



Cloud Computing: Beyond the Hype

Greg Bell, IT Division
EETD Lunch-Time Seminar
December 2, 2009

Two Major Topics

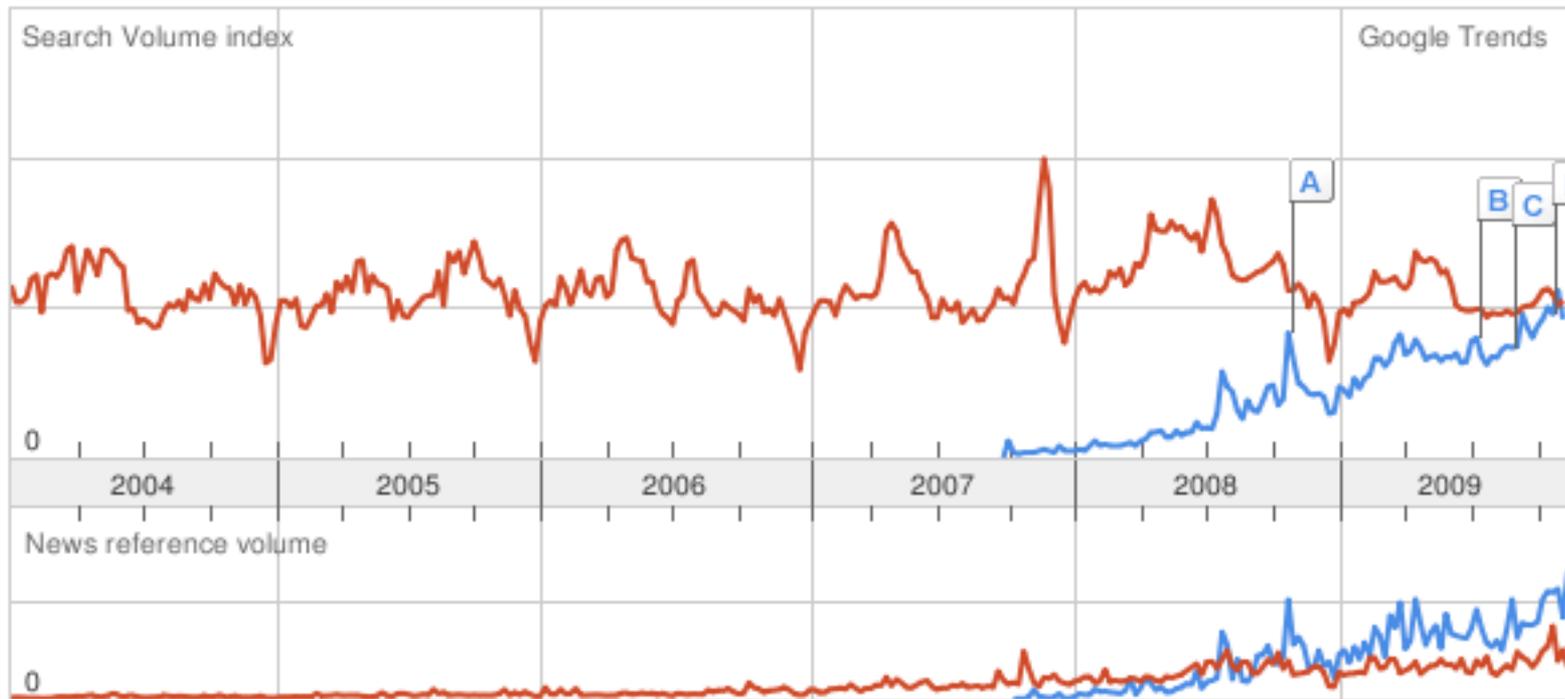
- what is cloud computing?
- how could it be relevant to EETD?
 - as consumers of Information Technology
 - as energy researchers

