



“LCG2 Operational Experience and Status”

Markus Schulz, IT-GD, CERN
markus.schulz@cern.ch





- Building LCG-2
- Impact of Data Challenges on operations
- Problems
- Operating LCG
 - how it was planned
 - how it was done
- Summary of the operations workshop at CERN



- Jan 2003 GDB agreed to take VDT and EDG components
- September 2003 LCG-1
 - **Extensive certification process**
 - **Integrated 32 sites ~300 CPUs first use for production**
- December 2003 LCG-2
 - **Deployed in January to 8 core sites**
 - **Introduced a pre-production service for the experiments**
 - **Alternative packaging (tool based and generic installation guides)**
- Mai 2004 -> now monthly incremental releases (not all distributed)
 - **Driven by the experiences from the data challenges**
 - **Balance between stable operation and improved versions (driven by users)**
 - **2-1-0, 2-1-1, 2-2-0, (2-3-0)**
 - **(Production services RBs + BDIs patched on demand)**
 - **> 80 sites (3-5 failed)**



- ... Group **or** Regional Operation Center
- Site [page](#)
- Sites [tool based installation](#)

worked **3-5 times**

- sites are daily re-certified and pre-ANNAH
- Experiments install their software and a
- **Adding new sites is now a quite smooth process**



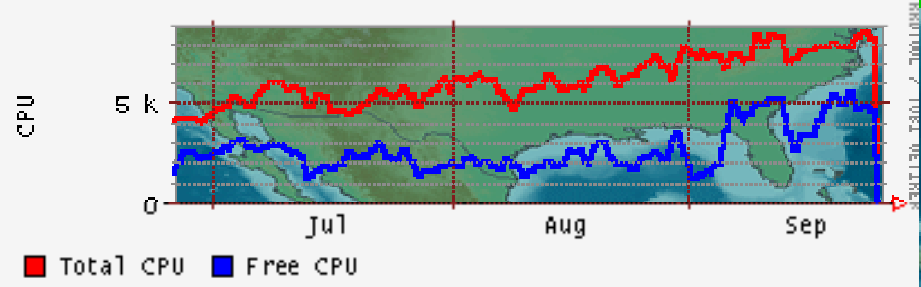
new interested sites should look here: [release](#)



LCG-2 Status 15/11/2004

Total:
87 Sites
~9700 CPUs
~6.5 PByte

● PK-NCP (Pakistan)



- Monthly process
 - Gathering of new material
 - Prioritization
 - Integration of items on list
 - Deployment on testbeds
 - First tests
 - feedback
 - Release to EIS testbed for experiment validation
 - Full testing (functional and stress)
 - feedback to patch/component providers
 - final list of new components
 - Internal release (LCFGng)
- On demand
 - Preparation/Update of release notes for LCFGng
 - Preparation/Update of manual install documentation
 - Test installations on GIS testbeds
 - Announcement on the LCG-Rollout list

