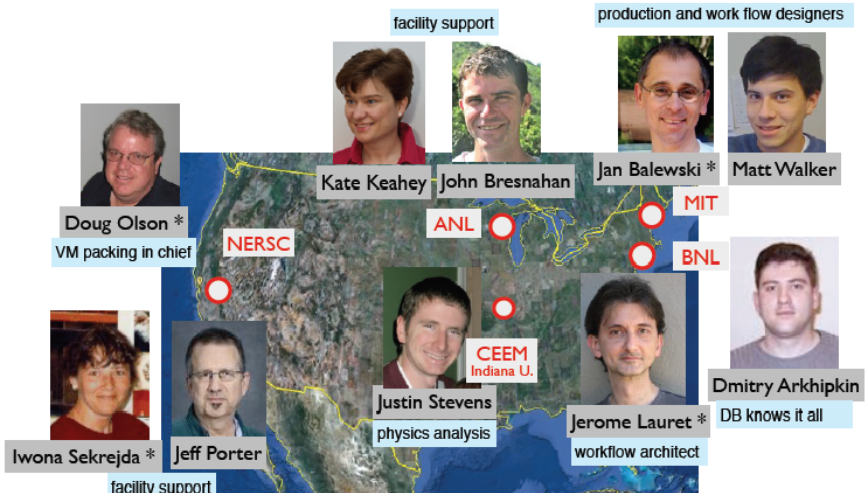


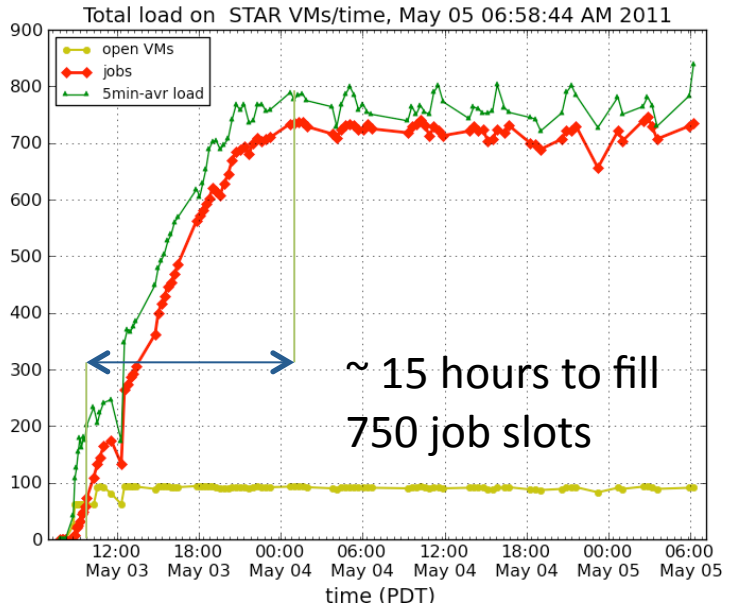


images courtesy of Jan Bielawsk, MIT

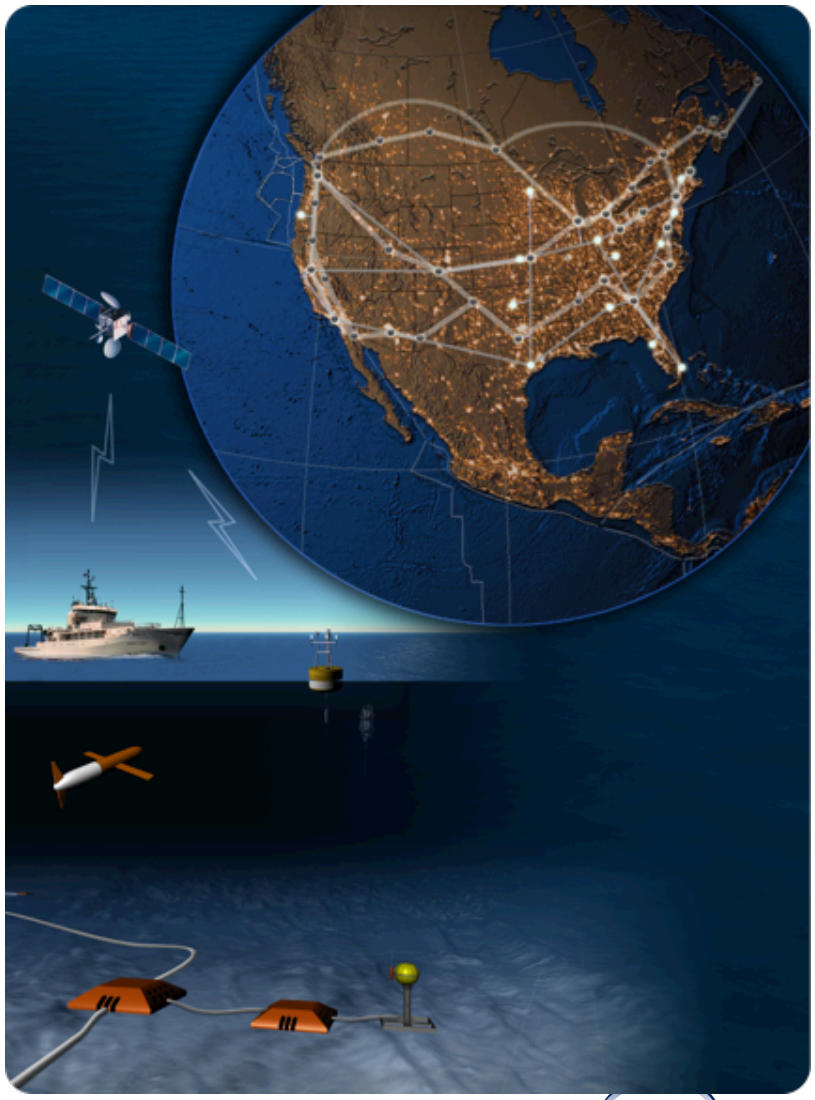
… and Its Challenges

- Cloud-related issues:
  - Not very on-demand
  - Involving 10 people
  - Hard to coordinate several clouds – manual and ad hoc
  - Lack of consistency, scalability, bad reaction time
  - No unified way to move/ store data
  - No autoscaling or easy adaptation for data transfer
  - Can't just define an appliance and have it work with several clouds
  - And many others…





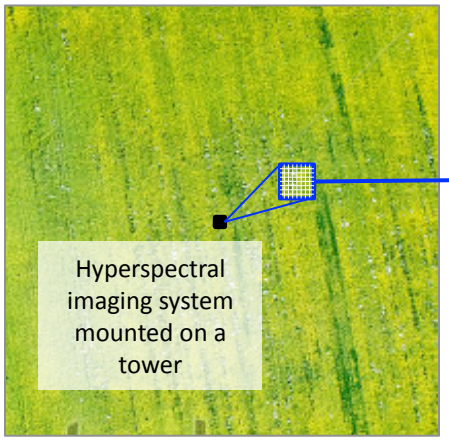
