Astrophysics and the Grid: Experience with EGEE

Fabio Pasian – INAF & VObs.it

IVOA 2007 Interoperability Meeting

Astro-RG session

INAF experience with the grid

- *(from the IVOA 2006 Interop):* In INAF
  - there is a community of astrophysicists (mainly theoreticians) already having "native" access to the "production Grid" facilities
  - training/dissemination activity for Grid-VObs users

- INAF experience with EGEE
  - As EGEE partners
  - Using EGEE
  - Expanding EGEE

EGEE project

egee
Enabling Grids for E-sciencE

**Flagship European grid infrastructure project
Now in 2nd phase (EGEE-II)**

- **Large-scale, production-quality grid infrastructure for e-Science & e-Business**
  - **Currently includes ~200 sites across 40 countries**
  - **~30000 CPUs, 7PB storage, 17 million jobs in 2006**
  - **~180 Virtual Organisations from wide-range of domains**
- **Middleware**
  - **Production quality middleware distribution with business friend open source licence**
- **User Support**
  - **Training & expertise in grid-enabling applications**
  - **Online helpdesk**
  - **Networking events (User Forum, Conferences etc.)**
- **Interoperability**
  - **Expanding geographical reach and interoperability with collaborating e-infrastructures**
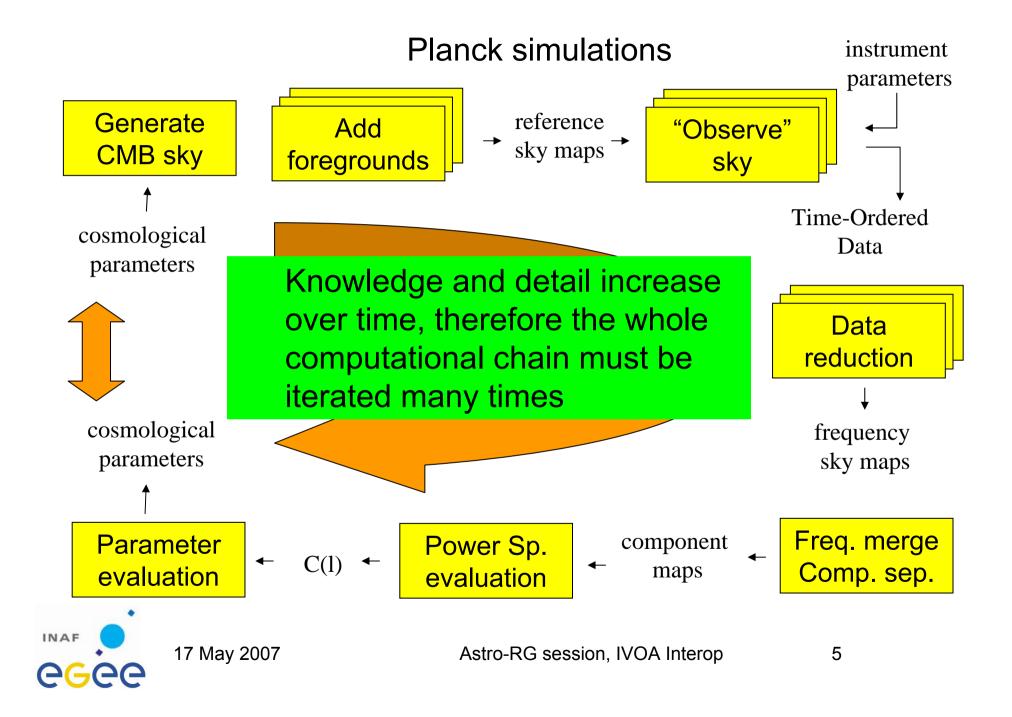
GEANT2

gLite
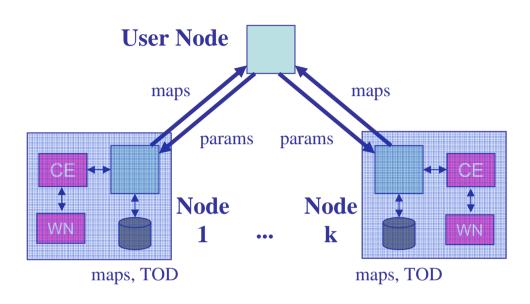Lightweight Middleware for Grid Computing

