

Accounting in EGEE/WLCG

◆ Pablo Rey, Javier López, Carlos Fernández

Resumen

La contabilidad de los recursos en computación distribuida y Grid tiene una gran importancia, ya que permite analizar cual es el uso que se está dando a la infraestructura, las necesidades de los usuarios y la calidad de su utilización, en el sentido de aprovechamiento de los recursos. Una de las mayores infraestructuras de contabilidad de este tipo se encuentra en el proyecto EGEE/WLCG, el cual engloba más de 240 sitios situados en más de 50 países diferentes y con más de 6.000 usuarios registrados. La arquitectura de contabilidad utilizada está basada en R-GMA y desarrollada en código abierto. Los registros de uso se generan a través de una aplicación denominada APEL que se encarga de recoger los datos y almacenarlos en una base de datos centralizada cifrando la información que pueda ser de carácter personal.

Posteriormente estos datos son procesados y analizados, lo cual permite reducir considerablemente el tamaño de la información. A partir de estos resultados se generan estadísticas que están disponibles a través de un portal web que permite diferentes niveles de acceso a los datos dependiendo del rol asignado a la persona que esté accediendo a los datos. Actualmente se está trabajando en añadir a esta infraestructura la contabilidad del uso de otros recursos como el almacenamiento.

Palabras clave: contabilidad, Grid, EGEE.

Summary

Resource accounting in distributed and Grid computing is very important, since it makes it possible to analyse the use made of the infrastructure, the needs of the users and the quality of use, in terms of taking advantage of the resources. One of the largest accounting infrastructures of this type is the EGEE/WLCG project, which covers over 240 sites located in over 50 different countries, with over 6,000 registered users. The accounting architecture used is based on R-GMA and developed in open source. The usage records are generated through an application called APEL, which collects the data and stores them in a centralised database, encoding any personal information. The data are then processed and analysed, which reduces the amount of information considerably. Based on these results, statistics are generated and published on a website that offers different levels of access based on the role assigned to the person accessing the data. Work is currently underway to add the use of other resources, such as storage, to this accounting infrastructure.

Keywords: accounting, Grid, EGEE.

1. Introducción

La infraestructura Grid de EGEE / WLCG[1, 2] ha alcanzado más de 40.000 CPUs, además en cuanto al almacenamiento, una media de 13 Petabytes (13 millones de Gigabytes) se están usando de 16 Petabytes disponibles, y mantiene 20.000 trabajos concurrentes de media. Hay más de 6.000 usuarios distribuidos registrados a más de 100 Organizaciones Virtuales (VOs). Con lo cual, es evidente que esta creciente infraestructura requiere una adecuada contabilidad con el fin de saber cómo se han utilizado los recursos y por quién, y proporcionar información para una asignación eficaz de recursos. La actividad de contabilidad recoge los datos de contabilidad de todos los sitios participantes en las infraestructuras de EGEE y WLCG, así como de los sitios pertenecientes a otras organizaciones Grid que están colaborando con EGEE. Entre otros, se incluyen el Open Science Grid (OSG), Nordugrid, INFN-Grid y GridPP.

Dentro de estos Grids, se han desarrollado diferentes sensores para recoger la información de contabilidad de los recursos usados por los trabajos enviados por los usuarios.

La información de contabilidad es enviada a una base de datos central (GOC) donde es procesada para generar resúmenes estadísticos que están disponibles a través del portal de contabilidad de EGEE/WLCG.

Estas estadísticas están disponibles para su uso, en diferentes vistas, por: todas las comunidades científicas.



