

Interoperabilità tra i progetti PON dell'Avviso 1575/2004

Documento Tecnico Operativo

(a cura del gruppo di lavoro tecnico operativo¹)

10/12/2007

1. TAG DI RUNTIME.....	3
DEFINIZIONE DELLE REGOLE PER LA PUBBLICAZIONE DEI TAG.....	3
TAG MIDDLEWARE.....	3
TAG SOFTWARE APPLICATIVO.....	3
TAG PON TOOL KIT.....	3
TAG SISTEMI OPERATIVI ED ARCHITETTURE.....	4
Tabella dei TAG da utilizzare per la variabile CE_RUNTIMEENV.....	4
Tabella dei TAG per i Sistemi Operativi da utilizzare nelle variabili d'ambiente CE_OS.....	6
Tabella delle variabili d'ambiente dei CE.....	8
Tabella delle variabili d'ambiente ti tipo SITE.....	9
2. QUEUE.....	10
3. SERVIZI COLLECTIVE CENTRALI E DISTRIBUITI.....	11
Servizi distribuiti e replicati nelle varie sedi.....	11
Servizi centralizzati.....	11
4. SERVIZI DI MONITORAGGIO.....	12
Strategia Generale.....	12
GridICE.....	12
GStat.....	12
SAM.....	13
Real-Time Monitoring	13
Real-Time Monitoring	13
HGSM (GOCdb).....	14
5. INSTALLAZIONE SOFTWARE APPLICATIVO.....	15
Software comune.....	15
Utenti SGM e PRD.....	15
6. INTERFACCE E PROTOCOLLI PER LO STORAGE.....	15
Protocolli di trasferimento.....	15
Interfacce SRM e configurazioni.....	15
7. ABILITAZIONE DELLE VO COMUNI.....	15
8. ACCOUNTING.....	16
Dgas.....	16
HLR.....	16
9. GESTIONE DI TICKET.....	17
10. SERVICE LEVEL AGREEMENT.....	17
11. INTEROPERABILITA' CON ALTRE INFRASTRUTTURE.....	17

¹ S. Pardi, F. Serio, M. Scognamiglio, D. Bottalico, V. Boccia, G. Tortone, R. Catania, G. Platania, G.M. Ricciardi, G. Passaro, A. Falzone, G. Bracco, C. Scio, A. Santoro, D. Mura, G. Mereu, con la collaborazione di R. Barbera.

1. TAG DI RUNTIME

DEFINIZIONE DELLE REGOLE PER LA PUBBLICAZIONE DEI TAG

Per garantire la piena interoperabilità tra le infrastrutture Grid dei progetti dell'Avviso 1575/2004, occorre uniformare le informazioni pubblicate dai CE relativamente al sito e alle risorse e le informazioni di run-time, al fine di permettere all'utente finale di esprimere *constraint* e *requirement* sulle risorse nei propri JDL.

Si propongono quindi le seguenti convenzioni da adottare:

TAG MIDDLEWARE

Dal punto di vista generale i middleware con le loro versioni saranno indicati utilizzando il nome del middleware maiuscolo seguito da un segno meno (-) e dalla versione utilizzando l'underscore (_) come simbolo di separazione tra i numeri della versione:

NOMEMIDDLEWARE-X_Y_Z

Esempi sono LCG-2_1_0, GLITE-3_0_0

TAG SOFTWARE APPLICATIVO

Per i software applicativi e per le librerie si propone di utilizzare il seguente schema: nome del pacchetto o libreria in maiuscolo seguito da un segno meno (-) e dalla versione utilizzando questa volta il punto (.) come simbolo di separazione tra i numeri della versione:

NOMETOOL-X.Y.Z

Esempi sono IDL-6.4, ABAQUS-6.7

TAG PON TOOL KIT

Per i PON TOOL KIT che rappresentano l'insieme dei software necessari per l'esecuzione dei job degli utenti dei differenti PON (vedi sezione INSTALLAZIONE SOFTWARE APPLICATIVO) si propone l'utilizzo di 4 tag così composti:

CRESCO-TOOL-KIT-x.y

CYBERSAR-TOOL-KIT-x.y

PI2S2-TOOL-KIT-x.y

SCOPE-TOOL-KIT-x.y

Per quanto riguarda i singoli software componenti i 4 tool kit ed installati mediante utenti sgm, si propone l'utilizzo della seguente nomenclatura utilizzata dai tool di installazione lcg-asis.

