University of Colorado **Boulder**

Analysis of Research Cloud Computing Requirements at CU Boulder
In support of
Research Cloud Computing for the CU Boulder Campus
Fall 2017

1. Charge

In September 2016, Vice Chancellor for Research & Innovation Terri Fiez convened a group of subject matter experts from around the CU Boulder campus to look at how cloud computing might be included in the campus research computing structure. The charge to the group included:

(1) Create a plan for a cloud computing service model and cloud deployment model in support of research, especially research incorporating big data analytics. The plan will include requirements gathering to support the launching of any technical directions;
(2) Identify a service and deployment model in support of instruction and training;
(3) Generate a plan for an organization and staffing that provides open and equitable access to interested users;
(4) Explore a sustainable financial model for the operation and support of cloud resources; and
(5) Design guidelines and policies that are amenable to the adoption of cloud computing for researchers at CU Boulder in the future.

An initial report outlining a vision for research and education cloud computing, definition of success for the effort, results of a benchmarking exercise, identification of possible organizational structures, and recommendations for next steps was submitted in November, 2016. In a subsequent report, we summarized a picture of current computing workflows, support used, and computing needs, based on survey and focus groups of CU Boulder researchers and research IT support personnel.

In this report, we developed several areas, including:

- **Infrastructure requirements**, including key partners, data security, legal and contractual processes, networks, data management and accessibility, and tools.
- **Evaluation of Commercial Cloud providers** matched against institutional requirements
- **Support needs**, including classes, training, networks.
- **Identification** of a potential structure that would meet institutional needs in a cost-effective manner.

In particular, we focused on ways to provide new services and infrastructure that would meet the largest reasonable cross-section of researchers' cloud computing and storage needs as identified from the previous report.

Although this report is focused on cloud computing within the campus research computing structure, there are important similarities and synergies between this and the cloud computing

needs of the administrative and teaching & learning sides. This is particularly true in the areas of infrastructure and business applications.

2. Working Group Members

Ken Anderson, Computer Science
Jim Dykes, IBS
Orrie Gartner, OIT
Dirk Grunwald, Computer Science
Sangtae Ha, Computer Science
David Hamrick, OIT
Thomas Hauser, OIT/RC
Brian Johnson, NSIDC
David Kohnke, Leeds
Michael Paul, Information Science
Larry Levine, OIT
Ben Shapiro, ATLAS/Comp. Science
Kurt Maute, AES
Doug Smith, CEAS
Joe McManus, ITP

Supported by: Ligea Ferraro, OIT
Facilitator: Emily CoBabe-Ammann, RIO

3. Overview of findings and recommendations

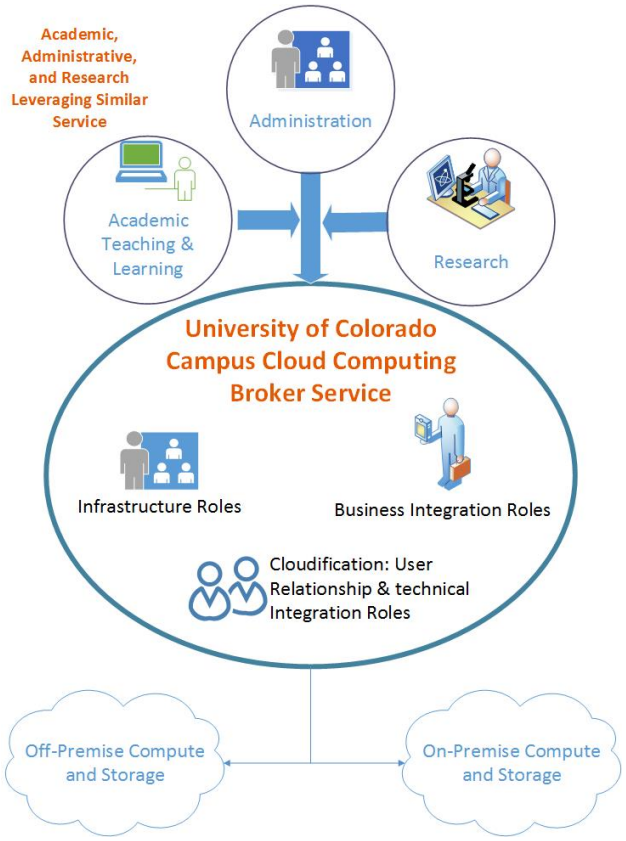


Figure 1. *Schematic of proposed campus-wide "Cloud Computing Broker Service"*

The research computing snapshot conducted earlier in 2017 showed significant demand for cloud computing and storage, beyond existing on-premise options. Given the ever-increasing sizes of datasets, the need to synthesize and integrate more and more heterogeneous data and the increasing numbers of disciplines turning to data to support their research efforts, the demand by researchers for increasing computing and storage capabilities is expected to continue to escalate at a rapid rate, outpacing the current campus capabilities. We need to explore multiple on- and off-premise options to meet campus demand, particularly solutions that provide researchers with the necessary capabilities and flexibility to meet their research needs. As part of the current effort, we explored several possibilities for integrating new cloud services into the portfolio offered by OIT and the Research Computing (RC) group.

One option would be to build a new on-premise elastic compute environment, with an associated storage facility, for the use of CU Boulder researchers. However, since several external companies have well-developed cloud computing offerings that are becoming a de-facto standard, trying to replicate that environment locally appears to be an expensive re-invention of the wheel and thus this option was abandoned.

The existing on-premise computing and storage infrastructure that supports CU Boulder research overlaps with and is in some cases tightly linked to the education and business computing infrastructure of the campus. The same is true when looking at cloud computing. While each of these areas do have unique cloud computing needs (outlined in this paper), CU Boulder is in a strong position to leverage across the campus to provide both on- and off-premise cloud computing services. Indeed, the campus would benefit from a holistic non-duplicative effort and result in overall cost savings. **We thus propose the creation of a campus-wide "Cloud Computing Broker Service" that supports the needs of researchers, and also provides infrastructure to meet the cloud-related requirements of other campus entities (see Figure 1).**

The CU Boulder Cloud Computing Broker Service is designed to augment and improve existing OIT On-Premise Cloud and RC Compute and Storage services, which will continue to be

offered. The roles required for the Cloud Broker Service would be integrated with several existing teams within Research Computing, OIT, OCG, and other campus departments.

For research computing, the leap to cloud computing will require some new investments, both for staffing requirements and infrastructure – though some initial staffing needs can somewhat be met through refactoring, retraining and modifying existing roles. That said, the rapidly increasing computing needs of researchers at CU Boulder will require new investments, regardless of approach. In particular, the research computing snapshot concluded in spring of 2017 made it clear that cloud computing does provide better and more flexible service for a considerable portion of the research community and that more support for cloud computing is needed.

We recommend that CU Boulder focus on implementing a contract agreement with Amazon Web Services (AWS) as its initial off-premise commercial cloud computing provider. This choice is based on:

- analysis of the capabilities and costs of commercial Infrastructure as a Service (IaaS) (see Section 4),
- the results from earlier focus groups involving CU Boulder researchers,
- the existence of an increasing number of relevant publicly-accessible data sets within the Amazon environment, and
- the current needs of the Office of Data Analytics, plus the availability of OIT/ODA's recently-funded AWS-focused cloud engineer position which can also lay the groundwork for the proposed Cloud Broker Service.

The cloud broker service would be extended to include other commercial cloud providers as needed in a later phase of CU's cloud service rollout.

Our recommendations do not take the place of the current efforts and resources expended by Research Computing for compute and storage or OIT's "private cloud". Rather, it is clear that the needs of computing by researchers on campus is growing rapidly. In order to support this growing need, CU Boulder will have to anticipate increased spending to support research computing across the spectrum. In addition, it is clear that a significant cross-section of CU Boulder researchers is better served by working in the cloud, and many have already moved into off-prem cloud computing and storage, paying for those cycles themselves. We are suggesting that now is an opportune moment for CU Boulder to move to add additional cloud support and services to meet this growing need.

4. Capabilities, support and costing of major commercial cloud services

As a first step of this work, we assessed the 3 primary off-premise cloud providers (Amazon Web Services, Microsoft Azure, and Google Cloud Platform (GCP)) for how they might meet CU Boulder research computing needs. A complete assessment of these three entities is found in Appendix I. Unsurprisingly, each of the primary off-premise cloud providers have their own strengths and capabilities along with shortcomings. For example, Google Cloud Platform provides significant flexibility defining network topology and provides the ability to create an

interregional private network that connects all instances across different Google cloud regions. However, it does not have capability to allow customers to metadata tag or group assets and correlate bills against those tags which is truly problematic for large deployments where bills need to be segregated based on user or department. Here we present a summary of those findings.

4A. Capabilities

To meet the needs of researchers, we evaluated cloud providers based on their ability to provide the following benefits and capabilities, following the guidelines provided by the Gartner Group's evaluation of Infrastructure as a Service (IaaS) (2017):

- Scalable and elastic compute
- Fractional consumption
- Eliminates need to purchase hardware for "peak" consumption
- Improves speed of delivery
- Allows for innovation – quick successes and quick fail trials
- Shortened time to market
- Improves service levels
- Improves access to data

AWS, Microsoft Azure and Google Cloud Platform, the three leaders in this space, were each evaluated across their compute, storage, network, security, service offerings, support, management and financial aspects leveraging Gartner Group's analysis and specialists (full analysis can be found in Appendix I). The evaluation grouped these various criteria into three categories: required, preferred and optional. The categories of each are as follows:

Required: Must have features needed to develop, deploy and manage a broad range of use cases including production applications in a cloud IaaS environment. Includes items such as rapid virtual server self provisioning, data encryption, scalable storage, customer defined LAN topology and detailed billing

Preferred: Features not necessary to meet the minimum requirements of a typical large organization but frequently desired to address specific needs such as larger scale, better management and improved availability. Includes items such as single-tenant VMs, tiered block storage, real-time network performance visibility, traffic encryption and billing alert notifications.

Optional: Features necessary for specific deployment scenarios but not needed in all deployments. Includes items such as bare metal provisioning, internet-accessible file shares, dedicated hardware security module (HSM) per customer and auction-priced tier offerings.

Each cloud provider's service offerings were ranked based on the criteria from each of the above categories (see Figure 2). AWS, the market share leader in the IaaS space, met 92% of Gartner's required criteria and scored the highest of all three providers. Strengths of AWS include network offerings, "up the stack" services and large-scale capacity and scalability while weakness were

few but included load balancing limitations and lack of published SLAs for all services. Azure was a strong second place contender with strengths in integration with Microsoft technologies, including O365, identity and access management and premier support (a support service CU Boulder currently purchases). Weaknesses included support for non-Microsoft technology and less availability options than AWS. Google Cloud Platform came in a distant third behind AWS and Azure with strong network offerings and configurations but lacking detailed billing capabilities, a crucial need for a campus cloud service.

