



Interactive European Grid

The int.eu.grid experience as an interoperable infrastructure between Portugal and Spain

**G. Borges, J. Gomes, M. Montecelo, M. David,
A. López, P. Orviz**

LIP

on behalf of Int.EU.Grid Collaboration



II IBERGRID Conference, Porto, Portugal, May 2008

MPI and Interactivity on the Grid

- VOs and users are anxious to get it
 - ▶ Needed by a wide set of applications in different scientific domains
- MPI has been neglected by larger Grid projects...
 - ▶ Aimed to sequential jobs...
 - ▶ How to properly set “Matchmaking” and “Brokering” for parallel and interactive tasks on a Grid Environment?
 - ▶ How to manage/control local cluster MPI support?
 - ▶ How to set central MPI support?

Int.EU.Grid

Provide an advanced **grid empowered infrastructure** for scientific computing targeted to support demanding **interactive** and **parallel applications**.

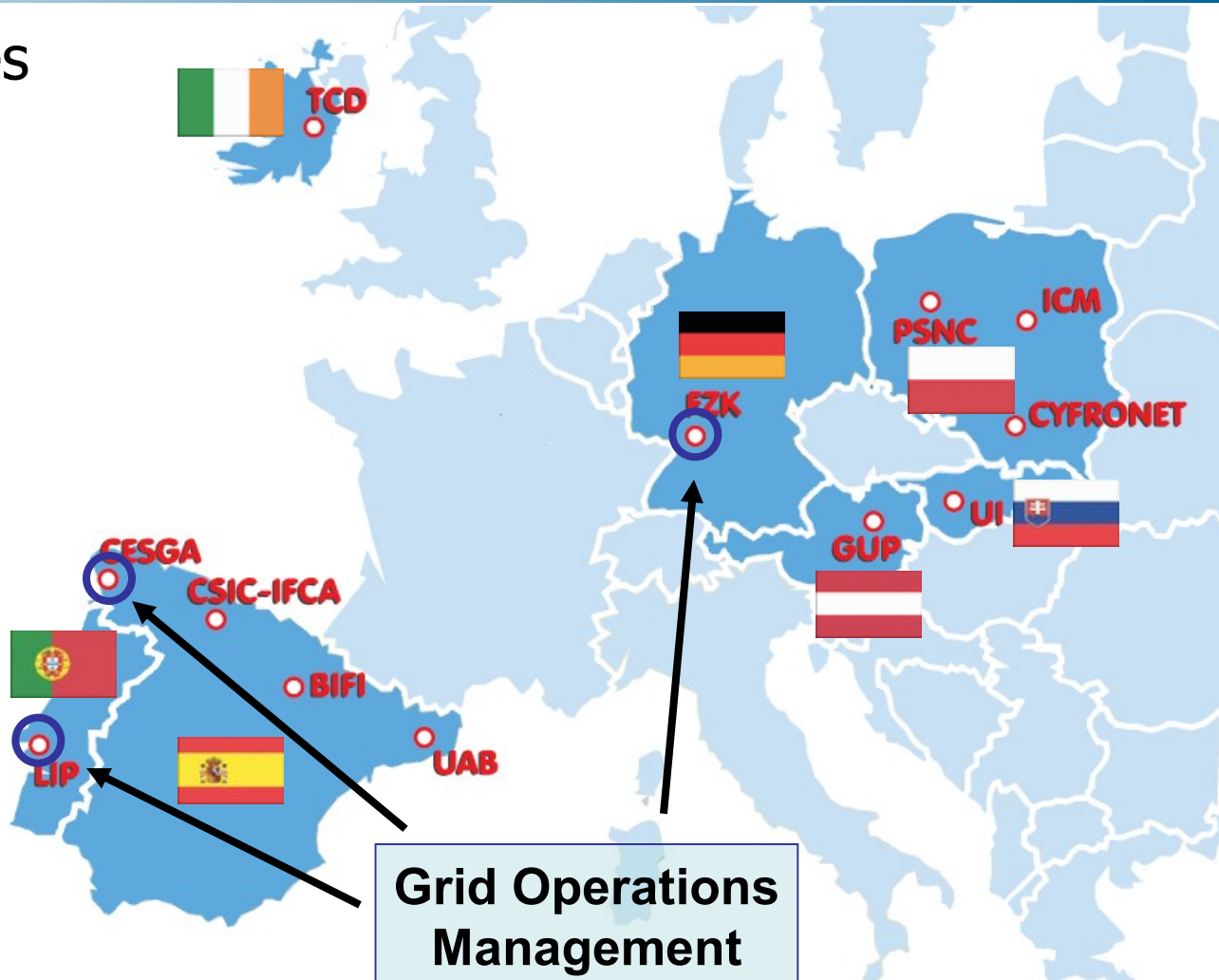
Int.EU.Grid grid infrastructure

- 12 sites, 7 countries
 - ▶ 9 in production
 - ▶ 4 in development

- ~ 900 CORES
 - ▶ Xeon
 - ▶ Opteron
 - ▶ Pentium

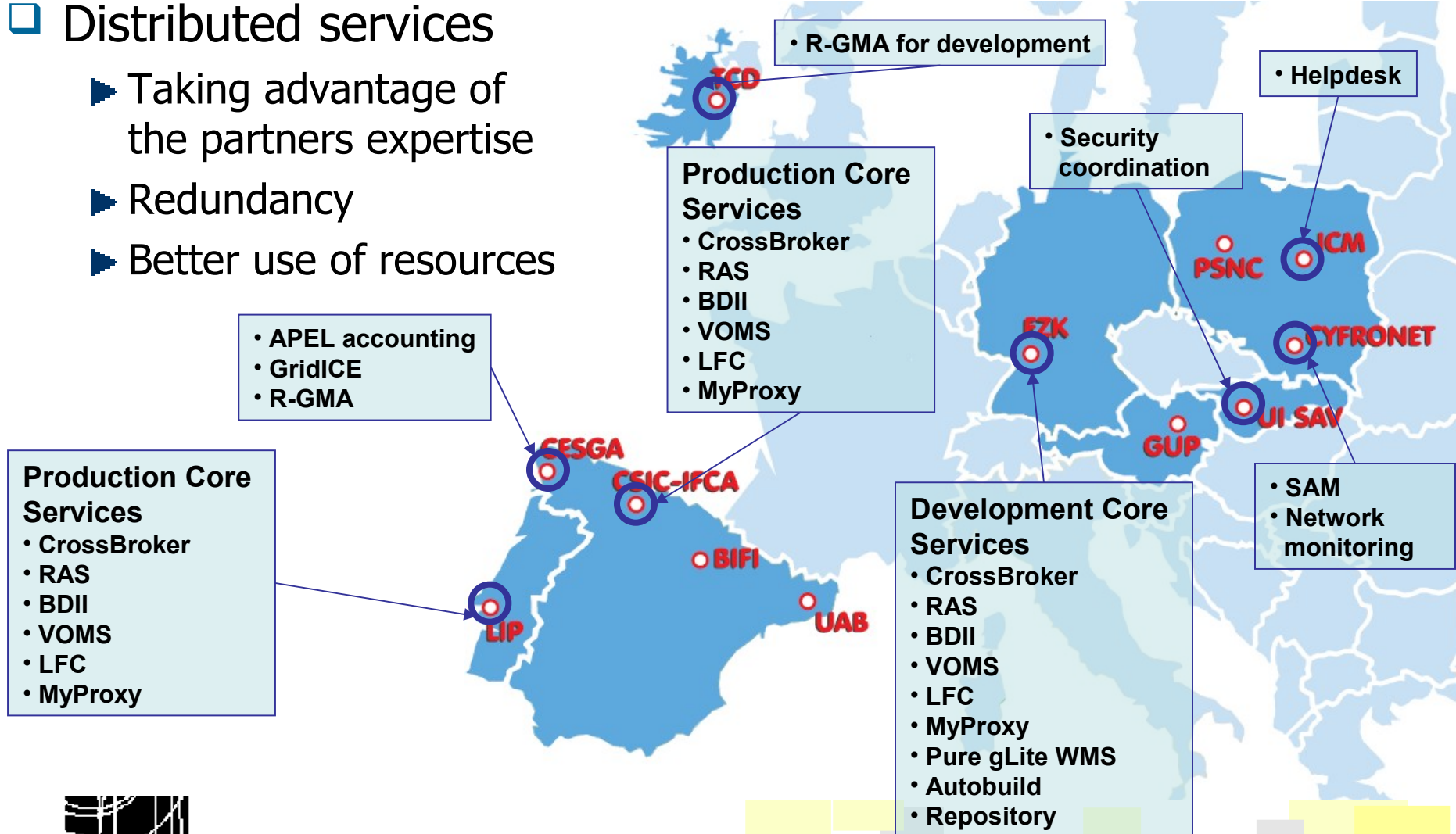