As EGEE partners

- EGEE is a project funded by the EU 6th Framework Programme (70M€ in 4 years)

- INAF is a partner within EGEE-II. It participates in:
  - NA3: training
  - NA4: porting of applications
  - Claudio Vuerli (OATs) is the reference person for the astro applications within NA4

- There is the wish to create an "astro community" in EGEE, building on the Planck VOrg experience
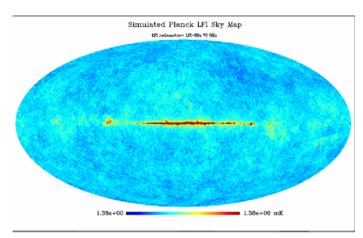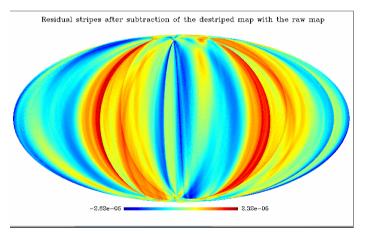
Planck simulations

instrument parameters

Generate CMB sky

Add foregrounds → reference sky maps → **"Observe" sky**

Time-Ordered Data

cosmological parameters

Knowledge and detail increase over time, therefore the whole computational chain must be iterated many times

Data reduction

frequency sky maps

cosmological parameters

Parameter evaluation ← C(l) ← **Power Sp. evaluation** ← component maps ← **Freq. merge Comp. sep.**

INAF

egee

Planck simulations in EGEE

User Node

maps   maps

params   params

CE   CE

WN   WN

Node 1   ...   Node k

maps, TOD   maps, TOD

de-striped "observed" sky

de-striping residuals map

- Simulation of the whole mission at the maximum LFI resolution (TOD ~ 1.3 TB)

- Once parameters have been set, each node builds an ideal sky, observes it in a LFI channel and produces a TOD, then rebuilds the map through de-striping techniques
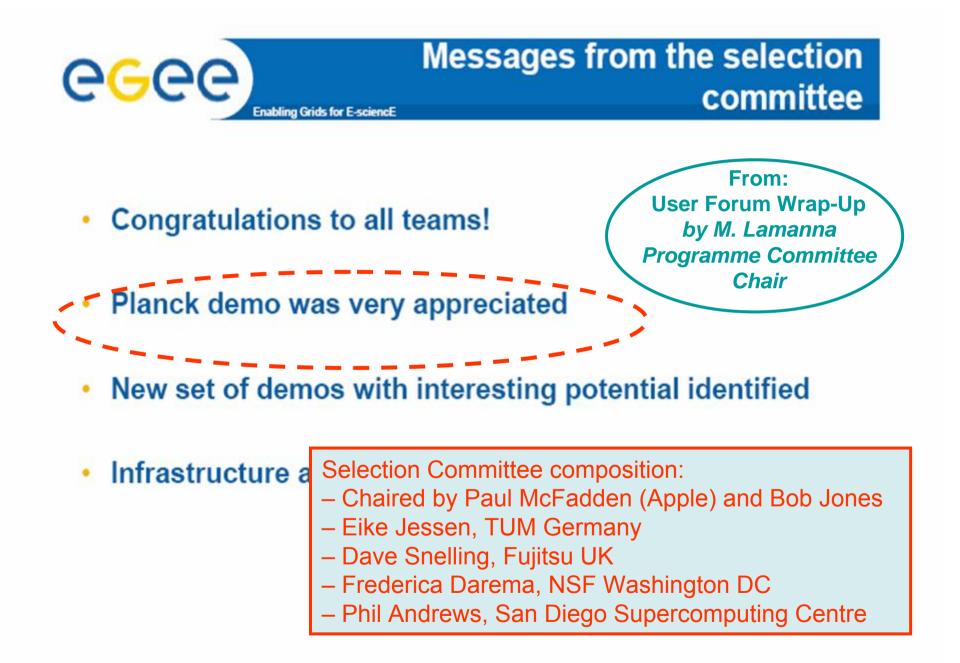
Planck simulations in EGEE

- Demo at Manchester EGEE User Forum:
  - How to submit a Planck simulation to the Grid using the Genius portal
    - How to prepare the Grid environment
    - How to set up some basic parameters
    - How to further refine the simulation by tuning cosmological and instrumental parameters
  - How to inspect the simulation status
  - How to retrieve results
  - How to visualize the output products of my simulation