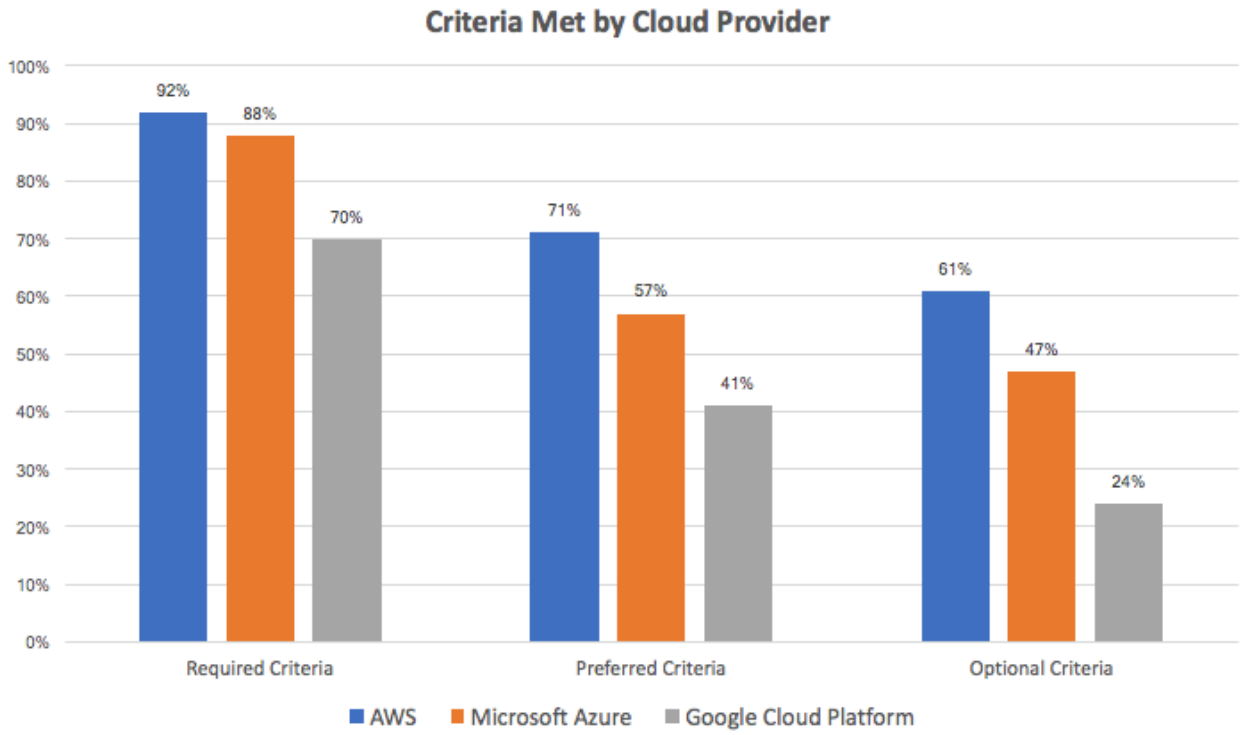


Figure 2. *Summarized ranking of select cloud providers' service offerings*

4B. Support

The University of Colorado Boulder should provide IT broker services to the off-premise cloud, providing continuity and stability for the campus. From a support perspective, the roles and responsibilities for the cloud provider and the organization are generally elucidated along the lines represented in Figure 3. The cloud provider supports the cloud

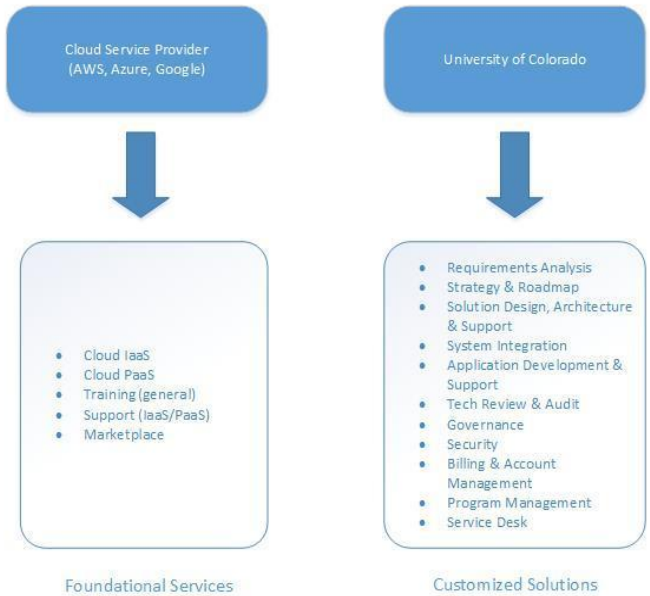


Figure 3. *Roles and responsibilities for the cloud provider and the organization*

IaaS and Platform as a Service (PaaS), as well as general training and support. The campus provides governance, security, account management and a wide range of support. This allows campus researchers to focus on their core research and have a single point to interface with when leveraging cloud computing. The Campus Cloud Computing Broker Services illustrated in Figure 4 shows where the critical cloud broker service interfaces with campus and various on and off premise providers.

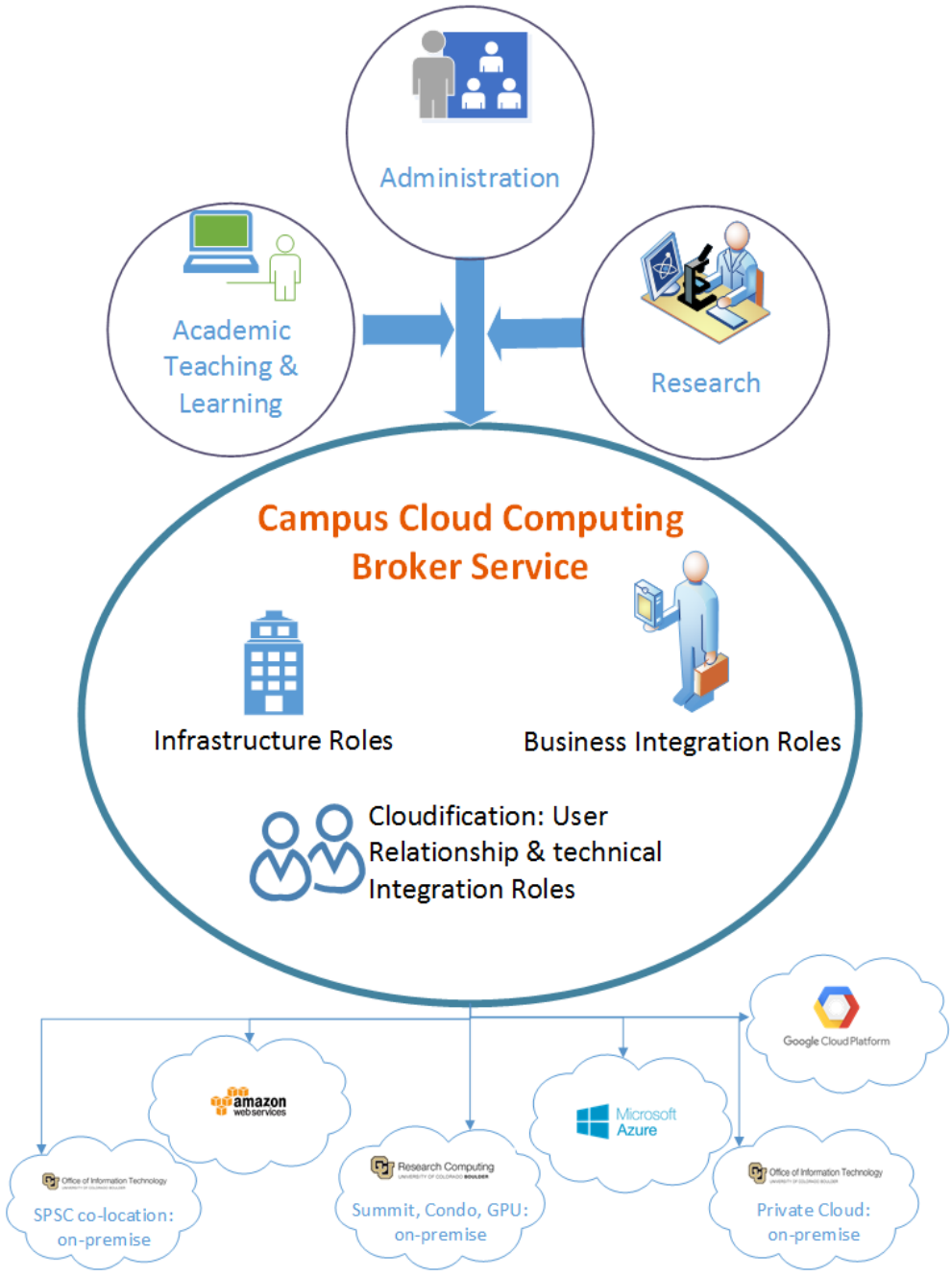


Figure 4. *Critical cloud broker service interfaces*

The majority of end user research support would be designed to fall upon the CU Boulder cloud computing broker service rather than the on- or off- premise cloud vendors. The CU Boulder cloud broker service would of course leverage external support from off-premise vendors as required. Based on the Gartner Group's in-depth assessment reports of each of the three major cloud providers, AWS and Microsoft Azure rate higher in support than GCP. GCP lacks centralized billing options, cloud offboarding support and is not a member of TSANet, an organization for empowering collaboration between multi vendors.

Support for existing on-premise options (OIT Cloud and RC compute and storage resources) is provided by OIT and RC staff. See Appendices II and IIIA for details.

4C. Costing of off-premises commercial cloud

It is difficult to perform a direct comparison of on-premise cloud to off-premise cloud as the service offerings are not always identical. On-premise offerings include the OIT on-premise cloud (also known as the OIT private cloud) as well as Research Computing's Summit supercomputer and large-scale storage offerings via the PetaLibrary service. This section provides an overview of the costs associated with on-premise cloud offerings followed by estimated costs for off-premise commercial cloud.

4C.1. Current costs of purchase, maintenance and support for On-Premise Cloud and Compute/Storage options at CU

CU Boulder researchers have access to several on-campus compute and storage resources, which are briefly described here as a comparison with the commercial options outlined below.

The OIT On-Premise Cloud was designed for small scalable compute needs, particularly web application services, for the academic, administrative and research spaces. Researchers often use this environment for custom web applications that either assist in their research or provide a means to display research results to the web. While the on-premise cloud is a VM environment, it differs from commercial IaaS in that its VMs (including operating system configuration and software stack) are fully managed and supported by OIT staff. Its price for CU Boulder users is based on a hardware cost-recovery model; the support and management staff costs are covered by OIT. The on-premise cloud was not designed for high performance computation as Research Computing's on-premise offerings are intended to meet that need. (More details regarding OIT On-Campus capabilities can be found in Appendix II.)

Research Computing provides large-scale data storage and high-performance computing to any CU Boulder researcher. The Summit supercomputer is capable of running compute jobs that scale to thousands of cores. It differs from AWS, Azure, and GCP options in that those are aimed at single-node workloads (up to about 24 cores). However, AWS, Azure, and GCP allow much greater user control over the VM configuration and virtually unlimited on-demand availability.

The total annual cost to CU Boulder for its portion of the Summit supercomputer is about $1,135,000 including acquisition, data center, and staff costs. Since Summit delivers at least 60,000,000 core-hours/year to CU Boulder researchers, each core-hour costs CU Boulder about $0.02. Note that CU Boulder users are not charged to use Summit. Compare off-prem HPC at $0.08 - $0.15/core-hour or off-prem non-HPC cloud at $0.05 - $0.08/core-hour.

Research Computing provides on-campus data storage space via the PetaLibrary service. The PetaLibrary Active service costs end-users $65/TB/yr, while the PetaLibrary Archive option costs $35/TB/yr. The analogous Amazon storage options, S3 and Glacier, cost at least $275/TB/yr and $48/TB/yr respectively. Data storage on Google Drive and MS OneDrive is currently free with CU's enterprise agreement, but transfer bandwidth and file operation limits apply. (More details about current Research Computing capabilities and cost breakdowns can be found in Appendix III.)

4C.2. Estimated Costs for Off-Premises Commercial Cloud

All three major cloud providers have an OpX cost model with payment made on an as-use basis. Although up-front costs would be required to build the campus infrastructure to provide the IaaS service to campus, these costs would be part of the start-up for the service. Campus customers using the service would leverage off-premise cloud pricing published by the vendor. Some things to note:

- For at least one provider, a potential exists for a percent discount off all services.
- Cost calculators exist for each provider so as to estimate OpX costs though these can be cumbersome to understand. As such, the IaaS service should provide billing and account management support, so users can understand expected costs as part of the IT Service Broker model.
- Based on the Gartner Group's in-depth assessment report of each of the three major providers, AWS and Microsoft Azure meet all price and billing requirements, whereas Google Cloud Platform does not as it lacks tagging and granular billing per user account. This would make it very difficult at this point in time to provide Google Cloud Platform in an IT Service broker model.
- Egress network fees may be applicable depending on the Enterprise Agreement that is reached with each off-premise cloud vendor; however, a main goal of an Enterprise Agreement should be to negotiate a massive discount - if not waiving - of any data egress fees.
- The cost per core-hour for AWS or Azure on-demand VM instances similar to a Summit compute node range from $0.05 - $0.08. Off-premise HPC clusters cost $0.08 - $0.15 per core-hour. See Appendix III.C for a more detailed price comparison between on-premise HPC and off-premise compute options.