Hay más de 6.000 usuarios distribuidos registrados a más de 100 Organizaciones Virtuales (VOs)



Es evidente que esta creciente infraestructura requiere una adecuada contabilidad con el fin de saber cómo se han utilizado los recursos y por quién, y proporcionar información para una asignación eficaz de recursos



◆
En EGEE/WLCG, hay numerosas herramientas disponibles para recoger la información sobre los recursos usados

◆
Con el fin de proteger la identidad del usuario y de mantener la confidencialidad, APEL realiza el cifrado del UserDN antes de enviar los datos desde el sitio hasta la base de datos centralizada utilizando R-GMA

usando las mencionadas infraestructuras Grid, los propietarios y los administradores de los recursos, y los gestores de las infraestructuras Grid. Por ejemplo, la CPU total consumida por cada VO.

Las estadísticas disponibles a través del portal de contabilidad enriquecen nuestra comprensión de la utilización de los recursos por las diferentes VOs y usuarios/sitios. Pueden ser utilizadas para identificar cómo los recursos, tales como la memoria, el disco y la CPU, pueden ser asignados mejor para llevar a cabo las diferentes tareas científicas. Pueden aportar información, y se utilizarán para verificar los Service Level Agreements (SLAs). Técnicas de contabilidad apropiadas pueden también ayudar a descubrir y seguir abusos o problemas de configuración de la infraestructura.

2. Arquitectura

En EGEE/WLCG, hay numerosas herramientas disponibles para recoger la información –tanto en tiempo real como después del evento– sobre los recursos usados. Algunas ampliamente utilizadas, pero que no se tratan en este documento, son DGAS (INFN-Grid), Gratia (OSG) y SGAS (SweGrid). La mayor parte de los datos de contabilidad de EGEE son recogidos usando APEL, el cual utiliza R-GMA para transportar los datos a través de la red hasta un repositorio central.

2.1. Accounting Processor Event Logs (APEL)

APEL es una aplicación de procesamiento de logs la cual es utilizada para interpretar los ficheros de log del sistema (por ejemplo, los logs del gatekeeper y del sistema de colas) para producir registros de contabilidad.

Actualmente soporta los sistemas de colas PBS, LSF, SGE y Condor pero puede ser extendido para soportar otras variantes. APEL es un sistema de contabilidad simple que recoge información sobre el uso después que el trabajo ha finalizado.

APEL utiliza un sencillo modelo de cuatro capas:

- Recopilación de datos: procesamiento de los ficheros de log del sistema y creación de los registros de contabilidad.
- Transporte de datos: envío de los datos a una base de datos central.
- Procesamiento de datos: agregación y consolidación de los datos en las tablas de resumen.
- Visualización de datos: informes a partir de los datos.

Esta aproximación permite a otras herramientas de contabilidad enviar los datos recogidos a una base de datos central donde pueden ser agregados y mostrados en el portal de contabilidad.

Con el fin de proteger la identidad del usuario y de mantener la confidencialidad, APEL realiza sobre la marcha el cifrado del UserDN antes de enviar los datos desde el sitio hasta la base de datos centralizada utilizando R-GMA[3].

Una clave privada-pública asimétrica 1024-bit RSA fue generada. La clave pública se utiliza para realizar el cifrado, mientras que la clave privada se usa para descifrar los datos en el GOC.

Como muchos trabajos pueden pertenecer a un solo usuario, una función que añade un texto aleatorio es aplicada durante el cifrado para reducir la probabilidad de que los patrones repetibles surjan en el texto cifrado.

APEL se desplegó en el Grid de producción de WLCG a finales de 2005. Ha recogido más de 40 millones de registros de trabajos, ha registrado un pico de actividad de más de 180 sitios publicando datos, y un rendimiento por encima de un millón de registros publicados en la base de datos central en un solo día.

2.2. R-GMA

La recopilación de los registros de contabilidad se realiza a través de R-GMA, una implementación de la Grid Monitoring Architecture (GMA) propuesta por el Global Grid Forum (GGF). GMA modela la infraestructura de los sistemas de información y monitorización de un Grid como un conjunto de consumidores (que requieren información), productores (que proporcionan la información), un

registro, que conecta los productores y los consumidores, y un esquema que define la estructura de la información intercambiada entre los productores y los consumidores.

