Analytics Environments on Demand: Providing Interactive and Scalable Research Computing with Windows

Jason C. Christopher
Research IT
University of California, Berkeley
2195 Hearst Avenue
Berkeley, CA 94720 USA
jchristo@berkeley.edu

ABSTRACT

Historically, the experimental and observational sciences have been well served by traditional High Performance Computing (HPC). More recently, researchers from the life sciences and other domains have joined the HPC ranks. Cloud computing offers promising alternatives to HPC, yet neither HPC nor cloud are sufficient to the meet the computational needs of researchers in other academic domains -- those newer to research computing and big data -- for example, from the social sciences, digital humanities and from professional schools, such as Law and Business. This paper describes the development and practice of a research computing service that provides interactive and scalable computing in a Windows environment, including the technical and end-user support challenges that were overcome to provide the service.

CCS CONCEPTS

• ***Information systems***~*Computing Platforms*

KEYWORDS

Cloud computing, research computing, service management, virtualization

ACM Reference format:

Jason Christopher. 2017. Analytics Environments on Demand: Providing Interactive and Scalable Research Computing with Windows. In *Proceedings of Practice & Experience in Advanced Research Computing. New Orleans, Louisiana, USA, July 2017 (PEARC17)*, 5 pages.

PEARC17, July 09-13, 2017, New Orleans, LA, USA

ACM ISBN 978-1-4503-5272-7/17/07⋯$15.00
http://dx.doi.org/10.1145/3093338.3093369

1 INTRODUCTION AND CHALLENGES

High energy physics, astrophysics, cosmology and astronomy are examples of experimental and observational sciences served by traditional High Performance Computing, (HPC). More recently, domains from the life sciences have joined in the community served by HPC. This model of computation, ancient by technology standards, like long-lived redwoods, continues to serve and accelerate scientific research, but the model does not fully meet the computational needs of researchers in other domains -- those newer to research computing and big data, for example, from the social sciences, digital humanities and from professional schools, such as Law and Business.

This paper describes an approach to research computing, a research IT service, that complements traditional HPC and cloud services by providing scalable, interactive, web-accessible research computing in a Windows desktop environment. The service is referred to as Analytics Environments on Demand (AEoD) and is designed to meet the computational challenges of researchers with requirements unsuitable for or in which HPC and/or cloud environments do not provide a specific benefit. For example, researchers:

1) who need software that does not run on HPC (e.g. Windows-only, ArcGIS, 3Di, WCopyFind…)
2) without prior HPC or cloud experience, where the learning curve may be regarded by the researcher as too demanding
3) who have special environment/software requirements to which HPC is not adapted (e.g. interactive RStudio)
4) whose self-service/on-demand compute usage model does not fit into HPC scheduler configurations.

The service challenges are more than computational. End-user support of unmanaged Windows environments on a campus with thousands of researchers is impossible with a limited support staff (<1.0 FTE). Further, interactive computing comes with a higher expectation of system responsiveness and support.

This is cultural and cannot be dismissed or minimized during service design. To deploy and support the AEoD service effectively, we needed a service model that would appeal to and enlist local IT staff to participate in ongoing service development, capacity planning and user support. Sizeable as the technical challenges are, the support and service challenges are greater. Yet the potential benefits are obvious. By providing a form of computation beyond HPC or cloud, we equip unserved researchers with the computational tools they need to conduct and hopefully accelerate their research.

2 WHAT SPECIFIC NEEDS WERE IDENTIFIED?

In consultation with researchers and IT professionals on campus, we identified the following computational and support requirements. Researchers need:

1) analytics stacks that are complex or time-consuming to install and configure on multiple machines running different operating systems or OS versions, for example, in a workshop or advanced instructional setting
2) the ability to use modest resources during design/debug/exploration phases, and then, on occasion, to scale up to more CPU and RAM to run a full model or process large datasets
3) a Windows-based platform, for familiarity and ease-of-use
4) access to a variety of storage options including thumb drives, attached drives, storage infrastructure provided by a department or school, file servers, cloud storage from Box and Google Drive, and storage at remote locations
5) occasional use of expensive software
6) consulting, guided access to resources, do-it-yourself documentation
7) responsive, local end-user support
8) occasional access to secure computing and secure storage

Identity and access management for the service is assumed and is essential, if not articulated by researchers.

3 CONTENTION FOR AVAILABLE RESOURCES ON A SHARED SYSTEM

In some shared research computing environments, one user's compute task may grow to consume all the available resources (CPUs/RAM) of the system, leaving other users unable to start or continue work that is already in progress, or to experience an unexpected reduction in performance. This is referred to as resource contention. In some instances, it is difficult to predict when this may happen. For AEoD to be successful, it needed to provide a solution to the resource contention within and between groups of researchers. In discussion with campus IT professionals, we identified this as the most critical problem to solve for researchers.

