Integrating Practical: Session 41 – 46 9:00 to 18:00

Title: Search for Knowledge
Practical Leader: David Fergusson
Authors: teams from NeSC, Condor, GT4, OMII and GILDA.

Introduction

The final practical of the summer school is designed to give you the chance to use some of the technologies presented at the school in a more realistic way – demonstrating the efficacy of grid technology for solving scientific problems. The exercise will involve you working together in teams and draws on the (optional) advanced practicals offered by each technology.

The advanced practicals involve sampling and visualising a surface in order to identify features on the surface. This process is a simplified version of the common scientific process of identifying features in a sample of data which is contaminated (or even dominated) by noise. For example, data is obtained during an experiment in which signals are generated from many sources, not just the one you are interested in. The data must be searched to look for the signals which have the characteristics you are interested in.

In the integrating practical you will search a surface for 20 "pillars of wisdom". The pillars have rectangular cross-sections and flat tops. These pillars are small ($*10^{-8}$) compared with the area of the search space and their axes are aligned with those of the search space. On the top surface of each pillar is a rectangular raised plaque, on which text is either embossed or etched. If you obtain the text from all 20 pillars you have obtained all of the information, you will need to use all of the technologies to do this. If you interpret the information correctly you will have obtained the "knowledge", analogous to recognising the natural mechanism that governed the form of the signal. Off the pillars the surface is defined by a background function which simulates noise. (In a more realistic emulation of scientific data processing, the noise would also overlay the pillars.)

There is a choice of which technologies to use to perform the tasks during the practical. While the same "search for knowledge" is performed using each technology the details are different in each case and can be categorised in terms of the space searched and demonstrating a:

- *Computationally intensive task*: the computational expense of generating the surface samples is artificially increased. This means that generating and analysing very large samples of data in one go will be very inefficient. The available technologies, which use this scenario are GT4, Condor, & OMII GridSAM.
- *Data intensive task*: the generation of the surface samples is separated from the task of analysing the samples. The surface samples are first generated and stored, along with information on the samples – for example the area sampled and the step size between samples. The computational expense of generating the surface

samples is not artificially increased. gLite and OMII OGSA-DAI are the available technology for this scenario.
- You may also choose to generate and store surface samples. For half the surface area ($y < 0$) the data is only available in stored files which contain previously generated samples.

NB. Not all of the data space is accessible. Concentrate on finding pillars rather than covering the data space.

For the region only available as stored data, there is metadata available to find files and then "experimental data" in the files. The metadata will include: minimum_x, minimum_y, maximum_x, maximum_y, number_of_sample_points and file_name. The details as to how metadata and data are stored will vary for each technology.

The exact details of the practical and software provided for each technology are given separately.

Hints of the locations of pillars are given in terms of results from two previous research teams. These hints are contained within two databases which can be accessed using an OGSA-DAI client. It is generally believed that between them these researchers came close to locating all 20 pillars but they possibly made erroneous observations. Their technology was too poor to resolve information on the surface of the pillars and was liable to generate false positives. The hints are incomplete and may not be accurate.

As searches of the surface are performed, each team will be able to give details of their success so far. Each team can submit information into a database via OGSA-DAI on the area searched so far and pillars located. This information will be displayed via the web throughout the practical.

Effective team work will be vital to your success in this practical. Your team will have to decide on

- which technology (or technologies) to use,
- how to organise and perform the search.

A short introductory talk will be given at the beginning of the afternoon sessions (Wednesday 18th July) . Time is short, you will only have 1 day to complete the practical and discover "the knowledge". An opportunity to provide feed back on the success of your team will be provided at the end of the practical.