









Soizic Laguitton^{1,2} Denis Rivière^{1,3} Thomas Vincent¹ Clara Fisher^{1,2} Dominique Geffroy³ Isabelle Denghien³ Nicolas Souedet⁴ Yann Cointepas^{1,3}

A unified and simple interface to parallel computing resources

¹CEA, I2BM, Neurospin, Gif-sur-Yvette, France | ²CATI, France | ³IFR 49, Gif-sur-Yvette, France | ⁴CEA, I2BM, MIRCEN, Fontenay-aux-Roses, France

http://brainvisa.info/soma-workflow

E

S

0

- Parallel computing resources are...
 - → Highly available
 - \rightarrow Valuable assets for researchers to accelerate, test, validate and compare their methods

workflow

• In many cases the parallelization is **coarse-grained**: Many long sub-processes No communication between running sub-processes

- Soma-workflow deals with execution workflows (low level)
- Different from high level workflow software, which have:
- Higher level description of workflows

- **However**, parallel computing resources are... Not user friendly for non expert users
 - → Heterogeneous
 - (generally) lacking tools dedicated to coarse-grained parallelization

Overview

 Coarse-grained parallelized processes: →Workflows



Each job starts as soon as possible considering the execution dependencies

- Higher level features
- Soma-workflow bridges the gap between computing resources and...
- Non expert users (extensive documentation, Python API made to be simple, GUI...) High level workflow software



High leve

orkflow software



- Submission, control and monitoring of workflows on different resources: Python API
 - and/or
 - \rightarrow **GUI**: valuable for workflow monitoring
- Remote computing resource access: →Transparent remote access
 - \rightarrow File transfer and file path mapping tools
 - Client disconnections allowed



3

4 Python functions

Computing resources

puting resource functionalities

Maximize the compatibility

→Creating a new interface

= implementation of

Job status

- Soma-workflow has built-in interfaces: \rightarrow Ready to use on multiple core machines Interface with the Distributed Resource Management Application API^[1] (DRMAA)
- DRMAA
 - Standard of the Open Grid Forum Soma-workflow was used with success on clusters with the systems: Oracle grid engine, LSF, Torque and Condor

^[1] Tröger, P. (2007) Standardization of an API for Distributed Resource Management Systems, In proc of the 7th IEEE International Symposium on Cluster Computing and Grid.



Acceleration of a single data analysis

- Functional neuroimaging application: Joint detection estimation $(JDE)^{[2]}$
- Parcel-wise parallel analysis:



- Wrapping of existing code in a workflow
- \rightarrow A whole brain processing:
 - •10 hours on 1 CPU
 - ~ 15 mins on the cluster
- \rightarrow A group study of 20 subjects: \sim 1 day

^[2] Vincent, T (2010) Spatially adaptive mixture modeling for analysis of within-subject fMRI time series IEEE Trans. Med. Imag. (2010)

192 CPU cluster shared with a group of users

3 use cases



3 different steps in the research process

Robustness improvement with regression tests

- Morphologist pipeline of BrainVISA^[3] : extraction of the main brain structures from T1 MR images.
- Step by step morphologist was tested on about 1000 T1 MRI
- Classical regression test: 80 T1 MRI



- Workflows generated by BrainVISA
- → 23 hours on 1 CPU
- \rightarrow ~ 1h on the cluster
- → More frequent regression test

→ Test on larger database

^[3] Cointepas, Y. (2010) The BrainVISA project: a shared software development infrastructure for biomedical imaging research. Proceedings 16th HBM

Extensive validation and comparison

- Cross-validation of cortical sulci identification models^[4] and comparison with a newer method ^[5]
- Leave-one-out cross-validation of the models on the learning database:



→ 10000 individual jobs \rightarrow > 7 months on 1 CPU \rightarrow ~ 3 days on ~ 100 CPU

^[4] Rivière, D. (2002) Automatic recognition of cortical sulci of the Human Brain using a congregation of neural networks. Medical Image Analysis.

^[5] Perrot, M. (2011) Cortical sulci recognition and spatial normalization. In: Medical Image Analysis. In press