5. The Campus Cloud Broker Service

The role of the Campus Cloud Computing Broker Service is to enable researchers, developers and consumers of cloud services to quickly access the technology services of the cloud while

safeguarding the interests of the University through the application of centralized policies and procedures. There are Technical Infrastructure Roles, Business Roles and Cloudification or Cyberinfrastructure Facilitator roles, as detailed in the Figure 5 to provide the support for researchers leveraging cloud services. These requests include the necessary capabilities for the research and development required to effectively support cloud computing on campus.

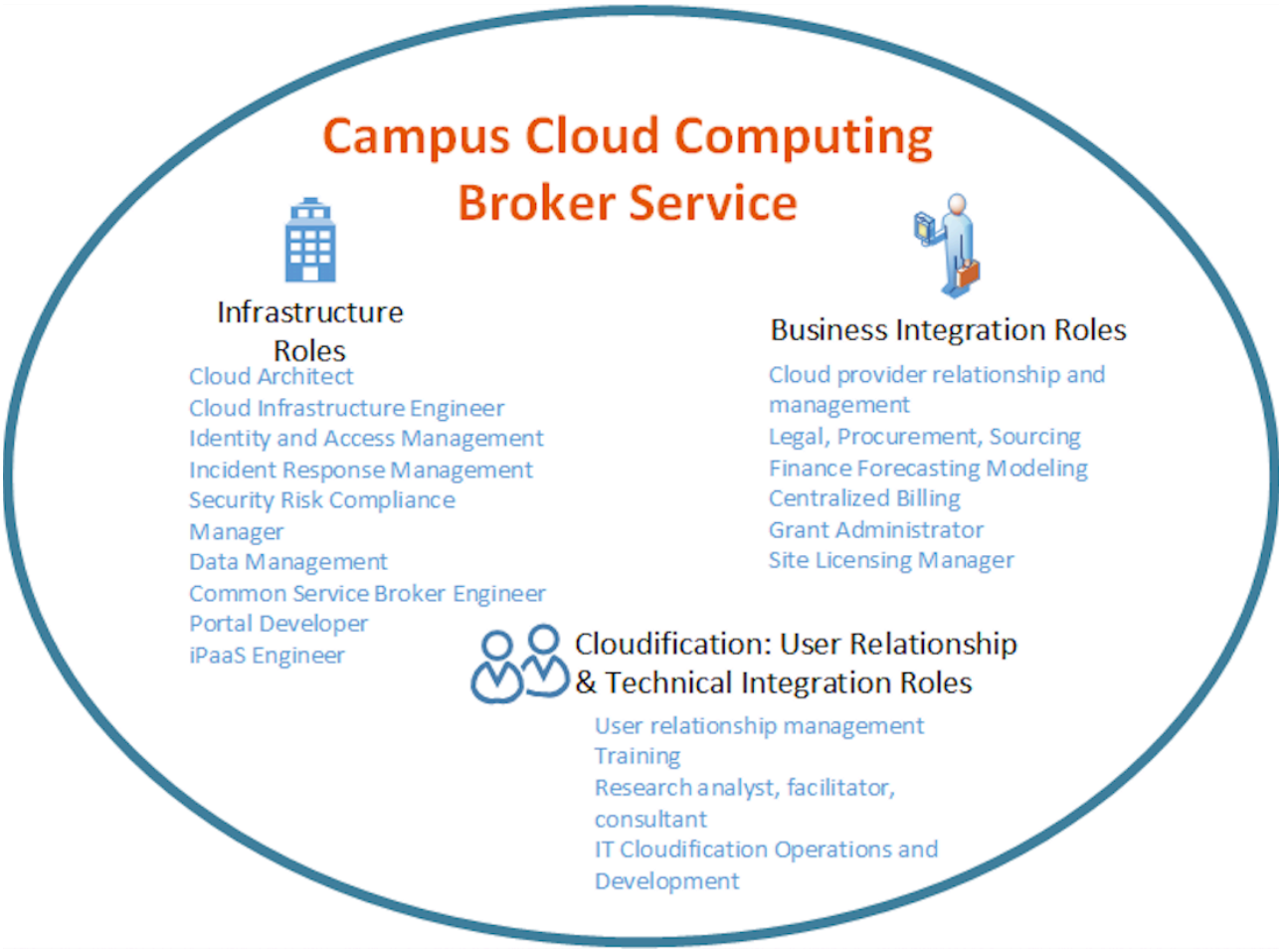


Figure 5. *Various roles supporting researchers leveraging cloud services*

Although the concept of the Campus Cloud computing broker service is new, the organizations providing the various services that combined form the Campus Cloud broker service currently exist. As such, the recommendation is to not create a new Campus Cloud organization but rather broaden existing organizations current service offerings to include services required to support this new offering. By leveraging existing organizations, as opposed to creating a stand-alone Campus Cloud broker unit, dependencies across organizations and teams will be created. These various teams will rely on others to deliver the overall Campus Cloud broker service. For example, Research Computing will rely on OIT Networking to provide services such as DNS,

DHCP and Networking to/from the cloud as opposed to having these resources in-house within the Research Computing organization. We believe having existing organizations focus on their core competencies related to the Cloud Broker service will result in efficiencies in delivery of the service and prevent the duplication of efforts.

Figure 6 below illustrates the existing organizations along with the identified roles within each organization. (Numbers in parentheses indicate number of positions. ***It must be noted that a role does not necessarily mean 1 full time employee (FTE). A role can be a partial or one or more FTE and a single FTE may have multiple roles.***)

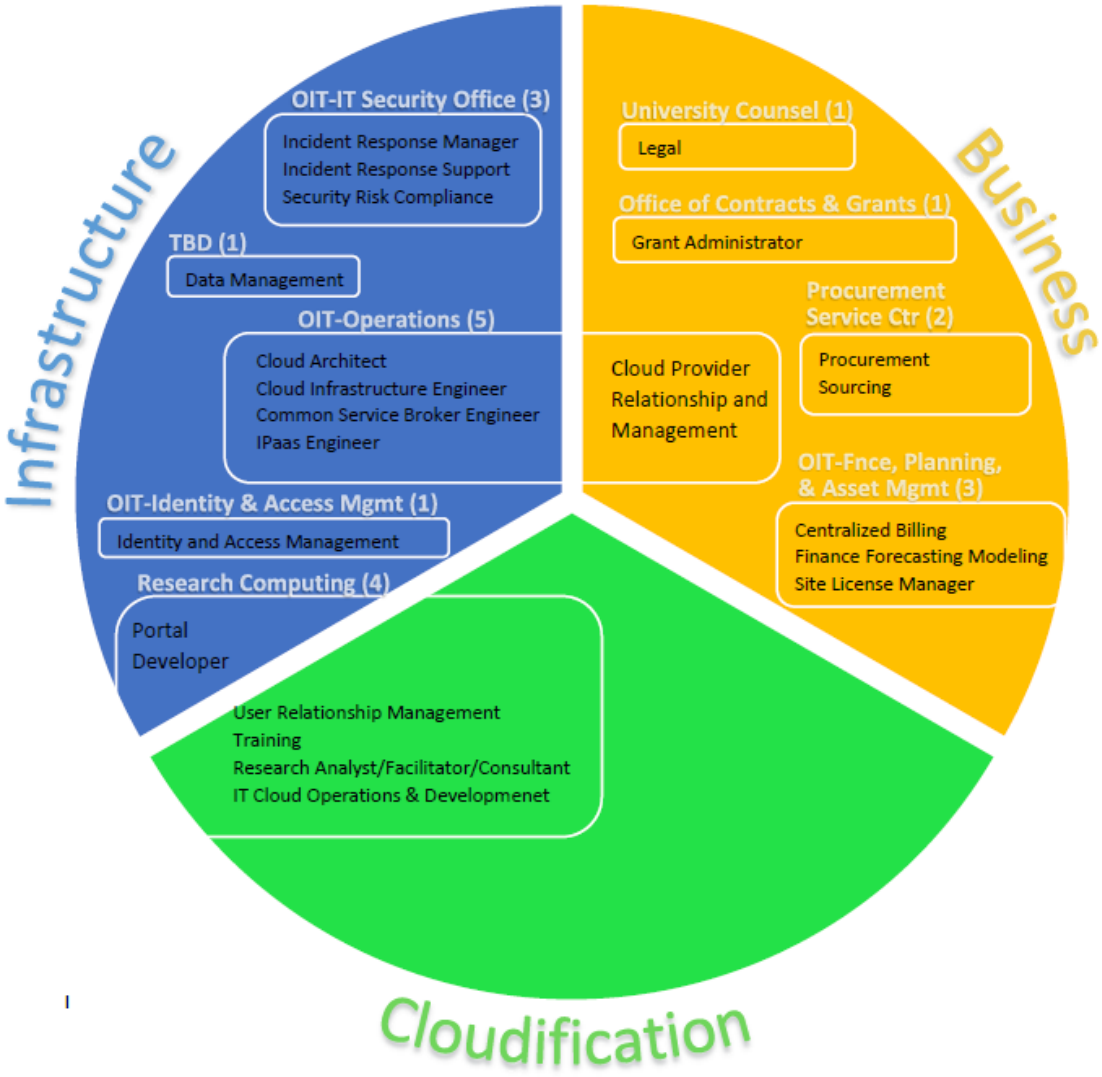


Figure 6. *Existing organizations and identified roles within each organization*

The role of central IT is evolving from just delivering technology to becoming the broker to researchers for all IT-based services including HPC, off-premise cloud and on-premise cloud. The existing organizations including OIT, Research Computing, IT Security Office, Office of Contract and Grants and others found in the above diagram will require additional resources to provide the services to support the Campus Cloud broker service. Software tools, although a large component of delivering the service, are not sufficient to implement a cloud broker service as new functional roles, skill sets, and procurement, finance and governance structures are required. These additional resources fall into the three large categories: infrastructure, business integration and cloudification/cyberinfrastructure roles. Within each category, roles are classified as Must Have's, Should Have's and Could Have's with definitions for each as follows:

Must Have: The service is required at the time of the launch of the campus cloud broker service

Should Have: The service is important but not necessary for delivery on day one and can be implemented within 12 months of launch.

Could Have: Requirements that would improve the user experience or customer satisfaction with the overall service but can be included 12 months or more after launch.

The identification of the specific roles within each of the three categories as well as additional resources required, approximate costs and the location within existing organizations where each should reside are found in the next sections.

Please see Appendix IV for greater detail about the tasks each role would undertake, the skills needed to fill the role, and the research/teaching/business needs that each could fulfill.

5A. Infrastructure Roles

Infrastructure Roles represent many of the IT components a researcher expects to "just exist" and be available to end users without additional cost or effort from an end user's perspective. The foundational elements include items such as network pathways (including DNS/DHCP), security tools and processes, user authentication mechanisms, logging and monitoring, data management and software interfaces. The specific roles needed vary, depending on whether the approach is on-premise or off-premise. Details on the specific infrastructure roles can be found in Appendix IV. A summary is provided in the Figure 7. Section 6 explains the services offered through a phased rollout and which specific infrastructure roles are must haves, should haves and could haves through each phase in order to provide the service.

Key for Figures 7, 8, 9 & Table 1:

IAM	Identity & Access Management	OCG	Office of Contracts & Grants	TBD	To Be Determined
ITSO	IT Security Office	PSC	Procurement Service Center	RC	Research Computing
OIT - OPS	OIT - Operations	OIT-FPAM	OIT Finance Planning & Asset Management (formerly FBO - Finance & Budget Office)	Counsel	University Counsel

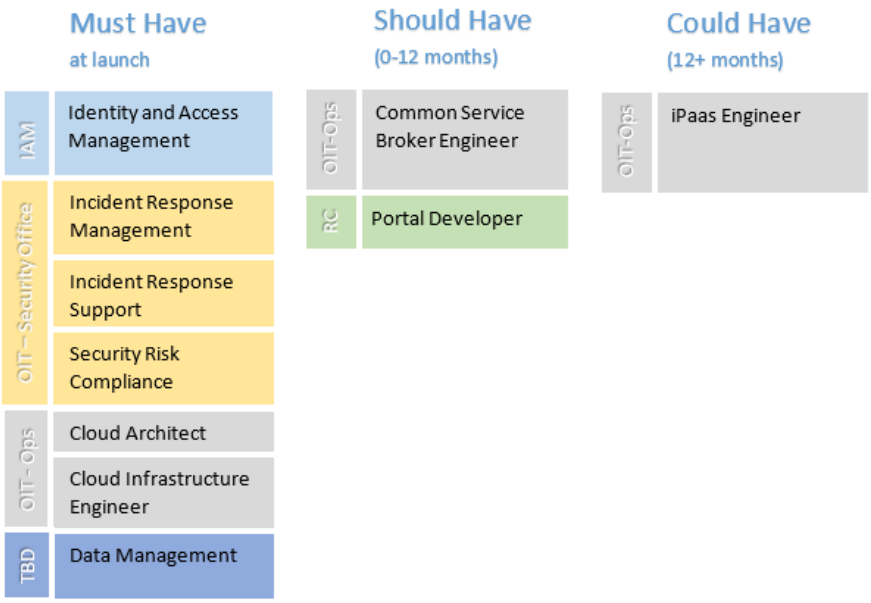


Figure 7. *Schematic illustrating infrastructure roles, priority and timing of need*

5B. Business Integration Roles

Business Integration Roles represent the management and back-office side of the cloud. These functions focus on vendor management, licensing, grants, finance, legal and procurement. The roles needed are the same, regardless of whether the approach is on-premise or off-premise. Details on the specific business integration roles can be found in Appendix IV. A summary is provided in the Figure 8. Section 6 explains the services offered through a phased rollout and which specific business integration roles are must haves, should haves and could haves through each phase in order to provide the service.

