ESA's Cloudscape: A review of projects using cloud technology in ESA

William O'Mullane

Gaia science operations development manager

Based on:
Final presentation of Study on Cloud Computing
ESRIN/Contract Nr. 22700/09/I-SB
Study manager: Jose Balseiro
Presented by: Jason Brazile and Ronnie Brunner

Why Cloud?
Govt. "Cloud first"

Example: US

- All CIO's **must define ≥ 3 projects by Q2 2011**
- By Q4, 1 **must be in operation**
- By June 2012, **all 3 must be**
- "Security concerns not enough"

FEDERAL TIMES.com

SEARCH

PAY & BENEFITS PERSONNEL AGENCY NEWS TRAVEL IT ACQUISITION FACILITIES, FLEET & ENERGY CONGRES

Start moving to cloud now: Launch 3 projects by June 2012

By NICOLE BLAKE JOHNSON | Last Updated: December 13, 2010

Comments (1) Recommend Like 6 Tweet 8 SHARE

Security concerns won't be enough to stop cloud computing.

Chief information officers have three months to identify a minimum of three systems they deem suitable for cloud operations and to create a strategy for moving them. A year from now, at least one of the three systems must be operating in a cloud environment, and by June 2012, all three must be.

s, OMB will require that agencies default to cloud-based
d option exists," according to the Obama administration
he CIO Council website.

maintaining data on costly, energy-inefficient agency-
e IT reform plan. Other steps call for aligning the budget
r program managers, increasing engagement with
ntability.

vek Kundra, federal chief information officer, told CIOs

tration and Agriculture Department have announced

g 21 different messaging and collaboration systems into
he cloud is expected to save the department $6 million a

nail accounts to Google Apps for Government. GSA CIO
mail systems on GSA-owned servers will lower costs by
reductions in staffing, licensing and program costs.

The Washington Post

Agencies to look for a 'cloud option'

By Marjorie Censer
Monday, November 22, 2010

The federal government is adopting a "cloud-first" policy, marking the administration's strongest statement yet in support of Web-based computing as i looks to overhaul the way it buys information technology.

Jeffrey Zients, the federal government's first chief performance officer, announced last week that the Office of Management and Budget will now require federal agencies to default to cloud-based solutions "whenever a secure, reliable, cost-effective cloud option exists."

THE WALL STREET JOURNAL. Digital Network | WSJ.com | MarketWatch | BARRON'S | All Things Digital | FINS | SmartMoney

MarketWatch

Latest News View All

8:32a BREAKING Treasury to sell warrants to buy Citigroup shares

8:31a CVS Caremark names Larry Merlo CEO

Front Page | News Viewer | Commentary | Markets | Investing | Personal Finance | Community | Games

Nov. 1, 2010, 7:02 p.m. EDT

Google sues U.S. over Microsoft preference

By John Letzing, MarketWatch

SAN FRANCISCO (MarketWatch) — Google Inc. has filed a lawsuit alleging that the U.S. Interior Department shut it out of bidding for a large contract, due to a preference for Microsoft Corp.'s technology, in a move that underlines the search giant's growing competition with Microsoft in the market for Internet-based software tools.

In a complaint filed last week, Google (GOOG 613.15, +1.32, +0.22%) and its government-reselling partner Onix Networking Corp. allege that the Interior Department issued a request for bids on an email contract, specifying that only Microsoft's (MSFT 28.14, +0.12, +0.43%) technology could be proposed — despite prior indications that Google would be able to compete for the deal.

Google argues in the complaint, filed in U.S. Court of Federal Claims, that the bid request violates the Competition in Contracting Act.

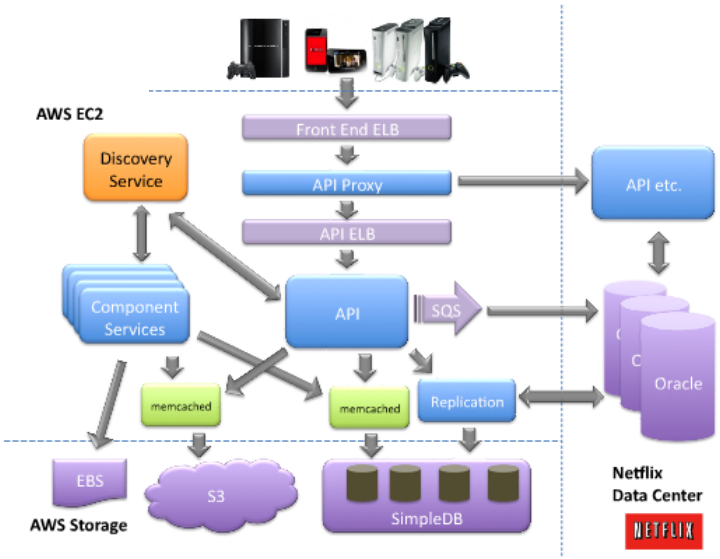
Slic

gaia European Space Agency

Why Cloud? Netflix → Amazon

Example:

- Gains: Agility, Reduced Cost
- Thousands of EC2 nodes
- Petabytes of S3
- Hadoop clusters
- Akamai/Limelight(CDN) use



Adrian Cockcroft, Netflix in the Cloud, Nov 2010

Netflix Turns From Oracle, IBM to Amazon to Save Cash

By Chris Kanaracus, IDG News Nov 24, 2010 8:00 pm

Netflix moved some of its most crucial IT operations over to Amazon Web Services' Elastic Compute Cloud in order to save money and gain flexibility compared to using more Oracle software and IBM iron.

"Our datacenter runs Oracle on IBM hardware, we could have switched to commodity hardware in a data center, but skipped that step by going to AWS," Netflix cloud architect Adrian Cockcroft told the consulting firm Cloudscaling in an interview posted Tuesday. "There are three points on cost, one is that Oracle on IBM is very expensive, so AWS looks cheap in comparison, and we have flat-lined our datacenter capacity."

