NORDUGRID
Grid Solution for Wide Area Computing and Data Handling

CSC

NorduGrid Tutorial

Linux Clusters for Super Computing Conference
Linköping, Sweden
October 18, 2004

Arto Teräs <arto.teras@csc.fi>
Juha Lento <juha.lento@csc.fi>

Tutorial Contents

- Part 1: Introduction to Grids and NorduGrid
  - Moving towards Grid: what changes?
  - NorduGrid collaboration and middleware
  - Authentication and authorization in Grids
  - Grid environment compared to local computational resources
  - Submitting jobs in NorduGrid, writing job description files, demo
- Part 2: Hands-on Exercises
  - Installing NorduGrid client software
  - Submitting jobs
  - Monitoring jobs using the Grid Monitor graphical interface

What is Grid?

- Hot term: many users, many definitions
- Uniform and secure access to geographically distributed heterogeneous systems
- Both the set of users and connected resources vary dynamically
- Grids go across multiple administrative domains!
- Collaboration and social networking, not only technology

Moving Towards Grid — What Changes?

- Importance of physical location of computers and data diminishes
- User accounts are replaced by certificates, number of passwords reduces
- Uniform interfaces to heterogeneous platforms, but library dependencies and other OS related problems remain
  - Virtual machines may bring improvements in the future
  - Error reporting and debugging currently rudimentary, will get better as middleware packages mature
- Users can still specify a target system if they need, but apps which work flexibly in different environments have an edge

