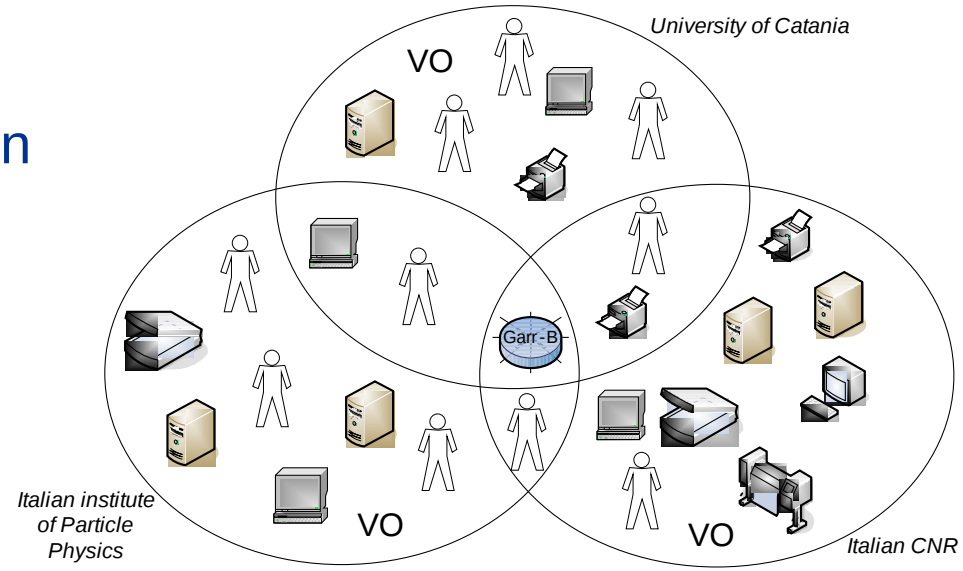
gLite Overview

Prof. Yudith Cardinale
Prof. Jesus De Oliveira

Introducción a Sistemas Grid

- **Escenario:**
  - Múltiples usuarios de diferentes organizaciones geográficamente distribuidas (VO) requieren alto poder de cómputo y compartimiento de datos para trabajo colaborativo
  - Múltiples recursos (computacionales y de almacenamiento) existen en diversas instituciones pero se usan de forma independiente



Introducción a Sistemas Grid

- **El objetivo de los sistemas y aplicaciones grid es**
  - Integrar
  - Virtualizar
  - Gestionar

- **Recursos y servicios a través múltiples VOs.**
  - El middleware grid es la capa que permite la agregación, compartimiento y gestión de recursos distribuidos de forma transparente.

Conceptos

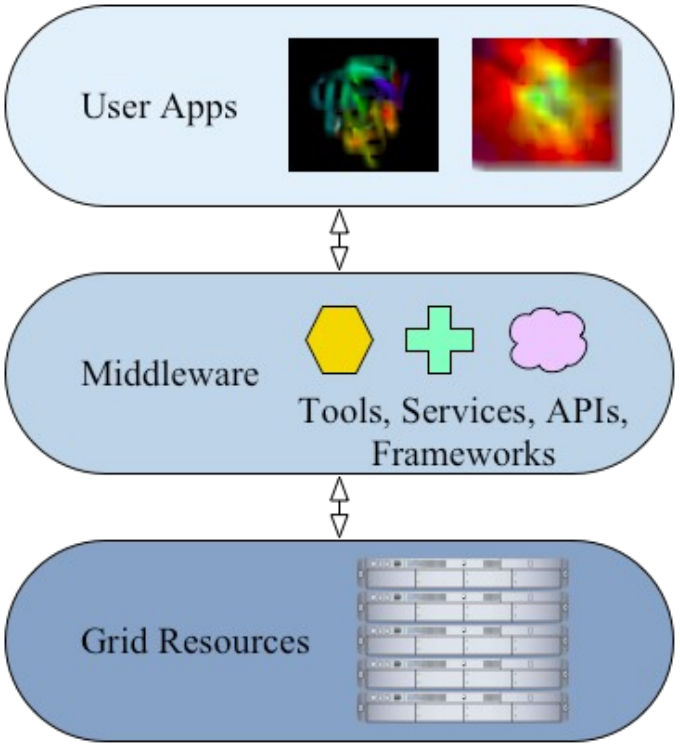
- **Recursos computacionales:** Máquinas donde los usuarios pueden ejecutar programas (denominados “jobs” o “trabajos”) y almacenar o acceder a archivos independientemente de su localización geográfica.

- **Job (trabajo):** es una tarea computacional (un ejecutable o un script) que el usuario desea ejecutar en el Grid, obteniendo sus resultados en su máquina local.

- **Job Submission (envío de trabajos):** Es la acción de delegar en el Grid la ubicación del mejor recurso computacional para ejecutar un Job y la “colocación” de éste para que sea ejecutado.

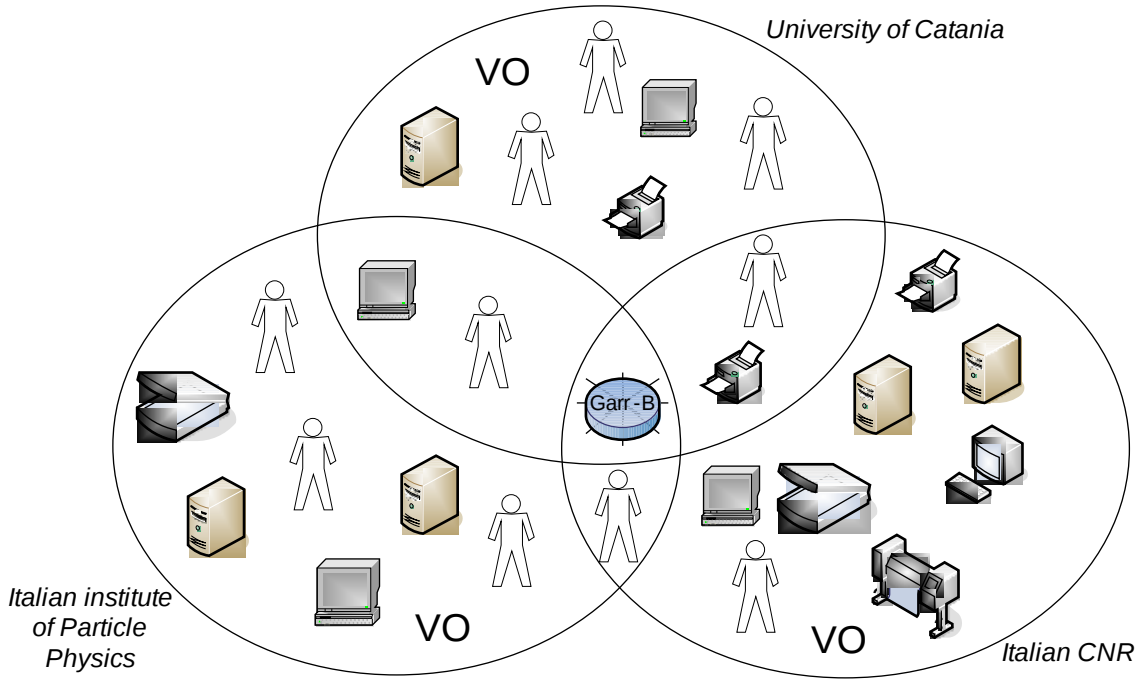
Conceptos

- **Middleware de Grid –** Capa entre aplicaciones de usuarios y recursos computacionales y de almacenamiento