- ~ 45 TB of storage space

- Interconnection by Geant



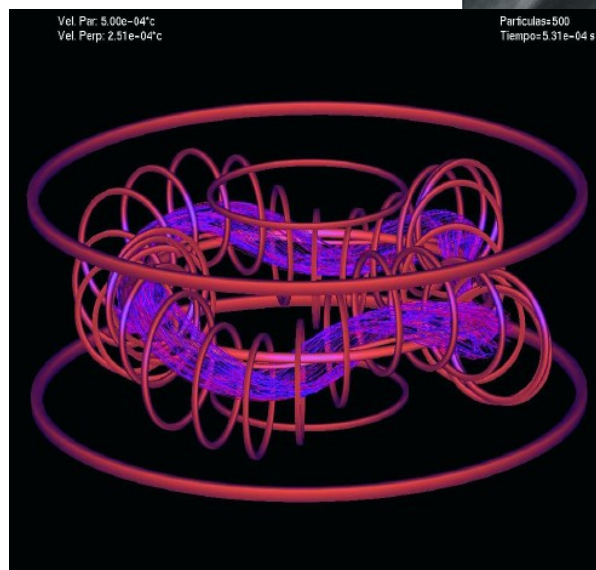
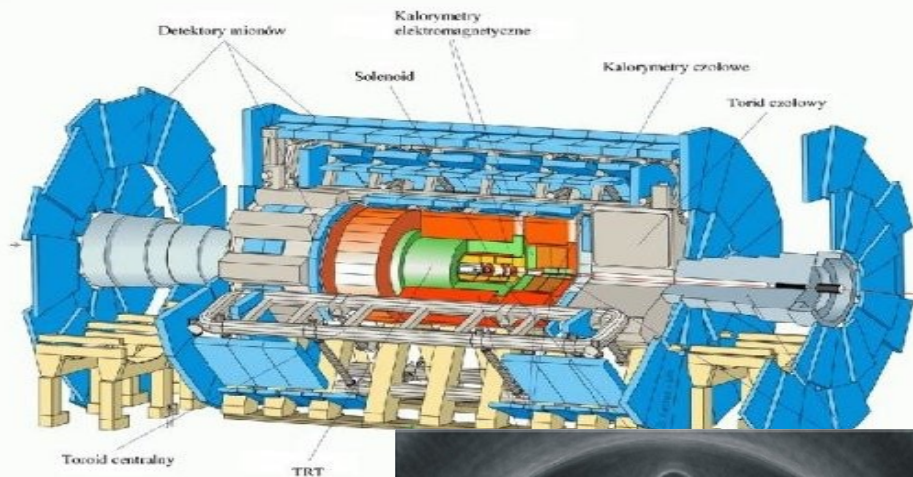
□ Distributed services

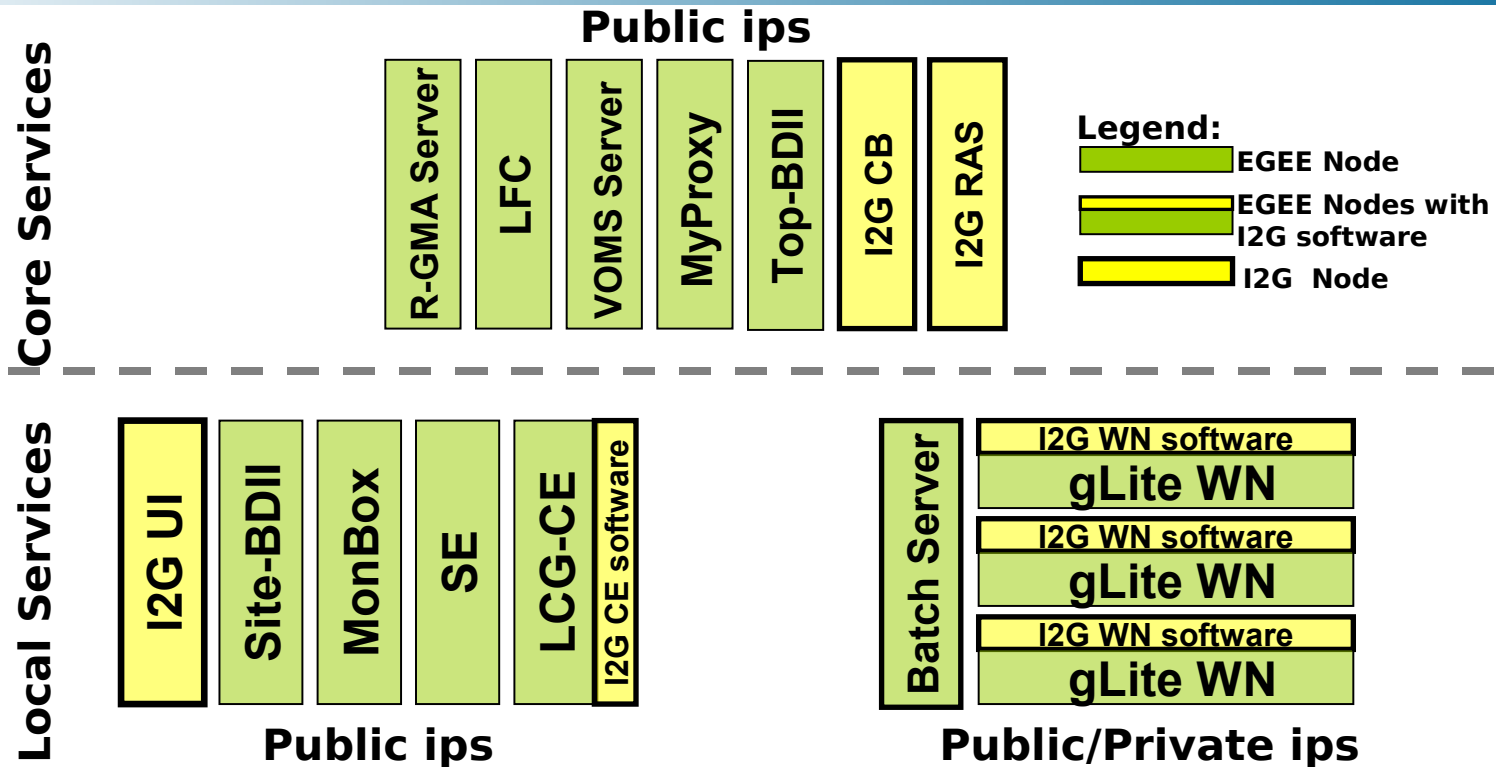
- ▶ Taking advantage of the partners expertise
- ▶ Redundancy
- ▶ Better use of resources



□ Applications

- ▶ ifusion
- ▶ ienvmod
- ▶ iusct
- ▶ ibrain
- ▶ ihep
- ▶ iplanck
- ▶ iwien2k
- ▶ icompchem
- ▶ imrt
- ▶ euforia
- ▶ ihidra
- ▶ icesga
- ▶ imain, imon, itut, itest





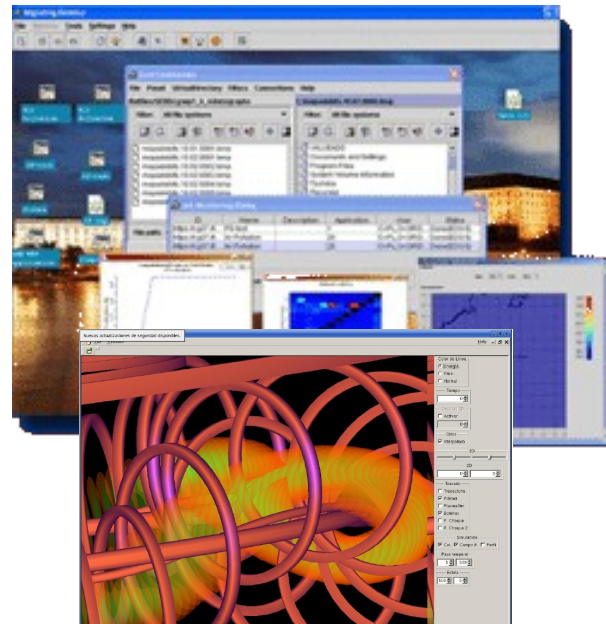
- Int.EU.Grid is a gLite based infrastructure but...
 - ▶ enhancing several services (lcg-CE, WN) and deploying new components (CB, RAS, UI) towards interactivity and MPI



Migrating Desktop and RAS

❑ Migrating Desktop (MD): User Friendly Grid Access

- ▶ Java based GUI; Hides the details of the grid
- ▶ Provides interactivity and visualization features
 - GView enables interactivity for OpenGL and X applications
- ▶ Allows to log-in in the GRID independently from
 - where you are (laptop, desktop, everywhere ...)
 - what kind of Computer/OS you are using (Windows, Linux)



