AWS Research Cloud Program

Accelerating Science and Innovation

Researcher's Handbook

(Version 2.0, Released 2017-10-01)

Edited by

Kevin Jorissen and Brendan Bouffler

with contributions by the

Research & Technical Computing Group

Amazon Web Services

Notices

This document is provided for informational purposes only. It represents AWS's current product offerings and practices as of the date of issue of this document, which are subject to change without notice. Users are responsible for making their own independent assessment of the information in this document and any use of AWS's products or services, each of which is provided "as is" without warranty of any kind, whether express or implied. This document does not create any warranties, representations, contractual commitments, conditions or assurances from AWS, its affiliates, suppliers or licensors. The responsibilities and liabilities of AWS to its customers are controlled by AWS agreements. This handbook is not part of, and does not modify any agreement between AWS and its customers. This handbook may include a set of suggested solutions but should not be construed as a legally-binding offer from AWS. For current prices of AWS services, please refer to the AWS website at www.aws.amazon.com.

FOREWORD

When I was asked to write the foreword for a book that explains how easy it is to use cloud computing for research, I jumped at the opportunity. Technology should never be an obstacle to researchers and their ability to gain new insights into our world.

For the past 20 years, I've run several research departments at the Fraunhofer Institute for Industrial Mathematics (ITWM) as well as leading our IT department for more than 10 of those years. I always felt it obvious to look for solutions that make the life of a researcher developing and using simulation software easier - and to transform IT to support this.

In 2001, I wrote a proposal called "I-LAB" (Internet-Lab), taking inspiration from the MOSAIC project that led to the first web browser. "*The browser was the basis for the initial success of the internet, and with I-LAB, the internet should become a ´problem solving environment´, first for researchers, then for large companies and finally for everybody*", I wrote at the time.

However, we understood then that the protocols and concepts of Grid Computing had been too complex to become a real success. At Fraunhofer, we started to use web services intensively to automate fault-tolerant execution, provisioning and large-scale parallelization. We started to develop what is today known as BeeGFS - a parallel file system – and we developed industry-grade applications based on asynchronous communication protocols. After Amazon Web Services launched Amazon S3 and Amazon EC2, we were ready to go, and within a few weeks, I ran several large parallel finance applications in the cloud directly from my laptop. I was even able to do this in real time – in fact, live during lectures I was giving on the subject!

Since then AWS has become the most flexible and versatile web services company and has set the standards for cloud computing. This book is an easy to read, step-by-step guide to AWS and its services. It describes the application ecosystem that is helping researchers around the globe to provision software components and set up their research IT infrastructure with just a few mouse clicks.

Today I head up the HPC Department at Fraunhofer ITWM where we develop applications that rely on highly specialized HPC platforms. We now use the cloud to run many of these applications at scale and today cloud and HPC coexist happily because for many problems, the limits of a supercomputer are where we begin our work.

In most universities and research institutes, it is considered a waste of valuable resources to run early stage developments, simple farming jobs or weakly coupled applications on highly specialized systems when there is a readily available cloud nearby that can scale up to whatever extent your science requires.

I'm still fascinated by the ongoing revolution that's come with the availability of cloud infrastructures and what this means for research and startups. For those like me that are interested in this broader view, I still recommend picking up Nicolas Carr´s book from 2008, "The Big Switch", where he describes the changes the use of cloud infrastructures will bring. Along with machine intelligence these topics will have a large impact on science and society.

I recommend this current document to get answers on why all this is important to you as a researcher, as well as practical knowledge on how to get started. This is a book written not for the IT guy, but for the researcher who's seeking a better way to get their work done.

Dr. Franz-Josef Pfreundt
Division Director, Fraunhofer Institute for Industrial Mathematics
Kaiserslautern, 8 December 2016

Contents

1 Introduction to the AWS Researcher's Handbook

As the authors and Amazon Web Services Scientific Computing experts, **we're delighted to welcome you to the cloud** with this handbook. We are ourselves life scientists, physicists and other researchers and PhDs who started using AWS for our scientific research during our academic careers. We grew to love AWS because of the flexibility it gave us to use computing resources right when we needed them, and to solve computational problems bigger than the constraints of a fixed 16-node department cluster. We appreciated no longer spending our time maintaining those aging clusters. Through experience, we found AWS to be a great place for sharing data and tools with our collaborators.

This Handbook is the "missing manual" for researchers that we wish we'd had at the time. It collects information on the services you're likely to need as a researcher, and it's written in a way that we think you'll appreciate – with the right amount of seriousness and technical depth. The Handbook allows researchers to "get to the science" as soon as possible, with the confidence that their data and budgets are safe in the cloud. It provides key information to use AWS for research, so that researchers who are not IT professionals can benefit from the AWS Cloud right away.

In this first chapter of the handbook we outline the reasons why the cloud is fundamental to changing the way computing is used for research. These reasons span from being able to access the right scale you need—large or small—to the agility you gain from being able to access resources you never planned for, just to find out if it solves a problem for you. The AWS Cloud allows you to speed up the cycles of modelling, data analysis and experimentation—a "failing fast" approach that ultimately leads to research success and a better understanding of the world. In the Cloud you are not required to pre-conceive all of the answers in advance.

In subsequent chapters we show you—step by step—how to create an AWS account and configure it so you can establish a secure and compliant place for your data. You'll also have confidence that you can predict, monitor and govern your expenditure accurately and transparently to help keep you on budget.

We provide **deeper dives on AWS services** that are commonly used by the research community, including computing, storage, analytics, and machine learning services.

The second half of this handbook introduces you to **the rich research community** thriving in the AWS Cloud. **We show you what's possible**, with an array of solution outlines and offerings from third party companies and AWS Partners that can immediately get you working with real, scientifically relevant solutions. For example, you can launch a large High-Performance-Computing (HPC) cluster with hundreds of day-to-day science applications pre-installed. You can create more complex projects building upon existing solutions, or create and share your own bespoke scientific architecture.

We have also included a **glossary** in the back of this Researcher's Handbook.

This Handbook is a part of the AWS Research Cloud Program, a program that provides researchers in the global scientific community with access to easy-to-use cloud resources. The **benefits of the program** include discounts on certain services, a fast

track to invoice-backed billing, and access to up-to-date information about techniques relevant to research. (See chapter 2.1.) Registering for this program also keeps you in the loop on new developments. We expect to **update this Researcher's handbook** several times a year, and we may share new technical solutions that we believe raise the bar on how research computing can be done, and will save you time and effort.

It is NOT necessary to have an AWS account to register for the program and receive this Researcher's Handbook.

Ultimately, our aim is to make the cloud more accessible for researchers, thereby creating a greater impact on the research world. If there is **anything you think of** that will help us improve the program, this handbook, or any of the services you use, please contact us at any time at aws-research-cloud@amazon.com.

1.1 What is Cloud Computing?

Before digging into the science, it is first important to understand why cloud computing provides a different model of IT delivery compared to traditional methods used for research.

Cloud computing is a quite simple idea. The primary difference between cloud computing and "traditional" computing or IT, is that in a cloud model you are not buying physical assets. Before cloud computing, if you wanted compute, storage and other IT services for research you first needed to buy physical servers, network equipment, racks, cabling, etc. Then in a secure, air conditioned and climate controlled room (something you had to organize yourself), you would take your recently purchased IT equipment and unpack it, install it, connect it, configure it, assign it, manage it, and monitor it. You paid the bills to power all of this infrastructure—*and*—every few years you had to replace your servers and infrastructure (assuming they lasted a few years), along with occasionally purchasing additional servers to meet projected increases in demand.

This wasn't an ideal situation, but if you wanted IT services it simply had to be done. Amazon did this for many years, and in that time developed unique skills in operating massive scale technology infrastructure and datacenters supporting Amazon.com. Then came the idea that other organizations could benefit from Amazon's experience and investment in running a hyper-scale, distributed, transactional IT infrastructure. This idea was cloud computing—powerful compute, storage, and other IT services running in Amazon's data centers. This allowed a new model of IT infrastructure delivery whereby customers would pay to tap into these services in a utility-style model—only paying for the resources they need.

1.2 What is the AWS Cloud?

The AWS Cloud is a broad set of global compute, storage, networking, database, analytics, application, deployment, management, developer, Internet-of-Things (IoT), Artificial Intelligence (AI), and security services, all of which are listed at

http://aws.amazon.com/products/. **Figure 1** on the following page is a simple view of AWS Cloud services.

AWS Cloud services are provided with a range of supporting components such as management tools, networking services, and application augmentation services, with multiple interfaces to AWS Application Programming Interface (API)-based services, including Software Development Kits (SDKs), Integrated Development Environment (IDE) toolkits, and Command Line Tools: http://aws.amazon.com/tools/.

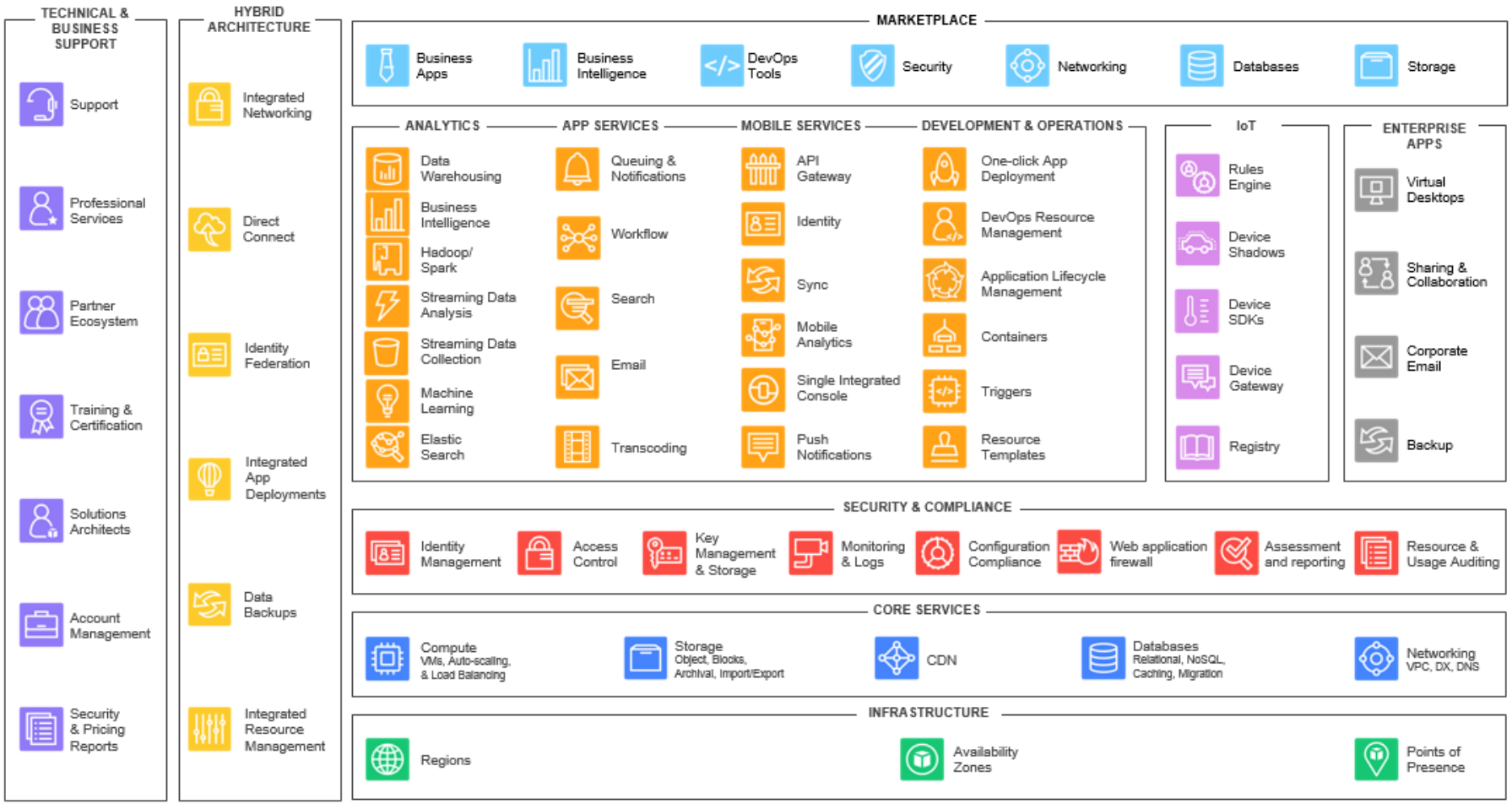


Figure 1 – High-Level View of the AWS Cloud.

In the cloud you are not required to pre-conceive all of the answers in advance, or be constrained by what you guessed might have been the right solution a year ago when you wrote your grant application. New techniques come along, and new services arise regularly. When advances in compute infrastructure better enable your research, the cloud allows you to take advantage of them immediately—rather than wait several years until your next on-premises hardware refresh. As an example of such innovation, AWS launched over 1,000 new features and services in 2016.