Ocean Observatory Initiative (OOI)

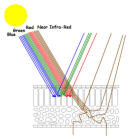


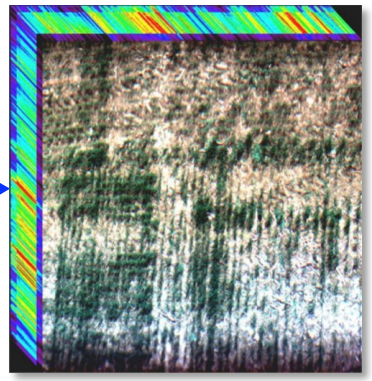
- Towards Observatory Science
- Sensor-driven processing
  - Real-time event-based data stream processing capabilities
  - Highly volatile need for data distribution and processing
  - An “always-on” service
- Nimbus team building platform services for integrated, repeatable support for on-demand science
  - High-availability
  - Auto-scaling
- From regional Nimbus clouds to commercial clouds

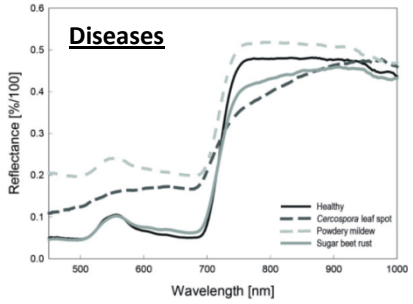
image courtesy of OOI Regional Scale Nodes program

