

Chapter 98

Gridifying Neuroscientific Pipelines: A SOA Recipe and Experience from the neuGRID Project

David Maset
maatG, France

The neuGRID Consortium

ABSTRACT

In recent times, innovative new e-Infrastructures have materialized all around the globe to address the compelling and unavoidably increasing demand on computing power and storage capacity. All fields of science have entered an era of digital explosion and thus need to face it with appropriate and scalable instruments. Amongst century's cutting-edge technologies, the grid has become a tangible candidate which several initiatives have harnessed and demonstrated the added value of. Turning the concept into a concrete solution for Neurosciences, the neuGRID project aims to establish a grid-based e-Infrastructure providing neuroscientists with a powerful tool to address the challenge of developing and testing new markers of neurodegenerative diseases. In order to optimize the resulting grid and to deliver a user-friendly environment, neuGRID has engaged the process of migrating existing imaging and data mining toolkits to the grid, the so-called gridification, while developing a surrounding service oriented architecture of agnostic biomedical utilities. This chapter reports on a preliminary analysis of the issues faced in the gridification of neuroimaging pipelines and attempts to sketch an integration model able to cope with the several and heterogeneous applications used by neuroscientists.

DOI: 10.4018/978-1-4666-2455-9.ch098

INTRODUCTION

Over the last decade, innovative new Information and Communication Technologies (ICT) have materialized into concrete e-Infrastructures. In particular, the so-called grid (Foster, Kesselman & Tuecke 2001), born in High Energy Physics, has been massively applied to harness distributed computing resources and thus address the digital explosion faced in all fields of science. Grid computing is an exciting concept promising to revolutionise many services already offered by the Internet. This new paradigm aims to provide rapid computation, large scale data storage and flexible collaboration by syndicating the power of a large number of commodity computers. The grid was originally devised for use in computing demanding fields but unsurprisingly was adopted in a number of ambitious medical and healthcare applications.

As of today, several projects around the world have been and still are exploiting grids to support biomedical research. In the US, most notably the caBIG™ initiative (<http://cabig.cancer.gov/>) founded by the National Cancer Institute (NCI) in 2004 to speed up discoveries in cancer research. In Europe, as key initiatives of the Framework Programmes of the European Commission, e-Infrastructures such as MammoGrid (Amendolia, Estrella, Hassan, Hauer, Manset, McClatchey, et al, 2004), Health-e-Child (Health-e-Child, The EU FP6 Information Societies Technology Project, 2008), @neurIST (<http://www.aneurist.org>) and many others have also demonstrated their added value.

Coming as a third generation grid, the neuGRID project has been recently launched to establish an international grid infrastructure specialized in the field of Neurosciences. neuGRID (NeuGrid Project, 2008) aims to interconnect major clinical research centres in Europe, ultimately supplying neuroscientists with the most advanced ICT to defeat Alzheimer's disease and neurodegenerative pathologies in general. In neuGRID, the collection

and archiving of large amounts of imaging data is paired with grid-based computationally intensive analyses to develop and test new disease markers. Leveraging the grid concept and technology being developed by the Enabling the Grid for E-science (EGEE) project (glite, A lightweight middleware for grid computing, 2008), neuGRID is pioneering an advanced Service Oriented Architecture (SOA) of biomedical research utilities mediating between user applications, backend and other facilities, while empowering them with the grid.

The main objective of this paper is to report on early experiences in the formalisation of an appropriate gridification model to allow neuroscientists from neuGRID to seamlessly run complex, data and computing intensive pipelines of neuroimaging algorithms. To do so, major design goals and underlying concepts are described while the requirements of such pipelines are precisely analyzed.

RATIONALE

Approach to Design

The major goal that guided the present design specification process was to establish a coarse-grained view of the system from the gathered users' requirements. This exercise was useful to identify major software layers, inner constituents and corresponding interfaces, while helping in better splitting the work and responsibilities among collaborators.

Similarly to the Service Oriented Modelling and Architecture (SOMA) (Arzanjani, 2004) process, coworkers aimed at identifying features and gradually grouping them into logical layers for future implementation. Thus, a meet-in-the-middle approach was adopted which reconciled requirements expressed by end-users with the bottom-up grid deployments of existing IT assets. The result of this work is here presented using a Service Oriented Architecture (SOA) (MacKen-

21 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/gridifying-neuroscientific-pipelines/73547

Related Content

Privacy-Preserving Hybrid K-Means

Zhiqiang Gao, Yixiao Sun, Xiaolong Cui, Yutao Wang, Yanyu Duan and Xu An Wang (2018). *International Journal of Data Warehousing and Mining* (pp. 1-17).

www.irma-international.org/article/privacy-preserving-hybrid-k-means/202995

Heterogeneous Text and Numerical Data Mining with Possible Applications in Business and Financial Sectors

Farid Bourennani and Shahryar Rahnamayan (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications* (pp. 816-836).

www.irma-international.org/chapter/heterogeneous-text-numerical-data-mining/73472

A Workload Assignment Strategy for Efficient ROLAP Data Cube Computation in Distributed Systems

Ilhyun Suhand Yon Dohn Chung (2016). *International Journal of Data Warehousing and Mining* (pp. 51-71).

www.irma-international.org/article/a-workload-assignment-strategy-for-efficient-rolap-data-cube-computation-in-distributed-systems/168486

Crime Analyses Using Data Analytics

Thanu Dayara, Fadi Thabtah, Hussein Abdel-Jaber and Susan Zeidan (2022). *International Journal of Data Warehousing and Mining* (pp. 1-15).

www.irma-international.org/article/crime-analyses-using-data-analytics/299014

A Survey of Extract–Transform–Load Technology

Panos Vassiliadis (2009). *International Journal of Data Warehousing and Mining* (pp. 1-27).

www.irma-international.org/article/survey-extract-transform-load-technology/3894