Harnessing cloud computing for high capacity analysis of neuroimaging data from NDAR

Submission Number:

8739

Submission Type:

Abstract Submission

Authors:

Daniel Clark[1], Christian Haselgrove[2], David Kennedy[2], Zhizhong Liu[3], Michael Milham[1], Petros Petrosyan[4], Carinna Torgerson[3], John Van Horn[3], Cameron Craddock[1]

Institutions:

[1]Child Mind Institute, New York, NY, [2]University of Massachusetts Medical School, Worcester, MA, [3]University of Southern California, Los Angeles, CA, [4]UCLA, Los Angeles, CA

Introduction:

The National Database for Autism Research (NDAR, http://ndar.nih.gov) and other NIH/NIMH data repositories are amassing and sharing thousands of neuroimaging datasets. With the availability of this deluge of data and the development of the NDAR infrastructure for its organization and storage, the bottleneck for applying discovery science to psychiatric neuroimaging has shifted to the computational challenges associated with data processing and analysis. Maximizing the potential of these data requires automated pipelines that can leverage high-performance computing (HPC) architectures to achieve high throughput computation without compromising on the quality of the results. A disadvantage of this approach is that it requires access to HPC systems that are not always available, particularly at smaller research institutions, or in developing countries. Cloud computing resources such as Amazon Web Services (AWS) Elastic Compute Cloud (EC2) offers a "pay as you go" model that might be an economical alternative to the large capital costs and maintenance burden of dedicated HPC infrastructures. Realizing this need, the developers of the Laboratory of Neuro Imaging (LONI) Pipeline, the Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC) Computational Environment (CE) and the Configurable Pipeline for the Analysis of Connectomes (C-PAC) have implemented pipelines in EC2 that interface with NDAR. Each pipeline was used to perform a benchmark analysis of 2,000 structural images from the NDAR database to establish the feasibility of this approach.

Methods:

Each of three pipelines were installed into Amazon Machine Images (AMIs) and customized to perform structural preprocessing on NDAR data. The LONI Pipeline (http://pipeline.loni.usc.edu) was enhanced to permit direct access to NDAR collections for workflow-based data processing (Torgerson, in press). Workflows can be created from a combination of commonly available neuroimaging processing tools represented as Pipeline Modules. With respect the benchmark analysis, specifically developed Pipeline Modules captured the results from FreeSurfer and FSL FirstAll, updated the NDAR with the results and returned them back to the NDAR Amazon Cloud storage server. C-PAC (http://fcp-indi.github.io) is a configurable pipeline for performing comprehensive functional connectivity analyses that was extended to include the Advanced Normalization Tools (ANTs) cortical thickness methodology (Tustison, 2014) and to interface it with NDAR (https://github.com/FCP-INDI/ndar-dev). Results of this workflow include 3D volumes of cortical thickness and regional measures derived from the Desikan-Killiany-Tourville atlas (http://mindboggle.info/faq/labels.html). NITRC-CE (http://www.nitrc.org/projects/nitrc_es/) is an AMI that is pre-installed with popular neuroimaging tools. A series of scripts were developed for NITRC-CE to interact with NDAR, calculate a series of quality assessment measures on the data, perform structural imaging analysis using FreeSurfer and FSL FirstAll results, and to write the results back to NDAR (https://github.com/chaselgrove/ndar).

Results:

Speeds obtained for processing structural data in EC2 were consistent with those obtained for local multi-core processors. For example, using an EC2 instance with 4 processors and 15 GB of RAM (m3.xlarge), the C-PAC pipeline was able to complete the ANTS cortical thickness pipeline in 8.5 hours per subject, in comparison to 9 hours on a local workstation with 12 processors and 64 GB of RAM. EC2 processing cost $1.94 per image for on demand instances and an estimated $0.26 per image when using spot instances.

Conclusions:

Analyzing data using cloud computing is an affordable solution, with low hardware and software maintenance burdens; this can be beneficial for smaller laboratories and when data is already in the cloud. Further reductions in cost can be obtained using lower costs spot instances, which fluctuate in price and may get shut down if demand gets too high.

Informatics:

Databasing and Data Sharing

Workflows [1]

Informatics Other [2]

Modeling and Analysis Methods:

Methods Development

Keywords:

Data analysis

Data Organization

Data Registration

Workflows

1|2 Indicates the priority used for review

Would you accept an oral presentation if your abstract is selected for an oral session?

Yes

Please indicate below if your study was a "resting state" or "task-activation" study.

Other

Healthy subjects only or patients (note that patient studies may also involve healthy subjects):

Patients

Internal Review Board (IRB) or Animal Use and Care Committee (AUCC) Approval. Please indicate approval below. Please note: Failure to have IRB or AUCC approval, if applicable will lead to automatic rejection of abstract.

Not applicable

Please indicate which methods were used in your research:

Structural MRI

For human MRI, what field strength scanner do you use?

3.0T

Which processing packages did you use for your study?

AFNI

FSL

Free Surfer

Other, Please list - ANTS

LONI Pipeline

Provide references in author date format

Torgerson, C.M., Quinn, C., Dinov, I.D., Liu, Z., Petrosayan, P., Kennedy, D.N., Toga, A.W., and Van Horn, J.D. (in press) "Interacting with the National Database for Autism Research (NDAR) via the LONI Pipeline Workflow Environment", Brain Imaging and Behavior.
Tustison, N.J., Cook, P.A., Klein, A., Song, G., Das, S.R., Duda, J.T., Kandel B.M., van Strien, N., Stone J.R., Gee, J.C., Avants, B.B. (2014) "Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements." Neuroimage. 2014 Oct 1;99:166-79. doi: 10.1016/j.neuroimage.2014.05.044. Epub 2014 May 29.