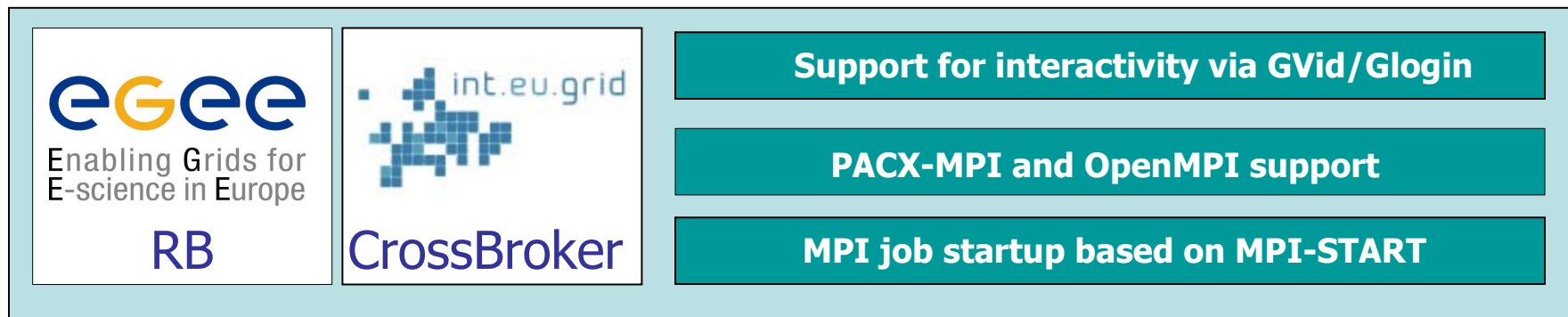
❑ Roaming Access Server (RAS):

- ▶ Gateway for Grid Access
- ▶ Performs actions on the grid on behalf of the MD



□ **CrossBroker: Int.EU.Grid meta-scheduler**

- ▶ Offers the same functionalities as the EGEE Resource Broker, plus:
 - Support for Interactive Applications
 - Interactive agent injection
 - Scheduling priorities;
 - Time Sharing
 - Full support for Parallel Applications
 - PACX-MPI and OpenMPI
 - Flexible MPI job startup based on MPI-START



- ❑ The user can decide which interactive agent to use through special JDL requirements
 - ▶ The CrossBroker can inject it transparently to the user

- ❑ If the job is recognized as interactive...
 - ▶ The CrossBroker treats it with higher priority

- ❑ There is in place a mechanism to use bandwidth measurements in the matchmaking process
 - ▶ Great for application needing visualisation
 - ▶ But not really implemented...

- ❑ If there are no available resources
 - ▶ Use a time sharing mechanism

Time Sharing: Glide-in mechanism

□ Main idea:

- ▶ Wrap every batch job with an agent (glide-in)
 - Agent will get control of the remote machine independently of its local resource manager.




□ Glide-in benefits on the interactivity framework

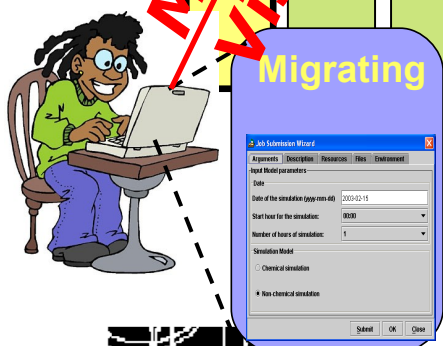
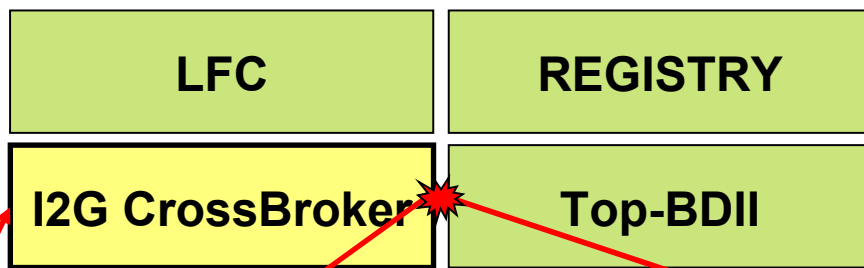
- ▶ Agents enable simple multiprogramming between interactive and batch jobs.
 - Interactive jobs may run even when no free resources are available.
- ▶ Agents can also be used as a fast start-up mechanism.
- ▶ Agent can control the amount of CPU that an interactive job gets according to QoS requirements expressed by the user in the JDL.

Int.EU.Grid basic workflow

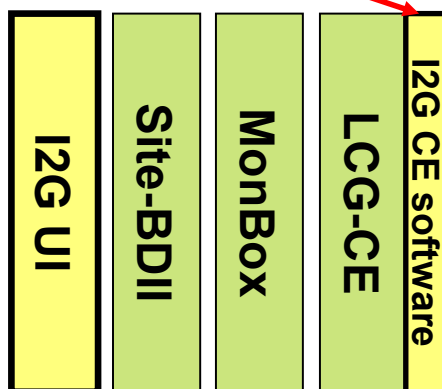
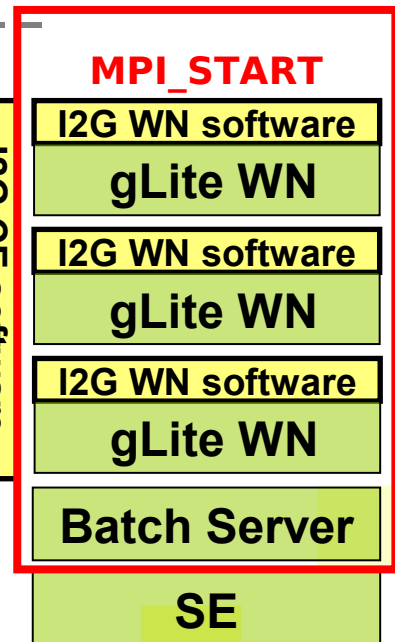
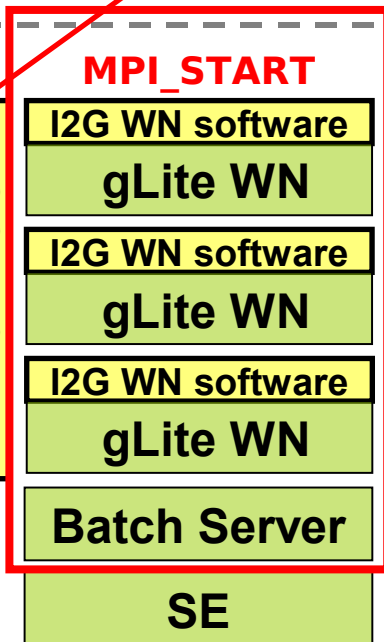
Local Services Core Services

Legend:

-  EGEE Node
-  EGEE Nodes with I2G software
-  I2G Node



MPI_VISUALIZATION



Int.EU.Grid added values

□ Int.EU.Grid offers

▶ **Interactivity and Visualization**

- "On the Fly" interaction, "On the Fly" response; Graphical

▶ **Inter and intra cluster parallel tasks**

- PACXMPI & OpenMPI

▶ **User friendly access to resources**

- The Migrating Desktop

□ Int.EU.Grid relies on the following components

▶ **CrossBroker**

- Scheduler and resource management
- Parallel tasks

Opportunity window to test interoperability with EGEE

... and applications supporting visualization
Int.EU.Grid software on top of the EGEE
(and gLite-WN)

1st interoperability approach: Share resources

- Many common sites between Int.EU.Grid and EGEE
 - ▶ How can we minimize the admin. effort on resource management?
 - Is it worthwhile to maintain separate clusters for different Grids?
 - What resources can be transparently shared?

- Local services
 - ▶ CE/MonBox are project dedicated to guaranty that accounting is correctly handled for each infrastructure
 - ▶ SE/LRMS may be transparently shared since they are configured on a VO basis
 - ▶ WNs may be transparently shared
 - After setting up a procedure to set project dependent environment variables (GFAL, LFC_HOST...)