Conceptos

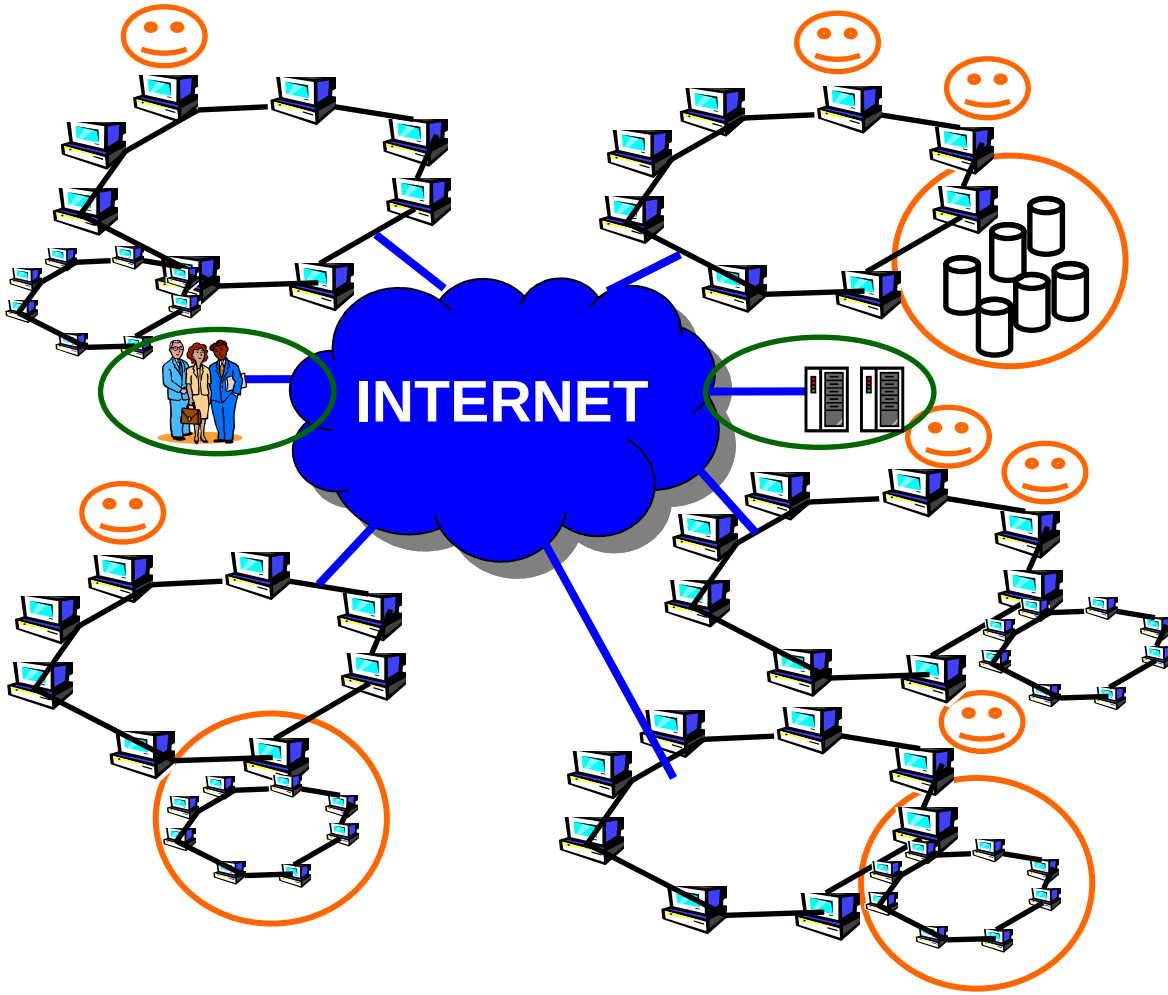
- **Organizaciones Virtuales (VOs):** Grupos de usuarios y recursos con objetivos de investigación comunes. Requieren herramientas que permitan el trabajo colaborativo entre sus miembros, que pueden estar geográficamente distribuidos.



Panorama general

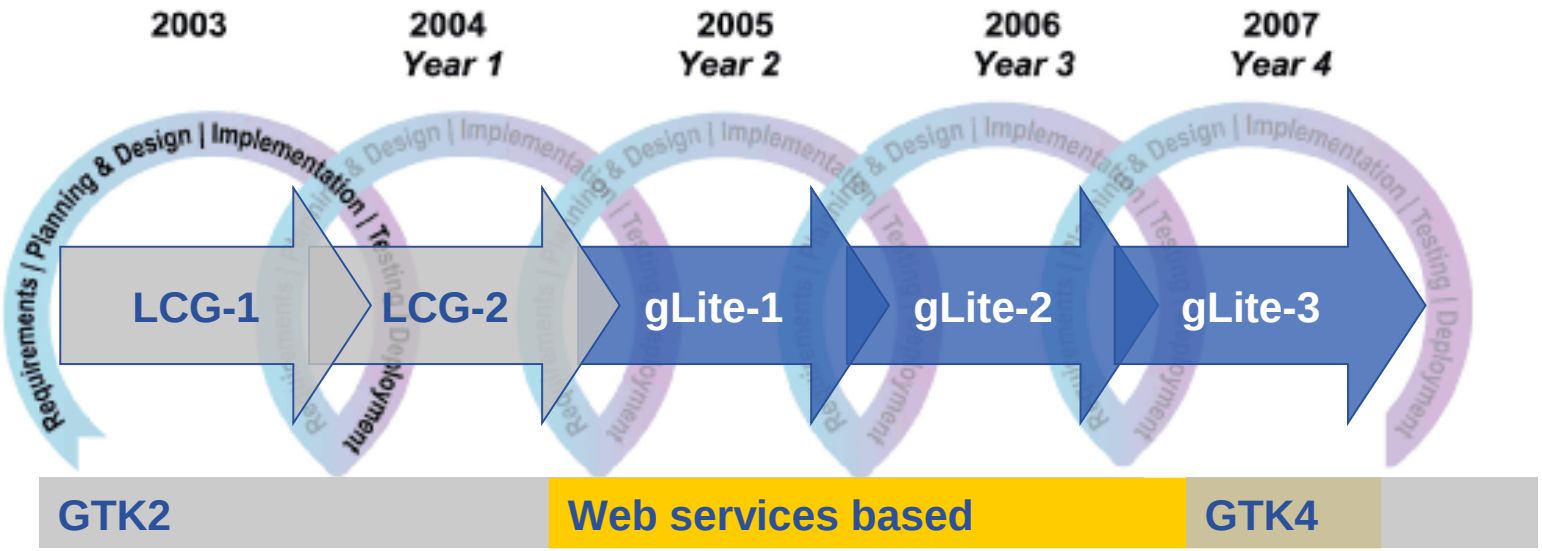
- **Los usuarios se unen a VOs**
- **Cada VO contribuye con recursos y negocia accesos.**
- **El middleware Grid permite el acceso y uso compartido de:**
  - "Elementos de almacenamiento (SE)".
  - "Elementos de cómputo (CE)".
- **Servicios adicionales (de personas y del middleware) potencian el grid.**
- **Resultado:**