A Short Video and a Simple Definition

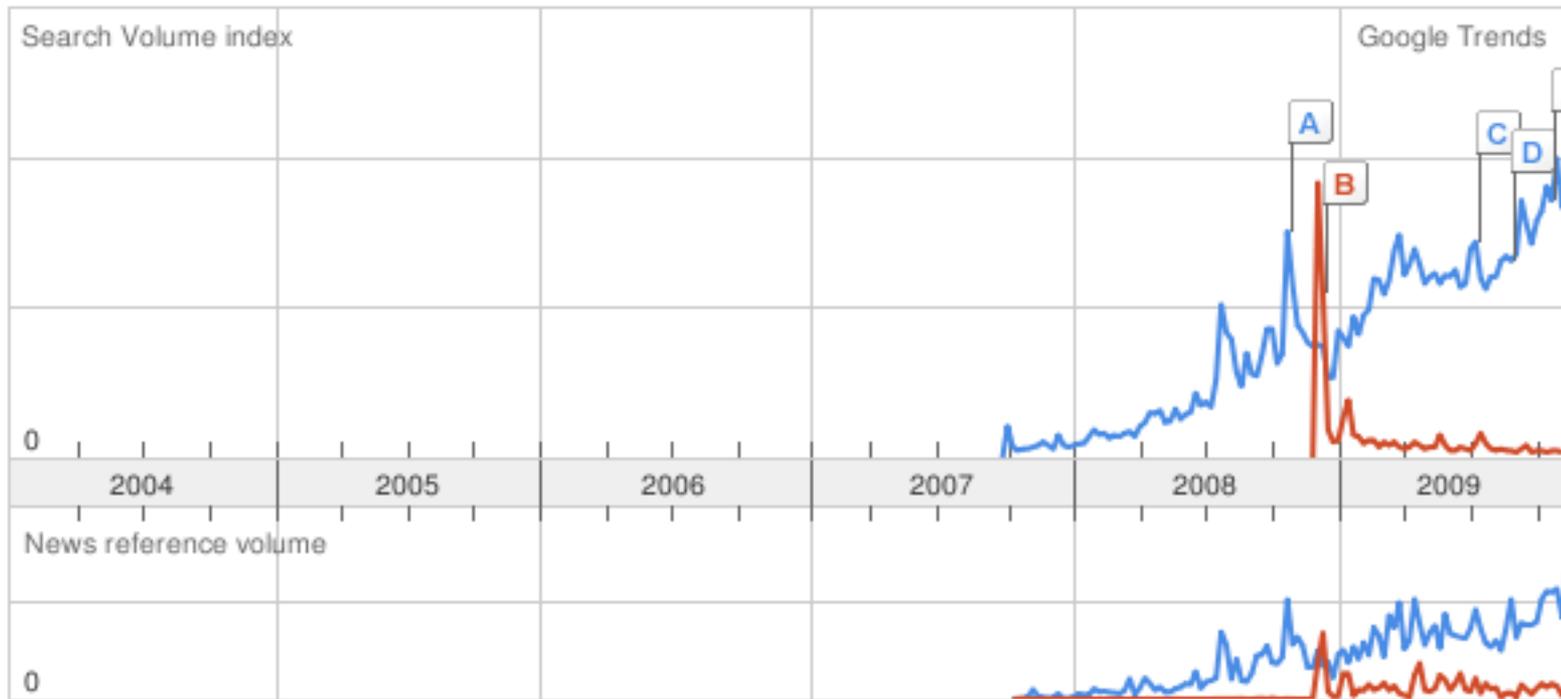
- GSA's cloud cartoon: <http://apps.gov>
- a useful definition from Urs Hoelzle, Senior VP for Operations at Google
 - 'cloud' implies that computing systems will eventually come in two sizes: tiny and huge
 - tiny: smart phones, net/notebooks, thin clients, tablets
 - huge: 'warehouse-scale computers': ~\$500M, ~50MW
 - your data & apps live here
 - SC09 Masterworks Talk: 'Warehouse Scale Computers'
- optimized for efficiency at either scale
 - battery constraints
 - OpEx @ 50MW

Using Cloud Services to Document Cloud Hype

● cloud computing ● solar power



● cloud computing ● steven chu



A Tale of Two Clouds



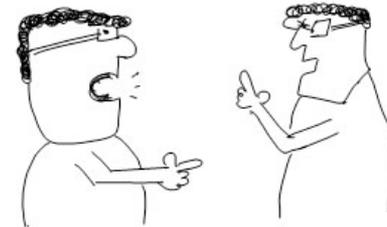
The Storm Cloud Will Bring...

- privacy disasters
- security breaches
- data and service lock-in
- complete dependence on the network
- lower quality of service (speed of c)
- lay-offs and corporate turmoil
- paradoxically higher prices



WHERE THE HECK
IS MY DATA?