Int.EU.Grid/EGEE shared WNs: Change Int.EU.Grid CE JobManager

- The JM in the Int.EU.Grid CE is changed to overwrite EGEE default environment variables in the WNs
 - ▶ Reads the user proxy and extracts the user VO
 - ▶ Reads a configuration file mapping VOs with the environment variables
 - ▶ If the user VO matches any of the VOs defined in the mapping file, it exports the corresponding environment variables

- Some of the environment variables identified up to now are:
 - ▶ LCG_GFAL_INFOSYS: Must point to the proper Top-BDII
 - ▶ VO_<USER_VO>_DEFAULT_SE: The default Storage Element
 - ▶ VO_<USER_VO>_SW_DIR: The VO software directory

- The VO environment mapping file
 - ▶ Placed in /opt/globus/lib/perl/Globus/GRAM/JobManager/vo_environment

CPU allocation EGEE/Int.EU.Grid

- Sharing WNs between several infrastructures
 - ▶ Don't "starve to death" one of these infrastructures
 - Avoid that long batch jobs from EGEE prevent "short" and interactive jobs from Int.EU.Grid to run
 - Properly configure your Batch System
 - Avoid that Int.EU.Grid job floods prevent EGEE jobs to enter
 - ▶ Mechanism
 - Slots/CPU reservation
 - If slots > CPUs, accounting and efficiency issues
 - Preemption mechanism
 - Data transfers and efficiency computation issues
 - Time Sharing "Glide-in" mechanism

- One batch server in the system
 - ▶ And (at least) two CEs
 - ▶ Each CE reports accounting information to different project MON Boxes
 - ▶ Each MON Box reports to the project dedicated central registry via R-GMA

- Batch system Logs
 - ▶ Must be shared by NFS with the CEs
 - ▶ Transferred/copy on a daily basis to the CE

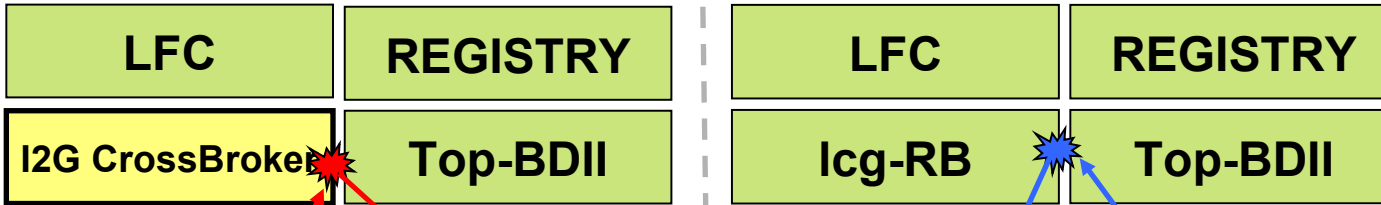
Scheme: Int.EU.Grid/EGEE shared WN

Core Services





Local Services

Int.EU.Grid Infrastructure

EGEE Infrastructure

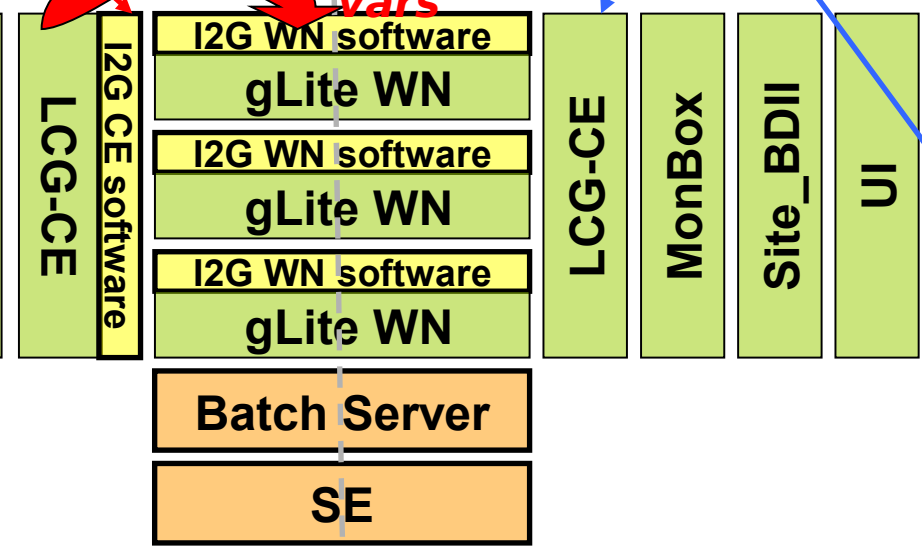
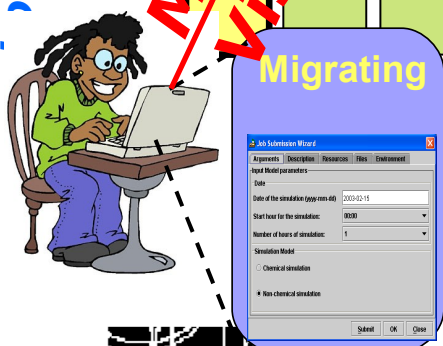


Legend:

-  EGEE Node
-  EGEE Nodes with I2G software
-  I2G Node
-  VO based sharing

Migrating
Visualization
MPI

JM env
vars



Int.EU.Grid &
EGEE VOs



2nd interoperability approach: EGEE VOs configured in Int.EU.Grid sites

- Int.EU.Grid/EGEE sites sharing WNs opens new possibilities:
 - ▶ EGEE users may have access to Int.EU.Grid software and features
 - ▶ VOs should start by establishing a SLA with the project

- Support EGEE VOs
 - ▶ At the core services level (CrossBroker) and local level (to be negotiated)

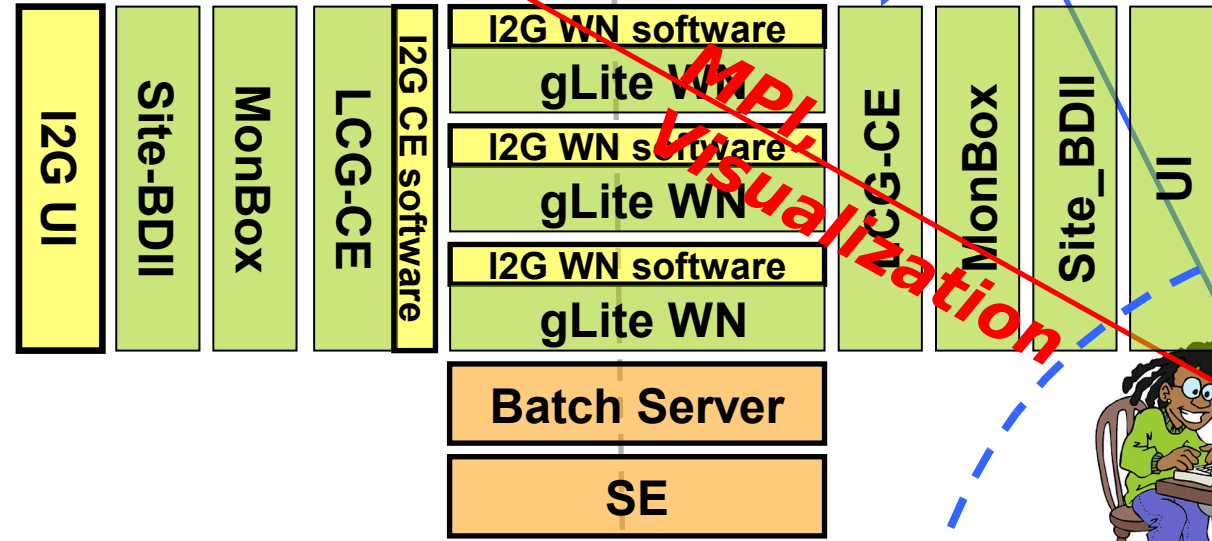
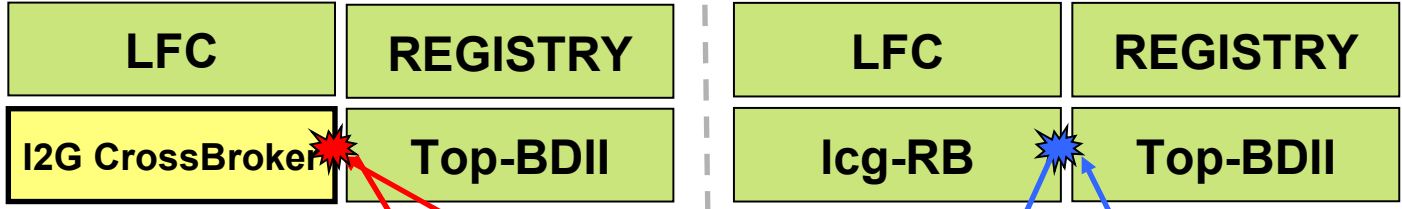
- Int.EU.Grid enhancements towards interoperability with EGEE
 - ▶ Set of packages to be installed on top of EGEE UIs
 - YAIM compatible
 - EGEE and Int.EU.Grid compliant: "edg-job-..." and "i2g-job-..." comand lines
 - ▶ CrossBroker/Top-BDII support
 - Configured EGEE VOs along with Int.EU.Grid VOs
 - Ldap contact strings from EGEE/Int.EU.Grid shared sites added to top-BDII