COLABORACIÓN



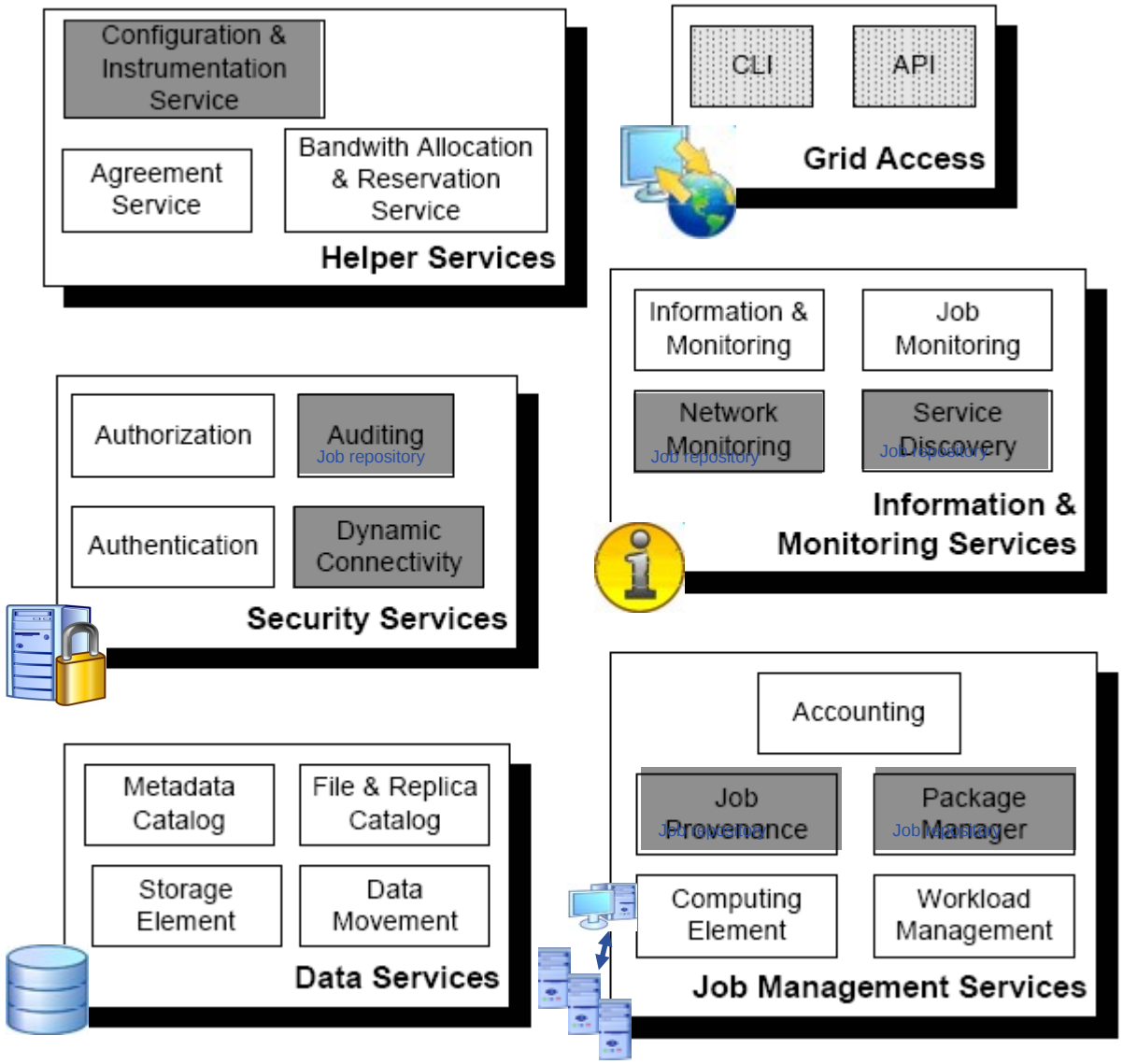
gLite

- Middleware ligero de última generación para computación grid.
- Basado en una arquitectura orientada a servicios (SOA)
- Nace de los esfuerzos colaborativos de instituciones de investigación académicas e industriales como parte del proyecto EGEE

