Case study:

Cloud computing for near-real-time access to satellite data

John D. Evans

Global Science & Technology, Inc. (GST)
Greenbelt, MD

… But first: a mostly hype-free intro to Cloud Computing

John D. Evans

Global Science & Technology, Inc. (GST)
Greenbelt, MD

Cloud Computing origins

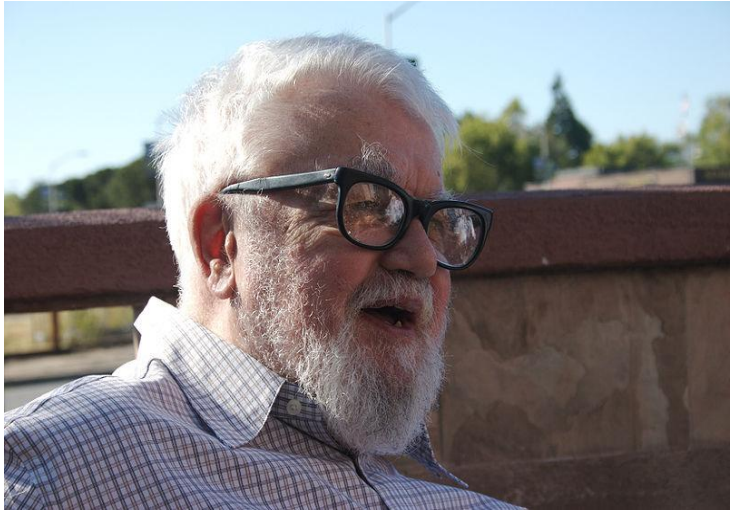


Photo by flickr.com/null0 via Creative Commons Attribution – Share Alike 2.0 Generic (CC BY-SA 2.0)

John McCarthy, 1961:

If computers of the kind I have advocated become the computers of the future, then ***computing may someday be organized as a public utility*** *just as the telephone system is a public utility... The computer utility could become the basis of a new and important industry.*

Cloud Computing origins

Ramnath Chellappa, 1997:
"Intermediaries in Cloud Computing:
A New Computing Paradigm"

"...where the boundaries of computing will be determined by economic rationale rather than technical limits."

http://www.bus.emory.edu/ram/

Plenty of enthusiasm for cloud computing

Photo: omb.gov

Vivek Kundra, 2010 *Federal Cloud Computing Strategy*:

"By leveraging shared infrastructure and economies of scale, cloud computing presents a compelling business model for Federal leadership. Organizations will be able to ***measure and pay for only the IT resources they consume; increase or decrease their usage*** *to match requirements and budget constraints; and leverage the shared underlying capacity of IT resources via a network."*

… and no shortage of critics

Richard Stallman, 2008:

"One reason you should not use web applications to do your computing is that you lose control. It's just as bad as using a proprietary program. Do your own computing on your own computer with your copy of a freedom-respecting program. If you use a proprietary program or somebody else's web server, you're defenseless. You're putty in the hands of whoever developed that software."

http://www.digitale-nachhaltigkeit.ch/2010/02/richard-stallman-en/

•… and no shortage of critics

Introducing Oracle's Exalogic "Cloud in a Box," Sept. 2010 *(Getty Images)*

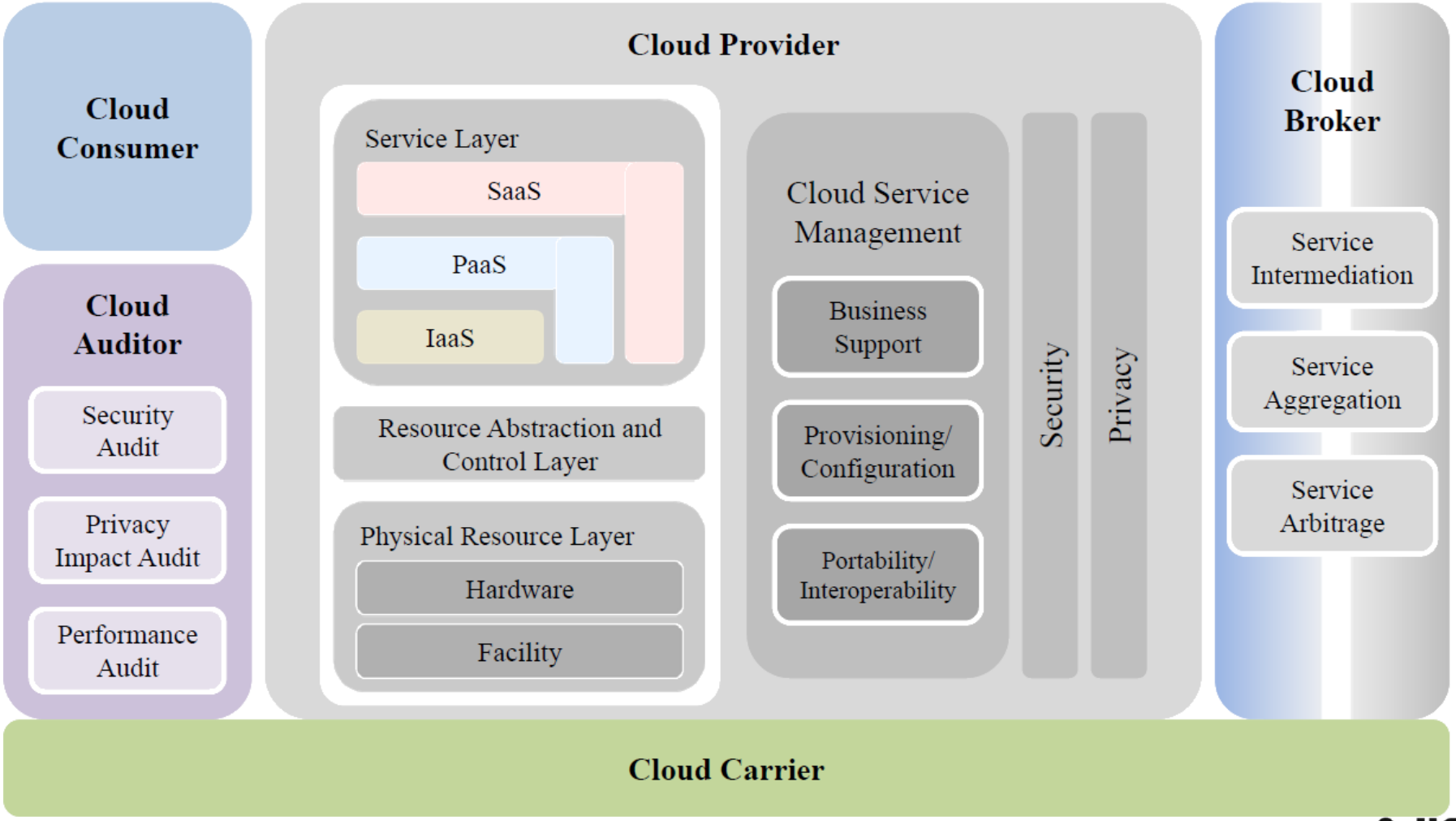
Larry Ellison, 2008:

'What is it? What is it? … Is it - 'Oh, I am going to access data on a server on the Internet.' That is cloud computing? … I don't know what we'll do differently in light of cloud computing other than change the wording on some of our ads."

