Universität Augsburg University

P2P & Cloud Computing

Prof. Dr. Jörg Hähner

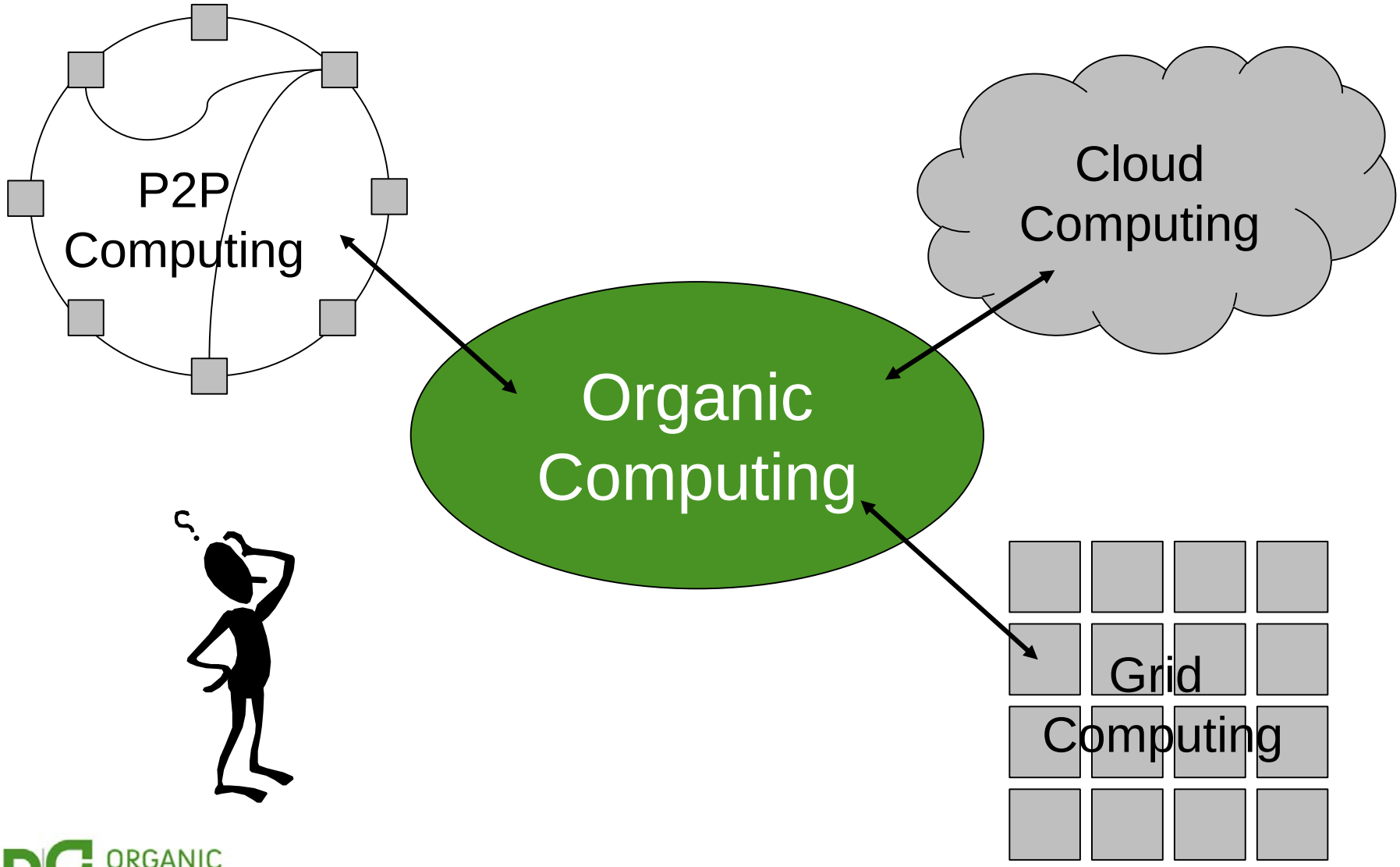
Part 1:

Organic, Cloud, Grid and Peer-to-Peer Computing?