C&T
Certification & Testing

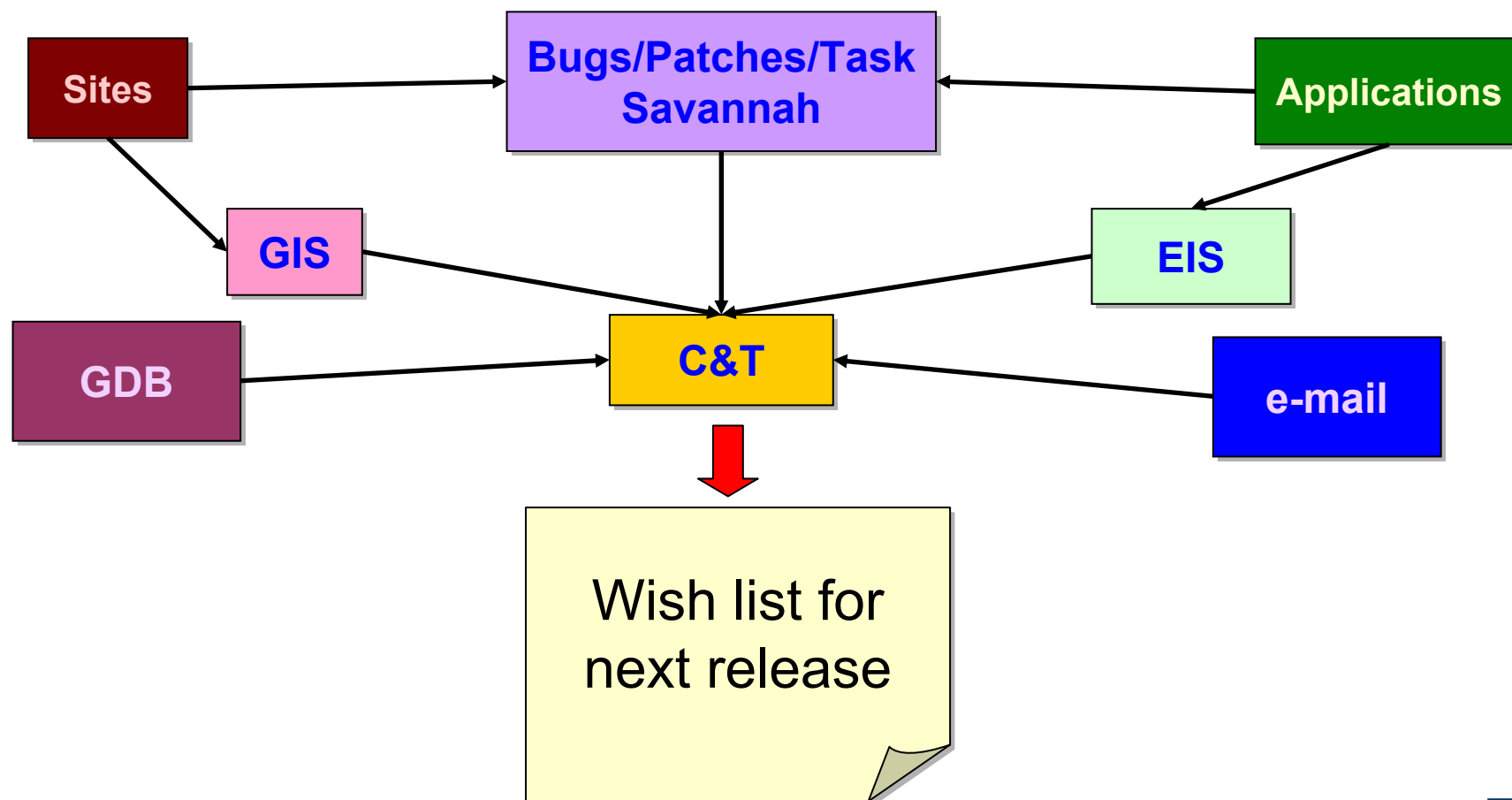
GDB
Grid Deployment
Board

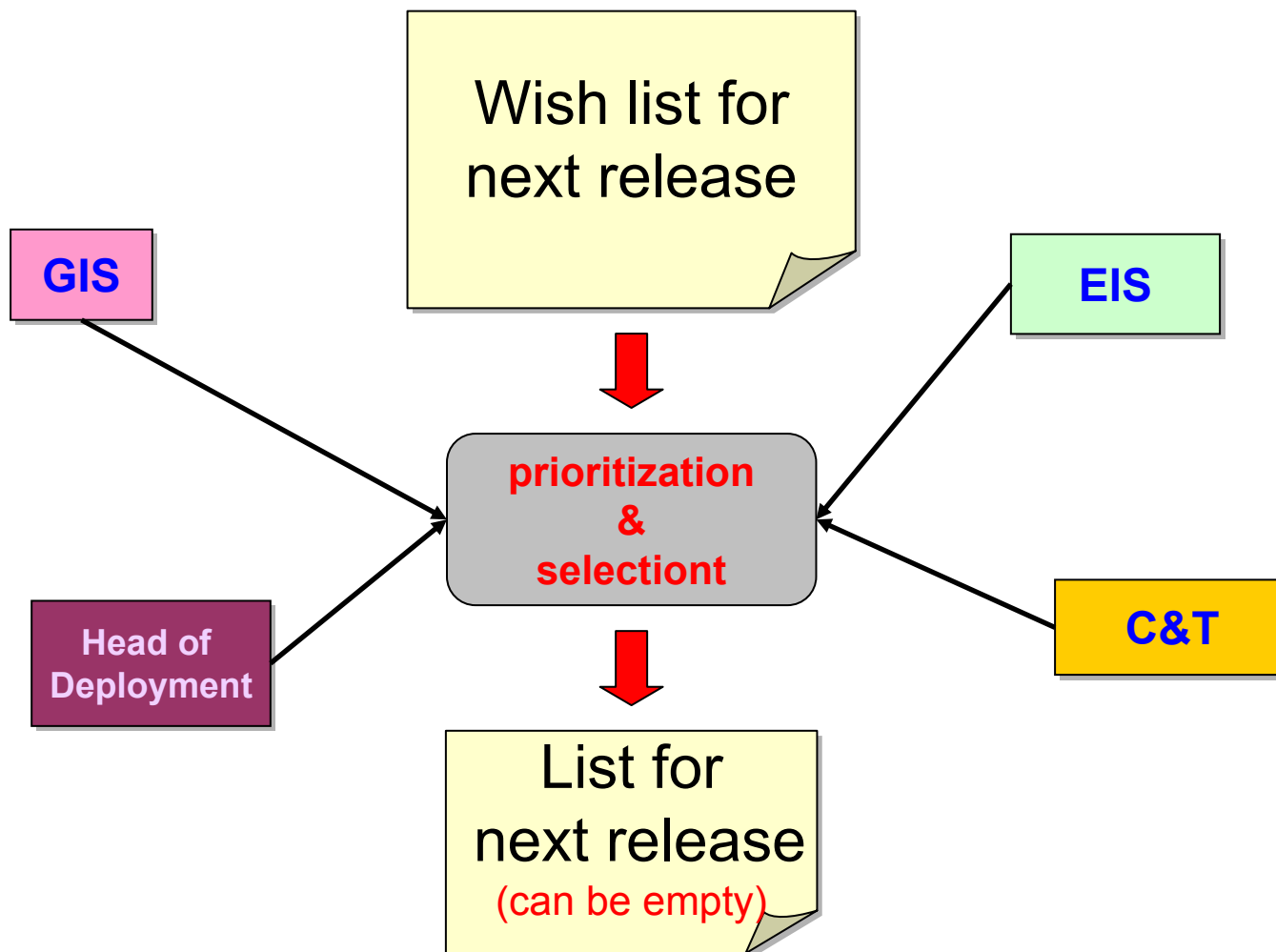
EIS
Experiment Integration
Support

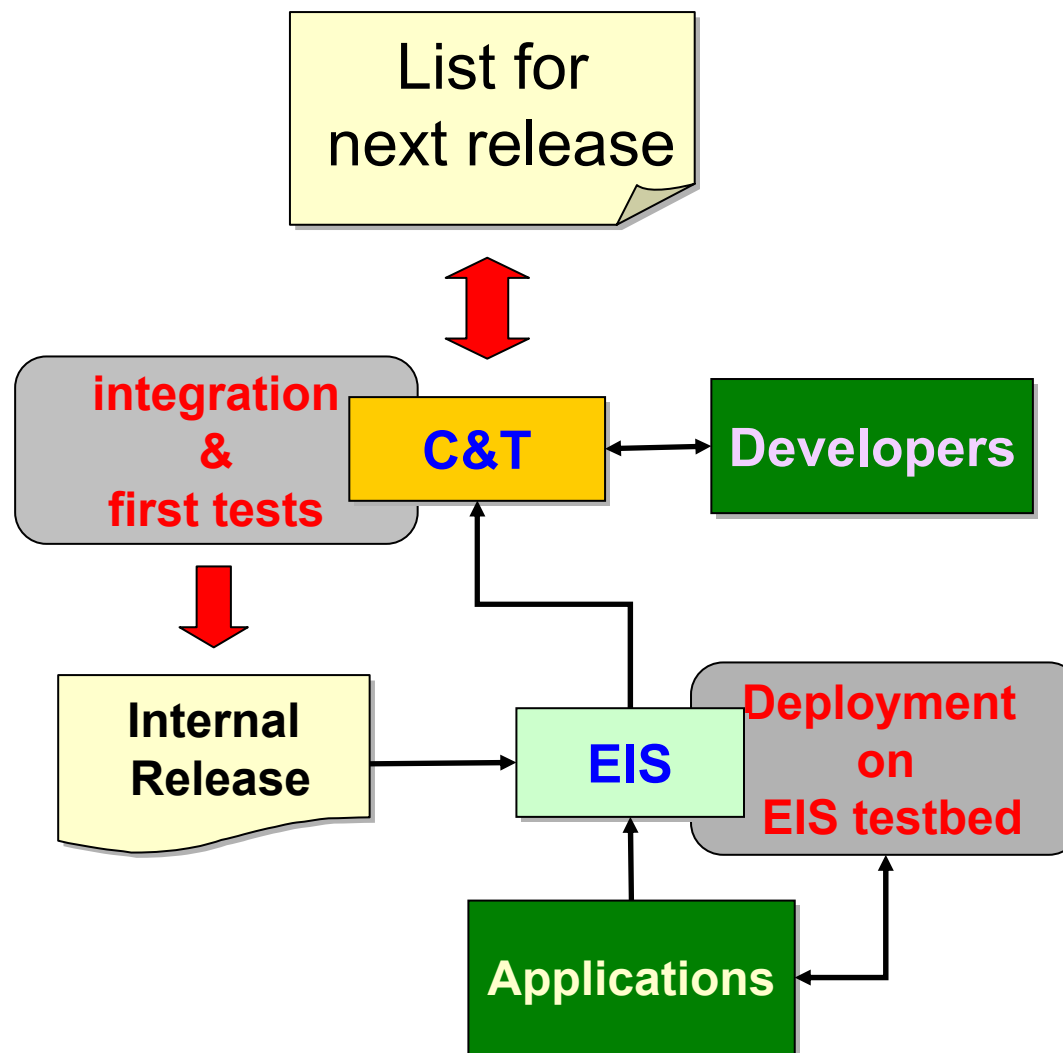
Applications

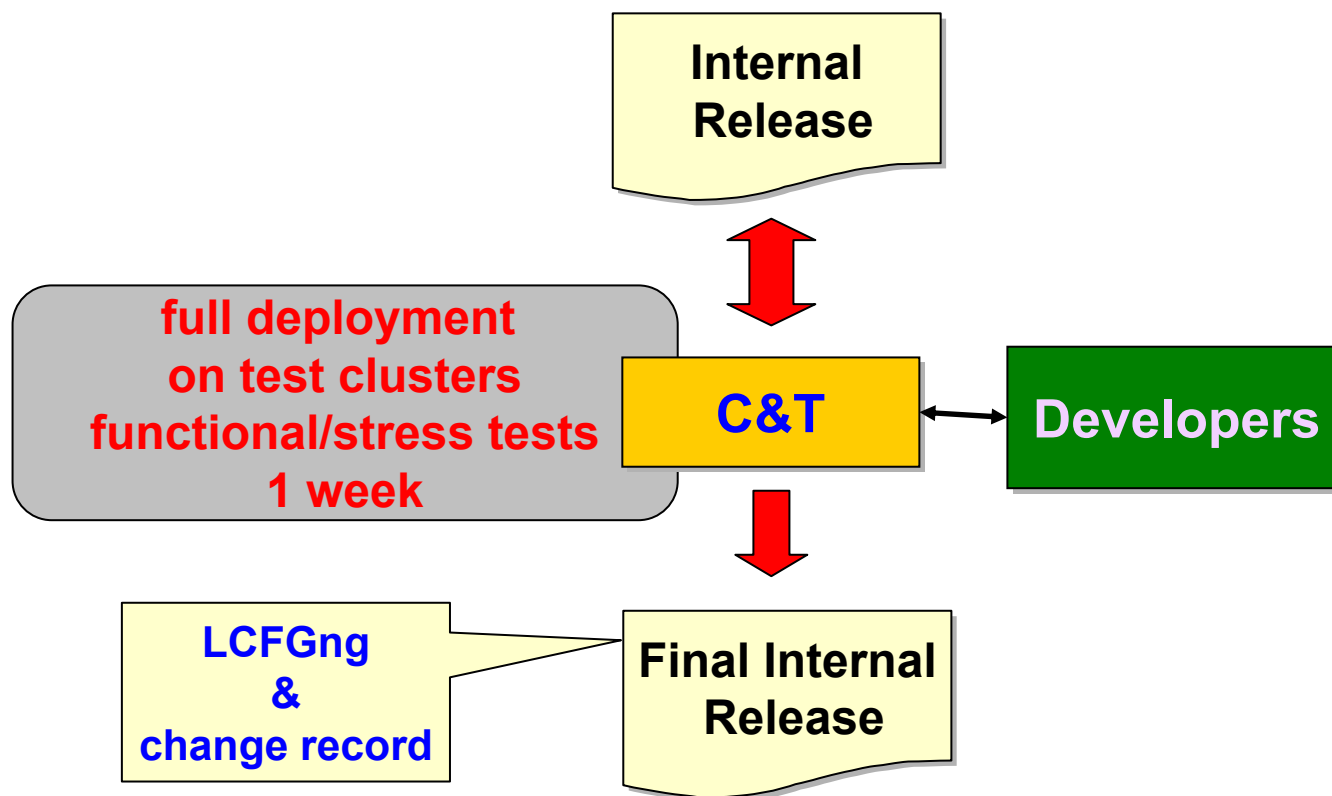
GIS
Grid Infrastructure
Support

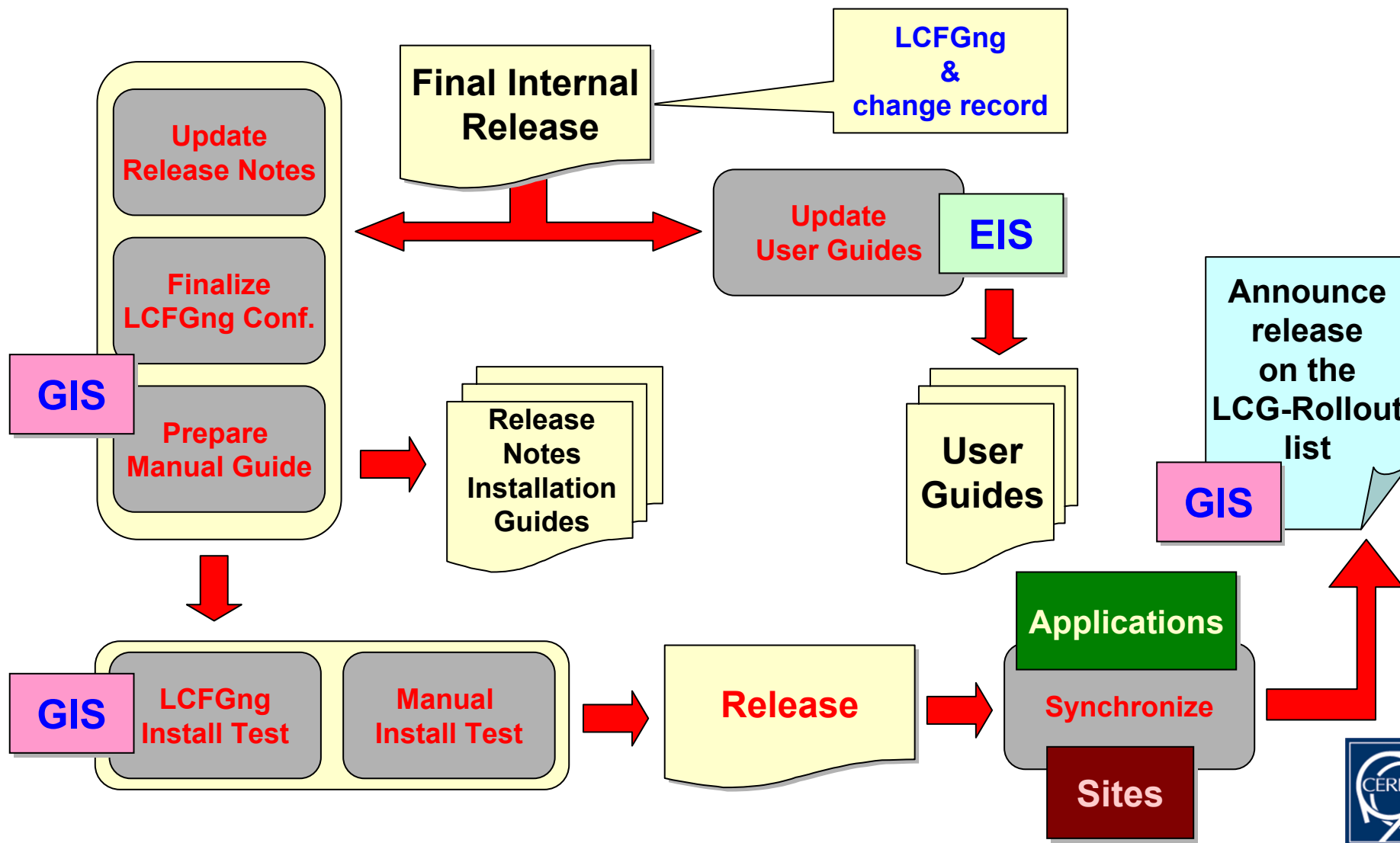
Sites

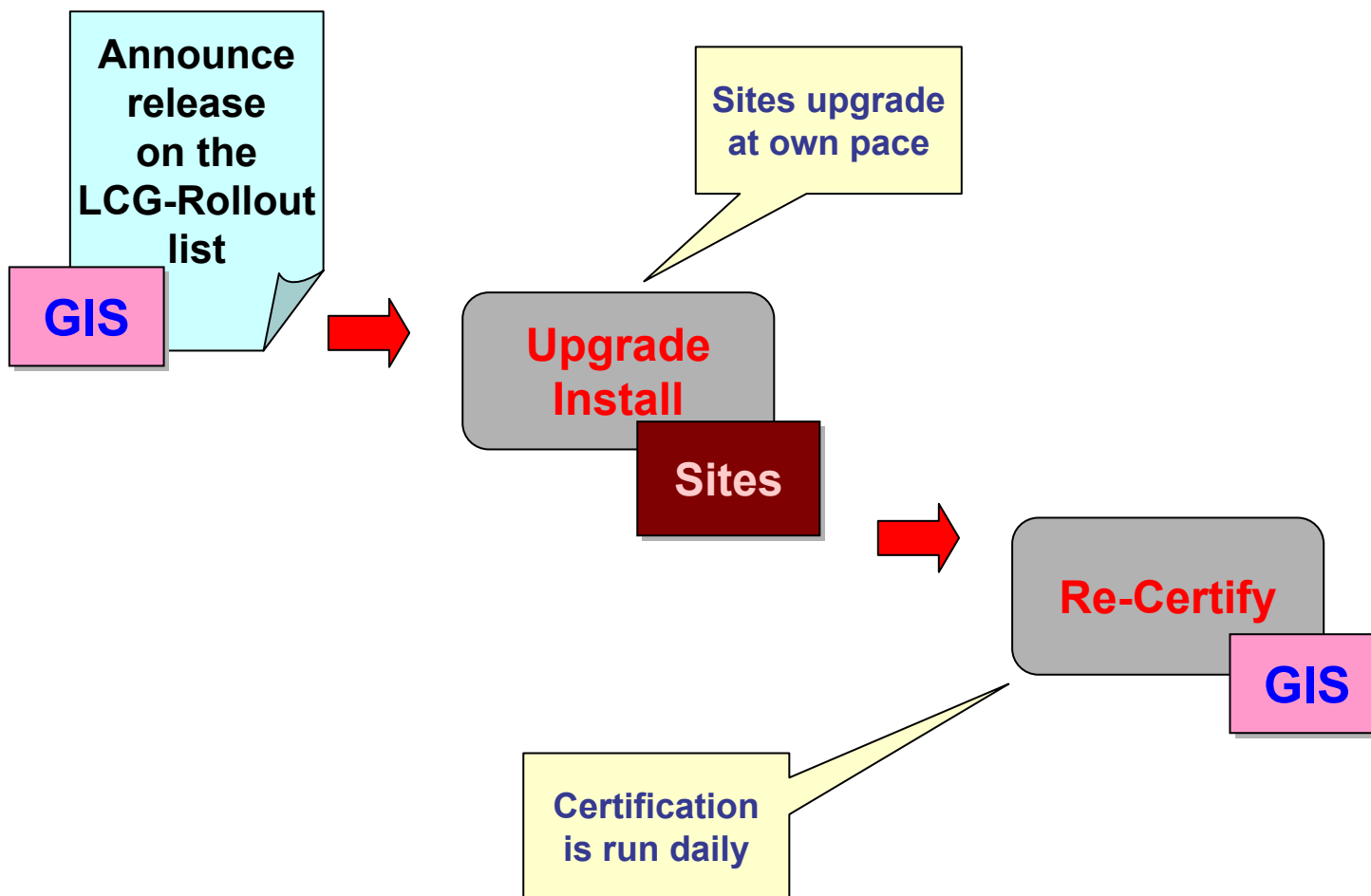






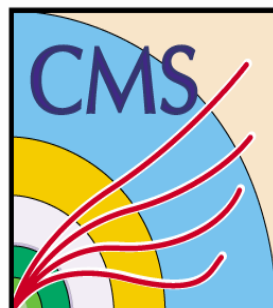






- The process was decisive to improve the quality of the middleware
- The process is time consuming
 - **There are many sequential operations**
 - **The format of the internal and external release could be unified**
 - **Multiple packaging formats slow down release preparation**
 - **tool based (LCFGng)**
 - **manual (+tar ball based)**
 - **All components are treated equal**
