



imath.cesga.es, the VO for the European mathematicians

M.T. Sánchez¹, R. Cao², J. Fernández⁴, M. García-Magariños³, F. García-Torre⁴, J.M. Gesto⁵, A. Gómez¹, W. González-Manteiga³, J.M. Gutiérrez⁴

¹CESGA, Fundación Centro Tecnológico de Supercomputación de Galicia. Santiago de Compostela, Spain

²Departamento de Matemáticas. Universidad de A Coruña. A Coruña, Spain.

³Departamento de Estadística e Investigación Operativa. Universidad de Santiago de Compostela. Spain.

⁴Departamento de Matemática Aplicada y Ciencias de la Computación. Universidad de Cantabria. Spain.

⁵Departament de Matemàtica Aplicada III. ETSECCPB, UPC, Barcelona, Spain.

<http://mathematica.nodo.cesga.es>

tsanchez@cesga.es

- i-MATH Project
 - Main objectives
 - Computing Platform

- GRID-MATHEMATICA Activity
 - imath.cesga.es VO
 - Infrastructure

- Three study cases

- Conclusions

- Ingenio MATHEMATICA (i-MATH) is a CONSOLIDER singular research project.

<http://www.i-math.org>

- To increase mathematical presence on the Spanish and worldwide scene.
 - From basic research to applications
 - How to understand the physical world
 - The essential computational support
 - Direct applications to society
- COMPUTING Platform
 - To promote the use of computational methods.
 - To increase the involvement of Spanish researchers in strategic and hot subjects in scientific computation

GRID-MATHEMATICA

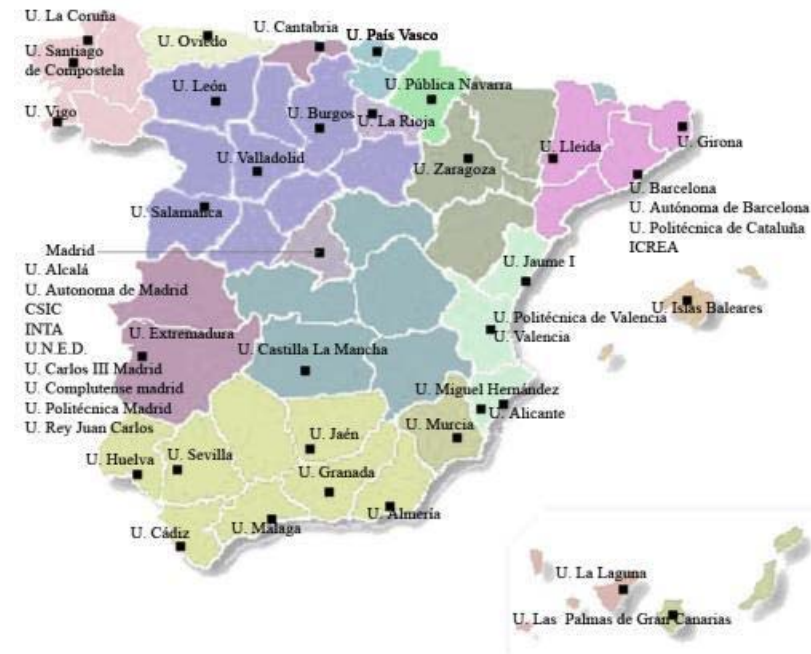
■ Objectives:

- To design a new **Virtual Organization VO** to use in i-MATH.
- To create a first validation prototype by means of executing in a new GRID infrastructure **three cases** of mathematical and computational interest.
- To **expand** this infrastructure to all the members of i-MATH.

Virtual Organization imath.cesga.es

- Creation of VO imath.cesga.es in EGEE infrastructure:
 - First VO for Mathematics Research.
 - 1,600 people
 - 300 research groups

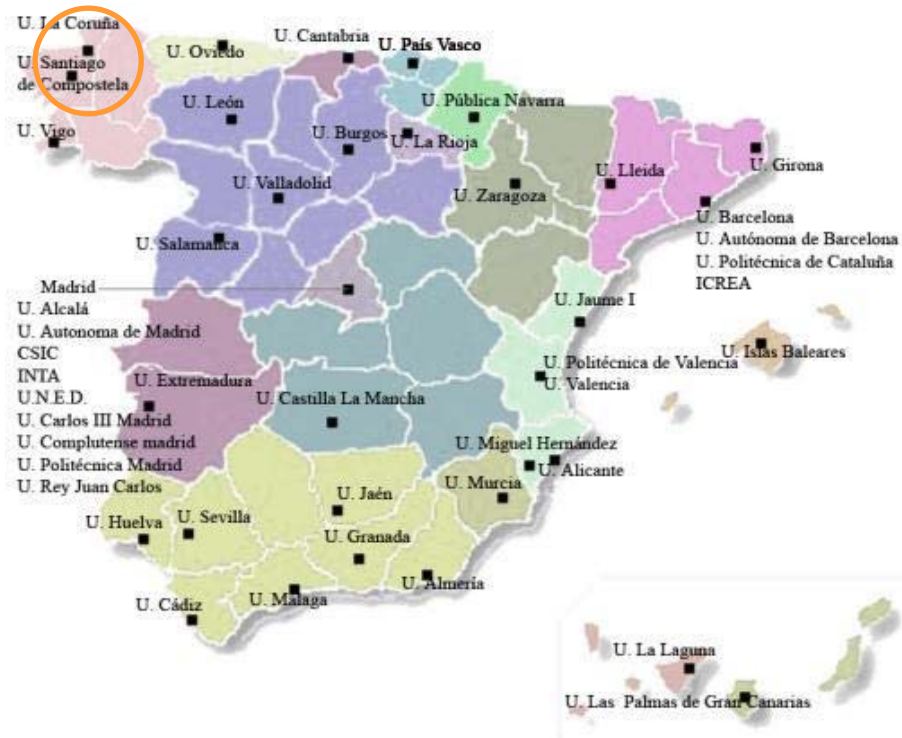
<http://imath.cesga.es>



Infrastructure:

Three sites:

CESGA

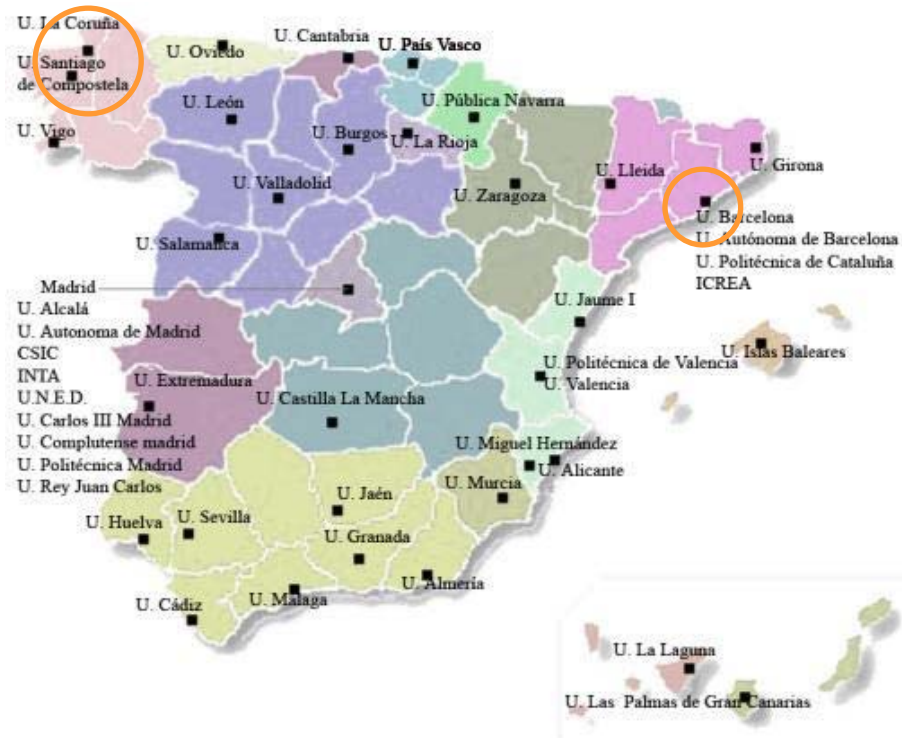


Infrastructure:

Three sites:

CESGA

IMUB



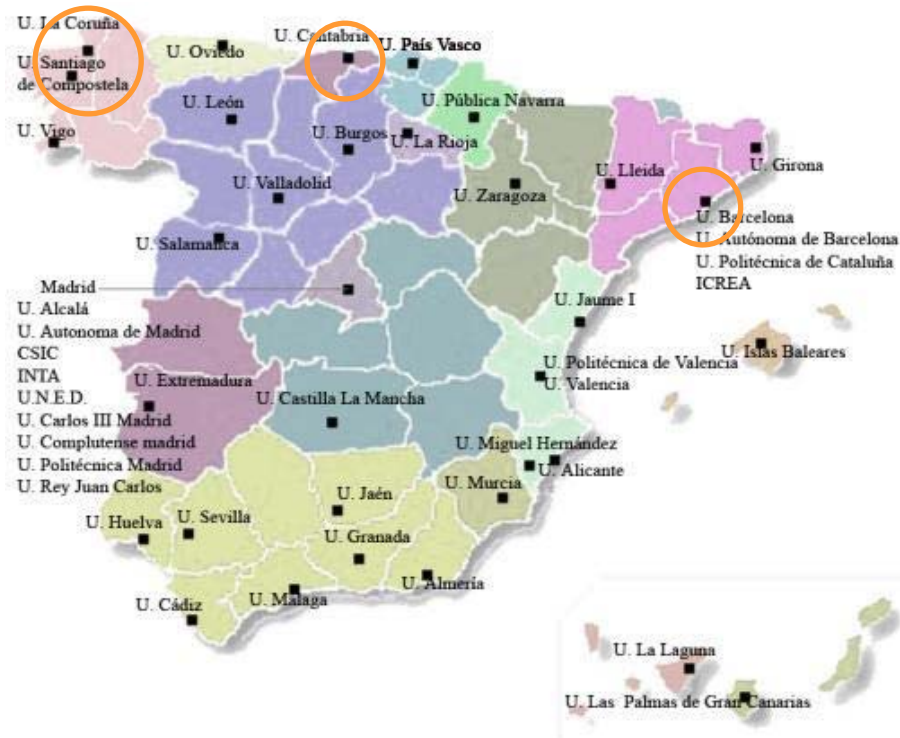
Infrastructure:

Three sites:

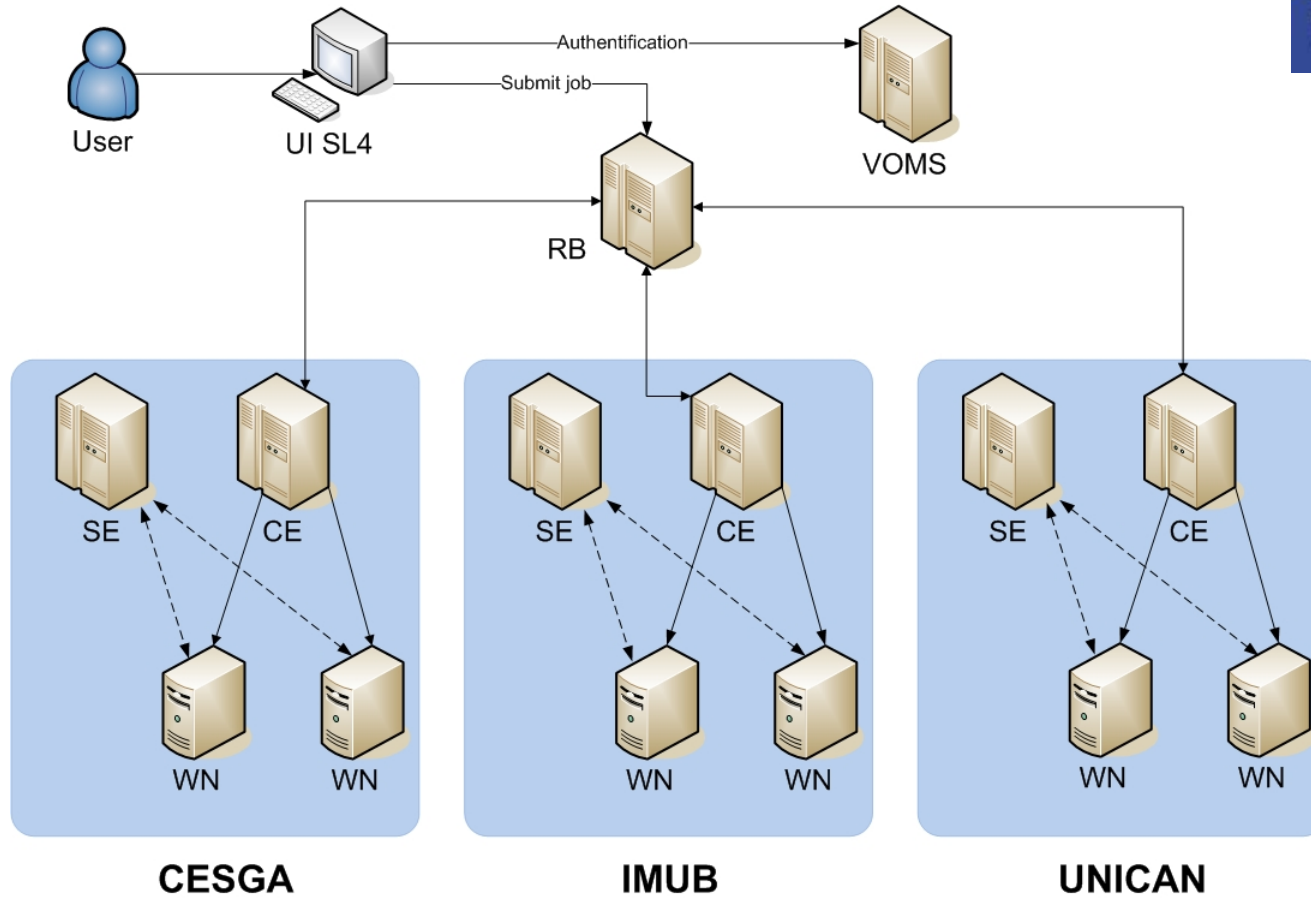
CESGA

IMUB

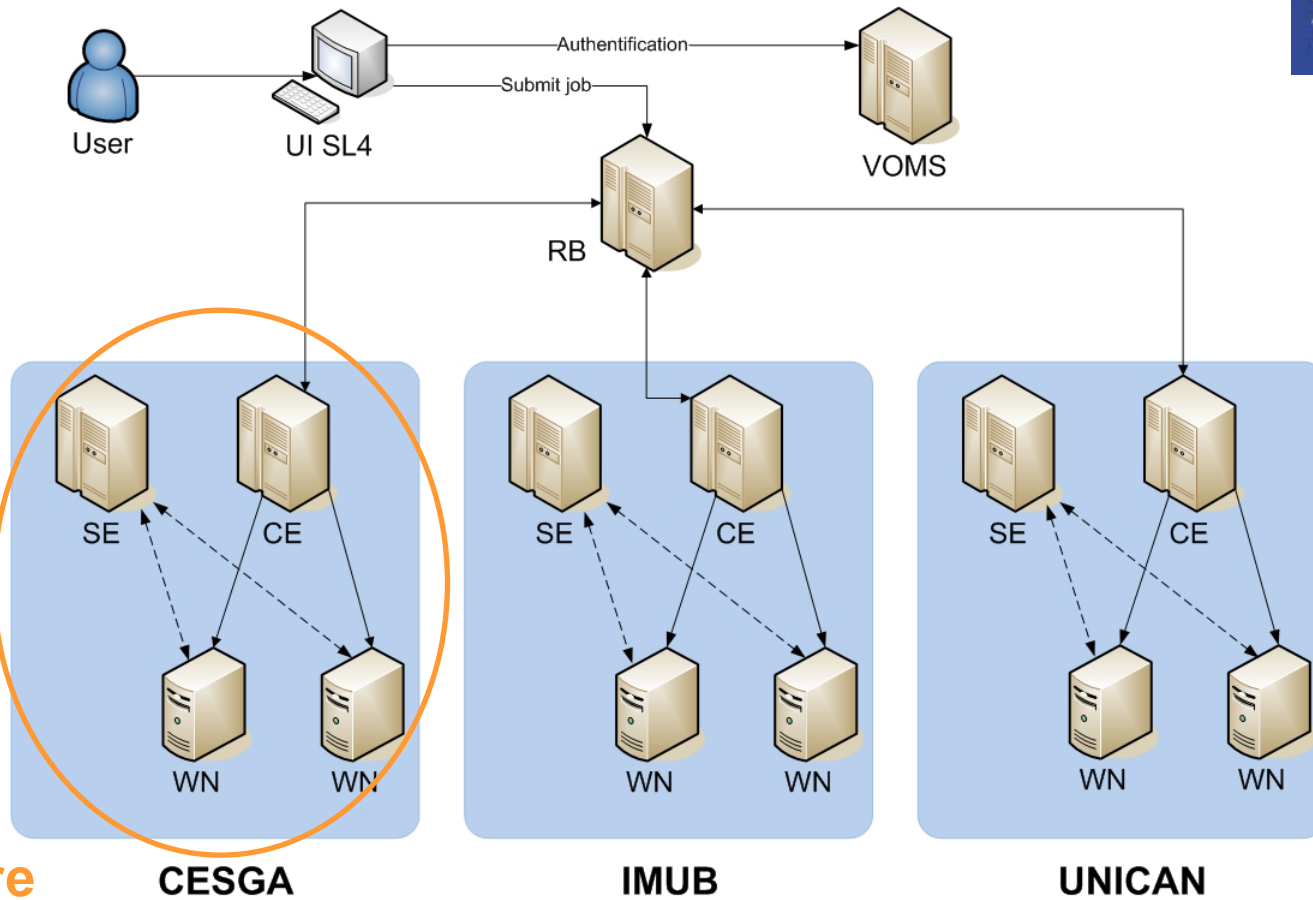
UNICAN



Infrastructure:



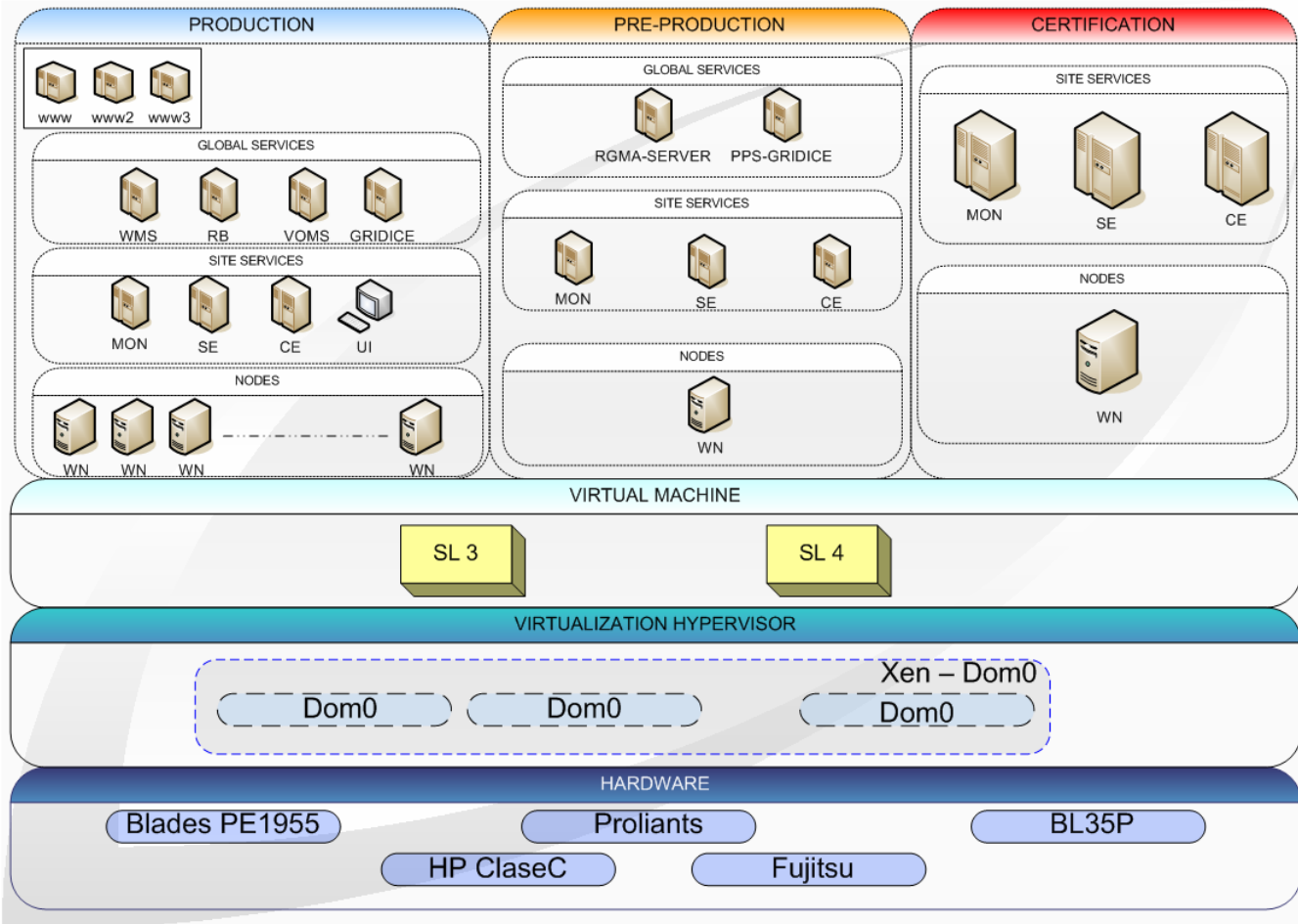
Infrastructure:



XEN VM

Dual Core Xeon

Infrastructure: CESGA-EGEE



356 cpus

1TB storage

VM nodes

Study cases

- To evaluate the applicability in Mathematics research of GRID infrastructure
 - Case 1: Support Vector Machines (SVMs) adaptation to case-control studies with SNP data
 - Case 2: Monte Carlo simulation of climate sensitivity
 - Case 3: Estimation of Fekete points

Study cases

- To evaluate the applicability in Mathematics research of GRID infrastructure
- Make easy to use GRID infrastructure
 - Testing different solutions (Nimrod/G, GridWay, ...)
- P-Grade Grid Portal
 - simplicity
 - powerful workflow



<http://uisl4.imath.cesga.es:8080/gridsphere/gridsphere>

Study cases



Case 1: Support Vector Machines (SVMs) adaptation to case-control studies with SNP data.

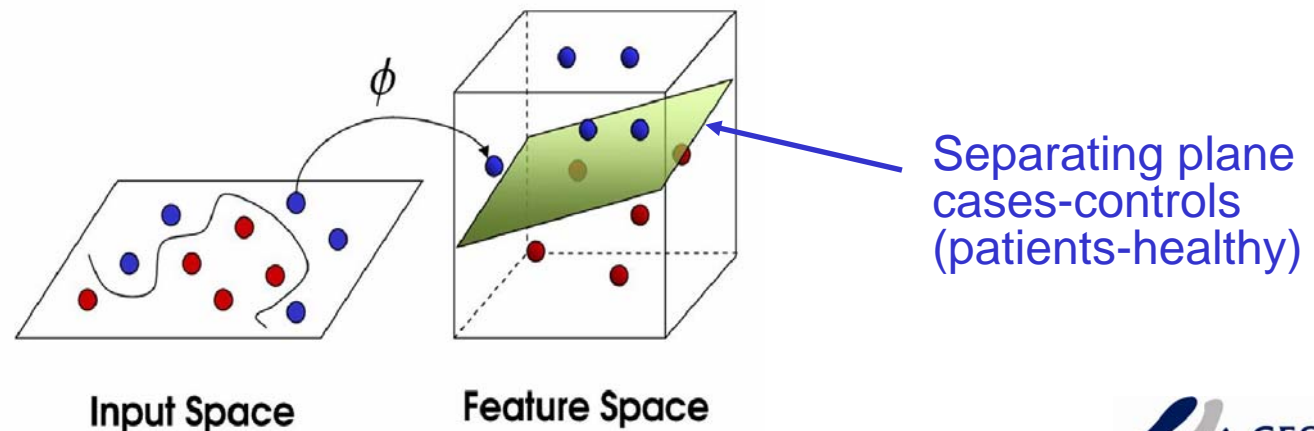
- Genetic association studies are thought as the biggest promise facing the discovery of the genetic basis of common diseases (diabetes, cancer, psychiatric, ...).
- Single Nucleotide Polymorphisms (SNPs) are responsible for around 90% of genetic variability among humans.
- Need to develop new statistical methodologies qualified to dissect (since the genetic point of view) these complex diseases.

Study cases



Case 1: SVMs, how do they work?

- Construction of a linear classifier in a high-dimensional feature space.
- This is made by means of kernels, similarity measures between individuals obtained from their SNP values.
- Compulsory adaptation to work with SNPs (categorical data).



