Variability detection, period search with GaiaGrid

S. Ansari, L. Eyer, D.W. Evans, M. ter Linden
VSWG – Genève

6th July 2005

Objective

Distribution

Simulating the data, making it available, then processing it.

Computation

Processing a representative number of stars with the computational power at hand 10 million stars

Collaboration

Access a common environment to examine the data and reprocess them if necessary

The Simulated Data

(D.W. Evans)

G-band fluxes for 10 million stars

Based in FORTRAN 95 code to generate epoch distributions provided by F. Mignard

Stars were selected uniformly across the sky

Three types were defined

a) Constant Stars
b) Pulsating Variables
c) Eclipsing Binaries

Current Infrastructure Status

7 Infrastructures in 5 countries (voluntary) [38 CPUs]:

- ESTEC [13 CPUs] (SCI-CI) + 1 Gigabit dedicated link to Surfnet
- ESAC [6 CPUs] (SCI-SD) + 8 Mb link to REDIRIS
- ESRIN [12 CPUs] (EOP) + 155 Mb link to GARR
- CESCA [4 CPUs] (Barcelona) + REDIRIS connectivity
- ARI [2 CPUs] (Heidelberg) + Academic backbone
- ULB [1 CPU] (Brussels) + Academic backbone
- DutchSpace [1 CPU] (Leiden) + Commercial link
- Univ of Geneve [1 CPU] (Geneva) + Academic Backbone
- IoA [1 CPU] (Cambridge) + Academic Backbone

2 Data Storage Elements:

- CESCA [5 Terabytes]
- ESTEC [2 Terabytes]
- IoA [0.5 Terabytes]
- ESAC [upto 4 Terabytes)

There is a single point entry to all these resources: esagrid11 controller installed at ESTEC

Results

The scientific results will be discussed by Laurent Eyer in a separate presentation

All the source code may be accessed through GATT at:

gaia.esa.int/algorithms/

Grid Preparation

The concept of doing the Variable Star analysis in a Grid environment is to cut the input data into small sections, distribute the sections to the Grid nodes, run the VarStar application on the Grid nodes and then copy the results back to the central node.

We defined three batches of simulated files:

Batch 1: 200,000 stars 1 file of 1 Gbyte

Batch 2: 500,000 stars 5 files of 0.5 Gbytes each 2.5 Gbytes

Batch 3: 10,000,000 stars 100 files of 0.5 Gbytes each 50 Gbytes

A small script was written to chop the files into small quantities in order to better distribute them amongst the 38 nodes available.

The code was installed on all Grid nodes.