1.2.1 Global Footprint

AWS's Cloud services are hosted within its global data center footprint, allowing customers to consume services without having to build or manage facilities or equipment. AWS Cloud services are offered in separate Regions in a number of separate geographic areas. Each Region contains multiple, isolated locations known as Availability Zones (AZs) that are engineered to be isolated from failures in other AZs. Each AZ in turn consists of 1 or more data centers.

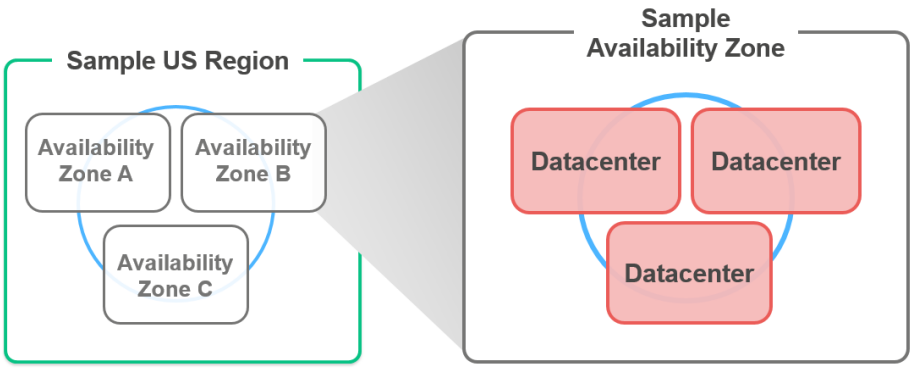


Figure 2 – AWS AZs consist of 1+ data centers, each with redundant power, networking, and connectivity—housed in separate facilities. Every AWS Region contains 2+ AZs. Some Regions have as many as 5 AZs.

Figure 3 displays AWS's 16 global regions and 44 Availability Zones.

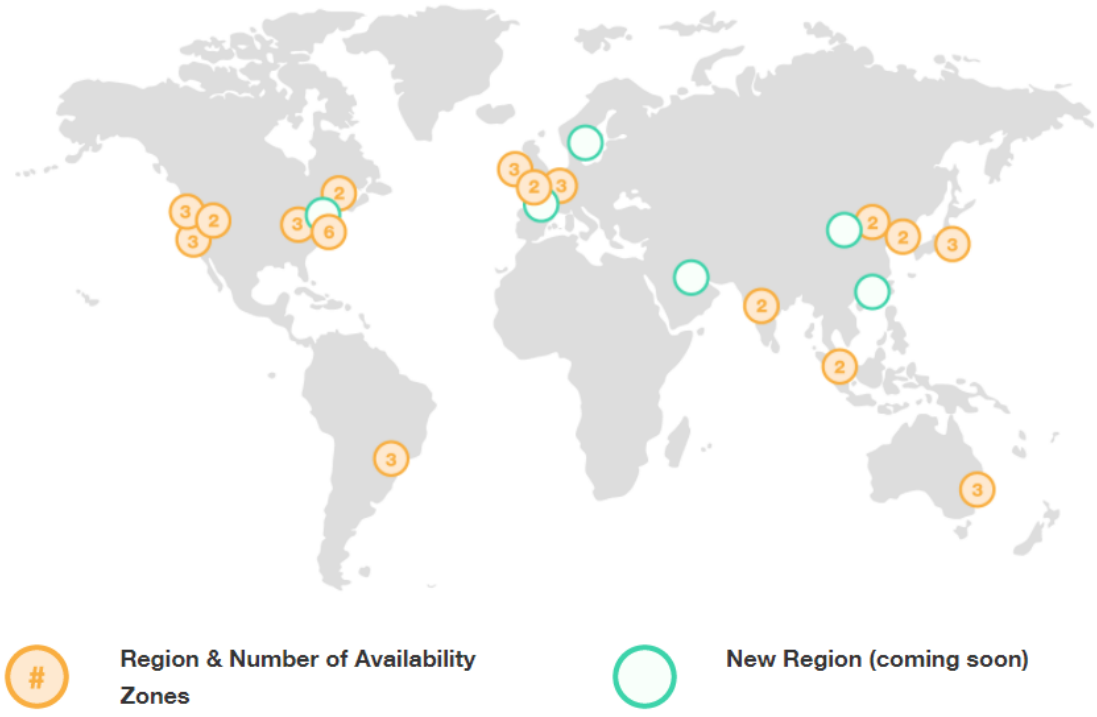


Figure 3 – AWS's global infrastructure consists of 16 regions and 44 Availability Zones (as of October 2017). At least 6 more AWS Regions (and 17 Availability Zones) in France, Sweden, Hong Kong, China, Bahrain, and a second AWS GovCloud Region in the US are coming online in the next year.

1.3 Why AWS for Science and Research?

The AWS Cloud allows researchers to quickly (and affordably) access the latest versions of many resources that may otherwise be difficult to access. This level of availability promotes deployment of new ideas or services that in other circumstances might have taken months or years to achieve. As a result, you can expedite prototyping an idea in order to discover its feasibility without incurring lasting costs or irreversible investment. Likewise, other stages of the scientific research cycle such as analysis and validation are expedited. By reducing the time to deploy and run your experiment, we increase the time you can spend developing new ideas or completing analysis, which allows scientific goals to be reached rapidly.

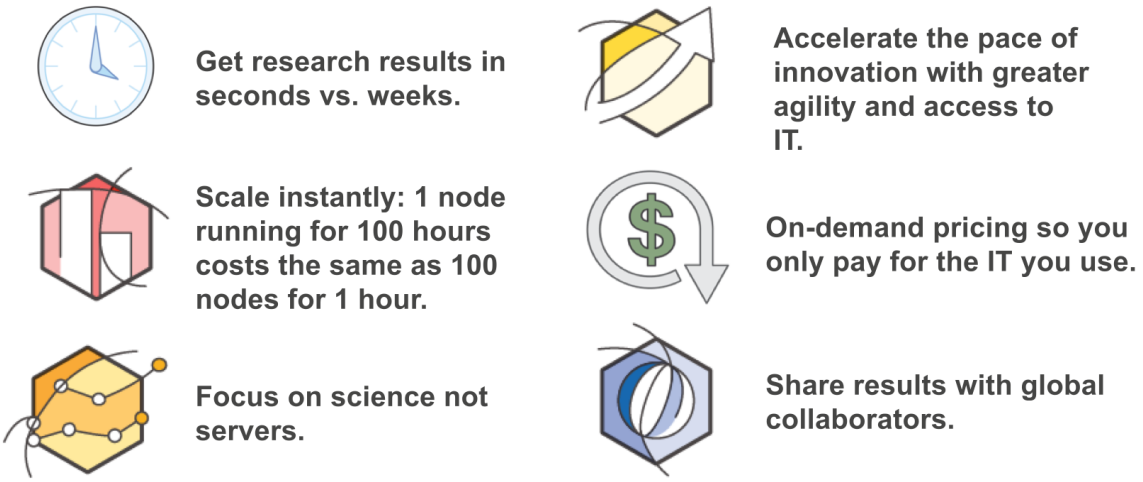


Figure 4 - Benefits of deploying research workloads in the cloud.

"Pay as you go" also means you can better match your compute capacity to your needs, in real-time. Historically, many researchers used an expensive workstation for real-time work, but even if you use it 9-5, Monday to Friday, you're only using it about one-fifth of the time. But in the AWS Cloud, you could have five times the compute performance (i.e., more cores, more memory) for that same amount of the time for the same cost, so when you **do** compute you get your results five times sooner.

Some of the key benefits of using AWS for research are:

- **Access to Innovation** – AWS constantly adds new services and new functionality, which you can use immediately, rather than wait for a 5-yearly on-premises hardware refresh. You can also tap into the innovation of our partner network, which further extends the power and breadth of AWS services.
- **Saves Time** – You can create an AWS account and your first computing machine—or cluster—in minutes.
- **Low Cost** – With pay-as-you-go pricing, there is no need to purchase and maintain expensive, space-consuming hardware. Researchers pay only for what they use and can lower costs further, for example with Amazon Elastic Compute Cloud (Amazon EC2) Spot Instances.

- **Flexible** – AWS has a large variety of compute, storage, database, and network offerings, meaning that you can match the right hardware to the job at hand.
- **Elastic and Scalable** – With AWS you scale up or down, paying only for what you use, when you need it.
- **A Space for Collaboration** – Collaborate securely with your colleagues around the corner or around the world. Our global footprint of cloud regions matches the global nature of research. You can safely share data, machine images, or workflows with just a few clicks.
- **Reproducible science** – Snapshot your entire working environment any time. You or your peers can spin up a live copy of the environment five years from now to verify your published results, or to revisit your data with the latest simulation or analysis tools.

Let's look at a few success stories of research organizations using AWS. You can find about a hundred more listed in chapter 10.4.

1.3.1 Astronomy: ICRAR / CHILES discovering neutral hydrogen galaxies

Higher performance at lower cost. A global radio astronomy consortium (led by a team at Columbia University) needed to process data from the Very Large Array (VLA) telescope in New Mexico. They required a 12-hour processing SLA, stemming from an opportunistic observing schedule at the observatory. Their partners at the International Center for Radio Astronomy Research (ICRAR) calculated that meant they needed around $2 million of HPC hardware – money that simply wasn't available. Working with AWS to exploit the EC2 Spot market in AWS's northern Virginia region, they were able to **deploy their HPC workload at a much larger scale** – meaning that they always beat their SLA (frequently by many hours), whilst averaging only $1,200 per month of EC2 compute resources. The project went on to smash the previous record for identifying a neutral hydrogen galaxy by nearly **twice the redshift** of its predecessor.

- https://arxiv.org/pdf/1511.00401.pdf
- http://www.icrar.org/astronomers-smash-cosmic-records-see-hydrogen-distant-galaxy/

1.3.2 High-Energy Physics: FermiLab in Chicago

50,000 cores for the weekend. AWS has collaborated with the U.S. Department of Energy's FermiLab in Chicago and the software engineering community that creates and maintains HT Condor, an HPC/HTC job scheduler, born in the global **particle physics** community. The result has been the "HT Condor Annex" – a set of extensions that have allowed FermiLab to demonstrate High-Energy Physics accelerator data analysis workloads running across several AWS regions and using **tens of thousands of cores** at a time. FermiLab relied on Amazon EC2 Spot market for access to potentially massive fleets of cores at very low cost, without any long-term commitment.

- https://aws.amazon.com/blogs/aws/experiment-that-discovered-the-higgs-boson-uses-aws-to-probe-nature/

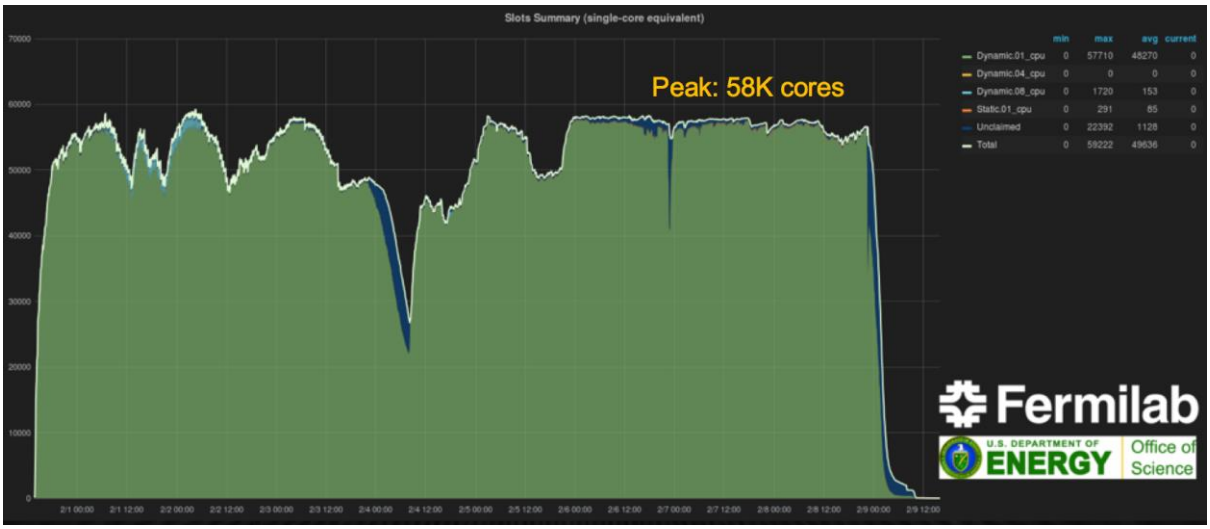


Figure 5 - High-Energy Physics: FermiLab in Chicago.

1.3.3 Natural Language Processing: Clemson University

Researchers at Clemson University took this approach even further and ran a massive Natural Language Processing study on 1.1 million vCPUs using AWS Spot instances – a compute fleet size comparable in core count to the largest supercomputers in the world. They conducted nearly half a million topic modeling experiments to study how human language is processed by computers.

