

Distributing Planck Simulations on a Grid Structure

Riccardo Smareglia, Fabio Pasian, Claudio Vuerli, Andrea Zacchei

INAF - Osservatorio Astronomico di Trieste, Trieste, Italy

Abstract. The production of Planck simulated data belongs to a class of problems which could get a great deal of advantage from their distribution on to a Grid-enabled environment, and has been identified by the Planck community as an ideal application to evaluate the power of the European Grid infrastructure. If the project is successful, its natural extension goes in the direction of studying the feasibility of porting large sections of the Planck data processing on the Grid.

1. Introduction - The Planck Mission

Planck is the third Medium-Sized Mission (M3) of ESA's Horizon 2000 Scientific Programme. The mission is designed to image the anisotropies of the Cosmic Background Radiation Field over the whole sky, with unprecedented sensitivity and angular resolution, and it will provide a major source of information relevant to several cosmological and astrophysical issues, such as testing theories of the early universe and the origin of cosmic structure. Planck is scheduled to be launched in February 2007, and two complete sky surveys are foreseen. The scientific development of the mission is directed by the Planck Science Team

Planck is composed of two instruments: the High Frequency Instrument (HFI) and the Low Frequency Instrument (LFI), operated by two dedicated Consortia through their Data Processing Centres (DPCs). The DPCs have in charge to integrate/run the simulation software contributed by HFI and LFI scientists, process the simulation results (the output of Planck receivers), build/test a number of pipelines to be used during operations to process both technical (House-Keeping) and scientific data. At the end of the mission DPCs will deliver the following scientific products:

- Calibrated time series data, for each receiver, after removal of systematic features and attitude reconstruction
- Photometrically and astrometrically calibrated maps of the sky in the observed bands (frequency maps)
- Sky maps of the main astrophysical components
- Catalogs of sources detected in the sky maps of the main astrophysical components
- Power spectrum of the CMB.

2. Planck simulations and the Grid

Some development steps are still underway for the Planck mission, one of which relates to the simulation of the mission outputs, necessary to forecast the behavior of satellite and instruments and to prepare the processing tools needed to prepare the scientific outputs of the mission..

In particular, there is the need to extend the simulation applications developed up to now to cope with new observational constraints. Sharing of simulation software, both existing and being developed, and of simulated data and computing resources to obtain them, are important as well for the success of a complex mission such as Planck.

One of the main challenges is the ability to design of a system, integrated while distributed across different sites, capable of generating at the production level different simulated skies depending on different instrumental configurations or behaviors. Planck simulations and processing are therefore an ideal application to evaluate the power of the European Grid Infrastructure.

2.1. A Possible Scenario

A possible scenario can be sketched for a Grid-enabled simulation environment for Planck: the Planck user requests to download, through a user interface, a specific set of all-sky simulated data under certain scientific hypotheses, and using a selected mission and instrument configuration; the environment understands if such a simulation has been already produced and, if so, it allows the user to access the data; if no data are available, then suitable computing facilities should be selected from a pool of available resources to produce the data the user will eventually be able to access; data could be processed locally or, if needed, in a distributed way throughout the Grid once again by selecting the computing facilities from those available on the Grid infrastructure.

2.2. The Grid-enabled application

The Grid-enabled application is based on a suite of already-existing applications developed by a set of institutes collaborating in the framework of the Planck ESA mission. Such applications, running on separate local computing facilities, simulate the microwave sky, produce time series of Planck data, and process the simulated data. The exercise would be generating a Gridified Simulation Pipeline whose components are the already-existing applications after they have been successfully ported on the Grid infrastructure.