Netflix Deployed on AWS

amazon web services

Content	Logs	Play	WWW	API
Video Masters	S3	DRM	Search	Metadata
EC2	EMR Hadoop	CDN routing	Movie Choosing	Device Config
S3	Hive	Bookmarks	Ratings	TV Movie Choosing
CDN	Business Intelligence	Logging	Similars	Mobile iPhone

Akamai Limelight NETWORKS

NETFLIX

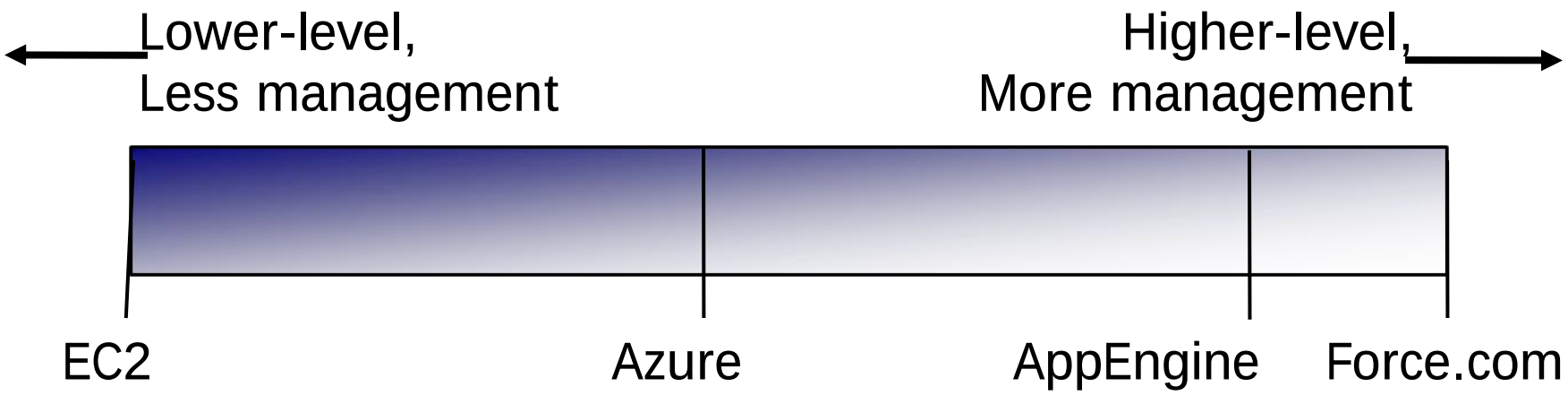
gaia European Space Agency

What a cloud is for me personally …

- Cloud Computing is:
  - Self-service
  - On-demand
  - Pay-as-you-go
- Not much different to a grid BUT..
  - No ‘gridware‘ – I can just have the machine
    - Hence no messing with security in my application
  - I can have ANY machine (within reason)
    - i.e. linux, windows, other obscure machine …
  - I pay per hour (cents per machine)
- Wikipedia says
  - Internet-based computing, whereby shared resources, software and information are provided to networked computers and other devices on-demand.

Most people agree on this ..

- Broadly Clouds come in 3 forms (services).
  - **Platform As A Service** (Google , also Ms Azure) develop against given API
  - **Infrastructure As A Service** (Amazon) just give me the machines I will do the rest …
  - **Software As A Service** (like Microsoft offering office, Salesforce.com) just use it
- Last most interesting for me/Gaia..



The Cloud Computing Stack

Cloud Enablers / Cross platform solutions

enStratus
RIGHTSCALE
cloudkick
SCALR
rPath
newScale
zimory
libcloud
nimsoft
From the Data Center to the Cloud

Software as a Service (SaaS)

Microsoft Office 2010 WEB APPLICATIONS
salesforce
slideshare Present Yourself
Basecamp Project Collaboration
Google
ZOHO Work. Online
NETSUITE ONE SYSTEM. NO LIMITS.
Taleo Talent Drives Performance
webex
GoToMeeting

Platforms as a Service (PaaS)

force.com platform as a service
hadoop
intuit
appscale open source Google App Engine
Engine Yard
Google App Engine
Windows Azure
heroku

Infrastructure as a Service (IaaS)

amazon web services™
rackspace
GOGRID
linode.com
bluelock
terremark
CloudSigma
Eucalyptus