Study cases



Case 1: SVMs, need of GRID techniques

- SVMs are itself computationally demanding.
- Our adaptation, carried out to work with SNPs, transform each individual genotype in a number of configurations that is two up to the power of a great number, increasing dramatically the computational efforts required.
- Large amounts of genetic data to be studied

Study cases



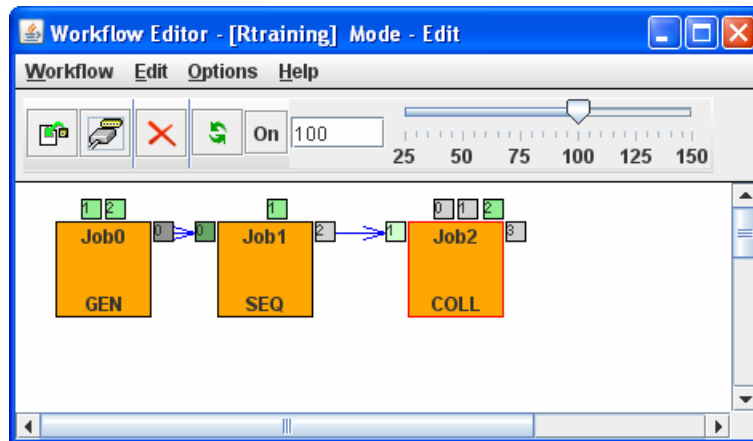
Case 1: SVMs, need of GRID techniques

- SVMs are itself computationally demanding.
- Our adaptation, carried out to work with SNPs, transform each individual genotype in a number of configurations that is two up to the power of a great number, increasing dramatically the computational efforts required.
- Large amounts of genetic data to be studied
- Needed software: Statistical Free Software R.
- Completely installed in CESGA-EGEE

Study cases



Case 1: Support Vector Machines, training-test



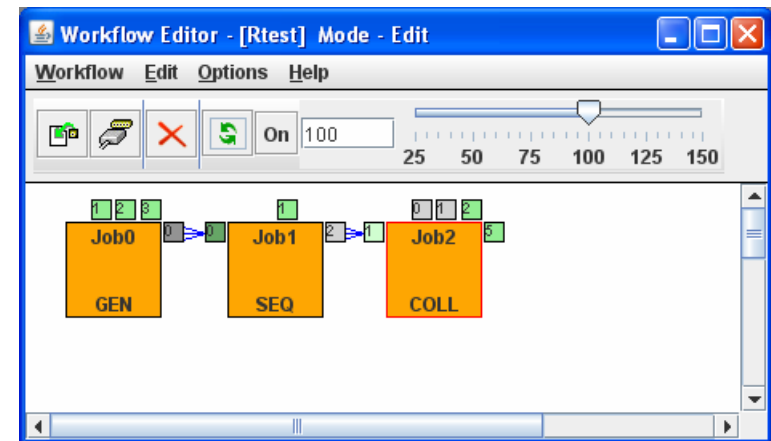
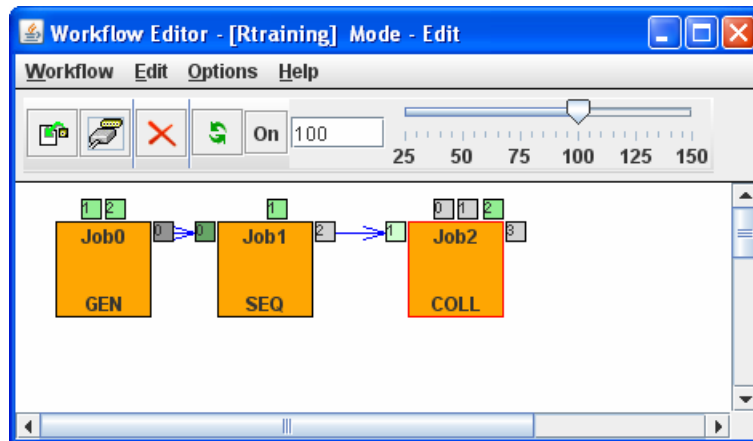
Training phase

Prediction case-control

Study cases



Case 1: Support Vector Machines, training-test



Training phase

Prediction case-control

Test phase

Method check with data

Study cases



Case 1: Support Vector Machines, training

The screenshot shows the PGrade Grid portal interface. The main content area displays the 'Workflow Manager' section with the following data:

PS workflow details						
PS Workflow	Status	[Output]	[Logs]	[Action]		
Rtraining	running	N/A	-	Suspend	Abort	Attach Delete

Jobs in generator phase						
Job	Status	Type	Gridname	Hostname	[Logs]	
Job0	running	generator	imath.cesga.es_LCG_2_BROKER	ce2.egee.cesga.es	--	Log

eWorkflow list Statistics						
Total	Init	Submitted	Rescue	Error	Finished	
-	-	-	-	-	-	

There is no submitted eWorkflow.

cpus: 20 Time: 140 min

Study cases



Case 1: Support Vector Machines, training

PGrade Grid portal - Mozilla Firefox

http://uis4.imath.cesga.es:8080/gridsphere/gridsphere?cid=77&gs_action=doRefreshList

Job	Status	Type	Gridname	Hostname	[Logs]
Job0	finished	generator	imath.cesga.es_LCG_2_BROKER	ce2.egee.cesga.es	Out Log

eWorkflow list					
Statistics					
Total	Init	Submitted	Rescue	Error	Finished
20	0	20	0	0	0

Workflow	Status	[Output]	[View]	[Action]
Rtraining.1	submitted	N/A	Details	Suspend Abort
Rtraining.10	submitted	N/A	Details	Suspend Abort
Rtraining.11	submitted	N/A	Details	Suspend Abort
Rtraining.12	submitted	N/A	Details	Suspend Abort
Rtraining.13	submitted	N/A	Details	Suspend Abort
Rtraining.14	submitted	N/A	Details	Suspend Abort
Rtraining.15	submitted	N/A	Details	Suspend Abort
Rtraining.16	submitted	N/A	Details	Suspend Abort
Rtraining.17	submitted	N/A	Details	Suspend Abort
Rtraining.18	submitted	N/A	Details	Suspend Abort

Terminado

cpus: 20 Time: 140 min

Study cases



Case 1: Support Vector Machines, training

The screenshot shows the PGrade Grid portal interface in Mozilla Firefox. The main content area is titled 'Workflow Manager' and displays the following information:

PS workflow details

PS Workflow	Status	[Output]	[Logs]	[Action]
Rtraining	running	N/A	-	Suspend Abort Attach Delete

Jobs in generator phase

Job	Status	Type	Gridname	Hostname	[Logs]
Job0	finished	generator	imath.cesga.es_LCG_2_BROKER	ce2.egee.cesga.es	Out Log

eWorkflow list

Statistics

Total	Init	Submitted	Rescue	Error	Finished
20	0	0	0	0	20

There is no submitted eWorkflow.