The tasks provided by the application (the Gridified Pipeline) is planned to be: the generation of a microwave sky by tuning a set of input parameters (e.g. noise and systematic effects introduction) the extraction of time series simulating Planck “observations” of the generated sky

The key purpose is to make possible a profitable usage of the Grid infrastructure by the Planck collaboration to run the simulations code, and, if possible, the whole data processing for the mission during operations. The main objectives of the application are:

1. To prove the feasibility of the interfacing and porting of the simulation applications on a Grid environment
2. To prove the feasibility of new applications interfaced with the Grid

3. To prove the feasibility of a system allowing management of the existing simulated data and production of new ones
4. To prove the feasibility of porting the whole Planck data processing structure, or a fraction thereof, onto a Grid environment
5. Training and dissemination activities to create grid competence and awareness in different groups of the Planck community

2.3. Added Value

There are a number of advantages in using Grid-enabled software to run Planck simulations. Simulations code and produced simulated data are transparently and easily accessible to the Planck community through the Grid UI. They can ask for specific simulated data and, in case, run the application (the Simulations Pipeline) to produce them. Pipeline runs may require considerable computing power capabilities: by using the Grid, pipeline runs will be disseminated over the continental Grid infrastructure so that computing power shortages of single institutes can be easily overcome. Moreover, computing resources exploitation is optimized. Simulated data may be remarkable in size (e.g. frequency and component maps, especially time series may be huge). Simulations results may be transparently spread over different Storage Elements (SEs) and, from them, retrieved by the gridified Pipeline. Because of their intrinsic parallelism, simulations applications should gain great advantage when run over the Grid infrastructure.

2.4. Problems/Requirements

There are some limitations in running Planck applications on a Grid-enabled environment. Access to simulations software and produced simulated data of the HFI and LFI Consortia is not free. The Planck-specific Integrated Data and Information System (IDIS) infrastructure provides a federation layer having in charge the access control to the Planck information system; only authorized users can access the information system. Each IDIS user has a user profile defining his/her privileges. Therefore, it is necessary that the IDIS federation layer is integrated with the Grid user certification and authorization mechanism.

Moreover, a web-based Grid User Interface (UI) could be necessary/desirable to make the simulated data retrieval and the simulated pipeline job submission easy. It shall be possible to interface several DBMS via the Grid-UI.

It is finally to be noted that, at present, simulated data are stored in FITS format, however in the near future the original data will be stored on a commercial database management system (Versant is the current baseline). This may imply some licensing issues.

3. Conclusions and Further Tasks

The production of Planck simulated data has been identified as an ideal application to evaluate the power of the European Grid infrastructure in solving this class of problems. It will therefore be proposed to the EGEE project as one of the applications of the astrophysical community to be used as a test-bed for the EGEE infrastructure. This is an astronomy application that is more in the

spirit of the computational grid, rather than the data grid, which fits more with the Virtual Observatory concepts.

A natural extension of the project goes in the direction of global processing. A further step is thus foreseen to be studying the feasibility of porting the whole Planck data processing infrastructure on to a Grid environment. In this case, the construction of photometrically and astrometrically calibrated frequency maps of the sky in the observed bands, the construction of sky maps of the main astrophysical components and the population of a catalog of sources detected in the sky maps of the main astrophysical components could be applications to be Grid-enabled in the near future.

Acknowledgments. The authors wish to thank a number of colleagues with whom the idea of distributing Planck simulations on a Grid environment was discussed. Among these, Anthony Banday, Matthias Bartelmann, Leopoldo Benacchio, François Bouchet, Kari Enqvist, Andrew Jaffe, Bob Mann, Enrique Martinez-Gonzalez, Rafael Rebolo, Jean-François Sygnet. The authors are members of the Planck/LFI Consortium, led by Reno Mandolesi.

References

- Bennett, K., Pasian, F., Sygnet, J.F., Banday, A.J., Bartelmann, M., Gispert, R., Hazell, A., O'Mullane, W., Vuerli C. 2000, in: *Advanced Global Communications Technologies for Astronomy*, SPIE proceedings 4011, p. 2-10
- Lama, N., Vuerli, C., Smareglia, R., Gasparo, F., Pasian, F., Genghini, M. 2004, this volume, 400
- Pasian, F., Smareglia, R., Vuerli, C., Zacchei, A., Lama, N., Benacchio, L. 2003, *Mem. S.A.It. Supplements*, in press
- Pasian, F., Sygnet, J.F. 2002, in: *Data Analysis II*, SPIE proceedings 4847, p.158-169 p. 25-34