3. Portal de contabilidad

El portal de contabilidad de EGEE[4] es la interfaz gráfica de usuario de los datos de contabilidad de EGEE. Las estadísticas de contabilidad están disponibles a través de este portal para el análisis de los diferentes usuarios del Grid, administradores de VOs y administradores de sitios.

3.1. Información mostrada

Los cantidades disponibles son entre otras:

- El número de trabajos finalizados.
- El tiempo de CPU consumido. Esta cantidad está disponible en forma pura (según lo reportado por el sistema de colas) y normalizada utilizando una escala de referencia de 1000 SpecInt2000.
- El tiempo de espera consumido, disponible tanto en forma pura como normalizada.
- Eficiencia de CPU definida como $SUM(CPU) / SUM(WCT)$.

Todas estas cantidades pueden ser agrupadas en función de VO, sitio o tiempo.

3.2. Casos de uso

Se utiliza un árbol para ordenar los distintos sitios permitiendo fácilmente navegar por Tier del proyecto LCG, país o Grid (EGEE o OSG) e incluso se pueden seleccionar los sitios en función de si están en Producción o Pre-producción.

Los sitios que se muestran en la rama del árbol correspondiente a EGEE están certificados o suspendidos y, en la infraestructura de Producción o Pre-Producción (PPS) con uno o más Computing Element (CE) registrados en el Grid Operations Centre Database (GOODB). Las agrupaciones correspondientes al Open Science Grid (OSG) y los sitios no registrados también son mostradas en el portal.

Las vistas o agrupaciones existentes en el portal son las siguientes:

- **Tier1:** permite ver solamente los datos asociados a los sitios de primer nivel dentro de EGEE.
- **Tier2:** permite ver solamente la información asociada a los sitios de segundo nivel dentro de EGEE.
- **Countries:** permite acceder a una vista geográfica de los sitios de acuerdo a su país.
- **EGEE:** permite obtener una vista global de la información de las infraestructuras de Producción y Pre-Producción dentro del proyecto.
- **OSG:** muestra la información de contabilidad publicada por los sitios del Open Science Grid (OSG).
- **Unregistered:** permite ver los datos correspondientes a los sitios que no están registrados en el proyecto.
- **VO_Discipline:** muestra la información agrupada por la disciplina de cada VO.
- **VO_Metrics:** identifica aspectos relacionados con la utilización de la infraestructura por las VOs.

3.3. Contabilidad por usuario

La información del DN de usuario está disponible en los registros de contabilidad de cada trabajo pero para proteger la privacidad de los usuarios esta información se almacena cifrada y una clave privada apropiada es requerida para descifrarla. Usando esta información es posible crear estadísticas acerca de la contabilidad a nivel de cada usuario individual.

Para acceder a las páginas de estadísticas por usuario, los usuarios necesitan autenticarse usando un certificado X.509 firmado por una autoridad certificadora reconocida por la IGTF. Una vez que el usuario es autenticado correctamente, la autorización puede ser concedida con arreglo a los siguientes papeles o roles: administrador de una VO, miembro de una VO, administrador de un sitio, usuario y desarrollador.

Dependiendo del rol asignado al usuario podrá ver diferentes estadísticas que le permitirán ver por



El Portal de Contabilidad de EGEE es la interfaz gráfica de usuario de los datos de contabilidad de EGEE



Las agrupaciones correspondientes al Open Science Grid (OSG) y los sitios no registrados también son mostradas en el portal



El análisis de los datos de contabilidad ayuda a comprender mejor como está siendo usado el Grid y descubrir los diferentes patrones de uso entre las VOs

Estudiar las tendencias de una determinada VO puede ayudar a establecer los períodos de mayor actividad durante el año

ejemplo, las estadísticas correspondientes a los usuarios que más han consumido en una VO o sitio si es un administrador de VO o sitio respectivamente, o bien ver la información sobre los trabajos que el propio usuario ha ejecutado.

