Software Design
Grid Computing

Hasso-Plattner-Institut Potsdam
Software Architecture Group
http://www.hpi.uni-potsdam.de/swa
Marko Röder
WS 2011/2012

Motivation

- The huge number of PCs in the world (> 1 billion) [Gartner]
  + the ever-growing number of other computing devices
- Supply computing power to science
  ➥ What does your computer do at night?
- Enable scientific research that could not be done otherwise
- In contrast: supercomputers
  - Extremely expensive
  - Available only for applications that can afford them

➡ (Desktop) Grid Computing / Volunteer Computing

What is Grid Computing?

- Term is taken from grids: power *grid*, water system
  ➥ Standard, reliable, and low cost access to associated transmission and distribution technologies

- **Vision:** a comparable network for computer systems
  - Computing power out of the wall socket ➡ Transparency
  - Ubiquitous access to computational resources
  - Cheap and widely available computing power

- Never got to work as expected -> buzzword/marketing

*"A computational grid is a **hardware** and **software** infrastucture that provides dependable, consistent, pervasive, and inexpensive **access to high-end computational capabilities**"*

-- Ian Foster, Carl Kesselman

Types of Grids

Computing grid

- Accumulates computing power of different computers
- Supplies this power to problems that need a lot of (CPU) time
- Applications: weather prediction

➥ left to ***Grid Computing***

Database grid

- Unified access large datasets (that cannot be handle on one machine)
- Queries get distributed and "the grid" collects answers
- Applications: single distributed or federated database systems

➥ Web services / SOA

Resource grid

- Provide resources for temporary or permanent usage
- Applications: distribute storage

➥ Cloud Computing

[1][5]

Use Casses for Grid Computing

Parameter Tests

- Scientific simulation of complex systems (e.g.physics, chemistry)
  ➥ Intensive calculations
- Example: Computer-Aided Drug Discovery

Map/Reduce

- (Storage and) Processing of large data sets
  ➥ Storing huge quantities of data and executing calculations
- Examples: Medical images, LHC

[5]

Basic requirement

Application is divisible into a large number of independent jobs

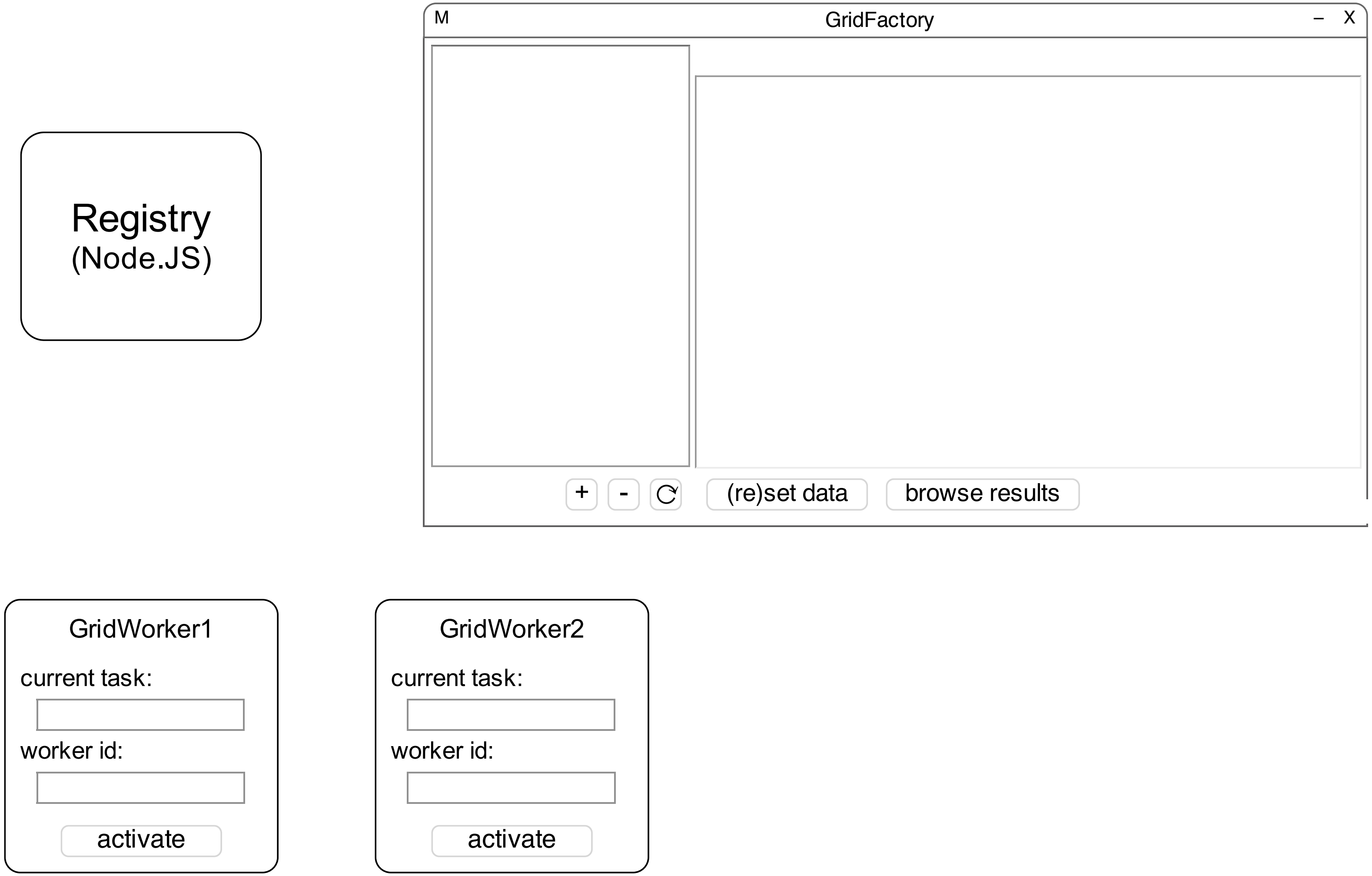
➥ Abstract execution model of parallel applications

