

# Heads in the Cloud: A Primer on Neuroimaging Applications of High Performance Computing

## Supplementary Issue: New Concepts in Magnetic Resonance as Applied to Cellular and In Vivo Applications

Anwar S. Shatil<sup>1,2,\*</sup>, Sohail Younas<sup>1,2,\*</sup>, Hossein Pourreza<sup>3</sup> and Chase R. Figley<sup>1,2,4,5,6</sup>

<sup>1</sup>Biomedical Engineering Graduate Program, University of Manitoba, Winnipeg, MB, Canada. <sup>2</sup>Neuroscience Research Program, Kleyesen Institute for Advanced Medicine, Winnipeg, MB, Canada. <sup>3</sup>Western Canada Research Grid (WestGrid) and Compute Canada, University of Manitoba, Winnipeg, MB, Canada. <sup>4</sup>Division of Diagnostic Imaging, Winnipeg Health Sciences Centre, Winnipeg, MB, Canada. <sup>5</sup>Department of Radiology, University of Manitoba, Winnipeg, MB, Canada. <sup>6</sup>Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD, USA. \*These authors contributed equally to this work.

**ABSTRACT:** With larger data sets and more sophisticated analyses, it is becoming increasingly common for neuroimaging researchers to push (or exceed) the limitations of standalone computer workstations. Nonetheless, although high-performance computing platforms such as clusters, grids and clouds are already in routine use by a small handful of neuroimaging researchers to increase their storage and/or computational power, the adoption of such resources by the broader neuroimaging community remains relatively uncommon. Therefore, the goal of the current manuscript is to: 1) inform prospective users about the similarities and differences between computing clusters, grids and clouds; 2) highlight their main advantages; 3) discuss when it may (and may not) be advisable to use them; 4) review some of their potential problems and barriers to access; and finally 5) give a few practical suggestions for how interested new users can start analyzing their neuroimaging data using cloud resources. Although the aim of cloud computing is to hide most of the complexity of the infrastructure management from end-users, we recognize that this can still be an intimidating area for cognitive neuroscientists, psychologists, neurologists, radiologists, and other neuroimaging researchers lacking a strong computational background. Therefore, with this in mind, we have aimed to provide a basic introduction to cloud computing in general (including some of the basic terminology, computer architectures, infrastructure and service models, etc.), a practical overview of the benefits and drawbacks, and a specific focus on how cloud resources can be used for various neuroimaging applications.

**KEYWORDS:** analysis, big data, cloud, cluster, grid, MRI, neuroimaging, supercomputer

**SUPPLEMENT:** New Concepts in Magnetic Resonance as Applied to Cellular and In Vivo Applications

**CITATION:** Shatil et al. Heads in the Cloud: A Primer on Neuroimaging Applications of High Performance Computing. *Magnetic Resonance Insights* 2015;8(S1) 69–80 doi:10.4137/MRI.S23558.

**TYPE:** Review

**RECEIVED:** September 21, 2015. **RESUBMITTED:** December 7, 2015. **ACCEPTED FOR PUBLICATION:** December 11, 2015.

**ACADEMIC EDITOR:** Sendhil Velan, Editor in Chief

**PEER REVIEW:** Four peer reviewers contributed to the peer review report. Reviewers' reports totaled 1125 words, excluding any confidential comments to the academic editor.

**FUNDING:** This work was supported by the University of Manitoba and the Winnipeg Health Sciences Centre Foundation. The authors confirm that the funder had no influence over the study design, content of the article, or selection of this journal.

**COMPETING INTERESTS:** Authors disclose no potential conflicts of interest.

**COPYRIGHT:** © the authors, publisher and licensee Libertas Academica Limited. This is an open-access article distributed under the terms of the Creative Commons CC-BY-NC 3.0 License.

**CORRESPONDENCE:** chase.figley@umanitoba.ca

Paper subject to independent expert blind peer review. All editorial decisions made by independent academic editor. Upon submission manuscript was subject to anti-plagiarism scanning. Prior to publication all authors have given signed confirmation of agreement to article publication and compliance with all applicable ethical and legal requirements, including the accuracy of author and contributor information, disclosure of competing interests and funding sources, compliance with ethical requirements relating to human and animal study participants, and compliance with any copyright requirements of third parties. This journal is a member of the Committee on Publication Ethics (COPE).

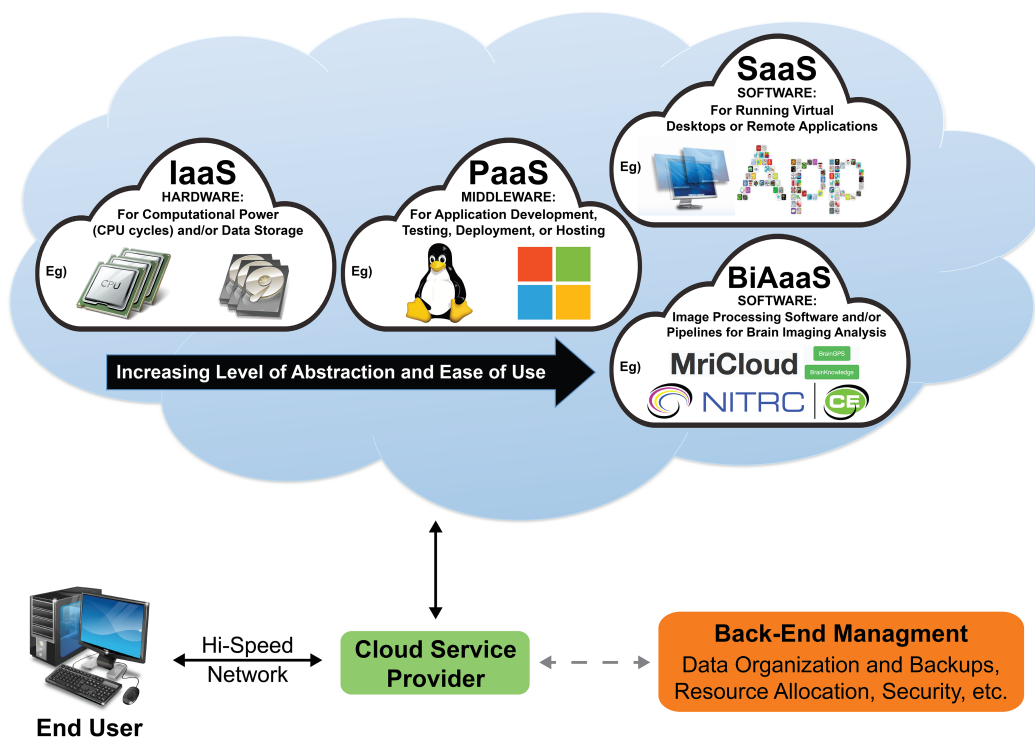
Published by Libertas Academica. Learn more about this journal.

## Introduction

Continued technological advances in magnetic resonance imaging (MRI) hardware that promote increased spatial and temporal resolution imaging, as well as the increasing sophistication of post-processing pipelines for advanced neuroimaging modalities—eg, functional MRI, diffusion imaging, myelin water imaging, etc.—continue to increase the computational resources required for the subsequent data analysis. Particularly in large-scale, multi-center studies such as the Human Connectome Project (<http://humanconnectome.org>)<sup>1–3</sup>, the sheer amount of imaging data (which is expected to be on the order of 100TB by project's end)—and therefore the computational resources needed to perform the corresponding analyses—are extremely large. Considering the ever-increasing need for storage and computing resources, a relatively new concept called cloud computing was introduced.<sup>4</sup> Cloud computing hides most of the details about the underlying computer

resources and their allocation, and offers end-users an illusion of unlimited resources through simplified interfaces (Fig. 1). In the past decade, several companies (eg, Microsoft, Google, Amazon, etc.) have started offering various cloud computing solutions, and an increasing number of government research institutes and research universities now have access to cluster and/or grid computing platforms.