VO-<VO_NAME>-<NOME_TOOL>

Ad esempio la libreria SAXON-8.7.3 del tool kit di SCOPE

TAG SISTEMI OPERATIVI ED ARCHITETTURE

Per i sistemi operativi è opportuno utilizzare tuttavia la modalità *case sensitive* (anche per uniformarsi alle altre infrastrutture dove ad esempio si usa ScientificCERN/SLC). Vale tuttavia per le versioni la regola del punto come separatore così come utilizzato per i software e le librerie.

Si propone inoltre di segnalare il tipo di architettura i386, i586, i686 o x86_64 sia nel TAG **CE_RUNTIMEENV** che nel TAG **CE_OS_ARCH**.

Di seguito proponiamo una serie di TAG da utilizzare per risolvere in maniera inequivocabile la presenza o meno di un particolare tool o libreria validi sia per le variabile **CE_RUNTIMEENV** che per le altre variabili d'ambiente del CE (ad esempio **CE_OS**, **CE_ARCH**).

Tabella dei TAG da utilizzare per la variabile CE_RUNTIMEENV

Di seguito la tabella dei TAG attualmente concertati tra i vari PON, tale elenco è comunque soggetto a successivi aggiornamenti per l'aggiunta di eventuali componenti software o di architettura

TAG	SPECIFICA
MIDDLEWARE	
LCG-2	Supporta versione del middleware LCG-2
LCG-2_1_0	Supporta versione del middleware LCG-2_1_0
LCG-2_1_1	Supporta versione del middleware LCG-2_1_1
LCG-2_2_0	Supporta versione del middleware LCG-2_3_0
LCG-2_3_0	Supporta versione del middleware LCG-2_3_0
LCG-2_3_1	Supporta versione del middleware LCG-2_3_1
LCG-2_4_0	Supporta versione del middleware LCG-2_4_0
LCG-2_5_0	Supporta versione del middleware LCG-2_5_0
LCG-2_6_0	Supporta versione del middleware LCG-2_6_0
LCG-2_7_0	Supporta versione del middleware LCG-2_7_0
GLITE-3_0_0	Supporta versione del middleware GLITE-3_0_0
GLITE-X_Y_Z	Supporta versione del middleware GLITE-X_Y_Z
R-GMA	Supporta R-GMA
ANAGRAFICA	
CITTA' (XXXX)	Città dove è situato il sito
PROJECT-NAME (XXXX)	Nome del progetto
SITE-NAME (XXXX)	Nome del sito
LIBRERIE E SOFTWARE	
MPICH	Libreria MPICH
MPICH2	Libreria MPICH versione 2
MVAPICH	Libreria MVAPICH
MVAPICH2	Libreria MVAPICH2
MPI_HOME_SHARED	Architettura MPI con directory Shared tra i worker node
MPI_HOME_NOTSHARED	Architettura MPI con directory NON Shared tra i worker node

IDL-X.Y	Supporto per IDL versione X.Y
ABAQUS-X.Y	Supporto per ABAQUS versione X.Y
PON TOOL KIT	
CRESCO-TOOL-KIT-x.y	Supporto per il tool kit del progetto CRESCO
CYBERSAR-TOOL-KIT-x.y	Supporto per il tool kit del progetto CYBERSAR
PI2S2-TOOL-KIT-x.y	Supporto per il tool kit del progetto PI2S2
SCOPE--TOOL-KIT-x.y	Supporto per il tool kit del progetto SCOPE
ARCHITETTURE	
SI00MeanPerCPU=xxxx	Valore medio del Benchmarkt SI00 tra tutti i WN
SF00MeanPerCPU=yyyy	Valore medio del Benchmarkt SF00 tra tutti i WN
i386	Architettura i386
i686	Architettura i686
X86_64	Architettura a 64 bit x86_64
IA64	Architettura Itanium a 64
PowerPC_G4	Architettura G4
PowerPC_G5	Architettura G5
NETWORK	
INFINIBAND-1X	Rete in infiniband tra i Worker Node 1X
INFINIBAND-4X	Rete in infiniband tra i Worker Node 4X

Tabella dei TAG per i Sistemi Operativi da utilizzare nelle variabili d'ambiente CE_OS

Di seguito la tabella riportata all'indirizzo

http://goc.grid.sinica.edu.tw/gocwiki/How_to_publish_the_OS_name per la definizione dei valori da impostare per le variabili d'ambiente pubblicate dal CE, relative al sistema operativo in uso.

