Cloud Computing – How is it different from Grid Computing?

Trial lecture in conjunction with Ph.D. thesis defense

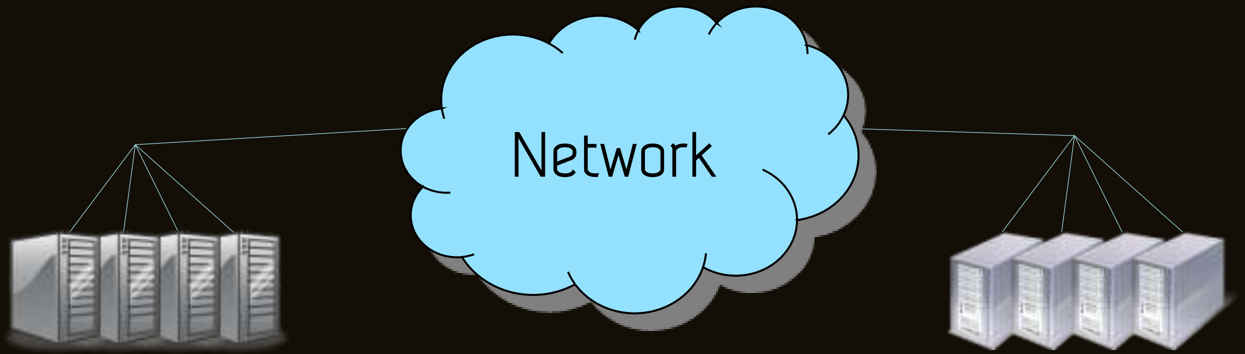
André Rigland Brodtkorb
2010-12-17

Outline

- Distributed and Grid Computing
- Cloud Computing and differences from the Grid
- Summary and Future Trends

Distributed Computing

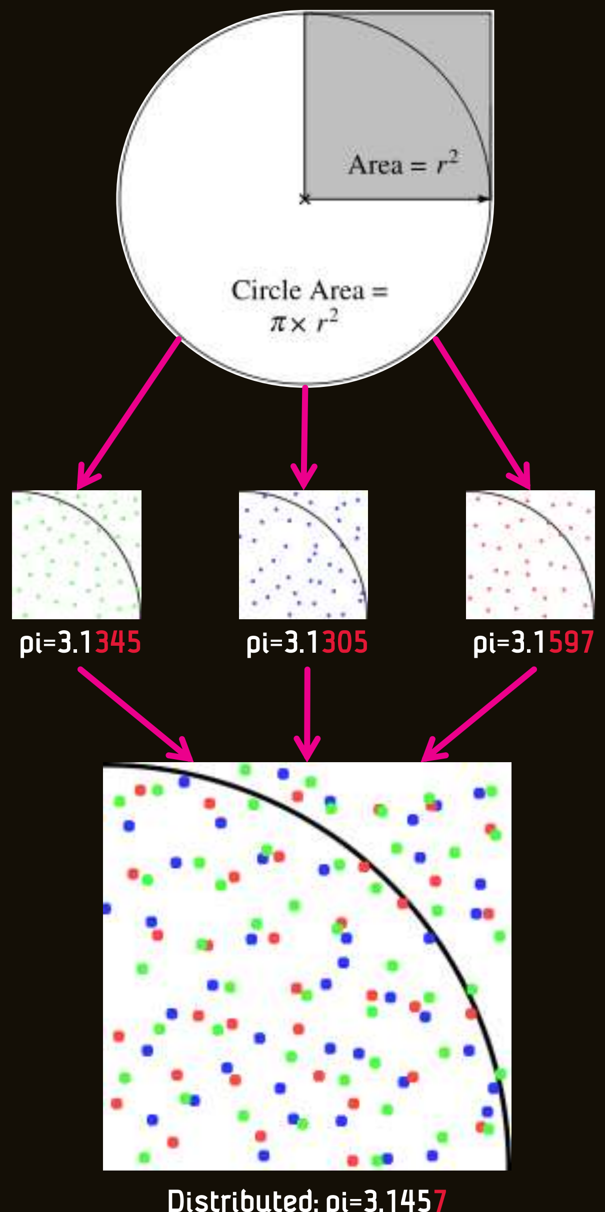
A distributed system consists of ***multiple autonomous computers*** *that* ***communicate through a*** *computer* ***network****.*
-- Wikipedia: Distributed computing



- Autonomous computers
  - Independent machines such as laptops, desktops, clusters, supercomputers, etc.
- Computer Network
  - The internet, (gigabit) ethernet, wifi, infiniband, etc.