Inicialmente, el UserDN no será mostrado directamente sino que se mostrará cifrado, debido a temas de privacidad y protección de datos que pueden afectar a algunos países involucrados en EGEE/WLCG. El UserDN cifrado es una versión modificada del UserDN que nos permite identificar al usuario, y cuya finalidad es agrupar su actividad.

4. Análisis de utilización del Grid

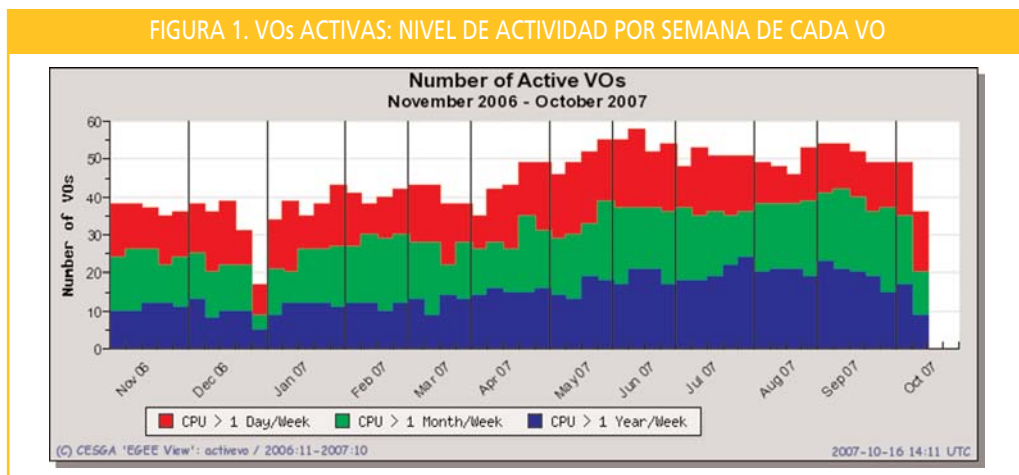
El análisis de los datos de contabilidad ayuda a comprender mejor como esta siendo usado el Grid y descubrir los diferentes patrones de uso entre las VOs. Esta información puede ayudar a los proveedores de los recursos y los administradores de las VOs a optimizar la utilización de esos recursos. Por ejemplo, para detectar el exceso de utilización o la infra-utilización y adoptar las medidas adecuadas.

Estudiar las tendencias de una determinada VO puede ayudar a establecer los períodos de mayor actividad durante el año. Durante estos períodos, los Centros de Recursos deben tratar de aumentar los recursos en línea lo máximo posible. Por otro lado, sabiendo de antemano los períodos de menor consumo ayudará a los sitios a programar periodos de mantenimiento, así como a diseñar alternativas para la utilización de los recursos libres durante estos períodos.

4.1. ¿Cuántas VOs están usando el Grid?

Hay aproximadamente 210 VOs que han publicado datos en la base de datos de contabilidad del GOC pero sólo 121 están registrados en la base de datos del Core Infrastructure Center (CIC). Pero, ¿cuántas de estas VOs están realmente utilizando el Grid? Si nos fijamos en la Figura 1 podemos ver que sólo una pequeña fracción de estas VOs están utilizando en realidad el Grid, por ejemplo, menos de 60 usan más de un día de tiempo de CPU durante una semana. Teniendo en cuenta un criterio un poco más exigente de consumo de un mes de tiempo de CPU durante una semana, el número de VOs activas se reduce a menos de 42.