Modeling Crops

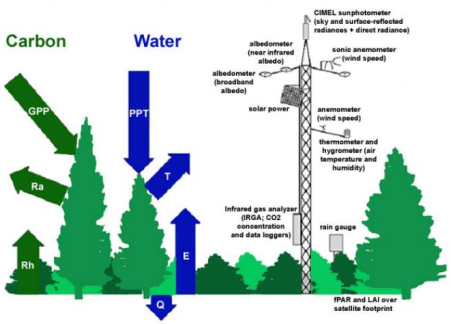




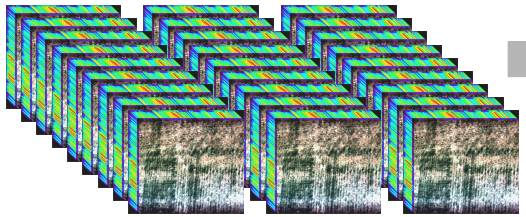


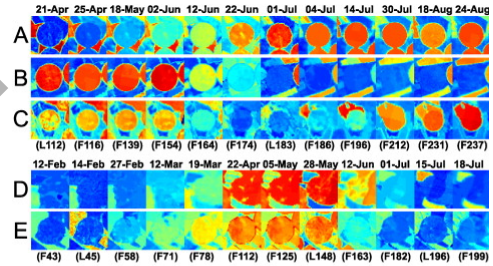


Spectral Indices*
NDVI • SAVI • EVI • NDWI



Multi-Temporal Hyperspectral Cubes and Spectral Index Images





images courtesy of Yuki Hamada, ANL

… and Their Challenges

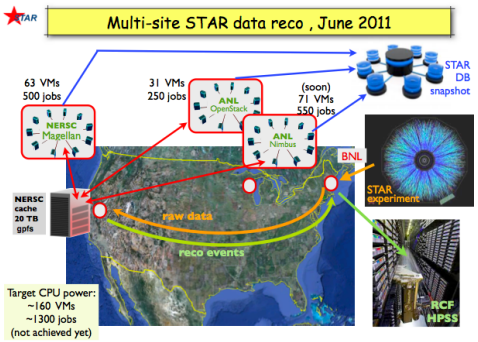
- Always-on
  - Managing deployment complexity
  - Automatic failure recovery
- Auto-scaling
  - Reaction times, costs, smart matchmaking
- Escalation pattern: from private to public/ commercial clouds
- Varied frequency and data streaming