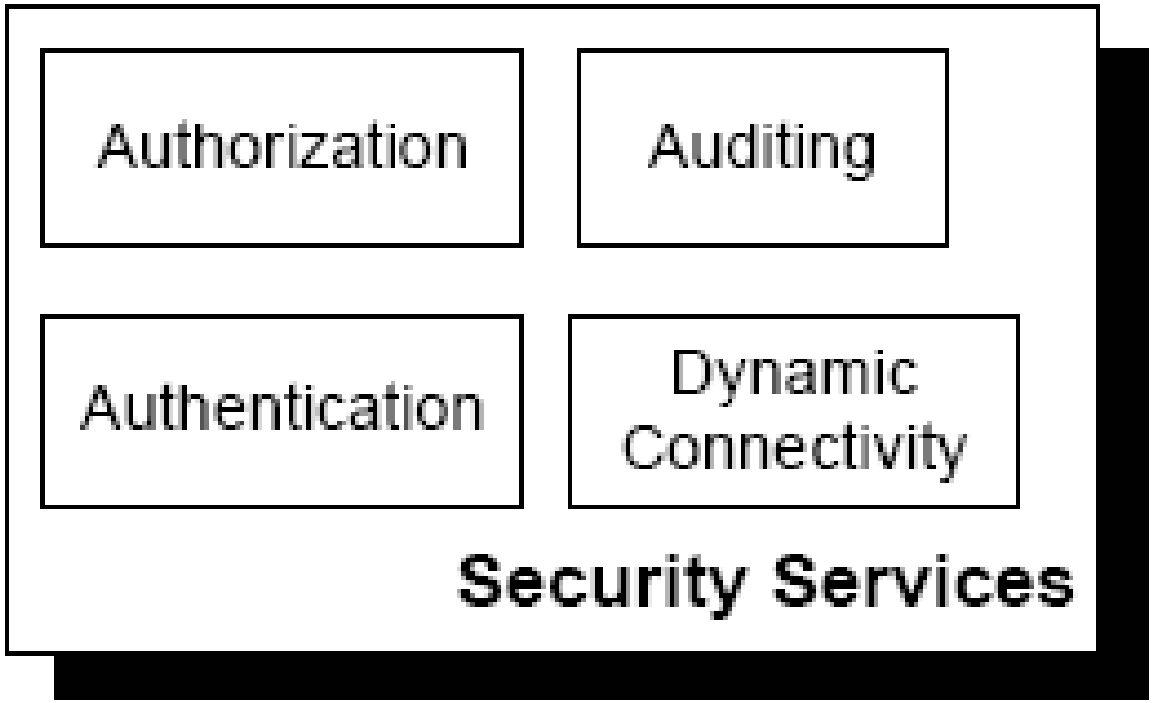


gLite – Descomposición de servicios

5 servicios de alto nivel + CLI & API



gLite – Servicios de seguridad



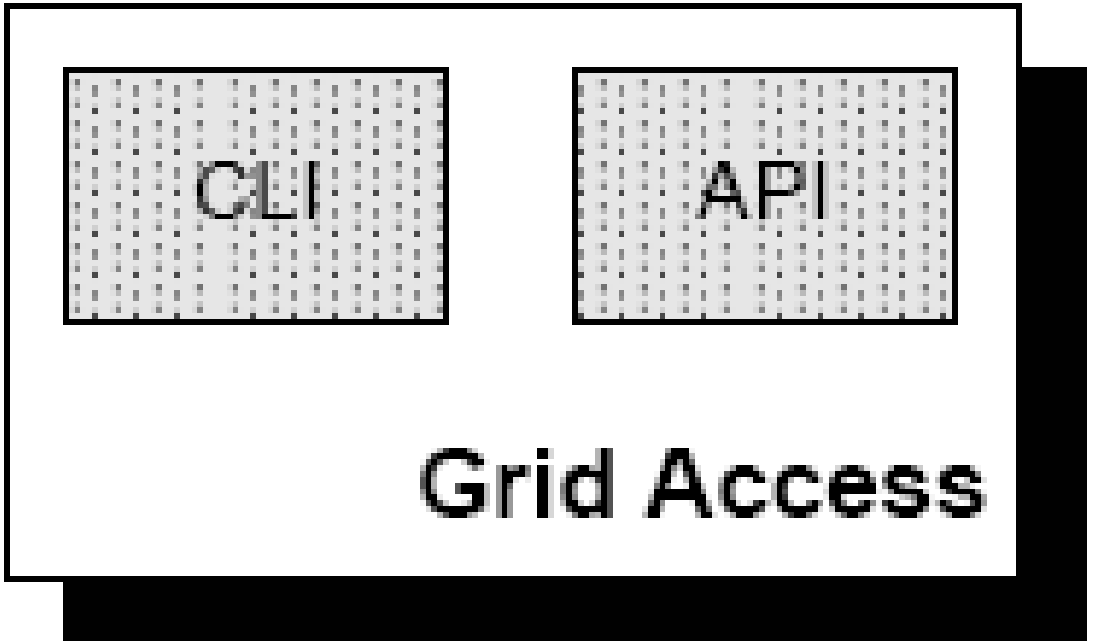
- **Autenticación basada en la infraestructura de clave pública X.509 (PKI)**
  - Autoridades certificadoras (CA) generan certificados que identifican individuos (p.e. un pasaporte)
  - Se establecen relaciones de confianza entre multiples CAs
  - Para reducir vulnerabilidades, la identificación de usuarios ante los servicios del grid se realiza usando proxies de sus certificados (con validez delimitada)
- **Los certificados proxy pueden:**
  - Ser delegados a un servicio, para que este actúe como si fuera el usuario original.
  - Incluir atributos adicionales (como info de VO a través del VO Membership Service - VOMS)
  - Ser almacenados en un almacén externo (MyProxy)
  - Ser renovados (en caso de que estén a punto de expirar)

- **Autenticación**
  - Un usuario recibe un certificado firmado por una CA
  - Se conecta al UI via SSH
  - Descarga e instala el certificado
  - **Crea el proxy (single sign on) – el grid emplea el proxy para identificar al usuario entre sus componentes**

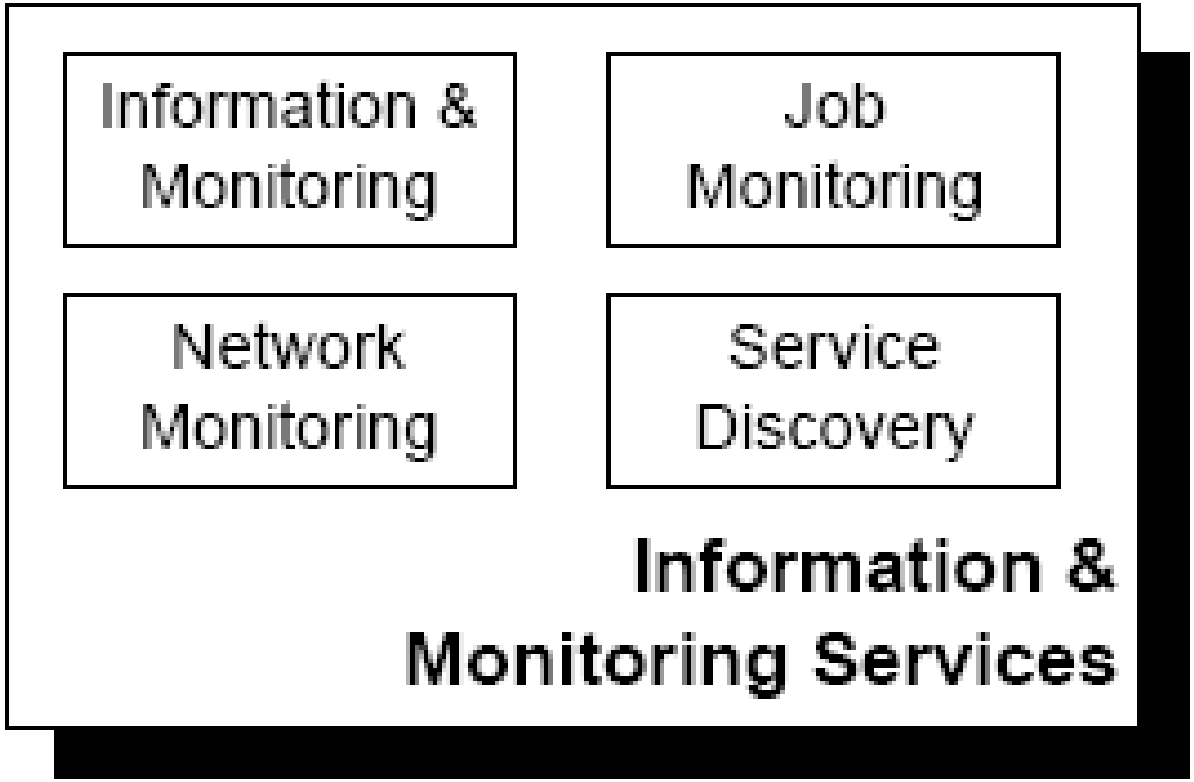
- **Autorización**
  - El usuario se une a una VO
  - VO negocia acceso a nodos y recursos del grid
  - En el CE se verifica si el usuario, siendo miembro de la VO, tiene acceso al recurso.
  - A través del “gridmapfile”, se hace la correspondencia entre el usuario y una cuenta local en el recurso

Dos posibilidades: **APIs** y **CLI**.