Cloud Computing Definition (NIST)

- "**Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources ... that can be rapidly provisioned and released with minimal management effort or service provider interaction.**"
  - On-Demand Self-Service
  - Broad Network Access
  - Resource Pooling
  - Rapid Elasticity
  - Measured Service
- Software as a Service *(SaaS – e.g., IRS E-File, Gmail)*
- Platform as a Service *(PaaS – e.g., Microsoft Azure, Google vis)*
- Infrastructure as a Service *(IaaS – e.g., Amazon Web Services, Rackspace)*

Cloud Computing Reference Architecture



NIST
National Institute of Standards and Technology
U.S. Department of Commerce

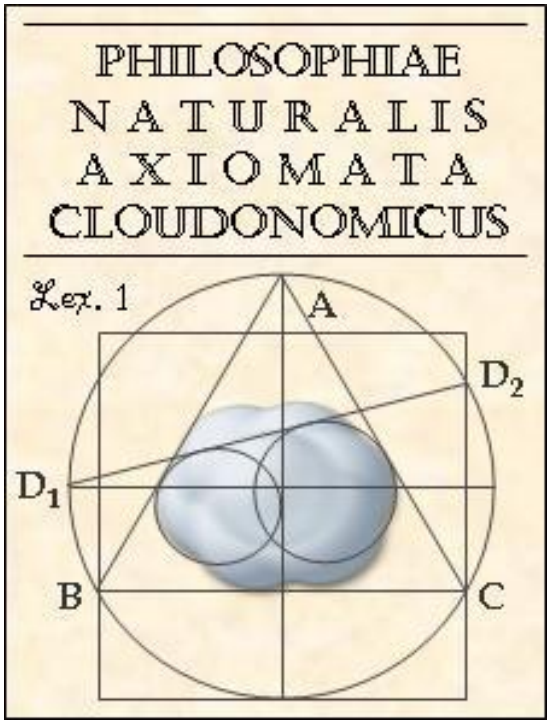
FGDC GeoCloud pilot projects

- DOI – National Wetlands Inventory: Wetlands Mapper
- USDA – CDSI Web Services
- USDA – Geospatial Public Catalog
- EPA – Lakes and Ponds
- USGS – GEOSS Clearinghouse Catalog
- Census – TIGER/Line Downloads
- USGS – VGI OpenStreetMap service
- NOAA – Particles in the Cloud (dispersion/diffusion simulation)
- NOAA – Integrated Ocean Observing System Catalog & Viewer
- NOAA – ERDDAP (CoastWatch / Environmental Research Div.)

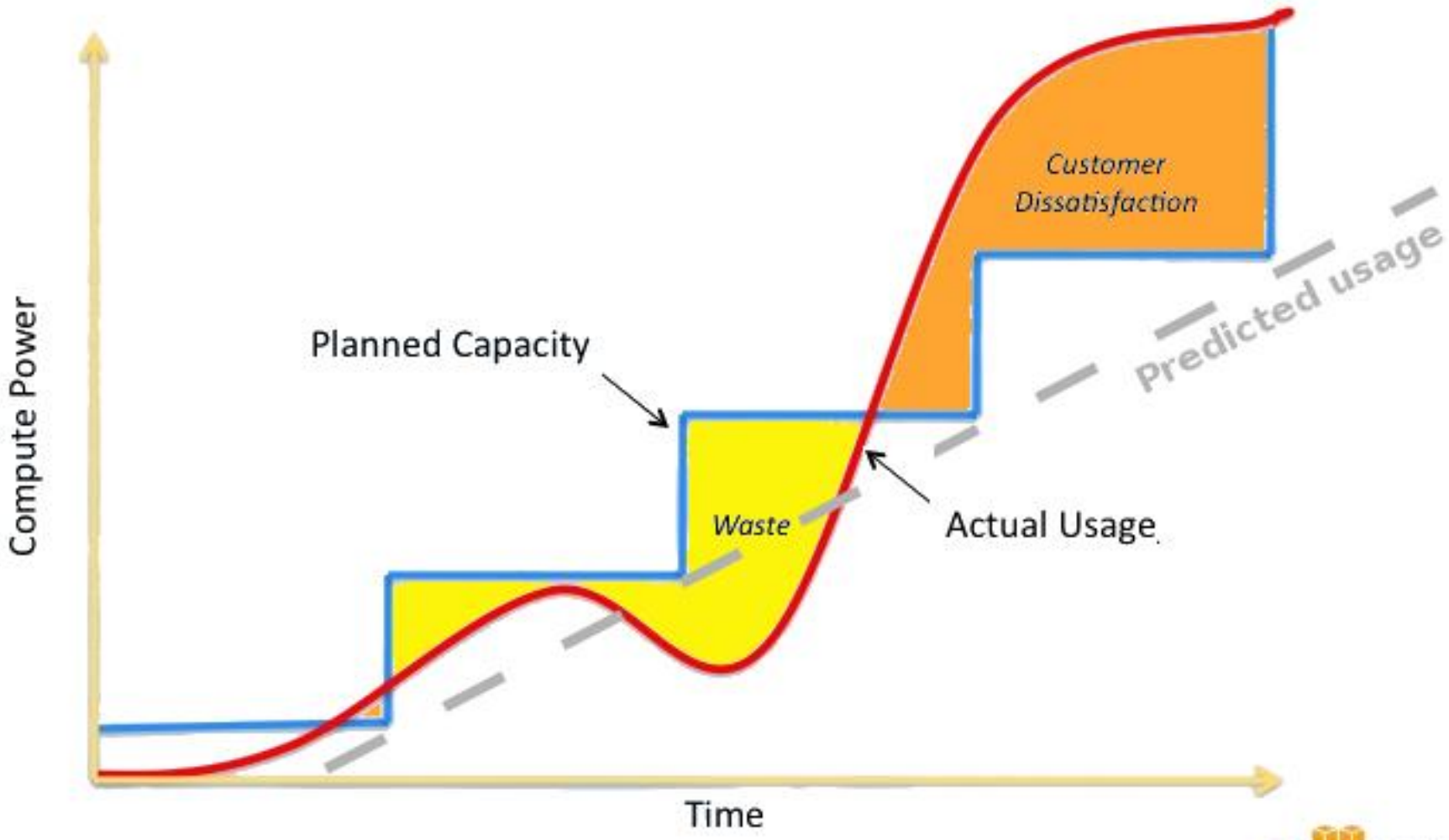
"Cloudonomics"

de Vermont – http://gigaom.com/2008/09/07

- Utility services cost less even though they cost more *(per unit)*
- The peak of the sum is never greater than the sum of the peaks
- Average unit costs are reduced by distributing fixed costs over more units of output *(economies of scale)*
- Superiority in numbers *(resilience to attack)*
- Aggregate demand is smoother than individual
- **On-demand trumps forecasting**
- **Space vs. time is a continuum**
- Dispersion is the inverse square of latency
- Don't put all your eggs in one basket.
- An object at rest tends to stay at rest.