Although cloud computing has obvious benefits for storing and analyzing extremely large data sets—eg, freely available data sets from the OpenfMRI database (<https://openfmri.org/>), NeuroVault database (<http://neurovault.org/>), Human Connectome Project (<http://humanconnectome.org>), 1000 Functional Connectomes Project ([http://fcon\\_1000.projects.nitrc.org/](http://fcon_1000.projects.nitrc.org/)), etc.—there are also other neuroimaging analysis techniques for which cloud-based computing may be more efficient, more expedient, and perhaps necessary, even for smaller datasets or single subjects. Therefore, some of these



**Figure 1.** Basic diagram of cloud- or grid-based computing. Each of the service models discussed in this primer—ie, IaaS (Infrastructure as a Service), PaaS (Platform as a Service), SaaS (Software as a Service), and BiAaaS (Brain Imaging Analysis as a Service)—are depicted, with examples, from the lowest level of abstraction (left) to the highest level of abstraction (right) within the cloud. It should be noted that as a result of this abstraction, PaaS includes the underlying IaaS, and both SaaS and BiAaaS (which is really just a specialized instance of SaaS) include the underlying PaaS and IaaS necessary to run the particular application or analysis.

more routine applications of cloud computing will also be addressed in this review, along with some examples and a practical discussion about how most neuroimaging researchers—even those with relatively limited technical proficiency—can access and leverage high-performance computing resources to improve their day-to-day workflow.

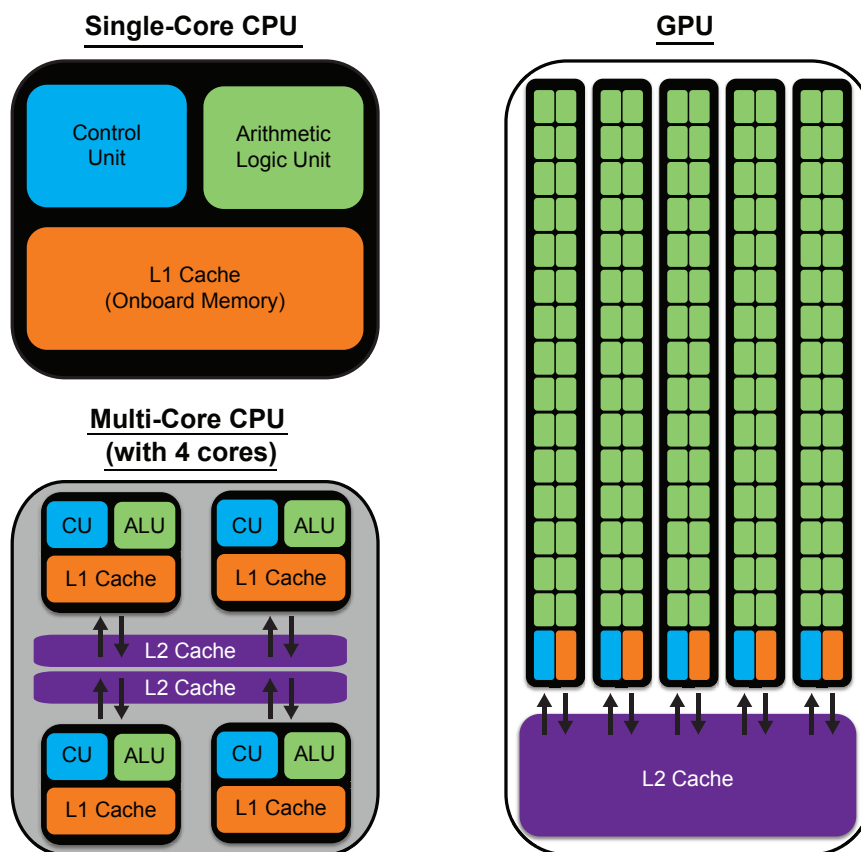
### CPU, GPU, Clusters and Clouds

Since this article is aimed primarily at neuroimaging end-users (eg, cognitive neuroscientists, psychologists, psychiatrists, neurologists, neuroradiologists, etc.) who may not have much in the way of formal computer science training, it is perhaps worth briefly reviewing some of the basic terminologies and principles regarding computer hardware.

In colloquial terms, the central processing unit (CPU) is the brain that serves as the computational hub of most computers. It is the primary circuit responsible for carrying out the basic arithmetic, logic, control and input/output (I/O) operations that are required to run most software (ie, computer programs). Although commonly thought of as such, CPUs are not technically single entities, but rather integrated circuits that comprise separate sub-components, including: 1) an arithmetic logic unit (ALU) that performs all of the arithmetic and logical operations, 2) several very fast but very small storage units (eg, a few kilobytes) called processor registers and level 1 (L1) cache that store outputs from the ALU, and 3) a control unit

(CU) that communicates with the memory and coordinates the operations of the ALU (Fig. 2); however, practically all modern CPUs are integrated circuits in which these components are combined on the same silicon chip.<sup>5</sup> Taking this concept even further, multi-core processors (eg, dual-core, quad-core, etc.)—where each CPU actually has multiple CUs and ALUs within the same circuit—are now becoming increasingly popular due to their enhanced parallel processing abilities and lower power consumption (and heat generation) per clock cycle compared to conventional (ie, single-core) CPUs.

While CPUs consist of either one or a relatively small number of very powerful processing units, and are optimized for performing large, serial (ie, sequential) operations, graphics processing units (GPUs) have massively parallel architectures that are composed of hundreds or thousands of less powerful cores. For these reasons, GPUs are much less expensive to produce and much more energy-efficient than CPUs on a per core basis, and therefore offer an efficient solution for running multiple—albeit smaller—tasks in parallel (ie, simultaneously). Although combining GPUs with CPUs was initially developed to handle ultra-high-definition graphics (hence their name), the application of GPU-accelerated computing is rapidly growing in popularity for other data intensive applications that can be run in parallel, including image processing, genomic analyses, and many other scientific/engineering applications.<sup>6</sup>



**Figure 2.** Simplified diagrams to highlight the main features of (and differences between) single-core CPUs, multi-core CPUs, and GPUs. Although each of the ALUs in single- and multi-core CPUs are more powerful than those found in contemporary GPUs, the large number of ALUs (ie, hundreds or thousands) make GPUs well-suited for processing large numbers of operations that can be run in parallel.

From an infrastructure standpoint, hybrid CPU-GPU (ie, combined CPU + GPU) workstations can either be standalone towers or rack-mounted systems, and can range in price from approximately \$1000 USD (for a relatively simple tower set-up) to tens of thousands of dollars for top-of-the-line systems with several multi-core CPUs, multiple GPUs, and relatively large amounts of random access memory (RAM).

In contrast to CPU and hybrid CPU-GPU standalone systems, there are also distributed high performance computing architectures that interconnect and therefore leverage the collective computational power of multiple machines (or nodes). For example, a *cluster* is a group of nodes (ie, CPUs or GPUs) that are connected by a high bandwidth and low latency local area network (LAN). While clusters can vary drastically in terms of their size and computational power (a cluster could technically be as small as two or three computers), the other common types of distributed systems, called *grids* and *clouds*, are typically larger in scale and are often geographically distributed. Overall, the concepts of grids and clouds are very similar, except for the fact that a grid is typically a collection of interconnected clusters that are each owned by different institutes or organizations—and are geographically distributed in multiple locations—to leverage the collective computational power of these shared

resources. One example of a well-known grid is SETI@home (<http://setiathome.ssl.berkeley.edu/>), which uses donated clock cycles from registered users' internet-connected computers to analyze radio telescope data in the search for extraterrestrial intelligence (SETI). On the other hand, a cloud is usually a collection of computers or clusters that are all owned by a single entity (eg, Amazon Web Service, IBM Cloud Computing, Google Cloud Platform, etc.). The primary advantage of clusters, grids and/or clouds is that they can offer levels of performance and processing speed that are not possible on standalone CPU or GPU workstations, particularly for computations or analyses that can be run in parallel. Because of these overarching similarities, the terms cloud, cluster, and grid will be used interchangeably for the remainder of this paper.