 - **same level of testing for core components and non vital tools**
 - **no difference for tools already in use in other project**
- Process of including new components not totally transparent
- Picking a good time for a new release is difficult
 - **conflict between users (NOW) and sites (planned)**
- Upgrading has proven to be a high risk operation
 - **some sites suffered from acute configuration amnesia**
- **Process was one of the topics in the “LCG Operations Workshop”**

- Large scale production effort of the LHC experiments
 - test and validate the computing models
 - produce needed simulated data
 - test experiments production frame works and software
 - test the provided grid middleware
 - test the services provided by LCG-2
- All experiments used LCG-2 for part of their production
 - **Details will be given by Dario Barberis**



- Reality is different from certification tests
 - Usage characteristics of real production can't be simulated
 - (Detailed) planning is of limited use
 - time better spend with small(er) scale trail DCs
- Time matters
 - Delays in support and operation actions are deadly during DCs
 - Several iterations needed to get it right
 - Communication between sites, operations, and experiments matters
 - not all players handled DCs with same priority (communication problem)
- Middleware matures quickly during DCs
 - Scalability, robustness, functionality
 - Several of the experiments will do continues production from now on
- Probing concurrently multiple “dimensions” of the system can be confusing
 - Problems with different components mix
 - leads to confuse cause and effect
 - Dedicated “Service Challenges” will help