ORGANIC COMPUTING

- Organic Computing
- P2P Computing
- Cloud Computing
- Grid Computing
- Comparison: Grid an Cloud Computing
- Conclusion and Literature

Organic- / P2P- / Cloud- / Grid-Computing?



- Organic Computing is based on the insight that we will soon be surrounded by:
  - large collections of autonomous systems,
  - which are equipped with sensors and actuators,
  - aware of their environment,
  - communicate freely,
  - and organise themselves in order to perform the actions and services that seem to be required.
- An "Organic Computing System" (or just an organic system) is a technical system, which is equipped with sensors (to perceive its environment) and actuators (to manipulate it). It adapts autonomously and dynamically to the current conditions of the perceived environment. This adaptation process has impact on the system's performance, which is continuously optimised by the organic system itself. To allow for such an adaptive behaviour, it is characterised by self-X properties and learning capabilities.

- Organic Computing
- P2P Computing
- Cloud Computing
- Grid Computing
- Comparison: Grid an Cloud Computing
- Conclusion and Literature

"Peer-to-Peer systems are distributed systems consisting of interconnected nodes able to self-organise into network topologies with the purpose of sharing resources such as content, CPU cycles, storage and bandwidth, capable of adapting to failures and accommodating transient populations of nodes while maintaining acceptable connectivity and performance, without requiring the intermediation or support of a global centralised server authority."

[S. Androutsellis-Theotokis, Diomidis Spinellius: A Survey of Peer-to-Peer Content Distribution Technologies. ACM Computing Surveys, 36(4), 2004]

Note: variations from this basic principle exist!

Self-organising system (the basic principle)

- No central control or coordination of resources
  - Peers access other peers' resources without coordination by central server
- Structure of P2P network emerges through distributed decisions of peers
  - Peers autonomously build and maintain the P2P network
- System constantly adapts to dynamic resources
  - Peers joining, leaving, failing

Coping with a large number of autonomous and dynamic peers is a key distinction of a P2P system to other distributed systems.

- Similarities:
  - Large collections of autonomous entities
  - Communicate freely
  - (Organise themselves)

- Differences:
  - OC has a much more broader scope.
    → P2P is „just“ an application, while OC tackles the whole system!
  - P2P systems serve *one* specific purpose only (e.g. computation, storage).
    → OC systems might perform different tasks simultaneously!
  - P2P clients are typically normal computers.
    → OC systems are all kinds of sensor/actuator systems!
  - P2P systems are aware of their direct neighbours.
    → OC systems have a broader awareness-level (i.e. environment- and self-awareness)!
  - OC systems are computationally more complex and sophisticated.

- Organic Computing
- P2P Computing
- Cloud Computing
- Grid Computing
- Comparison: Grid an Cloud Computing
- Conclusion and Literature

"One of the appealing aspects of Cloud Computing is that it hides the complexity of IT technology from users and developers. No need to know details of how a service is generated – it is the service provider's job to provide a corresponding abstraction layer."

[Source: Christian Baun: Cloud Computing: Web-Based Dynamic IT Services, Springer Verlag 2011]

- **Virtualisation:**

  The ability to run multiple operating systems on a single physical system and share the underlying hardware resources.
  [VMware white paper, Virtualisation Overview]

- **Cloud Computing:**

  "The provisioning of services in a timely (near on instant), on-demand manner, to allow the scaling up and down of resources".
  [Alan Williamson, quoted in Cloud BootCamp March 2009]

- Cloud Computing is based on several existing techniques, e.g.:
  - Virtualisation
  - Service-oriented architectures (SOA)
  - Web-services
- Virtualisation
  - … allows an abstract, logical view on the physical resources.
  - … includes servers, data stores, networks, and software.
  - Physical resources are pooled and managed as a whole.
  - Individual requests are served as currently required.
  - I.e.: dynamically generate a certain platform for a specific application exactly in that moment when it is needed.
  - Instead of real machines, virtual ones are used.