### Types and Availability of Cloud Computing Resources

Generally speaking, cloud resources may be offered in various service models, and of these service models, the three most common include:

- Infrastructure as a Service (IaaS) is the most basic service model (ie, the lowest level of cloud computing; Fig. 1),



- where the cloud service provider simply supplies computer hardware (eg, CPU and/or GPU clock cycles for computational processing, hard disk space for data storage, or other infrastructure components) to its users, and takes responsibility for handling the routine system maintenance, backups, etc. IaaS customers typically pay on a fee for service, pay-as-you-go model (eg, by the hour, week, month, or based on the number of clock cycles, the amount of hard disk space occupied, or the amount of data uploaded/downloaded from the cloud); however, it should be noted that members of many government research labs, private research institutes and academic institutions (ie, universities and colleges) may have access to free IaaS resources if their organization is affiliated with a supercomputing grid. Either way, however, the IaaS model eliminates the up-front expenses of purchasing in-house hardware and the ongoing expenses associated with system repairs, upgrades and administration. Moreover, since IaaS platforms offer highly scalable resources, they are exceptionally well suited for users with infrequent or intermittent high-performance computing needs. Examples of leading IaaS providers include Amazon Web Services, Google Compute Engine, Google Cloud Storage, DropBox, Microsoft Azure and IBM SmartCloud, etc. Also, it should be noted that most government and academic computer clusters and research grids offer IaaS for their members.
- b) Platform as a Service (PaaS) is an intermediate cloud service model, in which a cloud service provider delivers both the hardware and basic software tools (eg, an operating system and basic middleware such as Java, C++, etc.) necessary to create, develop and test new applications in a run-time environment. We mention PaaS here for completeness; however, because PaaS platforms are almost exclusively used by software developers, and are rarely (if ever) used by neuroimaging researchers, they will not be discussed further.
- c) Software as a Service (SaaS) is a software distribution model in which applications are hosted by service providers and made available to customers over a network, which is why SaaS products are sometimes also referred to as *hosted applications*. In this model, a vendor can either host commercially available software for customers and deliver it over the web, or give customers network-based access to a single copy of an application created specifically for SaaS distribution. In contrast to the traditional model of software distribution (ie, when software is purchased for and installed on standalone computers), the SaaS model removes the need for organizations to manage software installations and also offers several other benefits, including automatic updates, better compatibility for collaborative projects (ie, since users will all be using the same software

version), global accessibility, and reduced cost (ie, compared to buying permanent standalone versions). Again, the potentially reduced cost may be attractive, especially for software that is infrequently or only intermittently used over a fixed time period. Examples of SaaS models include Adobe Creative Cloud, Microsoft Office 365, Google Apps, Google Docs, Google Translate, Concur, Cisco WebEx, etc. It should be noted that basic versions of SaaS software are sometimes free (or free for a limited “trial period”), and that many government institutions and research universities have subscriptions or site licenses for at least some commercial SaaS applications. For this reason, it is advisable to check for basic, trial, and/or institutional licenses before purchasing a particular SaaS product. Brain Imaging Analysis as a Service (BiAaaS) is an emerging sub-class of SaaS that provides end-to-end solutions—including software, image processing pipelines, brain atlases and/or data sets, and also integrates the underlying platforms and infrastructure—for cloud-based neuroimaging analyses. Two recent examples of BiAaaS include MRICloud and NITRC-CE, which will be discussed in greater detail below.

Furthermore, each of these cloud service models can be deployed in publicly accessible clouds that are owned and operated by Cloud Service Providers to offer cloud resources (usually for a fee) or private clusters/grids/clouds that are operated by research institutions and are usually free for institutional members. Each of these infrastructure models will be briefly discussed below.

**Publicly accessible clouds.** Google and Amazon are two examples of publicly available cloud service providers whose resources can be utilized by anyone, including neuroimaging researchers, for a number of different applications. The Google Cloud, for example, provides everything from storage capabilities (eg, for data storage, static web hosting, etc.) via Google Cloud Storage (<https://cloud.google.com/storage/>) to scalable ultra high-performance computing infrastructure for data analysis via the Google Compute Engine (<https://cloud.google.com/compute/>), which bills by the minute with no upfront costs, so users do not pay for unused computing time. Similarly, the Amazon Web Services offers a full range of storage solutions via the Amazon Simple Storage Service (Amazon S3) (<https://aws.amazon.com/s3/>) to ultra-high performance computing via the Amazon Elastic Compute Cloud (Amazon EC2) (<https://aws.amazon.com/ec2/>). In both cases, large data sets can be stored and large-scale workloads can be run on the cloud provider’s infrastructure. A summary of cloud-based services provided by Google and Amazon is given in Table 1, and a partial list of available cloud solutions and tools is provided in Table 2.

**Private clouds.** In addition to public clouds, there are also private or semi-private grids that are accessible to users



**Table 1.** Service models of Google and Amazon cloud computing solutions.

SERVICE MODEL	GOOGLE	AMAZON
IaaS (Infrastructure as a Service)	Google Compute Engine (LP), Google Cloud SQL, Google Cloud Storage, etc.	DynamoDB, ElastiCache, RDS, etc.
SaaS (Software as a Service)	Google Docs, Google Prediction API, Google PageSpeed, Google Translate, Google BigQuery etc.	Cloud Search, SES, SNS etc.
PaaS (Platform as a Service)	Google App Engine	Appstream

from particular institutions and/or geographical regions. For example, Compute Canada (<https://www.computecanada.ca/>) offers freely available cloud computing and storage solutions to researchers affiliated with Canadian Universities. The infrastructure and support for these kind of systems are often provided at no charge to individual researchers, and therefore offer tremendous value. For this reason, researchers who are interested in getting involved in cloud computing for data storage and/or analysis purposes should: 1) find out whether their institutional affiliation entitles them to access any private research grids, 2) determine what, if any, costs are associated with using these resources (often they are free), 3) familiarize themselves with the capabilities of the available clusters and/or grids, and 4) determine whether

there is any local IT support that may be able to assist with initial setup, etc.

If institutional access is available, these free (or ultra low-cost) private clouds will be sufficient for most neuroimaging applications, and therefore an attractive option because of the low cost and abundant technical support. Using Compute Canada as an example, the available IaaS platform is completely free to institutional members, is scalable and capable of handling extremely large computing and data storage requirements, and is flexible enough to accommodate various research themes (eg, drug development, meteorological modeling, genomic analyses, advanced image processing, etc.). Additionally, the Compute Canada infrastructure platform is also combined with local and regional support services

**Table 2.** Partial list of cloud computing solutions and tools, their primary applications, and cost models (adapted in part from Tsafaris).<sup>59</sup>

SOLUTION NAME	DESCRIPTION	PRIMARY APPLICATION(S)	COST MODEL
Cycle Computing	Offering HPC clusters at an hourly rate based on Amazon Web Service (AWS) solutions	<ul style="list-style-type: none"> <li>• Big jobs capable of using thousands of CPUs</li> <li>• No tool to build your own cloud-based infrastructure</li> </ul>	<ul style="list-style-type: none"> <li>• Hourly billing</li> </ul>
StarCluster	An open source toolkit to automate building, configuring, and managing clusters of virtual machines on Amazon cloud	<ul style="list-style-type: none"> <li>• Starting scientific computation in a cloud</li> <li>• Users should install specialty software packages</li> </ul>	<ul style="list-style-type: none"> <li>• N/A</li> </ul>
OpenCPU	The OpenCPU server provides HTTP API to run R programs remotely	<ul style="list-style-type: none"> <li>• To run R programs on the cloud by installing the OpenCPU server</li> </ul>	<ul style="list-style-type: none"> <li>• N/A</li> </ul>
CPUsage	Provides an environment to install and configure applications. It then packages and deploys the local environment to the cloud	<ul style="list-style-type: none"> <li>• Starting scientific computation in a cloud</li> </ul>	<ul style="list-style-type: none"> <li>• Per minute billing</li> </ul>
Wakari	A Python-based data analytics tool hosted in the cloud. Each Wakari node is configured with extensive Python packages and it is browser accessible	<ul style="list-style-type: none"> <li>• To do Python-based data analysis</li> </ul>	<ul style="list-style-type: none"> <li>• Basic free and monthly</li> </ul>
CBRAIN	A data processing platform specifically designed for neuroimaging research	<ul style="list-style-type: none"> <li>• To do neuroimaging research</li> </ul>	<ul style="list-style-type: none"> <li>• Free for Compute Canada users</li> </ul>
Compute Canada	A cloud solution enabling users to install their own operating system and software	<ul style="list-style-type: none"> <li>• Start scientific computing in a cloud</li> </ul>	<ul style="list-style-type: none"> <li>• Free for Compute Canada users</li> </ul>
PiCloud	Acts as middleware between users and Amazon Web Service (AWS), with the goal of streamlining and simplifying scientific computing in the cloud for end-users	<ul style="list-style-type: none"> <li>• Run Python-based scientific computing in a cloud</li> </ul>	<ul style="list-style-type: none"> <li>• Sub-second billing for CPU time</li> <li>• Additional charges for data storage and download</li> </ul>



in order to help new users get started and to ensure that all users are making efficient use of the shared resources. Compute Canada's regional partners include ACENET in Atlantic Canada, Calcul Quebec, Compute Ontario, and WestGrid in Western Canada, each of which works together with their associated institutions in order to ensure that everyone has access to both the platform and local support.