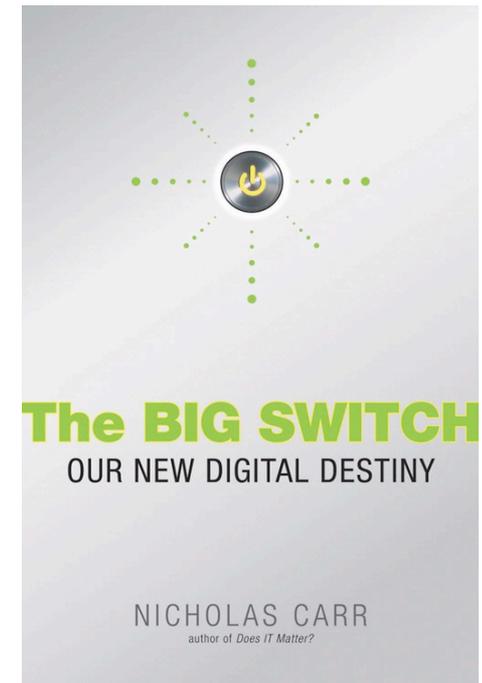
ITS THERE, UP
IN THE CLOUDS.



Brainstuck.com

The Fluffy Cloud Will Bring...

- much better IT services
- much faster innovation
- greater resiliency
- renewed focus on important problems
- more productive collaboration
- reduced travel
- an end to the PC as we know it
- lower prices over time
- profound & appropriate disruption



The Nicholas Carr Thesis

“What happened to the generation of [electrical] power a century ago is now happening to the processing of information. Private computer systems, built and operated by individual companies, are being supplanted by services provided over a common grid - the Internet - by centralized data-processing plants. Computing is turning into a utility...”

- Nicholas Carr, *The Big Switch*

More Complete Description of Cloud Computing

- IT sourcing model (not a technology)
 - exploits virtualization, fast networks, web-service standards and APIs, enormous data centers
- ‘5 essential characteristics’ per NIST
 - self-service provisioning
 - services delivered over network
 - multi-tenancy (resource sharing)
 - scales up/down rapidly (elasticity)
 - granular metering, micro-payments

Cloud Services Layers, From Bottom Up

- **Infrastructure as a Service [IaaS]**
 - customer rents low-level resources (cycles, storage)
 - sysadmin, integration, scaling tasks remain
 - the most flexible but labor-intensive option
 - quick demo (nagios on EC2)
- **Platform as a Service [PaaS]**
 - customer rents an application platform
 - customer provides code
 - no sysadmin tasks
- **Software as a Service [SaaS]**
 - customer rents an app
 - customer provides configuration + data



Are Customers 'Like Us' Using Cloud Services?

- cautious adoption in research, education, .gov
 - many institutions 'studying' cloud services (UCCTF)
 - early adopters include Genentech, city of LA, District of Columbia, ASU, UCD, USC (eg <http://getgoogle.usc.edu/>)
 - universities seem more willing to experiment with *student* email
- strong support from the Feds
 - Federal CIO Vivek Kundra promoting cloud very aggressively
 - remember <http://apps.gov>
- vendors refining services to entice .gov customers
 - Google 'federal cloud' + FISMA compliance in 2010
 - another vendor offering separate .gov hosting facility (*cloudish*)

What Does Cloud Computing Cost?

- hard to say: market is changing constantly
- two major pricing models at the moment
 - free
 - complicated
- feature parity (and therefore the possibility of simple apple-to-apple comparisons) is rare
 - to complicate matters, institutions don't always fully understand their local costs
- a few data points follow
 - but remember, cost is not the only dimension of service

Data Points: Amazon EC2 VM Service

United States	Europe		
Standard On-Demand Instances		Linux/UNIX Usage	Windows Usage
Small (Default)		\$0.085 per hour	\$0.12 per hour
Large		\$0.34 per hour	\$0.48 per hour
Extra Large		\$0.68 per hour	\$0.96 per hour
High-Memory On-Demand Instances		Linux/UNIX Usage	Windows Usage
Double Extra Large		\$1.20 per hour	\$1.44 per hour
Quadruple Extra Large		\$2.40 per hour	\$2.88 per hour
High-CPU On-Demand Instances		Linux/UNIX Usage	Windows Usage
Medium		\$0.17 per hour	\$0.29 per hour
Extra Large		\$0.68 per hour	\$1.16 per hour