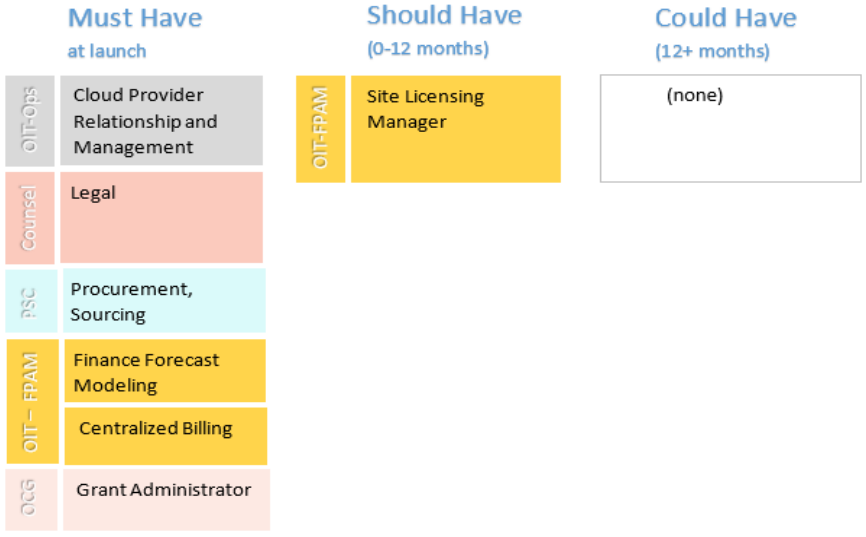


Figure 8. *Schematic illustrating business integration roles, priority and timing of need*

5C. Cloudification/Cyberinfrastructure Facilitator Roles

Cloudification/Cyberinfrastructure Facilitator Roles represent the end user facing aspects of cloud computing. These functions focus on assisting researchers in their move to the cloud and, once migrated, potentially the operational aspects of managing their environment in the cloud. The roles include user relationship management, training, consulting and working alongside researchers to create or migrate existing or new computational workflows into the cloud - a process known as "cloudification" or "cyberinfrastructure facilitation". The roles needed are the same, regardless of whether the approach is on-premise or off-premise. Details on the specific cloudification or cyberinfrastructure facilitator roles needed can be found in Appendix IV. A summary is provided in the Figure 9. Section 6 explains the services offered through a phased rollout and which specific cloudification/cyberinfrastructure facilitator roles are must haves, should haves and could haves through each phase in order to provide the service.

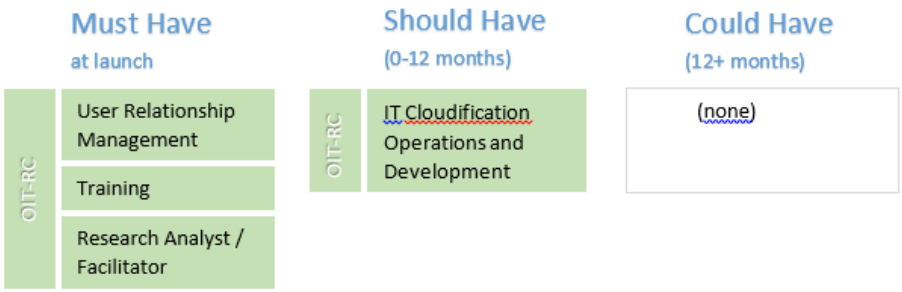


Figure 9. *Schematic illustrating cloudification/cyberinfrastructure facilitator roles, priority and timing of need*

5D. Costs of Cloud Computing Broker Service

The ongoing support costs for the campus IaaS service include costs to provide the services outlined in Figure 5, including software, potential hardware and people. Since the Campus Cloud Computing Broker Service can be leveraged by multiple disciplines across campus (academic, research and administration), re-examining current organizational structures and potential for cost sharing is a potential opportunity.

There are some identified upfront costs required to build out the campus infrastructure to provide an IaaS service to campus, in alignment with the vision and strategy for cloud computing. Table 1 below gives estimated one-time and continuing costs, all new expenditures:

	Description	FY19	FY20	FY21
OIT - Ops	***One-Time costs:*** Consulting for Cloud Infrastructure and Requirements (infrastructure - consulting)	$100,000	N/A	N/A
	Cloud Infrastructure Implementation including SDN (infrastructure - software)	$75,000	$350,000	$350,000
	Hybrid Cloud Infrastructure Integrations (infrastructure – hardware)	$0.00	$130,000	$130,000
	Training for Cloud Infrastructure (infrastructure - training)	$75,000	N/A	N/A
	Continuing costs: Cloud Architect & Infrastructure Engineer (infrastructure - personnel) 1 FTE	$135,000	$135,000	$135,000
	Administration of Networking Components (infrastructure - personnel) 1 FTE	$0.00	$128,000	$128,000
	Ingress/Egress networking (infrastructure - service)	$0.00	TBD	TBD
OIT – Security Office	Incident Response Manager (infrastructure - personnel)	$0.00	TBD	TBD
	Incident Response Support (infrastructure – personnel) student then 1 FTE	$39,500	$74,500	$74,500
	Security Risk Compliance (infrastructure - personnel) 1 FTE	$100,000	$100,000	$100,000
	Logging and Monitoring (infrastructure - software) (Vulnerability Management Licenses)	$2,000	$2,000	$2,000

OIT – Security Office	Logging and Monitoring (infrastructure – software) (LogRhythm SIEM Amazon Instance)	$35,000	$35,000	$35,000
	Network Security (if cannot leverage built-in AWS/Azure/GCP) (infrastructure - software services)	$0.00	$100,000	$100,000
OIT – FPAM	Cloud Billing Technician (business integration - personnel) FTE	$50,000	$50,000	$50,000
	Cloud Site Licensing Manager (business integration – personnel) 1 FTE	$0.00	$82,500	$82,500
	Cloud Billing software (business integration - software)	$50,000	$50,000	$50,000
OIT – Research Computing	Training (cloudification) and Research Analyst/Facilitator/Consultant (cloudification)	combined w/RC analyst	combined w/RC analyst	combined w/RC analyst
	Research analyst/facilitator/consultant (cloudification) 1 FTE	$150,000	$150,000	$150,0000
	IT Cloudification Operations and Development (cloudification) 1 FTE	$150,000	$150,000	$150,000
	IT Cloudification Operations and Development #2 (cloudification gateway developer) 1 FTE	$130,000	$130,000	$130,000
TBD	Data Management	TBD	TBD	TBD
OIT-IAM	Identity and Access Management - Integration & management of Access Solutions (infrastructure)	absorb in existing org	absorb in existing org	absorb in existing org
Counsel	Legal	absorb in existing org	absorb in existing org	absorb in existing org

PSC	Procurement Sourcing	absorb in existing org	absorb in existing org	absorb in existing org
OCG	Grant Administrator	absorb in existing org	absorb in existing org	absorb in existing org
	Total FTE Salary ()	**$639,500**	**$865,000**	**$865,000**
	One-time costs	**$250,000**	**$480,000**	**$480,000**
	Ongoing costs	**$87,000**	**$667,000**	**$667,000**

* Position funded by Kelly Fox effective October 2017 as part of ODA/AWS effort so salary not included in total
NOTE: All salary numbers do not include benefits (currently 37.6%)

Table 1. *Estimated one-time and continuing costs, all new expenditures*

It should be noted that these are based on various assumptions, including the ability to leverage existing campus personnel in existing organizations including OIT. Furthermore, the assumption of above costs is for a single cloud provider. A multi-cloud provider IaaS service offering would likely increase certain costs above. Ingress/egress networking fees are dependent upon specific cloud vendors and may or may not be applicable depending on the contract pathway agreements. A full in-depth cost model should be performed to gain further accuracy in costs.

Currently, users procuring off-premise cloud services use click-through agreements which are offered by all three major off-premise cloud providers. However, it would be in the best interest of the University of Colorado to establish individual Enterprise Agreements with each of these providers, before using their IaaS offering, as it would likely improve the service and terms the University would be subject to and address ownership of data and intellectual property issues not met in the standard end-user click-through agreements. The level of effort to establish an agreement will vary based on vendor (AWS, Microsoft, Google) as the process will be unique to each based on existing contract relationships and procurement. It is estimated that completely new agreements could take up to 6 months based on the experience of our peer institutions.

6. How a Rollout for Cloud might Look

The rollout for the cloud can be performed in specific phases so as to begin providing value to campus researchers as soon as possible. A proposed phased rollout could be broken down into seven phases. The estimated timeframe for phases one through three is 6 months. The timing for

phases four and beyond is fully dependent upon funding and prioritization and could range between 6 and 18 months.

Figure 10 below illustrates potential timeframes for each phase along with the roles and services provided at each phase.

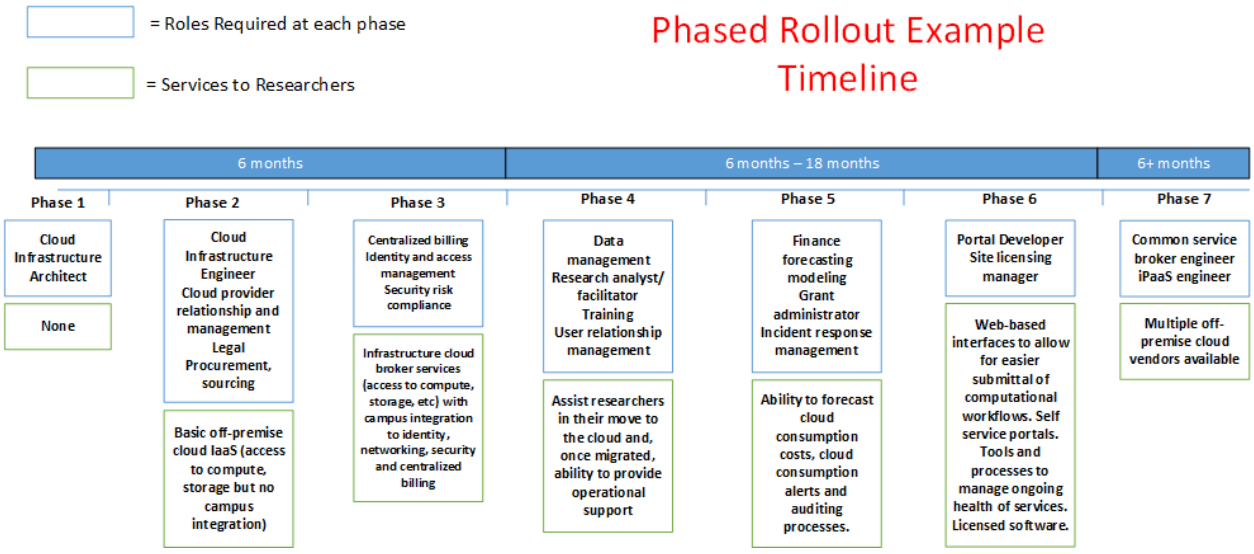


Figure 10. *Potential timelines by phase*

Each of the phases are described further, see Appendix V.