Although Compute Canada and other large-scale research grids offer enormous computational power at low or no cost, one potential disadvantage is that jobs can get backlogged in a queue depending on the overall demand for computational resources. In order to minimize this issue, Compute Canada invites prospective users to write a proposal every year and identify the number of CPU hours and the amount of storage that they would like to access. Depending on the proposal and past history of resource utilization by the researcher, national computing resources are then pre-allocated. However, especially if these resources have not been pre-allocated, one alternative is to run certain jobs within smaller grids (eg, WestGrid) or on local clusters, where there may be a shorter queue and/or jobs can be given a higher priority within the sequence. This will of course depend on the size and complexity of the proposed analyses, but we have had great success in the past running our large batch processes on the Grex cluster at the University of Manitoba, rather than using higher levels of the cloud (ie, WestGrid or Compute Canada). The cluster—which, in our case, has 316 nodes and approximately 4,000 total cores—may have a shorter queue for “walk-on” CPU time, and still offers enough computational power for all but the largest image-processing and analysis jobs (eg, as long as the number of subjects and/or time-points in the data set does not exceed the number of nodes in the cluster, they could all still theoretically be processed in parallel).

### MRI-Specific Cloud Resources

In addition to the generic types of cloud service and infrastructure models discussed above, there are also a growing number of MRI-specific cloud computing resources that are available for neuroimaging researchers. Although it is beyond the scope of this review to provide an exhaustive list of these resources (not least of which because the list is continually evolving and would therefore be almost immediately out of date), we would like to point to a few specific examples.

Given the unique and computationally demanding nature of brain imaging research, and all of the specialized software necessary to analyze advanced brain imaging data, a new category of cloud service model—referred to as Brain Imaging Application as a Service (BiAaaS)—has recently been proposed,<sup>7</sup> which combines the IaaS and SaaS models, and is specifically tailored for brain imaging analyses. In addition to the DiffeoMap toolbox within MRISudio (discussed below within the context of performing high-dimensional, non-linear normalizations), another MRI-specific cloud computing resource that falls under the BiAaaS model is the

recently released MRICloud (<https://mricloud.org/>), which is a web-based neuroimaging analysis platform that was developed to provide fully-automated, online diffusion tensor imaging (DTI) analysis pipelines that integrate quality control measures<sup>8,9</sup> and advanced preprocessing methods such as automated brain parcellation using Multiple-Atlas Likelihood Fusion.<sup>10,11</sup> The advantages of this or any other BiAaaS model are three-fold: 1) the developers can deploy upgrades to their pipelines and/or algorithms in real-time, 2) all of the computationally-demanding analyses automatically use supercomputing resources to drastically reduce the processing time compared to having individual users run the software on standalone workstations, and 3) BiAaaS platforms seamlessly integrate the software and hardware, and are relatively easy to use (compared to more general cloud computing alternatives). However, the downside of BiAaaS platforms is that they are usually tailored to a relatively small number of specialized neuroimaging purposes, and are therefore inherently limited to those specific applications.

However, in order to bypass this limitation, the Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC) has recently released the NITRC Computational Environment (NITRC-CE; [https://www.nitrc.org/projects/nitrc\\_es/](https://www.nitrc.org/projects/nitrc_es/)), which is an on-demand, cloud computing platform that has been pre-configured with a wide range of common neuroimaging software packages—eg, AFNI (<http://afni.nimh.nih.gov/afni/>), ANTS (<http://picsl.upenn.edu/software/ants/>), FSL (<http://fsl.fmrib.ox.ac.uk/fsl/>), FreeSurfer (<http://freesurfer.net/>), etc.—and allows users to add their own commercial or open-source software packages as well—eg, Matlab (<http://www.mathworks.com/products/matlab/>), SPM (<http://www.fil.ion.ucl.ac.uk/spm/>), etc. In this way, NITRC-CE offers extreme flexibility and the convenience of allowing the users to use whatever software packages they are already familiar with. Another nice feature of NITRC-CE is that it can be set up and run on either private (ie, institutional) or publicly-accessible (ie, commercial) clouds. If the users' university or research institution has a high-performance computing cluster or is part of a larger research grid that has enough computing power to run the analyses, then they may build their own computational environment on the local cluster or grid. (see [http://www.nitrc.org/plugins/mwiki/index.php/nitrc:User\\_Guide\\_-\\_Building\\_Your\\_Own\\_NITRC-CE](http://www.nitrc.org/plugins/mwiki/index.php/nitrc:User_Guide_-_Building_Your_Own_NITRC-CE) for instructions). Alternatively, if the user requires more computing power than is available to them, it is also possible to launch and run NITRC-CE through commercial cloud providers. (see [http://www.nitrc.org/plugins/mwiki/index.php/nitrc:User\\_Guide\\_-\\_Launching\\_NITRC-CE\\_with\\_AWS\\_Marketplace](http://www.nitrc.org/plugins/mwiki/index.php/nitrc:User_Guide_-_Launching_NITRC-CE_with_AWS_Marketplace) for instructions).

### When Cloud Computing Might (or Might Not) Make Sense

It should be noted that since standalone desktop workstations provide sufficient computational power for many routine



tasks, these will continue to be a mainstay in neuroimaging analyses for the foreseeable future. Nonetheless, a subset of computationally intensive analyses may significantly benefit from (or require) cluster, grid, or cloud-level resources.

However, because of the setup time typically required to: 1) transfer the raw data to the cloud, 2) upload, modify and troubleshoot the analysis software for offline analyses, and 3) download the final results, there are inherent trade-offs between the convenience of running analyses on desktop workstations (ie, where one's data are likely located, one's directories are already set up, and one is familiar with the operating system and software) and the potential gains in computer processing speed. Therefore, when confronted with having to choose between mutually exclusive alternatives, given limited resources—in this case, the amount of time, effort and money required to analyze a particular neuroimaging data set—one must weigh the opportunity costs of each possible approach, as well as the trade-offs (eg, between the amount of extra human time and effort required to set up a cloud-based analysis vs the expected gains in computer processing time). Of course, factors such as the amount of time required to set up an offline analysis will critically depend on several factors (not least of which is the level of one's experience in running such analyses); however, in cases involving simple analyses with minimal computational demand (eg, jobs that are expected to take ~45 minutes or less on a standalone workstation), most hypothetical gains in computing time will be negated for even the most proficient cloud-computing users.

However, with the advent of multi-band (aka, *simultaneous multi-slice*) acquisition techniques such as simultaneous echo-refocused EPI<sup>12</sup> or controlled aliasing techniques such as CAIPRINHA<sup>13,14</sup>, it is now possible to acquire whole-brain fMRI data with temporal resolutions below 1 s/volume—ie, more than 600 images for every 10 minutes of scanning—and high angular resolution diffusion imaging (HARDI)<sup>15</sup> or diffusion spectrum imaging (DSI)<sup>16</sup> data sets containing several hundred diffusion-weighted images. As a result, it is now possible, even for single subjects, to acquire extremely large amounts of data that may necessitate correspondingly large increases in image-processing and analysis time; and even for smaller data sets, there are certain computationally-demanding analyses—eg, non-linear normalizations,<sup>17</sup> automated brain segmentation,<sup>18</sup> lesion segmentations,<sup>19</sup> etc.—which might benefit from more computing power than is available on conventional laptop or desktop workstations.