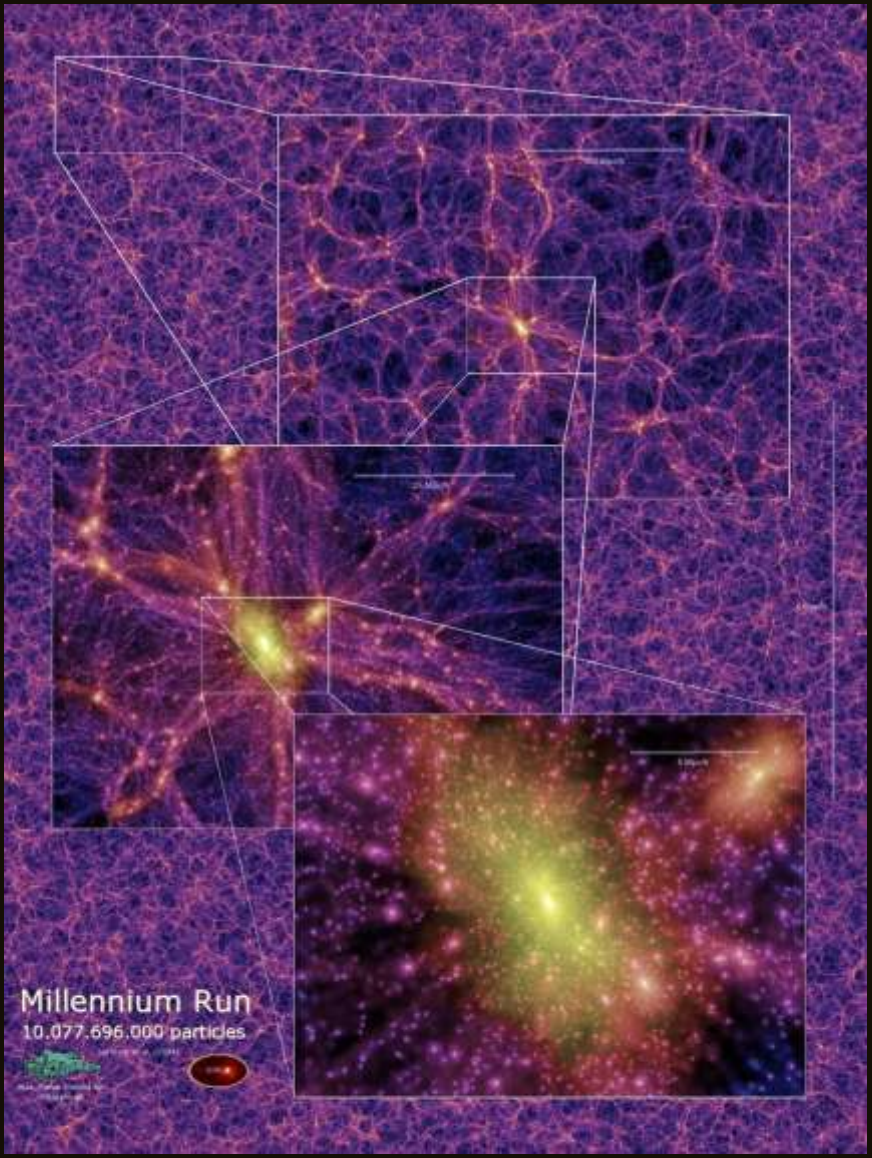
Distributed Computing

- No single machine has a full view of the problem
  - Each machine computes on a subproblem
- Example: Distributed computation of pi (3.14159…)
  - Disclaimer: this is a naïve way of calculating PI, only used as an example of distributed computing
  - Sample random points within one quadrant
  - Find the ratio of points inside to outside the circle
    - Pi ≈ 4 * ratio
  - Increase accuracy by sampling more points
  - Increase speed by using more nodes



Distributed Computing

- Some problems require distributed computing
  - No single machine is large enough
    to represent the full problem
  - Example: The Millennium Run (2005)
    - Simulated evolution of matter in the universe
    - 10 billion particles
    - 25 terabytes of output
  - No single machine is fast enough
    - Weather reports
    - Tsunami warnings



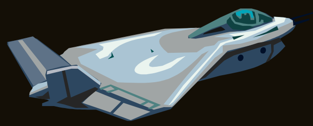
Grid Computing: Four examples [1]

- Placement of a new factory
  - Simulation runs at an application service provider
  - Historical data for the simulation hosted at a storage service provider
  - Application service provider uses a cycle provider for more "oomph"
- Feasibility study of supersonic aircraft
  - Multidisciplinary simulation
  - Different parts of the simulation runs on different participants machines
- Crisis management team for chemical spill
  - Local weather and soil models
  - Chemical reaction models
  - Planning and coordinating evacuation
- Large Hadron Collider – CERN
  - Pooling of computing, storage and networking resources
  - Analysis of petabytes of data.

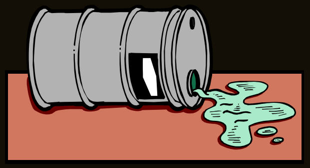
[1] I. Foster et al., The Anatomy of the Grid: Enabling Scalable Virtual Organizations, 2001

What did "The Grid" offer to each use case?

- Placement of a new factory:
  - Coupling of multiple simulations
  - Third-party storage
  - Third-party computations



- Feasibility study of supersonic aircraft:
  - Coupling of simulations at different locations

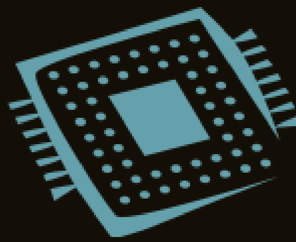


- Crisis management team for chemical spill:
  - Coupling of simulations at different locations

- Large Hadron Collider – CERN:
  - Storing and processing vast amounts of data

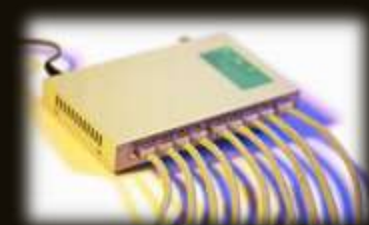
Grid Computing – The Grand Idea

- The four cases:
  - Do not exchange files/data – share resources



- Sharing of resources
  - CPU cycles
  - Software / Applications
  - Storage
  - Network
  - …

- Highly dynamic sharing of compute resources
  - Resource sharing "like a power grid"
  - Small and large providers implement a set of standard protocols
  - Mutually distrustful partners
  - (Potentially very) short time-span



Grid Computing Concepts

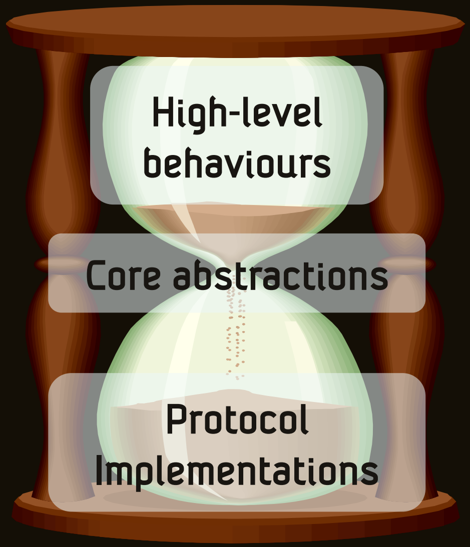
Grid computing is the "coordinated resource sharing [...] in virtual organizations"

-- I. Foster et al, **Anatomy of the Grid**, 2001

- Virtual Organizations – A key concept in Grid Computing
  - Partners who intend to share resources to reach a common goal
  - Vary dramatically in purpose, scope, size, duration, structure, etc.
  - Partners can be both producers and consumers
  - Each of the four use-cases are examples
- Grid computing facilitates resource sharing in *scalable* virtual organizations
  - No VPN connection

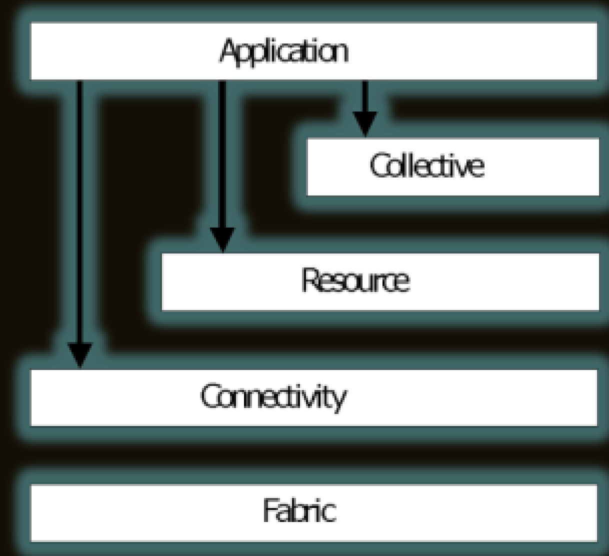
Grid Architecture: The hour glass model

- The number of abstractions must be limited
  - They should not over determine possibilities, or limit performance
  - They are required to capture the fundamental mechanisms for resource sharing
- Grid protocols are built on internet protocols
  - To be on "The Grid" you must be on the internet



Grid Architecture

- The Grid is a layered architecture
- The application layer consists of the applications a user will implement
- **The collective layer contains general-purpose utilities**
- **The resource layer specifies protocols for operating with shared resources**
- **The connectivity layer specifies protocols for secure and easy access**
- The Fabric consists of the resources to be shared
  - Network bandwidth, CPU time, Applications, etc.