FIGURA 1. VOs ACTIVAS: NIVEL DE ACTIVIDAD POR SEMANA DE CADA VO

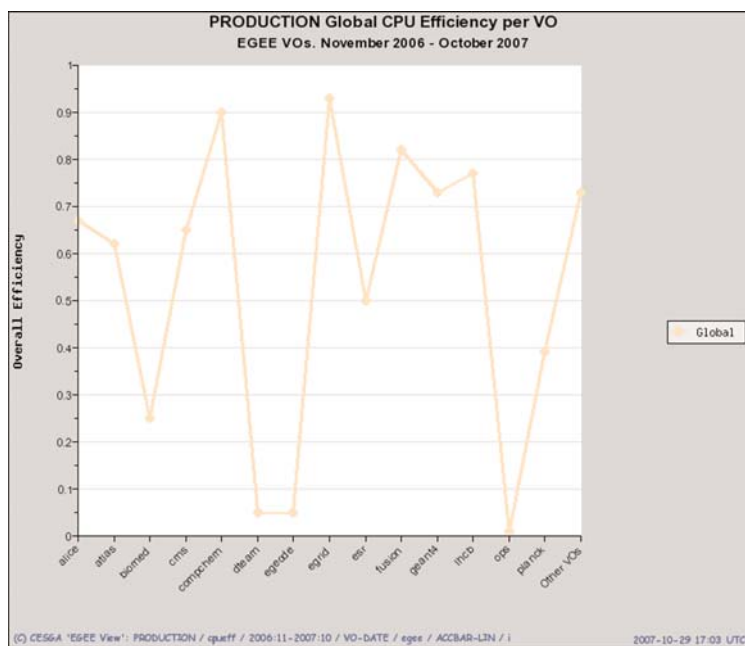


Un análisis de todos los datos publicados desde Enero del 2006 por los sitios de Producción y compuesto por casi 50 millones de registros de trabajos, muestra que el uso está dominado por las VOs de LHC y biomedicina que en conjunto consumen el 93% (equivalente más de 17.000 años de CPU) del tiempo de CPU reportado en EGEE.

4.2. ¿Cuán eficientes son los trabajos ejecutados en el Grid?

Hemos visto que sólo un número reducido del total de VOs están utilizando realmente la infraestructura, pero ahora la pregunta que queremos responder es si se está utilizando de manera eficiente o no. Para ello, definimos la relación entre el tiempo de CPU acumulado y el tiempo de espera acumulado de los trabajos ejecutados por una determinada VO durante un período de tiempo concreto. Los resultados pueden verse en la Figura 2. Se puede observar que sólo **cuatro** VOs tienen una eficiencia global superior a **0,75**. Las dos VOs con menor eficiencia son dteam y ops, lo cual es razonable, ya que son VOs de operaciones utilizadas sólo para monitorizar la infraestructura y no para realizar cálculos reales. Definitivamente las VOs que muestran una eficiencia por debajo de 0.50 deberían revisar sus trabajos en busca de posibles problemas.

FIGURA 2. EFICIENCIA GLOBAL DE CPU POR VO:
RATIO ENTRE EL TIEMPO DE CPU Y EL TIEMPO DE ESPERA



Sólo un número reducido del total de VOs están utilizando realmente la infraestructura

Podemos analizar cómo una determinada VO está utilizando el Grid buscando tendencias que nos ayuden a establecer los períodos de mayor actividad durante el año

4.3. Tendencia de uso de una VO

Finalmente podemos analizar cómo una determinada VO está utilizando el Grid buscando tendencias que nos ayuden a establecer los períodos de mayor actividad durante el año. Nos centraremos en la VO **Biomed** porque es la única para la que está disponible la información completa acerca de sus Data Challenges (DC)[5].