gaia European Space Agency

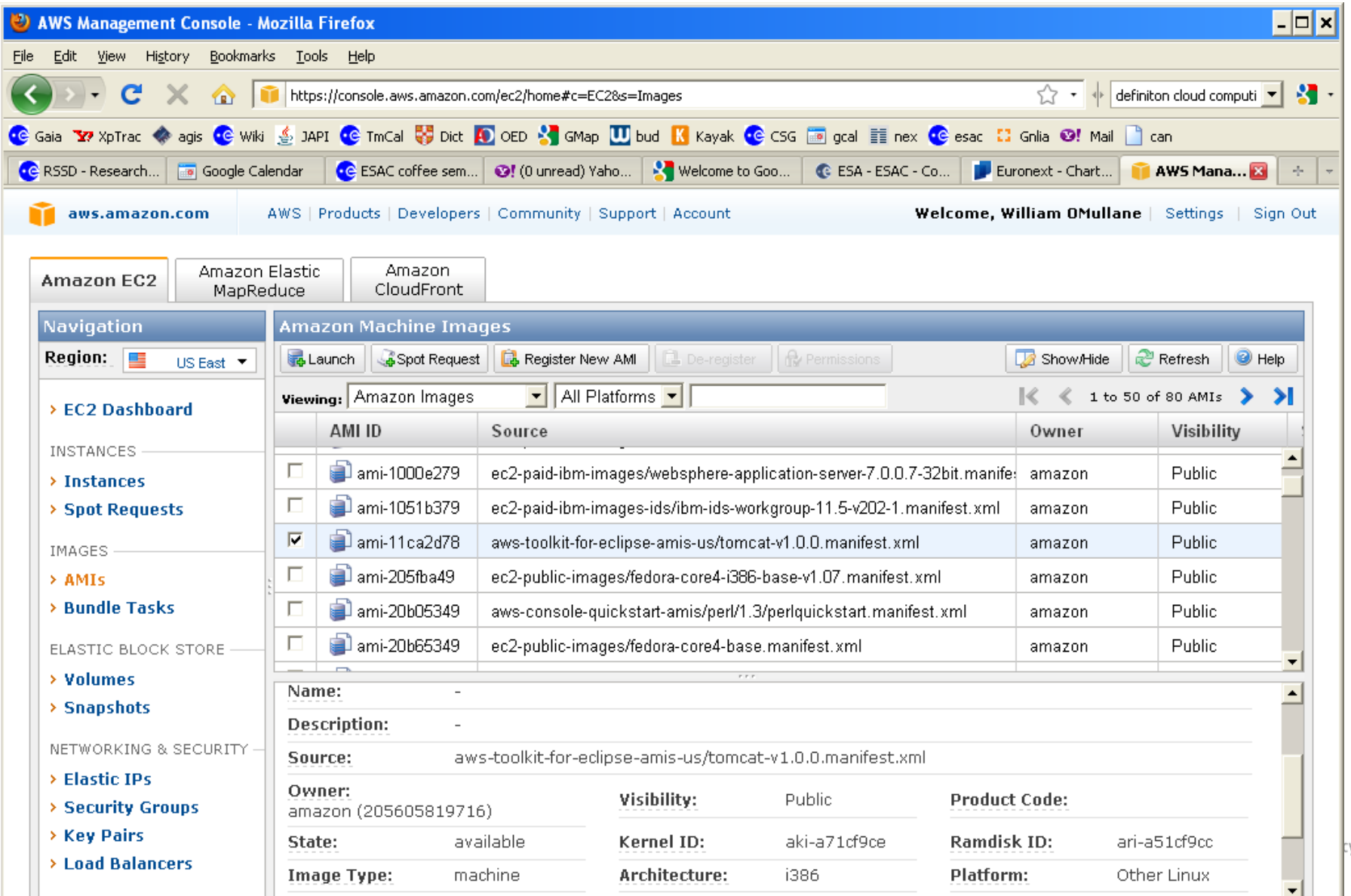
Getting a machine …

- How long does it take you to procure a machine ?

- It takes me at least six months !

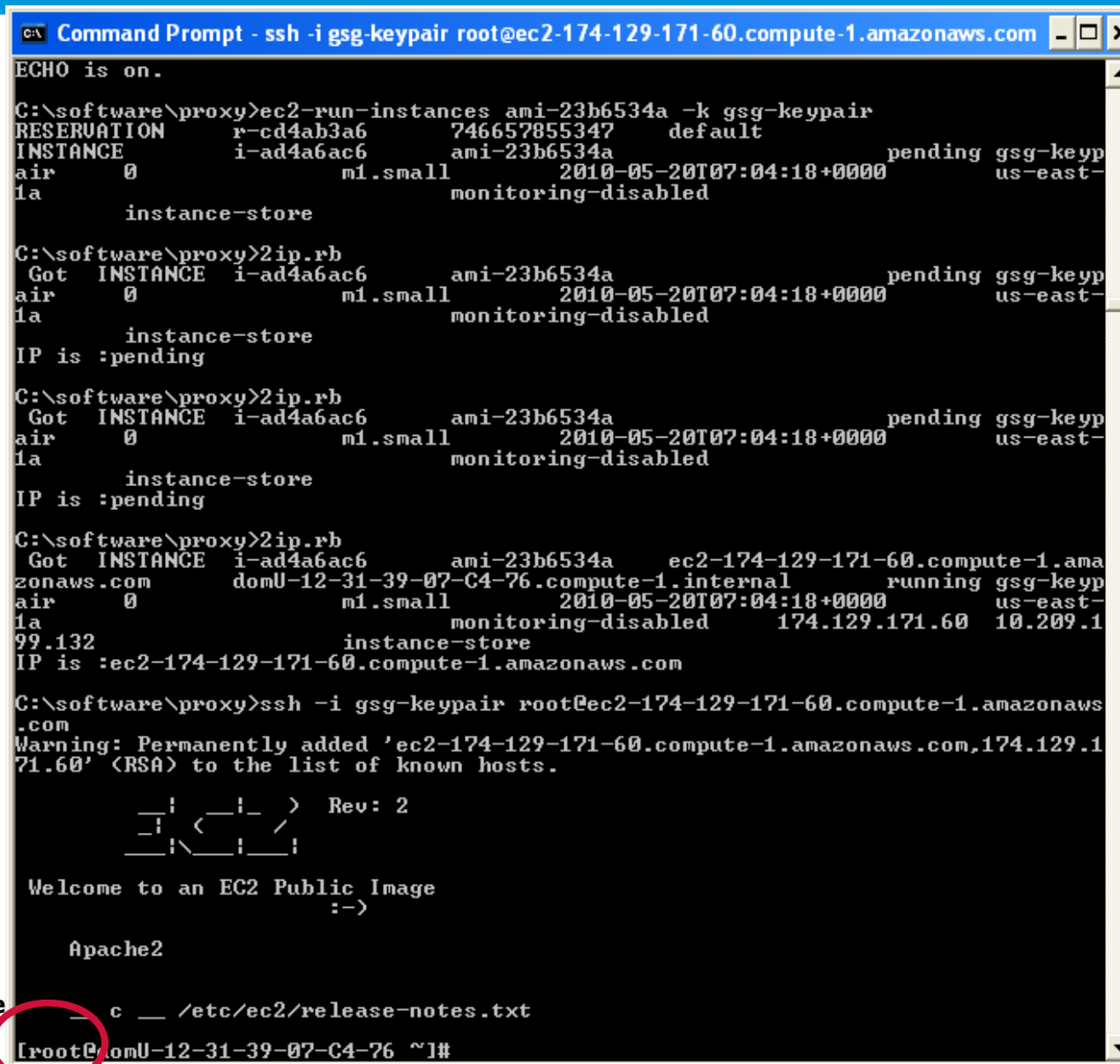
A machine in a minute

While on Amazon I can have one in minutes ..