Middleware

Figure from I. Foster et al., The Anatomy of the Grid: Enabling Scalable Virtual Organizations, 2001

Grid Architecture: Connectivity Protocols

- Communication and authentication protocols
  - Single sign-on
  - Authorization delegation to programs
  - Account management
- Example: GSI - Grid Security Infrastructure (built around existing TLS protocols)

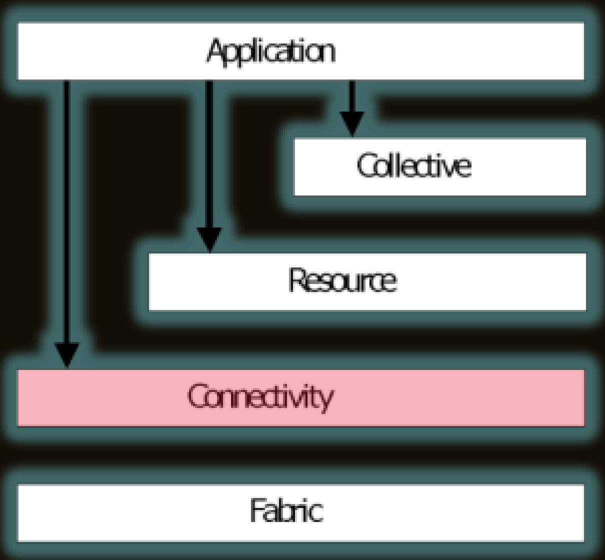


Figure from I. Foster et al., The Anatomy of the Grid: Enabling Scalable Virtual Organizations, 2001

Grid Architecture: Resource Protocols

- Give a unified way of accessing the fabric resources
  - Builds on the connectivity layer
  - Consists of information and management protocols
    - Information protocols offer information about shared resources
    - Management protocols to negotiate access to shared resources
- Example: GRIP - Grid Resource Information Protocol (Based on LDAP)
- Example: GRAM - Grid Resource Access and Management (Allocation and monitoring of resources)

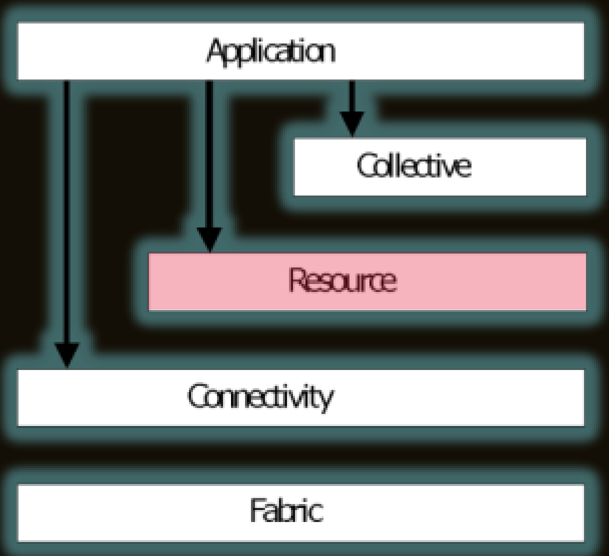


Figure from I. Foster et al., The Anatomy of the Grid: Enabling Scalable Virtual Organizations, 2001

Grid Architecture: Collective Layer

- Builds important utilities on top of resource and connectivity protocols
  - Co-allocation of resources
  - Monitoring
  - Data replication
  - Accounting

- APIs and SDKs:
  - MPI for grids: GridMPI, MPICH-G2, etc.
  - Checkpointing, job management, etc.

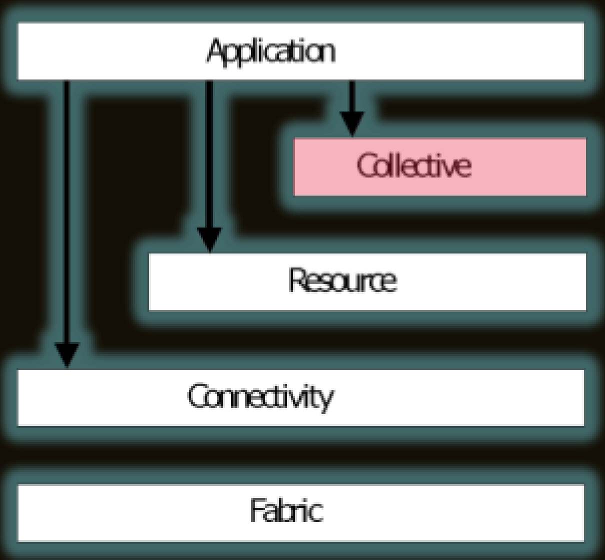
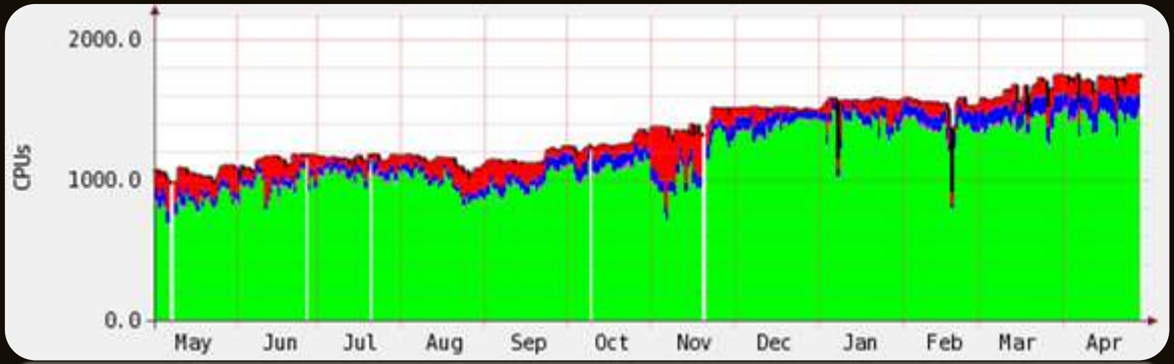


Figure from I. Foster et al., The Anatomy of the Grid: Enabling Scalable Virtual Organizations, 2001

