Grid Computing for LHC

Emrah Akkoyun
TUBITAK-ULAKBIM

TR-Grid
High Performance and Grid Computing Center

OUTLINE

- Who we are?
  - TR-Grid infrastructure
  - Leading Projects and HEP Studies
- Grid Computing
  - Introduction to Grid Computing Technology
  - The Architecture and Applications
- WLCG
  - Tiers of WLCG
- A brief introduction to interacting with grid middleware
  - Grid Job Management

Who we are?

- TUBITAK
  - The Scientific and Technological Research Council of Turkey (TUBITAK) established in 1963 is an autonomous institution and is governed by a scientific board
  - Nation-wide research is conducted in 15 different research institutes.
  - The national coordinating body of EU's Framework Program

Who we are?

- ULAKBIM
  - Turkish National Academic Network and Information Center is an institute of TUBITAK which is responsible for:
    - Managing national grid initiative and infrastructure (NGI)
    - Managing national academic network ULAKNET (NREN)
    - Acting as a (digital) library for academic publications

TR-Grid

- TR-Grid initiative was established in 2003 with a MoU between:
  - TUBITAK-ULAKBIM
  - Bogazici University
  - Bilkent University
  - İstanbul Technical University
- Then, it was extended with the participation of four universities
- SEEGRID, SEEGRID-II, SEE-GRID-SCI, EUMEDGRID, EUMEDGRID-Support, EGEE-II, EGEE-III and EGI are the international projects that TR-Grid have been involved
- Under the management of ULAKBIM, TR-Grid NGI has been coordinating national grid related activities for 7 years

EGI Certified Sites

	Number of Cores	Storage (Tbyte)	Memory per Cores (GB)	Interconnectivity
TR-01-ULAKBIM	5760	30	3	Infiniband (QDR)
TR-03-METU	312	250	1	Gigabit Ethernet
TR-04-ERCIYES	64	1	1	Gigabit Ethernet
TR-05-BOUN	64	1	1	Gigabit Ethernet
TR-09-ITU	64	1	1	Gigabit Ethernet
TR-10-ULAKBIM	384	740	3	Infiniband (QDR)
TOTAL	**6648**	**1023**		

TR-Grid HPC Resources

	Number of Cores	Storage (TByte)	Memory per Core (GB)	Interconnectivity	Theoretical Computing Performance (Tflops)
TR-01-ULAKBIM	5760	30	3	Infiniband	13
METU - CENG	360	12	2	Infiniband	3,8
YDU – IBM	1280	20	2	Infiniband	12

- The national HPC users can also utilize other computing centers, which are mostly managed by technical staffs on the universities

TR-Grid Leading Projects

- CERN Experiment
  - Both atlas and cms are supported with two production T2 centers
  - 9800 CPU (HEP-Spec) and 900 TB is dedicated in this year
- Earth Sciences
- Seismology
  - Data repository for seismic waveform obtained by geographically distributed stations
  - ~7.6 M files, ~2 TB data
  - National data as well as other countries in SEE region as part of SEE-Grid-SCI project
- Meteorology
  - ~20 TB data for processing
- Network Flow on WAN
  - ~18 TB data for storing
- Individual users scientific data

HEP Studies

- TR-Grid has two T2 centers for supporting ATLAS and CMS experiment.
  - TR-10-ULAKBIM for atlas
  - TR-03-METU for cms

		2009			2010			2011	
Turkey, Turkish Tier-2 Federation	**ATLAS**	**CMS**	**TOTAL**	**ATLAS**	**CMS**	**TOTAL**	**ATLAS**	**CMS**	**TOTAL**
CPU (HEP-SPEC06)	2800	2600	**5400**	5100	4700	**9800**	5100	4700	**9800**
Disk (Tbytes)	340	210	**550**	550	350	**900**	550	350	**900**