At the most abstract level, there are at least two general categories of neuroimaging analyses for which cloud computing is likely well suited (ie, worth the extra set-up time and effort). These include: 1) any large, computationally-intensive job that exceeds the computational limits of conventional workstations (eg, due to enormous physical memory requirements that cause software crashes and/or out-of-memory errors), and/or 2) relatively large numbers of analyses that could be run individually (ie, in serial) on conventional

workstations, but would collectively take a very long time (eg, days, weeks, or months).

An example of a large, computationally-demanding job that we routinely perform with cloud computing recourse is non-linear normalization of either  $T_1$ - or diffusion-weighted images via the DiffeoMap toolbox in MRISStudio (<https://www.mristudio.org/>, Johns Hopkins University School of Medicine, Baltimore, MD, USA). In this procedure, individual subject data are first warped to a standard template<sup>20–22</sup> using a 12-parameter affine (linear) transformation, followed by high-dimensional, non-linear warping with the large deformation diffeomorphic metric mapping (LDDMM) algorithm.<sup>23</sup> Since the second step of this procedure (ie, the portion involving the non-linear LDDMM processing) cannot be run on conventional workstations, the creators of DiffeoMap and MRISStudio have created and continue to support a platform for free remote processing (ie, cloud computing) using a high-performance computing cluster that comprises 256 eight-core processors with 64GB of dedicated memory per node, and over one petabyte of hard disk storage.\* In this way, the initial set-up parameters are entered into DiffeoMap on the user's workstation, and then are automatically uploaded as a job file (along with the subject and template images) to the remote cluster, which performs the LDDMM processing. For single-channel LDDMM processing (ie, normalization with one type of image contrast), the analyses can be performed in as little as 30 minutes (depending on the server demand and backlog) for standard 1mm cubic, whole-brain,  $T_1$ -weighted anatomical images; whereas multi-channel LDDMM processing (ie, based on two or three different image contrasts) typically takes on the order of 90 minutes. Once the processing has finished, the remote server automatically emails the user a secure link to retrieve their results. Since the LDDMM analysis is not possible on conventional desktop workstations, this is an obvious example of where high-performance computing is necessary—and in this case, the free DiffeoMap software provides an intuitive and user-friendly interface that seamlessly initiates the data transfer and then automatically runs the analysis in the cloud.

In addition to tasks that cannot be run on conventional workstations (eg, due to RAM limitations, etc.), there are certain types of neuroimaging analyses that require significant amounts of processing time to complete, and running these on a cluster or cloud may therefore be an attractive solution to speed up the analyses. One such example from our own lab that falls into this category is multi-component  $T_2$ -relaxation myelin water imaging, which uses a non-negative least squares (NNLS) method to estimate the myelin water fraction, which is given by the ratio of myelin water (ie, water trapped between the myelin layers) to the total water content within each voxel.<sup>24,25</sup> This method is relatively robust<sup>26</sup> and has been validated with histopathology,<sup>27,28</sup> but one drawback

\*1 petabyte = 1,000 terabytes = 1,000,000 gigabytes.



is that each whole-brain human analysis takes on the order of 24 hours to complete on contemporary desktop workstations with quad-core CPUs. Therefore, even after all of the pre-processing has been completed, the final step to generate the myelin water fraction maps for a moderately-sized study with 30 subjects would take an entire month, and analyzing larger or longitudinal data sets could require several months of uninterrupted computer time, which is not at all practical. Fortunately, we have been able to work around this limitation by taking advantage of free cloud computing resources because our institutional affiliation allows us to use WestGrid and Compute Canada resources at no cost. The enhanced power of these research grids allows us to analyze all of the data from different subjects and different time-points simultaneously (ie, in parallel) and retrieve our data in a fraction of the time (eg, less than one day instead of a several months), compared to running the same analyses on standalone workstations.

### Cost and Other Considerations Regarding Cloud Resources

As illustrated above, there are several potential benefits of using cloud computing resources for neuroimaging data storage and analysis; however, there are also several other considerations (and a few potential barriers) that we feel warrant at least a brief discussion as well. Cost is often a major consideration in research, and will be discussed. However, in addition to economics, there are several other factors that will be touched upon, including: data security, reliability, legal issues, privacy, open standards, compliance, and the long term validity of cloud computing,<sup>29,30</sup> as well as data transfer bottlenecks, software licensing, performance unpredictability, data confidentiality, and so on.<sup>31</sup>

**Cost.** In the current research funding climate, where grants are increasingly competitive and budgets are tight, one of the major considerations in terms of whether or not to use cloud resources at all (and if so, which resources to use) will be the cost.

Neuroimaging research labs have historically employed a local infrastructure model in which researchers use software and run analyses on their own workstation(s). However, these resources require: 1) large up-front costs (ie, to purchase computers, data servers and software), 2) dedicated work space (ie, for each computer system), and 3) a significant amount of low-level system administration (ie, to set up each system, manage software licenses, configure network permissions and manage users, perform routine backups and anti-virus checks, etc.).<sup>32</sup> For these and other reasons—including the ongoing power, cooling and maintenance costs<sup>33</sup> associated with running multiple systems—grid or cloud computing may actually offer a less expensive alternative to purchasing and supporting top-of-the-line standalone workstations.<sup>34,35</sup>

The other benefit of cloud computing is that it is almost infinitely scalable, and one only pays for what s/he uses on a publicly available cloud, and may not have to pay at all on a

local cluster or research grid. For these reasons, fewer and less expensive workstations can be purchased (ie, since ordinary laptop or desktop computers may suffice, rather than purchasing high-end multi-core CPU or GPU workstations), and the computing power can then be greatly increased later, depending on each user's specific needs. In essence, a laboratory can start relatively small and grow as it requires, instead of purchasing equipment based on what is thought will be needed a month, a year, or several years in the future. And because up-front capital costs are minimized (and replaced by operating costs that are lower and scalable), more grant money can be spent on data acquisition, additional graduate students, or hiring other personnel.

One potential pitfall, however, is that while cloud computing reduces the cost of infrastructure, the cost of data communication (ie, the cost of data transfer to and from the cloud) and the ongoing computing costs will likely be increased,<sup>36</sup> and these issues may be exacerbated when using a hybrid cloud model in which data are distributed among multiple public/private/community clouds.<sup>37</sup> One potential solution that avoids the high cost and bandwidth bottlenecks associated with extremely large data transfers is to physically ship the data on hard disks,<sup>38</sup> which is the model currently being employed to distribute more than 20 TB of data resulting from the Human Connectome Project (<http://www.humanconnectome.org/data/connectome-in-a-box.html>). Another potential solution would be using offline file transfer services such as Globus (<https://www.globus.org>). Globus is a fast, reliable, and secure file transfer service that maximizes bandwidth usage, manages the security configurations, and provides automatic fault recovery in case of transfer interruption. Compute Canada has a Globus portal (<https://www.computeCanada.ca/research-portal/national-services/globus-portal/>) where Compute Canada users can use their free Globus account to login and then transfer data within Compute Canada resources as well as between local storage and Compute Canada sites. Having said that, with the exception of extremely large data sets (eg, hundreds of GB or larger), bandwidth considerations are not likely to be a major factor for the majority of neuroimaging researchers—especially since most academic or government research institutions have robust, high-speed networks and neither monitor nor bill individual investigators for bandwidth use. Nonetheless, it should be noted that—all else being equal—file transfer (ie, upload and download) times will scale with file size, and this could become a factor in extreme cases.

Another thing to consider, if a free institutional cluster or grid is not available, is that using commercial cloud providers for sustained, long-term use may not be economically advantageous over investing in internal resources. For example, certain laboratories that anticipate running computationally-demanding tasks continuously over a long period of time (eg, protein or genetic modeling institutes) may be better served by investing in their own infrastructure or implementing a mixed-use strategy, in which some of the services and applications run





internally and the rest can be hosted in the cloud.<sup>39</sup> However, with the exception of the busiest and most prolific neuroimaging centers, this is not likely to be the case. The reality in neuroimaging research is that analyses are most often performed in relatively short and/or relatively infrequent intervals, which are exactly the types of analyses that would most greatly benefit from the scalability offered by grid or cloud computing.