Collective Layer: Brokers and Schedulers

- Grid infrastructure is expensive
  - Typically high-performance clusters
    => Want close to 100% resource utilization

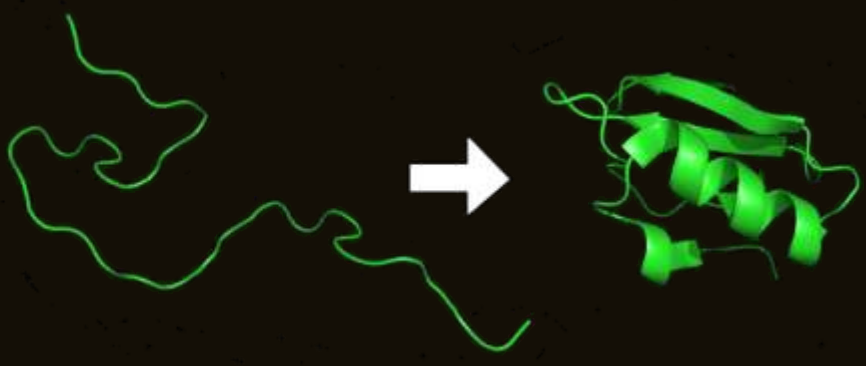
- Grids thus often use batch schedulers
  - Maintains a queue of incoming applications
  - GRAM - Grid Resource Allocation Manager
  - JSDL - Job Submission Description Language
  - Condor
  - SLURM - Simple Linux Utility for Resource Management



Statistics graph by Steve Rader, http://oss.oetiker.ch/rrdtool/gallery/index.en.html

Volunteer Computing – A part of Grid Computing?

- Prototypical use-cases
  - SETI: Search for Extra-Terrestrial Intelligence
  - Protein Folding for drug development
  - Calculation of Mersenne primes (cryptography)
- Volunteer computing properties
  - Embarrassingly parallel
  - Compute intensive
  - Little data communication



- Projects:
  - Cummulative BOINC project contribution: 11.71 Petaflops, 6 million computers
  - MilkyWay@home – 2.6 Petaflops, Study of stellar systems
  - Folding@home – 7.87 Petaflops, Protein folding
  - SETI@home – 769 Teraflops, Search for extraterrestrial intelligence
  - World Community Grid - 510 000 user accounts - FightAids@home, Dengue drugs, Genome comparison, clean energy, etc.

Protein folding.png from Wikipedia, DrKjaergaard

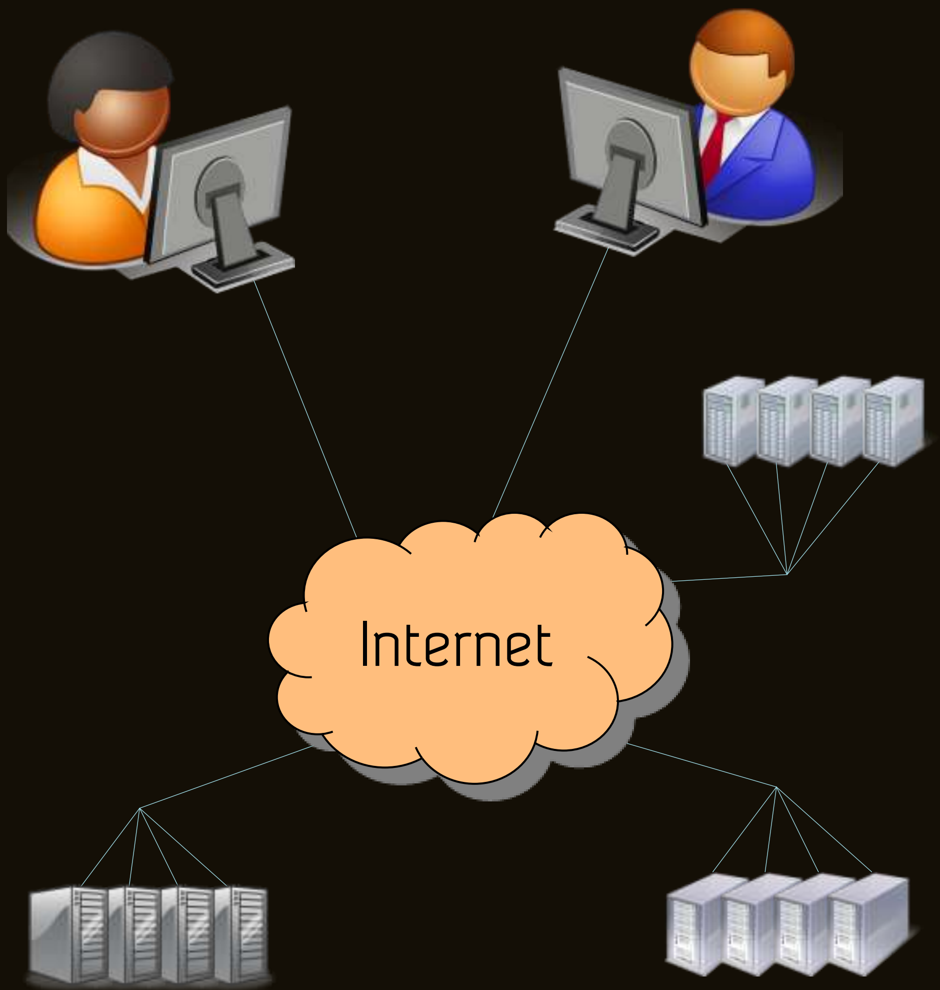
Is Volunteer Computing *Really* Grid Computing?

- Yes
  - A large number of autonomous computers connected through a network
  - A virtual organization joint for a common goal
  - Automatic resource management
- But not quite?
  - One-sided relationship: contributors get nothing (or "fame") in return for resource sharing
- What about peer-to-peer computing? [1]
  - Bittorrent, The Onion Router, etc.

