DATA AND NETWORK REQUIREMENTS

The application-specific work packages identified for the EUROGRID project are described in the following areas:

Bio Grid. The BioGRID project develops interfaces to enable chemists and biologists to submit work to high performance center facilities via a uniform interface from their workstations, without having to worry about the details of how to run particular packages on different architectures.

Metro Grid. The main goal of the Metro Grid project is the development of an application service provider (ASP) solution, which allows anyone to run a high resolution numerical weather prediction model on demand.

Computer-Aided Engineering (CAE) Grid. This work project focuses on industrial CAE applications including automobile and aerospace industries. It aims at providing services to high performance computing (HPC) customers who require huge computing power to solve their engineering problems.

The major partners in this work package are Debis SystemHaus and EADS Corporate Research Center. They are working to exploit the CAE features like code coupling (to improve system design by reducing the prototyping and testing costs) and ASP-type services (designing application-specific user interfaces for job submission).

High Performance Center (HPC) Research Grid. This HPC research grid is used as a test-bed for the development of distributed applications, and as an arena for cooperative work among major scientific challenges, using computational resources distributed on a European scale. The major partners in this work-package are the HPC centers.

The EUROGRID software is based on the UNICORE system developed and used by the leading German HPC centers.

European Union: Data Grid Project

DataGrid[15] is a project funded by the European Union that aims to enable access to geographically distributed computing power and storage facilities belonging to different institutions. This will provide the necessary resources to process huge amounts of data coming from scientific experiments in different disciplines.

The three real data-intensive computing applications areas covered by the project are:

- High Energy Physics
- Biology and Medical Image Processing
- Earth Observations

High Energy Physics (led by CERN, Switzerland)

One of the main challenges for High Energy Physics is to answer longstanding questions about the fundamental particles of matter and the forces acting between them. In particular, the goal is to explain why some particles are much heavier than others, and why particles have mass at all. To that end, CERN is building the Large Hadron Collider (LHS), one of the most powerful particle accelerators.

The search on LHS will generate huge amounts of data. The DataGrid Project is providing the solution for storing and processing this data. A multitiered, hierarchical computing model will be adopted to share data and computing power among multiple institutions. The Tier-0 center is located at CERN and is linked by high-speed networks to approximately 10 major Tier-1 data-processing centers. These will fan out the data to a large number of smaller ones (Tier-2).

Biology and Medical Image Processing (led by CNRS, France)

The storage and exploitation of genomes and the huge flux of data coming from post-genomics puts growing pressure on computing and storage resources within existing physical laboratories. Medical images are currently distributed over medical image production sites (radiology departments, hospitals).

Although there is a need today, as there is no standard for sharing data between sites, there is an increasing need for remote medical data access and processing.

The DataGrid project's biology test-bed is providing the platform for the development of new algorithms on data mining, databases, code management, and graphical interface tools. It is facilitating the sharing of genomic and medical imaging databases for the benefit of international cooperation and health care.

Earth Observations (led by ESA/ESRIN, Italy)

The European Space Agency missions download 100 gigabytes of raw images per day from space. Dedicated ground infrastructures have been set up to handle the data produced by instruments onboard the satellites. The analysis of atmospheric ozone data has been selected as a specific test-bed for the DataGrid. Moreover, the project will demonstrate an improved way to access and process large volumes of data stored in distributed European-wide archives.

TeraGrid

The TeraGrid[16] project was first launched by the NSF and was a multiyear effort to build and deploy the world's largest, fastest distributed infrastructure for open scientific research. The TeraGrid includes 20 teraflops[17] of computing power distributed at five sites, facilities capable of managing and storing nearly 1 petabyte of data, high-resolution visualization environments, and toolkits for Grid Computing. These components will be tightly integrated and connected through a network that will operate at 40 gigabits per second—this is the fastest research network on the planet today.

The major objective of this project includes creation of a high-speed network; grid services that provide data sharing, computing power, and collaborative visualization; and to provide facilities that create the technology requirements (e.g., data storage, bandwidth, etc.).

The five sites in the project are:

- National Center for Supercomputing Applications (NCSA) at the University of Illinois
- San Diego Supercomputer Center (SDSC) at the University of California
- Argonne National Laboratory in Argonne, Illinois
- Center for Advanced Computing Research (CACR) at the California Institute of Technology in Pasadena
- Pittsburgh Supercomputer Center (PSC)

The TeraGrid project is sometimes called a "cyberinfrastructure" that brings together distributed scientific instruments, terascale and petascale data archives, and gigabit networks.

Base Grid Services Layer (Resource Layer)

Some of the base services required for the TeraGrid are authentication and access management, resource allocation and management, data access and management, resource information service, and accounting. This layer forms the building block for the other high-level services.

Core Grid Services (Collective Layer)

With a main focus on coordination of multiple resources, core grid services include functionalities for data movement, job scheduling, monitoring, and resource discovery.

Advanced Grid Services

These are high-level application services, which provide super schedulers, repositories, categorization, resource discovery, and distributed accounting.

Based on the above architecture, the TeraGrid is defining protocols, schema, and interfaces at each layer of the above architecture but not implementation-specific details. These interfaces provide interoperability between the sites implementing the TeraGrid project.

Source : http://elearningatria.files.wordpress.com/2013/10/ise-viii-grid-computing-06is845-notes.pdf