- **All experiments encountered on LCG-2 similar problems**
- LCG sites suffering from configuration and operational problems
 - not adequate resources on some sites (hardware, human..)
 - **this is now the main source of failures**
- Load balancing between different sites is problematic
 - jobs can be “attracted” to sites that have no adequate resources
 - modern batch systems are too complex and dynamic to summarize their behavior in a few values in the IS
- Identification and location of problems in LCG-2 is difficult
 - distributed environment, access to many logfiles needed.....
 - status of monitoring tools
- Handling thousands of jobs is time consuming and tedious
 - Support for bulk operation is not adequate
- Performance and scalability of services
 - storage (access and number of files)
 - job submission
 - information system
 - file catalogues
- Services suffered from hardware problems (no fail over (design problem))



- Collection: [Outstanding Middleware Issues](#)
 - Important: 1st systematic confrontation of required functionalities with capabilities of the existing middleware
 - Some can be patched, worked around,
 - Those related to fundamental problems with underlying models and architectures have to be input as essential requirements to future developments (EGEE)
- Middleware is now not perfect but quite stable
 - Much has been improved during DC's
 - A lot of effort still going into improvements and fixes
 - Easy to deploy and operate space management still an issue
 - especially for Tier 2 sites
 - effort to adapt dCache
 - effort in simple Disk Pool Manager

- **Slow response from sites**
 - Upgrades, response to problems, etc.
 - **Problems reported daily – some problems last for weeks**
- **Lack of staff available to fix problems**
 - Vacation period, other high priority tasks
- **Various mis-configurations (see next slide)**
- **Lack of configuration management – problems that are fixed reappear**
- **Lack of fabric management (mostly smaller sites)**
 - scratch space, single nodes drain queues, incomplete upgrades,
- **Lack of understanding**
 - Admins reformat disks of SE ...
- **Provided documentation often not read (carefully)**
 - new activity started to develop “hierarchical” adaptive documentation
 - simpler way to install middleware on farm nodes (even remotely in user space)
- **Firewall issues –**
 - often less than optimal coordination between grid admins and firewall maintainers
- **PBS problems**
 - Scalability, robustness (switching to torque helps)

- Site mis-configuration was responsible for **most** of the problems that occurred during the experiments Data Challenges. Here is a non-complete list of problems:

integrated all common small problems into

**ONE
BIG PROBLEM**

- - Default user shell environment too big
-
- Only partly related to middleware complexity

- Multiple instances of core services for each of the experiments
 - separates problems, avoids interference between experiments
 - improves availability
 - allows experiments to maintain individual configuration (information system)
 - addresses scalability to some degree
- Monitoring tools for services currently not adequate
 - tools under development to implement control system
- Access to storage via load balanced interfaces
 - CASTOR
 - dCache
- Services that carry “state” are problematic to restart on new nodes
 - needed after hardware problems, or security problems
- “State Transition” between partial usage and full usage of resources
 - required change in queue configuration (faire share, individual queues/VO)
 - next release will come with description for fair share configuration (smaller sites)

- User (Experiment) Support:

- GD at CERN worked very close with the experiments production managers
- Informal exchange (e-mail, meetings, phone)
 - “No Secrets” approach, GD people on experiments mail lists and vice versa
 - ensured fast response
 - tracking of problems tedious, but both sites have been patient
 - clear learning curve on **BOTH** sites
 - LCG GGUS (grid user support) at FZK became operational after start of the DCs
 - due to the importance of the DCs the experiments switch slowly to the new service
 - Very good end user documentation by GD-EIS
 - Dedicated testbed for experiments with next LCG-2 release
 - rapid feedback, influenced what made it into the next release



- Installation (Site) Support:

- GD prepared releases and supported sites (certification, re-certification)
- Regional centres supported their local sites (some more, some less)
- **Community style help via mailing list (high traffic!!)**
- FAQ lists for trouble shooting and configuration issues: [Taipei RAL](#)

- Operations Service:
 - RAL (UK) is leading sub-project on developing operations services
 - Initial prototype <http://www.grid-support.ac.uk/GOC/>
 - Basic monitoring tools
 - Mail lists for problem resolution
 - Working on defining policies for operation, responsibilities (draft document)
 - Working on grid wide accounting
 - Monitoring:
 - GridICE (development of DataTag Nagios-based tools)
 - GridPP job submission monitoring
 - Information system monitoring and consistency check
<http://goc.grid.sinica.edu.tw/gstat/>
 - CERN GD daily re-certification of [sites \(including history\)](#)
 - escalation procedure under development
 - tracing of site specific problems via problem tracking tool
 - tests core services and configuration

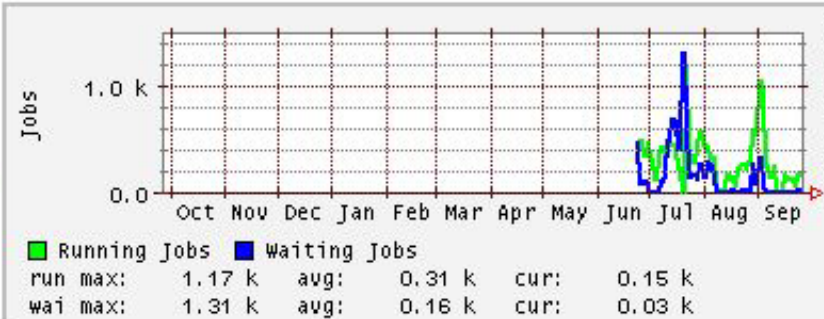
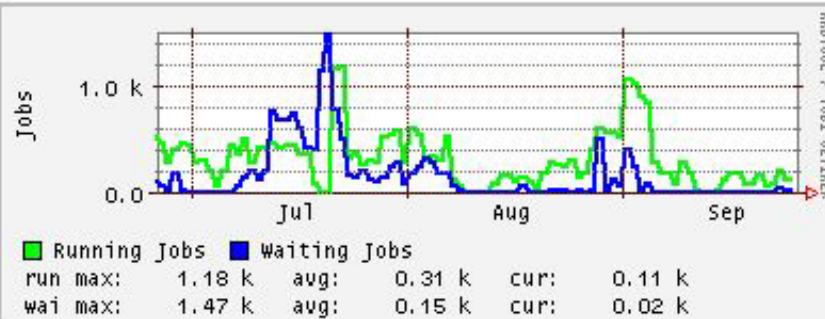
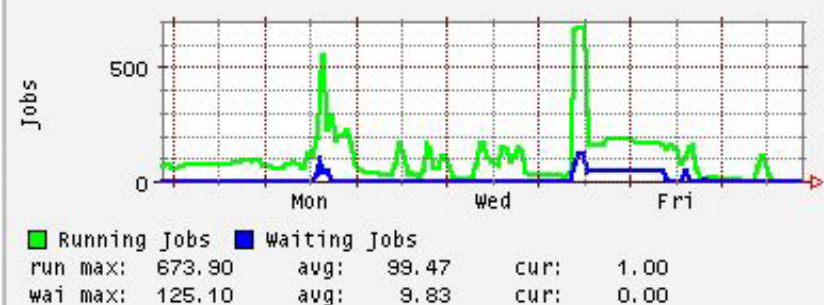
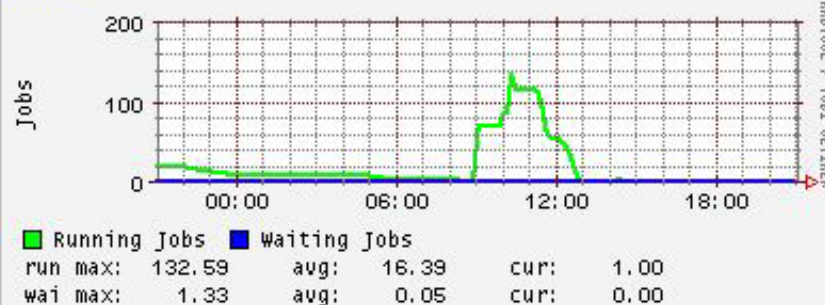