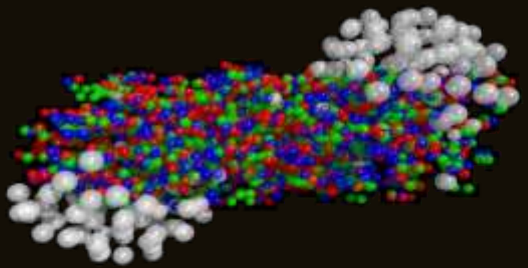
[1] I. Foster and A Iamnitchi, On Death, Taxes, and the Convergence of Peer-to-Peer and Grid Computing, 2003

Examples of Today's Grids

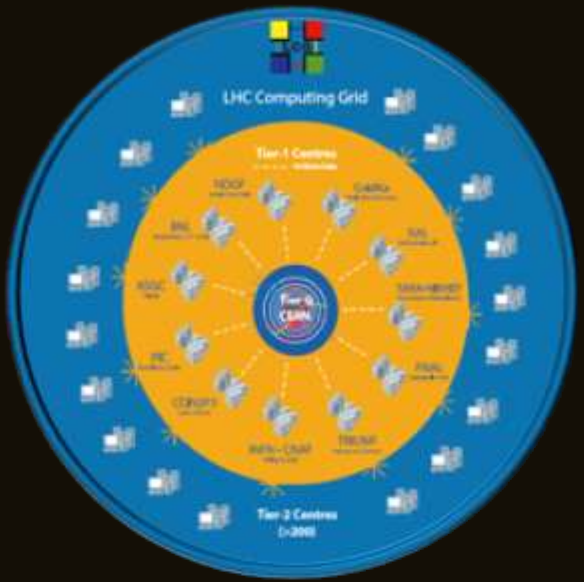
- WLCG - Worldwide LHC Computing Grid
  - Processing of data from LHC
- NDGF - Nordic Data Grid Facility / Nordunet
  - Collaboration mainly between Norway, Sweden, Denmark , and Finland.
  - Part of WLCG
- Open Science Grid
  - American collaboration of 72 institutions
  - 25000 computers (43k processors)
  - Part of WLCG
- TeraGrid
  - American collaboration of 11 institutions
  - Mostly Molecular Biosciences, physics and chemistry
- EGI - European Grid Infrastructure
  - Supersedes EGEE - Enabling Grids for E-sciencE
  - 50 institutions, 40 countries



WLCG – Worldwide LHC Computing Grid



- **The worlds largest grid**
  - Storage and processing of vast amounts of data
    - 15 million gigabytes per year
  - Too expensive to build one datacenter
  - 8000 researchers who access data
- Over 140 datacenters in 34 countries
  - Tier 0: CERN
  - Tier 1: Dedicated 10Gb/s fiber to 11 major datacenters
    - 63,723,713 GiB *online* Storage (plus additional tape storage)
    - 70,382 CPU cores
  - Tier 2: 140 sites
- Four virtual organizations (one for each LHC experiment)
  - ALICE, ATLAS, CMS, LHCb
- Middleware (Grid technology)
  - European Middleware Initiative (EMI -> ARC, gLite, UNICORE, dCache)
  - The Globus Toolkit
  - Open Middleware Infrastructure Institute
  - Virtual Data Toolkit



Top image from http://cdsweb.cern.ch/record/1221418
Bottom image from CERN, "An EGEE Comparative Study: Grids and Clouds – Evolution or Revolution?", 2008

NDGF - Nordic Data Grid Facility / Nordunet

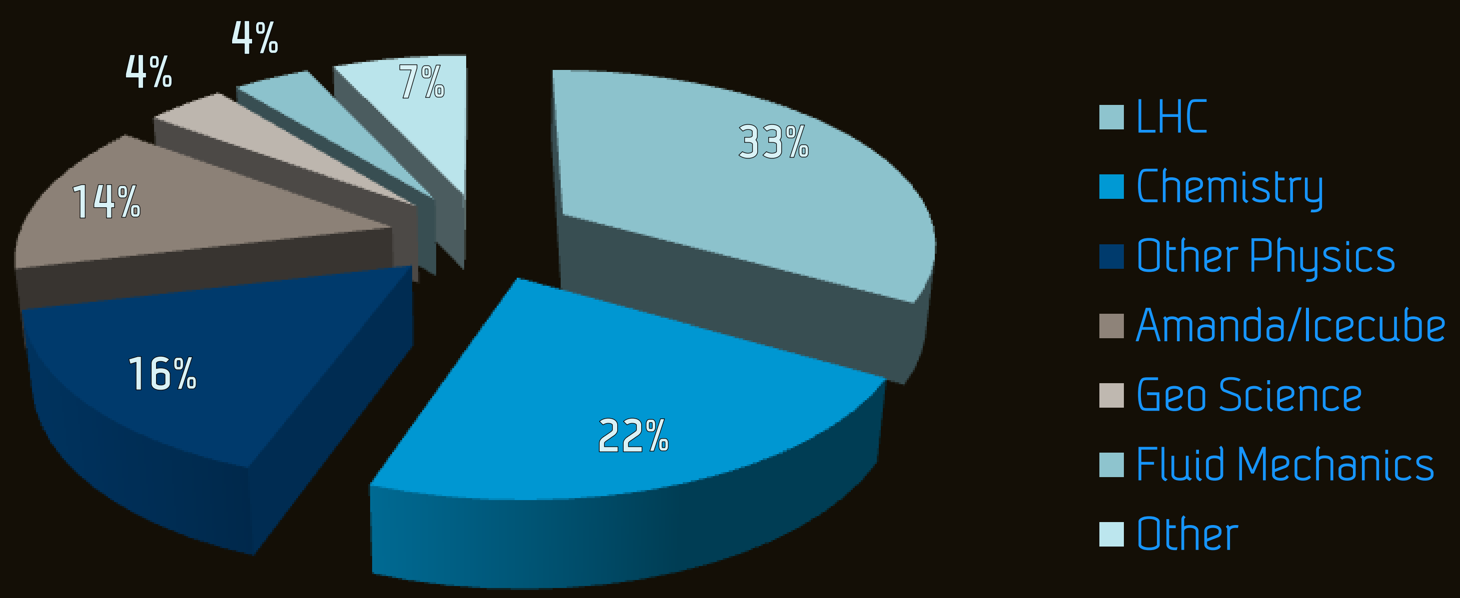
- Building large clusters and supercomputers is expensive!
  - Data centers in Nordic countries can share compute resources with each other to solve larger problems than would be possible alone.
  - Funded by Nordic research councils

- Creator of ARC middleware
  - Builds on standards, including the Globus toolkit