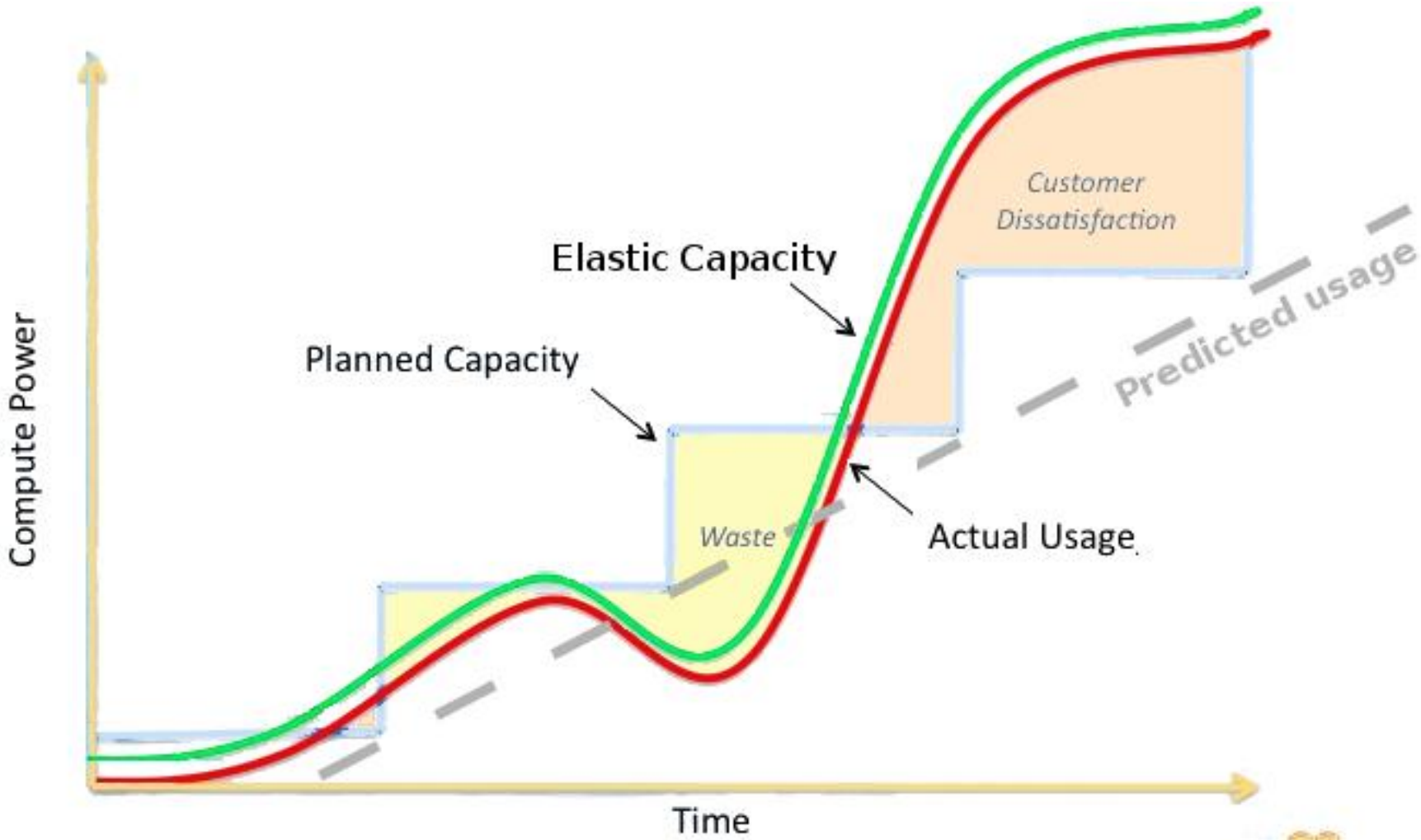


IT capacity forecasting



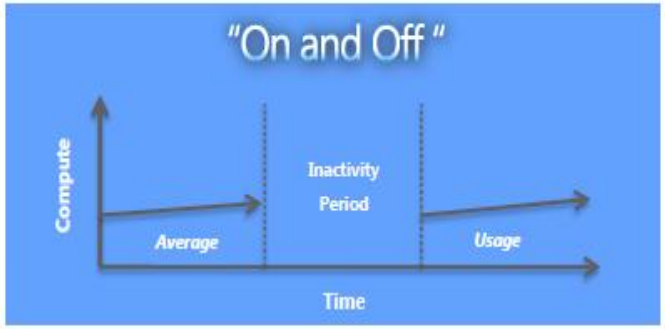
amazon
webservices

Elastic IT provisioning



amazon web services

Workloads optimal for cloud



- On & off workloads (e.g. batch job)
- Over provisioned capacity is wasted
- Time to market can be cumbersome

- Successful services needs to grow/scale
- Keeping up w/ growth is big IT challenge
- Complex lead time for deployment

- Unexpected/unplanned peak in demand
- Sudden spike impacts performance
- Can't over provision for extreme cases

- Services with micro seasonality trends
- Peaks due to periodic increased demand
- IT complexity and wasted capacity

Windows Azure

Key Concepts: Infrastructure as a Service

- Virtual Machine **Image**
  - Operating system + any installed server or application software
  - Complete machine state, bundled & stored (~ “hibernate file”)
- Virtual Machine **Instance Type**
  - “Resource bundles” of a particular size
    - CPU speed, number of cores, memory, root partition size
  - May be abstracted from hardware – or not
    - Bandwidth to storage; bandwidth between cores
    - Exclusive use of CPU, RAM, *etc.*
- Virtual Machine **Instance**
  - A Machine Image running on a particular Instance Type
  - A running OS w/ RAM, filesystem, IP address, services, etc.