El uso de web services permite la generación automática de APIs



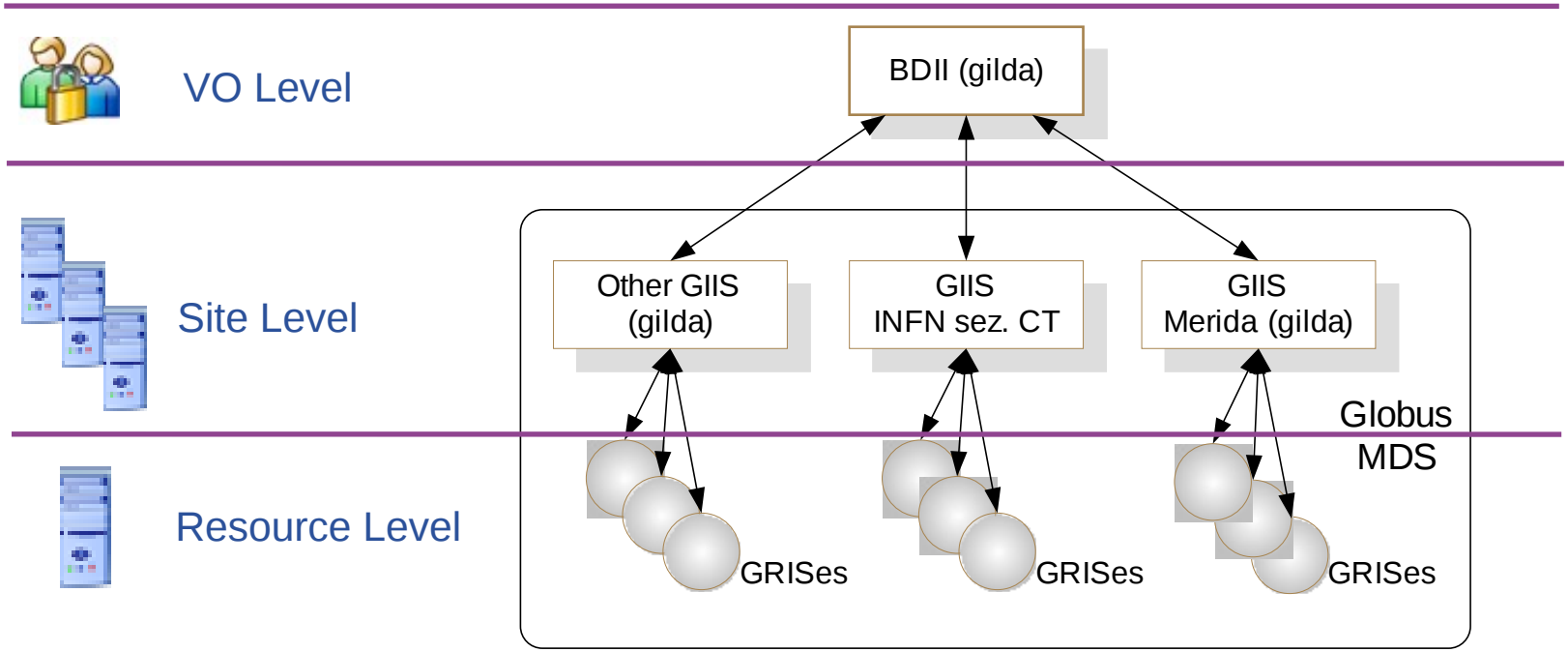
Los servicios de información son componentes vitales de bajo nivel del Grid



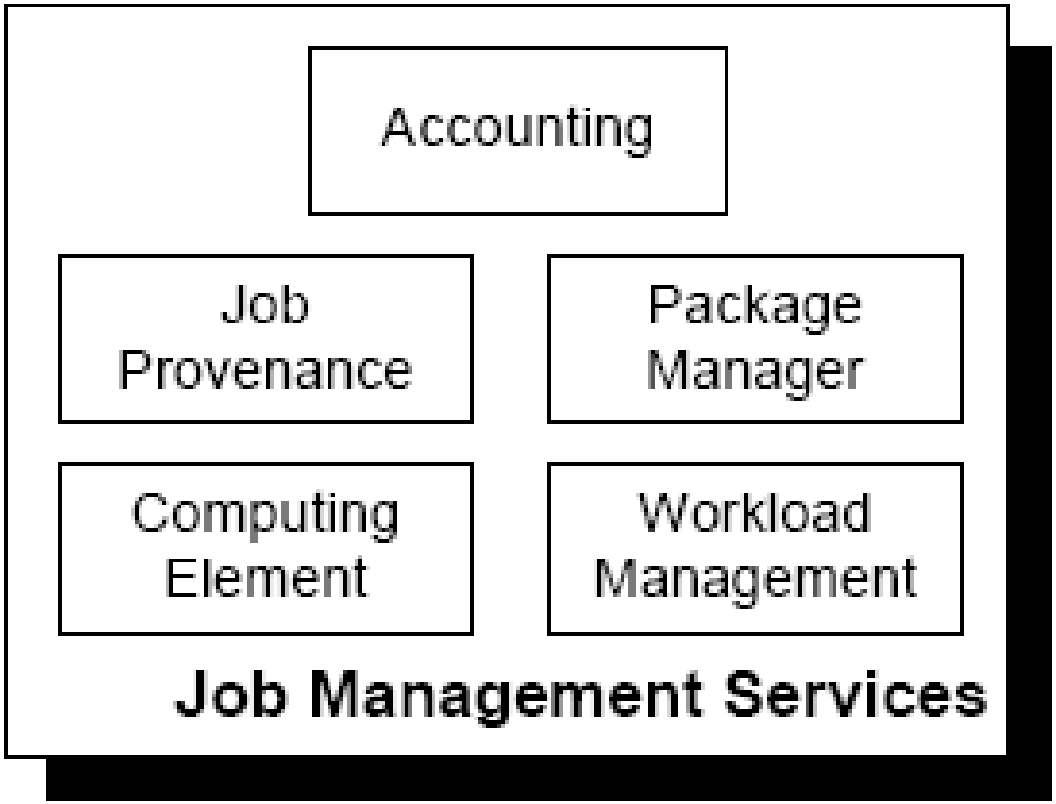
Berkeley Database Information Index (BDII)

La información es almacenada jerárquicamente bajo un modelo de árbol
(Implementación LDAP del esquema **GLUE**)

- **GRIS** Información a nivel de **recursos**
- **GIIS** Información a nivel de **sitio**
- **BDII** Información a nivel de **VO**



gLite – Job Management Services



- Acumula información sobre el uso de recursos por usuarios o grupos de usuarios (VOs).

- La captura de información de servicios/recursos del grid requiere de **sensores** (Medición de recursos, capa de abstracción de medición, registros de uso).

- Los registros son recolectados por el **Accounting System** (Consultas: Usuarios, Grupos, Recursos).

- Los servicios del grid deberán registrarse con un servicio tarificador cuando se requiera contabilización para propósitos de cobranza y determinación de costos.