- Applications:
  - LHC,
  - BioGrid,
  - Other academic uses

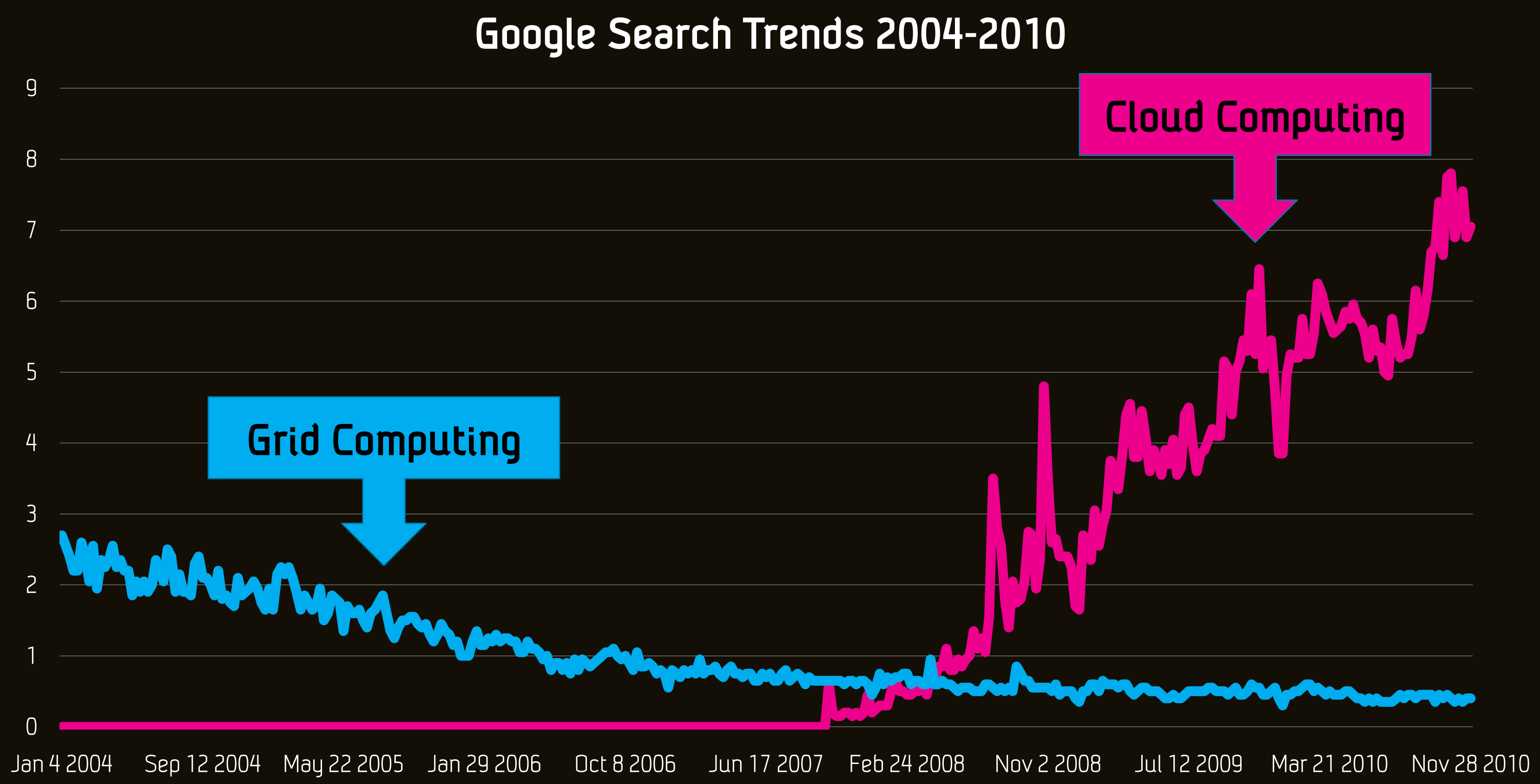
Image from http://www.nordugrid.org/

SweGrid Use



Data from Oxana Smirnova, How new communities can get access to a Grid infrastructure, 2009

From Grids to Clouds



Data from Google Search Trends, http://www.google.com/trends, 2010

What is the Cloud?

*"When talking about the cloud, it is **mandatory** to use at least one fitting quote"*

-- A. R. Brodtkorb, Ph.D. Trial Lecture, 2010

The Mandatory Quotes

*"The interesting thing about cloud computing is that **we've redefined cloud computing to include everything we already do**.... I don't understand what we would do differently in light of cloud computing other than change the wording of some of our ads."*

– Larry Ellison, CEO Oracle, 2008

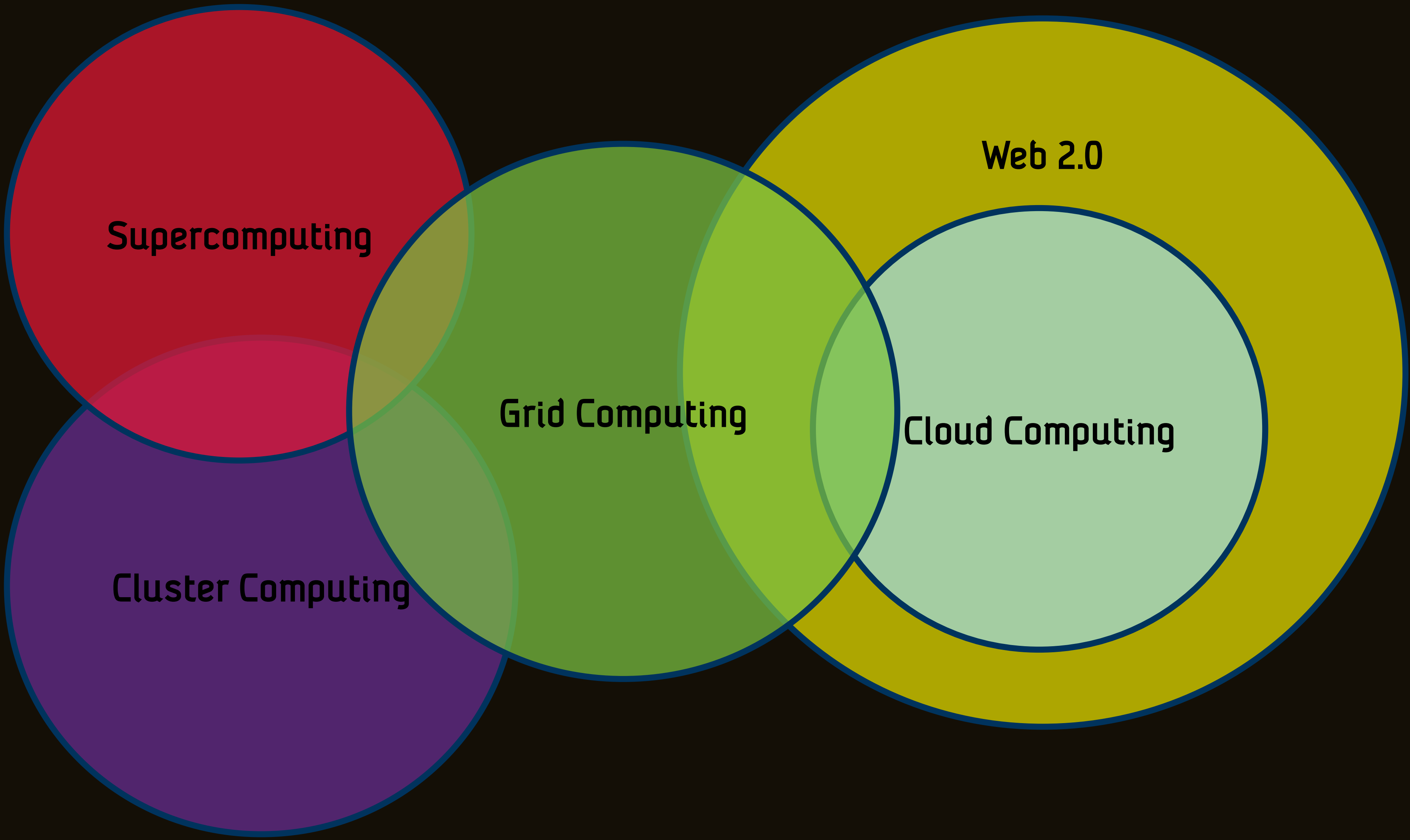
*"A lot of people are jumping on the [cloud] bandwagon, but **I have not heard two people say the same thing about it**."*