Jobs in collector phase

Job	Status	Type	Gridname	Hostname	[Logs]
Job2	running	collector	imath.cesga.es_LCG_2_BROKER	ce2.egee.cesga.es	.. Log

7 de mayo de 2008

cpus: 20 Time: 140 min

Study cases



Case 1: Support Vector Machines, training

The screenshot shows the PGrade Grid portal interface. The main content area displays the following information:

Workflow Manager

Refresh Back

PS workflow details

PS Workflow	Status	[Output]	[Logs]	[Action]
Rtraining	finished	✓	-	Submit Attach Delete

Jobs in generator phase

Job	Status	Type	Gridname	Hostname	[Logs]
Job0	finished	generator	imath.cesga.es_LCG_2_BROKER	ce2.egee.cesga.es	Out Log

eWorkflow list

Statistics

Total	Init	Submitted	Rescue	Error	Finished
20	0	0	0	0	20

There is no submitted eWorkflow.

Jobs in collector phase

Job	Status	Type	Gridname	Hostname	[Logs]
Job2	finished	collector	imath.cesga.es_LCG_2_BROKER	ce2.egee.cesga.es	Out Log

7 de mayo de 2008

Terminado

cpus: 20 Time: 140 min

Study cases



Case 1: Support Vector Machines, test

The screenshot shows the PGrade Grid portal interface in Mozilla Firefox. The main content area displays workflow details for a job named 'Rtest'. The job status is 'submitted'. Below this, the 'Jobs in generator phase' section shows a single job 'Job0' with a status of 'finished'. The 'eWorkflow list Statistics' section provides a summary of the workflow: Total 20, Init 6, Submitted 14, Rescue 0, Error 0, and Finished 0. At the bottom, a list of individual workflow steps (Rtest.1, Rtest.10, Rtest.11, Rtest.12) is shown, all with a 'submitted' status. The browser's address bar shows the URL: http://uis4.imath.cesga.es:8080/gridsphere/gridsphere?cid=77&gs_action=doRefreshList.

PS Workflow	Status	[Output]	[Logs]	[Action]
Rtest	submitted	N/A	-	Abort Attach Delete

Job	Status	Type	Gridname	Hostname	[Logs]
Job0	finished	generator	imath.cesga.es_LCG_2_BROKER	ce2.egee.cesga.es	Out Log

eWorkflow list Statistics					
Total	Init	Submitted	Rescue	Error	Finished
20	6	14	0	0	0

Workflow	Status	[Output]	[View]	[Action]
Rtest.1	submitted	N/A	Details	Suspend Abort
Rtest.10	submitted	N/A	Details	Suspend Abort
Rtest.11	submitted	N/A	Details	Suspend Abort
Rtest.12	submitted	N/A	Details	Suspend Abort

cpus: 20 Time: 60 min

Study cases



Case 1: Support Vector Machines, test

The screenshot shows the PGrade Grid portal interface. At the top, a table displays the status of Job0 as 'finished'. Below this, the 'eWorkflow list' section provides a summary of workflow statistics:

Total	Init	Submitted	Rescue	Error	Finished
20	0	20	0	0	0

The main workflow list shows 17 individual workflows (Rtest.1 to Rtest.17), all of which are currently in a 'running' state. Each workflow entry includes a 'Status' column (highlighted in red), an '[Output]' column (showing 'N/A'), a '[View]' column with a 'Details' link, and an '[Action]' column with 'Suspend' and 'Abort' buttons.

cpus: 20 Time: 60 min

Study cases



Case 1: Support Vector Machines, test

The screenshot shows the PGrade Grid portal interface. The main content area displays the following sections:

- Workflow Manager:** Includes 'Refresh' and 'Back' buttons.
- PS workflow details:** A table showing workflow status.

PS Workflow	Status	[Output]	[Logs]	[Action]
Rtest	running	N/A	-	Suspend Abort Attach Delete
- Jobs in generator phase:** A table showing job details.

Job	Status	Type	Gridname	Hostname	[Logs]
Job0	finished	generator	imath.cesga.es_LCG_2_BROKER	ce2.egee.cesga.es	Out Log
- eWorkflow list Statistics:** A summary table.

Total	Init	Submitted	Rescue	Error	Finished
20	0	0	0	0	20

There is no submitted eWorkflow.
- Jobs in collector phase:** A table showing job details.

Job	Status	Type	Gridname	Hostname	[Logs]
Job2	running	collector	imath.cesga.es_LCG_2_BROKER	ce2.egee.cesga.es	.. Log

The date at the bottom of the page is 7 de mayo de 2008.

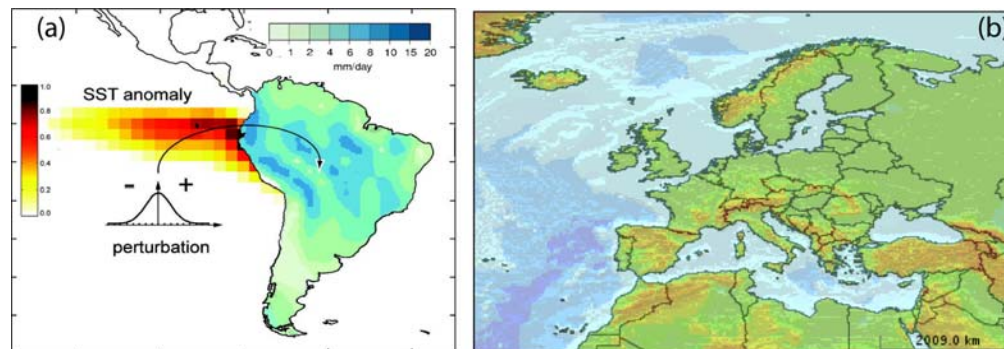
cpus: 20 Time: 60 min

Study cases



Case 2: Monte Carlo simulation of climate sensitivity

- EELA Project: Develop a simulation and analysis tool to predict local impacts of “El Niño” in Latin America
 - Sensitivity study SST-Precipitation
- Goal: To analyze the problem of soil moisture content over Europe

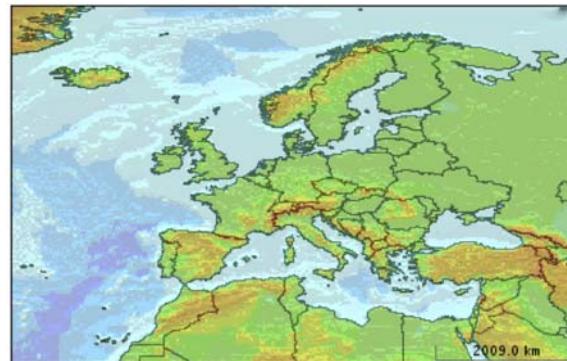


Study cases



Case 2: Monte Carlo simulation of climate sensitivity

- CAM: Global Atmospheric Model.
 - Produce information such as Precipitation or Temperature in a grid of points over the globe.
- Different realizations of the perturbation in a Monte Carlo approach.