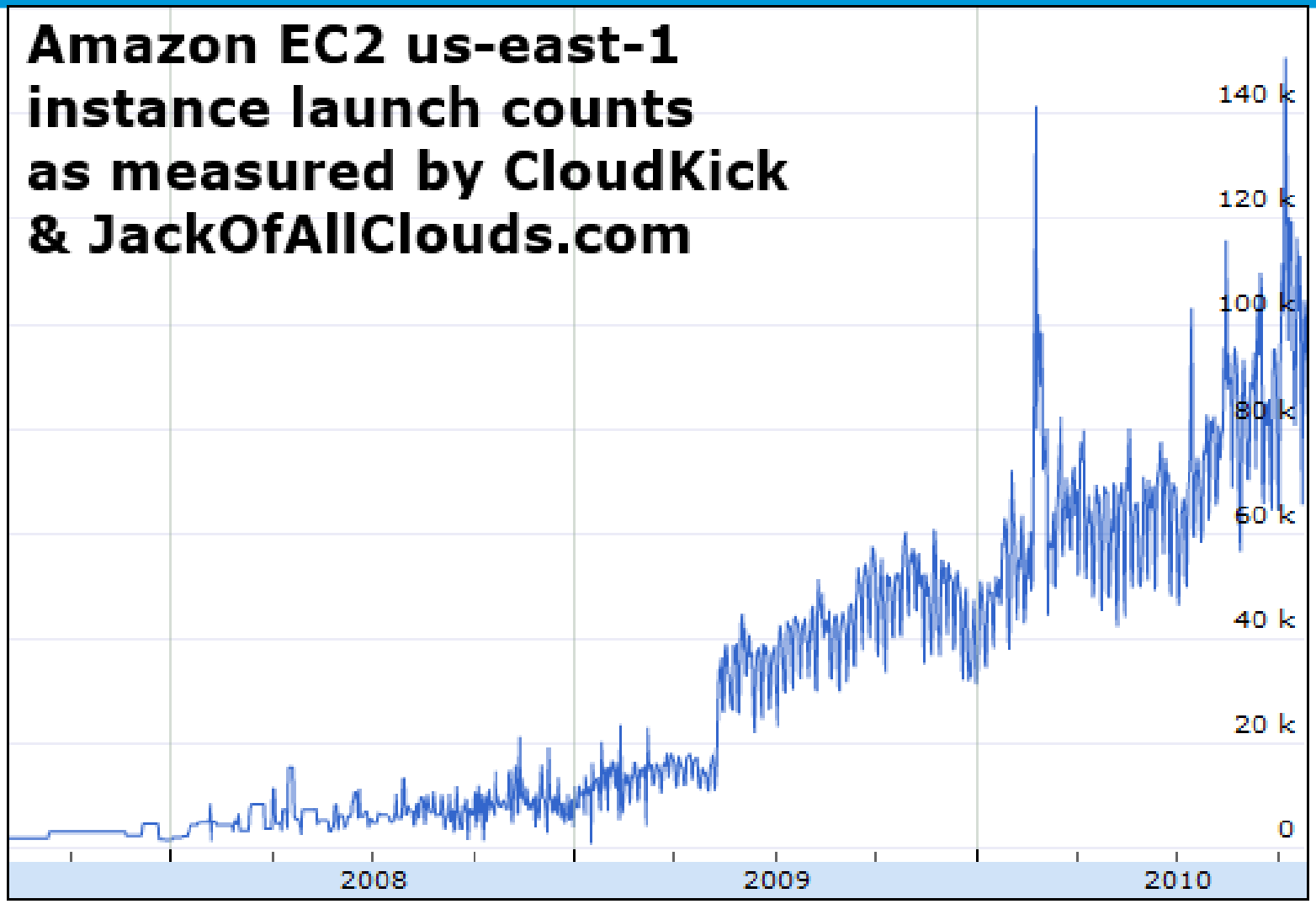


Command line too

With ROOT access!



Usage


Amazon EC2 us-east-1
instance launch counts
as measured by CloudKick
& JackOfAllClouds.com
140 k
120 k
100 k
80 k
60 k
40 k
20 k
0
2008
2009
2010

