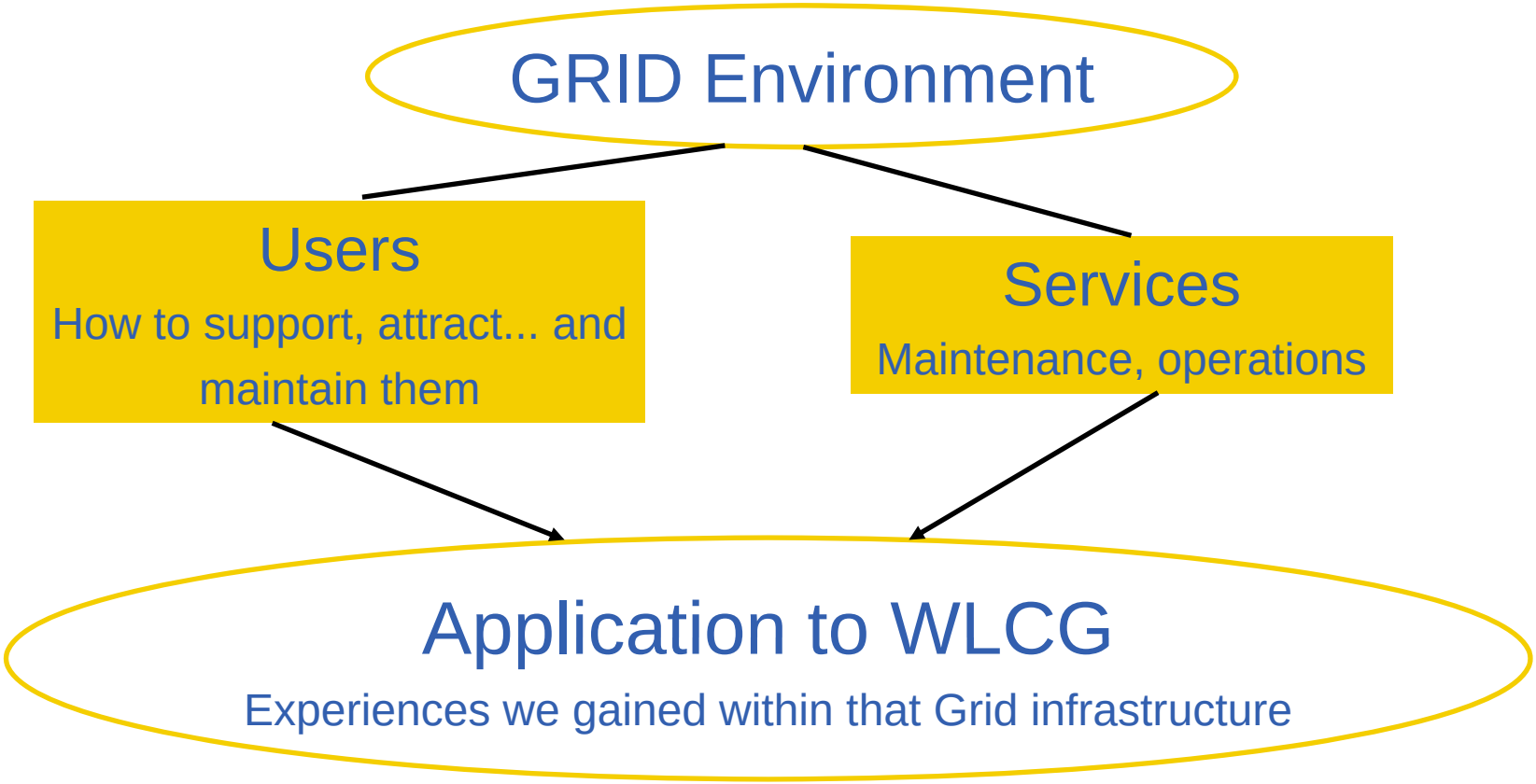
Increased productivity for merging Grid applications: the application support system

Patricia Méndez Lorenzo (CERN)

Andrew Maier (CERN), Hurng-Chun Lee (CERN), Jakub Moscicki (CERN), Adrian Muraru (CERN)