Introduction to Grid Computing

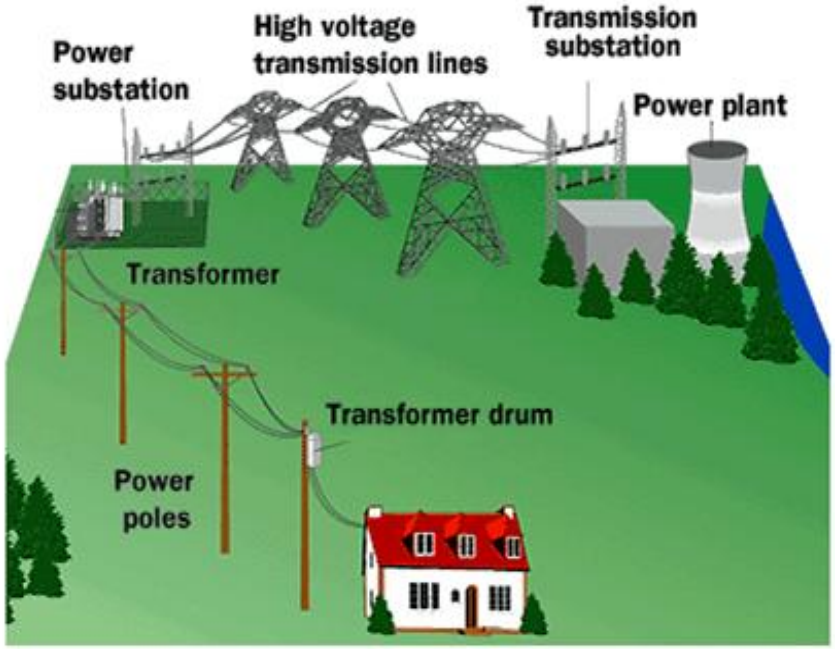
- World wide web uses the internet to share information
- Grid computing uses the internet to share computer power
- The purpose:
  - Forming a global network of computers that can be operated as one vast computational resources by integrating existing computing centers over the world which help a specific group of researchers

HPC and Grid Computing Model

- In a standard HPC model;
  - There exists many supercomputing and national centers in 1990s
  - To utilize computing resources, it is required to have an account and connect to a front-end system

- In Grid model;
  - The centers are integrated
  - A researcher can use a number of computing centers by using a single sign-on