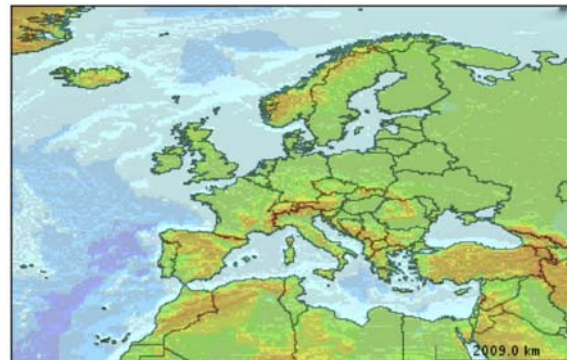
Study cases



Case 2: Climate sensitivity, requeriments

- Runs from days to weeks.
- AMGA metadata catalogs
- 500 GB storage

- Work in progress



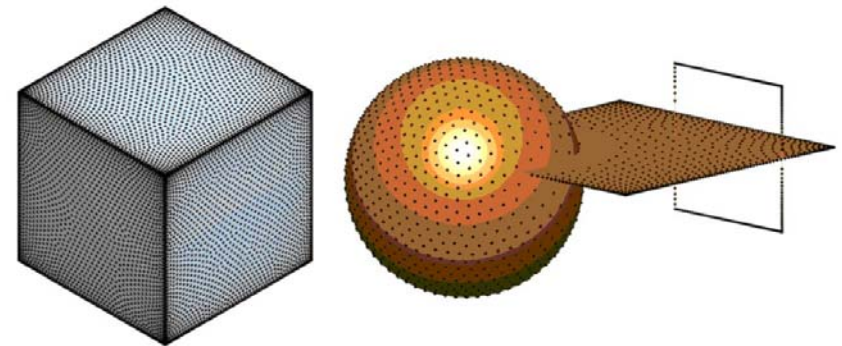
Study cases

Case 3: Estimation of Fekete points

N^{th} order Fekete points are the N -uples $\omega_N = \{x_1, \dots, x_N\}$, $x_i \in \mathbb{R}^d$ which minimize on a compact $S \subset \mathbb{R}^d$ different potential energy functionals \mathcal{I}_N depending on the relative distances among N points.

$$\mathcal{I}_N(x) = \sum_{1 \leq i < j \leq N} \log \frac{1}{|x_i - x_j|}$$

$$\mathcal{I}_N(x) = \sum_{1 \leq i < j \leq N} \frac{1}{|x_i - x_j|^s}$$



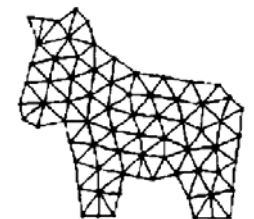
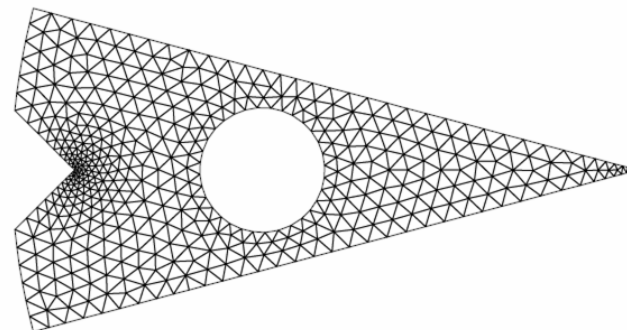
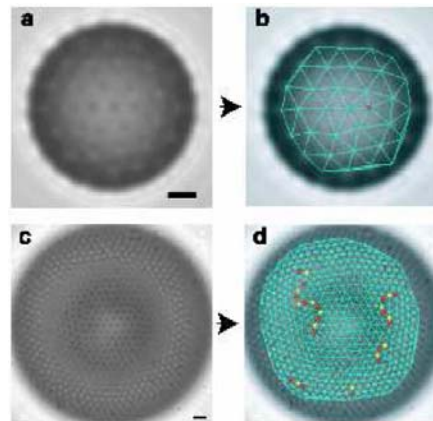
Study cases



Case 3: Estimation of Fekete points

Applications in:

- Physics
- Chemistry
- Numerical Methods
- Complexity Theory

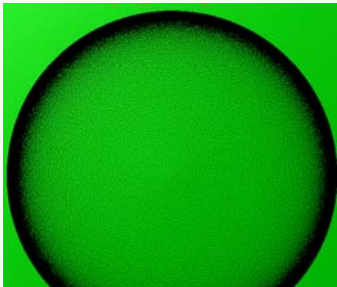


Study cases



Case 3: Estimation of Fekete points

- An algorithm developed for the numerical estimation of local minima.
 - Versatile, robust and efficient
 - Sphere case profusely analyzed
 - Considerable amount of experimental data needed for different geometries and energies.



Bendito, Carmona, Encinas, Gesto, Estimation of Fekete points.
Journal of Computational Physics 225 (2007) 2354-2376.

Study cases



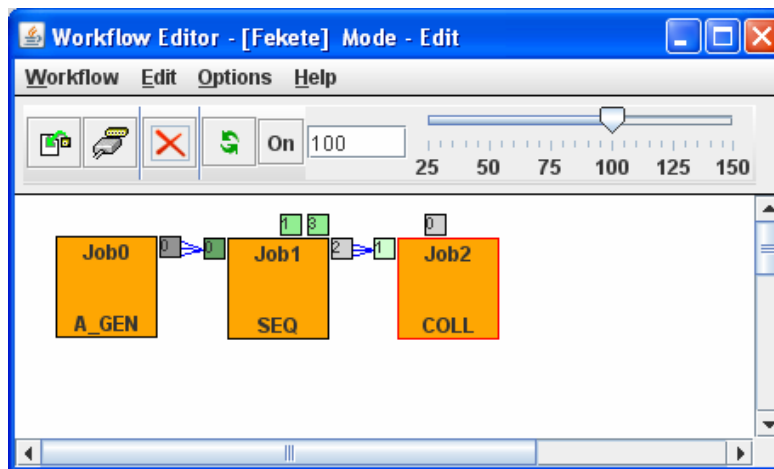
Case 3: Estimation of Fekete points, requirements.