3rd EGEE User Forum. Clermont, February 2008

Outlook



- User support and high reliable services are key aspects of any Grid project
- In this talk we will address the status of user support applied to different communities beyond HEP and the lessons we learnt with their gridification

Grid technologies: Introduction

- **Grid technologies has been adopted by many research communities as their computational model**
  - It covers computational and storage requirements for worldwide distributed communities
  - Ensure a good level of service at all sites
  - Provides the necessary manpower to ensure the required level of service
  - **Strong Points**
    - Excellent way of federating resources in quasi-homogeneous infrastructures
    - Matchmaking computational requirements in many applications
    - Decentralization of resources; small portion available on the lab
    - Worldwide access to data and resources
  - **Weak Points**
    - Reliability and scalability are open issues
    - Difficult to use, lack of user-friendly applications worldwide available

Grid: USERS point of view (I)

- **Potential use of the Grid**
  - **Provisioned use**:
    - Applications need quasi-dedicated resources
    - Offers an excellent long-term load ensuring the continuous use of the Grid
  - **Scheduled use:**
    - Many resources for a short time
    - Normally the resources are booked in advance for these communities
  - **Opportunistic use:**
    - Least demanding
    - Use of resources depending on their availability. Good examples available lately in theoretical physics applications
- **No individual centre will be able to assemble the resources and expertise level to satisfy any of these categories individually**

Grid: USERS point of view (II)

- **First Step: Getting known in the Grid**
  - Getting a certificate and mapping it to a certain VO
    - This procedure requires already a knowledge of the Grid procedures
      - *CA identification, Grid VO structure*
  - This procedure will require a mayor flexibility in order to include certain VOs
    - The security procedure will have to observe those VOs with non registered members
- **Second Step: "Gridification" process**
  - Continuous contact with a support person is required
    - Even to decide if Grid is the solution
    - Already from the 1st step, the support must already act
  - General procedures in terms of tools and infrastructures are mandatory in any Grid project
    - This procedure ensures a continuous flow of new applications with a limited number of members into the support team

Gridification structure

- **The Grid team at CERN helps different communities to be merged in the Grid environment in and beyond HEP**
- **Our "Gridification" structure is based on 3 mayor blocks (mostly dedicated to applications beyond HEP):**

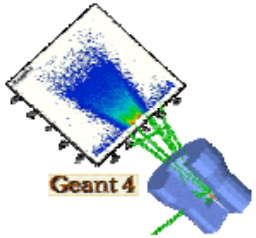
RESOURCES	SUPPORT	TOOLS
• Basic and full Grid infrastructure is guaranteed to the users • Built based on the experience gained with previous communities	• Follow up of the gridification and productions with the communities • Dedicated support person for each HEP community • Training and dissemination	• 1st Step: Basic set of tools to understand the requirements of the application • 2nd Step: Production • Ganga and/or DIANE are proposed as tools

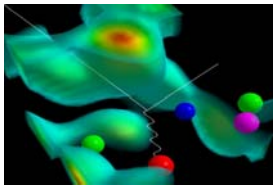
Communities

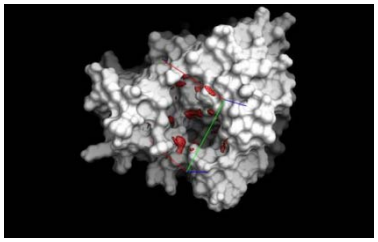
HARP

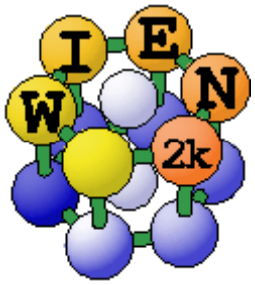
Garfield

Geant 4









Academia Sinica
Genomics Research Center

1. Resources , 2. Support…

- **Resources**
  - We provide a whole Grid infrastructure in terms of services
    - Based on the experience gained with the first big productions beyond HEP
    - Ready to be used by any application willing to test the Grid environment
  - VO structure: VO=gear
  - Central services placed at CERN: RB, LFC, VOMRS, SE
    - Not mayor interferences with big HEP productions are ensured
  - Resources ensured by several sites: queues, CPUs and software area are required

The idea is:"come and use the Grid"

