ISBN: 978-3-95450-209-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2019-MOMPR007

SCALABLE HIGH DEMAND ANALYTICS ENVIRONMENTS WITH HETEROGENEOUS CLOUDS

K. Woods, R. Clegg, R. Millward, Tessella Ltd, Abingdon, UK
F. Barnsley, C. Jones, STFC/RAL, Didcot, UK

Abstract

The Ada Lovelace Centre (ALC) at the Science and Technology Facilities Council (STFC) provides on-demand, data analysis, interpretation and analytics services to scientists using UK research facilities. ALC and Tessella have built software systems to scale analysis environments to handle peaks and troughs in demand as well as to reduce latency by provision environments closer to scientists around the world. The systems can automatically provision infrastructure and supporting systems within compute resources around the world and in different cloud types (including commercial providers). The system then uses analytics to dynamically provision and configure virtual machines in various locations ahead of demand so that users experience as little delay as possible. In this poster, we report on the architecture and complex software engineering used to automatically scale analysis environments to heterogeneous clouds, make them secure and easy to use. We then discuss how analytics was used to create intelligent systems in order to allow a relatively small team to focus on innovation rather than operations.

INTRODUCTION

The ALC at the STFC in the United Kingdom (UK) has been established as an integrated, cross-disciplinary data intensive science centre, for better exploitation of research carried out at large scale UK National Facilities including the Diamond Light Source (DLS), the ISIS Neutron and Muon Facility, the Central Laser Facility (CLF) and the Culham Centre for Fusion Energy (CCFE).

The ALC has the potential to transform research at the facilities through a multidisciplinary approach to data processing, computer simulation and data analytics. The impact will be felt across the many science disciplines and communities these facilities support, including industry and academia. However, for many National Facilities user communities, the computing infrastructure can be very difficult to use. Many scientists lack the specialist expertise to fully exploit the advanced computing infrastructure and services available to them. Scientists need to focus on analysing and interpreting the growing volume of data they obtain during their research; the mechanics of managing datasets and configuring environments can be an obstacle to their work. The Ada Lovelace Centre and Tessella have developed a set of tools to facilitate the scaling of computing infrastructure in order to respond flexibly to variations in demand.

PROBLEM

In supporting users of UK National Facilities, ALC faces several challenges:

- the very large and growing volumes of data generated by scientific instruments at facilities such as DLS, ISIS and CLF.
- the variety of scientific techniques (e.g. neutron scattering, x-ray scattering, laser scattering) employed by researchers requires different data analysis techniques.
- each data analysis technique must be made available to researchers in a consistent, repeatable manner.
- analysis of experimental data is typically performed at a scientist's home institute.
- demand for computing resources is variable and unpredictable.

Large datasets generated by experiments are most efficiently stored in data archives located at the National Facilities. It is frequently impractical to transfer such large datasets (often several hundred GB or more in size) to a scientist's home institute, either on physical media or via the internet. Therefore, there is a pressing need to ensure data can be moved quickly and efficiently between archives and compute clusters. Furthermore, scientists need to select the best tools to process and analyse their data, which requires that their experimental data is available on the most appropriate compute resources.

SOLUTION ARCHITECTURE

To address these problems, ALC has created the Data Analysis as a Service (DAaaS) platform. ALC and Tessella have created three key infrastructure components which underpin the operation of DAaaS.

External Cloud Provisioning

STFC operates a high-performance computing system, which runs the OpenStack [1] software platform. Within the STFC network this resource appears to researchers as a private cloud. DAaaS is a collection of cooperating utilities which run on top of this OpenStack instance.

ALC has access to additional compute clouds (for example, via STFC's IRIS eInfrastructure initiative [2]). Commercial cloud offerings also represent a potential source of compute resource that ALC can exploit, if necessary.

To support the operation of the DAaaS across multiple, heterogeneous clouds, we created a provisioning service. Provisioning DAaaS infrastructure on an external cloud is a multi-step process.

1. We use Ansible [3] playbooks and AWX [4] (running on the core DAaaS system in the ALC cloud) to provision the network infrastructure needed for the external DAaaS infrastructure with:
   - a Virtual Private Cloud (VPC) [5].
   - a public and a private subnet within that VPC.

ISBN: 978-3-95450-209-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2019-MOMPR007

- a Network Address Translation (NAT) Gateway [6] to allow internet traffic for the private instances.
- route tables for each of the subnets to direct to traffic to the NAT Gateway for the private subnet and the Internet Gateway [7] for the public subnet.

2. Ansible playbooks are used again to provision and configure the external cloud machines in the private subnet with an OpenVPN [8] client in order to route traffic back to DAaaS.
3. We then establish replicas of DAaaS infrastructure services in the private subnet of the external cloud including:
   - a Squid proxy [9].
   - a CernVM File System (CVMFS) proxy [10].
   - a Data Movement System (discussed below) web API.
   - a Data Movement Transfer System (also discussed below) endpoint, where the other endpoint is at ALC.
4. Use the VMM (also discussed below) to spin up preconfigured virtual machines for data analysis in the public subnet of the external cloud.