- https://aws.amazon.com/blogs/aws/natural-language-processing-at-clemson-university-1-1-million-vcpus-ec2-spot-instances/

1.3.4 Machine Learning: RISElab at Berkeley Collaboration

Machine Learning of the future. Amazon is a founding sponsor of the National Science Foundation backed RISELab at U.C. Berkeley. Its predecessor AMPLab published cutting-edge research on machine learning (ML) and built an open software stack for data analytics, including Apache Spark, which is now used all over the world to analyze massive amounts of research data. RISELab is building a next generation ML/AI stack for **real-time analytics on live datasets** – relevant for a broad range of today's and tomorrow's challenges, such as **smart cities** or fleet management of autonomous vehicles. RISELab builds on AWS compute and data analytics services.

New serverless HPC paradigm gives instant scalability. In one of their areas of work, RISELab and the Berkeley Center for Computational Imaging are making cloud infrastructure easier to use for scientists and engineers. The Pywren project supports a simple map-and-reduce-like interface, from python and running on top of AWS Lambda. Lambda functions are snippets of code that AWS executes for you on managed hardware. Pywren harnesses thousands of these workers in parallel, achieving **40 peak TFLOPS**[1] using Pywren on AWS Lambda, and over 60 GB/sec read/write performance to Amazon

[1] A teraflop (TFLOP) is a measure of computing speed equal to one trillion floating-point operations per second.

Simple Storage Service (Amazon S3) for data processing. You can run this from a Jupyter Notebook without ever explicitly touching a server, instead just providing the Python code to be executed. This is truly democratizing parallel scaling capabilities for nearly everyone, which used to be the sole preserve of large super-computing centers.

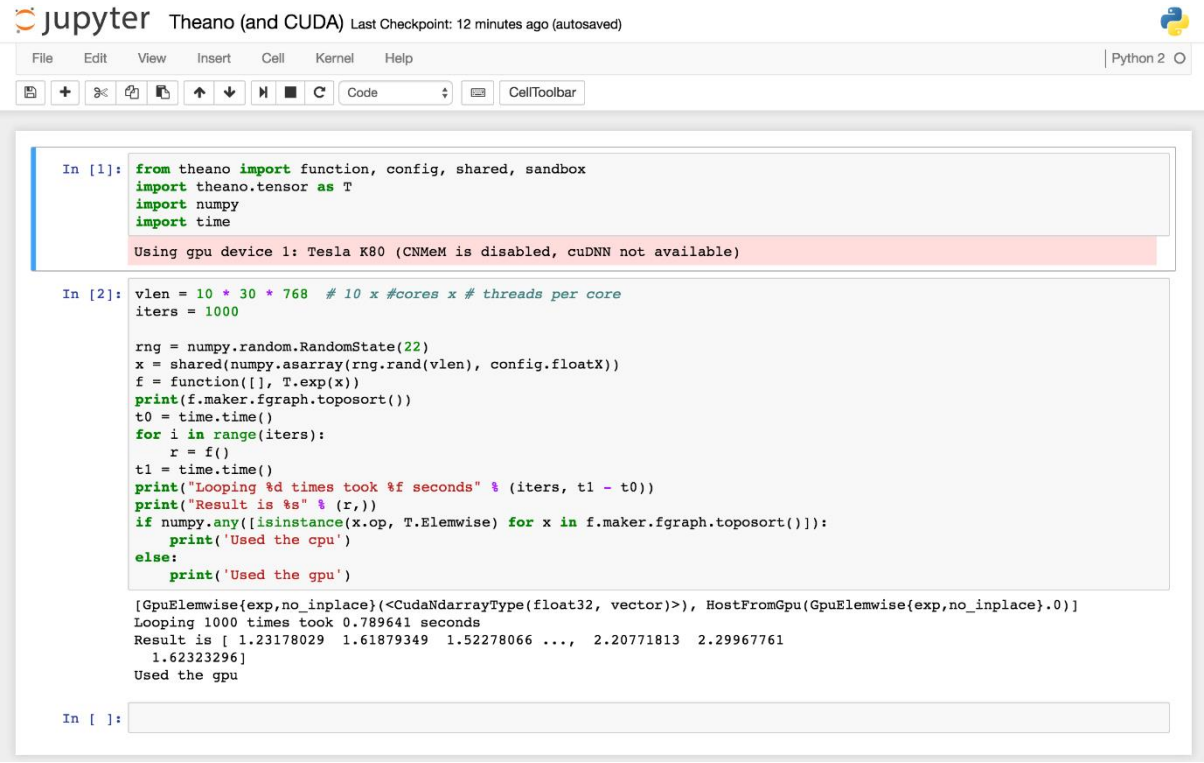


Figure 6 - Machine Learning: RISElab at Berkeley Collaboration.

- http://pywren.io/pywren.html
- https://arxiv.org/abs/1702.04024
- https://rise.cs.berkeley.edu

1.3.5 Genomics: Unilever

Faster results. Unilever augmented their existing HPC capacity with EC2, enabling the company to **process genetic sequences** twenty times faster than with their in-house system. “The key advantage that AWS has over running this workflow on Unilever’s existing cluster is the ability to scale up to a much larger number of parallel compute nodes on demand,” said Pete Keeley, eScience Technical Lead for R&D IT at Unilever.

1.3.6 Quantum Chemistry: Novartis

AWS enabled Novartis to massively accelerate pre-clinical R&D focused in the area of computational chemistry. “We completed the equivalent of **thirty-nine years of computational chemistry** in just under nine hours for a cost of around $4,200,” noted Steve Litster, Global Head of Scientific Computing, at Novartis.

- https://www.top500.org/news/sponsored/why-customers-are-moving-high-performance-computing-workloads-to-amazon-web-services-1/

1.3.7 Computational Fluid Dynamics: TLG Aerospace

Solve larger problems. Prior to moving computational fluid dynamic (CFD) simulations to AWS, TLG Aerospace couldn't run jobs of more than 1,000 nodes, resulting in lost opportunities. AWS afforded them the scalability to transcend that limitation. "We are definitely saving money by actively monitoring jobs to catch problems early and reduce rework," explains Andrew McComas, Engineering Manager at TLG. "We can also use it to reduce unnecessary cost in larger jobs that may otherwise run longer than required."

- https://www.top500.org/news/sponsored/why-customers-are-moving-high-performance-computing-workloads-to-amazon-web-services-1/

1.3.8 Medical Imaging: National Database for Autism Research (NDAR)

Collaborative Big Data research. The National Institute of Mental Health Data Archive (NDA) makes research data available for reuse. Data collected across projects can be aggregated and made available, including clinical data, and the results of imaging, genomic, and other experimental data collected from the same participants. In this way, **separate experiments** on genotypes and brain volumes can inform the research community on the over one hundred thousand subjects now in the NDA.

The NDA holds rich datasets (fastq, brain imaging) in object-based storage (Amazon S3). It supports the deployment of packages (created through the NDA Query tools) to an Amazon Web Service Oracle database. The NDA envisions real-time computation against rich datasets that can be initiated without the need to download full packages. Furthermore, a new category of data structure has been created called "evaluated data." This allows researchers using NDA cloud capabilities and computational pipelines to write any analyzed data directly back to the miNDAR database. Databases can also be populated with your own raw or evaluated data and uploaded directly back into the NDA for a streamlined data submission directly from a hosted database.

- https://ndar.nih.gov/cloud_overview.html

1.3.9 Genomics: GT-Scan2 from CSIRO in Australia

New HPC paradigms. In 2016, the Commonwealth Scientific and Industrial Research Organisation (CSIRO - a federal government agency for scientific research in Australia) used AWS Lambda functions to completely re-engineer an HPC workload called GT-Scan2 that had been developed in a traditional cluster setting and identifies **optimal CRISPR gene editing sites** through simulation. The re-casting of the code to use AWS Lambda, and other "serverless" functions in AWS, took only a few weeks for the developers at CSIRO.

AWS Lambda (see chapter 7.5) is a service that deploys software functions in a variety of languages into the cloud natively, triggered directly or driven by events in the cloud. The infrastructure (hardware, operating system and software environment) for AWS Lambda is managed by AWS and scales rapidly. This is crucial for using GT-Scan2 for personalized treatment, because the complexity of the targeted gene varies dramatically. A typical GTScan-2 job takes less than a minute, but the variation between jobs ranges from 1 second to 5 minutes. This fast fluctuation in load over minutes rather than hours,

and the need for rapid turn-around times meant that large amounts of server hardware could end up idle simply waiting for a job to arrive. A naïve EC2-based solution would also be limited, since new instances – which may take minutes to deploy - would come online too slowly to keep the runtime stable. But with AWS Lambda, the GTScan-2 runtime is stable at a few minutes per complete job, regardless of how many jobs (i.e. genetic samples) are sent to it, as new Lambda workers scale up promptly.

We see a future where the traditional model of HPC frequently breaks down in the face of highly optimized services like AWS Lambda, which allow developers to find a more optimized environment for their code to run – and at the scale and immediacy that suits their users – compared to traditional settings with fixed hardware capacities that are rationed by schedulers and governed by strict software and hardware limits where one size fits all.

- https://aws.amazon.com/blogs/aws/genome-engineering-applications-early-adopters-of-the-cloud/
- https://blog.csiro.au/cloud-technology-giving-us-a-crispr-view-to-help-cure-disease/

1.4 AWS Compute and Storage

You can create compute solutions quickly by starting virtual servers in Amazon EC2, attaching some elastic disk storage to them (Amazon Elastic Block Store [Amazon EBS]), and enclosing them within custom-designed networks called Virtual Private Clouds (Amazon VPCs) that include built-in firewalls to add layers of protection. **Figure 7** shows a 4-core compute instance ready to launch CentOS 7 and get to work after a small number of clicks. In a few minutes this user will be able to access the machine via Secure Shell (SSH) and begin running applications.

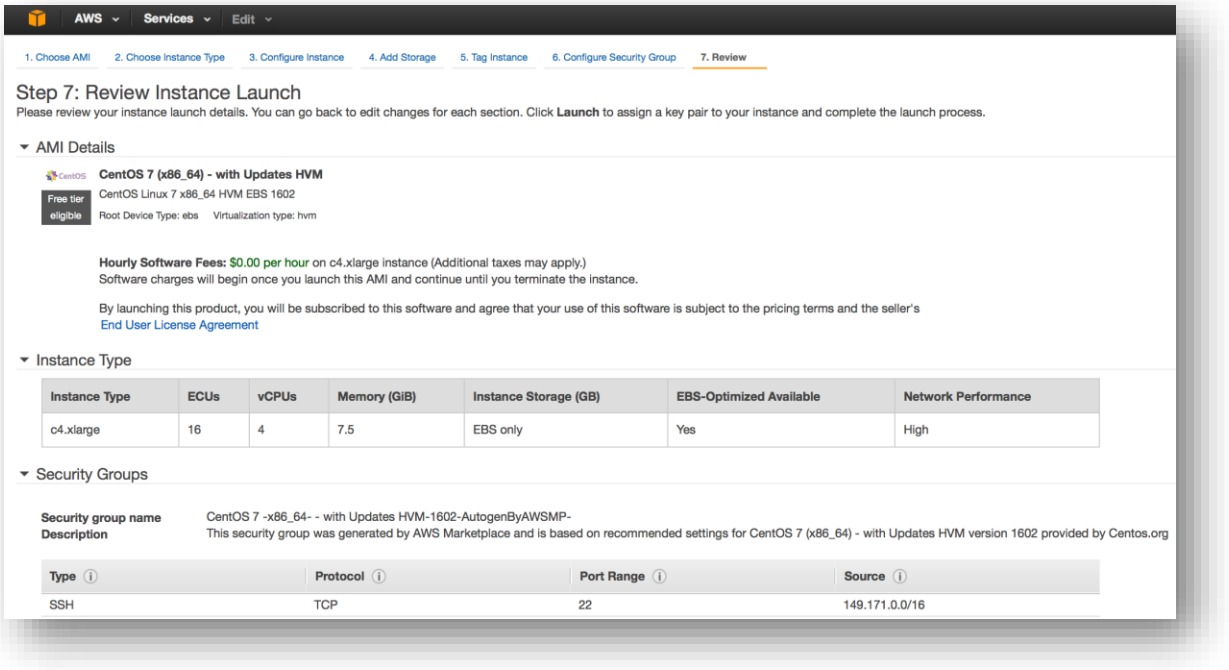


Figure 7 - Launching an EC2 instance only takes a few minutes and is easily customizable.

Launching an EC2 instance can be done in minutes, and happens so often inside the AWS Cloud that we've built many helpful tools that allow you to "freeze-dry" configurations and packages so you can reconstitute them in a flash.[2] Automating your activity is a good idea so you won't forget important steps like making sure the firewalls are on or attaching your dataset to your computer. Automation also makes it easier to share your scientific pipeline with collaborators.

While this can be done conveniently through the AWS Management Console, nearly everything can also be done via a command line interface or one of several Application Programming Interfaces (APIs; choose your favorite language from a long list). This means that repetitive tasks or complex architectures can be turned into shell scripts. There are more advanced levels of automation, but by now you'll get the idea that you can go from simple ideas to complex creations quickly.