4 SOLUTION - ARCHITECTURE

In light of the needs and challenges described above, we designed a service that would provide virtualized Windows research desktops using: VMware for virtualization tasks, Citrix for session management, easy web-accessibility and integration with Active Directory, and two tiers of block storage, as well as CIFS mountable storage. These platforms are common in enterprise computing, but are not configured to support our researchers' needs, nor do they include some desired features. Multiple conversations with our central IT staff partners and a period of experimentation indicated that most of our desired functionality and features *were* possible with configuration changes and the addition of a few new features that were available but not implemented in the current enterprise architecture, including a partial solution for resource contention.

By use of containers in the Citrix platform called delivery groups, we were able to assign specific users to a particular virtualized instance or instances. Members of a delivery group are still able to consume all available resources within their group, but are unable to contend for resources outside their delivery group. Thus, contention for resources is contained within boundaries that can be accepted by researchers and supported by end-user support staff. Researchers within a group can contend with each other, but not with those outside their group.

As to the other needs, they were addressed as follows: complexity of analytics stacks was solved in partnership with domain experts. Stacks can be cloned and shared with multiple users, saving time. Need for use of modest and occasionally larger resources was accomplished through provision of smaller and larger instances and via a CPU and Memory hot swap feature available in VMware that allows an administrator to increase memory on a running instance without a restart. CPU increases are also possible, but require a restart. Access to a variety of storage options was achieved by providing a full desktop environment with internet access and enabled web browser and access to local hard drives and peripherals. This poses some increased security risks, but these risks are mitigated by instance or group isolation. Interactivity, familiarity and ease-of-use were accomplished by use of a Windows desktop OS, in some instances Windows server. Access to prohibitively expensive software was accomplished through grant funding and potentially through cost sharing. For example, a license for a single concurrent use of a piece of software may be shared by many via virtualization and remote access to the instance. We employed this method to provide 20+ users across campus with access to an instance running advanced optical character recognition software.

4 SOLUTION – SUPPORT AND SERVICE DEVELOPMENT

As stated previously, interactive computing on Windows comes with a different expectation by users of system responsiveness and support, as opposed to computing on a "fire and forget" compute platform. If expected interactivity on the screen becomes slowed or paused at intervals, the user will likely perceive this as a problem and an occasion to request end-user support. Further, system administration tasks, including patching and security updates on unmanaged Windows systems, would overwhelm our group's available user services staffing.

In light of the above and other constraints described below, we elected to adopt a service model that provides AEoD in the form of a partnership between our group, Berkeley Research Computing (BRC), and a Partner Organization -- a school, department, research unit, etc. In short, BRC provides each partner with an allocation of resources -- 10 CPUs / 136GB of RAM at $0.00 cost to the Partner and maintains the backend infrastructure (VMware, Citrix, storage). The Partners, in turn, agree to designate an AEoD Partner Administrator who will distribute and manage the local allocation of resources and provide end-user support. The terms of the partnership are described in a lightweight MOU.

4 AEoD WORKING GROUP

An additional allocation of resources (CPUs/RAM) is made possible through an AEoD Working Group. This group is charged with providing capacity planning and governance for the service, and is composed of BRC staff, the Partner Administrators and other contributors from Partner organizations who actively use AEoD for research or advanced instruction. Participation in the group is required of all AEoD Partner Administrators. Partners are incentivized to participate fully in the working group and be good citizens so that they may more easily gain access to additional resources, as needed. Further, partnership allows for extended support across the Partner Organizations -- partners agree to support each other in the event of absences, vacations, sick leave, etc.

By partnering with academic departments, we shift the difficult question of how to fairly distribute available resources from our group to the Partner Organization, where it rightly should be. Alternatively, we would have to come up with some sort of fair use schema that seems likely to invoke frustration among researchers and between departments.

5 OUT OF SCOPE

Secure computing is outside the scope of AEoD, but we are aware of the growing need for this on campus. We have applied for seed grant funding from the Center for Information Technology Research at UC Berkeley (CITRIS) to support development of a secure version of the AEoD service.

5 AEoD IN PRACTICE AT UC BERKELEY'S HAAS SCHOOL OF BUSINESS

A cohort of students in Haas's Masters of Financial Engineering (MFE) program have been using AEoD to work on their Applied Finance Projects, the final projects in the MFE program. Five Applied Finance Project teams used AEoD resources, serving a total of twenty students.

The Applied Finance Projects employ quantitative finance tools and techniques learned in the MFE program and practiced in the program's internships. Students have the option of completing a one- or a three-credit project. The team that creates the top-ranked three-credit project receives the $5,000 Morgan Stanley Applied Finance Project Award.