How does it work? What do we need?

1. A **work generator** (factory) creates the job*
2. **Input (files)** is associated
3. Multiple instances (**tasks/work units**) of the job are created
4. **Server** (registry) dispatches the instances to different hosts
5. Each **worker** downloads its input files
6. The worker executes the job
7. The worker reports the job as completed & uploads **results**
8. A **validator*** checks the output files
9. An **assimilator*** handles the results

* A program the user should supply

Simple setup



Registry
(Node.JS)

GridFactory

GridWorker1

current task:

worker id:

GridWorker2

current task:

worker id:

Steps 1-3: Create a job, define input & tasks

Examples:

Square

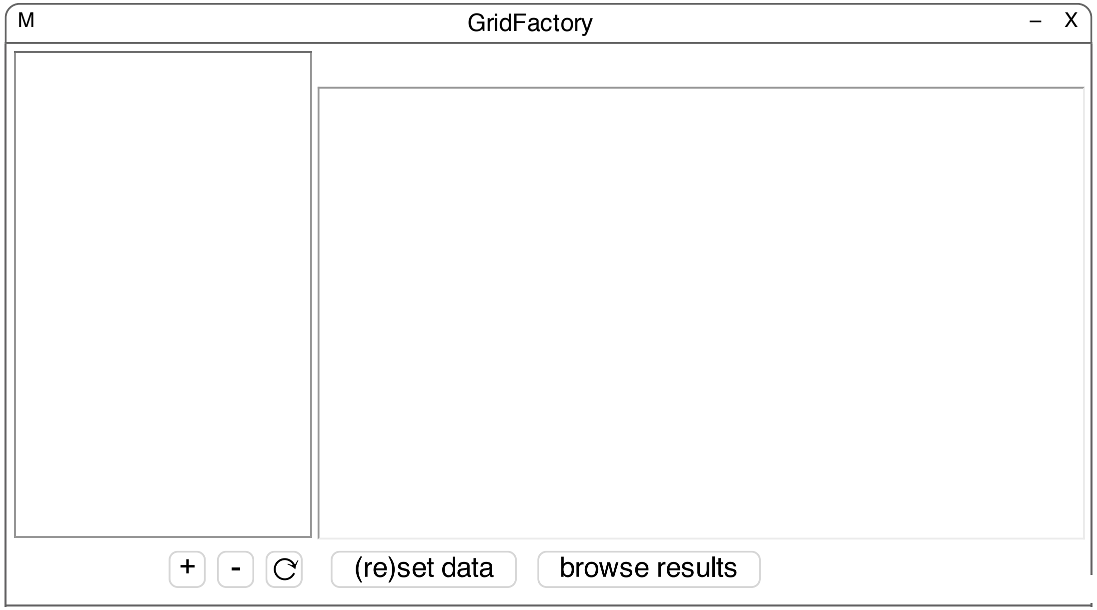
```
function square(a) {
    return a * a;
}
```