[as of 11/24/09]

Data Points: Amazon S3 Storage Service

United States

Storage

- \$0.150 per GB – first 50 TB / month of storage used
- \$0.140 per GB – next 50 TB / month of storage used
- \$0.130 per GB – next 400 TB /month of storage used
- \$0.120 per GB – storage used / month over 500 TB

Data Transfer

- \$0.100 per GB – all data transfer in
-
- \$0.170 per GB – first 10 TB / month data transfer out
 - \$0.130 per GB – next 40 TB / month data transfer out
 - \$0.110 per GB – next 100 TB / month data transfer out
 - \$0.100 per GB – data transfer out / month over 150 TB

Requests

- \$0.01 per 1,000 PUT, COPY, POST, or LIST requests
- \$0.01 per 10,000 GET and all other requests*

* No charge for delete requests

[as of 11/24/09]

Data Points: Joyent Infrastructure++ Service

Accelerator Name	Guaranteed CPU Resources*	Included Monthly Bandwidth	RAM	Storage	Monthly Fee	Yearly Fee
1/4 GB	1/16 Core(s)	10 TB	256 MB	5 GB	N/A	\$199
1 GB	1/4	10 TB	1 GB	15 GB	\$125	\$1,250
2 GB	1/2	10 TB	2 GB	25 GB	\$250	\$2,500
4 GB	1	10 TB	4 GB	50 GB	\$500	\$5,000
8 GB	2	10 TB	8 GB	100 GB	\$1,000	\$10,000
16 GB	4	10 TB	16 GB	100 GB	\$2,000	\$20,000
32 GB	8	10 TB	32 GB	100 GB	\$4,000	\$40,000

Setup fees are equal to the one-month charge for the accelerator.

*All plans can burst up to 8 cores

[as of 11/24/09]

Data Points: Google AppEngine Platform Service

Resource	Free Default Quota		Billing Enabled Quota	
	Daily Limit	Maximum Rate	Daily Limit	Maximum Rate
Requests	1,300,000 requests	7,400 requests/minute	43,000,000 requests	30,000 requests/minute
Outgoing Bandwidth (billable , includes HTTPS)	1 gigabyte	56 megabytes/minute	1 gigabyte free; 1,046 gigabytes maximum	740 megabytes/minute
Incoming Bandwidth (billable , includes HTTPS)	1 gigabyte	56 megabytes/minute	1 gigabyte free; 1,046 gigabytes maximum	740 megabytes/minute
CPU Time (billable)	6.5 CPU-hours	15 CPU-minutes/minute	6.5 CPU-hours free; 1,729 CPU-hours maximum	72 CPU-minutes/minute
Outgrowing the maximums? Request an increase.				

note: billing model actually far more complex than this summary table implies

[as of 11/24/09]

Data Points: Microsoft Azure Platform Service

Windows Azure

- Compute = \$0.12 / hour
- Storage = \$0.15 / GB stored / month
- Storage transactions = \$0.01 / 10K
- Data transfers = \$0.10 in / \$0.15 out / GB - (\$0.30 in / \$0.45 out / GB in Asia)

Microsoft SQL Azure

- Web Edition: Up to 1 GB relational database = \$9.99 / month
- Business Edition: Up to 10 GB relational database = \$99.99 / month
- Data transfers = \$0.10 in / \$0.15 out / GB - (\$0.30 in / \$0.45 out / GB in Asia)

[as of 11/24/09]

Data Points: Google Apps Email, Collab Service

Upgrading to the Premier Edition



The Premier Edition is an enterprise package of workplace applications. You can try Google Apps Premier Edition for free for 30 days. The 30-day free trial period starts from the time that you sign up or upgrade to the Premier Edition. It's not currently possible to select the Premier Edition for a subset of users at your domain, while keeping others on the Standard Edition.