– Andy Isherwood, HP VP Support & Services, 2008

*"**It's stupidity. It's worse than stupidity: it's a marketing hype campaign.** Somebody is saying this is inevitable — and whenever you hear somebody saying that, it's very likely to be a set of businesses campaigning to make it true."*

– Richard Stallman, Founder of the GNU Project, 2008

Grids and Clouds



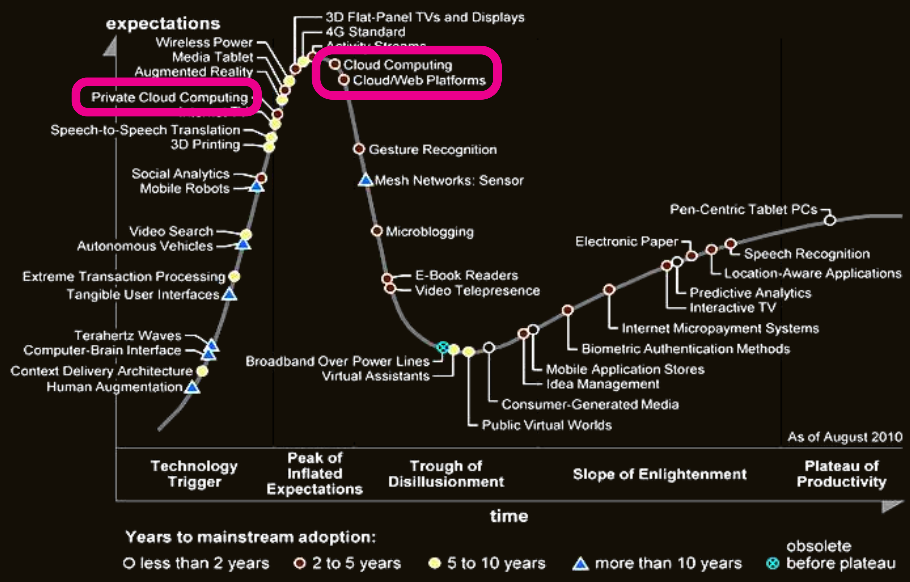
Adapted from *Cloud Computing and Grid Computing 360-Degree Compared*, I. Foster et al.,, 2009, DOI: [10.1109/GCE.2008.4738445](10.1109/GCE.2008.4738445)

Defining the Cloud

- The Cloud is a relatively new concept
  - A wide variety of definitions [1]
  - Garner says:
    Will hit mainstream market in 2-5 years [2]

- Important concepts and technologies
  - Web-based
  - Virtualization
  - On-demand
  - Service level agreements

- Important keywords
  - Platform as a Service
  - Software as a Service
  - Infrastructure as a Service



[1] Jeremy Geelan, "Twenty-One Experts Define Cloud Computing", 2009
[2] Gartner, "2010 Hype Cycle Special Report", 2010
Figure from [2]

Cloud Computing Core Concepts

- Enabling technology: Web 2.0
  - AJAX, SOAP, etc.
  - Used in: Google docs, web applications, etc.

- Enabling technology: Virtualization
  - Multiple virtual machines can occupy the same physical hardware
  - **Virtual Machines in cloud computing is the equivalent of threads of parallel computing**

- Enabling technology: Multi-core
  - Multi-core means that multiple virtual machines can easily run on the same physical machine

- Enabling technology: Idling server parks
  - Less than 10% server utilization [1, 2]

[1] Jeff Bezos to technologyreview.com, "Servers For Hire" by Wade Roush, 2006
[2] Vivek Kundra to datacenterknowledge.com, "Kundra: Fed Data Centers 7 Percent Utilized", 2010

Cloud Computing Keywords

- Platform as a Service
  - Offers a high-level programming API
  - Examples: Microsoft Azure, Google AppEngine
- Software as a Service
  - Examples: Google docs (Office applications), Salesforce.com (CRM - Customer relationship management)
- Infrastructure as a Service
  - Lease virtual machines
  - Example: Amazon EC2, Amazon S3, Dropbox

Cloud Infrastructure as a Service: Amazon EC2 and S3

- Amazon EC3: Elastic Compute Cloud
  - Buy virtual machines just-in-time
    - All you need is a credit card to get started
    - Standardized Service Level Agreements
  - Supply a virtual machine image of your choice
    - Or choose from a set of standard images
    - Machine goes online in less than two minutes
  - Supply input to running images
    - SSH, etc
- Amazon S3: Simple Storage Service
  - Buy storage just-in-time
  - Accessible from HTTP and BitTorrent

Amazon EC2 - Pricing Model (Service Level Agreements)

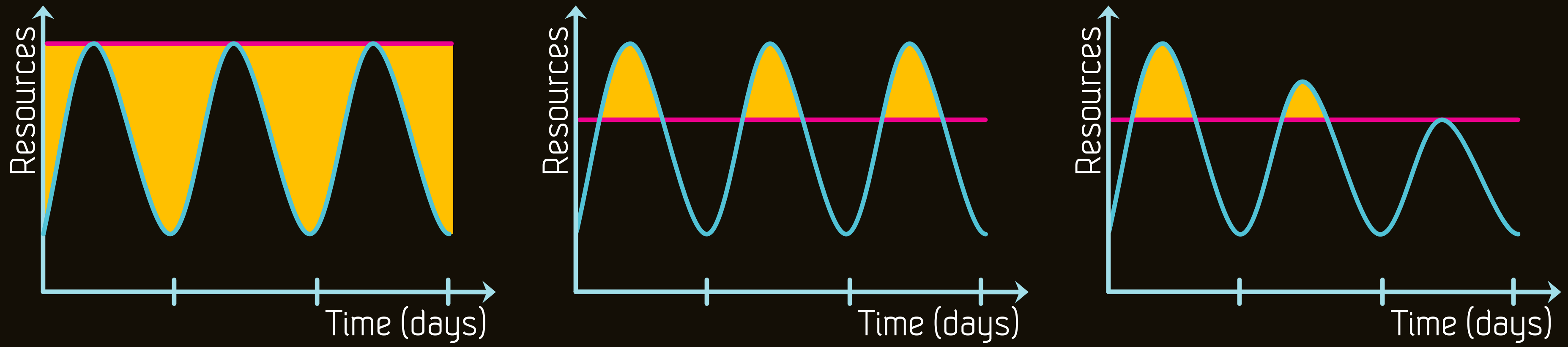
Service	Price
Micro Instance	$0.02 / hour
Large CPU Instance (7.5 GiB RAM, 2 cores)	$0.34 / hour
High CPU Instance (7.0 GiB Ram, 8 cores)	$0.68/ hour
Cluster Compute Instance (23 GiB, 8 cores, 10GiB Ethernet)	$1.60/ hour
Cluster GPU Instance (22 GiB, 8 cores, 2xM2050, 10GiB Ethernet)	$2.10/hour
High Redundancy S3 Storage	$0.14 / GiB / Mnth
Reduced Redundancy S3 Storage	$0.093 / GiB / Mnth
EBS Storage	$0.10 / GiB / Mnth
Small MySQL Instance	$0.22 / hour