Primes

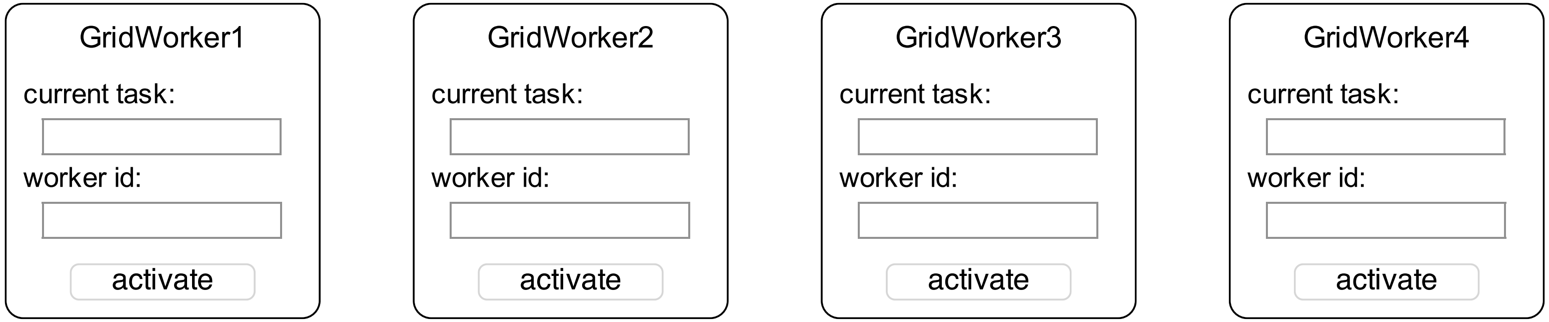
```
function isPrime(n) {
    if (isNaN(n) || !isFinite(n) || n%1 || n<2)
        return false;
    var m = Math.sqrt(n);
    for (var i = 2; i <= m; i++)
        if (n % i == 0)
            return false;
    return true;
}
```

Zero of a function

```
function zero(x) {
    return (Math.pow(x, 4) + Math.pow(x, 3) -
        4*Math.pow(x, 2) - 4*x) == 0;
}
```



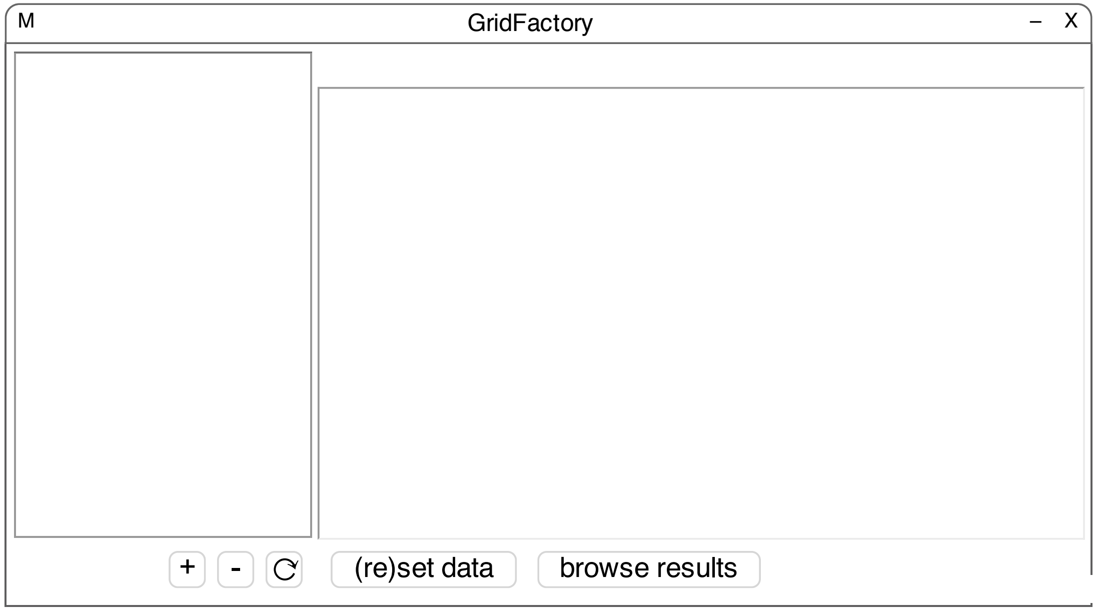
Steps 4-7: Dispatch, download, work & return



