The gLite middleware

Ariel Garcia
KIT

Overview

- Background
- The gLite subsystems overview
    - Security
    - Information system
    - Job management
    - Data management
- Some (my) answers to your questions and random rumblings

KIT – The cooperation of
Forschungszentrum Karlsruhe GmbH
and Universität Karlsruhe (TH)

SCC
Steinbuch Centre
for Computing

Ariel Garcia - gLite middleware

ESRF - 8 Dec. 2008

2/24

Intro

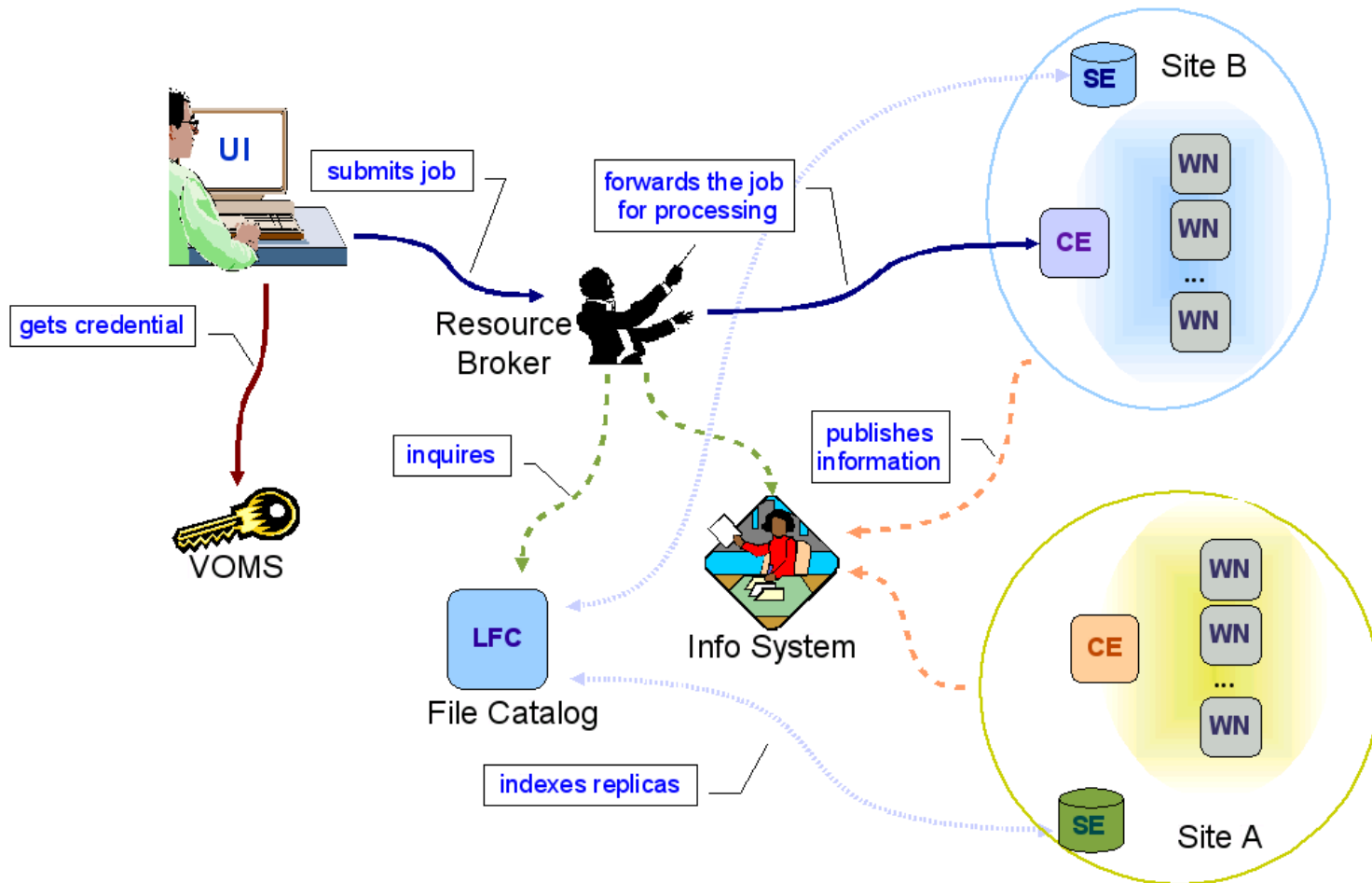- **KIT = FZK + Univ. of Karlsruhe**

About me

- 2002 – CrossGrid, interactivity and parallel jobs in LCG/gLite
- 2005 – GermanGrid, gLite middleware operation and support
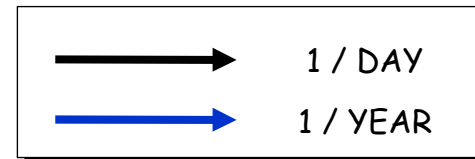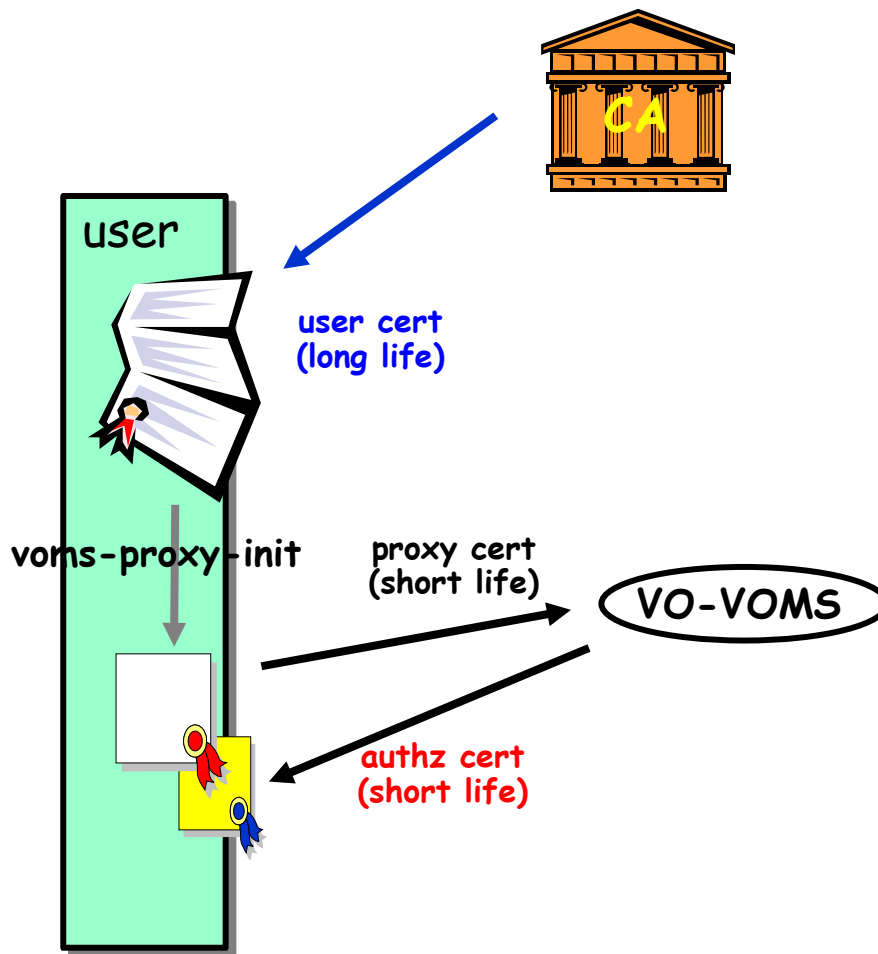- 2007 – gEclipse, GUI and toolkit for grid/cloud access

"The" Grid??

- **No such thing!**

- **(Too) many Middlewares! ("NIH" syndrome?)**
  - Globus Toolkit 4 (US)
  - gLite (EU)
  - Unicore (Germany)
  - ARC (NorduGrid)
  - GRIA (EU, UK, entreprise oriented)
  - ProActive (France, Java programming paradigm)
  - MyGrid (UK)
  - OurGrid (Brasil)
  - GOS (China)
  - Naregi (Japan)
  - ...

} D-Grid = German Grid

KIT – The cooperation of
Forschungszentrum Karlsruhe GmbH
and Universität Karlsruhe (TH)

SCC
Steinbuch Centre
for Computing

Ariel Garcia - gLite middleware

ESRF - 8 Dec. 2008   **4/24**

A bird's view of a gLite grid

KIT – The cooperation of
Forschungszentrum Karlsruhe GmbH
and Universität Karlsruhe (TH)

SCC
Steinbuch Centre
for Computing

Ariel Garcia - gLite middleware

ESRF - 8 Dec. 2008   **5/24**

How do I get authorized?