7. Summary

CU Boulder has reached a critical juncture, where its researchers have needs that can only be met by having access to both HPC and cloud computing opportunities, support and services. At the rate at which research computing needs are growing on campus, we suggest that now is the right time to have the next stage of growth include the option of off-premise cloud services.

Our analyses suggest that an expansion into the off-premise cloud, with the corresponding operational and support infrastructure, is the clearest path to providing researchers with the necessary capabilities and flexibility their expanding data needs require. Of the off-premise cloud providers we evaluated, AWS offers the best fit for our campus needs, though other providers may be brought on board at a later date.

The move into off-premise cloud, along with the associated investments, aligns research computing with other administrative units on campus, allowing us to take advantage of synergies and developing a comprehensive approach for the campus.

Appendix I. Summary of Strengths and Weaknesses of Off-Premise Cloud Providers

A. AWS

Figure 11 below shows the AWS feature snapshot based on the Gartner Group's criteria.

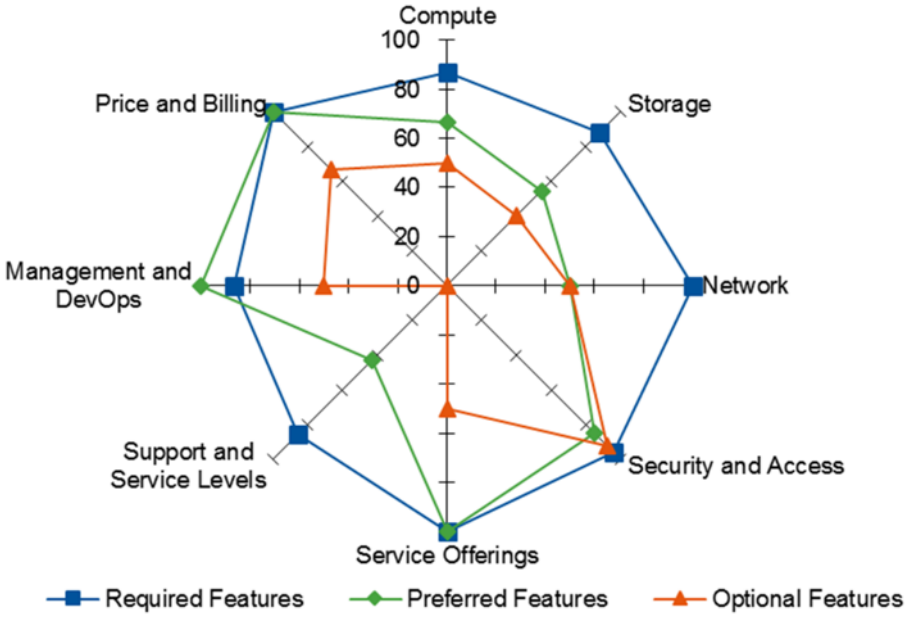


Figure 11. *AWS feature snapshot*

AWS strengths include:

- Network offerings and configurations: Great control and flexibility in defining network topology and private network connections. Robust load-balancing options for both external and internal facing applications.
- Security and access: Solid security foundations, including documentation, customer controlled firewalls, access lists, security groups, and comprehensive compliance certifications and reports. Broad role-based authorization controls for all services through AWS' Identity and Access Management (IAM)
- Availability options: Multiple availability zones (AZs) within its regions effectively providing multiple data centers in close proximity to one another making it easier to run applications across multiple AZs. Customer is responsible for architecting their application for high availability (HA).
- Management control and DevOps enablement: Includes comprehensive offerings through management console and API access. Customizable monitoring and alerting through CloudWatch.

- "Up the stack" features: PaaS-like services including Amazon Relational Database Service (Amazon RDS), NoSQL, real-time streaming and processing via Amazon Kinesis, DNS via Amazon Route 53, and many other services.
- Large-scale capacity and scalability offerings: Focused on building and delivering all services at large scale. Provides competitive high-performance computing (HPC) offerings under select conditions.
- Global geographic footprint: Provides cloud services across the globe.
- Financial management, analysis and billing flexibility: Offers industry's most robust set of provider-offered financial management options including TCO calculators.
- Broad ecosystem of partners and service catalog options: Service is extremely popular so the ecosystem is large and partner offerings are numerous.
- Many off-premise datasets are hosted in AWS, meaning that computational work in EC2 that uses that data already has it available "locally"
- All major Linux distributions are supported.
- Excellent support for containers, especially Docker.

AWS weaknesses include:

- Service levels: AWS meets 86% of Gartner's required criteria for support and service levels. Although AWS support plans offer plenty of flexibility, AWS's SLA is deficient in two ways:
    - Missing single-instance availability SLA
    - Does not offer an availability SLA that has a minimum allowed outage time less than or equal to 30 minutes in a month for its Simple Storage Service (S3)
- Expandable block storage volumes: Does not support increasing the size of Elastic Block Store (EBS) volumes on the fly. AWS does provide a facility to expand the storage space on an EBS volume by migrating data to the larger volume and then extending the file system on the OS to recognize the newly available space. After verifying the new volume is working, the old volume may be deleted.
- Dynamic vertical auto-scaling: Does not support dynamically resizing an instances resources (CPU, memory) when instance comes under heavy load. Can dynamically provision new instances when load decreases. Auto-scaling cannot increase or decrease the resources of an existing instance. Must power off, migrate to new instance size, and then power instance back on.
- Local load balancing – independent Internet Protocol (IP) address: Does not support static or elastic IP address assignment to an Elastic Load Balancer (ELB). AWS assigns a DNS name to an ELB. Some customers may have a requirement to connect to a static IP address instead of a DNS name.

B. Microsoft Azure

Figure 12 below shows the Microsoft Azure feature snapshot based on the Gartner Group's criteria.

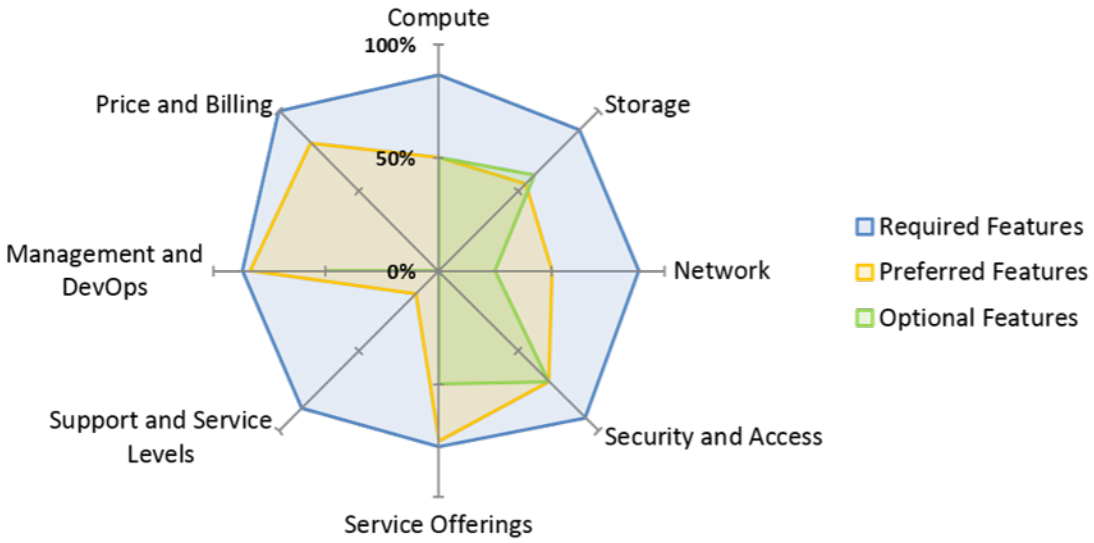


Figure 12. *Microsoft Azure feature snapshot*

Microsoft Azure strengths include:

- Integration with Microsoft technologies: Well integrated into existing Microsoft technologies such as Microsoft System Center for management and Microsoft Operations Manager Suite (OMS) for management of on-premises and cloud management using a common tool. Offers platform as a service (PaaS) for items like Microsoft .NET, SQL Server and BizTalk.
- Identity and access management: Several industry-leading advantages for integration, synchronization and conveyance of Active Directory (AD) domains or forests into Azure for seamless deployments. Appealing for organizations deeply invested in AD for user and account management.
- Price and billing options: Meets 100% of Gartner's required price and billing criteria. Existing enterprise agreement customers will find benefit by including Azure in broader Microsoft enterprise agreements and thereby benefiting from Azure promotional credits and enterprise discounts.
- Global geographic footprint: Provides cloud services across the globe.
- "Up the stack" features: Several value-added services outside of IaaS, including proactive cloud-based analytics, composition and orchestration of data services at scale, advanced key-value cache and store, Apache Hadoop and Windows apps as a service are available.
- Self-service templating: Offers self-service templating for automating deployment of stacks of infrastructure using Azure Resource Manager (ARM).
- Premier support: Organizations that are Microsoft Premier Support customers can take advantage of premier support for Azure services.

- Hybrid cloud capabilities: Through ExpressRoute service, a comprehensive hybrid cloud networking option exists.

Microsoft Azure weaknesses include:

- Availability options: Exposes a single data center to customers in each region. Customers can create WAN connections between regions (known as VNets) or set up cross-region replication, but Azure does not expose multiple data center offerings within a single region to support low latency, synchronous replication requirements.
- Support for non-Microsoft technology: Some notable gaps in supporting non-Microsoft technology exists. For example, Azure's relational database as a service (DBaaS) only supports MS SQL and not other options such as MySQL. Furthermore, some advanced functionality within Azure relies on leveraging PowerShell scripts or the Azure CLI which can be challenging for organizations preferring a comprehensive portal experience.

C. Google Cloud Platform

Figure 13 below shows the Google Cloud Platform (GCP) feature snapshot based on the Gartner Group's criteria.

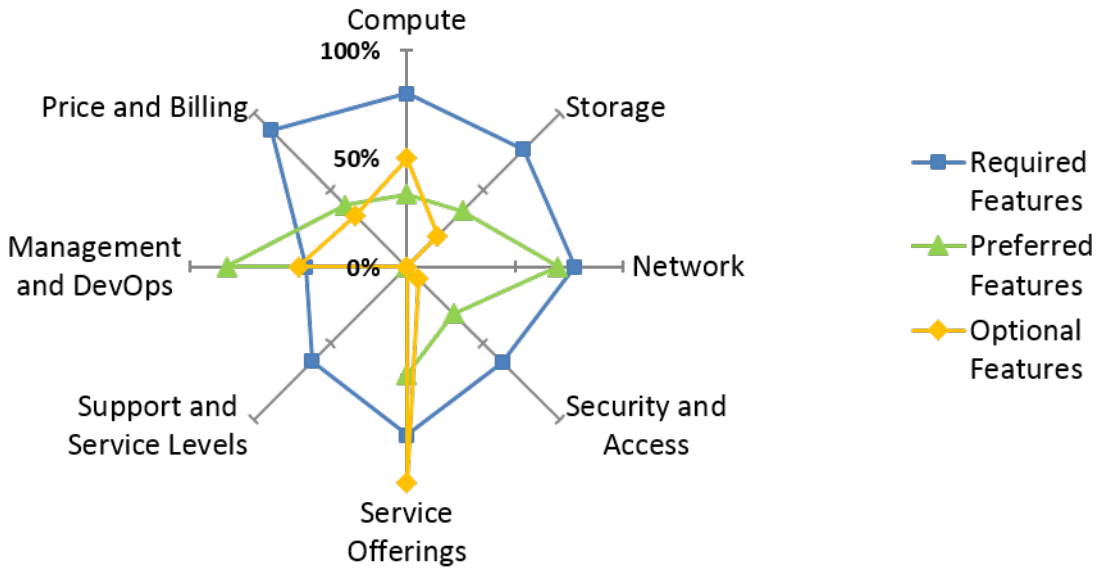


Figure 13. *Google Cloud Platform (GCP) feature snapshot*

GCP strengths include:

- Network offerings and configurations: Offers flexibility defining network topology. Also provides ability to create an interregional private network that connects all instances across different Google cloud regions using Google's internal high-performance global

network. Offers robust load-balancing options for external-facing applications with metrics-driven load balancing built into the service.
- Security and access management: Implemented good security documentation, a variety of compliance certifications, customer controlled firewalls/access control lists (ACLs) and Secure Sockets Layer (SSL) secured endpoints.
- Global geographic footprint: Provides cloud services across the globe.
- Live instance migration: Google Compute Engine (GCE) automatically live migrates customer workloads away from maintenance events so that the customer's applications continue to operate during any scheduled maintenance. The instance may experience a brief period of decreased performance, but it is ideal for workloads that require constant uptime.
- Block storage: Offers some differentiating block storage configurations when compared to its competitors. For example, the maximum block storage volume size is 64TB and one can share volumes across different compute instances with one instance in read/write mode and all others in read-only mode.
- Pricing discounts: Provides discounts on continued service usage, making a simple and attractive pricing model.