Steps 8-9: Validate & consolidate

Validate: Do it again! (one step back, ...)

Consolidate:



Examples

Great Internet Mersenne Prime Search (http://www.mersenne.org/)

- Formed in January 1996 (first volunteering project)
- Finding world record primes
- Mersenne primes are primes of the form 2^p-1 (only 46 known)

BOINC (http://boinc.berkeley.edu/)

- General-purpose grid computing solution
- For scientists (create a project), universities (virtual campus supercomputing), companies (desktop grid computing)
- Example: World Community Grid

Globus Toolkit (http://www.globus.org/)

- Standard (reference implementation) started in 1995
- Middleware technology to support and ease Grid Computing
- Aimed at high performance scientific computing

Others: distributed.net, SETI, Worldwide LHC Computing Grid

Concerns & Criticism

Projects perspective

- How can (all) challenges be described Grid Computing tasks?
  ➥ Many, many jobs and an abstract problem
- What is if a volunteer misbehaves in some way?
  ➥ Volunteers are anonymous and therefore not accountable

Volunteers perspective

- Can those task do damage my computer or invade my privacy?
  ➥ Security considerations, sandboxing, ...
- How is my work being used? (Truthfulness / Intellectual property)
  ➥ Can hackers use it as a vehicle for malicious activities?

References

[1] DUNKEL, J. et al.: Systemarchitekturen für verteilte Anwendungen. Hanser Verlag, First edition, 2008.

[2] FOSTER, I., KESSELMAN, C.: The Grid: Blueprint for a New Computing Infrastucture. Morgan Kaufmann, First edition, 1998.

[3] FOSTER, I., et al.: The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration. OGSI WG Global Grid Forum 22, 2002.

[4] PRODAN, R., FAHRINGER, T.: Grid Computing: Experiment Management, Tool Integration, and Scientific Workflows. LNCS Volume 4340, 2007.

[5] FERREIRA L., et al.: Grid Computing in Research and Education. IBM Redbooks, First edition, April 2005.

Web Resources

http://www.gartner.com/it/page.jsp?id=703807

http://boinc.berkeley.edu/

http://www.gridforum.org/

http://www.mersenne.org/

http://www.globus.org/ogsa

Bonus: Related technologies

Parallel Computing / Clusters

- Consumes more than a gigabyte of data per day of CPU time
- Large data, expensive and/or slow to send over internet

Service-oriented Architectures (SOA) / **Web Services**

- Aggregation of portable an reusable programs called services
- Can be accessed by remote clients over network
- Language and platform independent

Peer-to-peer Architectures (P2P)

- Aggregation of equivalent programs called *peers*
- Provide functionality and share part of their hardware resources without the involvement of a central server
- Benefits the participants, no notion of a 'project'
- High degree of scalability and fault tolerance