SISTEMA OPERATIVO	CE_OS	CE_OS_RELEASE	CE_OS_VERSION
CentOS 3.5	CentOS	3.5	Final
CentOS 3.6	CentOS	3.6	Final
CentOS 3.7	CentOS	3.7	Final
CentOS 3.8	CentOS	3.8	Final
CentOS 4.2	CentOS	4.2	Final
CentOS 4.5	CentOS	4.5	Final
Debian sarge	Debian	3.1	sarge
Debian etch	Debian	4.0	etch
FedoraCore 4	FedoraCore	4	Stentz
Gentoo 2006.0	Gentoo	2006.0	n/a
RedHat Enterprise	RedHatEnterprise	x.y.z	xxxx
RHEL AS3 Taroon Update 6	RedHatEnterpriseAS	3	TaroonUpdate6
RHEL AS4 Nahant Update 3	RedHatEnterpriseAS	4	NahantUpdate3
Rocks Linux 3.3	linux-rocks-3.1	Rocks Linux	?
Rocks Linux 4.1	linux-rocks-4.1	Rocks Linux	?
ScientificLinux 3.0.2	Scientific Linux	3.0.2	SL
ScientificLinux 3.0.3	Scientific Linux	3.0.3	SL
ScientificLinux 3.0.4	Scientific Linux	3.0.4	SL
ScientificLinux 3.0.5	Scientific Linux	3.0.5	SL
ScientificLinux 3.0.6	Scientific Linux	3.0.6	SL
ScientificLinux 3.0.7	Scientific Linux	3.0.7	SL
ScientificLinux 3.0.8	Scientific Linux	3.0.8	SL
ScientificLinux 3.0.9	Scientific Linux	3.0.9	SL
ScientificLinux 4.2	ScientificSL	4.2	Beryllium
ScientificLinux 4.4	ScientificSL	4.4	Beryllium
ScientificLinux 4.5	ScientificSL	4.5	Beryllium
ScientificLinuxCern 3.0.4	Scientific Linux CERN	3.0.4	SL
ScientificLinuxCern 3.0.5	Scientific Linux CERN	3.0.5	SL
ScientificLinuxCern 3.0.6	Scientific Linux CERN	3.0.6	SL
ScientificLinuxCern 3.0.8	Scientific Linux CERN	3.0.8	SL
ScientificLinuxCern 4.3	ScientificCERNSLC	4.3	Beryllium
ScientificLinuxCern 4.4	ScientificCERNSLC	4.4	Beryllium

ScientificLinuxCern 4.5	ScientificCERN	SLC 4.5	Beryllium
Suse 9	SUSE LINUX	9	n/a
Suse 9.3	SUSE LINUX	9.3	n/a
SUSE Linux Enterprise Server 10 (ia64)	SUSE LINUX	10	n/a
openSuse 10.2	SUSE LINUX	10.2	n/a
Ubuntu 5.10	Ubuntu	5.10	breezy
AIX 5.2	AIX	5.2	AIX
Windows XP	Microsoft WINDOWS	XP Professional	SP2

Tabella delle variabili d'ambiente dei CE

Di seguito è riportata una tabella con le variabili d'ambiente da settare sui CE per una corretta integrazione:

VARIABILE	SPECIFICA
BATCH_BIN_DIR	Path del comando LRMS
BATCH_LOG_DIR	Log directory del Batch System
BATCH_VERSION	Versione del LRMS
BATCH_SERVER	Host che offer funzionalità di LRMS
CE_BATCH_SYS	Tipo di Batch System esistente su CE
CE_CPU_MODEL	Modello di Cpu usato dai WNs
CE_CPU_SPEED	Frequenza di Clock in Mhz
CE_CPU_VENDOR	Venditore di Cpu su WNs
CE_HOST	Hostname del Computing Element
CE_INBOUNDIP	True: se la connessione in entrata è abilitata False: se la connessione in entrata non è abilitata
CE_MINPHYSMEM	Dimensione della RAM in Mbyte per WN (non Cpu)
CE_MINVIRTMEM	Dimensione della Memoria Virtuale in Mbyte per WN (non Cpu)
CE_OS	Nome del Sistema Operativo
CE_OS_RELEASE	Release del Sistema Operativo
CE_OS_VERSION	Versione del Sistema Operativo
CE_OS_ARCH	Architettura del sistema operativo (uname -m)
CE_OUTBOUNDIP	True: se la connessione in uscita è abilitata False: se la connessione in uscita non è abilitata
CE_SF00	Indice di performance di fabbricazione in SpecFloat 2000
CE_SI00	Indice di performance di fabbricazione in SpecInt 2000
CE_SMPSIZE	Numero di Cpu per WN
QUEUES	Nomi delle code del Computing Element
<QUEUE-NAME>_GROUP_ENABLE	Lista di VO names and VOMS FQANs che sono abilitate all'accesso alle code

Tabella delle variabili d'ambiente ti tipo SITE

Di seguito è riportata una tabella con le variabili d'ambiente di sito da impostare per facilitarne l'individuazione geografica e la raggiungibilità.

VARIABILE	SPECIFICA
SITE_EMAIL	Email del sito
SITE_SUPPORT_EMAIL	Email per il supporto
SITE_NAME	Nome del sito
SITE_LOC	Località nella forma: città, paese
SITE_LAT	Latitudine del sito con almeno 5 cifre decimali
SITE_LONG	Longitudine del sito con almeno 5 cifre decimali
SITE_WEB	Sito web del progetto o del sito
SITE_SUPPORT_SITE	Support Entry Point – Unique Id per il sito nel GOC DB e information system

2. QUEUE

Al fine di tener conto delle esigenze dei vari progetti si propone di utilizzare tre code job per ogni progetto, **short, long e infinite** con differenti *policy* di CPU TIME e WALL TIME e differenti priorità. A tali code si aggiungono le seguenti code di certificazione SAM: **poncert** per i job di certificazione comuni, più 4 code, per i job di certificazione specifici dei 4 progetti, **crescocert, cybrcert, pi2s2cert, scopecert**.

Nome Coda	TIPO	CPUTIME (minuti)	WALLTIME (minuti)	PRIORITY
jobmanager-<lsf/pbs>-poncert	Coda di Certificazione	2880	4320	Prioritaria
jobmanager-<lsf/pbs>-crescocert	Coda di Certificazione	2880	4320	Prioritaria
jobmanager-<lsf/pbs>-cybrcert	Coda di Certificazione	2880	4320	Prioritaria
jobmanager-<lsf/pbs>-pi2s2cert	Coda di Certificazione	2880	4320	Prioritaria
jobmanager-<lsf/pbs>-scopecert	Coda di Certificazione	2880	4320	Prioritaria
jobmanager-<lsf/pbs>-cresco_short	Coda Job Cresco	15	120	Alta
jobmanager-<lsf/pbs>-cresco_long	Coda Job Cresco	720	1440	Media
jobmanager-<lsf/pbs>-cresco_infinite	Coda Job Cresco	2880	4320	Bassa
jobmanager-<lsf/pbs>-cybr_short	Coda Job Crybersar	15	120	Alta
jobmanager-<lsf/pbs>-cybr_long	Coda Job Crybersar	720	1440	Media
jobmanager-<lsf/pbs>-cybr_infinite	Coda Job Crybersar	2880	4320	Bassa
jobmanager-<lsf/pbs>-pi2s2_short	Coda Job PI2S2	15	120	Alta
jobmanager-<lsf/pbs>-pi2s2_long	Coda Job PI2S2	720	1440	Media
jobmanager-<lsf/pbs>-pi2s2_infinite	Coda Job PI2S2	2880	4320	Bassa
jobmanager-<lsf/pbs>-scope_short	Coda Job SCoPE	15	120	Alta
jobmanager-<lsf/pbs>-scope_long	Coda Job SCoPE	720	1440	Media
jobmanager-<lsf/pbs>-scope_infinite	Coda Job SCoPE	2880	4320	Bassa

3. SERVIZI COLLECTIVE CENTRALI E DISTRIBUITI

Per l' implementazione dei servizi collective, si propone che alcuni di essi siano replicati nelle varie sedi al fine di garantire l'autonomia la ridondanza e la massima robustezza dell'infrastruttura, altri servizi meno critici saranno invece centralizzati.