- 1 / DAY
- 1 / YEAR

- **X509 certificates**
  - Authentication (who are you?)
- **VO Membership Service**
  - basis for authorization
    - user <=> VO
  - Globus certificate
    - X509 extensions
    - signed by VOMS server
  - single sign on

How is access granted?

VOMS

- multiple VOs per user
- (sub)groups in VOs for organizing members
  - hierarchical structure
  - administration can be delegated (ACLs)
- finer grained authorization, multiple roles in VO
  - members can play roles

    /hep/Role=swadmin

    /hep/subgroup

    /hep/subgroup/Role=production

- Up to services to **enforce policies**
  - allows for local policy
  - resource providers grant access to members of VOs / groups / roles
  - resources map VO / group / role members to local accounts

KIT – The cooperation of
Forschungszentrum Karlsruhe GmbH
and Universität Karlsruhe (TH)

Ariel Garcia - gLite middleware

ESRF - 8 Dec. 2008

7/24

VOMS?

- **Pros**
  - enables the VO concept (Grid!)
  - spreading
    - available in ARC
    - some support in GT4, Unicore 6
    - (mandatory in gLite)
 - reasonable compromise between SSO and security

- **Cons**
  - SAML?
  - might not be secure enough
    - proxies traveling around

Which resources where?

Information system

- Aggregates service information from multiple grid sites
  - hosts
  - accepted VOs
  - resource availability and usage (CPU, storage)
- GLUE schema 1.3, LDAP
- Used by Workload Management (RB's) to know about sites
  - defines WMS's view of the Grid!
- Used by replica management tools to locate Storage, Catalogs
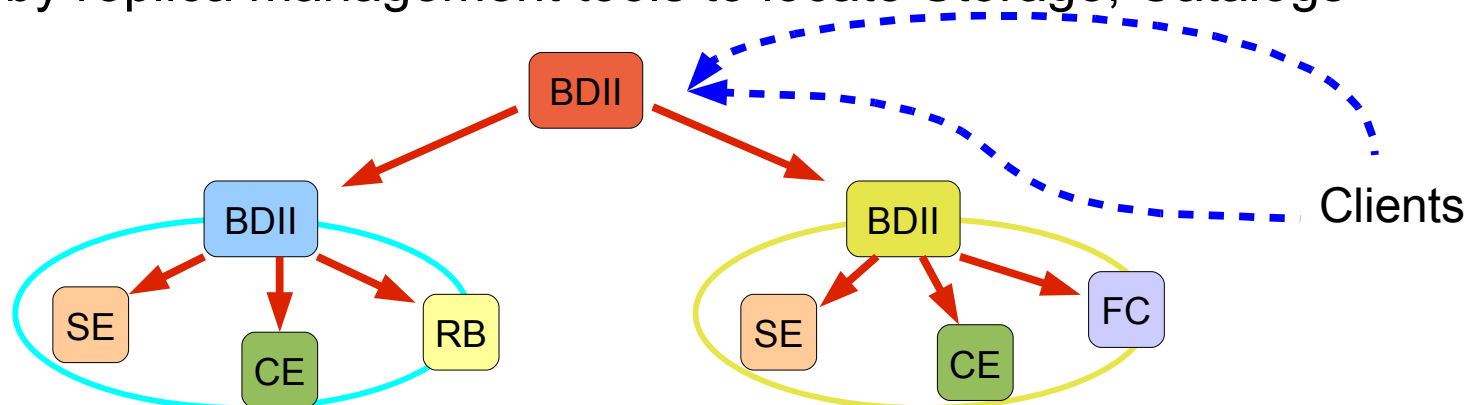
How are my jobs handled?
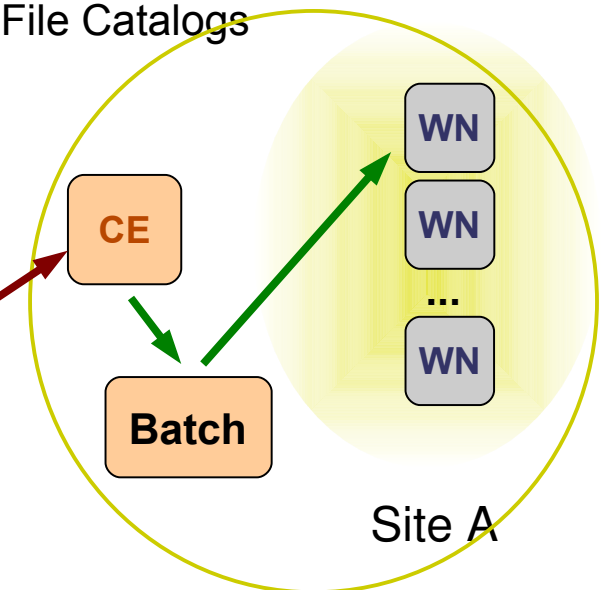
- **Computing Element**
  - gateway to local batch system & cluster
  - based on Globus 2 gatekeeper
- **Resource Broker (WMS)**
  - is primary job execution interface for users
  - find "best" location for a job
    - considering job requirements and available resources (CPUs, files)
    - gets resource information from InfoSystem and File Catalogs
  - L&B keeps track of job's status
  - supports
    - workflows (DAGs)
    - MPI, interactive IO
    - parametric jobs
    - job collections submission