Event-Driven Science

Sensor-based processing

Mobile devices

on-demand access
to data products and processing



Support for experiments

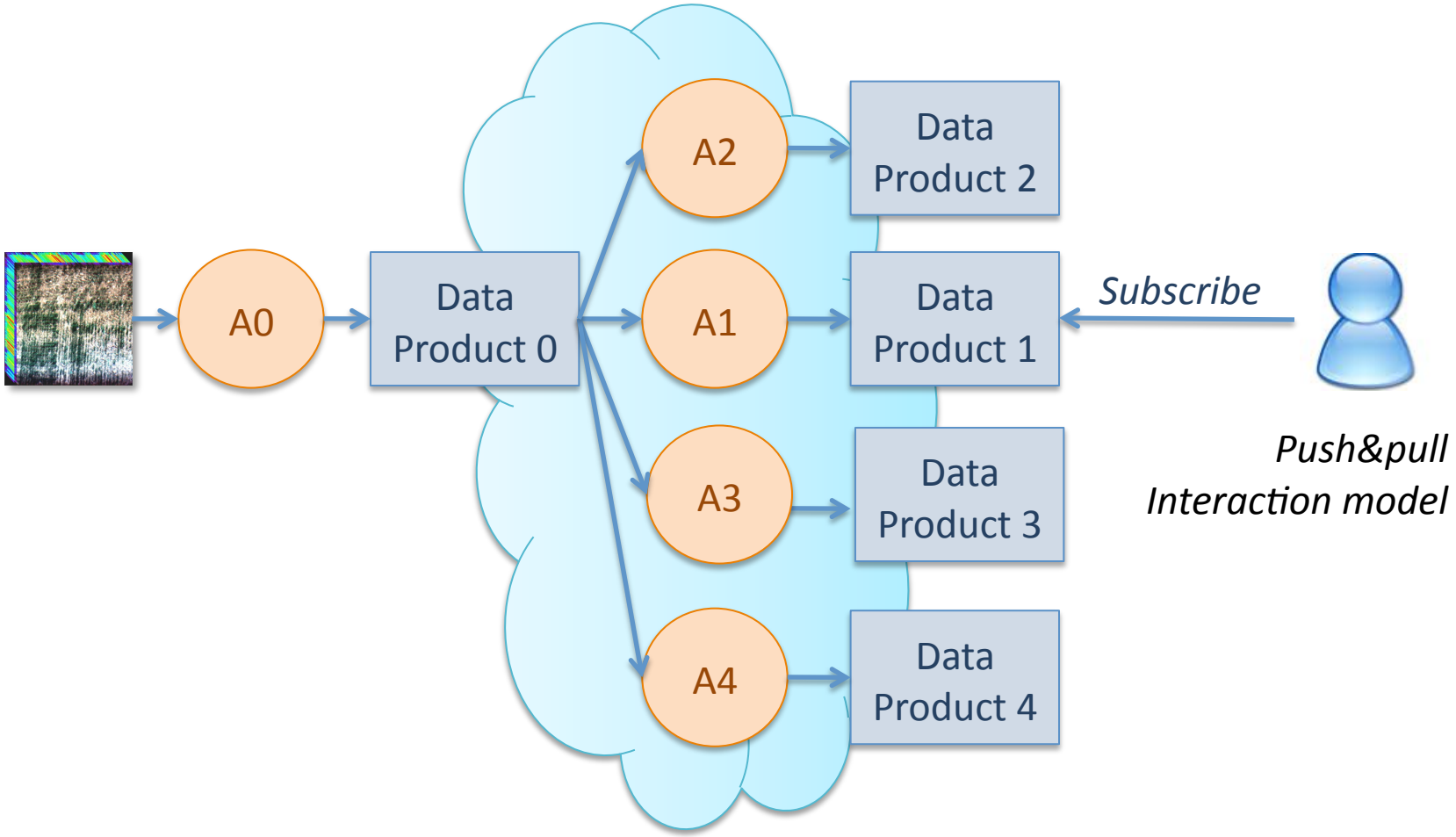
Brainstorming sessions

Electronic textbooks

Requirement Patterns

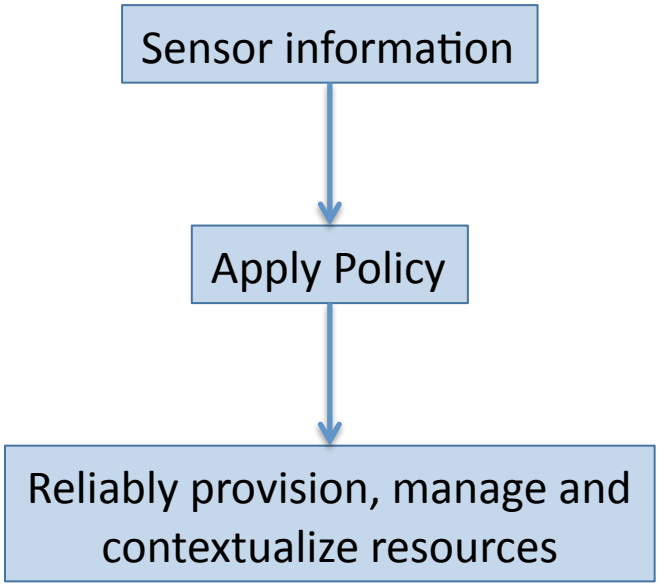
- On-demand, **elastic** processing…
  - Observatories, experiments, conference deadlines, fluctuating workload, growth management, mobile apps…
- … over **multiple providers**….
  - Risk-mitigation: not enough cycles, failure, market factors
  - From desktop to cloud
- … with emphasis on **ease of use**…
  - Automated provisioning of infrastructure resources
- …facilitates going from one-offs to **production** runs…
  - Steadily increasing in buy-in, complexity, and size
- …given the right **model**
  - Failure-prone environment
  - Achieving "uninterrupted power supply" or "computing power on tap