Case study:

Cloud computing for near-real-time access to satellite data

John D. Evans

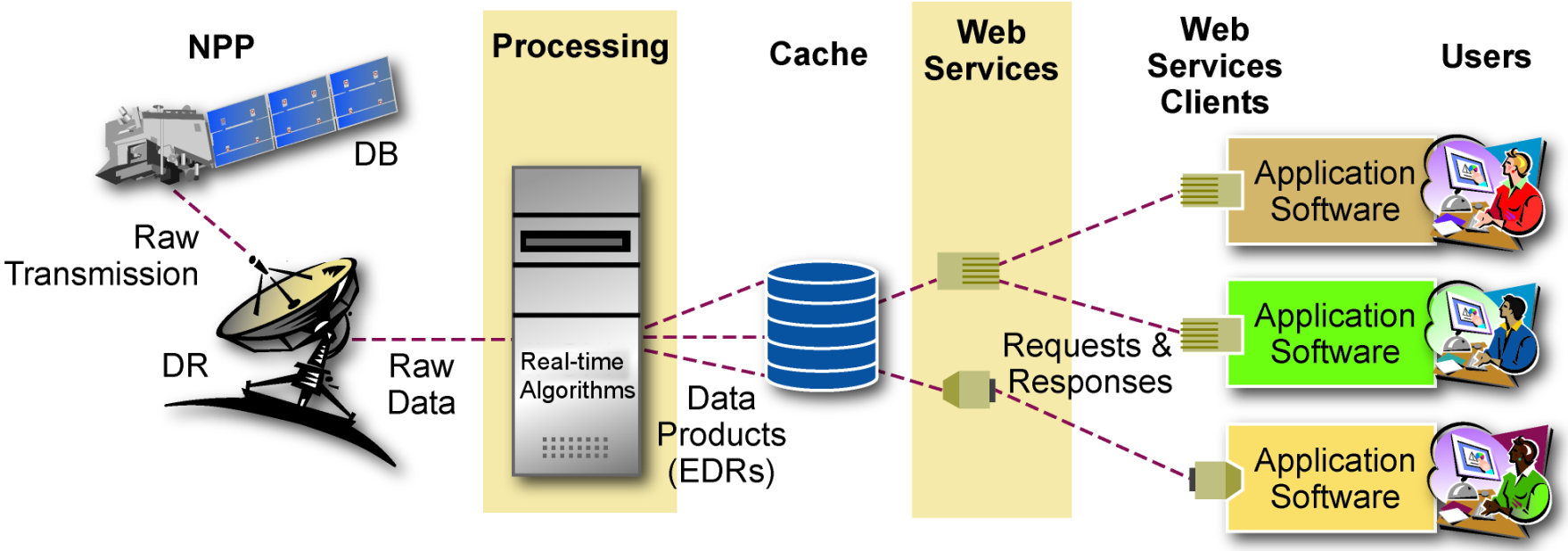
Eduardo G. Valente, Jr.

Samir R. Chettri

Global Science & Technology, Inc. (GST)

Greenbelt, MD

Case study: Near real-time workflow



Case study: data streams

- *Gearing up:* Aqua MODIS Direct Broadcast
  - Using data and software from NASA's Direct Readout Laboratory (DRL)

  http://directreadout.sci.gsfc.nasa.gov

  - *In particular:* International Polar Orbiter Processing Package (IPOPP)

- *Target:* NPP Direct Broadcast
  - NPP scheduled for Fall 2011 launch
  - NASA/NOAA ground system still in progress
    - This makes Direct Broadcast a vital alternative

Use case: weather forecasting

- Working with NASA's Short-term Prediction Research and Transition program

  http://weather.msfc.nasa.gov/sport

Why we went to the cloud

- Our in-house proof-of-concept quickly showed that
  - We would need a lot more bandwidth
  - We might need a full-blown data center for peak loads
  - Ensuring real-time throughput cost-effectively is hard
- Cloud Computing offered a way to
  - Get a lot of bandwidth on demand
  - Grow our computing capacity easily
  - Try out a variety of configurations without risk
- Infrastructure as a Service (IaaS) offered an easy transition
  - Amazon Elastic Compute Cloud (EC2)
  - NASA Nebula

Key concerns for our case study

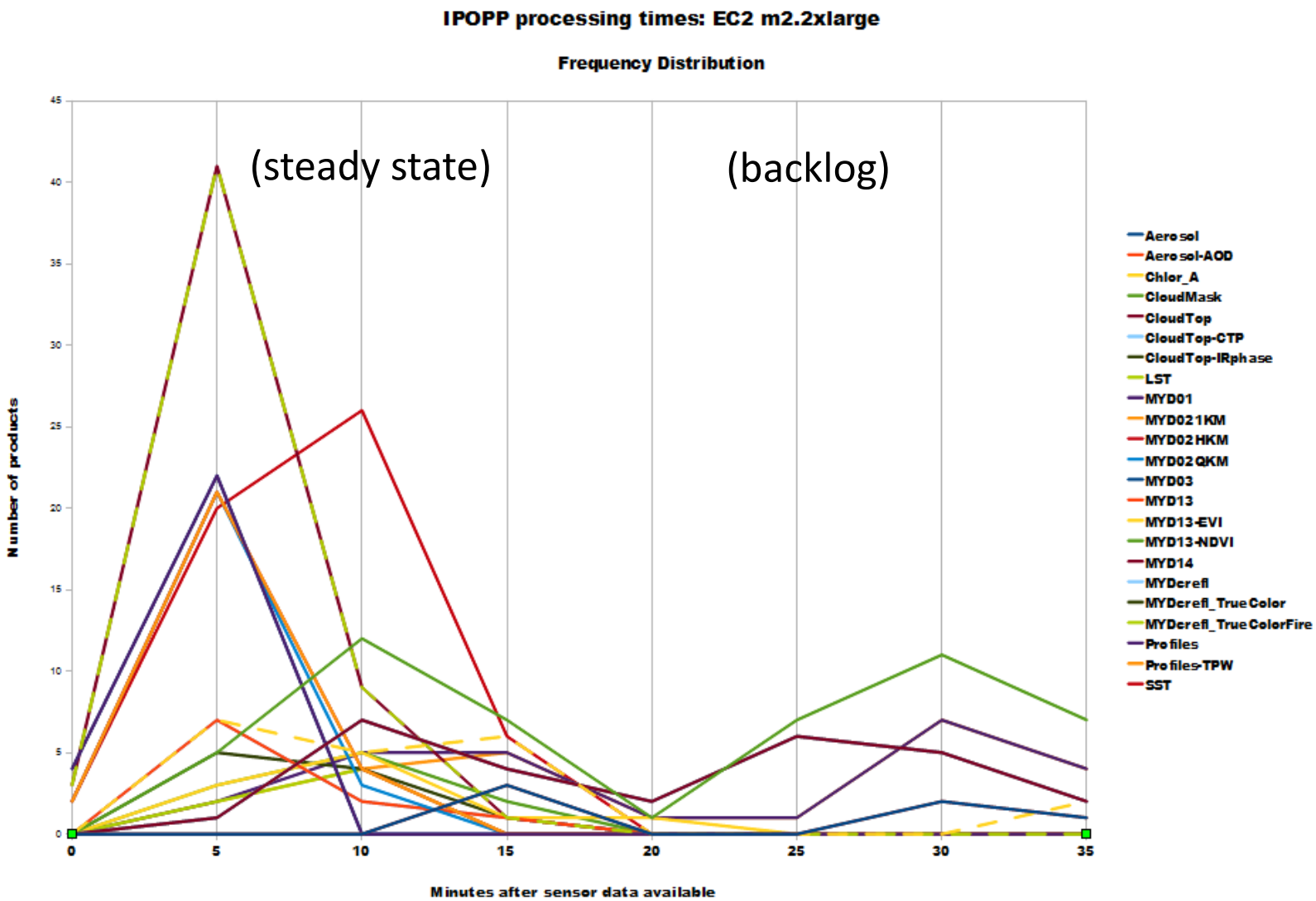
- Proof of concept
  - Could we replicate our processing chain in the cloud?
- Task effectiveness
  - Could we provide near-real-time processing and access via the cloud?
- Cost effectiveness
  - Would it be worth doing this in the cloud?