UI

RB

CE

Batch

WN

WN

...

WN

Site A

KIT – The cooperation of
Forschungszentrum Karlsruhe GmbH
and Universität Karlsruhe (TH)

SCC
Steinbuch Centre
for Computing

Ariel Garcia - gLite middleware

ESRF - 8 Dec. 2008   **10/24**

How do I define a job?

- **JDL Job description**
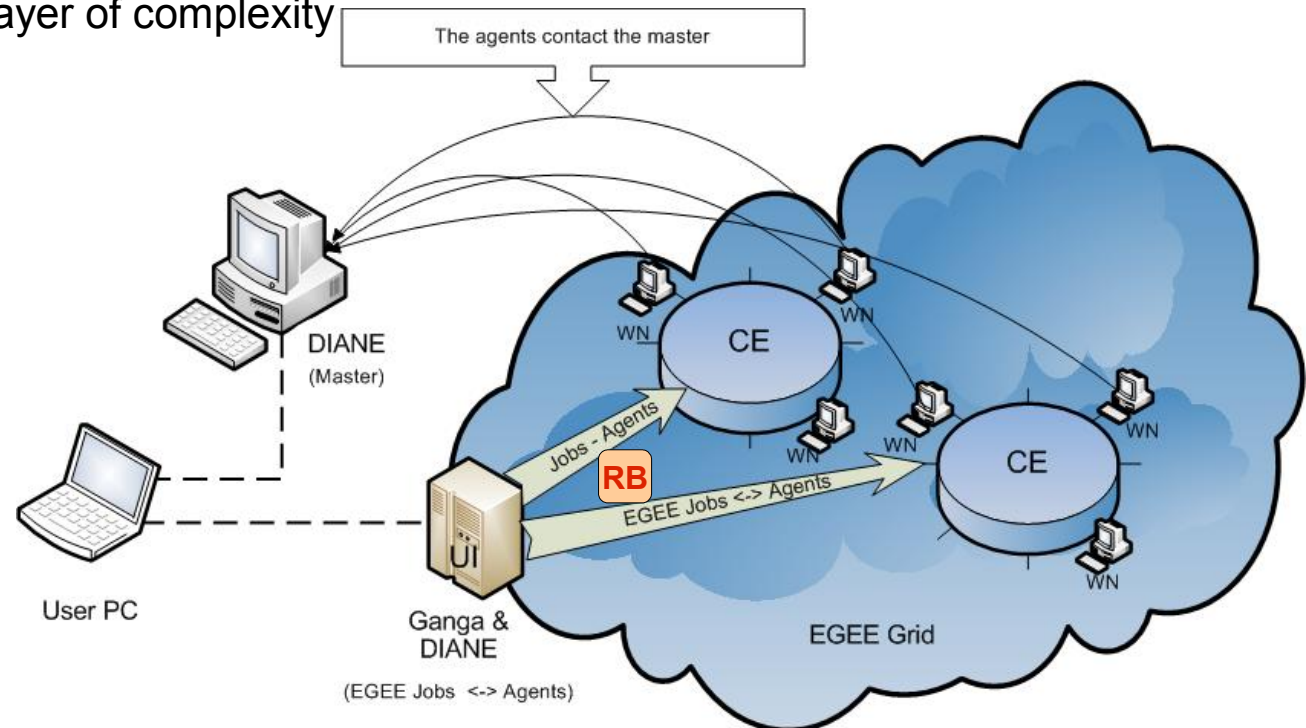
```
[
        Type = "Job";
        JobType = "Normal";
        RetryCount = 3;
        Executable ="myApplication";
        StdOutput = "output.txt";
        InputSandbox = { "myApplication", "input.txt" };
        OutputSandbox = { "output.txt" };
        OutputSandboxDestURI =
            { "gsiftp://se.domain.org:2811/dpm/domain.org/home/myVO/output.txt" };
        Requirements = other.GlueCEPolicyMaxCPUTime > 480;
        Rank = -other.GlueCEStateEstimatedResponseTime;
]
```