- Similarities
  - OC and Cloud systems try to take „the user out of the loop“ using autonomous concepts.
  - Clouds only perform task, are partially aware of neighbours/ environment: autonomous cloud management.

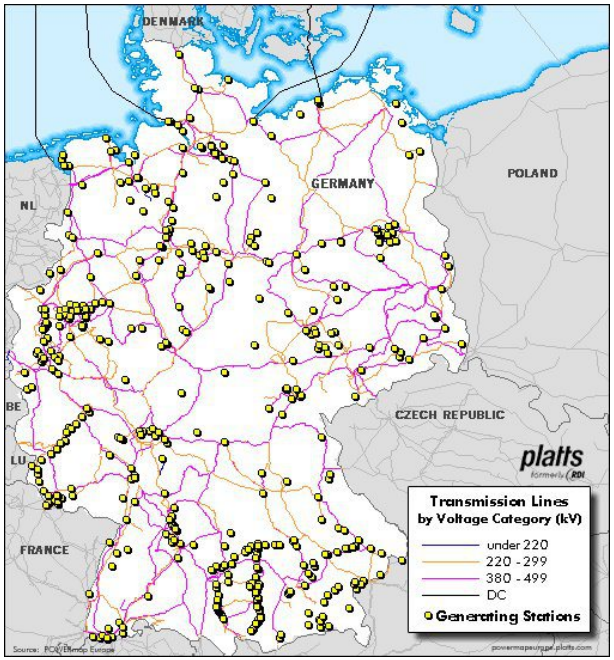
- Differences
  - OC has a much more broader scope.
    → Clouds are „just“ applications, while OC tackles the whole system!
  - Clouds serve one specific purpose only (e.g. computation, storage).
    → OC systems might perform different tasks simultaneously!
  - Clouds are typically hardware clusters.
    → OC systems are all kinds of sensor/actuator systems!

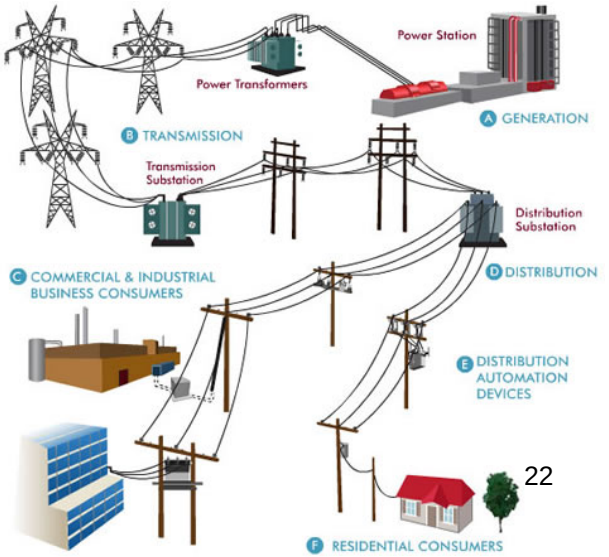
- Organic Computing
- P2P Computing
- Cloud Computing
- Grid Computing
- Comparison: Grid an Cloud Computing
- Conclusion and Literature

- Grid Computing – definition:
  - Grid computing is a form of distributed computing that involves coordinating and controlled sharing of diverse computing, applications, data, storage, or network resources across dynamic and geographically dispersed multi-institutional virtual organisations.
  - A user of Grid Computing does not need to have the data and the software on the same computer, and neither must be on the user's home (login) computer.

- Grid Computing
  - Term suggests a computing paradigm similar to an electric power grid.
  - A variety of resources contribute power into a shared "pool" for many consumers to access on an as-needed basis.