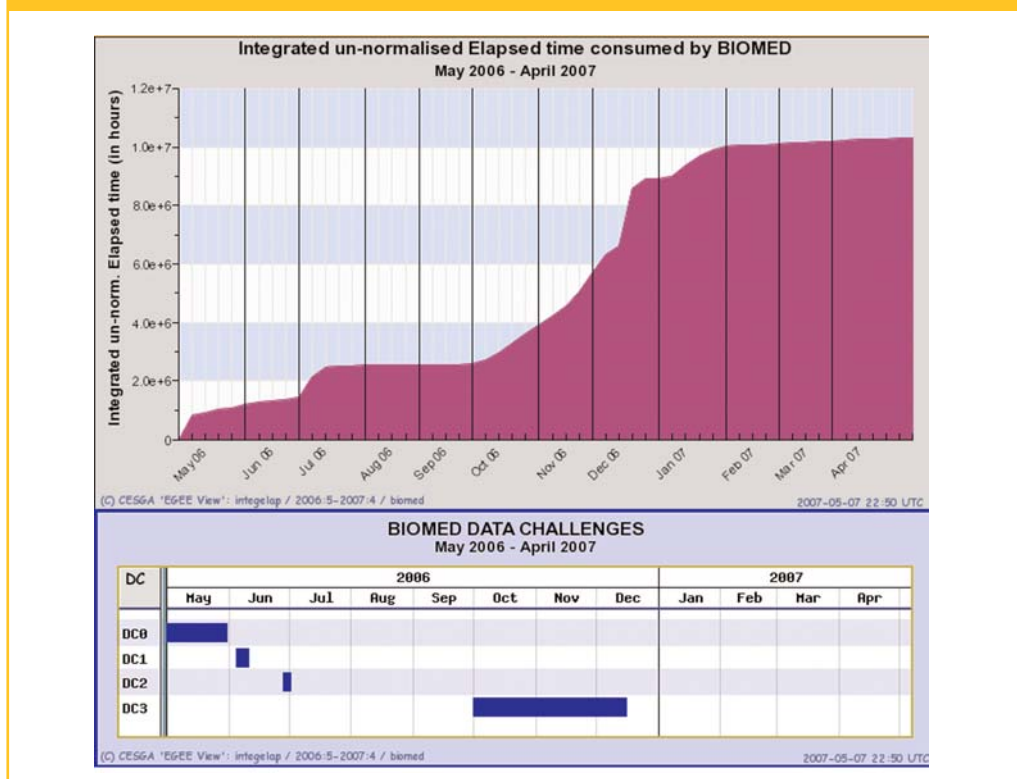
Si nos fijamos en la Figura 3 se puede observar que hay ciertos períodos de tiempo donde hay fuertes incrementos en el tiempo de espera acumulado normalizado consumido por Biomed. Esto indica períodos de alta actividad donde la demanda de recursos es muy alta. También hay períodos con casi ninguna actividad representada por las mesetas que aparecen en el gráfico. Comparando este gráfico con la información acerca de los DC de Biomed, mostrados en el diagrama de Gantt, puede verse que los períodos de alta actividad coinciden con los períodos de los DC. Otro hecho interesante es que la demanda de recursos se extiende algún tiempo después del fin oficial de los DC. Este es un factor importante a tener en cuenta por los centros cuando planean contribuir con sus recursos a un DC.



Los datos de cada sitio y VO pueden ser agrupados para formar resúmenes anónimos

Para mostrar esta información se ha creado un portal que nos permite ver datos históricos del almacenamiento (en disco y/o cinta) utilizado y disponible por cada sitio o a nivel de federación y desglosado por VO

FIGURA 3. TIEMPO DE ESPERA NORMALIZADO ACUMULADO CONSUMIDO POR LA VO BIOMED



5. Almacenamiento

El esquema de información GLUE (versión 1.2) nos proporciona información sobre almacenamiento a través de los atributos GlueSAStateAvailableSpace y GlueSAStateUsedSpace.

La información acerca de los recursos de almacenamiento en los sitios de EGEE es obtenida de un Top BDII (es decir, GlueSAStateUsedSpace por VO por sitio) e introducida en una base de datos MySQL. La información del BDII se obtiene a partir de sensores que se incluyen por defecto en muchas de las implementaciones disponibles para el Storage Element (SE): DPM, dCache y CASTOR...

Los datos de cada sitio y VO pueden ser agrupados para formar resúmenes anónimos. A diferencia de los registros de contabilidad del trabajos, la información del almacenamiento reportada a través de BDII es de muy alto nivel y totalmente anónima.