Site List

No	Site Reports	zone	GIIS Host	bdii	rgma	sanity	version	totalCPU	freeCPU	runJob	waitJob	seAvail TB	seUsed TB	maxCPU	avgCPU	DI
1	CAVENDISH-LCG2	pro,all	farm012.hep.phy.cam.ac.uk	ok		ok	LCG-2_1_1	11	5	5	832	1.57	0.02	11	10	lm
2	CERN-LCG2	pro,all	lxn1181.cern.ch	ok		ok	LCG-2_1_1	956	165	71	0	933.26	465.66	976	939	js
3	CIEMAT-LCG2	pro,all	lcg02.ciemat.es	ok		ok	LCG-2_1_1	2	1	1	0	0.05	0.00	2	1	ok
4	CNAF-LCG2	pro,all	wn-04-07-02-a.cr.cnaf.infn.it	ok		warn	LCG-2_2_0	1566	1479	93	0	94994.93	94994.90	1676	1436	ok
5	CYFRONET-LCG2	pro,all	zeus02.cyf-kr.edu.pl	ok		ok	LCG-2_1_1	24	5	19	18	1.28	0.71	24	22	ok
6	FNAL-LCG2	pro,all	hotdog46.fnal.gov	ok		warn	LCG-2_2_0	18	18	0	0			18	17	rm
7	IC-LCG2	pro,all	gw39.hep.ph.ic.ac.uk	ok		ok	LCG-2_2_0	66	36	30	0	0.08	0.32	66	63	wv

GIIS Monitor 21:14:45 09/25/04 GMT

home



GlueSEUniqueID:

```

castorftp-1.cr.cnaf.infn.it
castorftp-cms.cr.cnaf.infn.it
castorftp.cnaf.infn.it
diskserv-cms-2.cr.cnaf.infn.it
wn-04-01-36-a.cr.cnaf.infn.it
  
```

GlueServiceURI:

```

gridice ldap://wn-04-07-03-a.cr.cnaf.infn.it:2136/mds-vo-name=local,o=grid
  
```

TestZone tests reports for ce1.egee.f

History of results for site: ce1.egee.fr.cgg.com

Colours definition

Job list match failed	#ffcc39
Replica Management failed	#cc3cff
OK	#99ff99
Test job still waiting for execution	#ffff33
Job Submission failed (Job Manager)	#cc3c00
Wrong LCG version (too old)	#c0c0c0

Test date	Version	Software Version	BrokerInfo	CSH test	BDII LDAP
2004-09-25 07:05:02	LCG-2_2_0	LCG-2_2_0	OK	OK	ldap://lxn1189.cern.ch
2004-09-24 11:48:50	LCG-2_2_0	LCG-2_2_0	OK	OK	ldap://ce1.private.egee
2004-09-24 07:05:10	LCG-2_2_0	n/a	n/a	n/a	n/a
2004-09-23 07:05:55	LCG-2_2_0	LCG-2_2_0	OK	OK	ldap://ce1.private.egee
2004-09-22 07:05:33	LCG-2_2_0	n/a	n/a	n/a	n/a
2004-09-21 14:19:56	LCG-2_2_0	LCG-2_2_0	OK	OK	ldap://ce1.private.egee
2004-09-21 07:05:52	LCG-2_2_0	LCG-2_2_0	OK	OK	ldap://lxn1189.cern.ch
2004-09-20 07:05:29	FAILED	n/a	n/a	n/a	n/a
2004-09-19					

Site	URL	Test date	Version	Software Version	BrokerInfo	CSH test	BDII LDAP
CARLETONU-LCG2	lca02.physics.carleton.ca	2004-09-25 07:05:02	LCG-2_0_0beta	LCG-2_0_0	OK	OK	
CAVENDISH-LCG2	farm012.hep.phy.cam.ac.uk	2004-09-25 07:05:02	LCG-2_1_1	LCG-2_1_x	OK	OK	
CCIN2P3-LCG2	ccicqcel01.in2p3.fr	2004-09-25 07:05:02	LCG-2_2_0	LCG-2_2_0	OK	OK	
CERN-LCG2	lxn1184.cern.ch	2004-09-25 07:05:02	LCG-2_1_1	n/a	n/a	n/a	
CERN-LCG2	lxn1181.cern.ch	2004-09-25 07:05:02	LCG-2_1_1	n/a	n/a	n/a	
CGG-LCG2	ce1.egee.fr.cgg.com	2004-09-25 07:05:02	LCG-2_2_0	LCG-2_2_0	OK	OK	
CIEMAT-LCG2	lca02.ciemat.es	2004-09-25 07:05:02	LCG-2_1_1	LCG-2_1_x	OK	OK	
CNAF-LCG2	wn-04-07-01-a.cr.cnaf.infn.it	2004-09-24 11:48:50	LCG-2_2_0	LCG-2_2_0	OK	OK	
CNAF-LCG2	wn-04-07-02-a.cr.cnaf.infn.it	2004-09-25 07:05:02	LCG-2_2_0	LCG-2_2_0	OK	OK	