Grid Computing

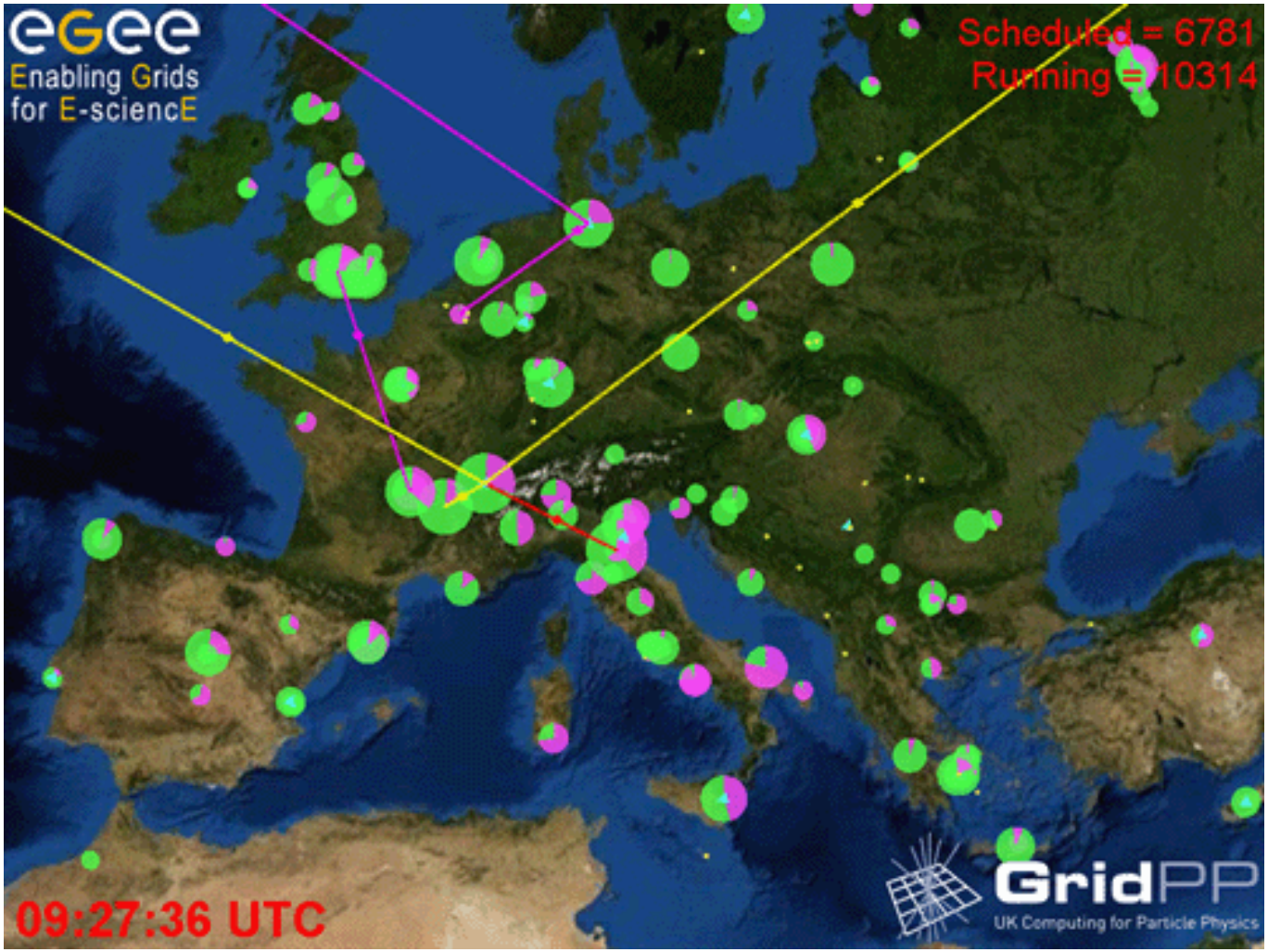


- "The Computational Grid is analogous to Electricity (Power) Grid and the vision is to offer a dependable, consistent, pervasive, and inexpensive access to high-end resources irrespective their location of physical existence and the location of access."

Global Resource Share

- Access to remote computing and storage resources, software and data
- Access and control of remote sensors, telescopes and other devices
- With the improvement on the network technology, a large amount of data can be sent to another computing center and analyzed

Real Time Monitoring - RTM



The Grid Architecture

- Network Layer
  - connects grid resources
- Resource Layer
  - computers, storage systems, sensors, telescopes
- Middleware Layer
  - Essential for WLCG to work
  - organizes and integrates the resource in grid to create a single and seamless computational grid
    - EMI
    - Globus Toolkit
- Application Layer
  - applications and development toolkits to support applications

High-Throughput Problems

- The large problem is divided into many independent tasks each of which is scheduled on different computer processors.
- It is possible to perform hundreds of tasks simultaneously

- Examples:
  - Large Hadron Collider: the analyses of thousands of particle collisions
  - WISDOM project: the analyses of thousands of molecules to discover a drug

Grid Applications

- High-Energy Physics: LHC, Tevatron, HERA
- Biology: Medical Images, Bioinformatics, Drug Discovery
- Earth Science: Hydrology, Pollution, Climate, Geophysics
- Astrophysics: Planck, MAGIC
- Fusion
- Computational Chemistry
- …

Grid computing for LHC

- In the beginning, the required computing power available at CERN was not sufficient for LHC data analysis
- Most of the universities and institutes involved in the collaboration had national and regional computing centre

  - Raised Question: Could it possible to integrate these facilities to provide a single LHC computing service?
  - The remarkable improvement on the wide area network
    - increasing capacity and bandwith with falling costs

The growth on the grid computing

- Computing power doubles every 18 months
- Network performance is doubling every nine months
- Data storage density is doubling every 12 months