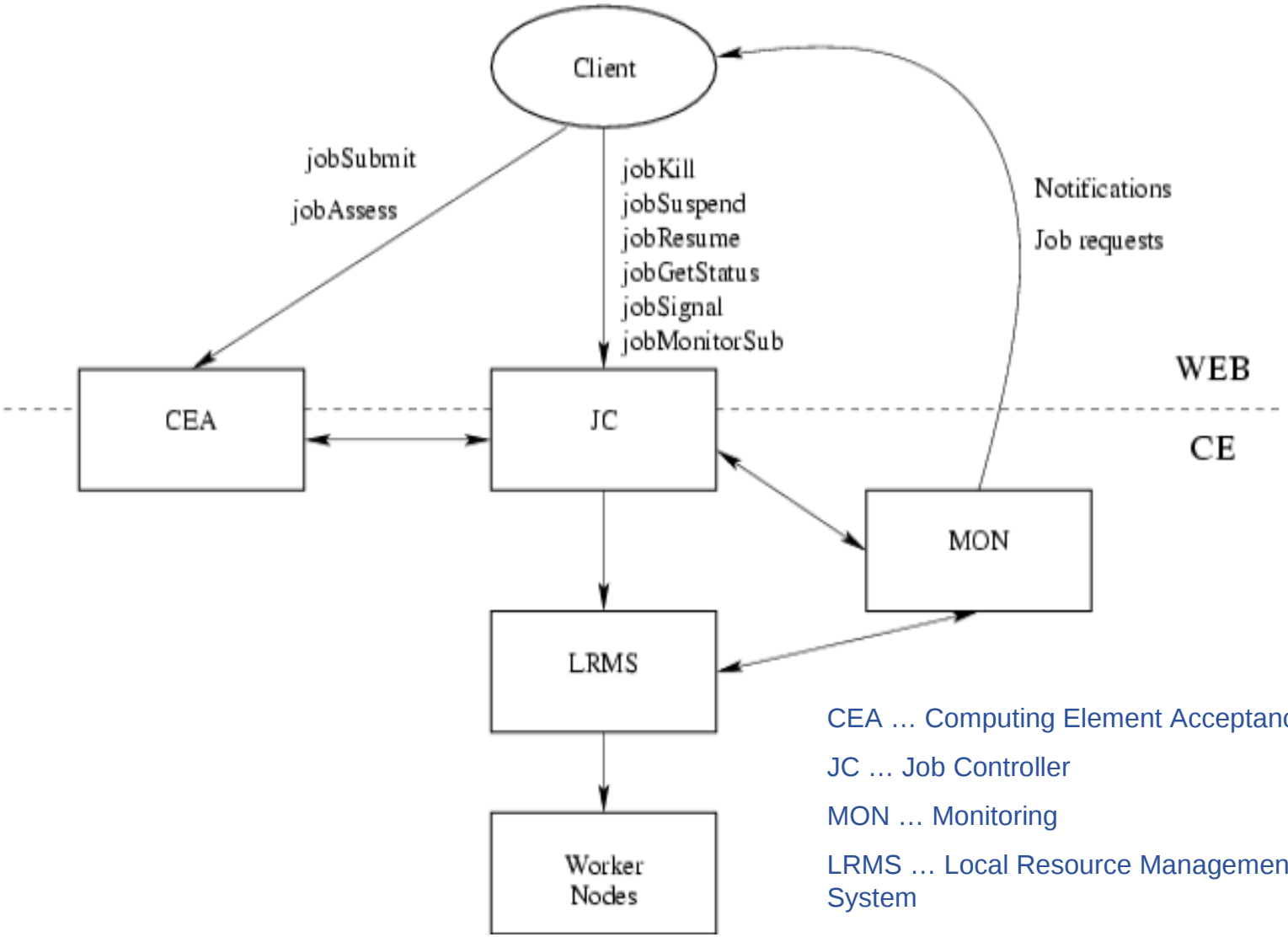
- Servicio que representa el recurso de cómputo que es responsable del job management: (envío, control, etc.).

- CEs se refiere a un conjunto o cluster de recursos de cómputo (WN) gestionado por un manejador de colas (LRMS), para despachar y ejecutar jobs que cumplan con los requerimientos del usuario.

- Dos modelos de job submission (de acuerdo a solicitudes del usuario o políticas del sitio):
  - **PUSH** (los jobs son enviados a la cola del CE),
  - **PULL** (los jobs son traídos a la cola del CE cuando está vacía)

- El CE es responsable de recolectar información de contabilización (accounting).

Computing Element (CE)



CEA … Computing Element Acceptance

JC … Job Controller

MON … Monitoring

LRMS … Local Resource Management System

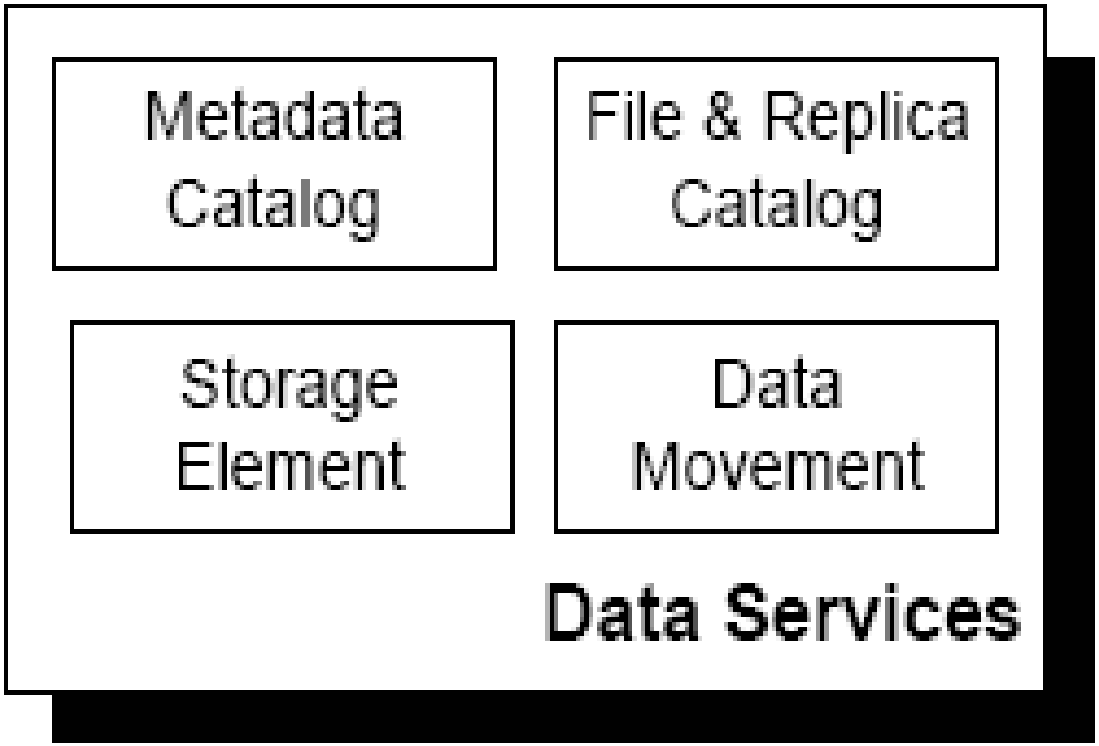
- **WMS** es el conjunto de componentes del middleware responsables de la distribución y gestión de jobs en los recursos del grid.