Instance types tested

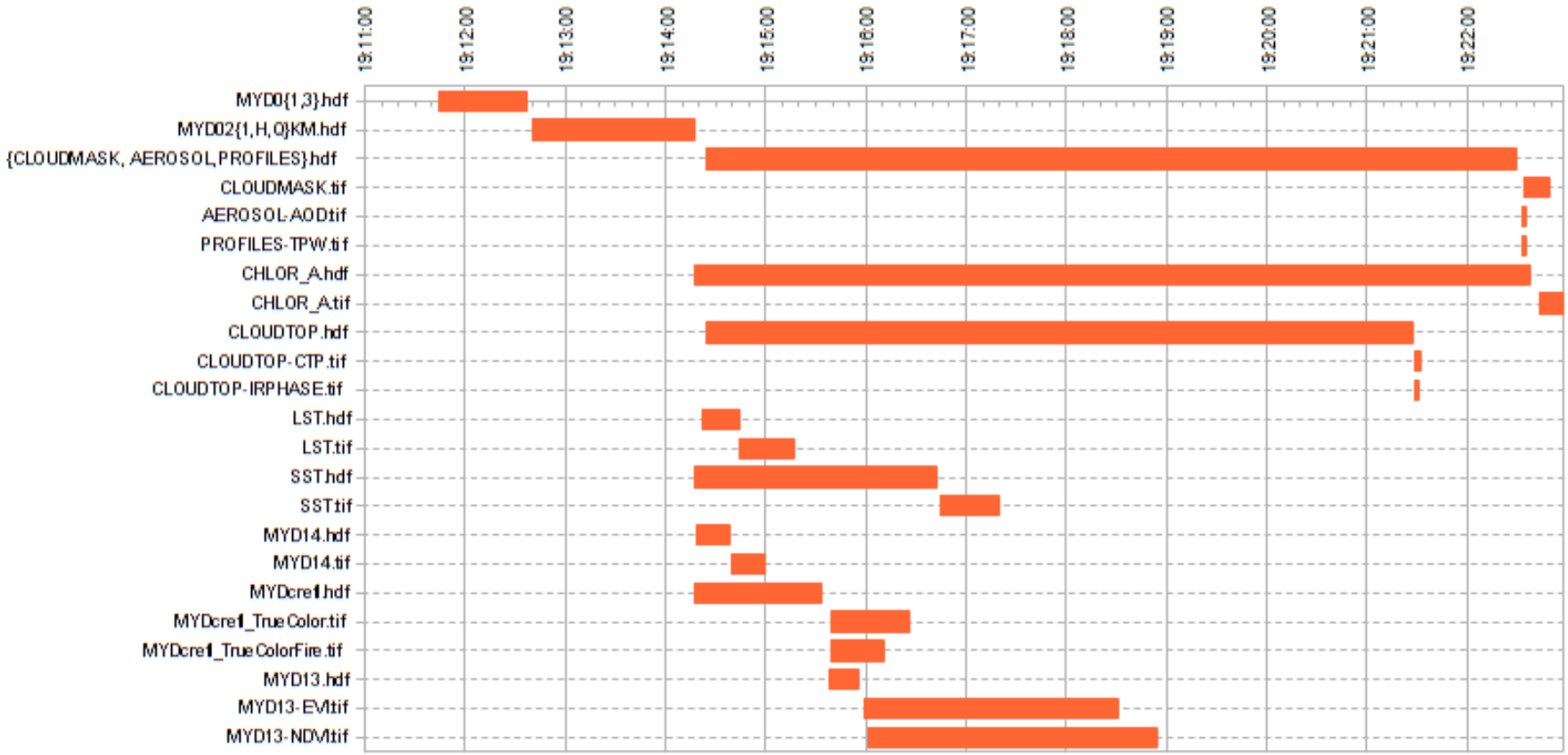
Instance Type	CPU type	PassMark Score	CPU cores	RAM (GB)	Cost ($/hr)
In-house server	Xeon X3220	3,114	4	8	
Amazon EC2 m1.large	Xeon E5430	4,158	2	7.5	$0.34
Amazon EC2 m1.xlarge			4	15	$0.68
Amazon EC2 c1.xlarge	Xeon E5506	3,374	8	7	$0.68
Amazon EC2 m2.xlarge	Xeon X5550	5,232	2	17.1	$0.50
Amazon EC2 m2.2xlarge			4	34.2	$1.00
Amazon EC2.4xlarge			8	68.4	$2.00
NASA Nebula m1.large	QEMU virtual CPU *(Nehalem / Windermere)*		4	8	
NASA Nebula m1.xlarge			8	16	