1.4.1 Amazon EC2 – Compute Instances

Amazon EC2 instances are virtual servers built from virtual machine images called Amazon Machine Images (AMIs) that you can run in minutes on a wide range of "instance types". Instance types represent differently scoped and sized computer servers with capabilities that

[2] See information on AMIs http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/AMIs.html and AWS CloudFormation templates https://aws.amazon.com/cloudformation/

you can select from a menu[3]. They range from tiny, single-core versions that might be great for running a wiki through to large instance types with terabytes of RAM or specialist GPUs.

The AWS website lists each instance family and type that you will become familiar with. Within each family, you choose at boot time what size instance you want to run to suit your application's needs. The size of each instance within a capability or instance family looks very much like a familiar t-shirt size-naming scheme.

For example, if you need 4 cores of our high-performance "C4" instance, after some studying of the Amazon EC2 instance family list, you might choose a "c4.xlarge." Later, if you wanted more cores for a larger problem, you'd reboot into a c4.2xlarge (8 cores) or a c4.8xlarge (36 cores). The same applies if you wanted to try running the same code with more memory—you can give yourself an upgrade in minutes. If it doesn't positively impact your throughput the way you'd hoped, you can always go back to the smaller instance or lower-memory instance and save your money.

1.4.2 Range of Storage Types

AWS storage is designed to scale in size, so if your prototype experiment proves to be a success, you can quickly scale it up without reformulating your scientific pipeline. AWS storage offerings are summarized in **Table 1** below.

Table 1 - AWS Storage Options.

If You Need:	Consider Using:
A scalable, durable service to make data accessible from any Internet location, for user-generated content, active archive, serverless computing, Big Data storage or backup and recovery	Amazon Simple Storage Service (Amazon S3)
Highly affordable long-term storage that can replace tape for archive and regulatory compliance	Amazon Glacier
Persistent local storage for Amazon EC2 instances, for relational and NoSQL databases, data warehousing, enterprise applications, Big Data processing, or backup and recovery	Amazon Elastic Block Storage (Amazon EBS)
A file system interface and file system access semantics to make data available to one or more EC2 instances, for content serving, enterprise applications, media processing workflows, Big Data storage or backup and recovery	Amazon Elastic File System (Amazon EFS)
A hybrid storage cloud augmenting your on-premises environment with Amazon cloud storage, for bursting, tiering or migration	AWS Storage Gateway
A portfolio of services to help simplify and accelerate moving data of all types and sizes into and out of the AWS Cloud	Cloud Data Migration Services

1.4.3 Services Built on Amazon EC2 and Storage

AWS offers many additional services beyond the core EC2 compute and storage services we've visited so far: for example, data analytics, data streaming, visualization, machine

[3] Beyond just using AMIs some researchers use the Amazon EC2 Container Service (Amazon ECS) to deploy and manage "Docker images", which package an application's code, configurations, and dependencies into simple building blocks. It helps applications to be deployed reliably regardless of the environment.

learning and deep learning. Many of these services integrate with the basic storage services, so you can build a complex architecture around your data stored in Amazon S3. The services are designed to manage the “undifferentiated heavy lifting” (i.e. drudgery) for you, freeing you to focus on the more scientifically relevant part of your work. It also lets us deliver you services that are run according to best practice, are extremely secure, and are operated and supported by a global team of professionals.

For example, Amazon Relational Database Service (Amazon RDS) offers multiple different types of databases running as a service. You choose which database type you want to use—we provide it almost instantly. This leaves you ready to import data and run queries. You never need to patch the database again or update the Operating System—we take care of that for you, and we’ll even handle synchronous replication between Availability Zones if you need a highly available service.

1.5 AWS Marketplace

While AWS offers many services ourselves, the real power of community is delivered through AWS Marketplace. AWS Marketplace is an application store for the cloud. It allows third-party groups (like companies or research organizations) to offer specialized solutions optimized for specific communities (see **Figure 8** below)

Some are virtual machines that others create and share in the cloud, like the CentOS Linux distribution, one of our most popular community AMIs.

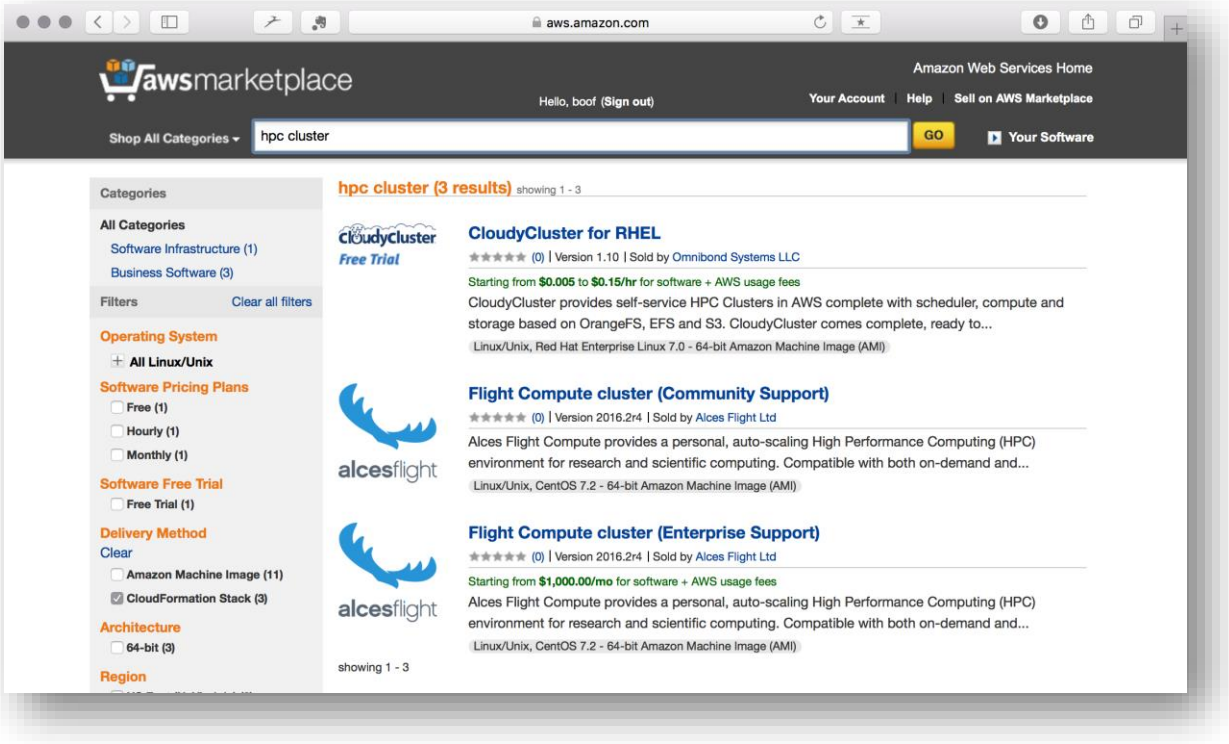


Figure 8 - AWS Marketplace makes it easy to find and deploy solution stacks built to specification and according to best practices, including a range of research-related workloads.

Another example is **Alces Flight**, which builds an HPC cluster in a few minutes. It contains hundreds of common scientific applications[4] (like Amber, NAMD, BLAST, etc.) that are ready to run immediately. The environment that will be very familiar to you if you've ever used a central supercomputing or campus facility before. Alces Flight makes use of AWS's Amazon EC2 Spot market[5] where $50 goes a long way. This is your own supercomputer, but with no RFP required, and with no queues. Intel's **Lustre**[6] and the Fraunhofer Institute's **BeeGFS** are also both in AWS Marketplace and extend the storage possibilities in AWS by building complex parallel file system solutions.

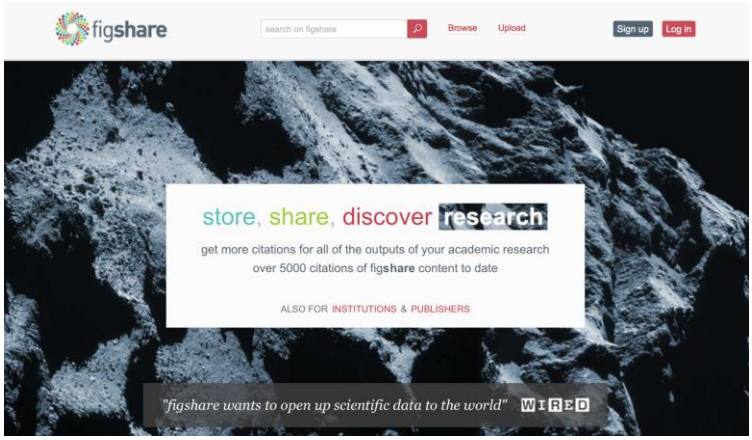
Finally, you pay for AWS Marketplace products on your AWS bill, which means you only have one procurement relationship to take care of, and all your costs can be monitored and metered through our budget management facilities, ensuring that you only spend what you intend.

1.6 Science as a Service

Some of our partners go one step further and provide managed services for science based on AWS but with their own portal access and user environments, completely abstracting the cloud management activities. Usually they offer pricing for services that are expressed in terms that are meaningful to the science, like € or $ per genome. This is referred to as Software as a Service (SaaS). For example:

- **DNAnexus**[7] is a company that offers genomic processing solutions via AWS. Their web portal access allows you to easily upload your data and execute customized pipelines for data analytics, while ensuring that your data management meets standards for encryption, privacy protection, and auditability that might be required by ethics boards or during compliancy audits. They achieve this by working in partnership with AWS to use our core services in specific ways. In doing so, they relieve users of the burden of worrying about IT and compliance, leaving them to focus on just the science.

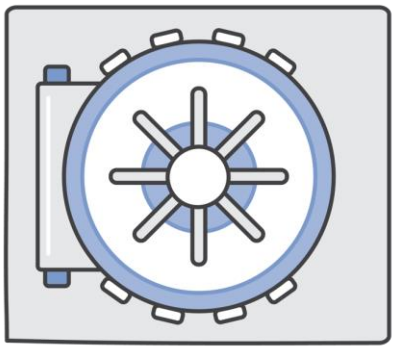
- **Figshare**[8] offers solutions to manage all your research outputs and make them available in a citable, shareable, and discoverable manner, including long-term storage (and archiving). Most journals (and funding bodies) are now mandating that wherever ethically possible, data and methods are shared in an open way to provide repeatability or falsifiability.



[4] You can see the Alces catalog of applications at http://docs.alces-flight.com/en/latest/apps/gridware.html .
[5] Amazon EC2 Spot market is an auction market for all our spare Amazon EC2 cores. We'll discuss it later in the handbook.
[6] See Intel's Lustre, BeeGFS, and other offerings in AWS Marketplace here: https://aws.amazon.com/marketplace/.
[7] www.dnanexus.com
[8] www.figshare.com

Figshare's service—running on AWS—enables academics, publishers, and institutions to easily adhere to these principles in the most intuitive and efficient manner.

1.7 Security



Cloud security at AWS is the highest priority. Helping to protect the confidentiality, integrity, and availability of our customers' systems and data is of the utmost importance to AWS, as is maintaining customer trust and confidence.

All AWS users benefit from data center architecture and network architecture built to satisfy the requirements of the most security-sensitive organizations. AWS and its partners offer hundreds of tools and features to help you meet your security objectives for visibility, auditability, controllability, and agility. This means that you can have the security you need, but without the capital outlay and with much lower operational overhead than in an on-premises environment.

Security is, nonetheless, a shared responsibility. AWS takes responsibility for securing the underlying infrastructure that supports the cloud, and you're responsible for anything you put in the cloud or connect to the cloud. This shared responsibility model can reduce your operational burden in many ways, and it may even improve your default security posture without additional action on your part.

As an AWS user you inherit all the best practices of AWS policies, architecture, and operational processes built to satisfy the requirements of our most security-sensitive customers.[9] AWS provides you with guidance and expertise through online resources, personnel, and partners. AWS also provides you with advisories for current issues, plus you have the opportunity to work with AWS when you encounter security issues.

AWS provides security-specific tools and features for network security, configuration management, access control, and data encryption. As an example of such tools and features, AWS Quick Starts are built by AWS solutions architects and partners to help you deploy popular solutions on AWS, based on AWS best practices for security and high availability. These reference deployments implement key technologies automatically on the AWS Cloud using AWS CloudFormation templates, often with a single click and in less than an hour. You can build your environment in a few simple steps, and start using it immediately.

AWS also offers access to additional third-party security tools to complement and enhance our customers' operations in the AWS Cloud. APN partners offer hundreds of familiar and industry-leading products that are equivalent to, identical to, or integrate with existing controls in a customer's on-premises environments.

[9] See AWS Security resources at https://aws.amazon.com/security/

Finally, AWS environments are continuously audited and have received certifications from accreditation bodies across the globe. In the AWS environment, you can take advantage of automated tools for asset inventory and privileged access reporting.

Security is our highest priority and we carefully document our approach—along with pointers to tools and security measures we recommend—In the AWS *Overview of Security Processes* whitepaper.[10]

1.8 Data Privacy

We deliver services to millions of active customers including financial services providers, healthcare providers, and governmental agencies, all of whom trust us with some of their most sensitive information.