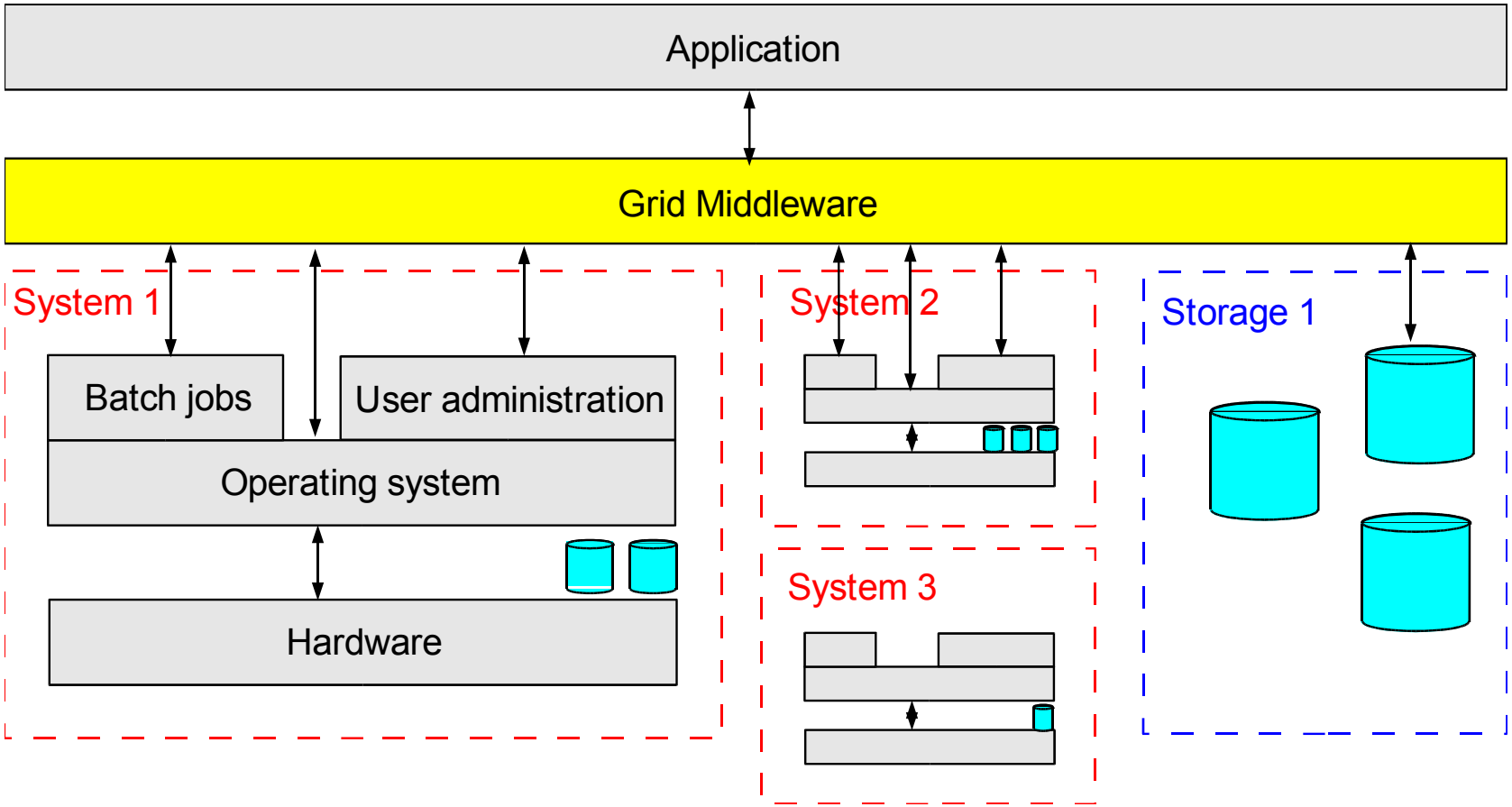
Common misconceptions

- Grid multiplies available resources
  - Common comparison with the World Wide Web doesn't work:
    - One web server serves 1000 users => price 0.001 x
    - One grid user wants to use 1000 servers => price 1000 x
  - Load balancing can bring some savings, but new services and easy access much more important
- Grid is a black box: everything inside happens automatically
  - Vision: computing power as electricity from the plug
    Reality: still quite far from it
  - Possibility to monitor job execution is important — trying to make a black box results easily in a black hole

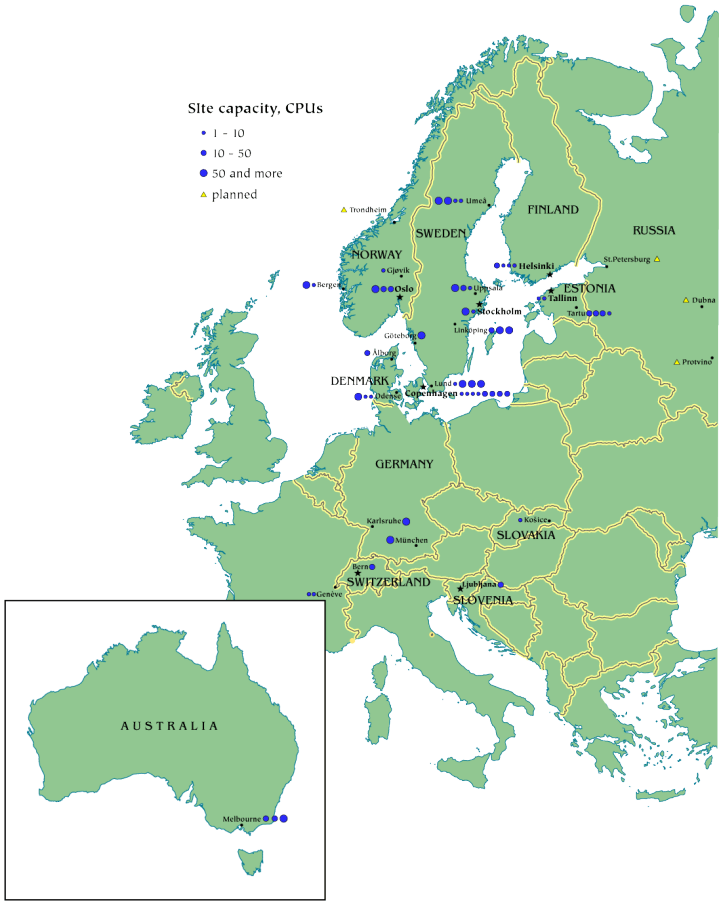
Does one need to change existing applications?