El sensor ofrece las siguientes funcionalidades:

- Obtención de la información desde un Top BDII, en lugar del BDII de cada sitio. Esto hace que el sistema sea más robusto ante fallos o períodos de mantenimiento de los sitios.
- Recoge la información del espacio utilizado y disponible para todos las VOS soportadas por un sitio.
- Recoge la información del espacio utilizado y disponible para todos los sitios de EGEE.
- Distingue entre el almacenamiento en disco y cinta.

Para mostrar esta información se ha creado un portal[6] que nos permite ver datos históricos del almacenamiento (en disco y/o cinta) utilizado y disponible por cada sitio o a nivel de federación y desglosado por VO.

6. Futuro

En los próximos meses, la contabilidad por usuario será nuestro principal objetivo. Una primera versión ha sido creada pero deberá ser actualizada y completada. Una de las principales dificultades

con las que nos encontramos es que actualmente muy pocos sitios están publicando la identidad del usuario por lo que la contabilidad por usuario no estará completa. Se deberá hacer hincapié en que los sitios publiquen esta información.

Otra de nuestras prioridades, será mostrar la información de contabilidad correspondiente a cada grupo y/o rol asociado a cada VO.

El desarrollo y el despliegue del Resource Usage Service (RUS)[7] permite el intercambio de los datos entre Grids de manera unificada (a diferencia de las actuales implementaciones que utilizan una variedad de mecanismos de transporte). La OGF-RUS define una hoja de ruta clara para la finalización de la especificación RUS, que se divide en dos partes: la especificación del núcleo del RUS y características avanzadas (agregación y la replicación de datos). Investigaremos prototipos basados en esta especificación para aplicarlos en la plataforma de contabilidad de EGEE.

Actualmente, la información de contabilidad que se posee en EGEE se corresponde con datos de trabajos enviados por usuarios Grid. Para completar los datos de contabilidad, se deberán recoger y mostrar los datos de los trabajos enviados por usuarios que no son del Grid (trabajos locales). Para EGEE/WLCG, es esencial poder distinguir entre la utilización Grid y no-Grid[8]. Otro de nuestros objetivos será incluir en la plataforma de contabilidad los trabajos locales.

La normalización del tiempo de CPU no siempre está bien caracterizada por los valores de SpecInt2000 publicados. Por eso se ha establecido un grupo de trabajo (HEPiX) para intentar encontrar una solución adecuada a este problema. Este grupo ha realizado una propuesta que intentaremos implantar en los próximos meses.

En cuanto al almacenamiento, se ha creado un primer prototipo que deberá ser completado. También se estudiará la posibilidad de integrar los dos portales en uno solo.

Referencias

- [1] <http://www.eu-egee.org/>
- [2] <http://lcg.web.cern.ch/LCG/>
- [3] <http://www.r-gma.org/>
- [4] http://www3.egee.cesga.es/gridsite/accounting/CESGA/egee_view.html
- [5] <https://cic.gridops.org/index.php?section=vo&page=datachallenges>
- [6] <http://goc02.grid-support.ac.uk/storage-accounting/view.php?queryType=storage>
- [7] http://omii-europe.com/OMII-Europe/News/OMII_Europe-MJRA1.4.pdf
- [8] https://twiki.cern.ch/twiki/pub/LCG/MbMeetingsMinutes/LCG_Management_Board_2007_06_05.htm
- [9] <https://edms.cern.ch/document/1726137>

Pablo Rey
Javier López
Carlos Fernández

Fundación Centro Tecnológico de Supercomputación de Galicia (CESGA)

◆
El desarrollo y el despliegue del Resource Usage Service (RUS) permite el intercambio de los datos entre Grids de manera unificada

◆
En cuanto al almacenamiento, se ha creado un primer prototipo que deberá ser completado. También se estudiará la posibilidad de integrar los dos portales en uno solo