The AEoD Service provided an MFE student and his Applied Finance Project team with computational resources to conduct their research. Specifically, each team was provided with a web-accessible, Windows-based, multi-tenant research desktop, with scalable compute power.

With AEoD, the project team created "a central place to aggregate large amounts of data from CRSP and COMPUSTAT, without needing to store it separately on each person's computer."

"The AEoD platform," says the student, "let us perform time-consuming computation on the cloud without interfering with our other tasks on our personal computers. Simultaneous logins helped us access the same data to perform different analyses at the same time." Saving time on setup and computation accelerated the completion and ultimately enhanced the success of the project.

Use of AEoD facilitated the students' research at Haas, but also allowed for a transformation in the way IT is provisioned to support the Program. Aaron Nebres, a technology manager at Haas, acts as a local AEoD Partner Administrator to coordinate and support the use of AEoD at his school and in the MFE Program. He describes the use of AEoD in the MFE Program as follows:

> "The administration of AEoD simplifies and reinvents the way MFE distributes server resources for students.
>
> Prior to AEoD, I had to re-purpose decommissioned servers as well as take servers out of our high frequency server pool to meet student demands. Maintaining and managing the software on each particular server was inefficient (and painstaking) as installs were done manually. The MFE director is keen on mimicking the group environments and standards of the finance industry, so it was essential to have students be able to log onto a dedicated server.
>
> Fast-forward to the present -- AEoD simplifies the server process as cloning from a base image with all the necessary software eliminates tedious installations for each respective server. In the time that it used to take to set up a physical server for one group, I am able to set

up five groups using the AEoD platform. Because the servers are virtualized within vSphere, I took advantage of the snapshots features to prep and troubleshoot software updates. User administration is as easy as dropping a username into the proper active directory security group. Also, as an added benefit, I am able to re-purpose existing servers (that particular student groups deemed unnecessary as their project needs have changed since then) to groups who have recently requested a server--having done the transition within minutes, exemplifies AEoD's ability to provide on-demand services.

Haas is preparing to expand the use of AEoD beyond the MFE Program by providing research desktops to a cohort of second-year PhD students.

Berkeley Research Computing has signed MOUs with two schools on our campus, the Haas School of Business and the Goldman School of Public Policy. We are currently exploring partnership with the Archeological Research Facility to provide geographical information systems (GIS) support to researchers from a variety of departments and schools on campus.

5 COMPLIMENTARY SERVICES AND MOBILITY OF COMPUTE

The AEoD service is designed to be complimentary with other research computing services offered in our group, BRC. Those services include HPC, cloud computing and a Consulting Service to match researchers with the best computational resources to fit their needs. Yet needs may change over the lifecycle of a research project.

In light of changing needs, our approach is to facilitate mobility between different computational resources, from a personal laptop to AEoD, from AEoD to cloud or HPC, with storage that can be accessed from multiple platforms. Our roadmap calls for completion of this storage component by end of Summer 2017.

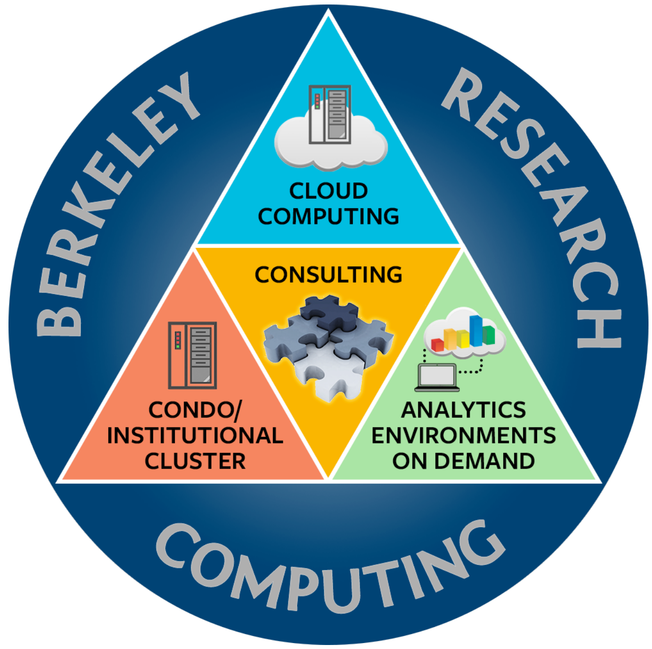

Figure 1: **Berkeley Research Computing - Computing resources and Consulting to match researchers with the optimal compute to meet their needs and facilitate mobility between resources.**

6 ASSUMPTIONS

The service model for AEoD is based on some fundamental assumptions about the changing nature of administrative and research computing at UC Berkeley, which we believe apply generally to many EDU and to some extent GOV environments.