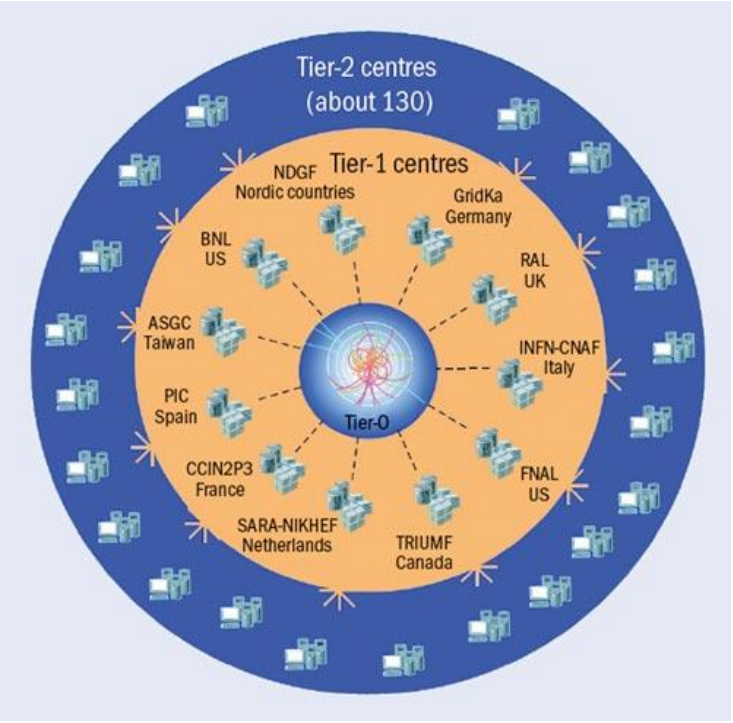
- RESULT:
  - Networks become faster and individual computers become powerful. As a result, grid concept becomes more feasible for solving increasingly complex problem.

Worldwide LHC Computing Grid (WLCG)

- a global collaboration linking to hundreds of data centers worldwide
- enables researchers to analyze 15 Petabytes of data annually generated by the LHC distributed stored over the world.
- a distributed computing infrastructure built by integrating thousands of computers and storage systems
- no matter where the resources are located
- more than 8000 physicists around the world can access and process the LHC data

WLCG - The World's Largest Computing Grid

- Based on two main global grids
  - EGI (European Grid Infrastructure) in Europe
  - OSG (Open Science Grid) in United States



Tiers of WLCG

- Tier-0
  - CERN Computer Centre to where all data from the LHC passes
  - Provides less than 20% of the total compute capacity
  - Responsibilities
    - safe-keeping of the raw data (first copy) and reconstruction
    - distribution of raw and reconstructed data over the Tier-1s

Tiers of WLCG

- Tier-1
  - Large computer centres with sufficient storage and computing capacity as well as high speed connectivity with other centres
  - There are 11 Tier-1 centres available.
  - Responsibilities:
    - storing proportional share of raw and reconstructed data
    - large scale reprocessing and storing generated output
    - distribution data to Tier-2 center and storing the simulation data generated by Tier-2 centres

Tiers of WLCG

- The list of Tier-1 centers over the world

Canada	**TRIUMF**
Germany	**KIT**
Spain	**Port d'Informació Científica (PIC)**
France	**IN2P3**
Italy	**INFN**
Nordic countries	**Nordic Datagrid Facility**
Netherlands	**NIKHEF / SARA**
Taipei	**ASGC**
United Kingdom	**GridPP**
USA	**Fermilab-CMS**
USA	**BNL ATLAS**

Tiers of WLCG

- Tier-2 centers
  - Typically universities and scientific institutes
  - 140 Tier-2 centers are currently available
  - Responsibilities:
    - Storing sufficient data and providing adequate computing power
    - Handling proportional share of simulated event production and reconstruction
- Tier-3 centers
  - Local small cluster at universities or an individual PC
  - No formal engagement between WLCG and Tier-3 resources

Interacting with Grid Middleware

- Installing a ssh client and having an account on UI
- Applying a grid certificate and generating private and public keys
- Being member of relevant virtual organization
- Creating a proxy
- Managing your jobs and data

Generating individual keys

```
feyza@lufer:~$ cp feyzaeryol.pl2 .globus
feyza@lufer:~$ cd .globus
feyza@lufer:~/.globus$ openssl pkcs12 -clcerts -in feyzaeryol.pl2 -out usercert.pem
Enter Import Password:
MAC verified OK
Enter PEM pass phrase:
Verifying - Enter PEM pass phrase:
feyza@lufer:~/.globus$ openssl pkcs12 -nocerts -in feyzaeryol.pl2 -out userkey.pem
Enter Import Password:
MAC verified OK
Enter PEM pass phrase:
Verifying - Enter PEM pass phrase:
feyza@lufer:~/.globus$ ls -lrt
total 12
-rw-r--r--  1 feyza trgridb 3938 Apr 29 23:24 feyzaeryol.pl2
-rw-r--r--  1 feyza trgridb 3574 Apr 29 23:25 usercert.pem
-rw-r--r--  1 feyza trgridb 1900 Apr 29 23:25 userkey.pem
feyza@lufer:~/.globus$ chmod 644 usercert.pem
feyza@lufer:~/.globus$ chmod 600 userkey.pem
```