- Three different approaches:
  1) Using the application as is: grid middleware will move it and data to target system
     - Linking statically helps in library dependency hell...
  2) Installing the application on the target system and using it via the Grid interface
     - Batch processing type applications normally work without changes, interactive apps and graphical interfaces more difficult
  3) Modifying the application to fully exploit a distributed environment
     - Distributing over a large geographical area is not practical unless the application can split a computation to independent subtasks

Role of Middleware



NorduGrid collaboration



- Started in 2001, originally for connecting resources in Scandinavia and Finland
- Has been later joined by groups in Estonia, Germany, Slovakia, Slovenia, Germany and Australia
  - Currently about 5000 CPUs total (mostly Linux clusters)
- NorduGrid ARC middleware
- Open for anyone to participate

Facts on NorduGrid

- Academic Grid, serves researchers and consists of academic resources
  - Largest user group is high energy physicists, but other applications are also constantly being run on the Grid
- Built from ”bottom to top”, connecting already existing resources to the Grid
  - Resources NOT dedicated to the Grid
- Production Grid available 24/7 since July 2002
  - Real users and applications, also a good platform for Grid research
- Dynamic: resources come and go

NorduGrid Middleware

- Called Advanced Resource Connector (ARC)
- Built upon Globus Toolkit 2 libraries, currently being ported to GT version 3.2.1
  - In particular, the security infrastructure comes from Globus
- However, many important differences to core Globus:
  - GRAM (Globus gatekeeper and job manager) replaced by NorduGrid ARC Grid Manager
  - Own gridftpd server
  - Own user interface and broker
  - Extended job description language and information model
- GPL license

What does it look like?

2004-10-17 CEST 01:14:08

Processes: Grid Local

Country	Site	CPUs	Load (processes: Grid+local)	Queueing
Australia	Atlas (UniMelb)	0		0+0
	Charm (UniMelb)	37	0+0	0+0
	Alfred (UniMelb)	90	0+8	18+1
Denmark	DistLab (DIKU)	10	0+0	0+0
	Aalborg Grid Gateway	50	32+0	0+0
	Niflheim (DCSC/DTU)	881	0+827	0+0
	Horseshoe (DCSC/SDU)	1199	0+884	0+182
	HEPAX1	1	0+0	0+0

- Grid Monitor shows currently connected resources
  - Take a look yourself at http://www.nordugrid.org