The first assumption is that IT in EDU will continue to shift toward consumption of IT services as commodities from external providers, and away from consumption of campus-grown, or locally produced products and services. Concurrently, departmental IT staff responsibilities will shift away from providing administrative computing toward supporting the IT that promotes the mission of the campus -- toward teaching and research. For example, with more centrally-managed endpoints via BigFix or other similar platform, there will be less need for local system administrators and more need for developers/integrationists and IT staff to provide consulting and access to external services. The staff that remain in campus departments will be charged with increasing their IT support for research. This change presents an opportunity for both central and local IT.

Researchers will continue to look for and rely upon local IT support from staff and students whenever possible for the majority of their computational needs, including administrative and research computing, despite repeated broadcasts and assurances of support from central IT sources. As such, central IT groups do well to engage local IT, where it exists, to promote and help structure the provision of their services. Failure to engage local IT in a constructive manner will diminish the chances of service success.

Research computing has historically been dominated by traditional HPC clusters and services with command line interfaces, schedulers and limited interactive capabilities. These characteristics are unwieldy and impose barriers to entry for researchers in academic domains that have not traditionally called for development of the technical skills necessary to use these systems. This is changing. Researchers *from within and without* domains traditionally served by HPC increasingly want robust, interactive, web-accessible or GUI-fronted tools. AEoD promotes this change.

These assumptions guided development of the AEoD service and continue to animate discussion as we consider improvements to the service.

7 LOOKING FORWARD

In consideration of our progress to date and in light of where we think research computing, as a field, is headed, our path forward with AEoD is defined in terms of: 1) consideration of options for service improvement and expansion gleaned from Partners, 2) integration of AEoD with existing consulting and community research support efforts we provide on campus, and 3) consideration of strategic, alternate methods for provisioning instances -- the "onDemand" portion of AEoD -- including off-premises, cloud-based methods.

Our existing Partners are exploring the possibility of expanding their use of AEoD with the purchase of additional AEoD computing resources. Such resources (nodes) would be added to our existing AEoD cluster as AEoD condo units. Condo owners would have priority but not exclusive access to the additional resources.

A large community of researchers from business school professors to hydrology researchers employs geographic information system (GIS) software tools. Our Partner development work with GIS research support staff will determine whether or not we can add a GIS cohort as an AEoD Partner. Success with staff to-date suggests a positive outcome, but we have learned from experience to be cautious. It can be difficult for organizations to change, even when the change is desired.

Our existing consulting and access program (which includes training and outreach) supports access to our local HPC cluster, Savio, as well as to cloud resources from XSEDE, Amazon, Google and Microsoft. We plan to integrate AEoD more fully with our consulting by training our consultants, expanding do-it-yourself AEoD materials for community consumption, and conducting training and outreach sessions across campus.

Our service model is a good fit for larger organizations on campus with ample support staff, such as professional schools, but some researchers we would like to serve do not have that luxury. This presents a support challenge to us. We are currently supporting some of these researchers ourselves, with a limited set of service options -- ephemeral instances with no permanent storage -- but we would like to be able to serve more researchers in these communities. We are discussing this issue with our Library partners, who have expressed an interest in potentially supporting groups without an obvious candidate for the AEoD Partner Admin role.

Our choice of a solution for the "onDemand" part of AEoD -- VMware +Citrix + local Storage -- effectively leveraged a substantive campus investment in Administrative IT, but budget discussion among senior IT management may result in an increase in the cost for these services, prompting us to consider other platforms. Concurrently, our central IT group and other large organizations on campus (College of Engineering) are actively considering migration to the cloud for some existing services. Likewise, we need to develop a clearer understanding of the strategic advantages possible to us using the cloud. Our approach will be to devise and test an architecture at AWS, Google or Azure as a basis upon which we can do a comparison study. Our goal is to be able to migrate within six months of a decision to do so.

AEoD is a new service, not yet well known on campus, but already we have been contacted by researchers who are interested in integrating AEoD into their scientific community gateways for research. We need to understand more about gateways and explore the possibilities that they enable, so that AEoD can potentially serve a vital role in that design if sensible.

After extended deliberation, we expressly ruled out consideration of providing Linux variants in AEoD during the first year of service. There are many options, both local and external for Linux-based research computing (Savio HPC cluster running Scientific Linux, XSEDE, cloud providers), but zero options, prior to AEoD, for Windows-based research computing. However, providing a Linux instance via AEoD would allow us to consider options for the mobility of an instance from AEoD to the cloud, or a very much larger resource, the prospect of which is compelling and, we feel, a likely thread in the future of research computing.