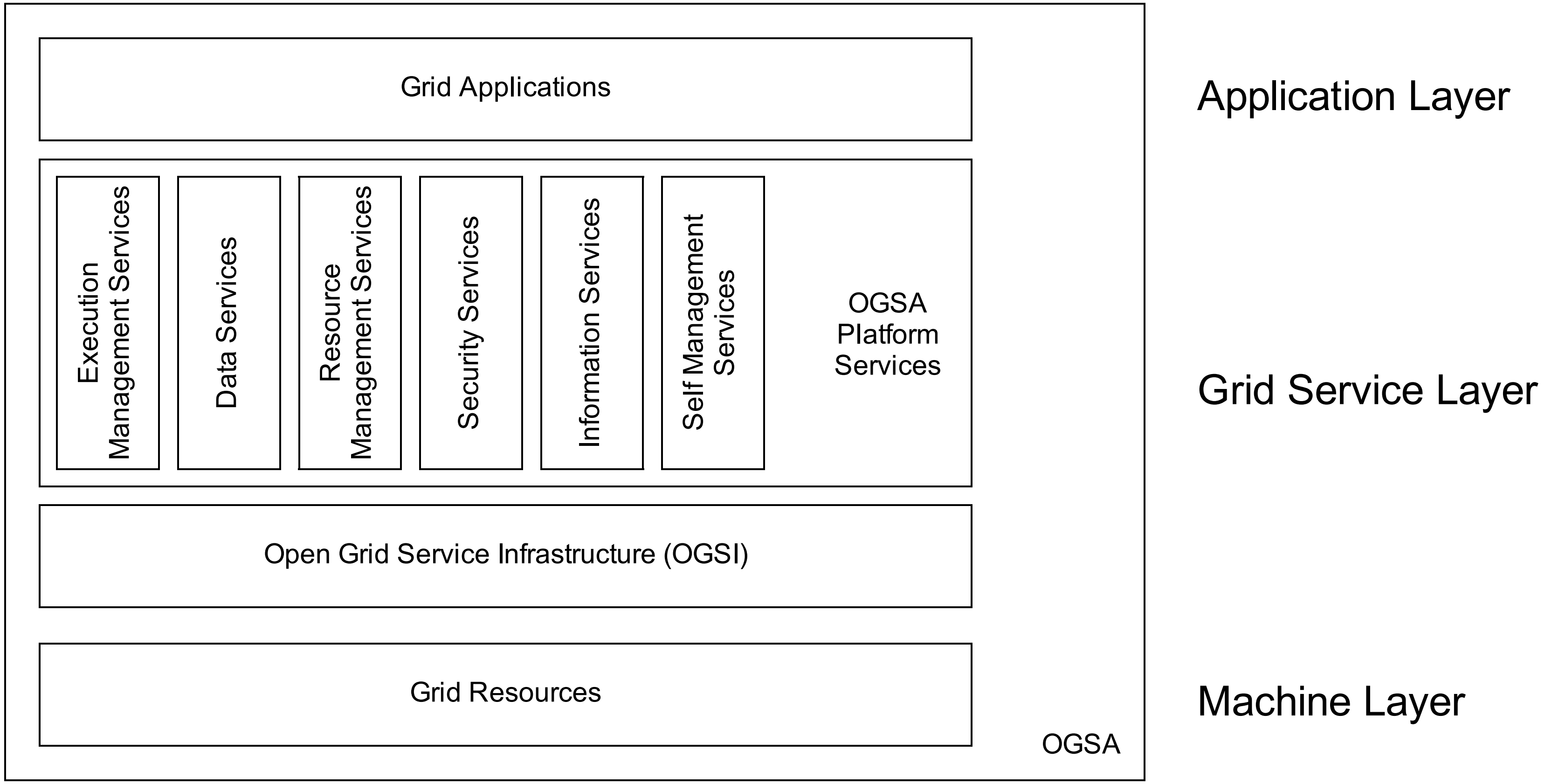
Cloud Computing

Bonus: Open Grid Service Architecture (OGSA)

Open Grid Services Architecture

- Official specification of the Global Grid Form (GGF)
  - Current version: 1.5 (from September 5, 2006)
  - http://www.globus.org/ogsa
- First prototype Grid service implementation January 29, 2002
- Globus Toolkit 3.0 and 3.2 offered an OGSA implementation
- Globus Toolkit 4.0 provides a OGSA capabilities based on WSRF
  - Open source, community-driven software project

Bonus: Components of the OGSA platform


Grid Applications
Execution Management Services
Data Services
Resource Management Services
Security Services
Information Services
Self Management Services
OGSA Platform Services
Open Grid Service Infrastructure (OGSI)
Grid Resources
OGSA
Application Layer
Grid Service Layer
Machine Layer