In terms of cost, the lowest cost—and in many cases free—options are either: 1) Brain Imaging Analysis as a Service (BiAaaS) tools like MRICloud for specific applications, or 2) distributed computing on an institutional cluster or grid. If these resources are not available (or are not sufficient for the desired analyses), it may be necessary to use a pay-per-use commercial cloud provider like Google or Amazon (perhaps combined with NITRC-CE, see above). However, even in the latter case, cloud computing may be more economical than purchasing top-of-the-line standalone workstations for neuroimaging applications.

**Data security.** In addition to cost, another critical consideration before adopting a cloud service is the security of the data. As a baseline, statistics show that lost or stolen devices (ie, laptops, hard drives, and other devices) account for approximately 33% of security breaches, while a further 16% are due to internal theft;<sup>40</sup> and although we would all like to think that our data is safe and secure, this is not necessarily the case. According to one popular report, the Medical/Healthcare industry single-handedly accounted for over 42% of all data breaches reported in 2014,<sup>41</sup> which was far higher than any other sector (ie, Banking/Financial, Business, Education, or Government/Military). Therefore, because neuroimaging data is a type of personal health information, it should (as a rule) be de-identified or anonymized whenever and to whatever extent possible, even on standalone workstations. For example, according to guidelines set out by our institutional privacy officer and research ethics board (REB), we are not allowed to electronically store patient names or initials, gender information, or dates of birth in connection with our imaging data (including in image header files, etc.).

However, since cloud computing involves the transfer of this data over a network, and storage on external servers, it is especially recommended (and may even be subject to more stringent institutional and/or jurisdictional policies) to de-identify datasets before uploading them to any off-site locations, including cloud services. Most, if not all, cloud providers now use data encryption and network architectures that have been designed to satisfy the requirements of security-sensitive users, but that does not mean that cloud services are free from hacking and/or data theft, as evidenced by several recent high-profile cyber attacks.<sup>42</sup> In a network, an organization's data and software are always vulnerable to data loss, phishing, botnet attacks, etc.<sup>37</sup> For example, hackers are using botnet that runs remotely on a collection of machines and provides cheaper options to start an attack on any cloud.<sup>43</sup> As a result, cloud data servers are often kept in different regions and data

are distributed across different servers to protect individual data from hacking. Additionally, some cloud providers offer data encryption while uploading, and this is likely one of the best solutions to prevent unwanted data security breaches.<sup>44</sup> Currently, most researchers also use encryption to ensure that their data are kept private in a cloud.<sup>45,46</sup> However, each potential cloud user should check with their local REB (and privacy officer, if applicable) ahead of time to make sure that their proposed cloud computing and/or data storage solutions comply with federal (eg, HIPPA: Health Insurance Portability and Accountability Act), provincial/state (eg, PHIA: Personal Health Information Act) and institutional regulations prior to transferring any data off-site.

**Data monitoring and seizure.** Because many cloud vendors house and maintain data servers in different countries throughout the world—in part to safeguard the data from third-party cyber attacks (see above)—there is also a possibility that, according to local laws, data stored on these servers can be legally monitored or seized by intelligence, security or other government organizations (eg, the CIA, FBI, etc.).<sup>47</sup> For this reason, some commercial cloud providers now provide users an option to choose where (geographically) their data will be stored. For example, Google Cloud Storage provides an option for users to restrict their data to servers in specific geographic locations—eg, Asia, Europe, or North America at a continental level, or even Eastern USA, Central USA, or Western USA at the regional level (c.f., <https://cloud.google.com/storage/docs/bucket-locations>). Therefore, although data should always be properly de-identified anyway, these storage options are available for researchers who wish to add another level of data security (ie, by restricting where their data will be stored) to safeguard against data monitoring and seizure.

**Data availability.** Data availability means that the resources of a system are accessible to the user whenever they are required,<sup>48</sup> and in the case of cloud computing solutions, security breaches<sup>48</sup> or other issues may impact the ability to access a particular server or database, which may in-turn affect data availability.<sup>49</sup> Also, while there is some redundancy in most cloud systems, power or network failures can cause network slowdowns and lead to interruptions in uploading, downloading or even data processing.<sup>7</sup> However, a more common occurrence with public grids and clouds is that there are numerous service requests to access the same resources, and this not only increases the CPU demand, but also increases waiting times and may (in the worst case) even drop service requests.<sup>50</sup> However, to tackle this problem, cloud providers are including technologies such as Adobe AIR, Google Gears and Curl, through which cloud based applications can run locally whenever there are disruptions in a network connection,<sup>49</sup> and it is now also possible to provide accessible resources in urgent cases by using multi-cloud (aka, *rain cloud*) models.<sup>51</sup>

**Compatibility.** Compatibility refers to the interoperability of the processed data and software within different platforms. Each cloud has its own data storage model, licensing model, and



security model,<sup>52</sup> and these may render cloud services from different providers incompatible with one another.<sup>53</sup> For example, some Google cloud products and services are incompatible with Microsoft cloud products and services (and vice versa),<sup>29</sup> which may preclude information from being exchanged between cloud systems. Furthermore, similar incompatibility issues may also occur between private and public clouds, which may prohibit users from running one portion of their analysis on an institutional cluster and then performing another portion of the analysis on a public cloud. For these reasons, data transformability into a suitable form for analysis has been a hindrance in the adaptation of big data.<sup>54</sup> Two of the possible solutions are to design an interoperable platform, or to develop a standard and open application program interface (API) so that data portability or software compatibility can be ensured.<sup>37</sup> To date, some cloud vendors have embraced others' APIs (eg, Hewlett-Packard Helion Eucalyptus is compatible with Amazon Web Service), and the Open Grid Forum continues to develop an open standard called the Open Cloud Computing Interface (OCCI) (<http://occi-wg.org/>) that focuses on developing interoperable and portable tools for various management tasks. Finally, some computing grids and cloud services—including Compute Canada and WestGrid—have started to support multiple operating systems (ie, Windows, Linux, etc.) and a variety of commonly used software platforms (eg, Matlab, Python, Java, etc.) to maximize utility, flexibility and compatibility.

**Data integrity.** Data integrity is making sure that the modification of data is done only by the owner or other authorized parties to prevent data misuse or corruption. There are a growing number of applications that allow users to manage and store their data on cloud servers, but these data may have different formats that eventually make it difficult to integrate multiple sources for analysis.<sup>55</sup> Also, these applications must make sure that the data integrity has been maintained,<sup>56</sup> and that data have been preserved.<sup>57</sup> Nonetheless, due to the remote access to data and a lack of support, verification of data integrity remains a challenge for cloud storage solutions, where it may be difficult to detect portions of missing or corrupted data in large data sets.<sup>56</sup>

**Technical expertise and IT support.** Although most of the aforementioned issues are relatively minor and will have little (if any) impact on most users interested in grid/cloud computing for neuroimaging data storage and/or analysis, perhaps the biggest challenge for new users is the lack of technical expertise and proper IT support. For example, first-time cloud users may experience difficulty with: 1) renaming and reformatting their files, 2) uploading their data to the remote server, 3) analyzing/processing their data, and 4) retrieving their results. Unfortunately, these and other technical challenges can lead to wasted time and extra frustration, which are ironically the two things that cloud computing is supposed to save users from. Fortunately, automated processes can improve data preparation and transport,<sup>58</sup> and improve compatibility of software and data formats in the cloud can decrease these

problems. However, the next section will go into further detail on how to get started with cloud computing and avoid some of the most common pitfalls.

## How to Get Started

Due to the growing number of research grids and cloud providers, and the equally wide variety of specific user requirements, it is unfortunately beyond the scope of this or any review to provide a set of instructions that will apply in every case and fulfill everybody's needs. However, the goal of this section is to briefly provide some general advice to new or novice cloud users and how they might be able to start leveraging the benefits of cloud-based resources.