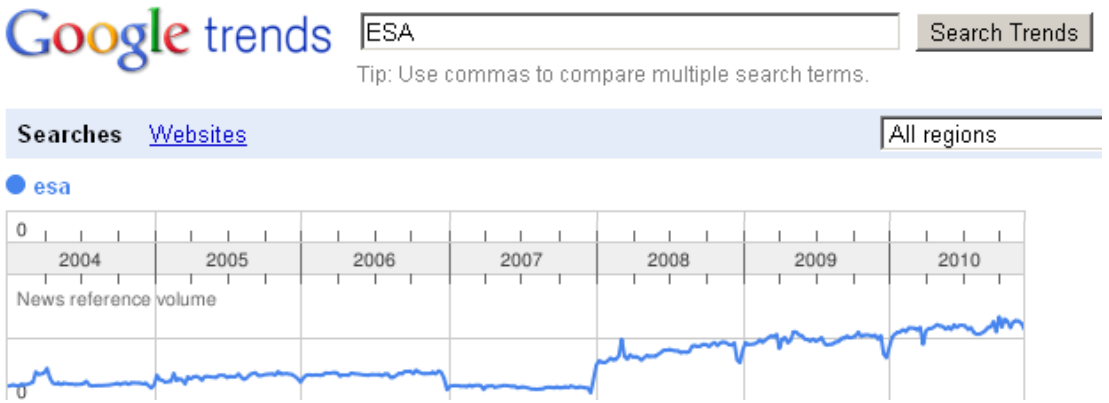
ESA Cloud Computing stories

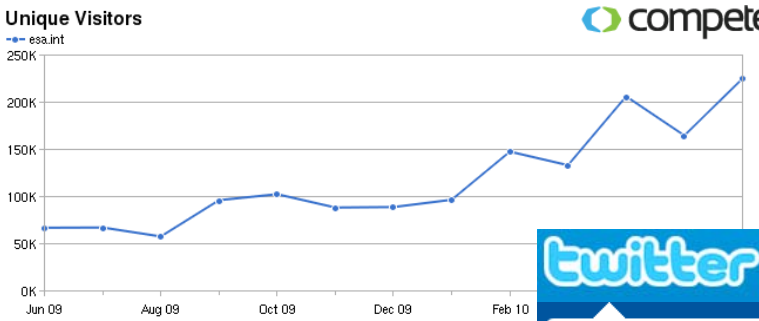
- There are already plenty of success stories some started in 2001 – all still consider using some mix of private and public clouds:
  - Corp. Comm: Portal Edge Caching, Media Distribution
  - GAIA mission: AGIS “Data Train”
  - G-POD Framework: Cloud prototype
  - Collaboration Tools
  - Supersites Geohazard Virtual Archive
  - SOA4GDS Software Development Environment, and others

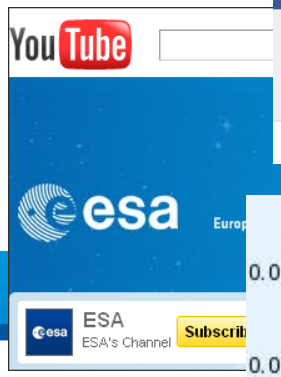
LEX-CCW's Portal Edge Caching, Media Distribution

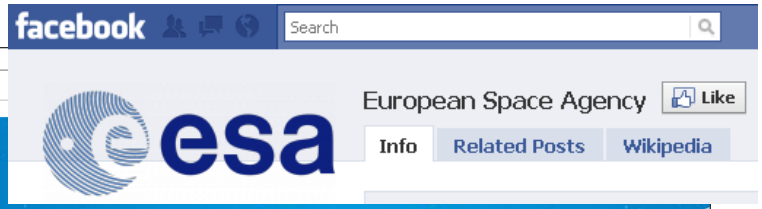
Since 2001…

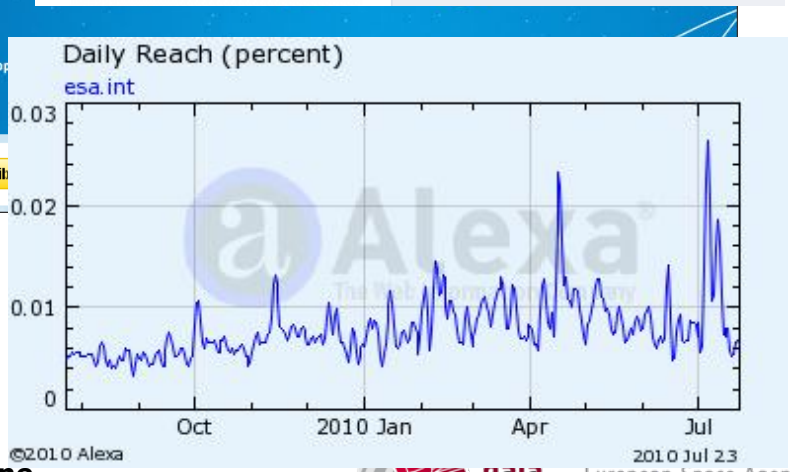
- Edge caching (Akamai, Highwinds)
- Image/Video dist.
- Content Mgmt