(For detailed cloud-computing benchmarks see http://cloudharmony.com*)*

EC2 m2.2xlarge performance

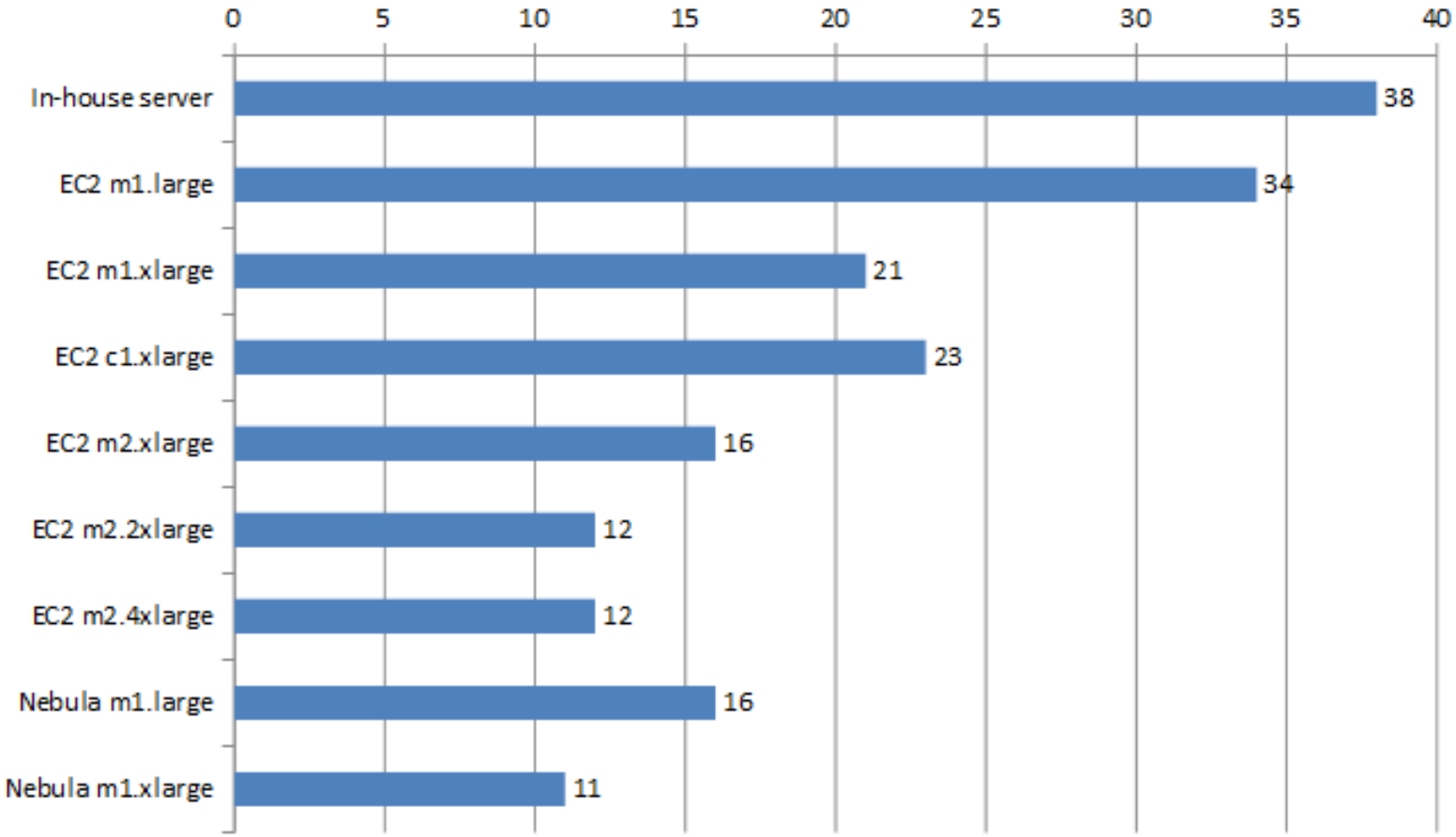


Simultaneous execution: a single run



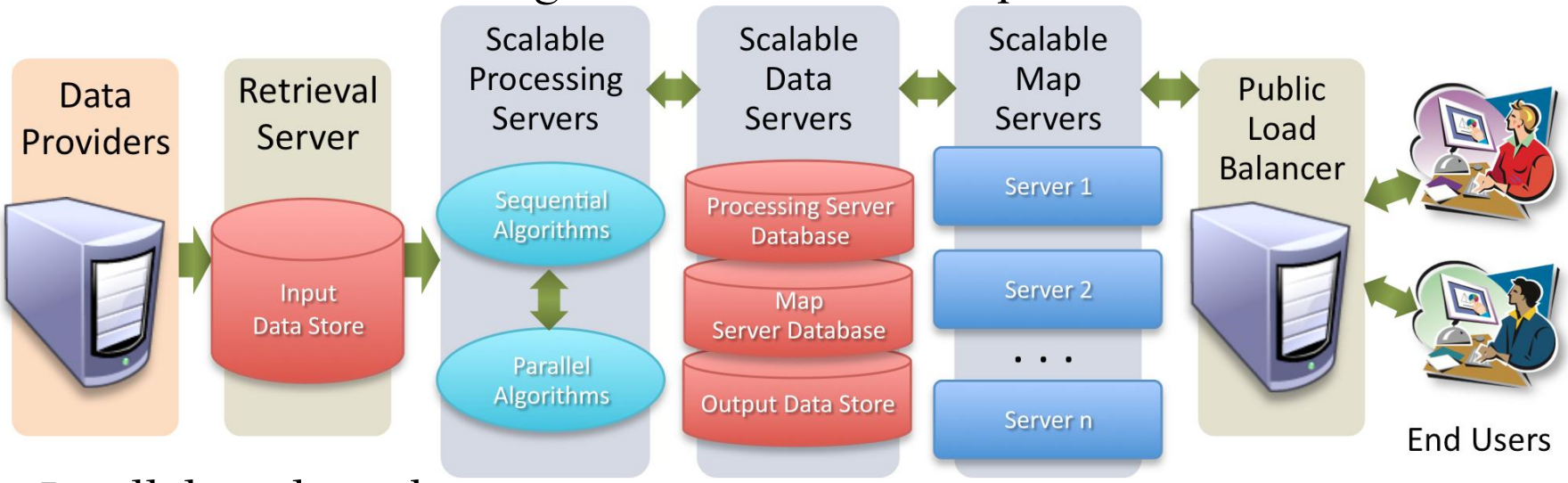
EC2 / Nebula performance *(steady-state)*

Processing latency (in minutes) for last product derived from largest (840MB) input images