- Está compuesto por dos componentes:

  - **WM (workload manager)**: **Acepta** y **satisface requerimientos.**
    El proceso de asignar el mejor recurso disponible es denominado **Matchmaking.**

  - **L&B (logging and bookeeping**: **Seguimiento** de la ejecución de jobs en términos de **eventos:** enviado, ejecutándose, finalizado,...).

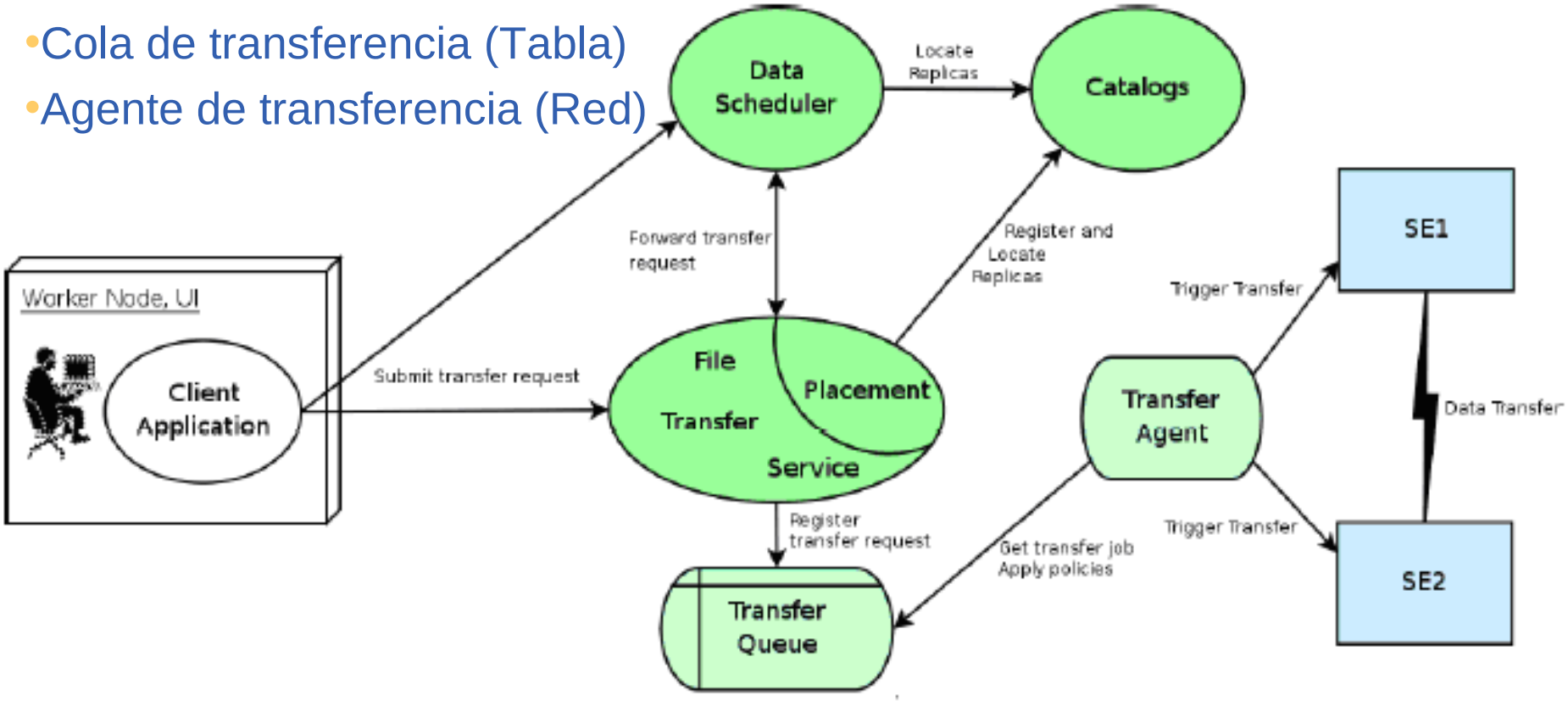
gLite – Servicios de datos



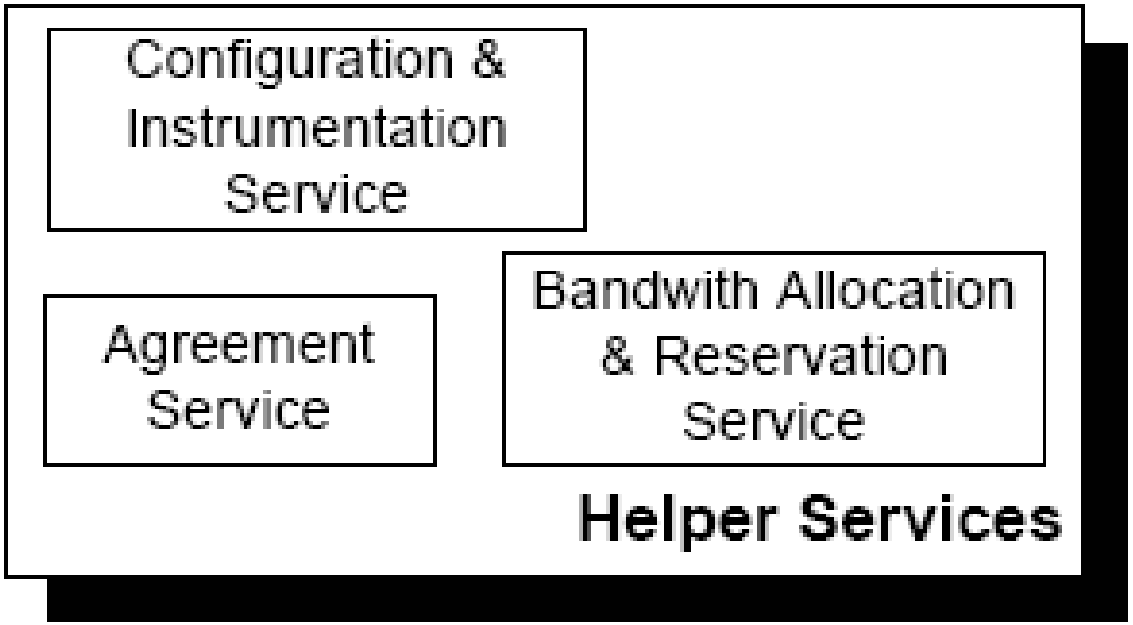
Los servicios requeridos para la gestión de datos son al menos los siguientes:

- Backend de almacenamiento (Drivers y Hardware)
- Interfaz del Storage Resource Manager
- Servicio de transferencia de datos (GridFTP)
- API de I/O para archivos POSIX-like (gLite-I/O)
- Servicios auxiliares de contabilización (accounting) y logging

- Data Scheduler (**DS**): hace seguimiento de las solicitudes de transferencia de usuarios y servicios
- File Transfer/Placement Service (**FTS/FPS**): localiza réplicas de los datos
- Cola de transferencia (Tabla)
- Agente de transferencia (Red)



gLite – Helper Services



Configuration and Instrumentation service – Consultan el estado de los recursos.