In particolare si propone:

Servizi distribuiti e replicati nelle varie sedi

Saranno replicati nelle vari sedi i seguenti servizi essendo cruciali per il corretto funzionamento delle singole infrastrutture.

- VOMS
- RB/BDII
- LFC
- TICKET

In particolare, ogni sito gestisce la propria VO sul proprio VOMS server ed abilita sul proprio RB le VO dei PON cresco, cybersar, cometa (pi2s2) e scope.

Servizi centralizzati

Per evitare un eccessivo numero di query ai BDII, si propone di centralizzare i Servizi di monitoraggio incaricando ciascuno dei quattro PON di gestire uno o più tools per conto di tutta la collaborazione, a partire dai software disponibili e stabili come GridICE, SAM, RTM ed altri.

Tale strategia consente di avere a disposizione tutte le possibili vedute offerte dai vari servizi di monitoraggio senza caricare ogni sito di dover implementare tutte le possibili .

I differenti tools di monitoraggio sono discussi nella sezione **SERVIZI DI MONITORAGGIO** di questo documento.

4. SERVIZI DI MONITORAGGIO

Strategia Generale

La strategia per i servizi di monitoraggio è quella di creare per ogni servizio un unico punto di accesso, dando a ciascun progetto il compito di gestire uno specifico tool.

Di seguito sono riportati i tools che si intende supportare per l'interoperabilità, e la strategia di deployment.

- Web LDAP Browser - Da installarsi sui BDII server di ogni sito
- GridICE - Centrale gestito dal progetto SCoPE
- GStat - Centrale gestito dal server goc.grid.sinica.edu.tw
- SAM +VO poncert - Centrale gestito dal progetto CYBERSAR
- Down Time-SYSTEM - Centrale gestito dal progetto PI2S2
- Real-Time Monitoring - Centrale registrandosi presso il database dell'Imperial College
- HGSM (GOCdb) – Centrale, utilizzare quello del CNAF

Web LDAP Browser

Descrizione: Servizio per consultare i BDII di ogni sito via web.

Si propone che ogni sito installi un web LDAP Browser sui propri top BDII per consentirne la rapida e semplice consultazione

Esempio di servizio web LDAP Browser

<http://scoperb01.dsf.unina.it>

GridICE

Descrizione: Tool di monitoring distribuito per Grid.

Si propone che ogni sito abiliti tutti i servizi per la pubblicazione delle informazioni su GridICE sulla porta ldap 2136. Tali informazioni saranno utilizzate da un servizio GridICE centrale.

Esempio di servizio Gridice

<http://scopegridice01.dsf.unina.it>

GStat

Descrizione: Tool per la creazione di statistiche e monitoraggio delle installazioni dei siti operativi.

Per il servizio GStat si propone di utilizzare il server goc.grid.sinica.edu.tw

Esempio ed indirizzo del server

<http://goc.grid.sinica.edu.tw/gstat/>

SAM

Descrizione: Tool per la certificazione dei servizi

Si propone l'installazione di SAM come tool di certificazione in un server centrale. A tal fine tutti i progetti dovranno creare sui propri CE delle code privilegiate con priorità maggiore delle altre, per l'esecuzione di Job di test da eseguire mediante il servizio SAM.

Deployment

1. Tutti i progetti abilitano una coda poncert sulle loro macchine, ed una coda di certificazione specifica del progetto: crescocert, cybrcert, pis2scert e scopecert, tali code avranno priorità maggiore delle restanti code job.
2. Si definisce una VO poncert unica sul server VOMS del progetto cybersar, server centrale nel quale vengono registrati un numero limitato di utenti privilegiati dei vari progetti.
3. Tali utenti vengono mappati su utenti locali di tipo poncert01, poncert02 e così via su tutti i progetti.
4. La coda poncert sarà accessibile solo agli utenti registrati alla VO poncert e verrà utilizzata per la sottomissione dei job di certificazione comuni, che dovranno essere stabiliti anche in funzione del Service Level Agreement.
5. Le code di certificazione di progetto, verranno utilizzate per l'esecuzione di job di certificazione specifici creati dai singoli progetti.