Although it may seem daunting to first-time users, getting started with cloud computing is relatively straight forward, and if there is already an established desktop workflow for a particular type of neuroimaging analysis, the same tasks can typically be done in the cloud with few (if any) changes. As previously noted, "new tools (some commercial and even public), have made it so that dealing with the cloud and running large-scale processing can be rather easy and efficient".<sup>59</sup>

A cloud solution usually offers an interface through which users can manage their accounts, monitor their usage, and set up new compute environment(s). Once the environment is set up (e.g., in either a Linux-based or Windows OS), the user interaction with the remote machine will be similar to regular desktop systems. An important factor to consider is the availability of software to carry out the neuroimaging tasks (ie, in terms of licensing, run-time environment, etc.). Once the proper environment is ready, typical steps involved in doing in-the-cloud computations are:

- Transferring (uploading) de-identified data and any analysis code to the cloud environment
- Performing the required image processing and data analysis (which may require installing and/or configuring the required software, but should then follow the same procedures as running the software on a standalone workstation)
- Completing any desired statistical analysis or other post-processing
- Transferring (downloading) the final results back to a local workstation
- Archiving or deleting the data and analysis code from the cloud

As previously mentioned, transferring data between the cloud and the local facilities could be a bottleneck in the case of large neuroimaging data sets. Therefore, using cloud-based storage (when possible) could enhance efficiency and simplicity, although the cost of cloud storage may factor into whether this is a suitable option.

Most cloud solutions, and all of the major research grids and cloud service providers, supply a detailed guide to get new



users up and running. Some cloud solutions come with all of the tools required to carry out certain types of neuroimaging analysis (eg, BiAaaS platforms), while others simply provide the infrastructure (IaaS) or the infrastructure and platform (PaaS), in which case users are responsible for installing their own analysis software. Although Compute Canada has a cloud solution for life science researchers called GenAp (<https://www.genap.ca/public/home>), most of Compute Canada's cloud solutions are of the latter type, where users need to install all of the necessary software (including the operating system). Therefore, in the absence of a BiAaaS model to perform a desired function, one possible solution would be to use an IaaS cloud model and then load NITRC-CE (discussed above) to configure the cloud environment and automatically load several of the most common neuroimaging analysis packages. Ultimately, the choice of cloud model will come down to each user, depending on their data set (eg, what types of neuroimaging data they wish to analyze), their specific analysis requirements (eg, how much data they have, what types of analyses they wish to perform, etc.), and a host of other circumstances (eg, their funding situation, whether there are suitable BiAaaS solutions available for the desired types of analyses, what institutional resources are available, etc.).

**A special note about clusters.** High-performance computing on local clusters, as opposed to user-friendly cloud solutions, may not provide the illusion of a private computing environment and may therefore require more user involvement and technical expertise. In many cases, users of computing clusters are aware of the shared environment and have to use non-interactive (or batch) modes to perform their analyses. A queuing system may still accept users' job submissions and order them based on predetermined factors (ie, to ensure fairness in the system). However, since most clusters are Unix/Linux based, familiarity of users with a command-based environment is essential (if not required), and installing software on a local cluster is not always as easy as in cloud-based solutions (and it may sometimes be impossible, due to licensing restriction). For these reasons, clouds or grids with user-friendly interfaces and dedicated IT support are likely a better option than local clusters for users with limited technical expertise. However, it should be noted, especially for government institutes and research university clusters/grids, that on-site IT support may be available to assist new users transfer their data and set up their analyses.

## Discussion and Conclusions

In recent years, high-performance grid/cloud computing has matured and grown considerably in popularity; as a result, there are now a relatively large number of academic, government and private computer grids/clouds that are available to various end-users. Nevertheless, while these resources have seemingly been embraced by astrophysicists, geneticists and researchers in other big-data fields,<sup>60</sup> there are—with the exception of a handful of imaging physicists, computer

scientists and biomedical engineers (ie, researchers who are developing new image processing techniques, pipelines and/or software packages)—relatively few radiologists, psychologists, psychiatrists, cognitive neuroscientists and other neuroimaging researchers have put these resources to use. The two possible explanations for this are that researchers in the latter group (ie, who are using imaging to address neuroscience questions) are:

- 1) generally informed about cloud computing resources and how these could potentially benefit their research, but they have made an informed choice not to use these resources; or
- 2) unaware of what cloud computing resources are available, unacquainted with how these could potentially benefit their research, or are uninformed (and likely intimidated) about how much it might cost or how hard it might be to make use of these resources.

Therefore, by introducing the main types of computer architectures, highlighting some of the cluster/grid/cloud resources that are available, discussing some of the considerations (including cost and some of perceived barriers) related to cloud computing, identifying when it might (and might not) make sense to use cloud resources, and then outlining how someone could get started, we hope this primer has addressed most of the questions neuroimaging researchers might have had, but were too afraid to ask about various high-performance computing options.

Although conventional desktops and/or laptop computers will continue to serve a major role in neuroimaging data analysis, we feel that a growing number of neuroimaging researchers might significantly benefit by incorporating cluster, grid and/or cloud computing solutions for certain compute-intensive applications. Especially for the many researchers who only run these sorts of computationally-intense analyses periodically, the scalable (or *elastic*) nature of grids and clouds offers unparalleled processing speed when it is needed, while in many cases, also offering a more cost-effective alternative compared to purchasing and maintaining high-end CPU or GPU workstations that would still pale in comparison. Based on a combination of factors (eg, cost, convenience, time, etc.), there will be some scenarios when it makes sense to run an analysis on the cloud and other scenarios when it does not. However, we advocate that being informed of both options, and therefore having the luxury to choose, is the ideal scenario.

## Acknowledgements

This work was supported by The Winnipeg Health Sciences Centre Foundation (HSCF), The University of Manitoba, WestGrid, and Compute Canada. In particular, we would also like to thank: 1) Dr. Melanie Martin, Dr. Benedict Albeni, and the rest of the *Magnetic Resonance Insights* editorial board for inviting us to submit this article; 2) Dr. Jennifer





Kornelsen, Teresa Figley, Kevin Solar, Elena Bilevicius, and Tiffany Kolesar for commenting on preliminary drafts of our manuscript. All trademarks are the property of their respective owners.

## Author Contributions

Reviewed and summarized the literature: ASS, SY, HP, and CRF. Wrote the first draft of the manuscript: ASS, SY, HP, and CRF. Contributed to the writing of the manuscript: ASS, SY, HP, and CRF. Agree with manuscript results and conclusions: ASS, SY, HP, and CRF. Jointly developed the structure and arguments for the paper: ASS, SY, HP, and CRF. Made critical revisions and approved final version: ASS, SY, HP, and CRF. All authors reviewed and approved of the final manuscript.