Premier Edition costs **\$50** per user account per year (example: If you request 50 user accounts for your domain, the cost is \$2500 annually)

[as of 11/24/09]

Data Points: Service-now.com SaaS Ticketing

Service-now.com - IT Service Management Suite

Welcome: Joe Employee Logout

Employee Self Service Refresh: Off

Self-Service

- Homepage
- Knowledge
- Change Password
- Help the Help Desk
- Incidents
- My Requests
- Requested Items
- My Profile
- Take Survey
- Akula_New_Application
- Akula_New_Application
- Leave Table
- library table
- My New Applicatio
- Name Change
- Requesting Organisation
- Test Data

News

View all items

Applications

Can't access SAP	2009-11-19
Application use from Home with VPN	2009-04-16
Starmosh login	2009-04-09

View all items

My Open Incidents

Number	Category	Short description
INC0010138	Inquiry / Help	testing
INC0010139	Inquiry / Help	
INC0010140	Inquiry / Help	
INC0010141	Inquiry / Help	
INC0010175	Inquiry / Help	testing
INC0010176	Inquiry / Help	

Can We Help You? Try this

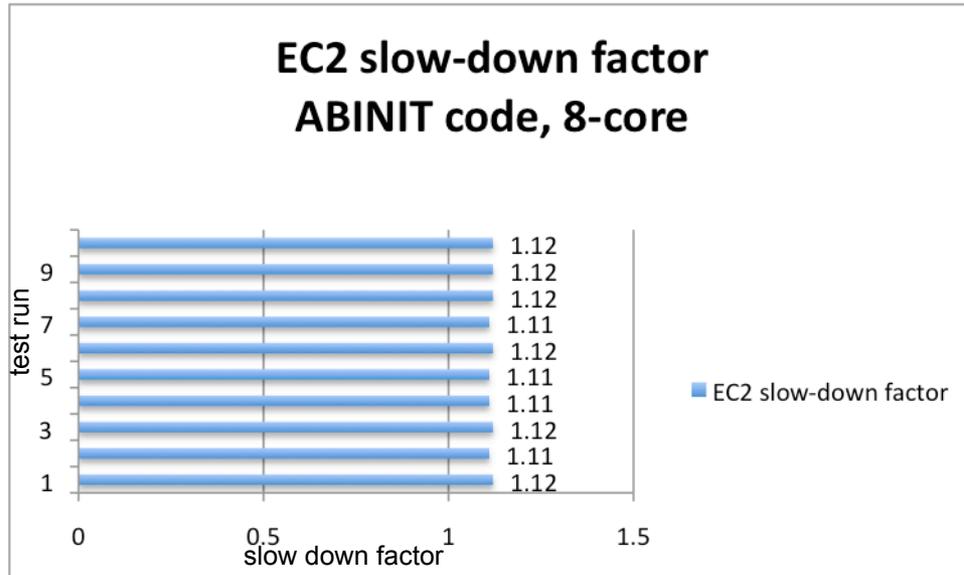
- Bond Trade Access Request
- Create a New Facilities Request
- Create a New Incident
- Request to Reset Router
- Reset a Password
- Schedule a Move
- Report a Performance Problem
- Report an Outage

[as of 11/24/09]

Even if You Understand the Pricing...

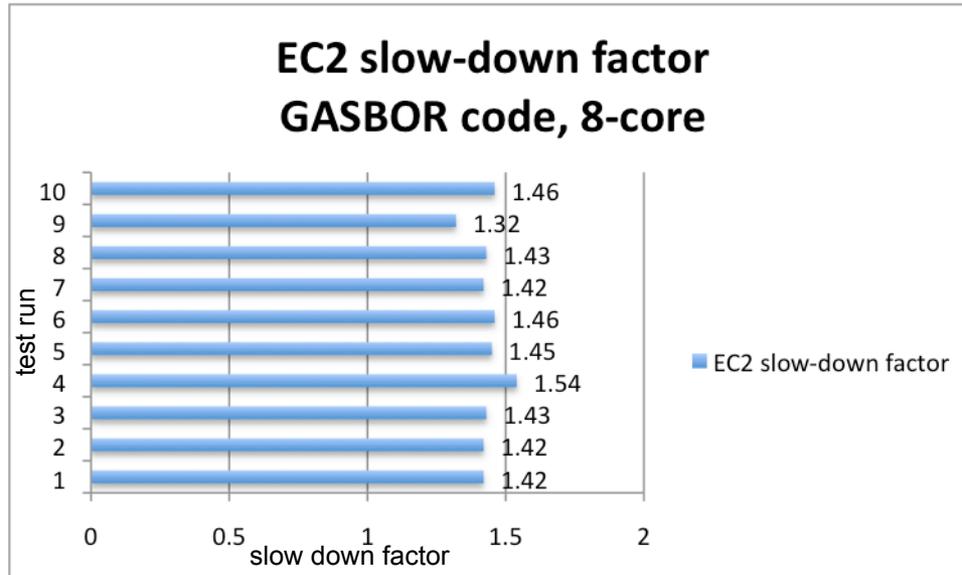
- do you understand the performance of your applications in the cloud?
- this Spring, the IT Division built virtual clusters in Amazon EC2, then ran a sampling of scientific codes
- some results were predictable, others surprising