We know users care deeply about privacy and data security. That's why AWS gives users ownership and control over their user content by design through simple but powerful tools that allow users to determine **where** their user content will be stored (for example, you can choose to deploy your AWS services and data exclusively in the London region, and AWS will not move customer content outside of London), to encrypt their user content **in transit or at rest**, and manage access to AWS services and resources for their users. We also implement responsible and sophisticated technical and physical controls designed to prevent unauthorized access to or disclosure of user content.

Customers maintain ownership of their customer content and select which AWS services process, store and host their customer content. **We do not access or use customer content** for any purpose other than as legally required and for maintaining the AWS services and providing them to our customers and their end users. We never use customer content or derive information from it for marketing or advertising.[11]

1.9 The Handbook and you

If someone else set up an AWS account for you, then you can probably skip ahead to Chapter 4. You'll learn about working with data on AWS, then go on to learn about computing, HPC, Jupyter notebooks, machine learning, and more. You'll skip chapters 2 and 3, where we talk about how to create an AWS account, how to configure it securely, and how to budget and set up billing alerts. If your lab director or IT administrator already did those things for you, then you can go straight to the good stuff.

If you need to set up an AWS account for yourself, your students, or the researchers you're supporting, then please do continue with chapters 2 and 3. They'll walk you through the process step by step.

If you are a PI or scientist at a research institution, and you're about to create an AWS account for yourself: now is a good time to **check if your institution already has a centrally administered AWS account**. If so, they may require that you sign up through them rather than go it alone. This can have many advantages for you:

- The central IT organization may provide support for your research work on AWS

[10] Which you can find here: https://d0.awsstatic.com/whitepapers/aws-security-whitepaper.pdf.

[11] More information on data privacy is found at https://aws.amazon.com/compliance/data-privacy-faq/

- You'll be in compliance with your institution's cloud policy
- Your institution may receive a volume discount on AWS services, stretching your research budget further

2 Setting up your AWS Account

The following sections detail the step-by-step processes through which you can move quickly to create, configure and use your AWS account.

Your first step is to create an AWS account and complete your registration for invoice-backed billing (no credit card required!) and the data egress waiver discount (no data download charges!). Sections 2.1-2.3 explains those steps. You can skip this if you've already completed registration for an AWS account.

What comes next?

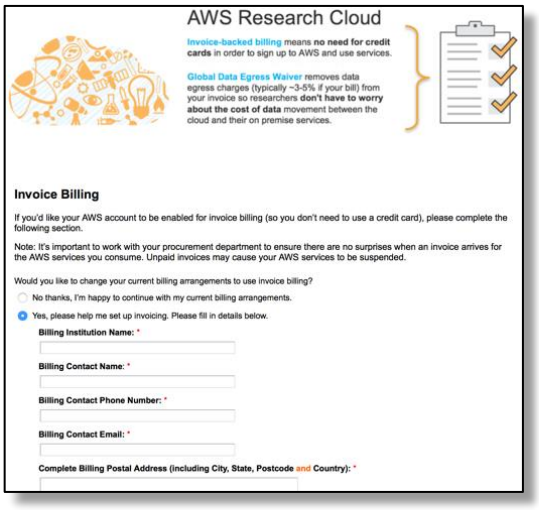
Starting with Section 2.4, we first want to guide you through some foundational things that we think you **should do** before you begin. These address basic security requirements to protect your data and your budget. Other steps cover installing tools that will generally make your life easier.

There are checkpoints at the end of each step to make sure you are ready to proceed to the next section. **We recommend following all of these steps**. At the end, you'll have an environment in the AWS Cloud that's capable of safely supporting some serious science.

2.1 Creating an AWS account and Registering for RCP Benefits

If you are a member of a public sector research or educational institution, you can receive additional program benefits:

1. **Invoice backed billing** – you won't need a credit card to pay for and manage your usage;
2. **Global Data Egress Waiver** –removes "data out" charges from your bill. Whilst these charges are usually very small (typically 3-5% of most users' bill) your bill is more predictable without them. There are additional qualification criteria (see chapter 4.4).



Which of the following describes you?

i. **If your AWS account was obtained via** a partner portal such as the Internet2 Net+ **portal** or a GÉANT portal, then your account has automatically inherited the above benefits and there's nothing more to do – congratulations, and please continue on to Section 2.4.

ii. **If your account was created directly with AWS**, then you'll need to go to https://aws.amazon.com/rcpstep2 to complete the registration process. You'll have to look up your 12-digit AWS Account ID first. Make sure the **account's email address is**

not a private address, but your university or work address.[12] Once complete, please continue on to chapter 2.4.

iii. **If you don't have an AWS account yet**, please follow the steps in Section 2.2 below to create an account now. The same distinction applies: if you create your AWS account via a partner portal, you will automatically obtain all the benefits. But if you create an account directly with AWS, the steps below will show you how to receive invoicing without using a credit card, and you will conclude by going to https://aws.amazon.com/rcpstep2 to apply to the data egress waiver discount program.

iv. If you are **not a member** of a public sector research or educational institution, then these additional RCP program benefits are not available to you. If you need to create an AWS account, please go to https://aws.amazon.com and click the yellow "Create an Account" button. Once you've completed the process, proceed to Section 2.4 to configure your new AWS account.

2.2 Creating a new AWS Account through a Partner portal (Internet2, Arcus, COMPAREX, SPARKLE)

Consult the table below to create an active account with invoice billing and the Data Egress Waiver benefit.

Table 2 - Process pathways for creating an AWS Account.

Partner Portal Account	or	Direct Account with AWS
Register with a partner		**Register with AWS**
US: Internet2 Net+ (operated by DLT) see 2.2.1 below. UK: Jisc Portal (operated by Arcus Global) see 2.2.2. EU: GÉANT portals (COMPAREX, SPARKLE, Arcus) see 2.2.3 below. ✓ Invoice-backed billing *included* ✓ Egress Waiver *Included* Potentially other additional terms, conditions and facilities.	**1**	Create a new Account at aws.amazon.com. Stop when you get to the stage for entering billing information. See section 2.3 below.
	2	**Discover your 12-digit Account ID** by looking at the top right corner of the screen in the Support Center Dashboard of your console.
	3	Record your 12-digit ID, and use it to: ✓ Request Invoice-backed billing ✓ Data Egress Waiver at https://aws.amazon.com/rcpstep2

[12] This is required for invoice-backed billing or the Data Egress Waiver. If you need to change an existing account's registered email address, you can follow the procedure here: https://aws.amazon.com/premiumsupport/knowledge-center/change-email-address/

2.2.1 Internet2 Net+ Operated by DLT (US users only)

The Internet2 NET+ Amazon Web Services was developed by a group of five universities through the NET+ Service Validation process and has additional enhancements to support enterprise usage and broad adoption across campus. This NET+ program provides significant technical and procurement benefits, and enables campuses to leverage AWS using a best-in-class offering.

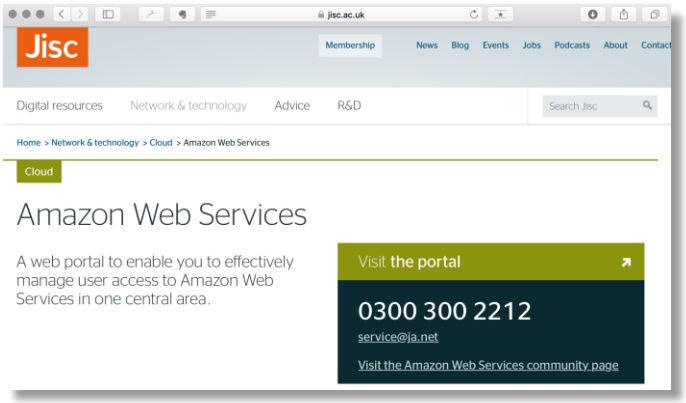
Features included:

- A community negotiated Business Associate Agreement (BAA) for HIPAA workloads containing patient health information (PHI);
- Increasing discounts of 3-5%, based on community usage;
- Detailed and granular billing with a variety of payment options providing visibility to campus AWS usage;
- 100+Gb/s of privately peered capacity to the Internet2 Network;
- Use *InCommon* credentials with the DLT Portal to request and transfer AWS accounts.

More information about getting an account via this method is available from the Internet2 Net+ page.

2.2.2 Jisc's Portal Operated by Arcus (UK Academic Users Only)

Jisc's web portal—operated by Arcus Global for academic customers in the UK—enables customers to procure AWS Cloud services with enhanced benefits (like the Global Data Egress Waiver) and to manage user access to AWS in one central location. It was developed specifically for research and education in collaboration with AWS.



The portal provides institutions with core administrative facilities and offers additional benefits to help manage budgeting and control costs:

- Capability to set budget limits for individual user accounts or departments.
- View and authorize all your institution's user accounts through one central portal.
- Monthly invoicing means you no longer need to use credit cards for payment.
- Administrators can retrieve service usage information.
- Volume discounts through aggregation across multiple educational institutions.

- It's possible to prepay via the portal, which may be needed in order to meet your grant's fiscal requirements that spending is complete by a set deadline. You can pay in advance and draw down against the balance as you go.

You can apply for an account here: https://www.jisc.ac.uk/amazon-web-services.

2.2.3 GÉANT Portals for European Researchers

GÉANT - Europe's leading collaboration on e-infrastructure and services for research and education will soon enable thousands of academic institutions in 36 European countries access to use AWS through a procurement contract vehicle.

The offering will be available in 2017 through Europe's National Research and Education Networks (NRENs) and selected AWS Resellers.

The procurement contract vehicle streamlines the procurement process for institutions, eliminating the need to conduct individual public procurement tenders. NRENs and AWS Resellers can offer customized research and education solutions that include: purchase order and prepaid billing, federated identity management, and a portal with administrative and reporting features especially designed for this community. All of this is built on underlying AWS infrastructure services.

Through COMPAREX:

1. Contact local COMPAREX Account Manager. Please see 'Contact Us' via comparex.co.uk/geant
2. Institution and COMPAREX develop and complete GÉANT call-off agreement with COMPAREX.
3. The flow of the initial order that comes afterwards will depend on the fact whether customer has an AWS account or not. In order to register a customer account or migrate it, COMPAREX needs some basic user information.
    - Option 1: Customer already has an AWS account. This account is already populated with the relevant information. Transfer of AWS account information to COMPAREX to be able to manage and offer the AWS services. This will be done by email invitation and customer acceptance.
    - Option 2: Customer does not have an AWS account. COMPAREX will initiate the customer enrolment, and will run a full onboarding process, gathering any necessary data and providing to the nominated AWS Account Administrator. Upon creation, COMPAREX will forward the login credentials to this email address.

COMPAREX will allocate dedicated resources to support every customer, including:

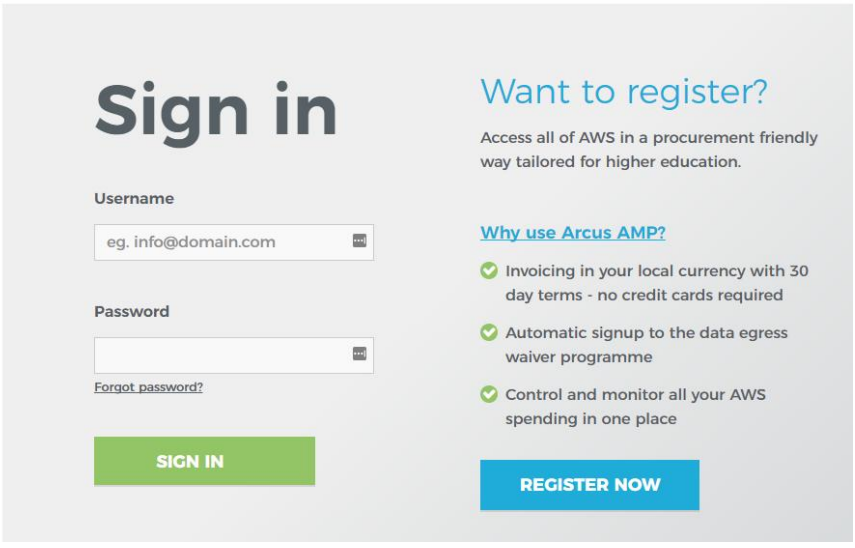
- Key Account Manager
- Cloud Technical Specialist
- Amazon Web Services Specialist Team

This team will guide institutions through the onboarding process, including guidance on AWS licensing structures, in order to identify the best value configuration for the institution's circumstances.

Through Arcus:

The Arcus AWS Management Portal (AAMP) provides an easy way to sign up for a new account, or bring in an existing account, under the GÉANT IaaS framework. Sign up today at https://aamp.arcusglobal.com or contact us at support@arcusglobal.com or +44 (0)1223 911 841.