Client SW on WNs

Info system

KIT – The cooperation of
Forschungszentrum Karlsruhe GmbH
and Universität Karlsruhe (TH)

SCC
Steinbuch Centre
for Computing

Ariel Garcia - gLite middleware   ESRF - 8 Dec. 2008   11/24

Want higher job throughput?

- Pilot jobs paradigm
  - Seti @HOME idea
  - Partly bypasses MW
    - OK for many finer grained jobs
    - But extra layer of complexity
  - e.g., Diane

The agents contact the master

DIANE (Master)

User PC

Ganga & DIANE
(EGEE Jobs <-> Agents)

Jobs - Agents

RB

EGEE Jobs <-> Agents

CE

CE

WN

EGEE Grid

What about the cluster?

- gLite delegates file operations to WNs (scalability)
    - client SW needed on WNs
    - users came to expect client tools
- Supported options
    - **Linux**, 32 or 64 bits
    - RPMs for SciLinux 4
    - tarball (SciLinux 5, SuSE, Debian,... *YMMV!*)
- Supported batch systems
    - Torque / PBS
    - LSF
    - SGE
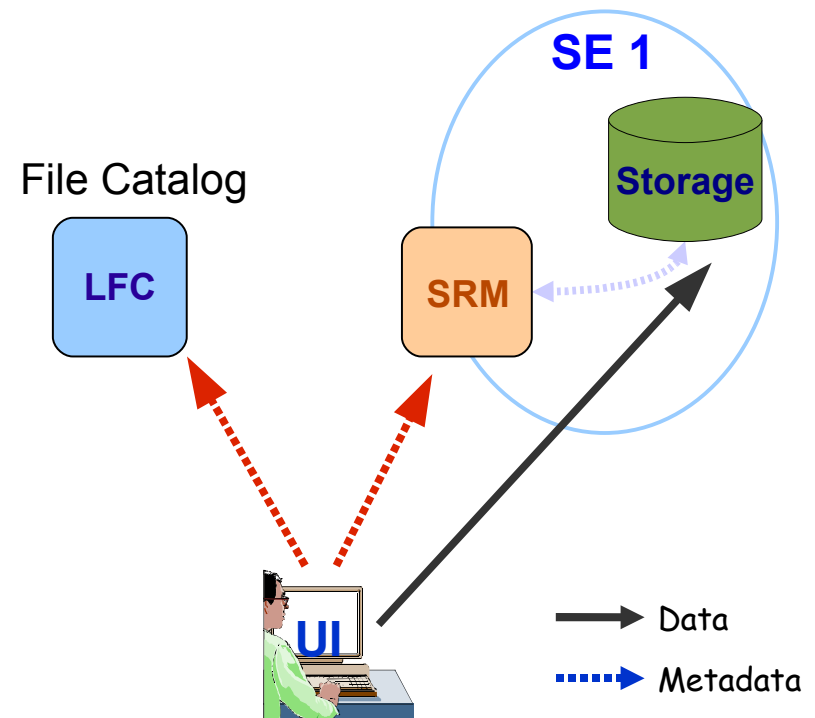   - Condor
- Networking
    - can be private, NAT'ed

Should I...?

- Do you need different archs?
- How long do your jobs run?

- What about MPI, workflows, interactivity?
- Do you want to offer and consume (application) services?

- How are your computing requirements (im)balanced?

SCC
Steinbuch Centre
for Computing

Ariel Garcia - gLite middleware

Where is my data stored?

- Storage Elements
  - gateway to local storage
    - disk
    - tape
  - SRM interface for metadata
    - WS "standard"
  - data transfer handled separately
    - PFNs point to actual storage location and access protocol
  - access protocols
   - gridFTP
    - rfio
    - gsidcap
    - ...