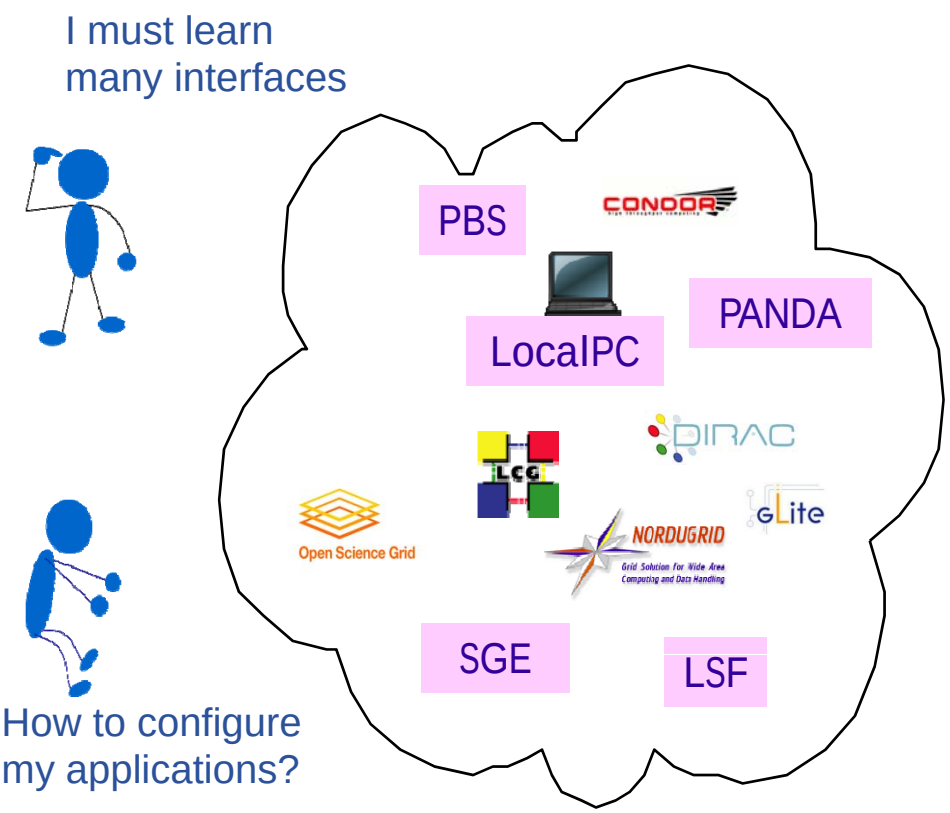
- **Support**
  - The full immersion and production is coordinated and supported by the support team
  - We become “experts” in each application to understand the requirements of the communities

… and 3. Tools

- **Proposed Tools**
  - Poor knowledge of the Grid infrastructure should be assumed
    - Grid can be a complicated environment for new comers
    - The support team proposes a set of tools to hide the Grid as much as possible and ensure a quick immersion
  - 1st set of tools: Delivered just to test the application in the Grid
    - Based in easy-to-use CLI to submit large bunches of jobs and control the outputs
    - During this phase the support is fundamental
      - *AIM: Creation of a self-contained application package*
  - 2st set of tools: Towards the Full Production
    - Ganga and DIANE
    - Remember... Standard tools able to be applied to any new community
    - The support team cannot scale the number of new communities

The UI: Ganga

- **Ganga: Gaudi/Athena and Grid Alliance**
- **User Interface with transparent and common interaction to different backends**
  - Submission purposes
  - Monitoring
- **Helps organizing the work**
  - Job history: keep track of what user did
  - Save job outputs in a consistent way
  - Reuse configuration of previously submitted jobs
- **Available in different views**
  - CLI
  - GUI
  - API