Scheme: EGEE VOs configured in Int.EU.Grid sites

Core Services
Local Services

Int.EU.Grid Infrastructure

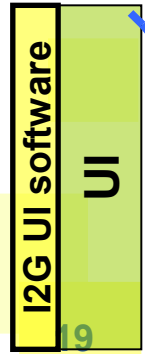
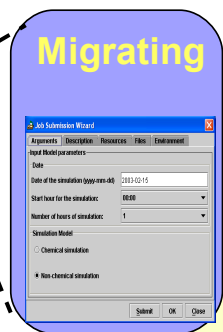
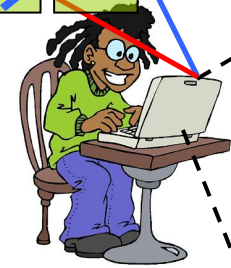
EGEE Infrastructure



- Legend:**
- EGEE Node
 - EGEE Nodes with I2G software
 - I2G Node
 - VO based sharing

API Visualization

Int.EU.Grid & EGEE VOs



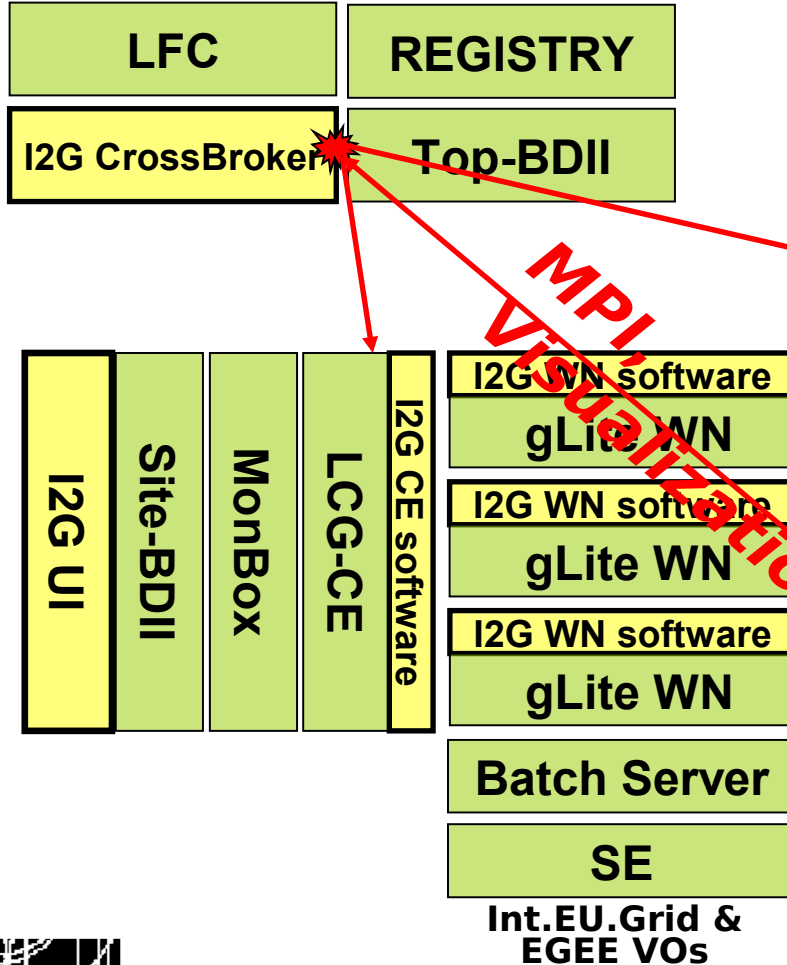
3rd interoperability approach: EGEE VOs bring their own resources

- EGEE sites already supporting those VOs may:
 - ▶ Install UI, CE and WN Int.EU.Grid middleware on top of the EGEE middleware
 - ▶ Add the site LDAP string in the CrossBroker Top-BDII
 - ▶ Start using Int.EU.Grid features
 - Run jobs inside the set of sites defined in Int.EU.Grid top-BDII

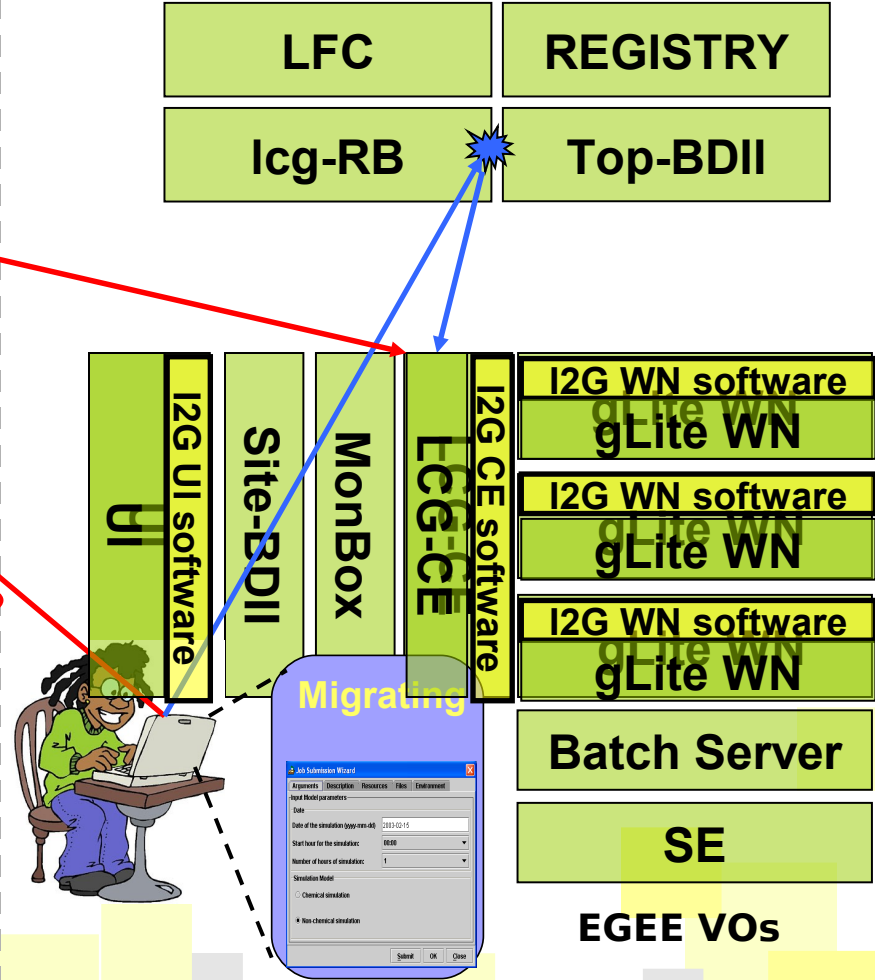
Scheme: EGEE VOs bring their own resources