- Explicitly mentioned by Selection Committee in EGEE User Forum final remarks

- **Congratulations to all teams!**

- **Planck demo was very appreciated**

- **New set of demos with interesting potential identified**

- **Infrastructure a**

From:
User Forum Wrap-Up
by M. Lamanna
Programme Committee
Chair

Selection Committee composition:
– Chaired by Paul McFadden (Apple) and Bob Jones
– Eike Jessen, TUM Germany
– Dave Snelling, Fujitsu UK
– Frederica Darema, NSF Washington DC
– Phil Andrews, San Diego Supercomputing Centre

Other applications using EGEE m/w

- High-resolution N-Body simulations (coupled with evolution of dark matter satellites orbiting inside more massive haloes using semi-analytical tools)

- From the above, generation of catalogues of cosmological dark matter halos (with known mass, position, velocity, merger history and angular momentum)

- Implementation of cfitsio driver over gridftp, and driver to access the EGEE filesystem

- Data processing applications
  - OmegaCam
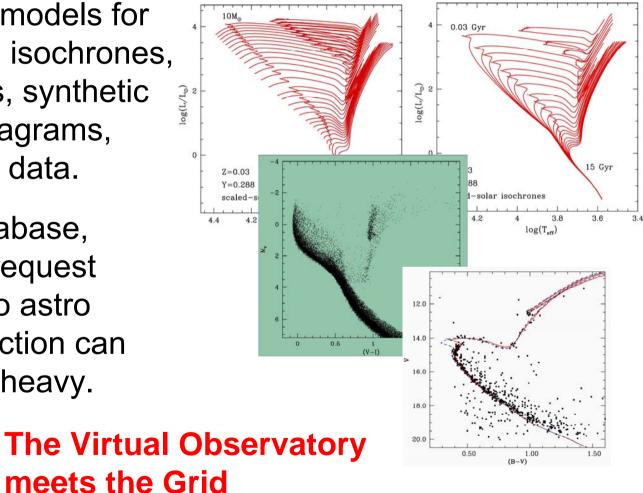 - test setup between ESAC and INAF-OATs

- BaSTI

Data on demand

BaSTI – numerical models for evolutionary tracks, isochrones, luminosity functions, synthetic color–magnitude diagrams, tables with relevant data.

BaSTI is also a database, which provides on request numerical models to astro users. Model production can be computationally heavy.

The Virtual Observatory meets the Grid

VO-DCA WP5

- VO Data Centre Alliance is a Euro-VO project (EU/FP6)
- WP5: Coordination with computational grid projects (INAF is WP leader). Goals are:
  - Achieve coordination between the VObs and the computational grid communities. Achieve the possibility for VObs users to exploit (through the Data Centres) the processing capabilities offered by computational Grids either nationally, or at the European level. Maximum priority given to coordination with EGEE.
  - Definition of standards in common between VO and Grid
  - Implementation in involved Data Centres
  - Knowledge acquisition; Coordination; Suggestions and guidelines; Dissemination
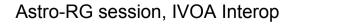- Interoperability with Data Centres is of maximum interest to EGEE

Extensions to EGEE: G-DSE

- The Grid limit: it is able to execute binary code or shell scripts and stores files
- Grid Data Source Engine, to cope with the following
  - I have a DB and I want to use it from my GRID.
  - I have a number of DBs and I want to use all of them.
  - Move the execution to the data and not data to the code.
- DB in the Grid? Extension of the existing Resource Manager of Globus for providing transparent access to heterogeneous DS and DSE
- Fully compliant with gLite
- Deal natively with asincronicity of DB queries!