[Source: toryaardvark.com]





- The idea of Grid Computing resulted from the confluence of three developments:
  - The proliferation of largely unused computing resources (especially desktop computers).
  - Their greatly increased CPU speed in recent years.
  - The widespread availability of fast, universal network connections (the Internet).

Key concepts of a grid system:

- Coordination of distributed resources that are not subject to centralised control.
  - Integration of resources from different administrative units (different units of one or different companies/organisations).
- Using standard, open, general-purpose protocols and interfaces:
  - Authentication / authorisation
  - Resource discovery / resource access
  - …

Key concepts of a grid system (continued):

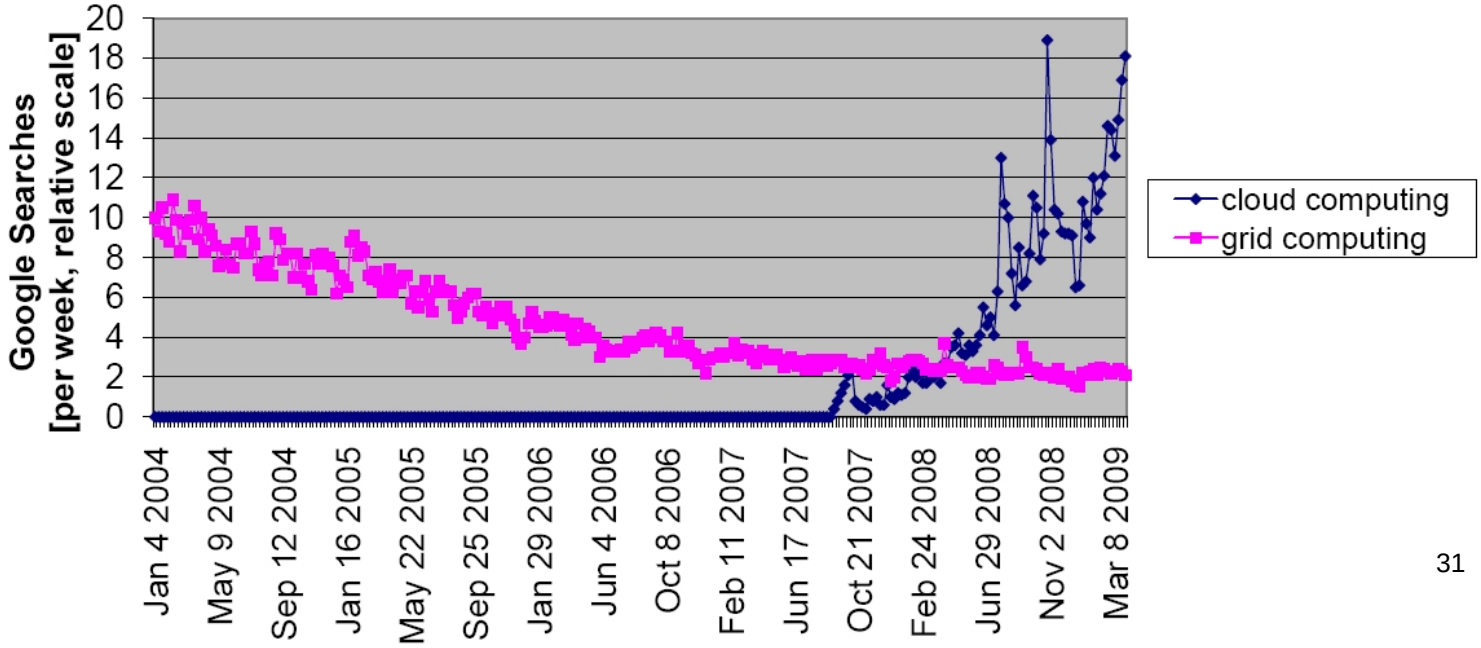
- Deliver nontrivial qualities of service to the user
  - Response time
  - Throughput
  - Availability
  - ...