File Catalog

SE 1

Storage

LFC

SRM

UI

→ Data

····▶ Metadata

SCC
Steinbuch Centre
for Computing

How to find my data?

- **LCG File Catalog**
  - only deals with data locations
  - files can be replicated on multiple SEs
  - Grid Unique ID for each registered file
  - LFNs are names that make sense to you

- **File Transfer Service FTS**
  - scheduling of transfers
    - (RB for "data jobs")
  - management of inter-site "channels"
    - point-to-point (Tiers 0, 1, 2)
    - queues, bandwidth
 - transfers between SEs with same protocol
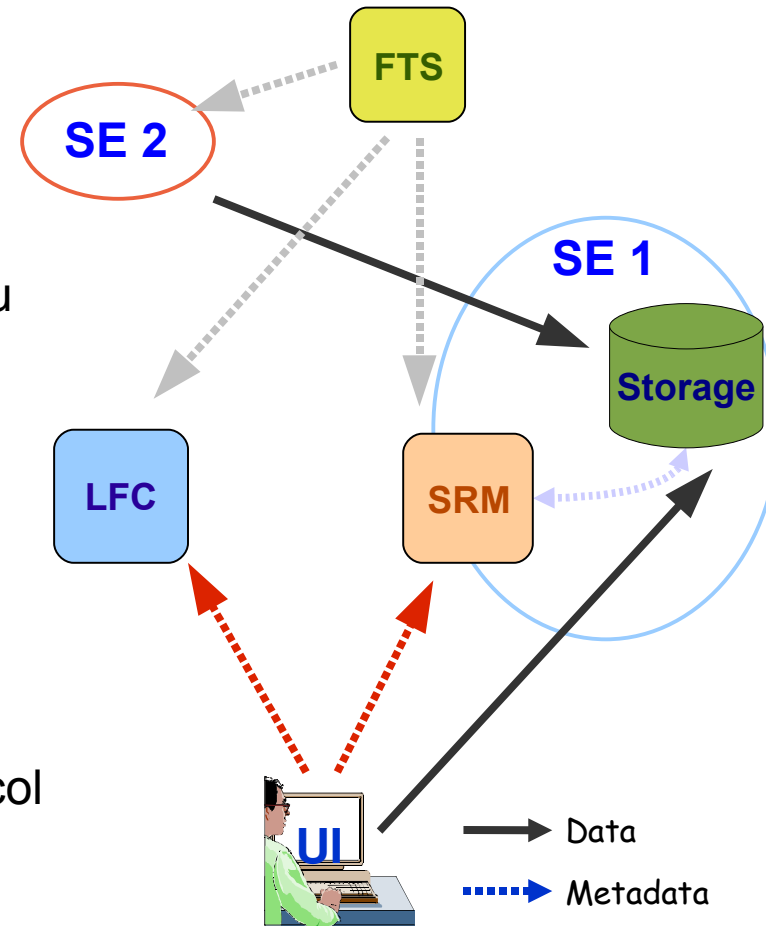
Other data components

- Amga Metadata Catalog
  - attributes (key-value pairs) on entries
  - hierarchy of entries
  - gridified DB (like OGSA-DAI)
    - X509/VOMS access, with ACLs
    - SQL-like queries
  - replication
  - WS + TCP-streaming access

- Hydra key store
  - data encryption
  - smart key management
    - splitted key, need M of N pieces to reconstruct ($1 < M < N$)

SCC
Steinbuch Centre
for Computing

Need to install your own resources?

- Admin overhead
  - avoid having many small sites!
  - or centralize! (I know... it's not "Grid" but... :-)