Data Distribution Network



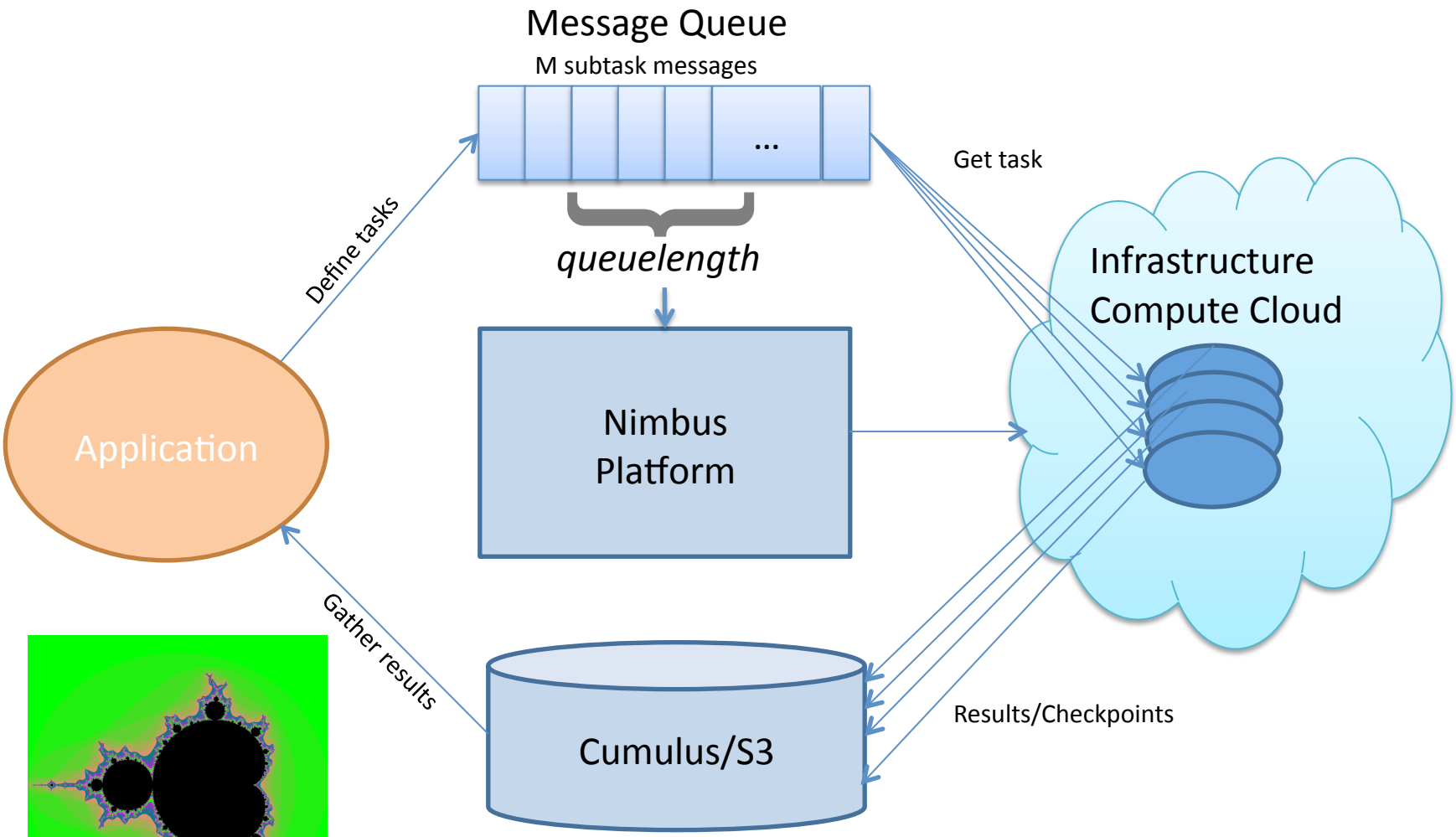
Scaling Resources with Nimbus

- Monitor scaling factors and failures
  - Generic/system qualities: deployment status, load, bank account, etc.
  - Application-specific qualities, e.g., a workload queue for ALiEn, PBS, AMQP, and others
- Evaluate against policies
- Scale and/or recover
  - For user components
  - For system components
  - Across different
- Release
  - Available resources