## REFERENCES

1. Glasser MF, Sotiropoulos SN, Wilson JA, et al. The minimal preprocessing pipelines for the human connectome project. *Neuroimage*. 2013;80:105–124.
2. Ugurbil K, Xu J, Auerbach EJ, et al; WU-Minn HCP Consortium. Pushing spatial and temporal resolution for functional and diffusion MRI in the human connectome project. *Neuroimage*. 2013;80:80–104.
3. Van Essen DC, Smith SM, Barch DM, et al. The WU-Minn human connectome project: an overview. *Neuroimage*. 2013;80:62–79.
4. Mell P, Grance T. The NIST Definition of Cloud Computing: Recommendations of the National Institute of Standards and Technology; 2011. NIST Spec Publ 800–145 Reports on:7.
5. Null L, Lobur J. *The Essentials of Computer Organization and Architecture*. 4th ed. Burlington, MA: Jones & Bartlett Learning; 2014.
6. Nickolls J, Dally WJ. The GPU computing era. *IEEE Micro*. 2010;30:56–69.
7. Tafula SMN, da Silva NM, Rozanski VE, et al. ABRIL—advanced brain imaging lab: a cloud based computation environment for cooperative neuroimaging projects. In: Conf Proc IEEE Eng Med Biol Soc. 2014:534–537.
8. Lauzon CB, Asman AJ, Esparza ML, et al. Simultaneous analysis and quality assurance for diffusion tensor imaging. *PLoS One*. 2013;8:e61737.
9. Li Y, Shea SM, Lorenz CH, et al. Image corruption detection in diffusion tensor imaging for post-processing and real-time monitoring. *PLoS One*. 2013;8:e49764.
10. Tang X, Yoshida S, Hsu J, et al. Multi-contrast multi-atlas parcellation of diffusion tensor imaging of the human brain. *PLoS One*. 2014;9:e96985.
11. Tang X, Crocetti D, Kuttan K, et al. Segmentation of brain magnetic resonance images based on multi-atlas likelihood fusion: testing using data with a broad range of anatomical and photometric profiles. *Front Neurosci*. 2015;9:1–13.
12. Feinberg DA, Moeller S, Smith SM, et al. Multiplexed echo planar imaging for sub-second whole brain fMRI and fast diffusion imaging. *PLoS One*. 2010;5:e15710.
13. Breuer FA, Blaimer M, Heidemann RM, et al. Controlled aliasing in parallel imaging results in higher acceleration (CAIPIRINHA) for multi-slice imaging. *Magn Reson Med*. 2005;53:684–691.
14. Breuer FA, Blaimer M, Mueller MF, et al. Controlled aliasing in volumetric parallel imaging (2D CAIPIRINHA). *Magn Reson Med*. 2006;55:549–556.
15. Tuch DS, Reese TG, Wiegell MR, et al. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magn Reson Med*. 2002;48:577–582.
16. Wedeen VJ, Hagmann P, Tseng W-YI, et al. Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. *Magn Reson Med*. 2005;54:1377–1386.
17. Ashburner J, Friston KJ. Nonlinear spatial normalization using basis functions. *Hum Brain Mapp*. 1999;7:254–266.
18. Khan AR, Wang L, Beg MF. FreeSurfer-initiated fully-automated subcortical brain segmentation in MRI using large deformation diffeomorphic metric mapping. *Neuroimage*. 2008;41:735–746.
19. Lao Z, Shen D, Liu D, et al. Computer-assisted segmentation of white matter lesions in 3D MR images using support vector machine. *Acad Radiol*. 2008;15:300–313.
20. Mazziotta JC, Toga AW, Evans A, et al. A probabilistic atlas of the human brain: theory and rationale for its development. *Neuroimage*. 1995;2:89–101.
21. Mazziotta J, Toga A, Evans A, et al. A probabilistic atlas and reference system for the human brain: international consortium for brain mapping (ICBM). *Philos Trans R Soc Lond B Biol Sci*. 2001;356:1293–1322.
22. Mori S, Oishi K, Jiang H, et al. Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. *Neuroimage*. 2008;40:570–582.
23. Beg MF, Miller MI, Trounev A, et al. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int J Comput Vis*. 2005;61:139–157.
24. MacKay A, Whittall K, Adler J, et al. In vivo visualization of myelin water in brain by magnetic resonance. *Magn Reson Med*. 1994;31:673–677.
25. Whittall KP, MacKay AL, Graeb DA, et al. In vivo measurement of T2 distributions and water contents in normal human brain. *Magn Reson Med*. 1997;37:34–43.
26. Meyers SM, Vavasour IM, Mädler B, et al. Multicenter measurements of myelin water fraction and geometric mean T 2: intra- and intersite reproducibility. *J Magn Reson Imaging*. 2013;38:1445–1453.
27. Laule C, Leung E, Lis DKB, et al. Myelin water imaging in multiple sclerosis: quantitative correlations with histopathology. *Mult Scler*. 2006;12:747–753.
28. Laule C, Kozlowski P, Leung E, et al. Myelin water imaging of multiple sclerosis at 7 T: correlations with histopathology. *Neuroimage*. 2008;40:1575–1580.
29. Popović K, Hocsenski Z. Cloud computing security issues and challenges. In: MIPRO, 2010 Proc 33rd Int Conv. 2010:344–349.
30. Yang J, Chen Z, Definition A. Cloud computing research and security issues. In: Computational Intelligence and Software Engineering (CiSE), 2010 International Conference on. Wuhan: 2010.
31. Armbrust B, Griffith R, Joseph AD, et al. A view of cloud computing. *Commun ACM*. 2010;53:50–58.
32. Markram H. Bioinformatics: industrializing neuroscience. *Nature*. 2007;445:160–161.
33. Belady CL. In the data center, power and cooling costs more than the it equipment it supports. *Electron Cool*. 2007;13(1):24–27.
34. Ross JW, Westerman G. Preparing for utility computing: the role of IT architecture and relationship management. *IBM Syst J*. 2004;43:5–19.
35. Rosenthal A, Mork P, Li MH, et al. Cloud computing: a new business paradigm for biomedical information sharing. *J Biomed Inform*. 2010;43:342–353.
36. Leinwand A. The Hidden Cost of the Cloud: Bandwidth Charges. 2009.
37. Dillon T, Wu CWC, Chang E. Cloud computing: issues and challenges. In: Adv Inf Netw Appl (AINA), 2010 24th IEEE Int Conf. 2010:27–33.
38. Gray J, Patterson D. A conversation with Jim Gray. *ACM Queue*. 2003;1:8–17.
39. Talukder AK, Zimmerman L, Prahalad HA. Cloud economics: principles, costs, and benefits. In: Cloud Computing: Principles, Systems and Applications. 2010:343–360.
40. Mills E. Cloud Computing Security Forecast: Clear Skies. 2009.
41. Identity Theft Resource Center. Data Breach Reports. [http://www.idtheftcenter.org/id-theft/databreaches.html] 2014:1–192.
42. Walters R. *Cyber Attacks on U.S. Companies in 2014*. Washington, DC: Heritage Foundation; 2014:1–5. Herit Found Issue Br No. 4289.
43. Chen Y, Paxson V, Katz RH. What's New About Cloud Computing Security? University of California, Berkeley; 2010:1–8. Rep No UCB/EECS-2010-5.
44. Clark C, Fraser K, Hand S, et al. Live migration of virtual machines. In: Proc Symp Networked Syst Des Implement. 2005:273–286.
45. Cao N, Wang C, Li M, et al. Privacy-preserving multi-keyword ranked search over encrypted cloud data. In: Proc—IEEE INFOCOM. Shanghai: IEEE; 2011:829–837.
46. Lin HY, Tzeng WG. A secure erasure code-based cloud storage system with secure data forwarding. *IEEE Trans Parallel Distrib Syst*. 2012;23:995–1003.
47. Tankard C. Big data security. *Netw Secur*. 2012;2012:5–8.
48. Zissis D, Lekkas D. Addressing cloud computing security issues. *Future Gener Comput Syst*. 2012;28:583–592.
49. Nazir M. Cloud computing: overview & current research challenges. *IOSR J Comput Eng*. 2012;8:14–22.
50. Subashini S, Kavitha V. A survey on security issues in service delivery models of cloud computing. *J Netw Comput Appl*. 2011;34:1–11.
51. Lee S, Park H, Shin Y. Cloud computing availability: multi-clouds for big data service. *Commun Comput Inf Sci*. 2012;310:799–806.
52. Yang X, Wallom D, Waddington S, et al. Cloud computing in e-Science: research challenges and opportunities. *J Supercomput*. 2014;70:408–464.
53. Mont MC, Pearson S, Bramhall P. Towards accountable management of identity and privacy: sticky policies and enforceable tracing services. In: 14th Int Work Database Expert Syst Appl 2003 Proceedings. 2003.
54. Akerkar R, ed. *Big Data Computing*. Boca Raton, FL: CRC Press; 2014.
55. Assunção MD, Calheiros RN, Bianchi S, et al. Big data computing and clouds: trends and future directions. *J Parallel Distrib Comput*. 2014;79–80:3–15.
56. Hashem IAT, Yaqoob I, Badrul Anuar N, et al. The rise of “big data” on cloud computing: review and open research issues. *Inf Syst*. 2014;47:98–115.
57. Sravan Kumar R, Saxena A. Data integrity proofs in cloud storage. In: 2011 3rd Int Conf Commun Syst Networks, COMSNETS. 2011.
58. Shahand S, Benabdalkader A, Huguet J, et al. A data-centric science gateway for computational neuroscience. In: Proceedings Int Work Sci Gateways. Zurich: 2013.
59. Tsafaris SA. A scientist's guide to cloud computing. *Comput Sci Eng*. 2014;16:70–76.
60. Drake N. Cloud computing beckons scientists. *Nature*. 2014;509:543.