Checking the certificate

```
egitiml@lufer:~$ grid-cert-info
Certificate:
    Data:
        Version: 3 (0x2)
        Serial Number: 3132 (0xc3c)
        Signature Algorithm: sha1WithRSAEncryption
        Issuer: C=GR, O=HellasGrid Demos, OU=Certification Authoritites, CN=HellasGrid Demo CA 2006
        Validity
            Not Before: Apr 13 07:35:02 2009 GMT
            Not After : Apr 28 07:35:02 2009 GMT
        Subject: C=GR, O=HellasGrid Demos, OU=People, L=Turkey - Grid User Training for Local Commun
        Subject Public Key Info:
            Public Key Algorithm: rsaEncryption
            RSA Public Key: (1024 bit)
                Modulus (1024 bit):
                    00:ce:8b:60:8c:6d:22:7c:cf:4d:73:e8:4c:9f:a6:
                    02:6d:11:8f:3e:83:e0:f2:32:f2:3c:09:cc:20:3a:
```

Creating a proxy

- To be authorized to utilize the resources for 12 hours, it requires to create a proxy
  - $ voms-proxy-init –voms sgdemo
- To obtain information about the generated proxy
  - $ voms-proxy-info --all
- It is possible to extend the validation time of created proxy by using myproxy

```
egitiml@lufer:~$ voms-proxy-init --voms sgdemo
Cannot find file or dir: /home_palamut2/egitim/egitiml/.glite/vomses
Enter GRID pass phrase:
Your identity: /C=GR/O=HellasGrid Demos/OU=People/L=Turkey - Grid User Training for I
Creating temporary proxy ..................................................... Done
Contacting  voms.irb.hr:15012 [/C=HR/O=edu/OU=irb/CN=host/voms.irb.hr] "sgdemo" Done
Creating proxy ............................................................... Done
Your proxy is valid until Sat Apr 18 10:45:49 2009
egitiml@lufer:~$
```

Questions should be answered before submission of jobs

- The following points should be decided
  - Which executable, codes will be submitted?
  - Which data will be used?
  - The amount of data will be sent
  - Whether the executables are depend on the Operating System or Library

Job Description Language (JDL)

- JobType – Normal (simple, sequential job), Interactive, MPICH, Checkpointable
- Executable – The command will be executed
- Arguments – The arguments passing to command
- StdInput, StdOutput, StdError – The input, output and error files
- Environment – Environment variables
- InputSandbox – The input files that application needs
- OutputSandbox – The output files after job completion
- Requirements – Application specific requirements

An example of JDL

```
Executable = "/bin/sh";

Arguments = "HelloWorld.sh";

Stdoutput = "stdoutput.txt";

StdError = "stderror.txt";

InputSandbox = {"HelloWorld.c","HelloWorld.sh"};

OutputSandbox = {"stdoutput.txt","stderror.txt"};

Requirements  = (other.GlueHostOperatingSystemName == “linux");

Rank = other.GlueCEStateFreeCPUs;
```

Listing the appropriate sites

- With the JDL, a user defines the needs of the jobs. The list of sites which satisfy the needs of the application can be found as follow:
  - $glite-wms-job-list-match –a <job.jdl>
- A user can specify a single site where the jobs will be submitted
  - Requirements = other.GlueCEUniqueID == "ce.ulakbim.gov.tr:2119/jobmanager-lcgpbs-sgdemo"

Listing the appropriate sites

```
egitiml@lufer:~/is-gonderme/examplel$ glite-wms-job-list-match -a  HelloWorld.jdl

Connecting to the service https://wms.ulakbim.gov.tr:7443/glite_wms_wmproxy_server

==========================================================================

                     COMPUTING ELEMENT IDs LIST
 The following CE(s) matching your job requirements have been found:

         *CEId*
 - ce.iiap-cluster.sci.am:2119/jobmanager-pbs-sgdemo
 - ce.iiap-cluster.sci.am:2119/jobmanager-pbs-sgdemo
 - ce.seegrid.hpcc.sztaki.hu:2119/jobmanager-lcgpbs-sgdemo
 - ce.ulakbim.gov.tr:2119/jobmanager-lcgpbs-sgdemo
 - ce01.grid.renam.md:2119/jobmanager-lcgpbs-sgdemo
```