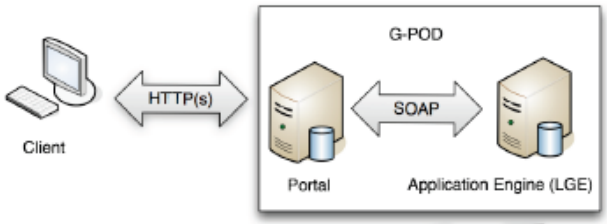


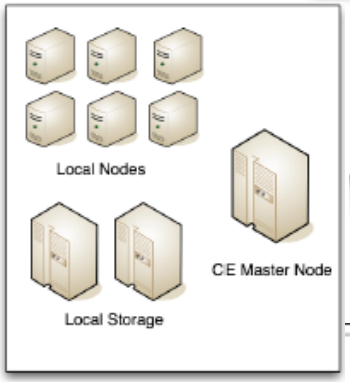


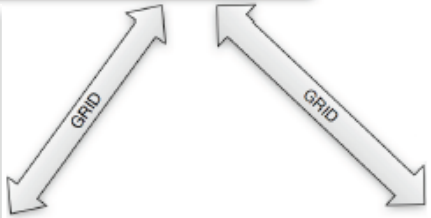
gaia European Space Agency

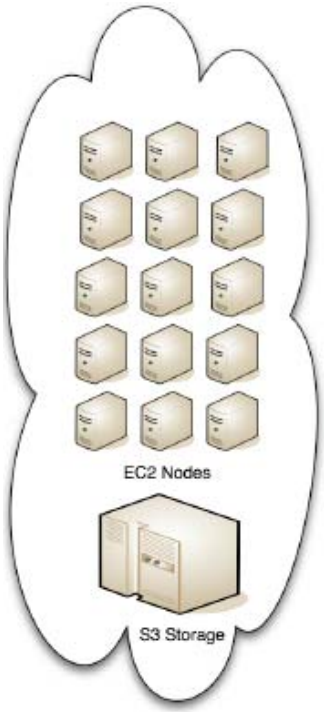
EO's G-POD Framework

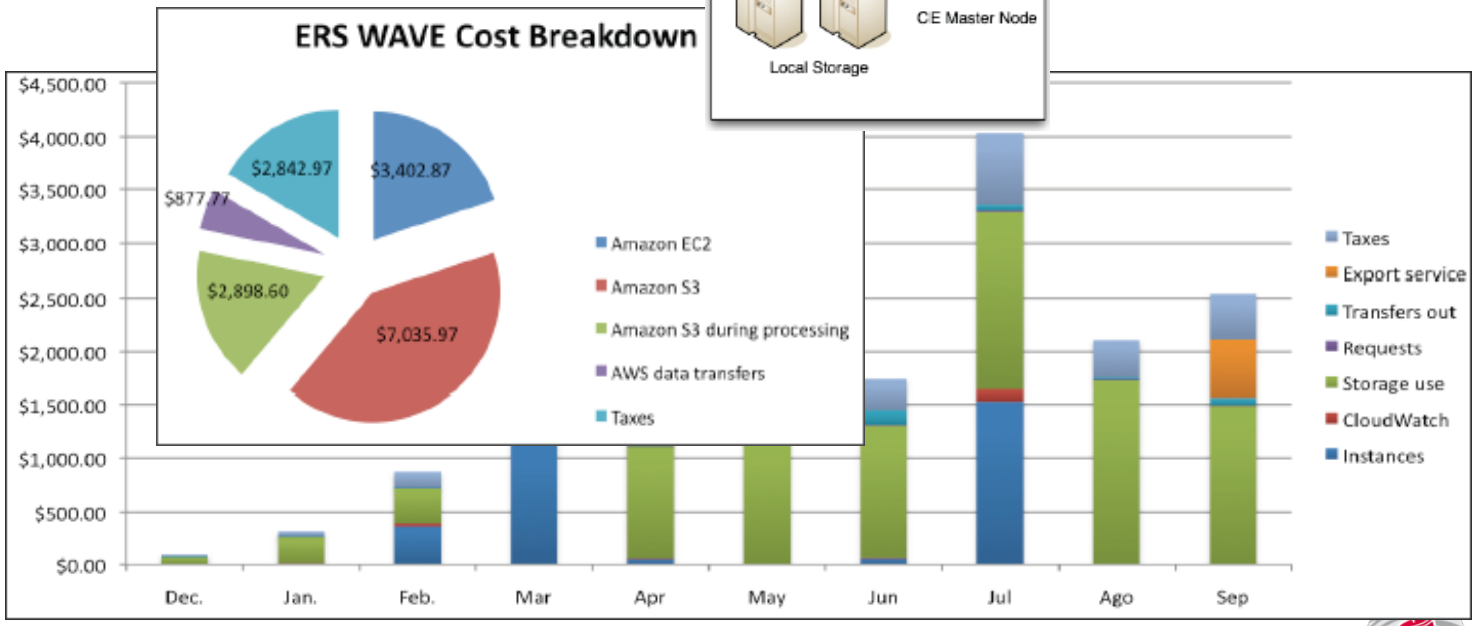
- Since 2009 (prototype)…
  - Amazon EC2 / S3
  - Grid and Cloud
  - Export service











Brito: A 10K reprocessing campaign for ERS Wave, Nov 2010

Corporate IT's Collaboration Tools

- Since 2009 (prototype)
  - Virtual Meetings/Desktop and Application Sharing
  - Recordings for meeting absentees

  Benefits
  - Improved productivity of remote workers
  - Expanded collaboration also with external partners
  - Reduced travel costs

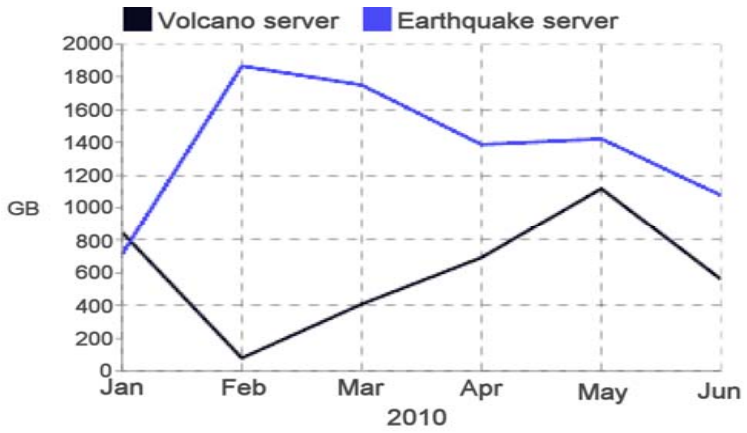
The yearly cost of WebEx is offset if 500 staff use WebEx instead of traveling once a year.