Site test report

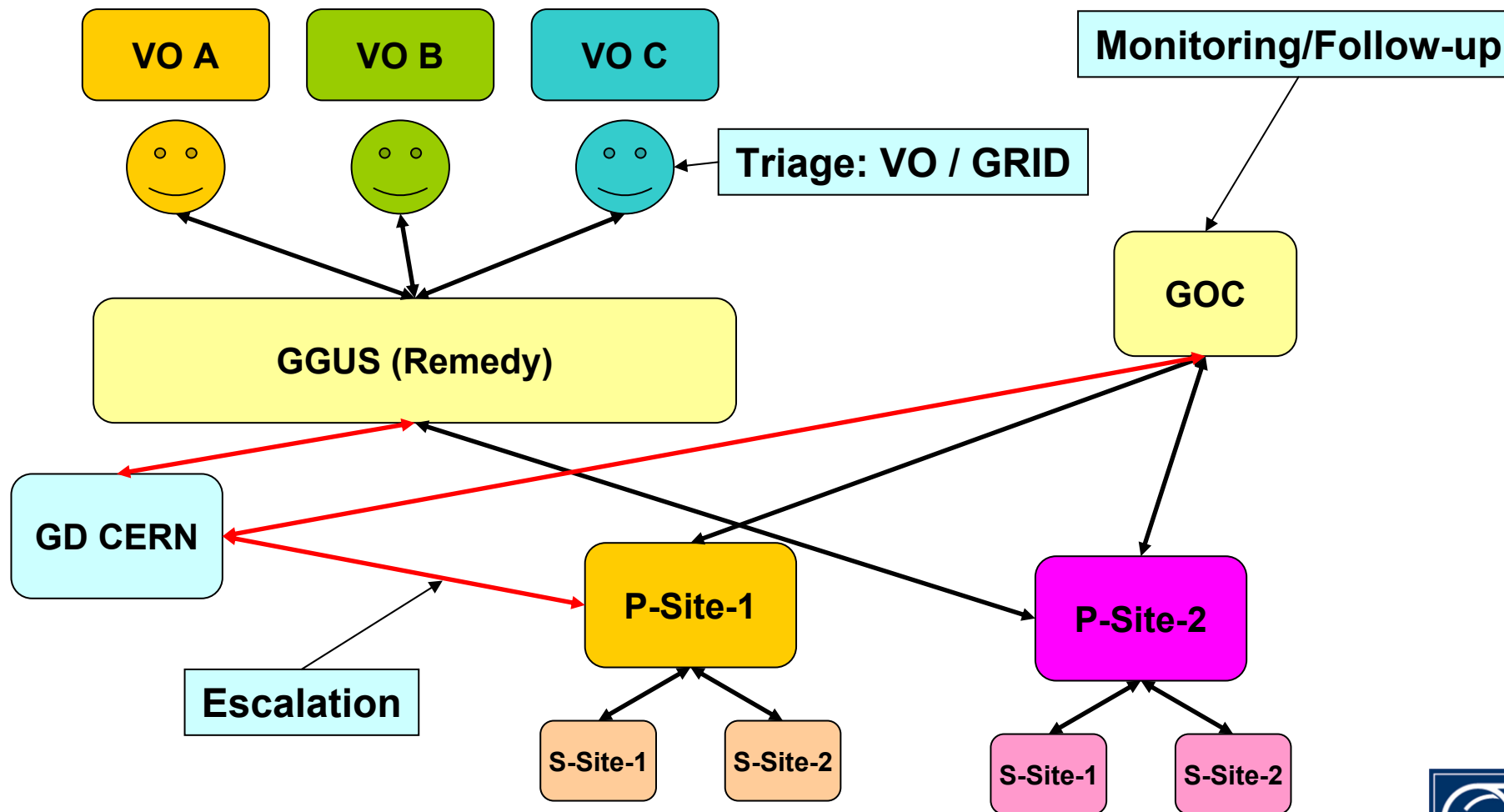
WN: kirin.physics.carleton.ca

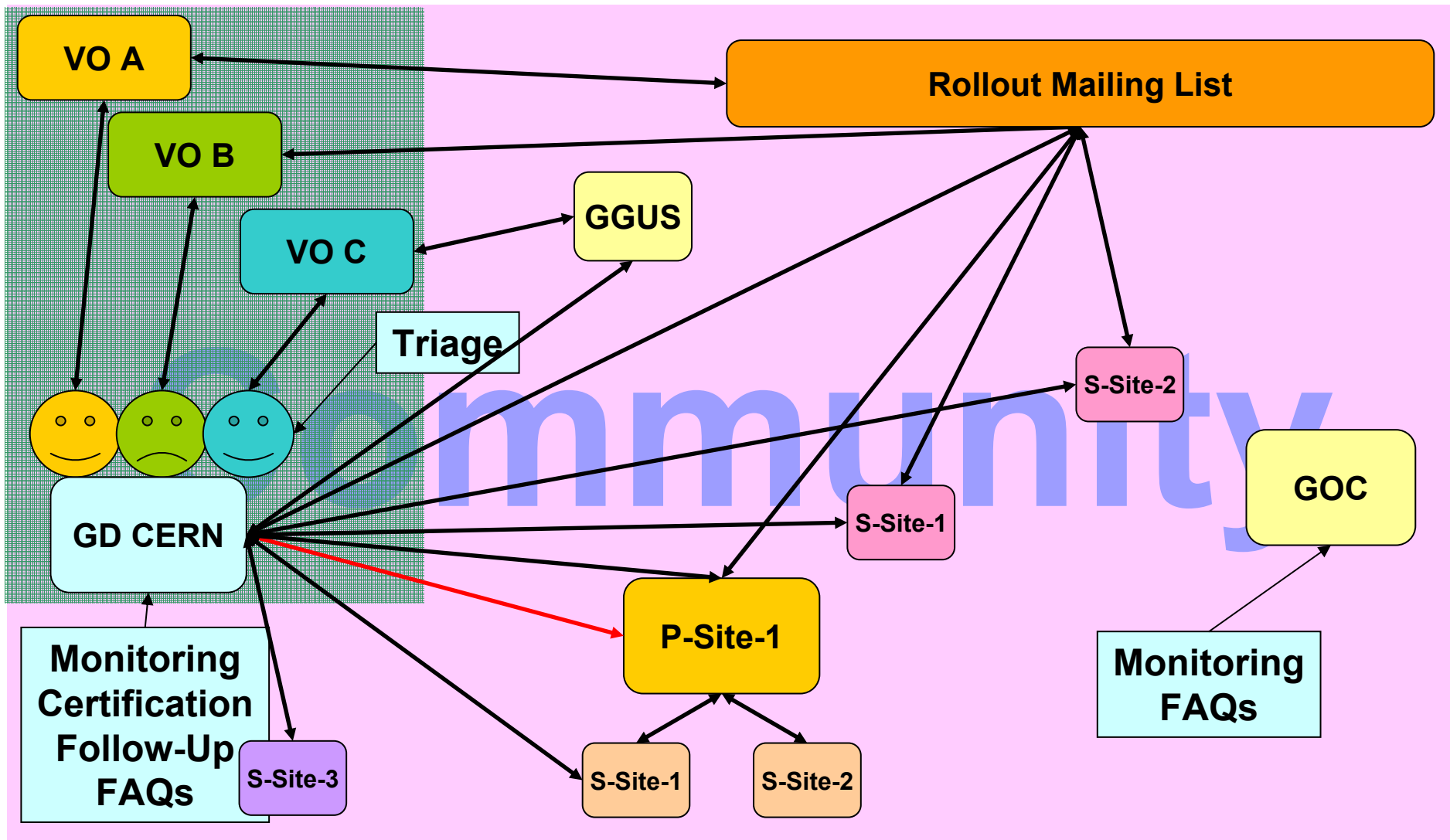
date: Sat Sep 25 03:54:55 EDT 2004

Table of contents

- Configuration of the Worker Node

Test	SE	Rep.	3rd Party cp	Delete Replica	GFAL sys	lcg-cr -> defaultSE	ld
Environment variables (bash)	OK	OK	OK	OK	OK	OK	OK
Checking if CSH works	OK	OK	OK	OK	OK	OK	OK
Difference between variables for bash and csh	OK	OK	OK	OK	OK	OK	OK
Software paths for the experiment	OK	OK	OK	OK	OK	OK	OK
Mountpoints on WN	OK	OK	OK	OK	OK	OK	OK
Accessibility of software paths	FAILED	FAILED	FAILED	OK	OK	OK	OK
Replica Manager configuration	FAILED	FAILED	FAILED	OK	OK	OK	OK
Installed RPMs list	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Installed software version	OK	OK	OK	OK	OK	OK	OK
Replica Manager tests	FAILED	FAILED	FAILED	FAILED	OK	OK	OK
Checking if we can see the SE at CERN	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Checking printinfo command	OK	OK	OK	OK	OK	OK	OK
Checking copyAndRegisterFile command	OK	OK	OK	OK	OK	OK	OK
Check copyFile command	OK	OK	OK	OK	OK	OK	OK
Checking replication to SE at CERN	OK	OK	OK	OK	OK	OK	OK
Checking 3rd party replication	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Checking copyFile for 3rd party replicated file	OK	OK	OK	OK	OK	OK	OK
Checking deleteFile command on replicated	OK	OK	OK	OK	OK	OK	OK
Replica Manager tests using LCG tools	OK	OK	OK	OK	OK	OK	OK
Checking information system (GFAL)	OK	OK	OK	OK	OK	OK	OK
Checking if default SE is defined	OK	OK	OK	OK	OK	OK	OK
Checking lcg-cr command	OK	OK	OK	OK	OK	OK	OK
Check lcg-cp command	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Checking lcg-rep to SE at CERN	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Checking 3rd party replication	OK	OK	OK	OK	OK	OK	OK
Checking lcg-cp for 3rd party replicated file	OK	OK	OK	OK	OK	OK	OK
Checking lcg-del command on replicated file	OK	OK	OK	OK	OK	OK	OK







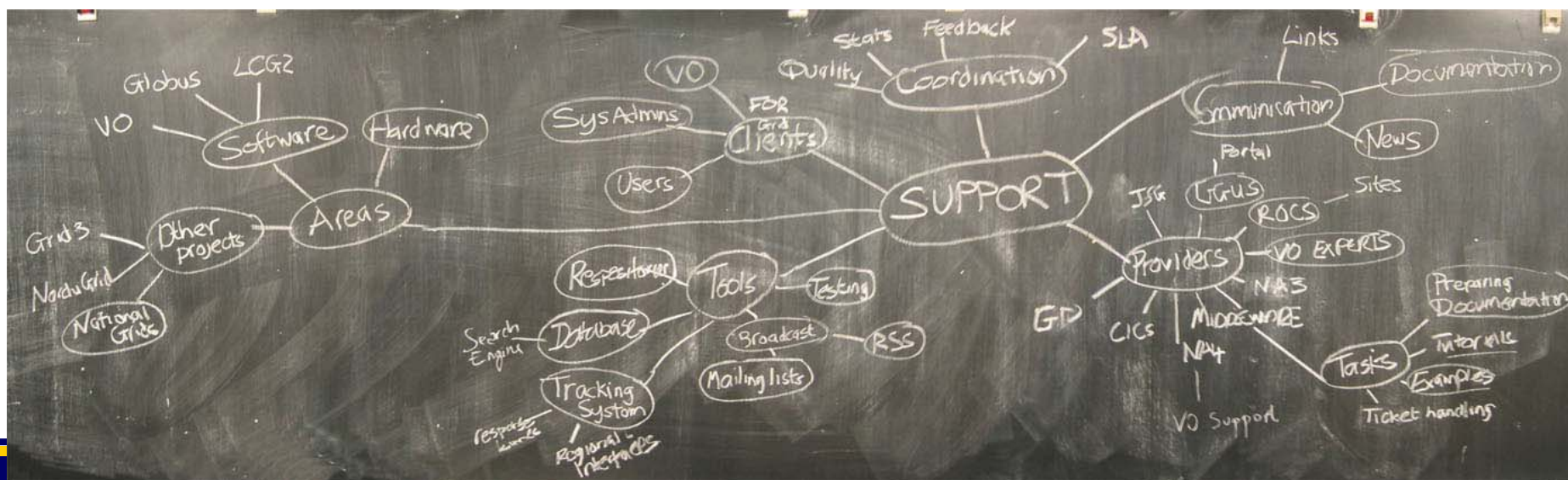
- Motivation
 - LCG -> (LCG&EGEE) transition requires changes
 - Lessons learned need to be implemented
 - Many different activities need to be coordinated
- 02 - 04 November at CERN
 - >80 participants including from GRID3 and NorduGrid
 - Agenda: [Here](#)
 - 1.5 days of plenary sessions
 - describe status and stimulate discussion
 - 1 day parallel/joint working groups
 - very concrete work,
 - results into creation of task lists with names attached to items
 - 0.5 days of reports of the WG



- Operational Security