Local Services Core Services

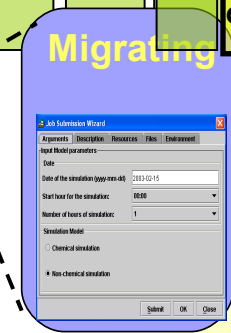
Int.EU.Grid Infrastructure



EGEE Infrastructure



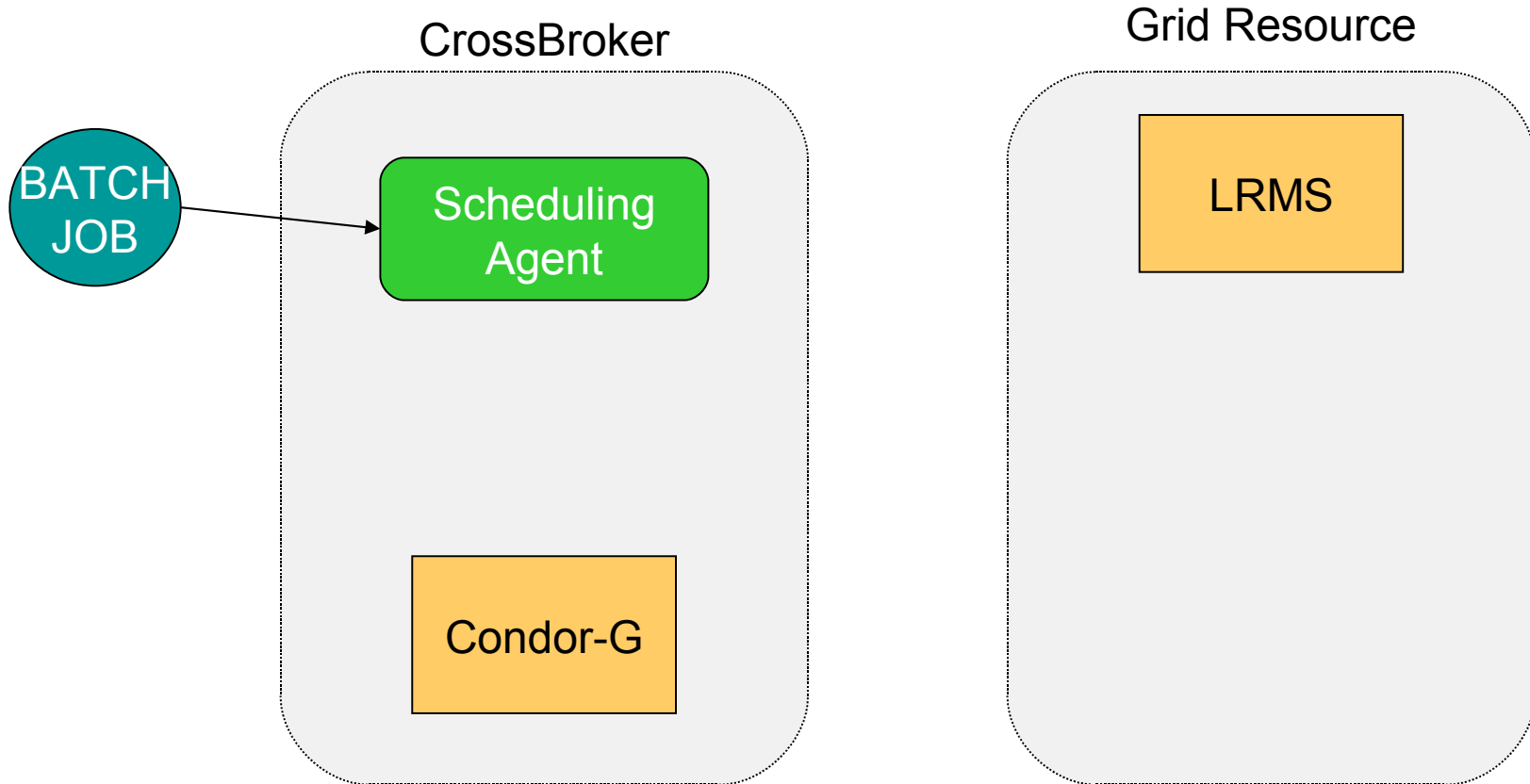
MPI
Visualization

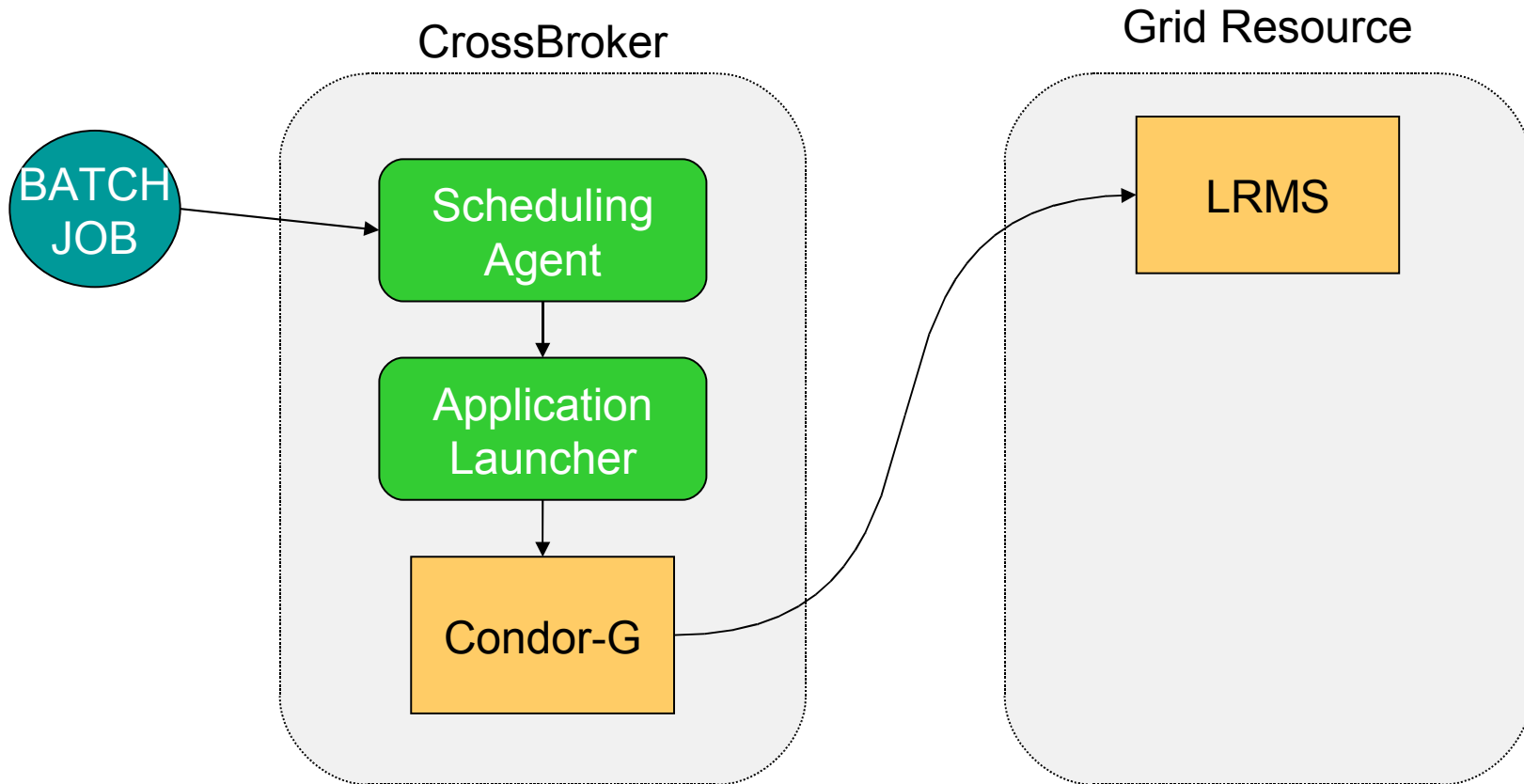


- We propose different approaches for sites wishing to operate between Int.EU.Grid and EGEE infrastructures:
 - ▶ The same institutions with separate sites in both projects may
 - Share the same physical Worker Nodes
 - Jobs from both Grid are scheduled to the sites by different Brokers
 - Are submitted to the cluster via dedicated CEs but managed by a common Batch system
 - ▶ EGEE sites may take advantage of Int.EU.Grid features
 - Without the need to join Int.EU.Grid project
 - Through the installation of some Int.EU.Grid nodes and via the deployment of specific Int.EU.Grid software
 - ▶ Full documentation under

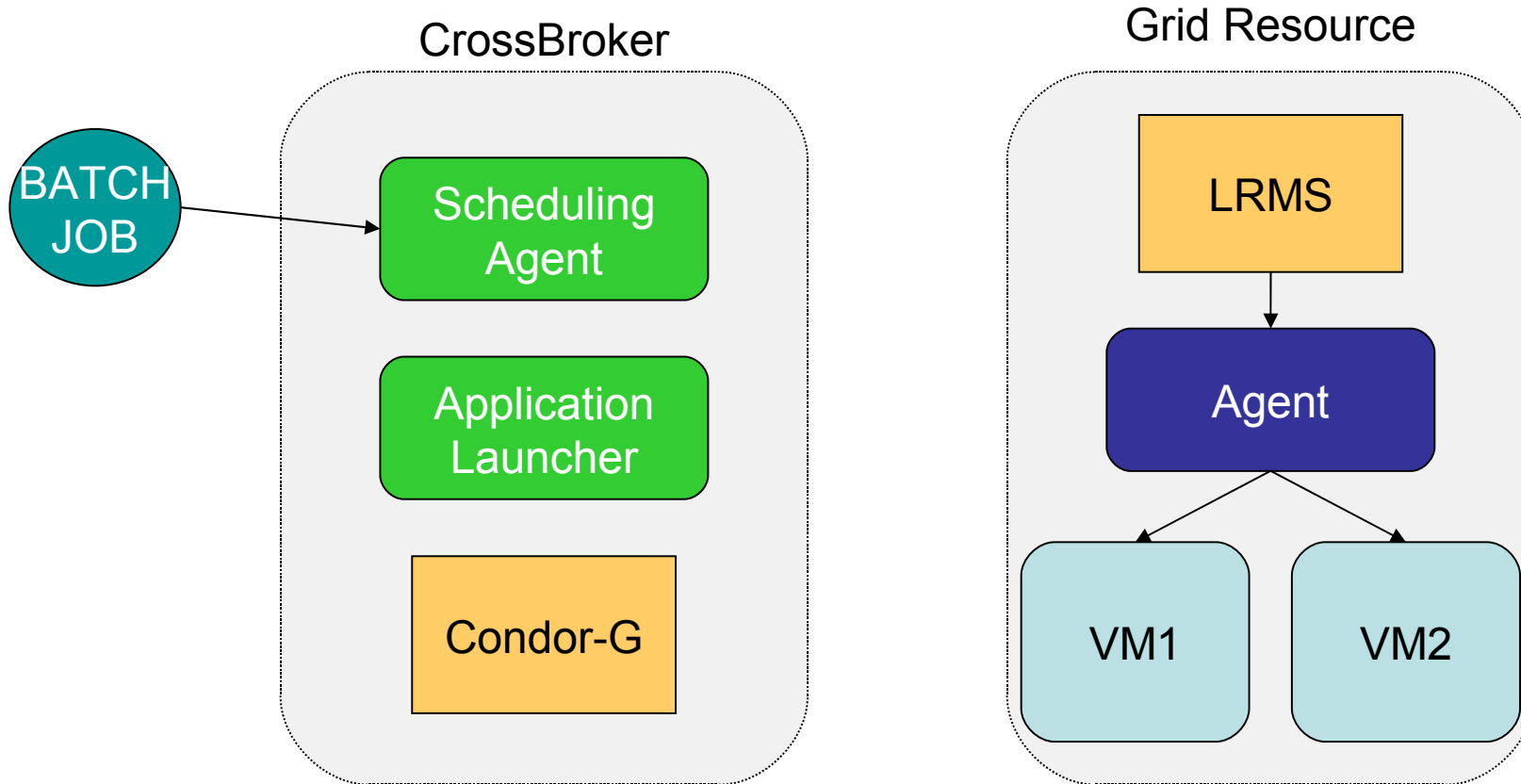
https://wiki.fzk.de/i2g/index.php/inter-operability_between_I2G_and_EGEE