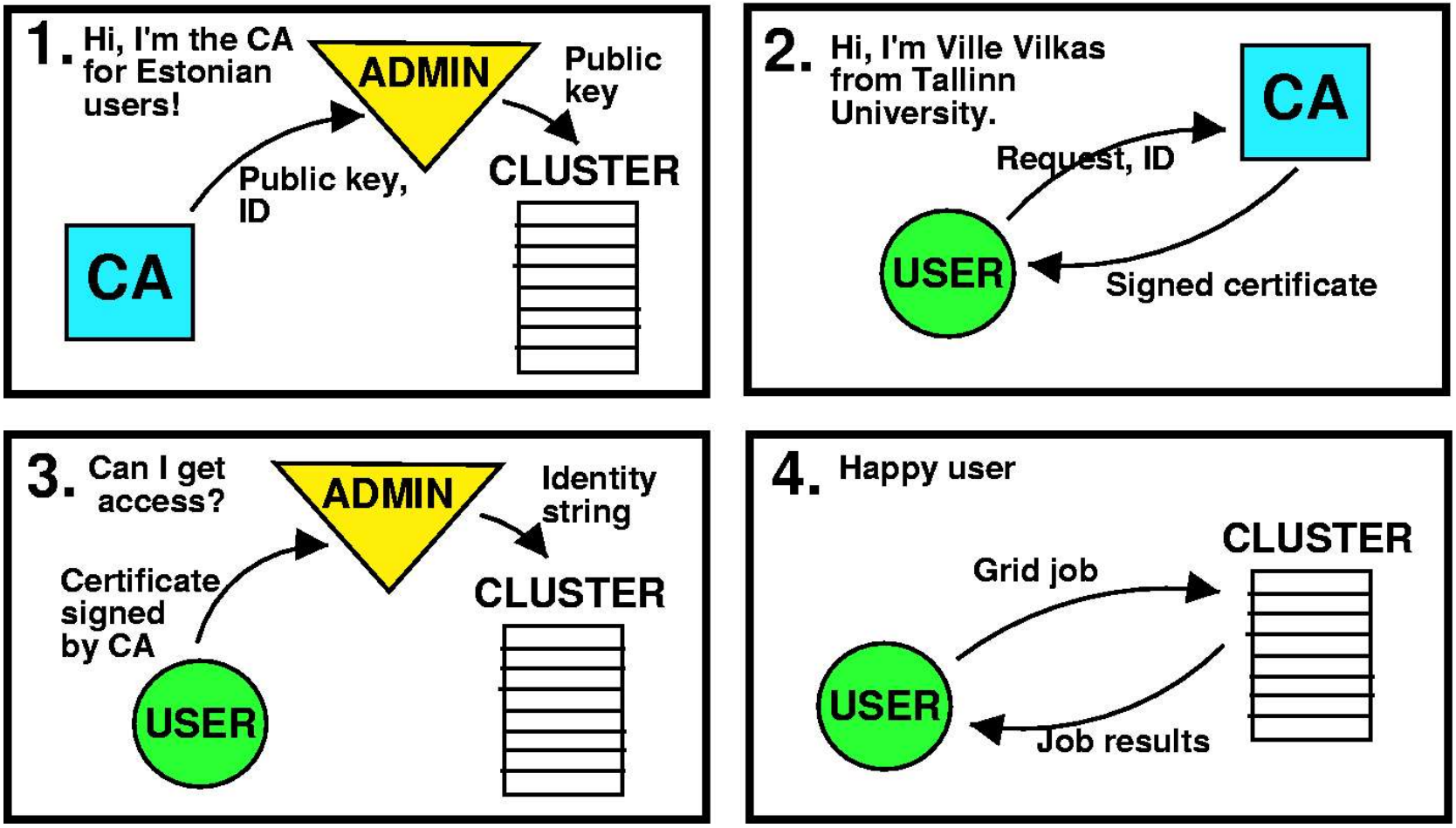
Grid Security

- ”It's like when the PC came...” (Urpo Kaila, manager, information security, CSC)
  - Grid account is a pass to computers beyond organizational domains!
  - New risks, IT staff often afraid
  - Great power => great damage
- On the other hand: development started when viruses and worms were already around — security aspects have been considered
  - Strong authentication and encryption: no plain-text passwords
  - Identity tied to a certificate: revocation blocks access in the whole Grid

Certificates

- User authentication is based on X.509 certificates in almost all Grid middleware, including NorduGrid
- Certificate represents user's identity in the Grid
  - Each user has his/her own personal certificate
  - Based on public key encryption, tried and tested technology
  - Does not automatically give access to resources
- Trusted third party called Certificate Authority (CA)
  - Gives his guarantee of the identity of the user by signing the certificate
  - Maintains a list of revoked certificates

Certificate Trust Chain



Virtual Organizations (VO)

- Lists of user identities in the Grid
  - Allow to manage users as groups
- Typically, access to a resource is granted to a virtual organization, meaning that all members of that VO can use the resource
  - Similar to a group visa in physical world
  - Implies trust towards the organization managing the VO
- Implemented as simple text files, LDAP servers or databases specially designed for VO management

Questions?

- Next:
  - Grid environment compared to local computational resources (Juha Lento)
  - Submitting jobs in NorduGrid, writing job description files (Arto Teräs)
  - Demo (Juha Lento)