Try it out on FutureGrid!
www.nimbusproject.org/phantom

Using Nimbus Platform



NIMBUS www.nimbusproject.org

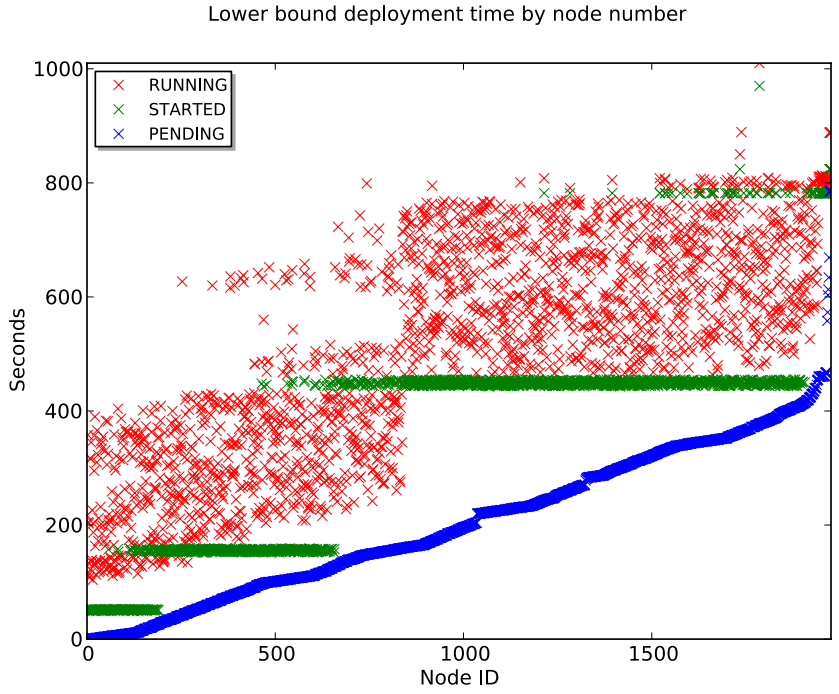
Adventures in Availability

Mean time between failures

$$A = \frac{MTBF}{MTBF + MTTR}$$

Mean time to repair

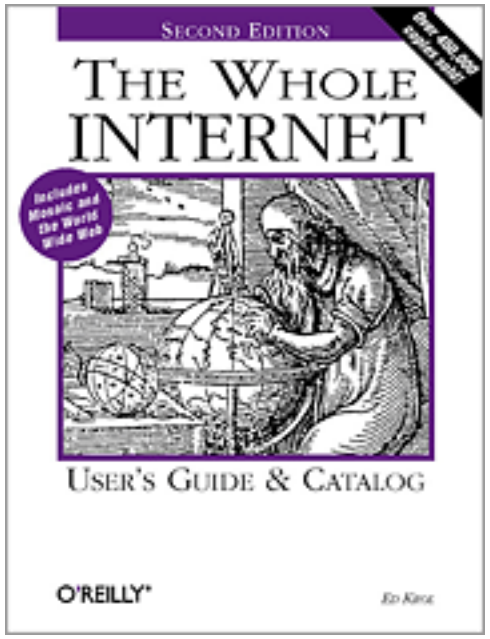
- Time to repair (TTR)
  - Diagnosis
  - Time to scale (TTS)
    - PENDING (request)
    - STARTED (deployment)
    - RUNNING (contextualization)



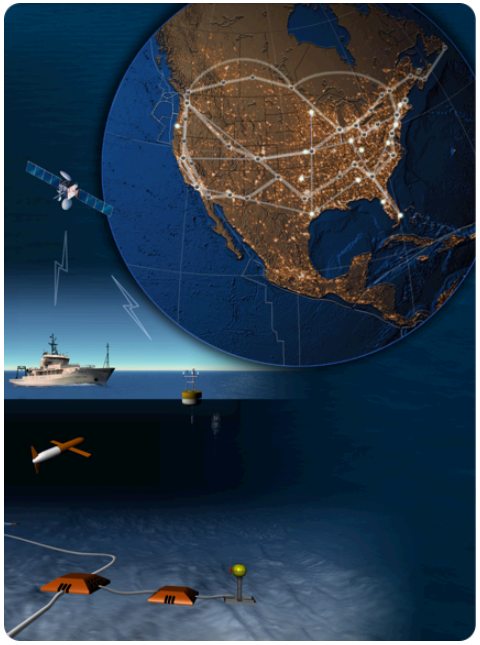
TTS: preliminary results for 2,000 VMs provisioned on AWS EC2

Parting Thoughts

Mature Innovation: the Internet



Maturing Innovation: the cloud



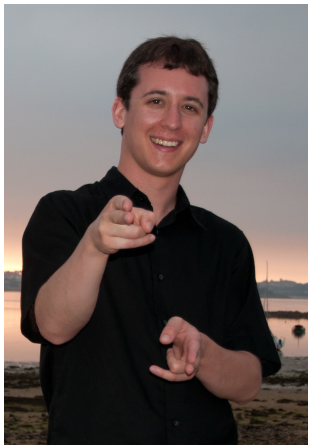
"Easy" is at least as important as "possible"

Nimbus Team

Kate Keahey
Argonne National Lab
University of Chicago

Patrick Armstrong
University of Chicago

David LaBissoniere
University of Chicago



Pierre Riteau
University of Chicago