Time Sharing

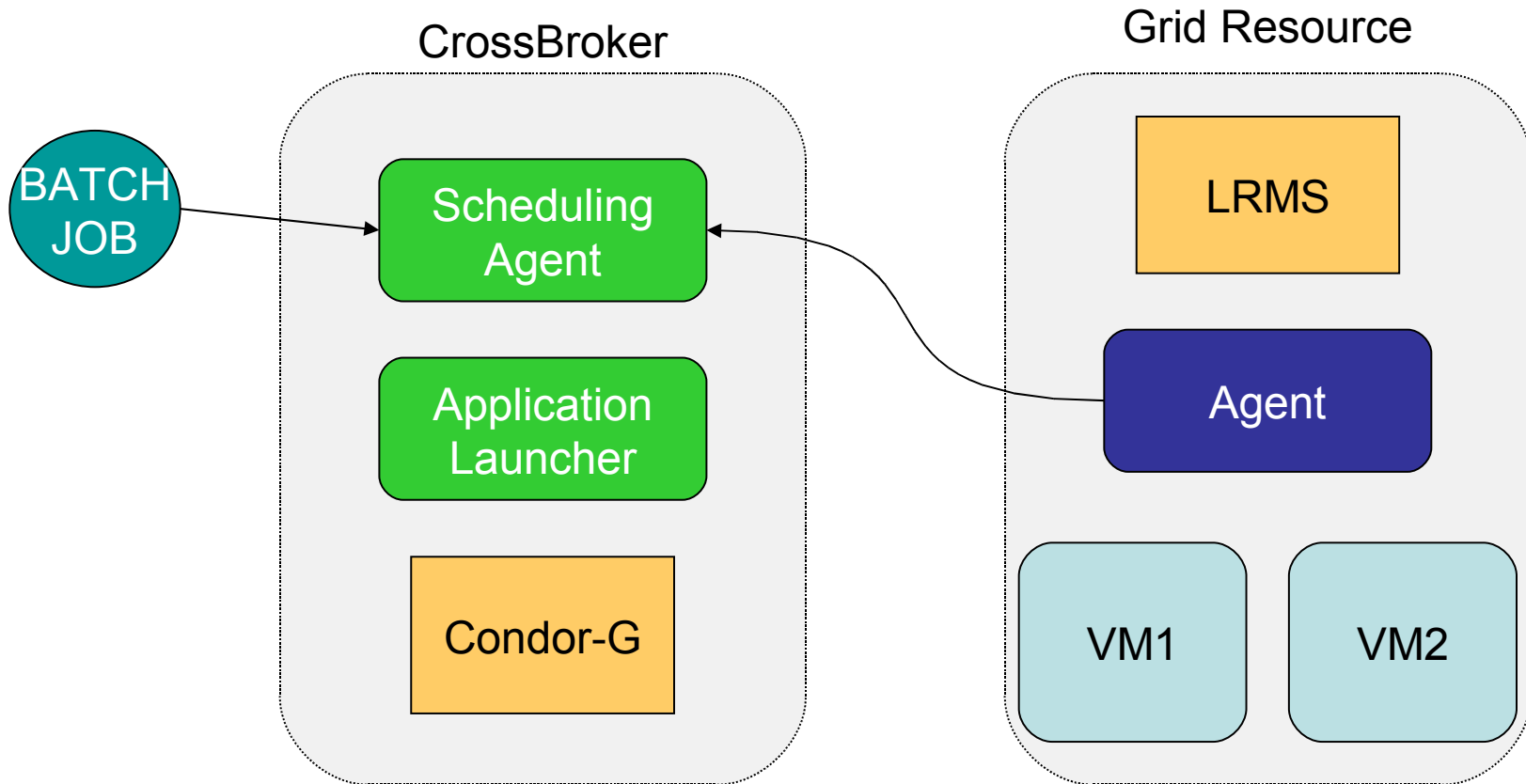


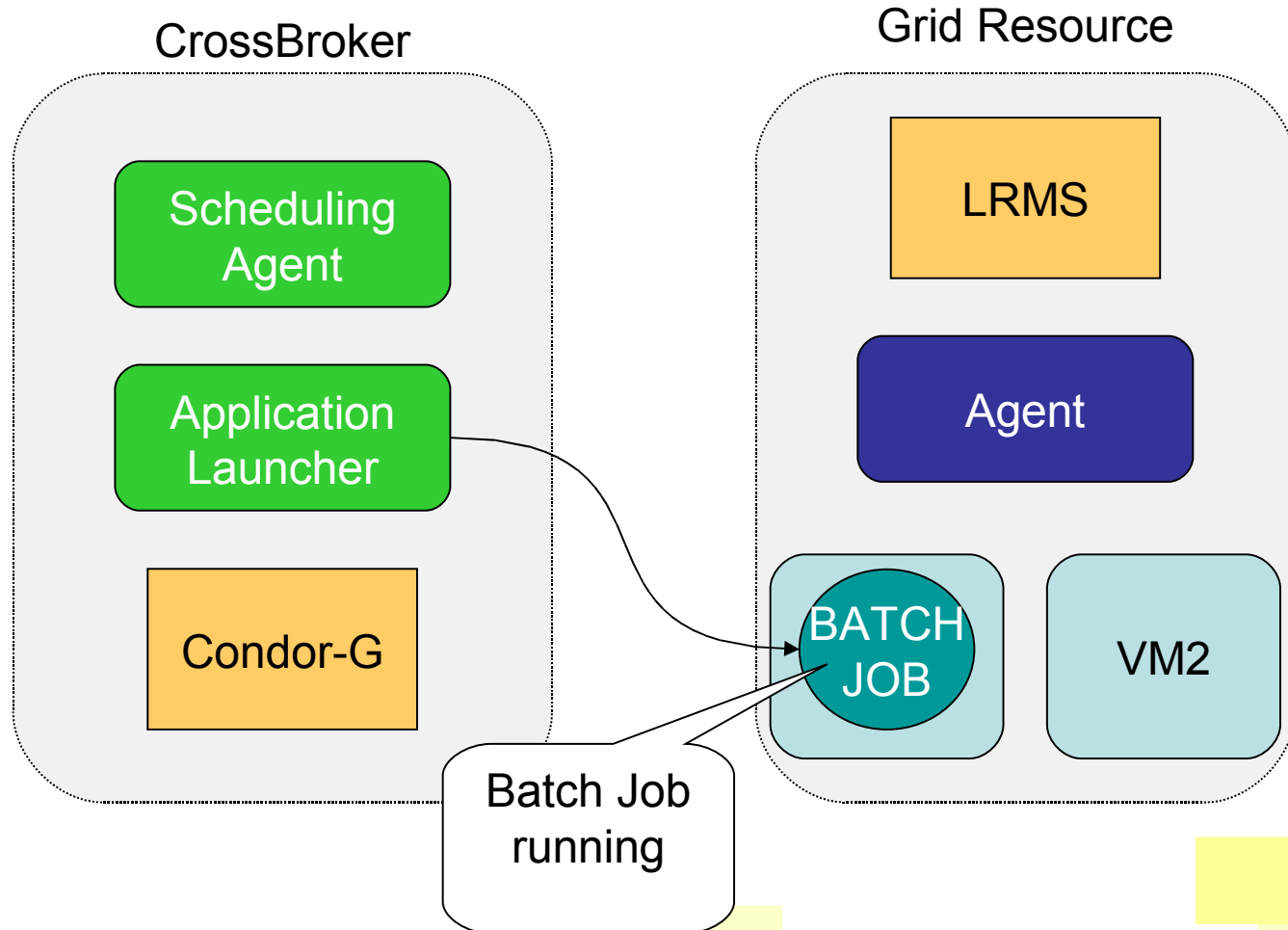


Time Sharing

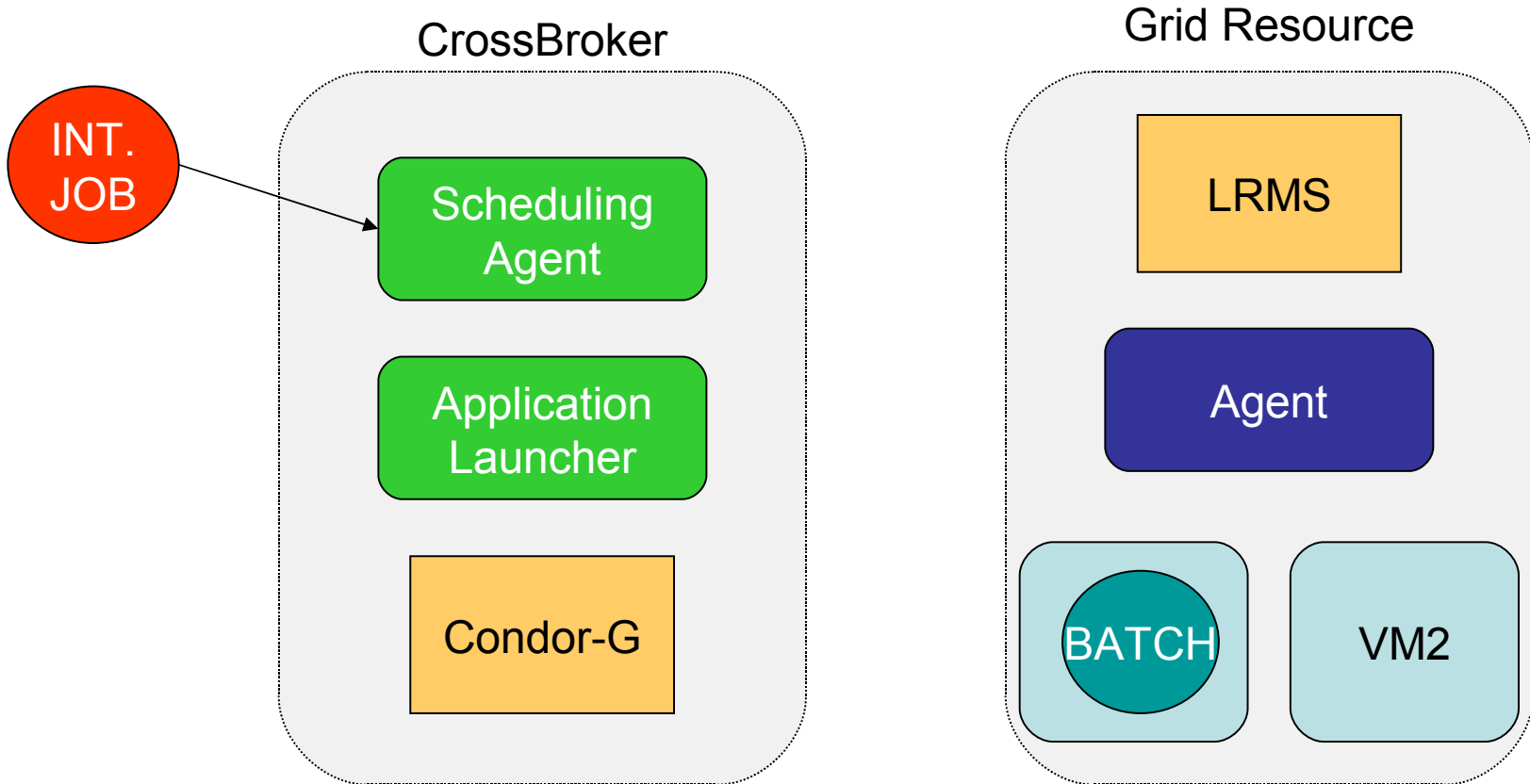


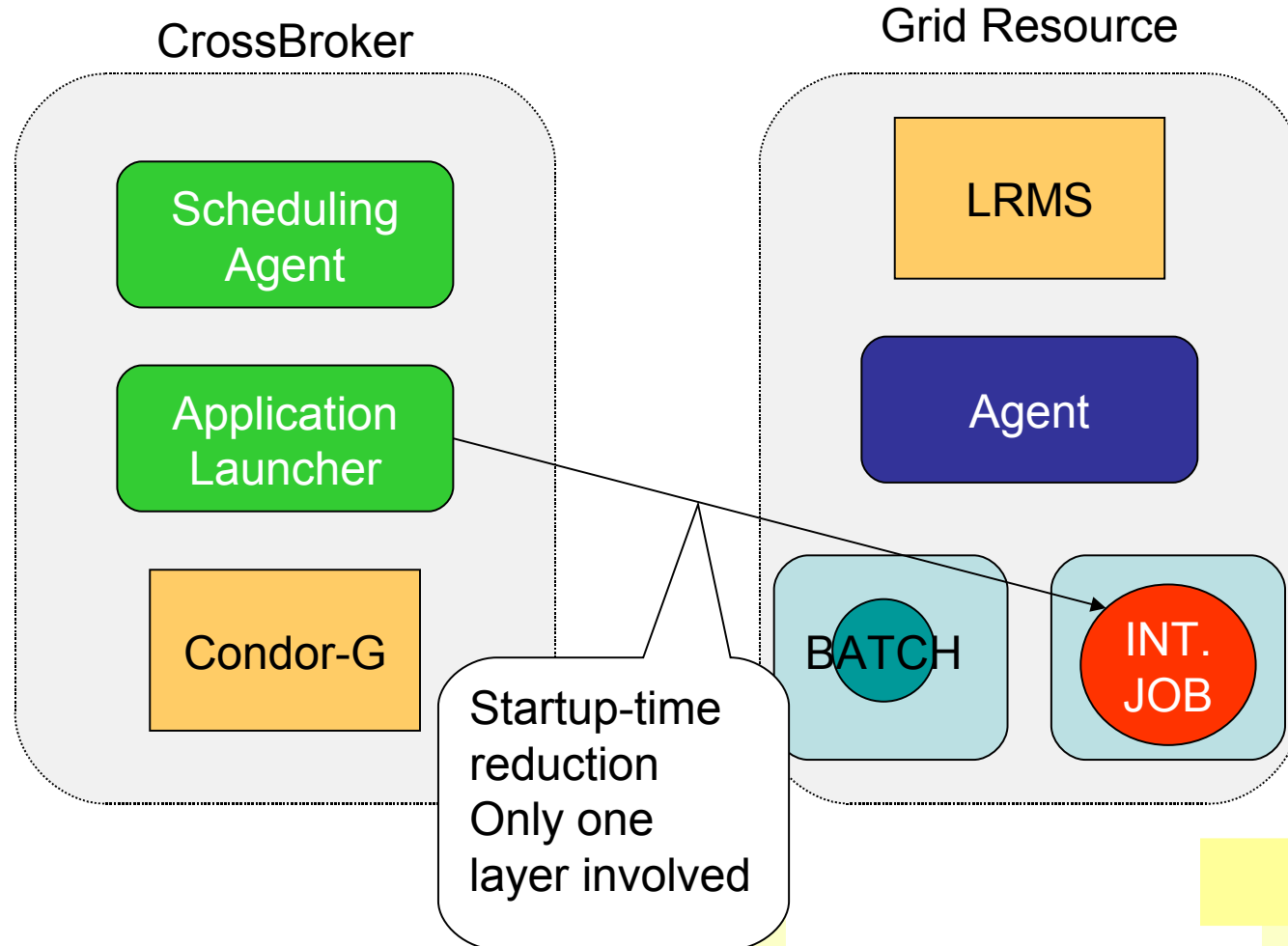
Time Sharing





Time Sharing





CrossBroker

CE + WN

Mechanism	Resource Searching	Resource Selection	Submission	
			Campus Grid	Remote Site
Free machine submission	3s	0.5s	17.2s	22.3s
Glidein submission to free machine	3s	0.5s	29.3s	33.25s
Virtual Machine submission	0.5s		6.79s	8.12s