- Similarities:
  - Organises efficient access to data / resources.
  - Caching, migration, replication are important issues.
  - Structured P2P systems are a step towards Grid systems.
- Differences:
  - P2P (especially unstructured) is more a basic technology used by Grid systems than a comparable approach!
  - In other words: Grid is situated at the application layer, while P2P serves as basis.
  - Grid is mainly used for computational tasks, P2P has a broader focus.
  - Grid includes authentication mechanisms, while most P2P do not.
  - Recovering and monitoring of jobs is seldom found in P2P systems.

- Organic Computing
- P2P Computing
- Cloud Computing
- Grid Computing
- Comparison: Grid an Cloud Computing
- Conclusion and Literature

- One distributed computing buzzword per decade:
  - Parallel Computing (~1978)
  - Meta Computing (~1987, L. Smarr)
  - Grid Computing (~1997, I. Foster / K. Kesselman)
  - Cloud Computing (~2007, E. Schmidt)

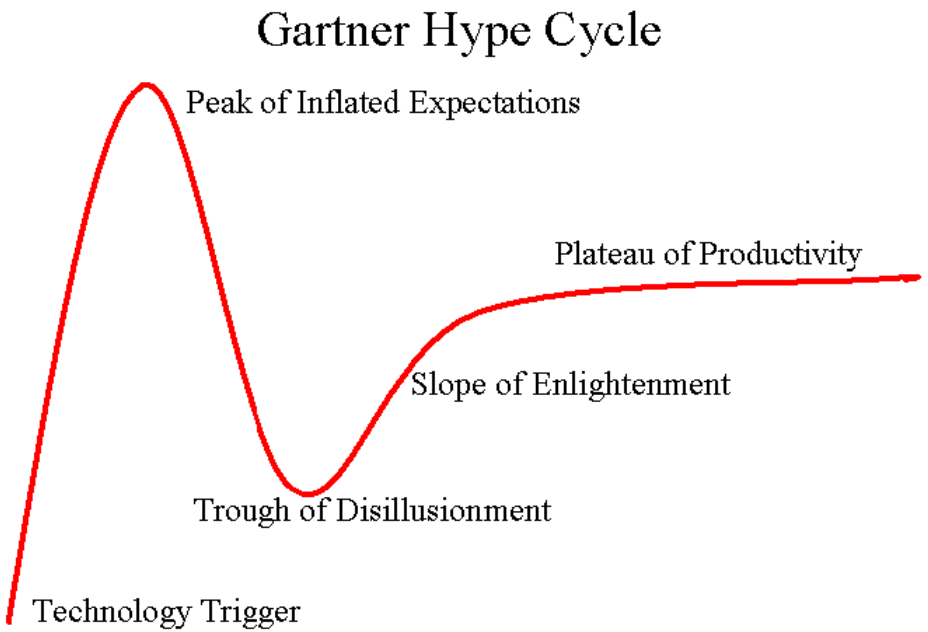
Grid vs. Cloud: Google Searches



The Current Cloud Hype?

"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
The Guardian, September 29, 2008

"The interesting thing about Cloud Computing is that we've redefined Cloud Computing to include everything that we already do… I don't understand what we would do differently in the light of Cloud Computing other than change the wording of some of our ads.
Larry Ellison
Wall Street Journal, September 26, 2008



- Service Grids
  - are systems that federate, share and coordinate distributed resources from different organisations which are not subject to centralised control, using standard, open, general-purpose protocols and interfaces to deliver non-trivial qualities of service. Service grids are used by Virtual Organisations, thematic groups of users crossing administrative and geographical boundaries.
- Cloud Computing
  - is an on-demand service offering a large pool of easily usable and accessible virtualised resources (such as hardware, development platforms and/or software services) in a pay-per-use model. Clouds are usually commercial and use proprietary interfaces.