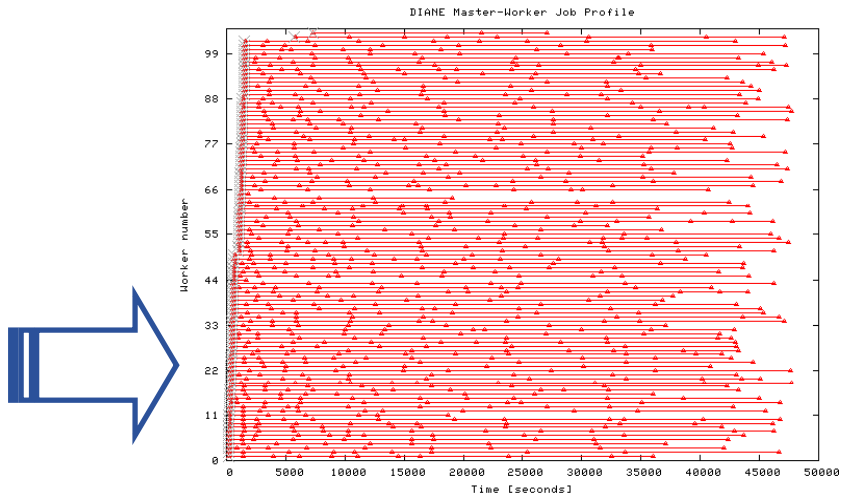


Optimizing resources use: DIANE

- **DIstributed ANalysis Environment**
  - Started at CERN in 2001 in a prototyping phase
  - Applied as resource optimization layer in many gridifications projects

Para ver esta película, debe disponer de QuickTime™ y de un descompresor TIFF (LZW).

- **DIANE Arquitecture**
  - Dynamic user-level scheduler in a beyond master/worker model
    - Job splitting and merging, dynamic load balancing, costumizable error recovery
  - Application independent
  - Job execution profile automatically generated



The 1st community: Geant4

- **General purpose toolkit for simulating the tracking and interaction of particles through matter**
- **Currently used in production in several particle physics experiments (BaBar, HARP, ATLAS, CMS and LHCb). Also used in other areas as space science, medical applications and radiation studies**

- Why they need GRID?
  - Geant4 performs 2 mayor releases per year: June and December
  - Before releasing the software has to be tested following a “regression” strategy
    - Set of physical observables are defined and compared between the new candidate and the previous released one
    - The comparison is performed through a large number of configurations based in different energies, particles, physics configurations, geometries and events
      - *This means about 7000 jobs per candidate*
      - *They can provide several candidates during the testing phase*
      - *This is very CPU demanding: about 4 years of CPU*
  - We have only two weeks to perform the whole production before releasing

Many jobs in a short time: Ideal candidate for Gridification

The Geant4 Production

- **Geant4 was the 1st community gridified from scratch**
- **It began to run in 2004 and was fully recognized as WLCG/EGEE community in December 2005**
  - It has its own VO: geant4
  - About 20 sites support the Geant4 production
- **The VO=gear follows exactly the same infrastructure in terms of resources**
- **Indeed the Geant4 experience was critical to create the gridification infrastructure we provide**
- **From December 2006 the production is totally performed using Ganga/DIANE**

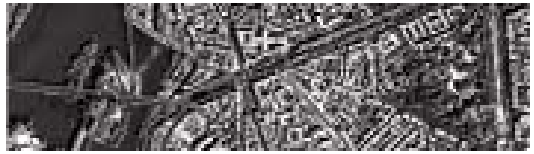
The HARP gridification follows exactly the same procedure as Geant4

UNOSAT is a United Nations Initiative

- **Objectives**
  - Provide the humanitarian community with access to satellite imagery and Geographic Information System services
    - Reduce disasters and plan sustainable development
  - Ensure cost-effective and timely products
- **Core Services**
  - Humanitarian services
  - image processing

VEGETATION – 1 Km

IKONOS – 1m



- **The UNOSAT Gridification project**
  - Bringing the Grid to mobile devices
    - Via GPS the user can know the coordinates of his physical position
    - Using this information as input the user obtains an image of his placement previously stored in the Grid
  - Hiding completely the Grid using web services

Final Remarks

- **The list of "Gridifications" is much larger**
- **ITU**
  - May 2006: The International Telecommunication Union organized a world conference to establish a new frecuency plan for the introduction of digital broadcasting in Europe, Africa, Arab States and former Russian Federation States
  - The software developed by the European Broadcasting Union (EBU) performs compatibility analysis between digital requirements and existing analogue broadcasting stations
  - The CPU of each job is not predictable and the duration of the jobs variates from few seconds until several hours
- **Garfield**
- **We have been able to create a common infrastructure and procedure for new communities**
  - In terms of dedicated resources for new communities, a full support and a set of tools already used by many communities