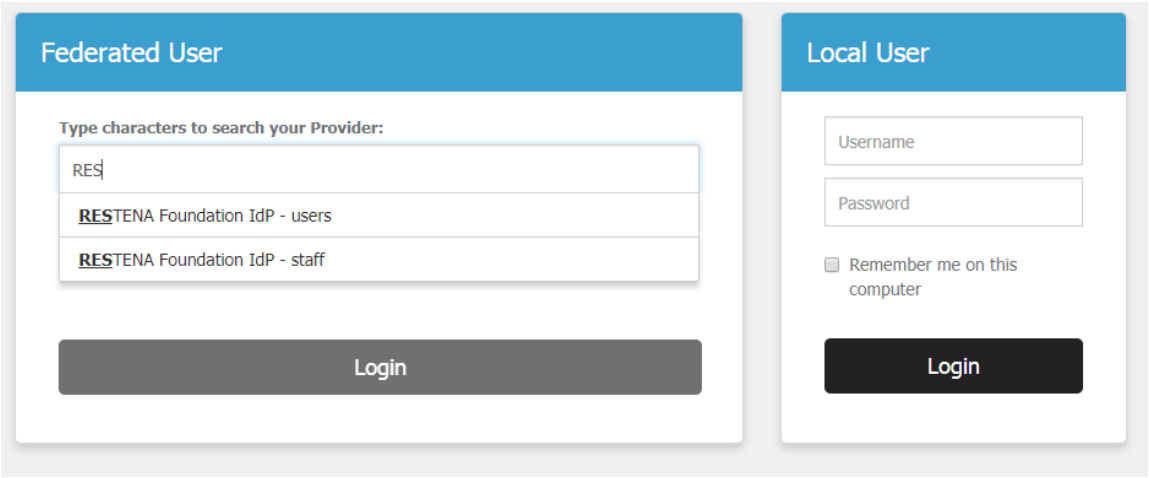
- If you go to https://aamp.arcusglobal.com you can use the simple registration process to guide you through sign up.
- The process even offers you an easy way to enter your existing 12-digit AWS account ID to speed up the process of bringing an existing account on board without down time or impacting your current configurations.
- Don't worry if you don't have an account as we will create one for you if you leave this blank!
- When your accounts are live and in the system, you will be able to see how your account spending tracks against the budgets that you set
- You will also be able to look at the daily spend across your accounts to get a clearer picture of usage and identify any spikes or unexpected activity
- The forecast function provides a view of potential future costs
- Combined with simple, per portal user, configurable spend threshold alarms and notifications you can easily keep track of costs and make sure that you do not go over budget!

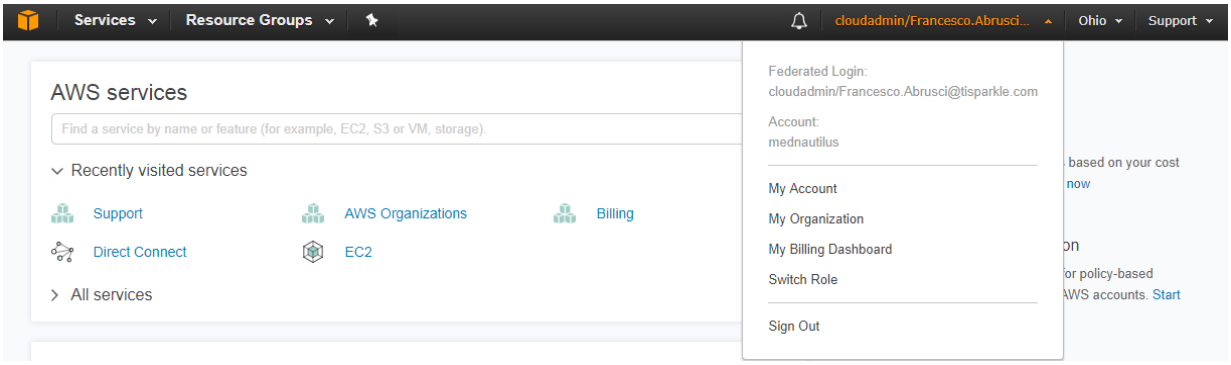


Through Sparkle:

In order to use Sparkle's AWS services your organization (NRENs, education & research institutions) must have a call-off agreement, as defined within the GÉANT framework agreement. Please contact our commercial representative at GEANT-CLOUD@tisparkle.com to register your organization. The steps necessary to provision user access to AWS cloud services are:

1. The organization's designated representative will request the creation of one or more AWS accounts via email to GEANT-CLOUD@tisparkle.com. In the near future the Account registration and provisioning will be performed through Sparkle's cloud automation portal, under development.
2. Your Identity Provider (IdP) will enable for your organization to allow Edu Gain federated access for your designated users. The IdP information (meta-data) for your organization will be configured in Sparkle's authentication servers.
3. The administrator at the user's organization must authorize users to access the IaaS resources. The administrator first needs to consider the structure of the different roles within the AWS accounts (administrator, budget holders, project head, users, etc.)
4. Now users can access Sparkle's authentication portal for AWS at http://aws.cloud.tisparkle.com/.
    a. First, the user selects her Identity Provider and clicks Login.
    b. The Login page appears. She uses her organization's authorized login and password.
    c. She is logged into the proper AWS Console with associated IAM roles. (A request is submitted to her administrator if she is not yet authorized.)

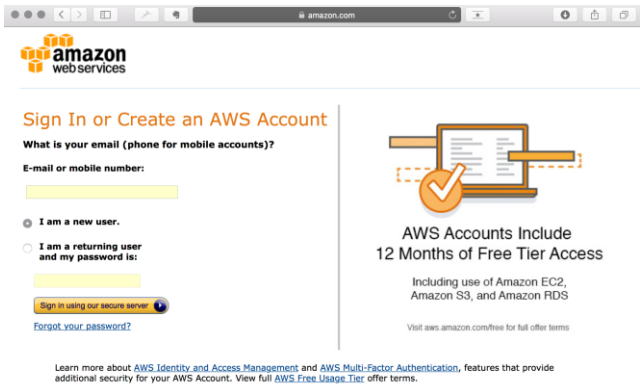




2.3 Getting a new AWS Account directly with AWS

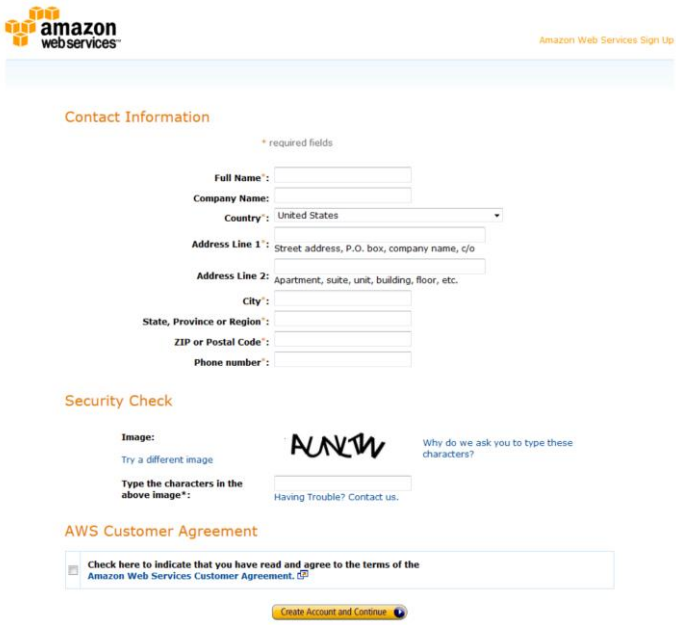
Follow the steps below to create a new account.

At the last step, you can stop before supplying a credit card, then register with the AWS Research Cloud Program and switch to invoice-backed billing. See Table 2 above.

To get started, follow the link below to the **AWS registration page**: [AWS Registration]	
Enter your work email address. It's important to note the following: 1. You must use your **institutional email address** (@*.ac.uk or @*.edu, for example). Please don't use personal email addresses that fail to identify your institution or employer. 2. **You cannot use the same address twice** for any AWS account. But a new AWS account will be completely separate from your Amazon online shopping account with the same e-mail address, with separate passwords, multifactor authentication settings, and payment methods. **Finally**, be sure to select "I am a new user."	

Next, you'll need to provide your personal name and the name and contact details for you **at your institution**.

This will be important information for later on, when you register for the Research Cloud Program and ask for Invoice-backed billing.



When you've finished reviewing the terms of our customer agreement, check to accept and click "Create Account and Continue."

Your account will be created using the email address you typed in.

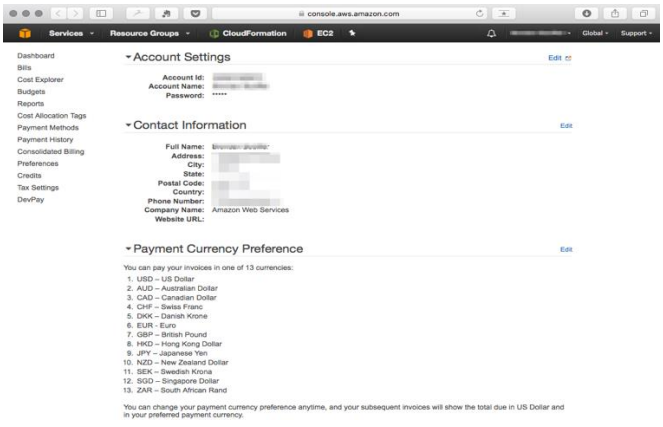
To avoid completely using a credit card, you can **stop** the account creation process at this point**.**

Finally, to find your account ID, close the sign-up page and log in to the AWS console: https://aws.amazon.com/console

with the account information you just provided.

1. Click on **your name** in the menu bar (top right);
2. Select "My Account"
3. You'll find your 12-digit account ID in the first section of the dashboard;

Record this number below.



Finally, go to https://aws.amazon.com/rcpstep2 and complete the registration.	
You will be contacted shortly to confirm invoice billing and data egress waiver discounts.	

Checkpoint

Have you **created an AWS account using your institutional email address and contact details**?	☐
Have you **found your 12-digit root account ID number**? If so, record it here:	☐
Have you **registered for invoice back billing and the Data Egress Waiver**?	☐

2.4 The Root Account

The first AWS account you create (either with your credit card or invoice backed) is what we refer to as your **root account**. We recommend that the root account's registered email address be a shared account, if possible. Having more than one person in a position to react to unforeseen events is good practice.

We also strongly recommend that you **not** use this master account for anything other than billing. Instead, we recommend that you set up privileged single-user accounts for day-to-day usage of AWS. This is known as an **AWS Identity and Access Management (IAM) login** (see below). IAM logins are useful since they protect your billing information, and you can create (and delete) new ones at will. They also allow you to create restrictions

on the AWS services that can be launched. It's best to get into the practice of creating IAM logins for everyone (including yourself) and preserve the root account console for budgeting and billing activity only.

As an example, suppose that you have a lab with three post docs: one computational biologist who needs access to robust computing resources, and two bench scientists who only require access to storage and a few web apps. In this case, you can create:

- One IAM login for yourself that is able to view billing data and usage reports
- One IAM login for your computational biologist that is able to launch a compute cluster via a command line interface or the AWS Web Console. It also has read, write, and delete access to all your data resources on our object storage service, Amazon S3 (we will cover Amazon S3 later).
- Two IAM logins that are only able to read and write (but not delete) data on S3. Since they only need to *use* the web applications, and not *manage* any underlying computational resources, they don't need IAM access to compute resources.

In the following section we will walk through an example of setting up IAM logins for yourself and for others.

2.5 Creating IAM Logins

You create IAM logins and set the permissions for your AWS account using the IAM console. We have a whitepaper that describes several scenarios in detail[13], but for now, we'll guide you through creating an Administrator IAM login for yourself, as well as any additional user logins for people who need to access resources that you control, such as researchers in your group or collaborators.

First, open your browser and navigate to the AWS Management Console (https://aws.amazon.com/console) and log in using your **root account** credentials.

Under the AWS Services menu, look for "Identity and Access Management."

In the navigation pane, choose "**Users**."

Click on the "Create New Users" button at the top and begin entering login names for your users (including a login for you).

[13] http://aws.amazon.com/whitepapers/setting-up-multiuser-environments

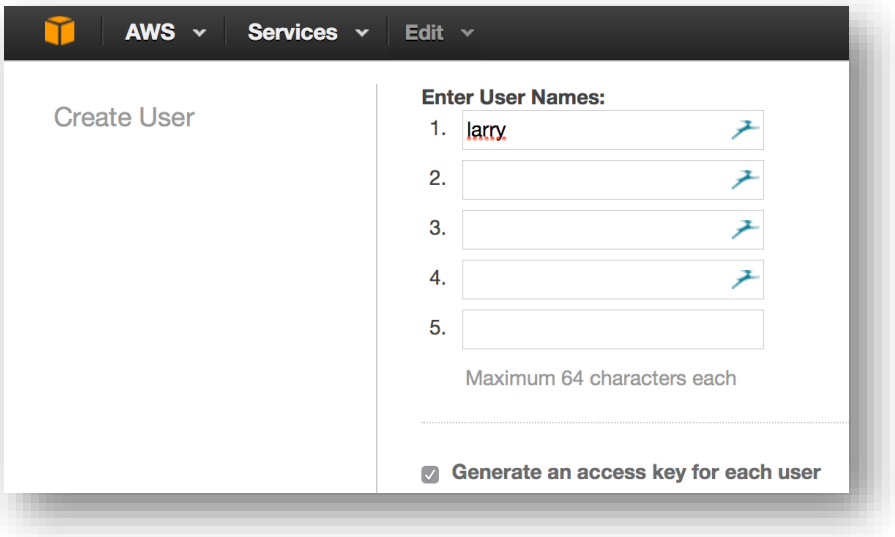


Figure 9 – Creating Users with IAM.