To aid security, the infrastructure machines (CVMFS, Squid, etc.) are located inside the private subnet; a bastion server [11] in the public subnet is required to provide restricted access to the private machines.

The end result of this process is illustrated in Fig. 1:

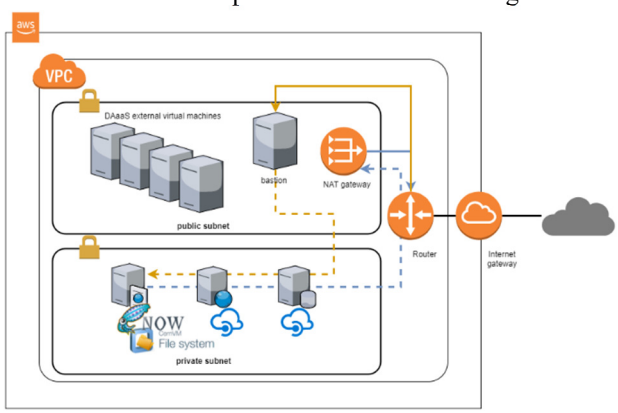

Figure 1: DAaaS provisioned for external cloud.

Ansible includes support for different cloud providers (e.g. OpenStack, Amazon Web Services (AWS) [12] or Azure [13]). We exploit this to make the playbooks cloud agnostic from a system administrator's point-of-view. Thus, the provisioning service enables DAaaS to establish itself in clouds other than ALC's own.

This ability confers great flexibility on DAaaS. It can exploit external compute resources:

- when it has insufficient capacity to meet spikes in demand,
- when portions of STFC's infrastructure are not available (e.g. for routine maintenance)
- when it needs to satisfy special requirements (e.g. GPUs or very large amounts of RAM).

Virtual Machine Manager

DAaaS provides scientists with a range of virtual machines to perform data processing and analysis of experimental data. The virtual machines are pre-provisioned with appropriate data analysis tools. Using a GUI, the scientist simply selects the virtual machine most appropriate to their needs. The GUI passes the request for a virtual machine to the Virtual Machine Manager (VMM).

The VMM was created to manage multiple pools of virtual machines. Each pool contains a number of identical pre-provisioned virtual machines. The size of the pool is configurable, according to the expected demand for virtual machines of that type. Different pools contain different types of virtual machine configured for different types of data processing and analysis. For example, one pool of virtual machines can be provisioned with Mantid [14] for general analysis of neutron scattering data, a second pool could again be provisioned with Mantid, but configured for many CPUs (or GPUs) and/or large memory requirements, a third pool can contain virtual machines provisioned with other data analytics tools. Virtual machines are allocated to users on demand.

The VMM is built with libcloud [15], which hides differences between different cloud provider APIs, providing a means of managing different cloud resources through a unified API. Consequently, the VMM is capable of managing pools of virtual machines spread over more than one cluster, enabling DAaaS to operate on multiple clusters or over multiple clouds.

It is important to emphasize that the VMM simply manages pools of virtual machines. The content of the virtual machines (i.e. the installed data analysis environments and software tools) is completely irrelevant to the VMM. In other words, the virtual machines can be tailored for any type of data analysis task, whether that be in the physical sciences, biological sciences or any branch of data science.

Data Movement System

The large datasets generated by experiments are most efficiently stored in dedicated data storage archives, which are not always co-located with compute resources. The Data Movement System (DMS) is used to efficiently transfer large data files, using a third-party transfer technique, from the source storage on the target storage whilst also preserving permissions. The DMS comprises three parts:

1. a FUSE [16] client, which forms part of a virtual machines used for data analysis. FUSE provides a Unix-like file system for the end-user.
2. a web API, which orchestrates the data transfer.
3. the Data Movement Transfer System (DMTS) which performs the data transfer.

The DMTS employs a third-party transfer mechanism. Data is transferred directly from the source (e.g. a data archive) to the target (e.g. a compute cluster), to avoid the overhead of transferring data via the service making the request. When transferring data, the DMTS opens multiple connections, each controlled by a dedicated worker thread, to maximise the use of the available bandwidth (we assume

ISBN: 978-3-95450-209-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2019-MOMPR007

that facilities running DAaaS employ high bandwidth, high performance wide area networks). To preserve data confidentiality, the connections use secure sockets (but could be configured as normal sockets).

The design of the DMTS utilises data transfer adapters. We have created several adapters to implement different data transfer strategies. For example, one adapter transfers just the contents of the current directory (there is no point in transferring files from other directories if a researcher is not using them); another adapter transfers small files before transferring larger files. These techniques help ensure that the DMS is fast, efficient and is perceived as responsive by end-users.

The DMS operates as a part of the DAaaS infrastructure: it interacts with other DAaaS services and will accept requests only from authorized services. To this end the DMS verifies the security certificate of a requesting service against a known certificate authority (in our case STFC). Such a de-centralized security model facilitates horizontal scaling.

SOFTWARE ENGINEERING CHALLENGES

Security

It is important that DAaaS infrastructure on external clouds, especially public clouds, be secure to ensure the confidentiality of research data. A uniform model across all clouds is essential to ensure a consistent approach to managing security. Our design uses a bastion machine, automatically provisioned using Ansible, to ensure minimal access to the core DAaaS infrastructure through the use of certificates and IP restrictions.