G-DSE → QE
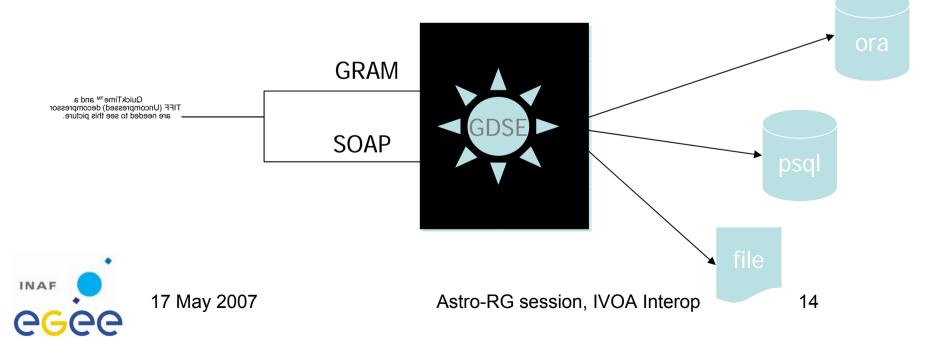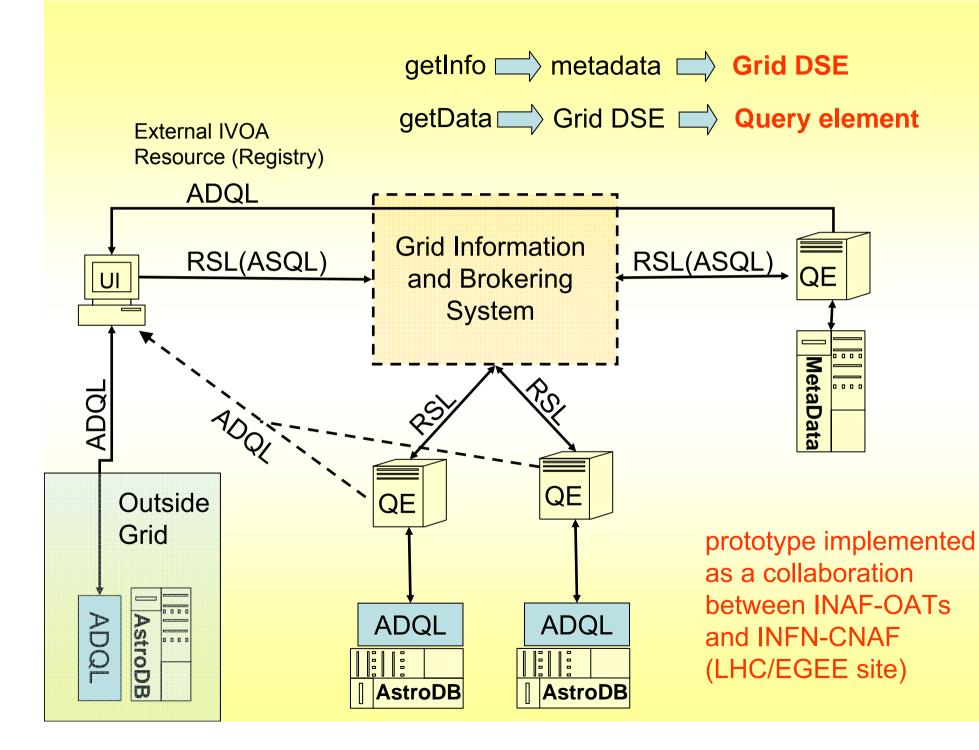
- The Grid Resource Framework Layer, Information System and Data Model are extended so that a software virtual machine as a Data Source Engine becomes a valid instance for a Grid computing model.

- A new Grid component (G- DSE) that enables the access to a Data Source Engine and Data Source, totally integrated with the Grid Monitoring and Discovery System and Resource Broker is defined

- A new Grid Element, the Query Element, can be built on top of the G-DSE component.

QE implementation

- Runs on any linux/unix flavor: GT>=2.4.3
- Backends: any DB vendor (MySQL, Oracle, PostgreSQL etc…) + flat files
- Two protocols: GRAM or WS
- API: C, C++, python, Java, perl
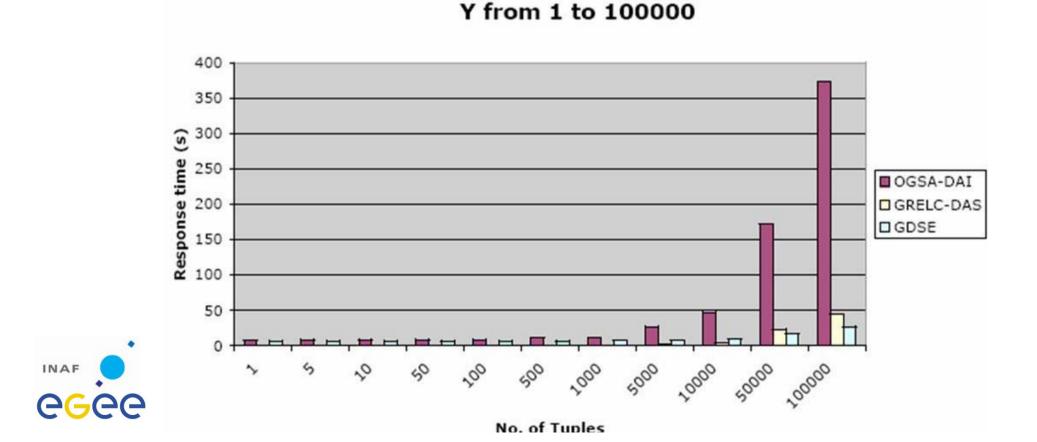- If it works with Globus it works with G-DSE