Grid Job management

- To submit a job to grid:
  - glite-wms-job-submit –a [--vo <VO>] [-o <file_name>] <job.jdl>
  - where --vo indicates the virtual organization and -o indicates the file where the job ID will be stored
- To monitor the status of the submitted job:
  - glite-wms-job-status [–i <dosya_ismi>] [job ID]
  - where -i indicates the name of the input file where the job ID is stored
- To get the output of the jobs
  - glite-wms-job-output [–i <dosya_ismi>] [job ID]
- To cancel a submitted job
  - glite-wms-job-cancel <işNumarası>

Job Submission

```
egitiml@lufer:~/is-gonderme/examplel$ glite-wms-job-submit -a -o examplel HelloWorld.jdl

Connecting to the service https://wms.ulakbim.gov.tr:7443/glite_wms_wmproxy_server


====================== glite-wms-job-submit Success ======================

The job has been successfully submitted to the WMProxy
Your job identifier is:

https://wms.ulakbim.gov.tr:9000/3QIRrHwp6IM6gosSNUnRAQ

The job identifier has been saved in the following file:
/home_palamut2/egitim/egitiml/is-gonderme/examplel/examplel

==========================================================================


egitiml@lufer:~/is-gonderme/examplel$
```

Monitoring Job

```
egitiml@lufer:~/is-gonderme/example1$ glite-wms-job-status -i example1

------------------------------------------------------------------
1 : https://wms.ulakbim.gov.tr:9000/rNMyB1syFnA5F68cAuCi_w
2 : https://wms.ulakbim.gov.tr:9000/3QIRrHwp6IM6gosSNUnRAQ
a : all
q : quit
------------------------------------------------------------------

Choose one or more jobId(s) in the list - [1-2]all:2



*************************************************************
BOOKKEEPING INFORMATION:

Status info for the Job : https://wms.ulakbim.gov.tr:9000/3QIRrHwp6IM6gosSNUnRAQ
Current Status:     Done (Success)
Logged Reason(s):
    - Job got an error while in the CondorG queue.
    - Job terminated successfully
Exit code:          0
Status Reason:      Job terminated successfully
Destination:        node001.grid.auth.gr:2119/jobmanager-pbs-sgdemo
Submitted:          Fri Apr 17 22:54:19 2009 EEST
*************************************************************

egitiml@lufer:~/is-gonderme/example1$
```

Getting Job Results

```
egitiml@lufer:~/is-gonderme/example1$ glite-wms-job-output -i example1
----------------------------------------------------------------------
1 : https://wms.ulakbim.gov.tr:9000/rNMyBlsyFnA5F68cAuCi_w
2 : https://wms.ulakbim.gov.tr:9000/3QIRrHwp6IM6gosSNUnRAQ
a : all
q : quit
----------------------------------------------------------------------

Choose one or more jobId(s) in the list - [1-2]all (use , as separator or - for a r

Connecting to the service https://wms.ulakbim.gov.tr:7443/glite_wms_wmproxy_server


================================================================================

                        JOB GET OUTPUT OUTCOME

Output sandbox files for the job:
https://wms.ulakbim.gov.tr:9000/3QIRrHwp6IM6gosSNUnRAQ
have been successfully retrieved and stored in the directory:
/tmp/jobOutput/egitiml_3QIRrHwp6IM6gosSNUnRAQ

================================================================================
egitiml@lufer:~/is-gonderme/example1$ ls -lrt /tmp/jobOutput/egitiml_3QIRrHwp6IM6gosSNUnRAQ
total 4
-rw-r--r--  1 egitiml seismo 12 Apr 17 23:00 message.txt
-rw-r--r--  1 egitiml seismo  0 Apr 17 23:00 stderror
egitiml@lufer:~/is-gonderme/example1$
```

QUESTIONS ?

More Information

- GridCafe, http://www.gridcafe.org/
- WLCG Public Website, http://lcg.web.cern.ch/lcg/public/default.htm
- TR-Grid, www.grid.org.tr