Ways to improve performance

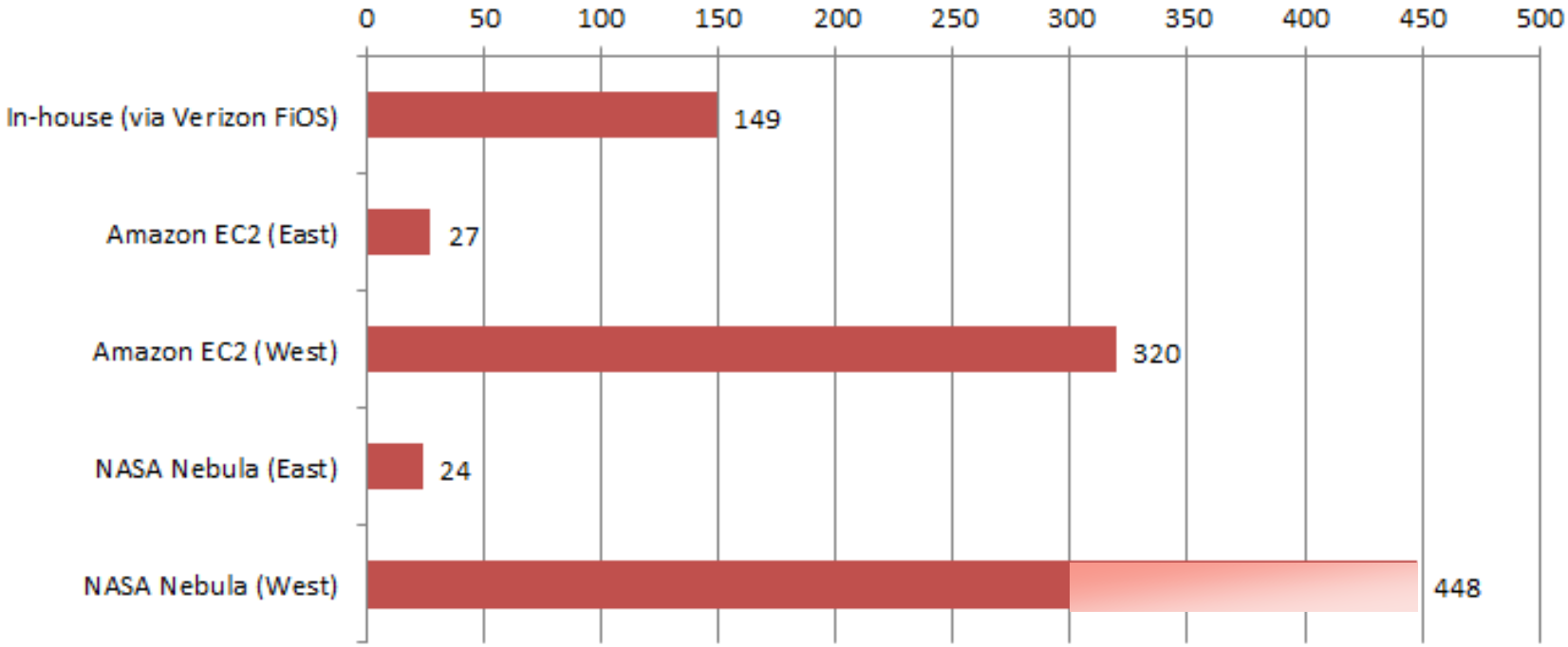
- Optimize IPOPP algorithm scheduling
- Use lots of fast CPU cores to handle backlogs
- Distribute services & algorithms across multiple machines



- Parallelize algorithms
- Compile algorithms to run on GPUs
- Streamed data / streamed processing

Cloud performance: bandwidth

Time (in seconds) for transferring largest (840MB) input images from receiver (DRL)



Cloud computing costs *(Amazon EC2)*

- CPU usage *(see table)*
  - Pennies per hour add up!
  - Note 100:1 cost ratio
- Data Storage
  - EBS: $100/TB/month
  - S3: $37-140/TB/month
- Data Transfer
  - In: $100/TB
  - Out: $80-150/TB
- Other
  - SQL queries; I/O requests
  - Snapshot GETs/PUTs

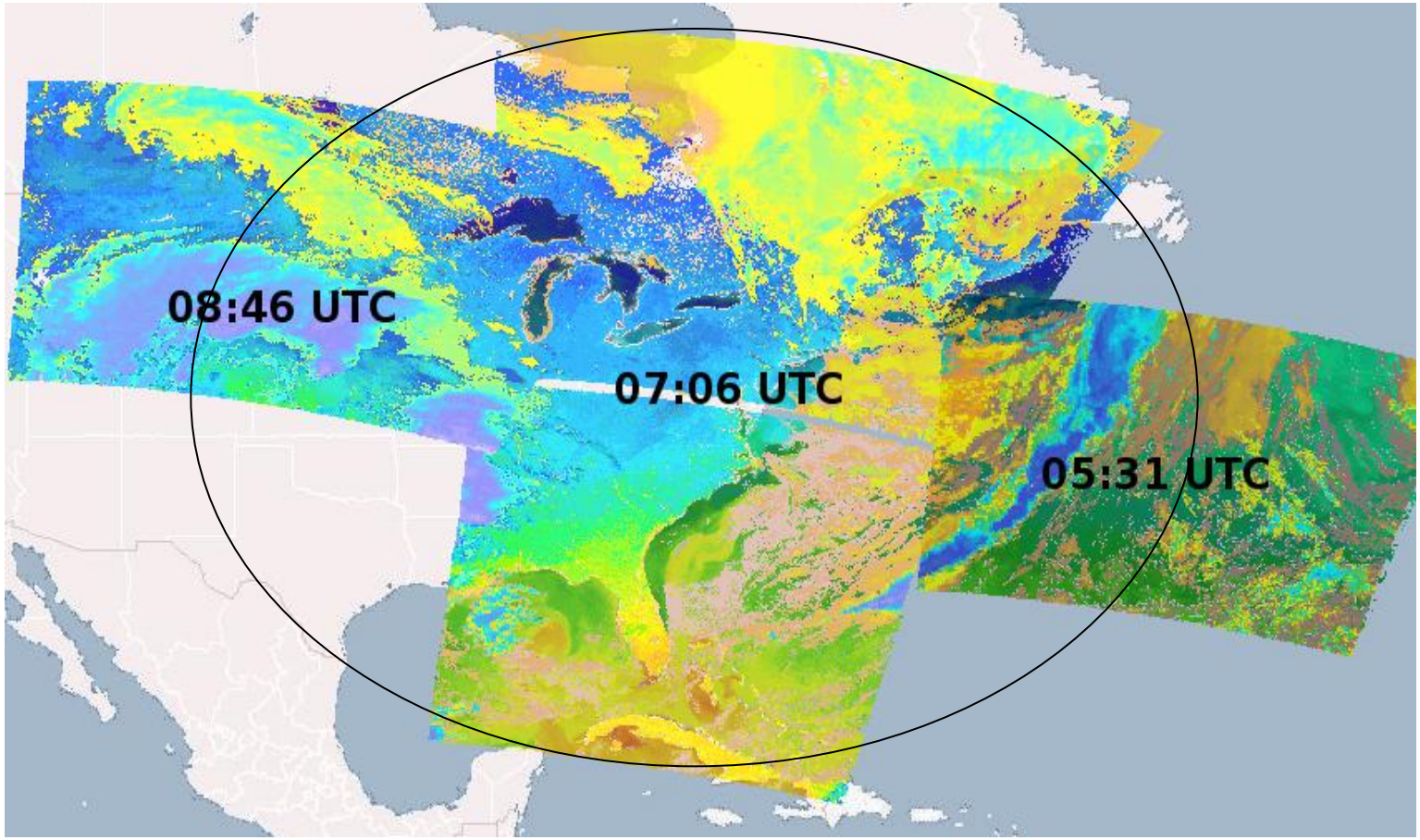
Instance type	cost/hr	/day	/mo	/yr
t1.micro	$0.02	$0.48	$15	$175
m1.small *(default)*	$0.085	$2	$62	$745
m1.large	$0.34	$8	$248	$2,978
m1.xlarge	$0.68	$16	$496	$5,957
c1.xlarge	$0.68	$16	$496	$5,957
m2.xlarge	$0.50	$12	$365	$4,380
m2.2xlarge	$1.00	$24	$730	$8,760
m2.4xlarge	$2.00	$48	$1,460	$17,520
cc1.4xlarge *(cluster compute)*	$1.60	$38	$1,169	$14,026
cg1.4xlarge *(GPU cluster)*	$2.10	$50	$1,534	$18,409