ABINIT Code in EC2: Best-Case Performance



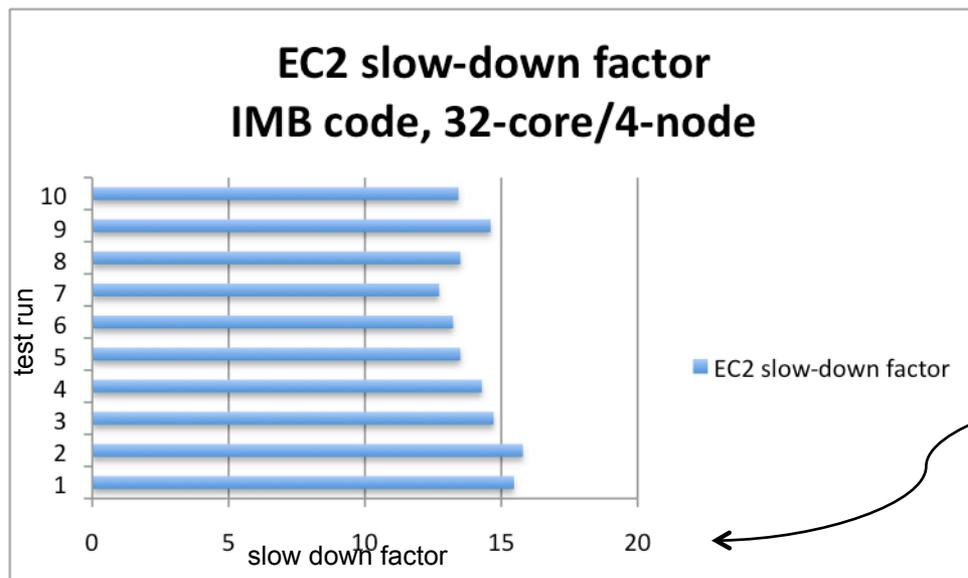
- using all cores on 8-core node
- graphing 'slow-down' vs local cluster performance
 - varies significantly by app
- this code is not latency sensitive; 2-hour run

GASBOR Code in EC2: a Bit More Expensive



- using all cores on 8-core node
- code is not latency sensitive; 10-hour runs

IMB Code in EC2: Definitely Cost-Prohibitive



- using 32 cores on 4 nodes
- code is latency-sensitive
- also ran VASP code (modeling single-molecule junctions in nanostructures) with similar slow down

Tentative Conclusions About Pricing

- it's hard to generalize about the cost of cloud services
- we need a new cost model for each cloud-service use case (IaaS and PaaS)
 - even then, it's necessary to prototype and verify
- for SaaS, vendors have widely-differing models

Does Cloud Computing Save Energy?

- favorable factors
 - increased server utilization (due to virtualization)
 - larger, newer, denser data centers
 - smarter integration of building systems & IT
 - obvious financial incentives for cloud providers to conserve
 - resilient architectures could be well-suited for demand-response or 'follow-the-sun' applications
- potentially unfavorable factors
 - increased network utilization
 - increased data replication
 - availability of new services may stimulate demand (eg: hulu)

A Few Scenarios and Conjectures

- HPC moves from legacy data center to cloud
 - almost certainly a big win
 - no virtualization benefit, but cloud data center more efficient
- email infrastructure moves from legacy data center
 - probably a win, but there are many variables
 - network utilization, resiliency, data replication
- data-intensive application moves to cloud
 - might be a close call
 - benefit depends on workflow: does increased network utilization offset other benefits?
- many variables + variance in each scenario
- thanks to Bruce Nordman, Rod Tucker for insights

Cloud Projects Within the IT Division

- [launch Google Docs for current version of project list]
- by the way, Google Docs is a useful collaboration tool you can access today as LBNL staff
 - not feature-rich
 - but good for simple tasks
 - terrific for collaborative authorship



Cloud Projects Relevant to EETD

- collaborative and desktop services (potentially helpful to entire Division)
 - GMail
 - GDocs / Sites
 - Dropbox
 - desktop backup/restore (task force soon completing)
 - more to come: iTunesU, GCal?