GRAM

SOAP

GDSE

ora

psql

file

getInfo ⟹ metadata ⟹ **Grid DSE**

getData ⟹ Grid DSE ⟹ **Query element**

External IVOA
Resource (Registry)

ADQL

RSL(ASQL)

Grid Information
and Brokering
System

RSL(ASQL)

UI

QE

MetaData

ADQL

ADQL

Outside
Grid

ADQL

AstroDB

RSL

RSL

QE

QE

ADQL

AstroDB

ADQL

AstroDB

prototype implemented
as a collaboration
between INAF-OATs
and INFN-CNAF
(LHC/EGEE site)
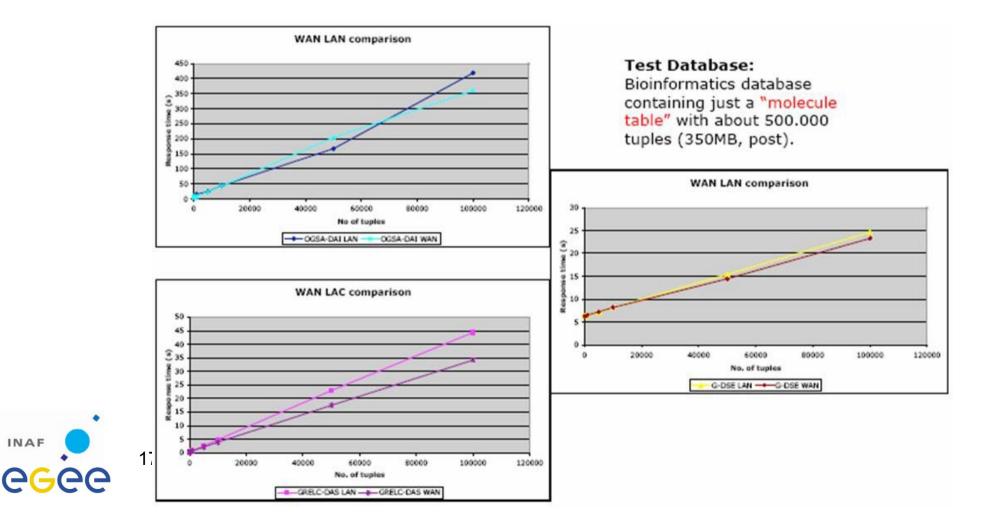
Benchmarks

- Test carried out between OGSA-DAI (full Globus), GRelC (WS) and G-DSE

- Preliminary results in a multi-disciplinary environment (shown at EGEE User Forum).

"select * from molecule where id < Y"
Y from 1 to 100000

Benchmarks

- Test carried out between OGSA-DAI (full Globus), GRelC (WS) and G-DSE

- Preliminary results in a multi-disciplinary environment (shown at EGEE User Forum).

Test Database:
Bioinformatics database
containing just a "molecule
table" with about 500.000
tuples (350MB, post).

Recycling work on G-DSE to enable Grid-enabled Instrument Control (G-ICS)

- IN Globus, and not ON Globus
- Provide a suitable formal definition of a Grid Abstract Computing Machine using Formal specification language
- Review the architectural definition of a Data Source Engine (DSE), to build an Instrument Control System
- Provide ICS integrated in the existing Grid Resource Layer Framework (G-ICS)
- Extend GIS capabilities to understand and handle monitoring and control requests to ICS
- Extend GIS capabilities to describe ICS Resources
- Provide a new Manager for ICS integrated in Globus
- Extend semantic and provide it to RSL to use G-ICS
- Etc…