EO and UNAVCO's Supersites Geohazard Virtual Archive

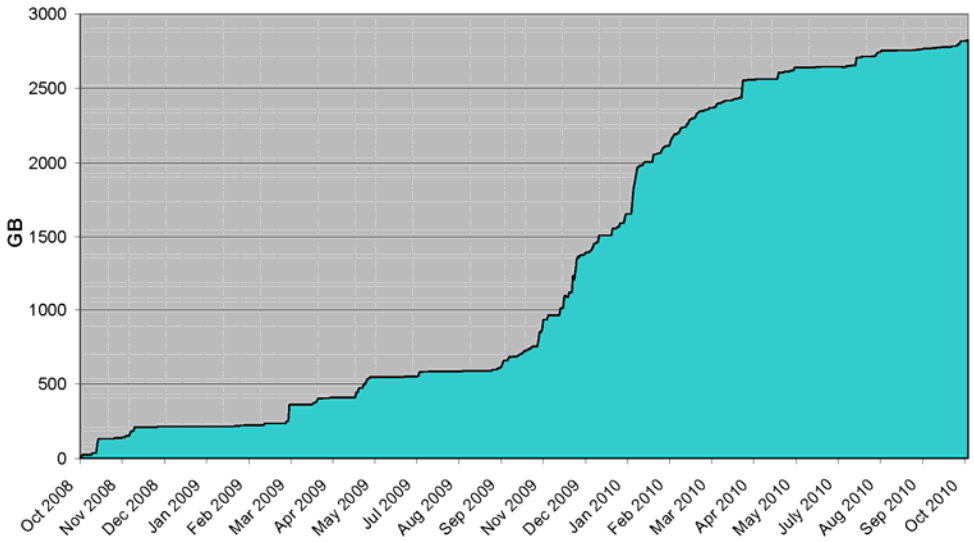
- Since 2008 (prototype)…
  - CDN large file distribution
  - Collaboration with ≥ 20 organizations to pool disaster observation data

GROUP ON EARTH OBSERVATIONS

Network bandwidth traffic



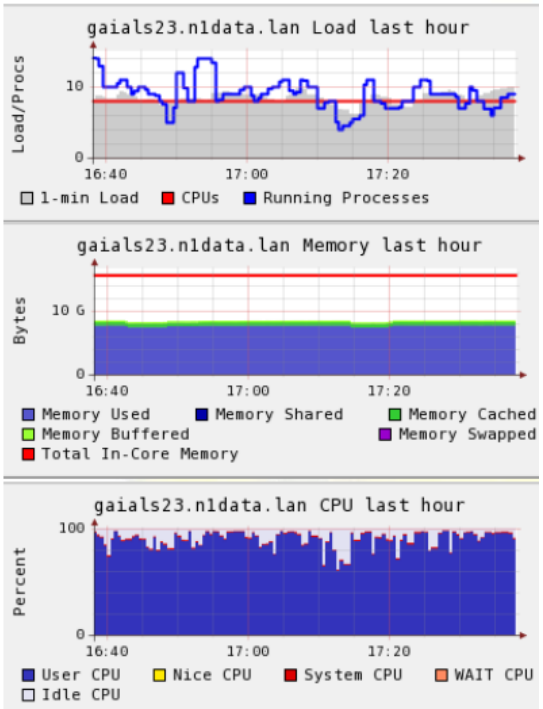
Monthly Storage Growth

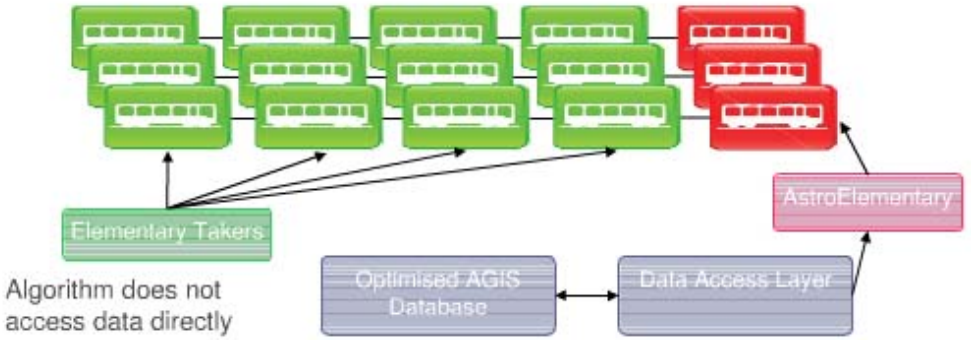


European Space Agency

ESAC's GAIA/AGIS "Data Train"

- Since 2009 (prototype)…
  - Amazon EC2 / S3
  - Oracle as a service





O'Mullane, GAIA Data Processing and Challenges, June 2010

We consider this successful compared to SDSS experience

But < 1TB data
No Users !!

And not all rosy this year.

O'Mullane, GAIA Data Processing and Challenges, June 2010

Parsons et al., Cloud Science or Astrometric Data Processing in Amazon EC2 May 2009

Lessons so far

	IaaS (computation)	CDN	PaaS	SaaS
Benefits	▪ Easier migration than expected ▪ Computation costs lower than expected ▪ Helped find scalability issues	▪ Agility / reach ▪ Much better Latency/Bandwidth ▪ Reduced network transit costs	No real experience yet	▪ Twitter ▪ Facebook ▪ Flickr ▪ YouTube ▪ Webex ▪ SharePoint
Caveats	▪ Storage costs at times higher than expected ▪ High volume data transfers slow/costly ▪ Inconsistent network performance ▪ Manual architecting needed	▪ Most not really pay-as-you-go, self-service, on-demand ▪ Most complex product / pricing structure		▪ Often needs "digital natives" involved in design (especially for social media) ▪ Learning curve varies greatly
Notes	▪ Mature yet still innovating Standardization "ad hoc"	▪ Mature ▪ New offerings coming		▪ "Just the beginning" ▪ Provider change quite difficult ▪ Mostly hard to generalize

Risks and their Consequences

Risk	Examples	Result
Re-invention of wheel	Portal proliferation; User account mess	Poor services, inefficiency
Individual “contracts” via credit card	Critical service is down because key person‘s individual credit card expires	Service failure, data mess (where’s what?)
Single actor can chose wrong direction quickly	Introduction of a proprietary SaaS solution that (only) provides a quick fix	Unmanaged service portfolio, not reaching strategic goals
Costs can‘t be tracked well	Monthly bills unpredictable due to irregular demand. Lots of hard to track small transactions with many providers	Financial exposure and uncertainty
Costs slowly increase	Nobody cleans up hard disks or gets rid of unused virtual machines	More expensive over time, unclear what‘s still needed
Data gets leaked	Data protection violation, leak of industry partner’s (or member state‘s) secrets	Financial liability, loss of trust
Data loss	NASA‘s moon landing tapes, hacker data vandalism, Provider default	Image/brand damage

EIROforum – Science Cloud

- EIROforum is a collaboration between eight European intergovernmental scientific research organisations that are responsible for infrastructures and laboratories: CERN, EFDA-JET, EMBL, ESA, ESO, ESRF, European XFEL and ILL.