GCP weaknesses include:

- Support and service levels: Meets only 62% of Gartner's required criteria for support and service levels. It is deficient in multiple ways including:
    - Notification window for customers to submit notice of a missed SLA is less than two billing cycles
    - Missing a single-instance SLA
    - Does not guarantee 90-day notice for any SLA changes
    - Does not publish an SLA version history
- Management and DevOps: Meets only 47% of Gartner's required criteria for management and DevOps. It is deficient in multiple ways including:
    - Comprehensive labeling and tagging resources, which is required in order to properly allocate individual cloud assets to various initiatives or applications
    - Log management services for various events like account management activities, provisioning catalog actions and security configurations are only in beta status
- No session-affinity support for load balancing: Currently in development.
- No IPv6 support: Does not support IPv6.
- No back-end load balancing capability: Load balancing within GCP can only be performed with public IP addresses.
- Content Delivery Network (CDN): Currently in beta status.
- Granular billing: Does not have capability to allow customers to metadata tag or group assets and correlate bills against those tags. This is problematic for large deployments where bills need to be segregated based on user or department.

Appendix II. The OIT On-Premise Cloud has the following capabilities:

Compute:

- 24 dual-socket (20 real core) compute with 256GB RAM per system

- VMWare vSphere environment
- User-supplied VM appliances supported
- Managed Support Services on all VMs
- Unmanaged VMs available
- 146TB storage available spread across SSD, 10K SAS and 7.2k nearline SAS/SATA

Network :

- shared 40GB between compute and storage
- shared 10-20GB between compute and campus backbone

Support :

- 2 FTE Virtual Infrastructure Administrators
- 7 FTE Linux Administrators providing full support services
- 4 FTE Windows Administrators providing full support services
- 2 FTE Database Administrators providing full support services
- Managed VM Support Services include but not limited to:
    - Operating system installation
    - Operating system upgrades
    - 24/7 system monitoring
    - Security monitoring
    - Patch management
    - Anti-Virus/Malware Scanning
    - Installation of 3rd party applications
    - Software package management
    - Software installation, integration and potential configuration
    - Account management
    - End user tech support
    - Application Support
    - Logging, Monitoring, and Trending Administration

Cost :

- Standard VM: 1 CPU, 2GB RAM, 100GB storage and backups: $90/month
    - Additional CPU: $65 per CPU/month
    - Additional Memory: $2.50 per GB/month
    - Additional Storage: $0.10 per GB/month
    - Additional Backups: $0.10 per GB/month
- Managed and Unmanaged VMs are the same cost to the end user (cost is for infrastructure, not FTE)
- Consulting: Available at the rate of $71.50/hour

Training:

- Training is not available as part of the OIT On-Premise Cloud Service. Users are recommended to work with their vendor who often provides the application training required.

Appendix III: Capabilities, support and costing of on premise RC offerings

A. Capabilities -

Compute (Summit Supercomputer):

- 500 dual-socket (24 real core) compute nodes available at no charge via a queue system to any UCB researcher; minimum of 128 GB RAM/node; 11 nodes have high-end GPUs
- High-speed (100 Gb/s) Omni-Path network to all nodes for extreme multi-node computational performance
- 1.2 PB parallel high-IOPS scratch storage system
- User-supplied VMs not supported, but containers (Singularity/Docker) are supported

Compute (Visualization Cluster):

- 5 dual-socket (24 real core) visualization nodes with high-end GPUs
- Remote visualization software enables high-performance 3D rendering directly to the end-user's desktop or laptop

Compute (Condo Cluster):

- Researchers buy and own individual compute nodes that are aggregated into a single cluster for the use of all condo members
- Node owners get substantial prioritization on their nodes
- Designed mainly for single-node workloads
- Containers (Singularity/Docker) supported

Network (Science Network / DMZ):

- 80 Gb/s ethernet backbone network connecting several data centers across campus
- Minimum of 10 Gb/s to individual departments, RC storage servers, and RC data-transfer nodes
- Dedicated 10 Gb/s network border connection to national research networks

Storage (Core):

- 250 GB free to any UCB researcher with an RC account (snapshots and backups enabled)

Storage (PetaLibrary):

- "Active" and "Archive" options available, depending on access frequency of data
- Redundant storage servers on 10 Gb/s or 40 Gb/s network

- Researchers pay for the incremental cost of media; most data center, storage infrastructure, and staff costs are subsidized
- High-efficiency data transfer available via a Globus endpoint; sftp also supported

B. Support -

User support and consulting:

- 4.5 FTE plus 1-2 student employees
- Full range of expert services from help desk to HPC workflow to data management to application optimization

Training:

- Over 50 training sessions of various types per year, including hour-long tutorials, new-user seminars, and multi-day workshops
- Topics include a variety of HPC and CI topics, including programming, visualization, batch/queue, and data management/transfer

C. Cost to users -

Compute:

- Summit supercomputer - no direct charge to researchers; core-hours are allocated based on an application process
- Condo cluster - researchers pay for their compute nodes; network and support are fully subsidized

Storage:

- Core storage - no direct charge to researchers
- PetaLibrary - Basic "Active" storage (all on-disk), $65/TB/yr ; Basic "Archive" storage (hierarchical storage disk/tape), $35/TB/yr.

Support/training:

- Consulting is free to all users
- Most training is free, although multiday sessions sometimes cost around $50/person

The acquisition cost to CU Boulder for its portion of the Summit supercomputer, which has an expected lifetime of 5 years, was $870K in matching funds to meet the requirements of the NSF grant (which contributed $2.03M). The matching funds consisted of a combination of cash and the staff time needed to install and configure Summit. Acquisition and other costs to provide Summit are summarized in Table 2 below.

Category	**Total cost**	**Years provided**	**Cost/year**
Summit matching funds	$870,000	5	$175,000
Datacenter maintenance (incl power and cooling)	$200,000	1	$200,000
Uninterruptible Power Supply	$600,000	10	$60,000
High-Performance Computing Facility	$4,000,000	20	$200,000
Staff (system administration and user support)	$550,000	1	$550,000
Total			**$1,135,000**

Table 2. *Acquisition and other costs to provide Summit*

Since Summit is expected to provide about 60,000,000 core-hours/yr to CU researchers, the resulting cost per core-hour is about $0.02.

Cost to users (commercial cloud options) -"HPC" (including high-speed interconnect)

Advanced Clustering Technologies http://www.actnowhpc.com/pricing/

$0.10 per core hour

Microsoft Azure HPC https://azure.microsoft.com/en-us/pricing/details/virtual-machines/linux/

$0.11 per core hour

Nimbix https://www.nimbix.net/nimbix-cloud-demand-pricing/

$0.15 per core hour

Penguin Computing https://pod.penguincomputing.com/pricing

$0.08 per core hour

R Systems http://rsystemsinc.com/pricing/

$0.09 per core hour

Rescale http://www.rescale.com/pricing/

$0.08 per core hour

Sabalcore http://www.sabalcore.com/services/pricing/

$0.12 per core hour

"Elastic compute"

- Amazon EC2 and Azure offer literally dozens of different node configurations, all at different base costs. For comparison with existing CU Boulder on-prem HPC, we chose "on-demand" node instances that most closely match Summit nodes. Note that these do not include high-performance internode networking. (Source: https://aws.amazon.com/ec2/pricing/on-demand/ and https://azure.microsoft.com/en-us/pricing/details/virtual-machines/linux/)
- Amazon EC2 "r3.4xlarge" (16 cores, 122 GB memory, 320 GB SSD) : $1.33 / node-hr = $0.08 / core-hr
- Amazon EC2 "c4.4xlarge" (16 cores, 30 GB memory, no SSD) : $0.80 / node-hr = $0.05 / core-hr
- Azure "D14v2" (16 cores, 112 GB memory, 800 GB SSD) : $1.20 / node-hr = $0.075 / core-hr
- Azure "F16" (16 cores, 32 GB memory, no SSD) : $0.80 / node-hr = $0.05 / core-hr
- Note that EC2 and Azure frequently offer much cheaper "spot pricing". During times of very low demand, VM instances can be available for less than 50% of the "on-demand" price. However, if demand rises, a spot instance can be terminated with little warning.

Storage

- Amazon S3 (analogous to RC PetaLibrary "Active") : $276/TB/yr
  - Egress (retrieval) out of Amazon to Internet : $10/TB
  - Access from and transfer to other parts of AWS : price varies

- Amazon Glacier (analogous to RC PetaLibrary "Archive") : $48/TB/yr
    - Egress (retrieval) out of Amazon to Internet : $90/TB
    - Access from and transfer to other parts of AWS : price varies
- Microsoft OneDrive : free with current CU Boulder Enterprise Agreement
- Google Drive : free with current CU Boulder Enterprise Agreement (bandwidth and file operation limits apply)
- (*Note that egress fees are often waived for Academic customers with Enterprise Agreements*)

Appendix IV. The Campus Cloud Broker Service Rolls Further Defined

With the Campus Cloud Broker Service roles identified, here we strive to further detail each role providing: 1) a description of the role, 2) its necessity, and 3) a potential location within the University IT organizations. ***It is important to note that a role does not necessarily translate to 1 FTE. An FTE may have multiple roles or a role may require multiple FTEs.*** [NOTE: It is not in-scope of this report at this time to identify if additional resources are required to provide these services. This will be determined through the cost modeling phase.]

A. Infrastructure Roles Further Defined

A1. Identity and Access Management (Must Have)

This is a technical role in the identity and access management area focused on linking persons and systems to campus services and data resources. This role should have skills in the following areas:

- Active Directory Services
- AWS IAM
- Azure Active Directory
- Google Cloud IAM
- Google Cloud Key Management Services
- LDAP
- Federated Identity Services (Shibboleth)

Some of the duties of this role include:

- Integration with current campus identity and access solution management systems
- Management and ongoing management of cloud identity management systems

A2. Incident Response Management (Must Have)

CU, as part of leveraging the off-premise cloud service offerings, will be responsible for cyber security incident detection, containment, and remediation for all aspects above the hosting infrastructure. This role will be responsible for replicating many of the existing on-premises

systems and processes in place for cyber security incident response management in the off-premise cloud infrastructure. This includes security event and incident management systems (typically a virtual appliance) which will provide long term storage of log data, real-time monitoring, correlation of events, and automation of incident containment. As the scale of the cloud operations increase, additional staff support will be needed to respond to alerts and incidents.

A3. Security Risk Compliance (Must Have)

As part of increasing the diversity of research funding, research is moving beyond basic research to more restricted research (health, export controlled, etc.). Additionally, effective December 31, 2017, grants and contracts which are deemed by the federal government to involve *confidential unclassified information (CUI)* will require compliance with additional security standards (NIST 800-171). 517 existing contracts were received from agencies which if awarded today would likely require 800-171 compliance. The total value of the grants exceeds $740 million. CU must either meet the security standards or not accept awards involving CUI, thus sacrificing our competitive edge to other universities who can meet these security standards.

To provide adequate services to CU Boulder researchers, cloud services must be able to support compliance requirements beyond what is required for basic research. Compliance with standards will require change in operating practices by research faculty, changes to information technology systems, and ongoing processes to document compliance. OIT and RIO have proposed a research information assurance compliance program (which includes training, compliance attestation, and consultation) managed by this role, a research information security risk assurance officer.

A4. Cloud Architect (Must Have)

This is a strategic role responsible for the overall cloud computing initiative within an organization and for directing the architectural aspects of a cloud-broker service across all aspects of IT and the business. A cloud architect should possess as many of the following skills as possible:

- Leadership
- Vision
- Strong communication and organizational skills
- Financial analysis
- Governance

Some of the duties of this role include:

- Lead cultural change for cloud adoption
- Develop and coordinate cloud architecture
- Develop cloud strategy and coordinate adoption
- Develop and coordinate cloud governance
- Provide education, evangelism and training
- Oversee cloud management and operations

This role is required for on-premise and off-premise cloud solutions. This role, for off-premise, will live within OIT's Operations unit. For off-premise, this role is 0.5 to 1.0 FTE and can be

filled by existing personnel. This role already exists for on-premise in both RC and OIT Operations.