- Lots of relatively small jobs
 - From minutes to some hours
- 2,000 computational hours for each geometry and energy
- 5 GB storage

Study cases



Case 3: Estimation of Fekete points



N=482

3000 runs

30 cpus

100 runs each cpu

Parametric studies

Study cases



Case 3: Estimation of Fekete points

The screenshot shows the P-Grade Grid portal interface. The main content area displays the 'Workflow Manager' section with the following data:

PS workflow details						
PS Workflow	Status	[Output]	[Logs]	[Action]		
Fekete	running	N/A	-	Suspend	Abort	Attach Delete

Jobs in generator phase						
Job	Status	Type	Gridname	Hostname	[Logs]	
Job0	running	auto-generator	imath.cesga.es_LCG_2_BROKER	Portal Server	---	

eWorkflow list Statistics						
Total	Init	Submitted	Rescue	Error	Finished	
-	-	-	-	-	-	

There is no submitted eWorkflow.

cpus: 30 Time: 210 min (in only a computer 4500 min)

Study cases



Case 3: Estimation of Fekete points

The screenshot shows the P-Grade Grid portal interface. The main content area displays workflow details for a job named 'Fekete'. The workflow status is 'submitted'. Below this, a table shows the 'Jobs in generator phase' with one job, 'Job0', which is 'finished'. At the bottom, an 'eWorkflow list Statistics' table shows a total of 30 workflows, with 30 initialized, 0 submitted, 0 rescued, 0 errors, and 0 finished.

eWorkflow list Statistics						
Total	Init	Submitted	Rescue	Error	Finished	
30	30	0	0	0	0	

cpus: 30 Time: 210 min (in only a computer 4500 min)

Study cases



Case 3: Estimation of Fekete points

PGrade Grid portal - Mozilla Firefox

http://uis4.imath.cesga.es:8080/gridsphere/gridsphere?cid=77&gs_action=doRefreshList

Statistics

Total	Init	Submitted	Rescue	Error	Finished
30	20	10	0	0	0

Workflow	Status	[Output]	[View]	[Action]
Fekete.1	submitted	N/A	Details	Suspend Abort
Fekete.10	submitted	N/A	Details	Suspend Abort
Fekete.11	init	N/A	Details	
Fekete.2	submitted	N/A	Details	Suspend Abort
Fekete.3	submitted	N/A	Details	Suspend Abort
Fekete.4	submitted	N/A	Details	Suspend Abort
Fekete.5	submitted	N/A	Details	Suspend Abort
Fekete.6	submitted	N/A	Details	Suspend Abort
Fekete.7	submitted	N/A	Details	Suspend Abort
Fekete.8	submitted	N/A	Details	Suspend Abort
Fekete.9	submitted	N/A	Details	Suspend Abort

Jobs in collector phase

Job	Status	Type	Gridname	Hostname	[Logs]
Terminado					

cpus: 30 Time: 210 min (in only a computer 4500 min)

Study cases



Case 3: Estimation of Fekete points

PGrade Grid portal - Mozilla Firefox

http://uis4.imath.cesga.es:8080/gridsphere/gridsphere?cid=77&gs_action=doRefreshList

Google

eWorkflow list						
Statistics						
Total	Init	Submitted	Rescue	Error	Finished	
30	0	30	0	0	0	
Workflow	Status	[Output]	[View]	[Action]		
Fekete.1	running	N/A	Details	Suspend	Abort	
Fekete.10	running	N/A	Details	Suspend	Abort	
Fekete.11	running	N/A	Details	Suspend	Abort	
Fekete.12	running	N/A	Details	Suspend	Abort	
Fekete.13	running	N/A	Details	Suspend	Abort	
Fekete.14	running	N/A	Details	Suspend	Abort	
Fekete.15	running	N/A	Details	Suspend	Abort	
Fekete.16	submitted	N/A	Details	Suspend	Abort	
Fekete.17	submitted	N/A	Details	Suspend	Abort	
Fekete.18	submitted	N/A	Details	Suspend	Abort	
Fekete.19	submitted	N/A	Details	Suspend	Abort	

Terminado

cpus: 30 Time: 210 min (in only a computer 4500 min)

Study cases



Case 3: Estimation of Fekete points

The screenshot shows the PGrade Grid portal interface in Mozilla Firefox. The main content area is titled 'Workflow Manager' and contains several sections:

- PS workflow details:** A table showing the 'Fekete' workflow with a status of 'running' (highlighted in red). It includes columns for 'Status', '[Output]', '[Logs]', and '[Action]' with buttons for 'Suspend', 'Abort', 'Attach', and 'Delete'.
- Jobs in generator phase:** A table showing 'Job0' with a status of 'finished' (highlighted in green). Columns include 'Job', 'Status', 'Type', 'Gridname', 'Hostname', and '[Logs]'.
- eWorkflow list Statistics:** A summary table with columns: Total (30), Init (0), Submitted (0), Rescue (0), Error (0), and Finished (30). Below it, it states 'There is no submitted eWorkflow.'
- Jobs in collector phase:** A table showing 'Job2' with a status of 'running' (highlighted in red). Columns include 'Job', 'Status', 'Type', 'Gridname', 'Hostname', and '[Logs]'.

The browser's address bar shows the URL: `http://uis4.imath.cesga.es:8080/gridsphere/gridsphere?cid=77&gs_action=doRefreshList`. The system clock at the bottom indicates the date '7 de mayo de 2008'.

cpus: 30 Time: 210 min (in only a computer 4500 min)

Study cases



Case 3: Estimation of Fekete points

The screenshot shows the 'Workflow Manager' interface in a Mozilla Firefox browser. The main content area displays the following information:

PS workflow details

PS Workflow	Status	[Output]	[Logs]	[Action]
Fekete	finished	✓	-	Submit Attach Delete

Jobs in generator phase

Job	Status	Type	Gridname	Hostname	[Logs]
Job0	finished	auto-generator	imath.cesga.es_LCG_2_BROKER	Portal Server	Out ..

eWorkflow list Statistics

Total	Init	Submitted	Rescue	Error	Finished
30	0	0	0	0	30

There is no submitted eWorkflow.

Jobs in collector phase

Job	Status	Type	Gridname	Hostname	[Logs]
Job2	finished	collector	imath.cesga.es_LCG_2_BROKER	ce2.egee.cesga.es	Out . Log

7 de mayo de 2008

cpus: 30 Time: 210 min (in only a computer 4500 min)

Conclusions

- imath.cesga.es VO devoted exclusively to Mathematics
- 3 distributed sites, basing front-ends on XEN
- Pilot to prove the applicability of using GRID in mathematical research.
- Three different study cases
 - Statistics, climatology and fundamental maths.
- Make easy the usage of GRID: P-Grade Grid Portal
 - Simplicity, powerful workflow but some limitations
- imath.cesga.es is open to Mathematics research groups in Europe at large!!!

Thanks for your attention!

<http://mathematica.nodo.cesga.es>

<http://imath.cesga.es>

tsanchez@cesga.es