- NEW Q4 2010: GPU instances
- Windows images cost slightly more (10-20 cent)
- Data transfer in and out of S3 datacenter is not free

Table adapted from J. Seland, Cloud Computing, 2010

Cloud Selling Point #1: Fine-grained Scalability

- Small startup company with a good idea (e.g., dropbox)
- How much infrastructure do you need?
- Using the cloud will probably be the least expensive alternative



- Capacity
- Demand
- Problematic demands

Adapted from *A View of Cloud Computing*, M. Armbrust et al.,, 2010, DOI: [10.1145/1721654.1721672](10.1145/1721654.1721672)

Cloud Selling Point #2: Pricing schemes

"1000 servers for one hour costs no more than using one server for 1000 hours" [1]

Ray tracing / rendering

- Time to delivery might be worth a lot
- Rendering Ratatouille took over 1500 CPU years
- Payroll calculations
- Other "peaky" workloads

Payroll

[1] Armbrust et al., A View of Cloud Computing, 2010

The Economies of Scale

"The GSA ***previously paid \$2.35 million in annual costs for USA.gov****, including \$2 million for hardware refreshes and software re-licensing and \$350,000 in personnel costs,* ***compared to the \$650,000 annual cost to host the site with Terremark****."* [1]

- Huge data-centers are cost effective [2]
  - Costs down a factor 5-7
- Clouds could mean more green computing [3]
  - Higher resource load means less servers
  - US Datacenters: 170 Mt CO2
  - Argentina: 142 Mt CO2
  - Netherlands: 146 Mt CO2
  - Malaysia: 178 Mt CO2

[1] Vivek Kundra to datacenterknowledge.com, "Kundra: Fed Data Centers 7 Percent Utilized", 2010
[2] Armbrust et al., A View of Cloud Computing, 2010
[3] Edward J. Lucente, The Coming 'C' Change in Datacenters , hpcwire.com, 2010

Cloud Uncertainties

- Who has access to your data?
  - Cloud provider?
  - Subcontractors?
  - Other customers?
  - Law enforcement agencies in another country?

- What happens if the cloud provider (or subcontractor)
  - Goes out of business?
  - Is bought by a third party?
  - Is brought down for political or legal reasons? [1, 2]
  - Is blacklisted because of another customer (Wikileaks, email spammer, etc.)?
  - Leaks its password database? (Gawker)

- Other problems
  - Vendor lock-in: can you export your data?
  - Debugging at large scales?
  - Software licenses?

[1] Brian Krebs, Amazon: Hey Spammers, Get Off My Cloud!, Washington Post, 2008
[2] Kim Zetter, Company Caught in Texas Data Center Raid Loses Suit Against FBI, wired.com, 2009

Is the "Cloud" something new?

"Computation may someday be ***organized as a public utility****."*
– John McCarthy, Inventor of Lisp, 1961

- No! Same sheep, different wrapping:
  - Grids: "Dynamic resource sharing"
  - Utility Computing: "Pay as you go"
  - The natural evolution of the grid

- Yes! Technology enablers have realized old ideas:
  - Appearance of infinite computing resources
  - No up-front commitment from users
  - Pay as you go on a short-term basis

Summary

In 2012 80% Fortune 1000 will pay for some cloud computing service; 30% will pay for Cloud infrastructure.

-- Gartner, 2008

- Most of us are users of "the cloud" already
  - Youtube, Dropbox, Google docs, etc.
  - Clouds will take over more and more of private computing
- But there are major obstacles:
  - Privacy
  - Data safety
  - Vendor lock-in
- Future of Grids:
  - Have built up a large user base and mature technology
  - Will continue to prosper in academic communities
- Future of Clouds:
  - Economic potential is huge => No up-front commitment
  - Economies of scale => Green computing
  - GPU Clouds => True GPU virtualization next?
  - High-performance computing and academic use of Clouds?

Some References and Further Reading

- I. Foster, **The Grid: A New Infrastructure for 21st Century Science**, Physics Today, Volume 55, Issue 2, pp. 42-47 (2002).
- I. Foster, C. Kesselman, S. Tuecke, **The Anatomy of the Grid: Enabling Scalable Virtual Organizations**, Proceedings of the 7th International Euro-Par Conference Manchester on Parallel Processing , 2001
- Ian Foster and Adriana Iamnitchi, **On Death, Taxes, and the Convergence of Peer-to-Peer and Grid Computing**, 2nd International Workshop on Peer-to-Peer Systems (IPTPS'03), 2003
- Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, Matei Zaharia, **A view of cloud computing**, Communications of the ACM, Volume 53, Number 4 (2010), Pages 50-58
- Foster, I. Yong Zhao Raicu, I. Lu, S. , **Cloud Computing and Grid Computing 360-Degree Compared**, Grid Computing Environments Workshop, 2008. GCE '08, Page(s): 1 – 10
- EGEE Report: **Grids and Clouds – Evolution or Revolution?**, 2008
- Open Cloud Manifesto, www.opencloudmanifesto.org
- Rajkumar Buyya, Chee Shin Yeo, Srikumar Venugopal, James Broberg, Ivona Brandic, **Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility**, Journal of Future Generation Computer Systems Volume 25 Issue 6, June, 2009
- Marios D. Dikaiakos, Dimitrios Katsaros, Pankaj Mehra, George Pallis, Athena Vakali, **Cloud Computing: Distributed Internet Computing for IT and Scientific Research**, IEEE Internet Computing, pp. 10-13, September/October, 2009
- Cloud Computing Use Case White Paper V4
- Open Grid Forum -- www.gridforum.org
- Grid Café – www.gridcafe.org
- I. Foster, C. Kesselmann, The Grid: Blueprint for a New Computing Infrastructure