Ambitious goals of science cloud

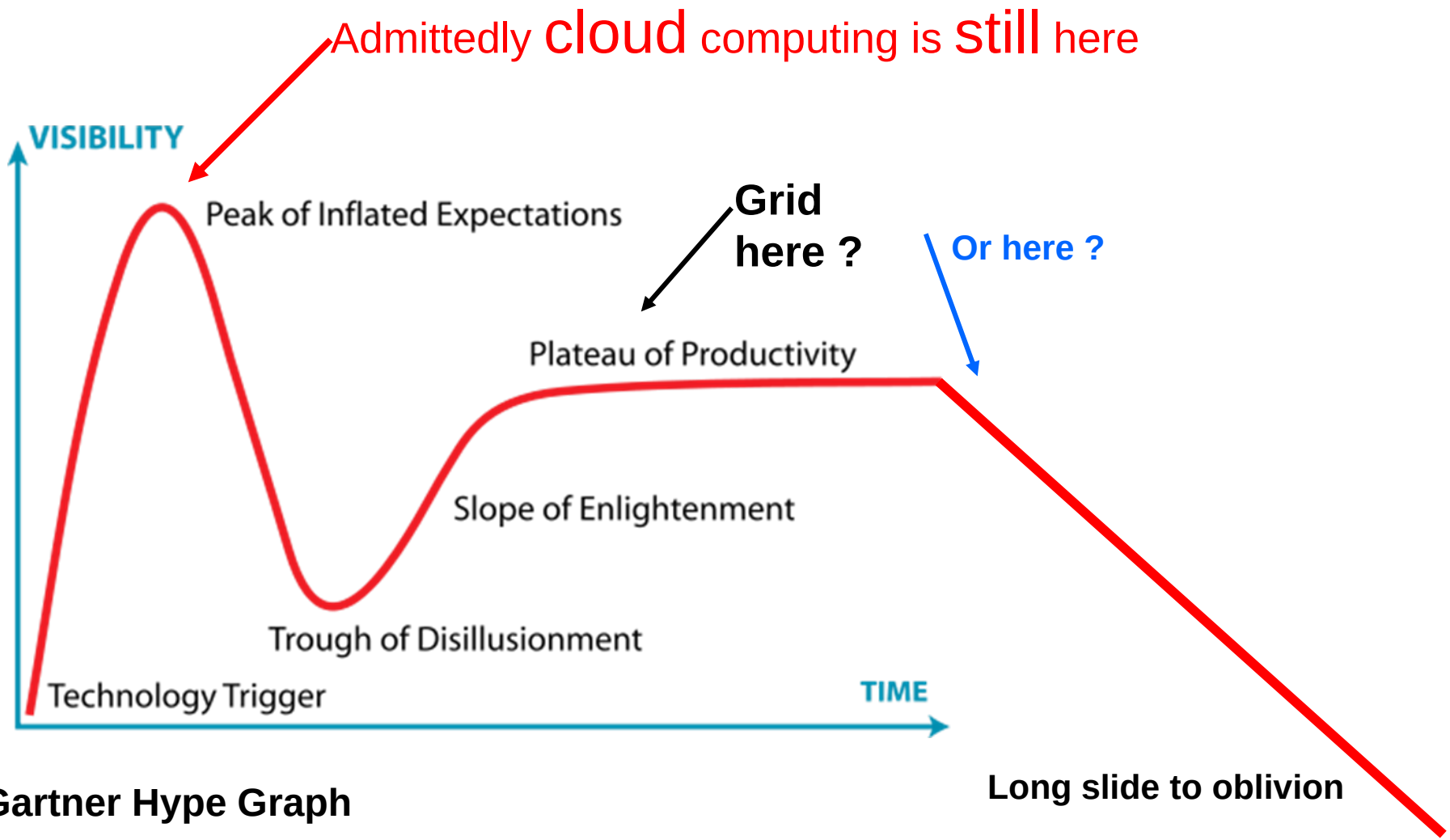
- By 2020, all scientists of all disciplines will choose the European Cloud Computing Infrastructure as their first option to store and access data, for data processing and analysis.
- This infrastructure will be considered as a natural infrastructure for the global science community similar to the road or telecommunication infrastructure for the general public today.
- This infrastructure will contain vast quantities of data, an unrivalled array of open source tools, and a literally infinite amount of computing power accessible and usable from any kind of computer, smart phone or tablet device.

European strategic plan to put functionality in place for 2020.

Finally – Virtualized Observatory?

- For Gaia looking at virtualization/cloud for complex data interactions
  - DBMS/Tap will work for many queries
  - But there are many more which will basically require data ‘trawl’ – bring data across wire will not be efficient
    - Virtualization could provide a way to run ‘my code’ in the archive
    - All those complex statistical operators you want on **ALL** data
    - Also could allow advanced user applications to run in archive
    - Easier if the whole Archive is in the cloud
    - Could also allow Pay as You Go clients then
- CANFAR / SKA already on this road – CADC in Gaia working group on archive
- Others also (hence this session at ADASS!)

Head in the clouds



Conclusion

- Cloud is a nebulous thing
- But it is here and now
- It is NOT for everyone and all things
- But you probably do not want to ignore it completely
  - Great for short projects and testing
  - Can be cheaper for PEAK processing
- You will still want to keep your data backed up on earth someplace.
- For development and debugging you probably want local machines also

Questions ?