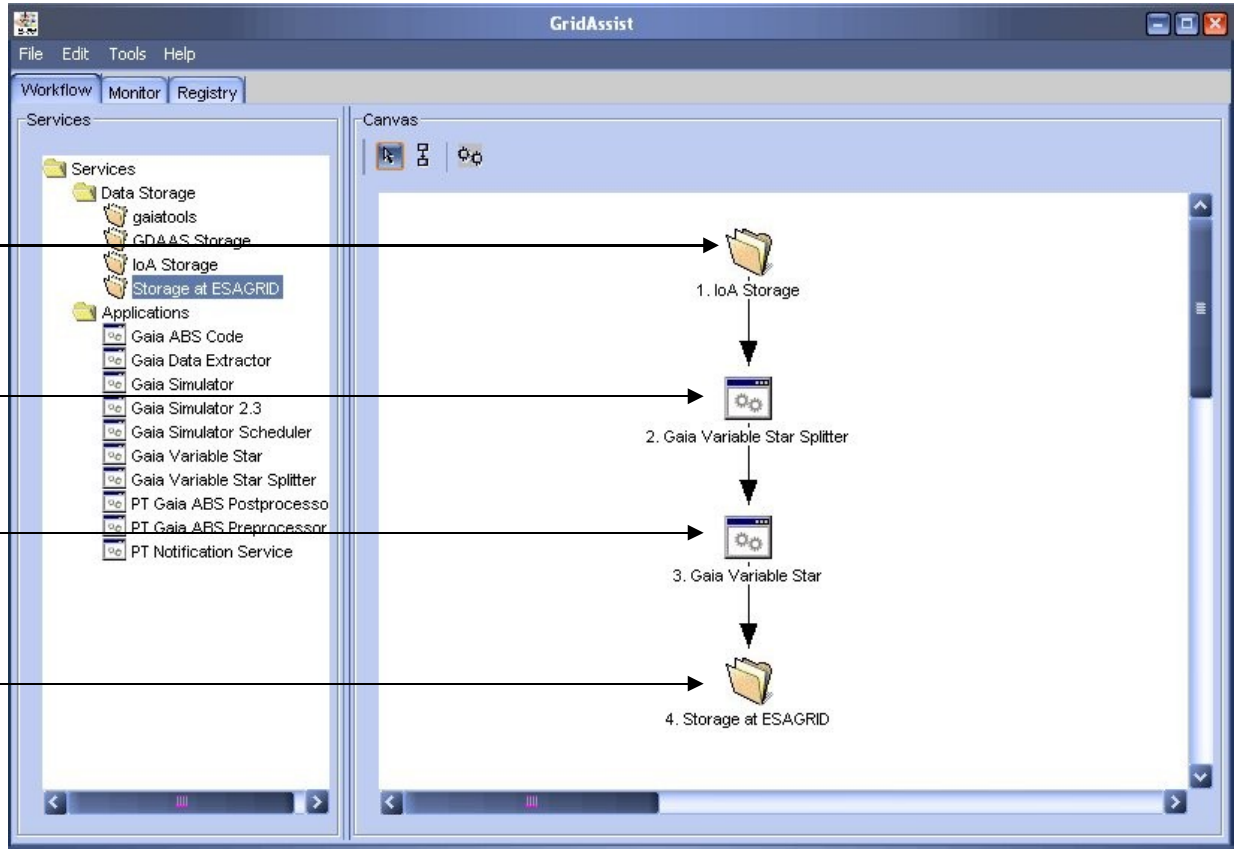
Workflow

Data is stored and retrieved from IoA Storage

The splitter script extracts a section of the data

The data is processed by the VarStar algorithm distributed amongst all GaiaGrid nodes

The results are stored in the ESAGRID Storage at ESTEC



Summary of Results

The processing time required for 100,000 stars is of the order of 12 hours

Some bottlenecks were encountered in the overruns of some jobs due to the expirey of certificates (currently defined at 48 hours)

The output of the analysis is minimal and is equal to a maximum of 30 MB/100,000 stars.

Conclusions

The use of flat files, in contrast to using a database management system, has its merits in that access and distribution is simplified to writing a script that will actually divide the large simulation file into smaller sections that are scalable, depending on the number of nodes to be used.

The CPU times are representative for an “ideal and simple” case. Based on the current estimates of analysing 100,000 and 1 million stars, the analysis of 10 million stars is of the order of 1.5 months (linear extrapolation).

Considering that the variability analysis is a task that will run possibly every six months, depending on the availability of core data, and that 1 billion stars will need to be analysed, and in addition that we can estimate a certain increase in computing power (based on Moore’s Law), the analysis of 1 billion stars (current simplified simulations, no noise, ideal case) may take up to 2 months of Grid power already.

Next Steps

Improve on the Grid environment

Analyse more complex simulations

Introduce classification schemes for Variable Stars (Univ. of Bonn work?)

Examine other potential programming languages?

This talk is summarised in a paper accesible via Gaia Livelink: GAIA-GRID-001

The Gaia Toolbox on GaiaGrid

Francesca de Angeli has started adding routines to the Gaia Toolbox that can be accessed from within Gaia Grid

If you have any suggestions of what you would like to see there, please send an email to Francesca:

fda@ast.cam.ac.uk