- Centralized Grid management (e.g. Grid Ireland)
  - site provides
    - batch system
    - user management
    - basic cluster OS and operation
    - storage
    - 2 hosts (Xen Host)
  - operation team
    - required poolaccounts
    - CE, SE installation
    - grid related operation
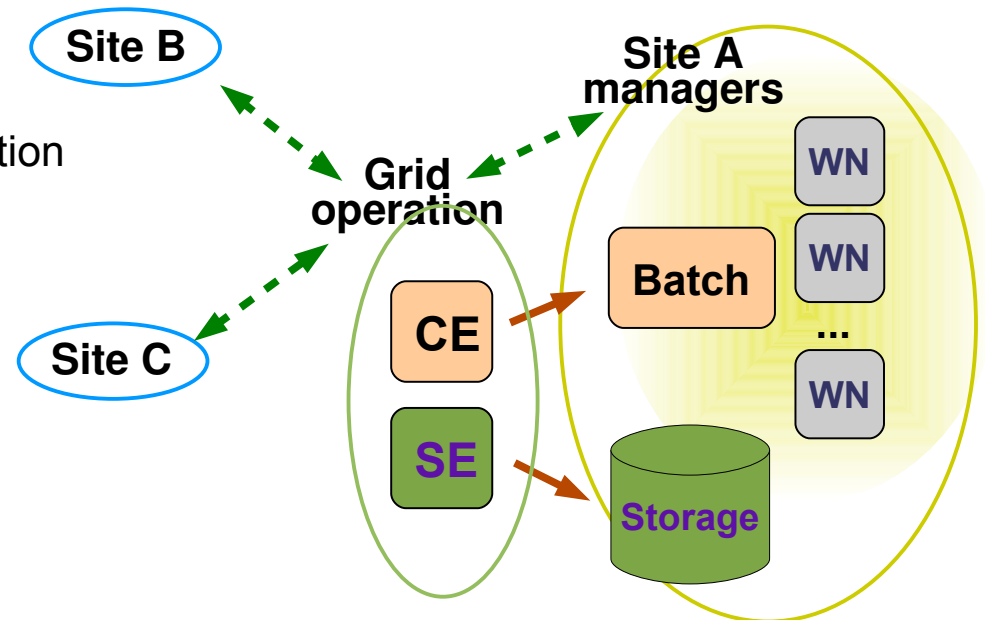
To Grid or not to Grid?

- **Computing perspective**
  - Underused or fully overcommitted resources?
    - => no Grid!
  - Unbalanced clusters?
    - => Grid
  - Somebody offers you resources? (special ones! supercomputer...)
    - => Grid - but determined by "provider"
  - Need to cope with high peak needs?
    - => go Cloud
- **Data perspective**
  - Across-borders collaboration wants to share data?
    - => Grid (security, GridFTP)
  - Want to unify access to data?
    - => Grid (file catalog)
- **Resource provider perspective**
  - Avoid dealing with single external users
    - => Grid (poolaccounts/VOMS in gLite)
    - but pay Grid MW overhead cost

Which grid middleware?

- Better: depends for what
- Good: probably none ;-)
- Evaluate: yes

- Behind FW
  - most will work fine, but some are more cumbersome
  - gLite → list of incoming (TCP) ports to open
- NAT:
  - no problem for the cluster
  - services need to be accessed!
- Networks:
  - no special needs
    - service hosts with multiple IPs/names → care with certificates
  - gLite: few public IPs for services with several ports open

KIT – The cooperation of
Forschungszentrum Karlsruhe GmbH
and Universität Karlsruhe (TH)

SCC
Steinbuch Centre
for Computing

Ariel Garcia - gLite middleware

ESRF - 8 Dec. 2008   **20/24**

Which grid middleware?

Take with tweezers!!

- Want VO concept?
  - gLite (VOMS, but not only)
- Big collaboration sharing data?
  - gLite (StoRM?)
  - or evaluate Globus + storage solution
- Heterogeneous platforms, no data grid?
  - go for Unicore
- Need API access, service oriented?
  - Globus (or use gEclipse :-)

- Need turnkey ready solution?
  - go for EGEE if you get the required resources
- Need to install your own resources?
  - delegate grid admin know-how! centralized solution (Grid Ireland)

gLite or not gLite?

- **Pros**
  - VO concept, out of the box
  - Out of the box functionality
    - brokering
    - file catalog, data management capability
    - SW management, some MPI support
    - ...
  - Single-shop Grid distribution
  - Available infrastructure (EGEE!)
  - Supported by gEclipse :-)

- **Cons**
  - Restricted set of platforms (cluster)
  - High admin overhead
  - Batch oriented (contraposition to: **services**)
    - needs pilot jobs framework for higher small jobs turn-around
  - Low standards compliance

Fazit...

- Don't believe in marketing!
- But also don't reinvent the wheel!

- Evaluate your needs
- Make sure what the different solutions offer to you
- Go for flexible solutions
- Think services
- Don't forget "cloud" (even if marketing buzzword)

- Don't believe in marketing ;-)

The end

?

Thanks for listening!