Once you've created these users, you can (and should) do a few key things for each of the people you're giving access.

1. **Download their credentials** and pass these on to them. The credentials are important digital keys that allow your users to use client-side scripts and API tools (scripts on their laptops, for example) to create or remove resources in the cloud. They're also essential for accessing any of the AWS resources via the AWS cli, such as SSH access to EC2 instances, the Amazon S3 management tools and more. Each user is identified by their own credentials, so be careful to share them discreetly. Also, **be aware** that you are only able to download these credentials once (i.e., if you lose them for some reason, you'll need to generate a completely new pair for your forgetful IAM user).

   We recommend that you create a file **`~/.aws/credentials`** on your local computer (this works on Mac and Linux) containing the Access Key and matching Secret Key as follows:

```
aws_access_key_id = SDFxxxxxxxxxxxxF89
aws_secret_access_key = 345xxxxxxxxxxxxxxxxx3454/sdfsxxxxxxxx/8
```

   This file will be used by the AWS command line tools (see below for more on that) and by tools like CfnCluster (see Chapter 7) that use API requests to execute your wishes. Placing your access and secret keys in a credentials file and/or assigning them to variables ($ACCESS_KEY_ID, $SECRET ACCESS_KEY) in your .bashrc_profile environment configuration is recommended best practice, rather than placing them directly in scripts. Also, we recommend not disclosing these

keys in any shared documents, scripts (i.e. on GitHub), or presentations. (We have even x'd out the real keys in this example.)

2. **Individually allocate some permissions** to each IAM user by clicking on the user name in the list under the "Users" pane and selecting the "Permissions" tab.
    a. You add permissions by "attaching policies" to a user.
    b. While you can create a custom policy, you probably want to use existing policies from the (quite exhaustive) catalog available, since these cover most scenarios you're likely to face while you're getting started.
    c. Almost all your team will probably need an Amazon EC2 policy (for starting instances) and an Amazon S3 policy (for creating buckets and storing data).
    d. For your own account, you should add the "AdministratorAccess" policy to make sure you retain full control.
    e. For others' accounts, we suggest following a principle of least privilege, only adding policies for resources on an as-needed basis. Once users have earned your trust, you can extend their privileges if you need to.
    f. For multiple users with the same access permissions, you can assign them to **Groups**, which can make managing the permissions for those individuals much easier.

3. **Generate passwords** for some users. If an IAM user requires access to the AWS Management Console (you'll definitely need this for your own IAM login), then you'll need to generate them a password. This is done under the "Security Credentials" tab when you click on an individual user login in the IAM console and is in the section called "Sign-In Credentials."

2.5.1 Naming Your Console

In the IAM console, you'll notice that you have the opportunity to customize the URL or link that your IAM users will use to sign in to the AWS Management Console. It's a good idea to do this now so you can circulate something to your team that is easy to remember. It needs to be unique, so common words like "lab" are probably already taken. You'll need a little imagination (e.g., try something like "Einstein-Lab").

Checkpoint

Have you **created an IAM role for yourself** with Administrator privileges?	☐
Have you **downloaded your IAM user's security credentials** (Secret Key and Access KeyID) and stored them safely?	☐
Have you **customized** the name of your console login address?	☐

Have you **created a password** so you can log in to the console?	☐

2.5.2 Billing and Cost Permissions for Your IAM Login

In order to be able to manage billing and budgets from the IAM login you just created for yourself, a few final steps are necessary. If you don't carry out these steps, you'll only have access to your billing and budgeting from the root account. We don't recommend this, since as a best practice you should not be logging in to the root account regularly.

2.5.2.1 Enabling IAM Control over Billing Features

To activate IAM user access to the Billing and Cost Management console:

1. Sign in to the AWS Management Console with your root account credentials (the email address and password that you used to create your AWS account). Don't sign in with your IAM user credentials.
2. On the navigation bar, choose your account name, and then choose My Account.
3. Next to "**IAM User Access to Billing Information**," choose "**Edit**."
4. Select the "**Activate IAM Access**" check box to activate access to the Billing and Cost Management pages.
5. You can now use IAM policies to control which pages a user can access.

By default, only IAM logins with either full access (Administrator privileges) or specifically itemized access to the billing functions will be able to access the Billing and Cost Management features.

Since your own IAM login (which you created above) has Administrator privileges, that account will automatically have permissions to access the billing and cost management functions.

2.5.2.2 Defining a Lab Policy to Control IAM Access to Billing Features

It's likely that other people in your lab or research group will need to access Administrator-*like* functionality in order to do the day-to-day work supervising others, but you'd like to restrict their access to billing details.

To cover a number of permutations around these preferences, we've placed some detailed sample IAM access control policies on the web here: http://docs.aws.amazon.com/awsaccountbilling/latest/aboutv2/billing-permissions-ref.html.

You can apply these to IAM users in your group. To do this, we'll show you how to create your own managed policy so you can create and manage it centrally and apply it to several users, as you choose.

1. Navigate to the page linked above and find the policy that best matches your needs. There are nine examples to pick from, and they're described on the page.

2. Open your AWS Management Console (http://aws.amazon.com/console) and go to the IAM console (under services).
3. Select “Policies,” choose “Create Policy,” and select the “Create Your Own Policy” option.
4. Paste in the policy text to the large editor window.
5. Give your policy a name and enter some description text that’s meaningful to you and others in the lab, like in the example below. The name must only be text, with no spaces or special characters. Putting the name of your lab or group at the front of the policy name can be helpful for finding it later.
6. Once you create this policy, you now have a privately managed policy you can reuse each time you need to apply it to a user as people come and go from your research group.

Figure 10 - Creating your own managed policy.

2.5.2.3 Applying Your Lab Policy to Specific Users

Now, you just need to **apply this policy to a specific user** (one or more):

7. In the IAM console, choose “users” from the left-hand menu.
8. Select the user you want to apply the new permissions to, and click on their “Permissions” tab.
9. Click on “Attach Policy.”
10. Find your policy in the (very long) list by searching for your lab name in the search box.
11. Select the policy and click “Attach Policy.”

If you ever need to modify this policy, you can do so from the “Policies” area of the IAM console, and this means you’ll only have to modify it once to have the changes apply everywhere.

Checkpoint

Have you **enabled IAM access to billing** so you (and potentially other Administrators you permit) can monitor billing and manage budget features?	☐
Have you **created a lab policy to restrict access to billing features** for specific users?	☐

2.5.3 Permissions Structure for a Research Team

The diagram below illustrates an example permissions structure for a research group.

The green person is the Principle Investigator (PI) or administrator and she has full access to the AWS account. She creates the three grey users, each of whom have access to different resources.

For example, the teaching assistant can fully use all the services such as compute and machine learning, but he can't see the billing information. The grad students can create compute instances (which incur costs), but they can't access data buckets that belong to other group members. Finally, undergrad students—who only need to do some course work on AWS—can access only the compute instances that were created by the teaching assistant and are tagged with the "student" label, but they cannot create new instances (which would incur extra costs).

It's possible to create quite complicated permissions structures if you need to, but for now, we'll settle with the structure you've just created.

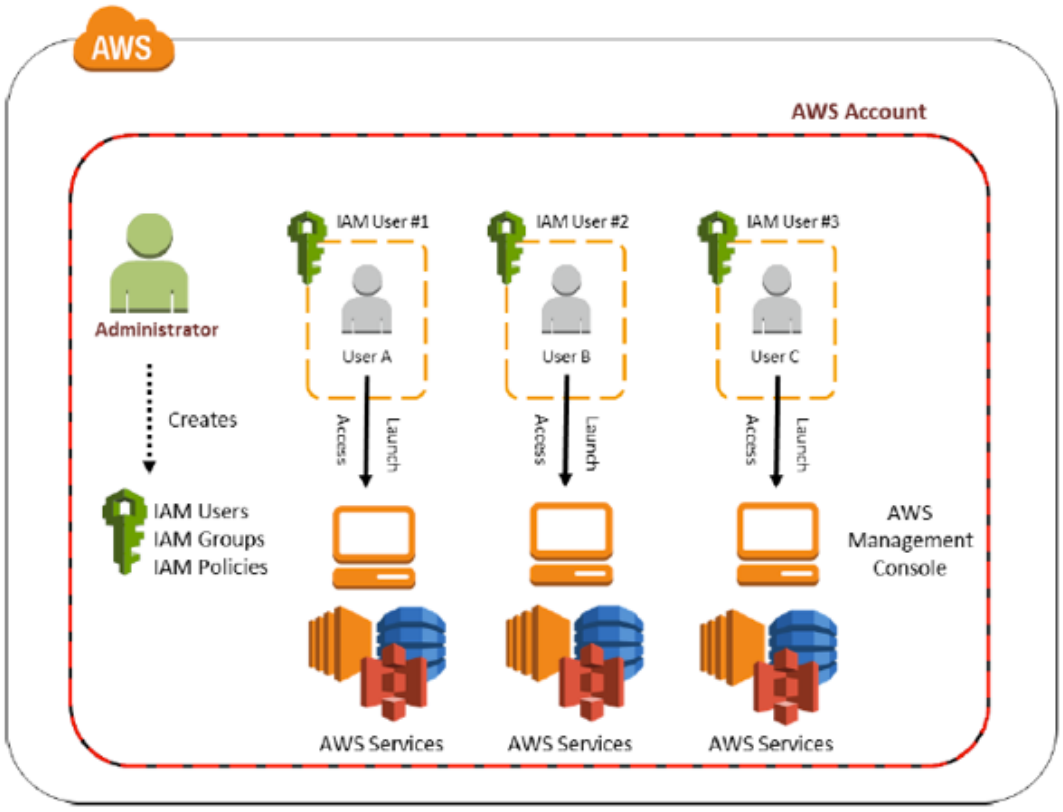


Figure 11 - Example of a permissions structure for a research team. Administrator (PI): can do anything. User A (Teaching Assistant): can use any AWS services, but cannot access billing information. User B (grad student): can create EC2 instances, can access only the data in his personal S3 buckets. User C (undergrad students): can access (but not stop or create) EC2 instances tagged as “student”.

Additional tutorials and documentation are available online:

- Introduction to IAM: http://docs.aws.amazon.com/IAM/latest/UserGuide/introduction.html
- Our recommended best practices: http://docs.aws.amazon.com/IAM/latest/UserGuide/best-practices.html

2.5.4 IAM groups

User privileges can also be managed by creating IAM user groups and attaching policies to the group. Each user who is a member of the group then automatically inherits the associated policies. If you edit or remove the group policies in the future, the changes will apply to all group members. You can also add or remove IAM users to and from groups when their job roles change. For example, in the previous example we could have created groups named “TAs”, “grad students”, and “undergrad students”, and then added each user to the relevant group(s).

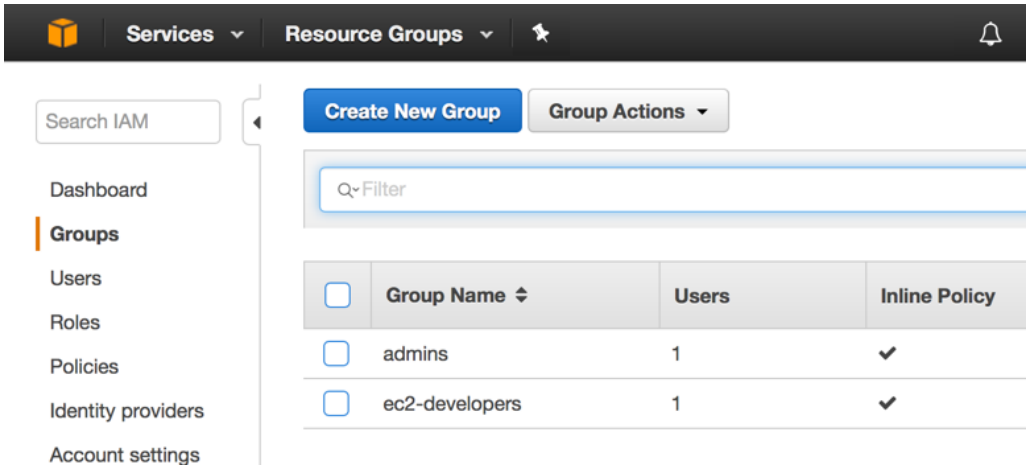


Figure 12 - Creating a New IAM Group.

2.5.5 Multi-Factor Authentication

Now that you've created an IAM login for yourself that can manage all your cloud permissions (this is referring to your Administrator privileges), it's time to lock down your **root account** just a little more.

We recommend that you configure Multi-Factor Authentication (MFA) to protect your AWS resources. MFA adds extra security because it requires someone logging in to enter a unique authentication code from an approved authentication device or SMS text message when they access the AWS Management Console or services.

You can add MFA to any or all your IAM logins, but at the very least, it's important that you **enable MFA for your root account**, since this is your last resort if anything goes wrong and, as you know, your root account has complete privileges to everything in your cloud account (including your billing details).