A5. Cloud Infrastructure Engineer (Must Have)

This is a tactical implementation and operational role of the strategy developed by the cloud architect. Unlike traditional infrastructure engineers which often focus on a specific silo, such as networking, storage or compute, a cloud infrastructure engineer must have much broader expertise with an understanding of all silos, with a strong emphasis on networking and security. A cloud engineer should possess as many of the following skills as possible:

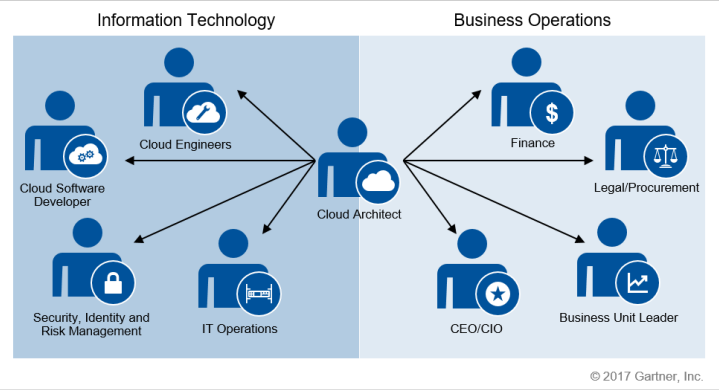


Figure 14. *The cloud engineer's relationship to the cloud architect*

- Scripting
- Automation and orchestration
- Configuration management
- Financial management understanding
- Source code management
- Cloud networking
- Understanding of IT security practices
- Understanding of service continuity and DR

Some of the duties of this role include:

- Translate cloud strategy and architecture into a highly available and secure technical implementation
- Work closely with existing subject matter experts from IAM, networking, security, virtualization and storage teams to develop the network and infrastructure solution to deliver the vision
- Continuously monitor and optimize the cloud implementation for technical and cost efficiencies
- Continuously support and collaborate with other teams and projects as needed
- Develop service continuity and DR plans

Figure 14 is an attempt to show the cloud engineer's relationship to the cloud architect.

This role is required for off-premise cloud solutions.

This role may live within OIT's Operations unit. 1.0 FTE required.

A6. IT Infrastructure Operations (Must Have)
This is a technical operational role handling design and operation of network, virtualization, compute and storage services in the off-premise and on-premise clouds. Skillsets in this role include:

- Enterprise network administration
- Enterprise virtual infrastructure administration
- Enterprise server/serverless administration
- Enterprise storage administration
- Enterprise Monitoring administration

Duties include:

- Integrate and facilitate network connectivity between systems and data
- DNS/DHCP administration
- Management of server/serverless infrastructure
- Management of firewalls, load balancers
- Integrate existing or new data transfer technology to achieve efficient data movement to/from on-campus storage (eg, local desktops, department storage systems, PetaLibrary) to/from cloud storage services (Data transfer hub).
- Liaison with Data Manager, Trainers, and CI Facilitators to provide second-level technical support.
- Hybrid-storage facility migration where storage can move between clouds without user being aware

This role is required for supporting on-premise and off-premise cloud offering.

This role may live within OIT Operations. Current NEO and SIS staff would require retraining with an additional 2 FTE required plus one-time consulting. Number of FTE required is related to the number of off-premise cloud providers within the broker service.

A7. Data Management (Must Have)
This role assists users with understanding options, policies, and techniques regarding data storage, transfer, sharing, and possibly curation.

Some of the duties of this role include:

- Be familiar with existing university policies regarding data security and retention and apply to cloud storage.
- Develop expertise in using various storage facilities at multiple off-premise cloud providers.
- Communicate and consult with users on data storage, transfer, and sharing technologies and best practices.
- Collaborate with other CI facilitators to support computational workflows that involve data storage or movement.
- Develop techniques for creating and managing metadata.
- Liaison with Library staff on data curation issues.

Necessary skills include:

- Experience with academic- or funding agency-related data management regulations and requirements
- Experience with metadata creation and management
- Experience with data sharing and transfer methods
- General understanding of data retention and curation methodology
- Excellent communication and interpersonal skills

This role is required for supporting off-premise storage.

This role may live in RC or CRDDS. Curation and metadata management would draw on existing expertise in the Libraries. 0.5 to 1.0 FTE required.

A8. Common Service Broker Engineer (Should Have)
Technical role plus significant software toolset to provide consistent user experience for service requests across multiple off-premise and on-premise cloud providers. Some of the duties include:

- Creation and maintenance of common portal interface for collecting and fulfilling requests across multiple off-premise and on-premise cloud providers, likely leveraging a multi cloud-capable cloud management platform (CMP).
- Ability to move services between off-premise and on-premise cloud providers based on specified criteria without disruption to service or end user

Some of the skills of this role include:

- Understanding of on-premise virtual infrastructure tools including VMWare and Summit services
- Skills in scripting and automation in a cloud environment
- Architecture and engineering cloud background

This role may live within OIT Operations. 1.0 FTE required plus software tool and one-time consulting.

A9. Portal Developer (Should Have)
Staff who provide "portals" are in a technical role that creates and maintains web-based interfaces for submitting computational workflows to off-premise cloud VM instances. Such portals may include existing domain-general applications including JupyterHub, Sandstone, or Rstudio; or existing or new domain-specific portals.

Some of the duties of this role include:

- Evaluate existing web interfaces for necessary functionality needed by researchers.
- Enhance existing web interfaces, or develop entirely new ones, as needed.
- Develop back-end "spawner" technology to send computational work to an appropriate off-premise cloud instance.

Skills

- Software and web development experience, including expertise with version control, formal QA/testing, and development methodologies
- Experience with programming languages and frameworks including python, javascript, etc etc
- Experience with interface-related user-experience testing
- Experience with batch scheduling systems
- Understanding of APIs for one or more cloud providers

This role is required for support of off-premise cloud services.

This role would most likely live in RC. Depending on whether existing interfaces (such as JupyterHub or Sandstone) are used, 0.5 to 1.0 FTE would be sufficient. If entirely new interfaces need to be developed and maintained then additional FTE needed. If domain-specific portals need to be developed then additional FTE needed. [Zeb Sampedro is a great contact for further discussion on this]

A10. iPaaS (integration PaaS) Engineer (Could Have)
The integration platform as a service (iPaas) engineer helps address integration of various cloud vendor services enabling development, execution and governance of integration flows connecting any combination of on premise and off premise cloud based processes, services, applications and data.

Some of the duties of this role include:

- Integrating cross cloud vendor services

Skills

- Expertise in off premise cloud services

This role is required for off-premise

This role may live in OIT.

B. Business Integration Roles Further Defined

B1. Cloud provider relationship and management (Must Have)
The is a service owner role who is responsible for the relationship and management of the relationship with the off-premise cloud provider.

The duties of this role include:

- provider selection
- overseeing contract negotiations
- onboarding

- monitoring quality of service delivery and SLA expectations
- billing details with off-premise cloud provider (not billing to users leveraging service as that is part of the Centralized Billing role).
- negotiation of volume price discounts and waiving of data egress fees
- Group cloud services training from vendor or partner

A cloud provider relationship and management role should include the following skills:

- leadership
- strong communication and organizational skills
- financial analysis
- security understanding
- ITIL Operations understanding

This role is required for off-premise

This role may live in OIT's SIS Virtual Infrastructure Team as they currently manage this relationship with regards to VMWare and the on-premise cloud. 1.0 FTE to start, 0.25 FTE to maintain.

B2. Legal, Procurement, Sourcing (Must Have)
This role handles the procurement, contracting and expense optimization.

Duties of this role include:

- Working through the procurement process with the vendor, both for initial signing and ongoing relationship
- Work with vendor and procurement as changes and modifications to the original agreement warrant
- Ensure compliance to appropriate legal requirements
- Work with legal and risk departments from other parts of the organization

A legal, procurement, and sourcing role should include the following skills:

- Negotiation skills with enterprise license agreements including Business Associate Agreements (BAA)
- Familiarity with enterprise level Familiarity with campus procurement processes
- Strong written and communication skills

This role is required for off-premise

This role likely is spread across multiple units including OIT's SIS Virtual Infrastructure Team, CU-Legal, PSC and OIT's Finance and Business Operations.

B3. Finance Forecasting Modeling (Must Have)
This is a service role to provide the forecasted costs to estimate a consumer's cloud bill prior to fulfillment so the end user does not need to manage the complexities of public IaaS financials themselves.

Duties of this role include:

- Design and develop a forecasting model for users to estimate their cloud consumption costs
- Ensure the currency of the forecasting model

A finance forecasting modeling role should include the following skills:

- Financial analysis skills
- Familiarity with campus budget rules and processes

This role is required for off-premise

This role may live in OIT's Financial and Business Operations team. It likely will rely on members of the IT Infrastructure Operations team.

B4. Centralized Billing (Must Have)
This role acts as the central point for billing for cloud services for each user/department/speedtype associated with cloud service purchases. This role is responsible for ensuring monthly charges are allocated appropriately.

Duties of this role include:

- Aggregation of billing from cloud provider and parsing out to departmental speedtype
- Ensures appropriate speedtypes are charged based on their cloud service consumption

A centralized billing role should include the following skills:

- Knowledge of accounting
- Familiarity with campus budget rules and processes

This role is required for off-premise

This role may live in OIT's Finance and Business Operations team.

B5. Grant Administrator (Must Have)
This is an administrative role focused on working with various entities to eliminate the overhead fees on cloud purchases.

Note: This is only a Must Have if we want to provide incentive to move to off-premise cloud. If we do not, it is not a Must Have.

A grant administrator role should include the following skills:

- Deep understanding of University grant proposal process

Duties of this role include:

- Working with University entities to eliminate overhead fees on cloud purchases

This role is required for off-premise

This role may live in the Office of Contracts and Grants

B6. Site licensing manager (Should Have)
Manages campus site licensing procurement for licensing software for use in the cloud which oftentimes is different than on-premise and requires vendor negotiation.

Duties of this role include:

- Identifying software products for which volume site licensing in the cloud is desired and will be cost effective
- Negotiate with vendors for price agreements

A site licensing manager should include the following skills:

- Knowledge of campus budget rules and processes

This role is required for off-premise

This role may live TBD.

C. Cloudification Roles Further Defined

C1. User relationship management - known point of contact (Must Have)
Interfacing with the consumers of the cloud service including researchers, developers, and end users. Often first point of contact for the service.

A user relationship manager should include the following skills:

- leadership
- strong communication and organizational skills
- understanding of cloud service definition and operations

Duties of this role include:

- Service-offering manager accountable for overall cloudification service
- Understanding research requirements
- Creation and operation of Cloud Community program

This role is required to support off-premise cloud services

This role will live in RC. No additional FTE needed as an analyst/facilitator/consultant would take on this role.

C2. Training (Must Have)
Trainers provide written documentation, on-demand learning materials (eg, video recordings), and in-person group tutorials aimed at teaching techniques and best practices for computation and data storage in the cloud.

Duties required of this role may include:

- Understand major use cases for cloud-based workflows, which could include all aspects of the cloud broker service (financial, security, data management/transfer/storage, research computation, etc.)
- Develop effective web-based documentation explaining these use cases.

- Develop and deliver in-person group training sessions covering specific aspects of computation and data storage in the cloud.

Skills

- Exceptional written and verbal communication skills
- Experience in developing both web-based and in-person training materials
- Expertise in delivering in-person training to groups of various sizes, including to remote attendees
- Sufficient hands-on knowledge of cloud-based technologies to enable development and delivery of related training

This role is required to support off-premise cloud services

At least part of this role would live in RC, which already has an active researcher-facing training service. 0.25 FTE required for researcher-facing training; not sure how much more needed for other customer groups.

C3. Research analyst / facilitator / consultant (Must Have)
Analysts/facilitators/consultants work directly with end users to understand their needs, recommend appropriate resources (either in off-premise or on-premise services) to meet those needs, and engage Cloudification Operations and Development if needed to help implement research workflows in the optimal location.