Ways to reduce cloud costs

- Reserved instances: fixed sum + lower hourly rate
  - Worthwhile if usage exceeds 11 hrs/day (1 year) … 8.5 hrs/day (2 years) … 5.5 hrs/day (3 years)
- Spot instances
  - Spot price: as little as 1/3 of regular hourly rate
  - Pick a rate ($/hr) ceiling; run until spot price exceeds ceiling
  - Useful when load-balancing across many identical machines
- **Scale up & down as needed**
  - A polar-orbiting satellite may let us save 60% or more by scaling up & down several times a day

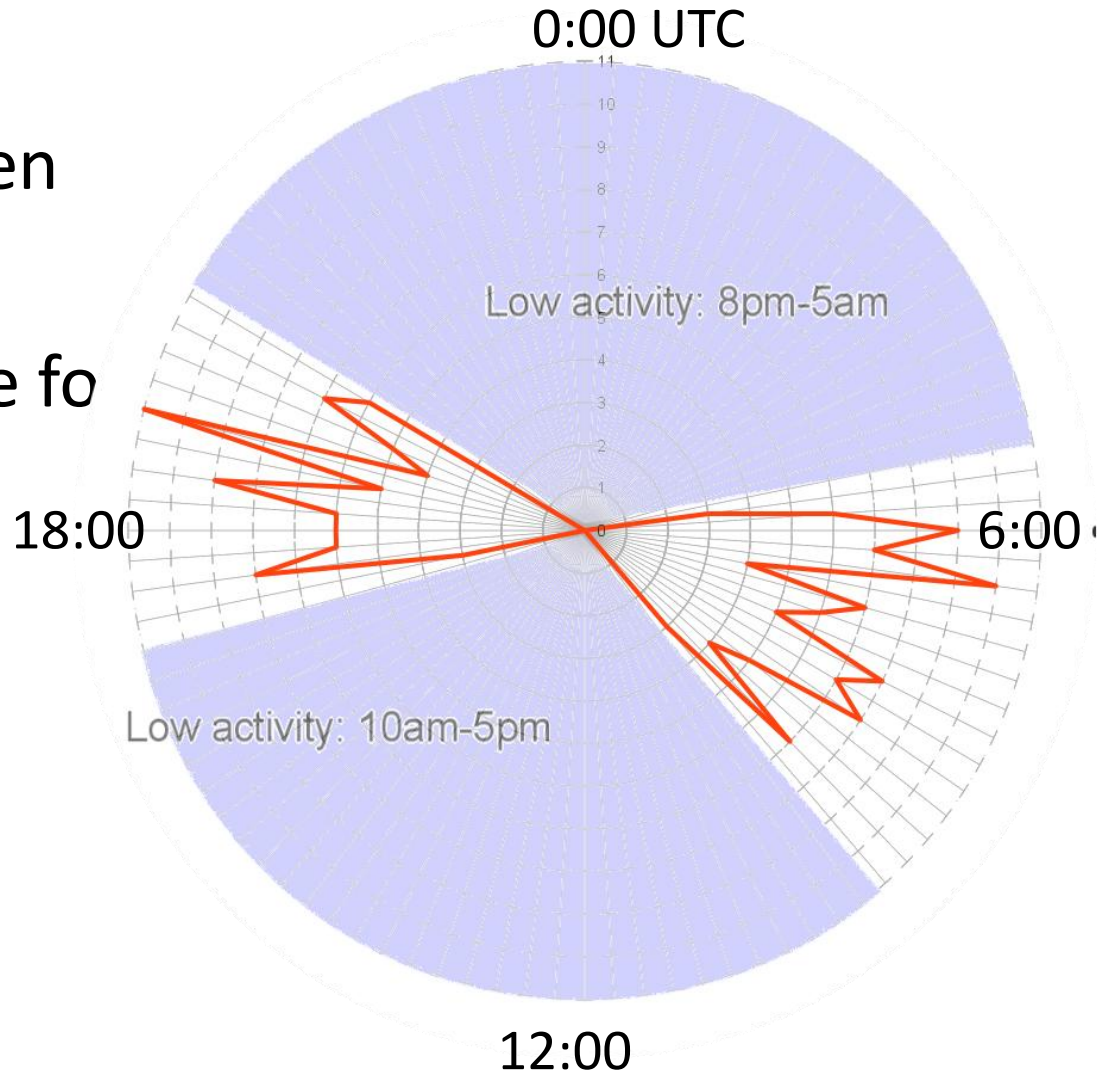
Daily cycle of data supply (& demand)

One morning's Direct Broadcast from NASA's Aqua satellite to a receiving station in Greenbelt, MD



Daily cycle of data supply (& demand)

- 2-3 overpasses, twice a day
- Scale down between overpasses
- Scale up just in time fo next overpass



Case study summary

- Proof of concept
  - Infrastructure as a Service made for a quick & easy transition
- Task effectiveness
  - Tried many different instance types @ low cost, no risk
  - Reduced latency from ~40 minutes to ~12
  - Significant add'l performance is clearly within reach
  - Gained significant bandwidth
- Cost effectiveness
  - Cloud computing costs are complex and significant

    (But so are data center costs)
  - Elastic provisioning and utility pricing worked well for us
    - Scaling up & down 2x/day will shrink our costs by 60+%