Esempio di servizio SAM

<https://sam-cybr.ca.infn.it/sam/sam.py>

Real-Time Monitoring

Descrizione: Tool per la pubblicazione dei periodi di sospensione dei servizi da parte dei siti (Down –Time)

Si propone di creare un unico punto dove vengano pubblicati i periodi downtime di tutti i progetti.

Esempio di servizio di Down-Time

<http://trigrid-advice.trigrid.it/support/calendar/>

Real-Time Monitoring

Descrizione: Tool di monitoraggio basato su immagini satellitari per la collocazione geografica dei siti distribuiti.

Si propone di registrare i siti nel database del servizio RTM dell'Imperial College (UK) al sito:

<http://gridportal.hep.ph.ic.ac.uk/rtm/>

HGSM (GOCdb)

5. INSTALLAZIONE SOFTWARE APPLICATIVO

Software comune

Il software comune dovrà essere selezionato per aree tematiche e dovrà comprendere le librerie ed i tool necessari per eseguire le applicazioni che si intende portare in interoperabilità.

A tal fine è necessario individuare le applicazioni e il software necessario per garantire il corretto funzionamento dei job utente, sulla base di questo lavoro saranno quindi creati dei TOOL_KIT tematici che dovranno essere installati su tutte le macchine che partecipano all'interoperabilità.

Tali toolkit saranno preparati per essere installati tramite job sfruttando gli utenti sgm

Utenti SGM e PRD

Il software di progetto che comprende librerie e tool applicativi, verrà installato nelle aree /opt/exp_soft delle singole infrastrutture.

Ciascun progetto dovrà abilitare sulle proprie risorse, un numero limitato (uno, al massimo due) di utenti sgm per le VO cresco, cybersar, cometa (pi2s2) e scope, in modo di garantire che ciascun progetto possa installare ed aggiornare i propri toolkit sulle altre infrastrutture.

6. INTERFACCE E PROTOCOLLI PER LO STORAGE

Protocolli di trasferimento

Per mantenere il sistema di autenticazione basato sui certificati, si propone che i progetti abilitino il protocollo di trasferimento grftp per il trasferimento dati.

Interfacce SRM e configurazioni

Per garantire piena compatibilità ed interoperabilità con le altre strutture internazionali si propone che gli Storage Element impiegati per l'interoperabilità dai progetti utilizzino interfaccia SRM supportando **srnv1.1** ed **srnv2.2**.

Ogni progetto abiliterà sugli storage element impegnati nell'interoperabilità, una quantità di spazio disco da definire creando le seguenti aree logiche: **/home/cresco**, **/home/cybersar**, **/home/pi2s2**, **/home/scope**.

7. ABILITAZIONE DELLE VO COMUNI

.....

8. ACCOUNTING

Dgas

HLR

9. GESTIONE DI TICKET

Si propone che ogni progetto utilizzi il proprio sistema di ticketing locale con il requirement che tale sistemi siano interoperabili tra loro.

Per la gestione dei ticket interprogetto, supponendo che un utente del progetto X abbia problemi con il progetto Y ci sono due possibilità da indagare:

Prima strategia

L'utente del progetto X apre ticket su una area dedicata del sistema di ticketing locale, quindi l'amministratore esegue il forwarding del ticket sul sistema di supporto del progetto Y.

Seconda strategia

L'utente del progetto X apre un ticket sul sistema di ticketing del progetto Y.

I sistemi di ticketing devono prevedere almeno 3 aree per la suddivisione dei ticket:

- **SITE** – per i problemi tecnici di installazione
- **VO** – per le questioni relative alla singola virtual organization, certificazioni etc.
- **APPLICATION** – per il supporto delle applicazioni.

10. SERVICE LEVEL AGREEMENT

.....

11. INTEROPERABILITA' CON ALTRE INFRASTRUTTURE

.....