Illustrating the difference:

Cooperative harvesting (Service Grid)

Contract harvesting (Cloud)

Clouds and Grids (4)

	Grids (e.g. EGEE)	Clouds (e.g. Amazon)
What?	Grids enable access to shared computing power and storage capacity from your desktop.	Clouds enable access to leased computing power and storage capacity from your desktop.
Who provides the service?	Research institutes and universities federate their services around the world.	Large individual companies.
Who uses the service?	• Research collaborations. • "Virtual organizations," comprising researchers located around the world.	• Small to medium commercial businesses. • Researchers with generic IT needs.
Who pays for the service?	Governments: providers and users are usually publically funded research organizations.	The cloud provider pays for the computing resources; the user pays to lease them.
Where are the computing resources?	In computing centers distributed across different sites, countries and continents.	In the cloud provider's private data centers, which are often centralized.
Why use them?	• You don't need to buy or maintain your own personal computer center. • You can complete more work and tackle more difficult problems. • You can share data with your distributed team.	• You don't need to buy or maintain your own personal computer center. • You can quickly access extra resources during peak work periods.
What are they useful for?	Grids were designed to handle large sets of limited duration jobs that produce or use huge quantities of data.	Clouds best support long-term services and longer-running jobs.

Clouds and Grids (5)

Universität Augsburg University

Benefits	• **Collaboration**: grids offer a federated platform for distributed and collaborative work. • **Ownership**: resource providers maintain ownership of the resources they contribute to a grid. • **Transparency**: the technologies used are open source, encouraging trust and transparency. • **Resilience**: grids are located at multiple sites, reducing the risk in case of failure at one site.	• **Flexibility**: users can quickly outsource peaks of activity without long-term commitment. • **Reliability**: provider has financial incentive to guarantee service availability (Amazon, for example, can provide user rebates if availability drops below 99.9%). • **Ease of use**: relatively quick and easy for non-expert users to get started.
Drawbacks	• **Reliability**: grids rely on distributed services maintained by distributed staff, often resulting in inconsistency in reliability. • **Complexity**: grids are complicated to build and to use, and users require some level of expertise.	• **Generality**: clouds do not offer many of the specific high-level services currently provided by grid technology. • **Security**: users with sensitive data may be reluctant to entrust it to external providers. • **Opacity**: the technologies used to guarantee reliability and safety of cloud operations are not made public. • **Rigidity**: the cloud is generally located at a single site, which increases risk of complete cloud failure. • **Oligopolism**: there is a risk of being locked in to services provided by a very small group of suppliers.

- Peer-to-Peer, Cloud, Grid and Organic Computing:
  - The basic concepts have many similarities and connections.
  - In some cases even similar technical concepts (e.g. P2P and Grid).
  - In all cases, the distributed and interconnected character plays a major role.
  - Although Cloud Computing is the current hype, all four communities gain high interest in academia and industry.
  - From these four concepts, Cloud Computing is the most industry-driven approach and closest to market maturity!

  - Current research tries to combine the advantages of all four disciplines and communities: a distributed P2P-like computing system, accessible like a Cloud, reliable like a Grid and autonomic (or self-organised) like organic systems.

- Further “organic” activities during the upcoming term
  - Seminar “Selbstorganisation in Verteilten Systemen” (Bachelor)
  - Lecture “Organic Computing II” (Master)
  - Lecture „Ad-hoc and Sensor Networks“
- Other topics
  - Praxis-, Forschungs- and Projektmodul
  - Theses (Master, Bachelor, Diploma, …)
  - “HiWi” jobs
  - PhD positions
- Interested? Just ask…

- Grid vs. Cloud
  - EGEE Comparative Study: Grids and Clouds – Evolution or Revolution?
  - e-IRG White Paper 2008 (Grid vs. Cloud Computing Chapter)
  - OGF-Europe Cloudscape Workshop (several Grid vs. Cloud talks)
- Cloud status
  - Above the Clouds: A Berkeley View of Cloud Computing
  - Let it Rise: Special Report on Corporate IT, The Economist
  - The Open Cloud Manifesto