 - Incident Handling Process
 - Variance in site support availability
 - Reporting Channels
 - Service Challenges
- Operational Support
 - Workflow for operations & security actions
 - What tools needed to implement the model
 - “24X7” global support
 - sharing operational load
 - Communication
 - Problem Tracking System
 - Defining Responsibilities
 - problem follow-up
 - deployment of new releases
 - Interface to User Support

- Fabric Management
 - System installations
 - Batch/scheduling Systems
 - Fabric monitoring
 - Software installation
 - Representation of site status (load) in the Information System
- Software Management
 - Operations on and for VOs (add/remove/service discovery)
 - Fault tolerance, operations on running services (stop, upgrades, re-starts)
 - Link to developers
 - What level of intrusion can be tolerated on the WNs (farm nodes)
 - application (experiment) software installation
 - Removing/Re-adding sites with troubles
 - Multiple views in the information system

- User Support
 - Defining what “User Support” means
 - Models for implementing a working user support
 - need for a Central User Support Coordination Team (CUSC)
 - mandate and tasks
 - distributed/central (CUSC/RUSC)
 - workflow
 - VO-support
 - continuous support on integrating the VOs software with the middleware
 - end user documentation
 - FAQs





- Very productive workshop
- Partners (sites) assumed responsibility for tasks
- Discussions very much focused on practical matters
- Some problems ask for architectural changes
 - gLite has to address these
- It became clear that not all sites are created equal
- Removing troubled sites is inherently problematic
 - removing storage can have grid wide impact
- Key issues in all aspects is to define split between:
 - Local,Regional and Central control and responsibility
- All WGs discussed communication





- LCG-2 services have been supporting the data challenges
 - Many middleware problems have been found – many addressed
 - Middleware itself is reasonably stable
- Biggest outstanding issues are related to providing and maintaining stable operations
- Future middleware has to take this into account:
 - Must be more manageable, trivial to configure and install
 - Management and monitoring must be built into services from the start on
- Operational Workshop has started many activities
 - Follow-up and keeping up the momentum is now essential
 - Indicates a clear shift away from the CERNtralized operation