Consistent Deployment

The analysis of a dataset might be spread over several sessions and there is no guarantee that a virtual machine will be running on the same physical hardware in any two sessions. Consistent creation and deployment of images is critical to ensure a consistent experience between sessions. We chose to use Packer [17] to automate the creation and configuration of virtual machine images. Packer ensures a codified infrastructure, ensures scripts are always executed in the correct sequence. This degree of control over the configuration of virtual machines enables analysis environments to be versioned.

Dealing with Scalability & Variation in Demand

Demand by scientists for compute resources is highly unpredictable. The DAaaS infrastructure need to be able to scale and respond to variations in demand. The number of virtual machines in pools manged by the VMM is configurable and, each time a virtual machine in a pool is allocated, the VMM will automatically provision a new virtual machine (up to a configurable maximum limit), so that it always has a reserve of available virtual machines to allocate. The configurable pool size also allows support teams to manage planned downtime - the pool size on other clouds can be increased to compensate for the temporary loss of resource. The VMM also monitors allocated virtual machines, to check if any have been abandoned (i.e. inactive for prolonged periods - the period is, of course, configurable). It will reclaim any abandoned virtual machines. And, finally, the VMM will also monitor for clouds experiencing communication or performance issues) and will attempt to spread the load to other clouds.

OPERATIONAL BENEFITS

The design of the DAaaS provides several operational benefits.

Response

DAaaS is responsive. From the perspective of an end-user, the response time is essentially the time required to boot a new virtual machine. The DMS transfers data from archive storage to the virtual machine only as required. There is no delay waiting for a large dataset to be transferred.

System Administration

In DAaaS, cloud-specific knowledge is encapsulated in the VMM and Ansible. We can manage different clouds and scales via Ansible's dynamic inventory capabilities – in essence, a single set of playbooks is all that is required to provision multiple clouds. This greatly simplifies the administration of the system, enabling ALC's DAaaS support team to spend a greater proportion of their time creating innovative developments and enhancements to DAaaS.

Monitoring & Analysis

The VMM monitors the levels of demand in the system. It will auto-start virtual machines in response to demand, it will cull abandoned virtual machines. The VMM logs the usage of individual virtual machines types and the overall load on the infrastructure. The log files can be analysed by the system administrators to determine how resources are being used and the results of the analysis used to optimize the future use of those resources.

Security

The uniform security model abstracts cloud-specific differences from system administrators. This reduces the scope for security breaches caused by error and misconfiguration.

CONCLUSION

We have created three key infrastructure components which together provide the foundation to support scalable, high-demand analytics environments across heterogeneous clouds.

Our external provisioning service ensures that we can extend the DAaaS infrastructure to utilise external cloud resources, enabling DAaaS to exploit additional capacity or specialist resources as required. External clouds can be made available through collaborative ventures, such as IRIS, or they can be commercial clouds, for example, Amazon Web Services or Azure.

ISBN: 978-3-95450-209-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2019-MOMPR007

The VMM provides a means to manage pools of virtual machines across heterogeneous clouds in a manner that is transparent to the end-user. The VMM is entirely agnostic to the content of the virtual machines used to perform data analysis, allowing DAaaS administrators to tailor the number and composition of virtual machine pools to respond to changing demand from researchers and to best suit available compute resources.

The DMS operates in tandem with the VMM to provide fast, efficient transfer of data from storage to the data analysis virtual machines. Using a 3rd-party transfer mechanism, the DMS makes maximum use of available bandwidth and flexible adapters to ensure data is quickly mounted on data analysis virtual machines and made available to scientists.

ACKNOWLEDGEMENTS

We acknowledge funding from UK Research and Innovation -STF (UK SBS IT 18160).

REFERENCES

[1] OpenStack, https://www.openstack.org/.

[2] IRIS, https://www.iris.ac.uk/.

[3] Ansible, https://www.ansible.com/.

[4] AWX, https://www.ansible.com/products/awx-project/.

[5] VPC, https://en.wikipedia.org/wiki/Virtual_private_cloud.

[6] NAT, https://en.wikipedia.org/wiki/Network_address_translation.

[7] Internet gateway, https://en.wikipedia.org/wiki/Gateway_(telecommunications)

[8] OpenVPN, https://openvpn.net/.

[9] Squid, http://www.squid-cache.org/.

[10] CVMFS, https://cernvm.cern.ch/portal/filesystem

[11] Bastion, https://en.wikipedia.org/wiki/Bastion_host

[12] Amazon Web Services, https://aws.amazon.com/.

[13] Azure, https://azure.microsoft.com/.

[14] O. Arnold *et al.*, "Mantid—Data analysis and visualization package for neutron scattering and μSR experiments", *Methods in Physics Research Section A.*, vol. 764, pp. 156-166. doi.org/10.1016/j.nima.2014.07.029

[15] libcloud, https://libcloud.apache.org/.

[16] FUSE, https://github.com/libfuse/libfuse

[17] Packer, https://www.packer.io/.