Cloud Projects Relevant to EETD

- scientific computing (potentially interesting to those researching data centers, networks, servers)
 - benchmarking commercial services
 - Magellan project
 - \$32M ARRA-funded collaboration (LBNL / ANL / ESnet) to explore feasibility of science cloud, with 100GE network
 - exploring performance, service models, architectures, costs
 - science pilots underway (all involving CRD)
 - BLAST on Amazon Hadoop
 - climate analysis on MS Azure
 - Supernova Factory on Amazon EC2

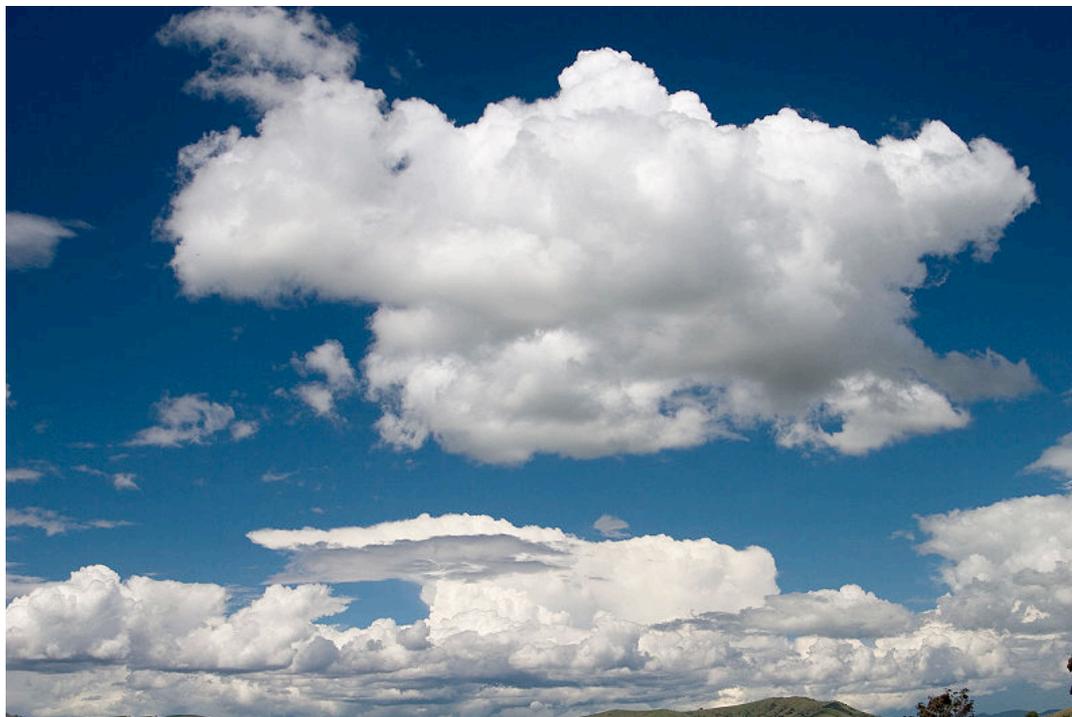
Cloud Projects Relevant to EETD

- application hosting (potentially useful to application owners)
- HES
 - empowers home owners to reduce their energy use
 - customer-facing app + licensed web service
 - hardware currently in Bldg 50A
 - EETD & IT will explore possibility of cloud-sourcing
 - which layer? PaaS, IaaS
- LEEP
 - HES for wet labs
 - implemented at nearly the same time on Google AppEngine, Appshosting, and conventionally (*terrestrially*) in J2EE
 - quick demo

To Summarize

- beware of generalizations about cloud computing
 - it's a barely coherent category
 - but having said that....
- I believe many IT functions now performed locally will be purchased as cloud services in 3-5 years
- this is good for science & good for the planet, but there will be trade-offs
- lower cost is not the only benefit
- general posture of IT Division
 - proceed optimistically but pragmatically, seeking data
 - understand range of services, policy issues
 - gain competence in cloud-sourcing, conduct pilots
 - use cloud services where appropriate

Thank you. Questions?



Greg Bell, grbell@lbl.gov