Agreement Service – Implementa el protocolo de comunicación para Service Level Agreements (**SLA**s).

Bandwidth Allocation & Reservation service (**BAR**) – Controla, Balancea y gestiona flujos de red.

- **Virtual Organization Membership Service**
  - Múltiples organizaciones virtuales (VOs).
  - Múltiples roles in cada VO
    - Extensiones X509 compatibles.
    - Firmadas por el VOMS server.
  - Interfaz administrativa web.
  - Soporta MyProxy.
  - Permite dar privilegios de acceso a los recursos por VO o por roles.
  - Cada sitio asocia los miembros o roles de una VO al mecanismo local de autenticación (unix users accounts).
    - Permite la implantación de politicas de seguridad locales a cada recurso.

- **MyProxy**
  - Permite la ejecución de jobs largos e incrementa la seguridad.
    - El WMS es responsable de renovar el proxy
    - Los usuarios no deben producir proxis “long lived”.
  - Permite seguridad en usuarios móviles.
    - Los usuarios no necesitan copiar o transferir manualmente sus claves privadas (globus keys).

Arquitectura

- User Interface (UI)
- Workload management system (WMS)
- Logging and bookkeeping service (LB)
- Virtual Organization Management service (VOMS)
- Information system (BDII), monitoring (MON)
- Computing element (CE) y worker nodes (WN)
- Storage element (SE) y File catalogue (LFC)

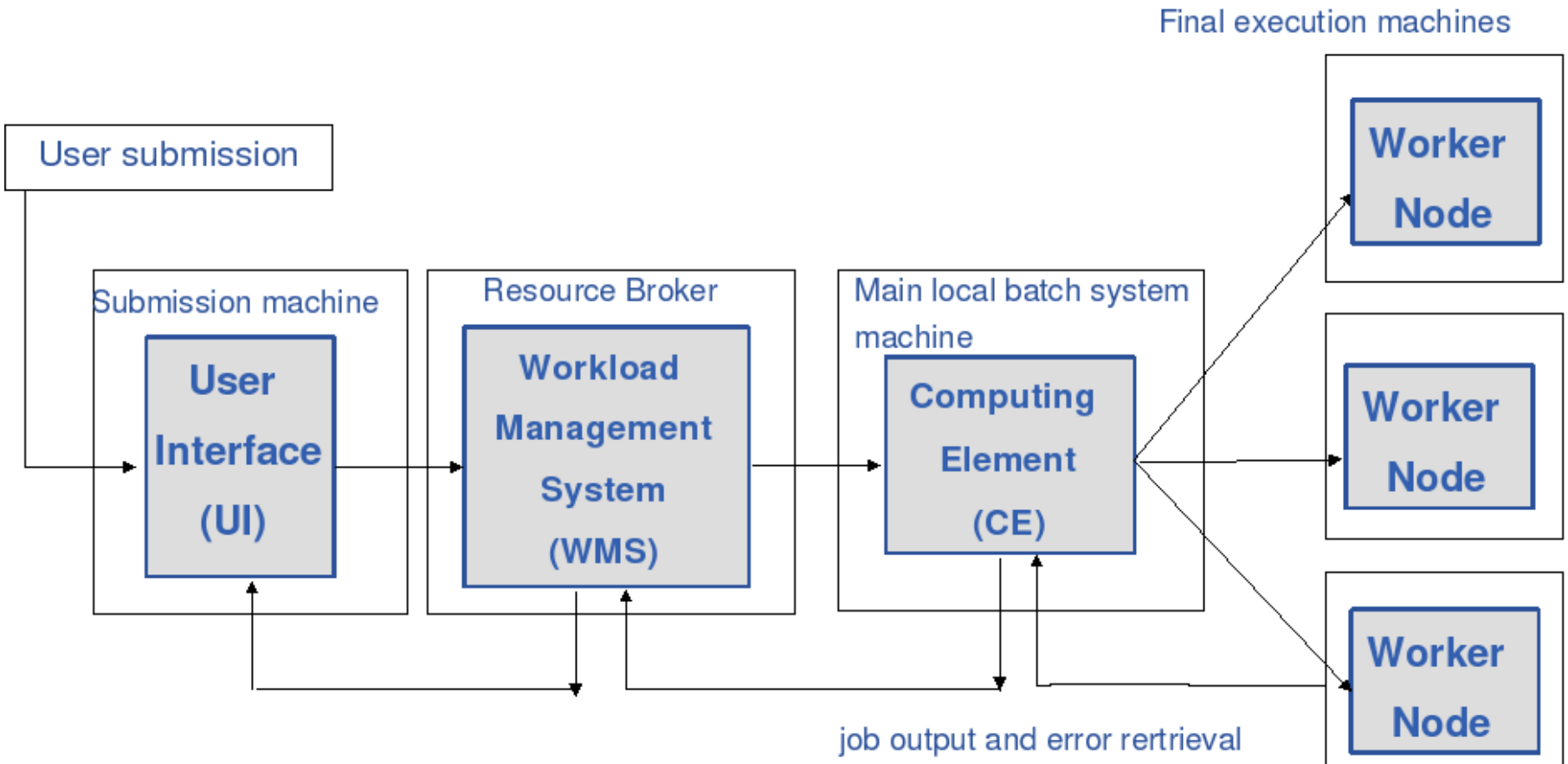
Componentes

- El **User Interface (UI)** es el punto de entrada al grid, es parte de la estación de trabajo del usuario. Se considera parte del WMS.
- El **Workload Management System (WMS)** es un conjunto de componentes cuyo objetivo es encontrar el recurso que cumple con los requerimientos de un job de un usuario entre los CEs disponibles. Esto es, encontrar la máquina donde finalmente el job será ejecutado.
- El **Computing element (CE)** es el punto de entrada al sistema de colas local (PBS,LSF, CONDOR).
- Los **Worker Nodes** son las máquinas donde los jobs son realmente ejecutados. Están enlazados con el CE a través del sistema de colas local, al cual los jobs son enviados.

Componentes

- El **Information System and Monitoring** (IS y MON), mantienen data sobre los recursos disponibles y el estado del sistema.
- El **Logging and bookkeeping service** (LB), hace seguimiento a los eventos que le ocurren a los jobs.
- El **Virtual Organization Management service (VOMS)** , provee mecanismos de autenticación y autorización en el acceso a los recursos.
- El **Storage element (SE) y File catalogue (LFC),** permiten gestionar transferencias de archivos grandes o facilitar la disponibilidad y localización de archivos de datos que los jobs requieren.

Ciclo de vida de un job



Life of a job, from the submission to the retrieval

Preguntas