Duties of this role will include:

- Develop an excellent understanding of computing and storage options and techniques on multiple off-premise and on-premise platforms.
- Work with researchers to develop plans for creating or migrating workloads to the platform that best meets the needs of the research group.

Skills

- Excellent written and verbal communication skills; excellent interpersonal skills
- Practical understanding of major cloud facilities available to CU researchers
- General understanding of research methods and needs

This role is required to support off-premise cloud services.

Note that this role is very labor intensive. Based on the experience of Earth Lab, one FTE is needed to support 2-4 research groups that do not already have extensive experience in advanced computing or cloud workflows.

This role will partially live in RC, which has an established team of facilitators and consultants to assist researchers. Can start with existing RC Consulting and Operations staff and hire additional staff as customer base grows. Additional FTE will be needed to support non-research customers.

C4. IT Cloudification Operations and Development (Should Have)
Cloudification dev-ops staff provide hands-on technical assistance to create or migrate computational workflows in the cloud.

Duties

- Provide expert assistance on basic and advanced cloudification tasks:
- Assist researchers in migrating existing computational or storage workflows to cloud environments.
- Assist researchers in developing new computational or storage workflows in cloud environments.
- Provide workflow and application optimization to help minimize both the time and financial costs of using commercial cloud services.
- Assist with developing new applications that are particularly designed to work in a cloud environment
- Understanding and respecting the boundary between enabling customers' use of the cloud and doing customers' jobs for them.

Skills (not all individual dev-ops staff need to have all of these, but the pool of staff should)

- Ability to work with customers from various research disciplines and with varying levels of technical experience
- Programming or scripting ability in bash, python, R or similar interpreted languages
- Intermediate to advanced Linux command-line abilities
- Intermediate Linux system and network administration experience
- VM (OS template or image) creation experience
- Container (Singularity or Docker) creation and maintenance experience
- Experience installing and optimizing scientific and engineering applications
- Understanding of major off-premise and on-premise cloud compute and storage services
- Possibly experience with compiled languages (C, C++, Fortran)
- Possibly knowledge of domain-specific applications

This role is required for off-premise and on-premise cloud services.

This role will initially live in RC, which has an established team of technical consultants to assist researchers.

Note that this role is very labor intensive. Based on the experience of Earth Lab, one FTE is needed to support 2-4 research groups that do not already have extensive experience in advanced computing or cloud workflows.

Number of FTE required is related both to the number of different cloud services being provided and to the number of research groups supported. Potential FTE count very difficult to estimate. At a bare minimum can start with existing RC Consulting and Operations staff and hire additional staff as customer base grows. A possibility would be to have research groups fund partial FTEs that would focus on their particular needs.

Appendix V. Project Timeline of Cloud Computing Implementation, in 6 phases

Phase 1: Plan Creation Phase

Phase 1 is the **creation of the cloud governance team** including identifying key roles and responsibilities.

ROLES REQUIRED FOR THIS PHASE:

- Cloud Infrastructure Architect (OIT Ops)

SERVICES OFFERED IN THIS PHASE:

- None, other than communication plan around project and upcoming service

DELIVERABLES IN THIS PHASE:

- Creation of cloud governance team
- Creation of project communication plan
- Definition of service offering

Phase 2: Cloud Provider Selection and Procurement Phase

Phase 2 will **define the criteria for selecting the first provider** to start with and **proceed with procurement**. At the conclusion of this phase the University will have an agreement with a single IaaS provider. The service available to campus will provide users access to a single IaaS off-premise cloud vendor permitting them to purchase off-premise cloud services directly from the vendor using the signed campus agreement. No support from on-campus resources will be available nor will campus network or security policies be in place. Centralized billing will not be available yet so purchasing of services may still be performed via procurement card or other local non-centralized mechanism.

ROLES REQUIRED FOR THIS PHASE:

- Cloud Infrastructure Engineer
- Cloud provider relationship and management
- Legal
- Procurement, Sourcing

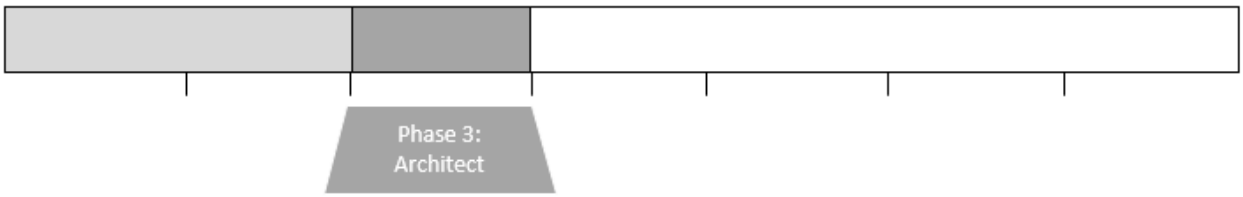
SERVICES OFFERED IN THIS PHASE:

- Basic off-premise cloud IaaS availability.

DELIVERABLES IN THIS PHASE:

- Criteria defined for provider selection
- RFP
- Procurement of contract
- Definition of service offering

Phase 3: Architect Cloud Services



Phase 3 implements some of the **core IT infrastructure** that makes CU Boulder unique including creating network paths to/from cloud provider, creation of cloud network architecture so networking to/from cloud is seamless, provides Identity and Access Management (IdentiKey) services for authentication and authorization, creation of a "tenancy" model for researchers to each have their own "space", ability to procure off-premise cloud services (virtual services, storage, etc.) through a centralized means and centralized security controls will be designed and developed.

ROLES REQUIRED FOR THIS PHASE:

- Centralized Billing
- Identity and Access Management
- Security Risk Compliance

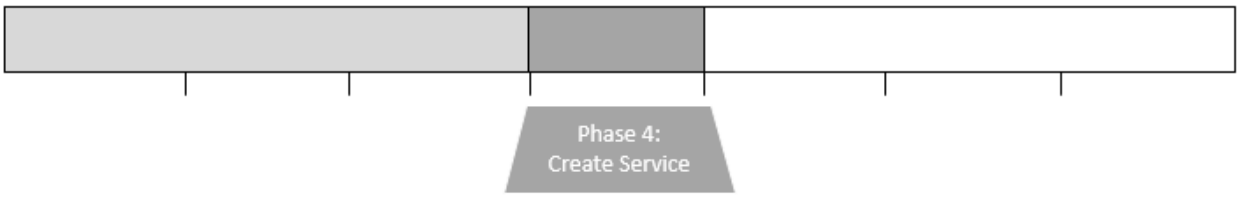
SERVICES OFFERED IN THIS PHASE:

- Infrastructure cloud broker services
- Centralized billing

DELIVERABLES IN THIS PHASE:

- Network egress/ingress pathways
- IdentiKey available for off-premise
- Purchasing ability of virtual services, storage and other off-premise cloud services under the "CU Boulder umbrella"
- Centralized billing service

Phase 4: Creation of Cloudification or Cyberinfrastructure Service



Phase 4 creates the **foundation for the cloud broker service** from an end user perspective also known as the "cloudification" service. This will assist researchers in their move to the cloud and, once migrated, the ability to provide ongoing operational support. This includes VM templates, configuration and deployment tools, automated security controls and other specific application software stacks.

ROLES REQUIRED FOR THIS PHASE:

- Data management
- Research analyst/facilitator
- Training
- User relationship management

ROLES RECOMMENDED FOR THIS PHASE:

- IT Cloudification Operations and Development

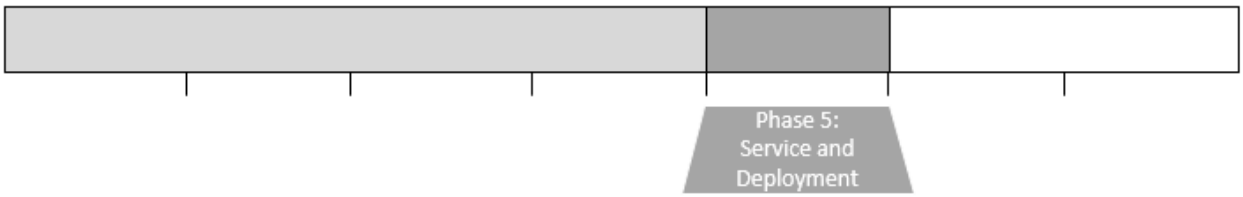
SERVICES OFFERED IN THIS PHASE:

- "Cloudification" service

DELIVERABLES IN THIS PHASE:

- VM and container templates
- Configuration and deployment tools
- Automated security controls
- Specific software stacks on top of VMs or serverless infrastructure
- Training and consulting for research groups on using cloud compute and storage effectively

Phase 5: Creation of Service and Deployment model



Phase 5 creates a **centralized forecasting** model to predict cloud consumption costs. In addition, it documents and publishes **governance strategy** including roles, policies, alerts, **incident response** and audit processes. Provides an **alerting service** to notify researchers when thresholds are reached, such as spending or resource limits. It also **addresses the overhead fees assessed to grants on cloud purchases** is part of this phase.

ROLES REQUIRED FOR THIS PHASE:

- Finance Forecasting Modeling
- Grant Administrator
- Incident Response Management

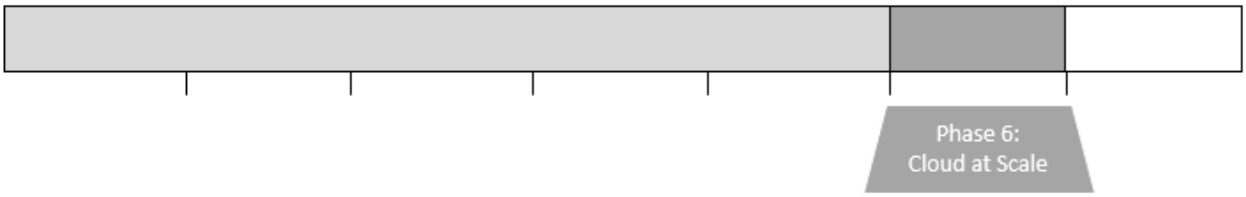
SERVICES OFFERED IN THIS PHASE:

- Centralized Forecasting model
- Governance strategy
- Incident response
- Alerting

DELIVERABLES IN THIS PHASE:

- Cloud consumption alerts
- Forecasted cloud consumption costs
- Auditing processes
- Addressing Overhead fees

Phase 6: Operating Cloud at Scale



Phase 6 focuses on the **creation of the tools and processes** required to manage the ongoing health of the cloud service. This includes oversight around operation and utilization as well as security validation and remediation tools. Self-service portal services are created that may include applications such as JupyterHub, Sandstone, or Rstudio as well as domain-specific portals.

ROLES REQUIRED FOR THIS PHASE:

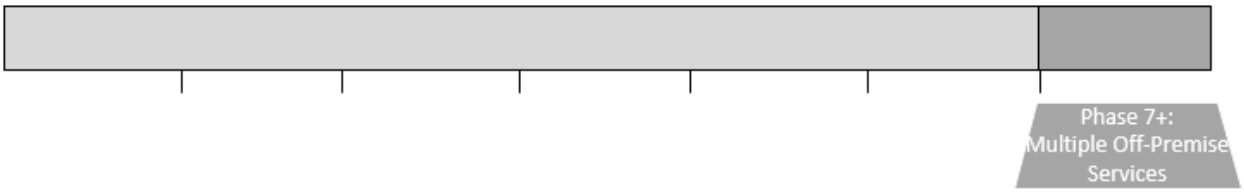
- Portal Developer
- Site licensing manager

SERVICES OFFERED IN THIS PHASE:

- Self-service portal

DELIVERABLES IN THIS PHASE:

- Tools and processes to operate and manage environment at scale are procured and deployed
- Web-based self-service portal interface for submittal of computational workflows to off premise cloud services
- Software licensing service and/or software license



After Phase 6: Multiple Off-Premise Cloud Services

This phase **adds** additional **off-premise cloud vendors** to the service allowing for **multi-vendor off-premise options**.

ROLES REQUIRED FOR THIS PHASE:

- Common Service Broker Engineer
- iPaaS Engineer

SERVICES OFFERED IN THIS PHASE:

- Multi off-premise cloud service

DELIVERABLES IN THIS PHASE:

- Multiple off-premise service providers can be leveraged
- Movement of workloads between different off-premise cloud providers based upon strengths each have to offer