To set up MFA, follow the steps in the MFA guide, which will also help you choose what secondary method you'll use for your MFA:
http://docs.aws.amazon.com/IAM/latest/UserGuide/id_credentials_mfa.html.

Checkpoint

Have you **configured MFA for (at least) your root account**?	☐
Have you **logged out from the root account** in the console? From now on, you should use your **IAM login only**.	☐

*From now on, you should use your **IAM login only**.*

2.6 AWS Organizations

So far we have set up your AWS account so that multiple people can use it, each with their own set of permissions. There is a higher-level framework to control access policies and billing across multiple AWS accounts: AWS Organizations. This is useful at the departmental or campus level. If your campus uses AWS Organizations, they may request that you "link" your AWS account to the Organization. You can find more information at https://aws.amazon.com/organizations/ . For a regular research group it's probably easier to keep track of multiple research projects using IAM users, "tags", and separate "budgets" (see Chapter 3) within your AWS account – rather than using multiple accounts and an Organization.

2.7 Setting up SSH Access Keys

It's likely that you'll soon be creating Amazon EC2 compute instances to run things on and also very likely that you'll want to log in to these instances using SSH.

This requires registering an "SSH key pair" with the Amazon EC2 service, which we explain below.

There are a few things to know before you start:

- **Keys belong to AWS Regions** – So carefully select your favorite region from the pull-down menu in the top right corner of the console. We won't distribute your keys to the other regions (remember we never copy your data without you actively being involved).
- **You can use an existing SSH key from your personal desktop** – This is probably the easiest choice, since it'll save you a lot of typing as time goes by. If you use a Mac or Linux desktop, check in your ~/.ssh directory to see if you have an existing key pair. If you don't, you can create one by running the command **`ssh-keygen -t rsa`**. Doing this just once will create a key pair you can use all the time.
- **Once generated we only store the PUBLIC key in the cloud** – This is the nature of public key encryption. Only *you* have the private key and so only *you* (or those with access to your account on your desktop) can then access the resources that are secured with these keys. So again, be careful with these keys.
- You may want to use **separate SSH keys for each IAM user** with access to your EC2 service. This will prevent users from interfering with each other's EC2 instances, and helps control who owns what.

2.7.1 Generating or Importing Your SSH Keys for Use in AWS

To get set up with your Amazon EC2 SSH keys, log in to the AWS Management Console (https://aws.amazon.com/console) using your IAM login credentials and navigate to the Amazon EC2 console, which looks like this:

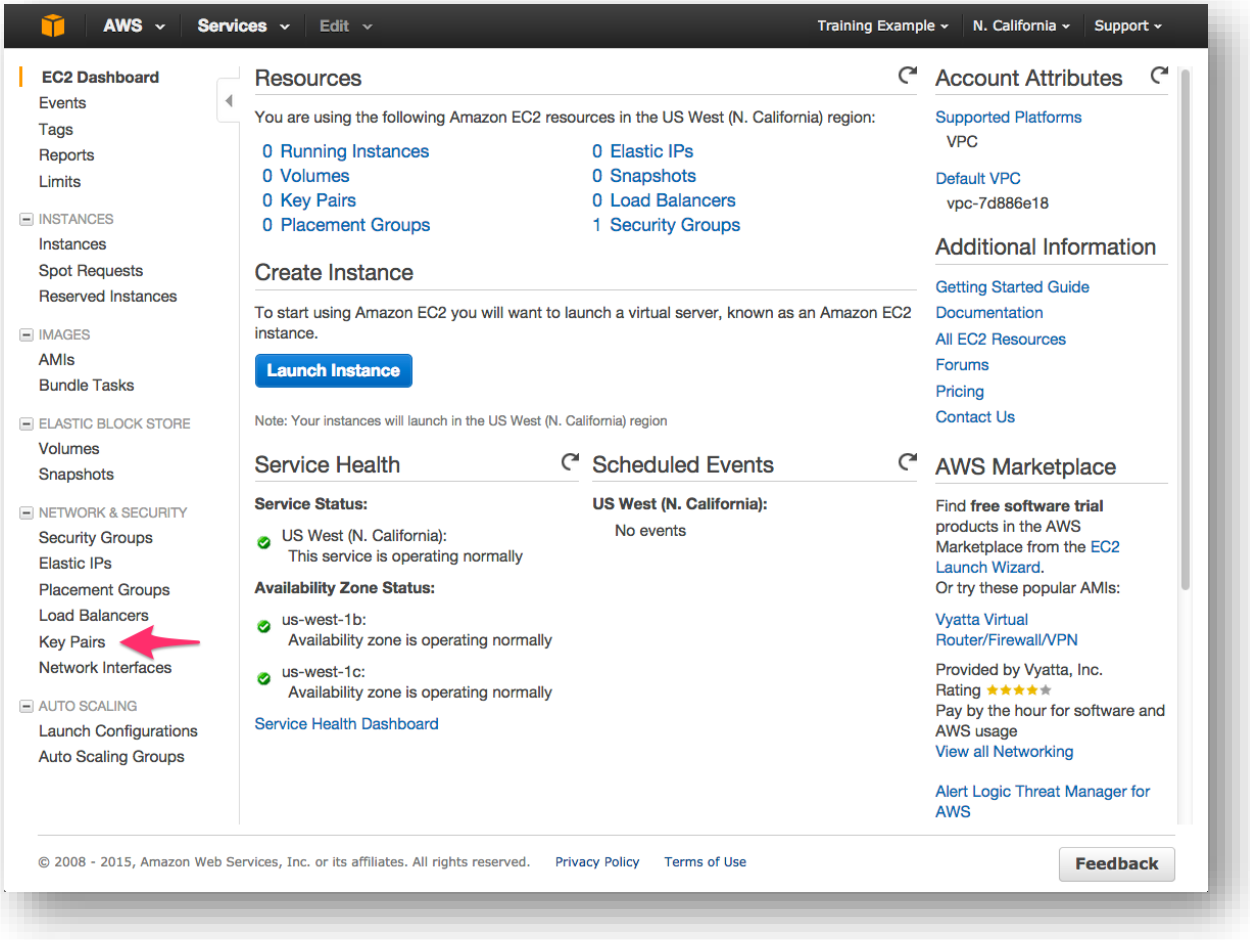


Figure 13 - EC2 Dashboard in the AWS Management Console.

Inside the Amazon EC2 console, click on "**Key Pairs**" under "Network & Security." You'll find something like this:

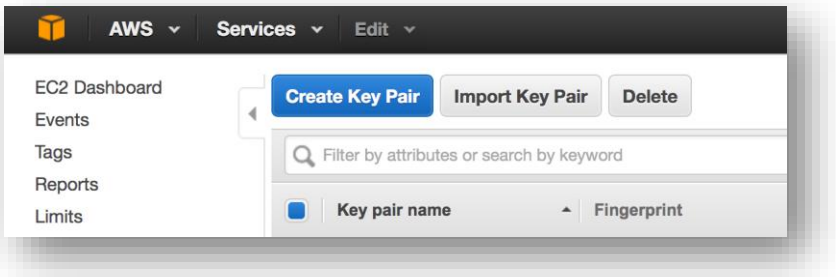


Figure 14 - Import an existing key or create a new one.

Now you need to choose between importing an existing key from your day-to-day desktop or creating a new key that you'll use for AWS separately.

To import an existing key from your desktop (Mac or Linux):

1. If you use an existing key pair, click on “Import Key Pair” and paste in the content of your PUBLIC key, which you’ll find in your **`~/.ssh`** directory on your desktop in a file probably named something similar to **`id_rsa.pub`**.
2. You’ll need to name your key—something meaningful will help a lot.
3. When you use SSH to connect to your instances that are secured with these credentials, you don’t need to do anything special, and you’ll be securely admitted entry to your systems, since the Amazon EC2 instance will be recognized by your key. For example:

```
ssh ec2-user@1.2.3.4
```

To generate a new, separate key: (Mac, Linux or Windows users):

1. Click on “Create Key Pair.”
2. Give it a name—again, best use something meaningful.
3. AWS will send to your browser a download file with a *.pem extension. You’ll need to **store this file somewhere safe** since it’s not only *your* credentials and your private encryption key, it’s also irreplaceable—if you lose it, AWS cannot give you another copy, since we don’t retain your private key.[14]
4. From now on, when you use SSH to connect to instances that are authenticated using this key, you’ll need to tell SSH which *.pem file to use by adding “-i /path/to/my/keyfile.pem” to your SSH command line, for example:

```
ssh –i ~/.ssh/my-ec2-private-key.pem ec2-user@1.2.3.4
```

The full details about your keys are in the *Amazon EC2 User Guide*[15].

2.7.2 Managing SSH Access for Windows Users

Since Windows versions prior to Windows 10 do not natively include an SSH client[16], the following instructions explain how to connect to your instance using PuTTY, a free SSH client for Windows. You can download PuTTY at http://www.chiark.greenend.org.uk/~sgtatham/putty/.

As you generated a key pair in the previous section, you’ll have a “.pem” file stored on your desktop (in a safe place). PuTTY does not natively support this file format; however, PuTTY has a tool named PuTTYgen, which can convert keys to the required PuTTY format (.ppk).

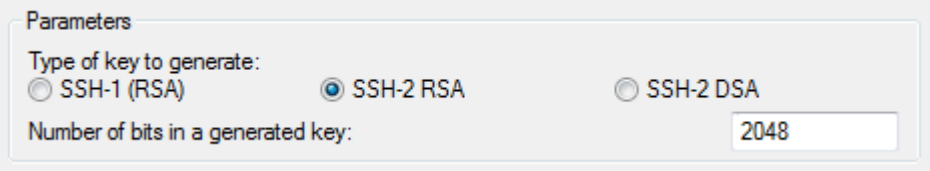
You must convert your private key into this format (.ppk) before attempting to connect to your instance using PuTTY.

To convert your private key:

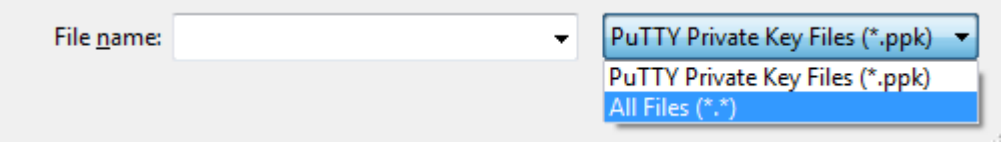
- Start PuTTYgen (for example, from the **Start** menu, click **All Programs > PuTTY > PuTTYgen**).

[14] If you lose a key, you can of course create a new one. But you will lose access to any EC2 instances associated with the lost key.
[15] http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ec2-key-pairs.html
[16] If you have Windows 10, you’re in luck. This version of Windows includes a BASH shell, allowing you to use the Linux SSH instructions and forget all about PuTTY.

- Under "**Type of key to generate**," select "**SSH-2 RSA**."



- Click **Load**. By default, PuTTYgen displays only files with the extension .ppk. To locate your .pem file, select the option to display files of all types.



- Select your .pem file for the key pair that you specified when you launch your instance, and then click **Open**. Click **OK** to dismiss the confirmation dialog box.
- Click **Save private key** to save the key in the format that PuTTY can use. PuTTYgen displays a warning about saving the key without a passphrase. Click **Yes**. **Note:** A passphrase on a private key is an extra layer of protection, so even if your private key is discovered, it can't be used without the passphrase. The downside to using a passphrase is that it makes automation harder because human intervention is needed to log in to an instance or copy files to an instance.
- Specify the same name for the key that you used for the key pair (for example, my-key-pair). PuTTY automatically adds the .ppk file extension.

Your private key is now in the correct format for use with PuTTY. You can now connect to your instance using PuTTY's SSH client.

For more on how to use PuTTY with AWS, see http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/putty.html.

2.7.3 Default usernames on Amazon EC2 Instances

When a new Amazon EC2 instance is spun up from standard AMIs, which you'll find in AWS Marketplace, or our own repositories, you'll need to log in initially using standard, predefined login names specific to the AMI to used.

Some of the most common login names:

- For an Amazon Linux AMI, the user name is ec2-user.
- For a RHEL5 AMI, the user name is either root or ec2-user.
- For an Ubuntu AMI, the user name is ubuntu.
- For a Fedora AMI, the user name is either fedora or ec2-user.

- For SUSE Linux, the user name is either root or ec2-user.

Otherwise, if ec2-user and root don't work, check with the AMI provider. Once you are logged in to the EC2 instance, you can create additional Linux users (if it's a Linux EC2 instance) or Windows users (if it's a Windows instance). But the default ec2-user account suffices for most straightforward use cases.

Checkpoint

Have you **created an SSH key pair for your IAM login**?	☐
Have you **saved your *.pem file to a known and safe place on your desktop** OR **imported an existing SSH public key into AWS**?	☐

3 Budgeting

There are a few key tools to help with budgeting and controlling your costs on AWS.

3.1 Setting Budgets in your AWS Account

Working in the cloud with a pool of **variable-cost infrastructure isn't hard** if you keep tabs on your usage with the AWS Budget tools—preferably ahead of time.

In your AWS Management Console, you'll find a menu at the top right that contains an item called "Billing & Cost Management."

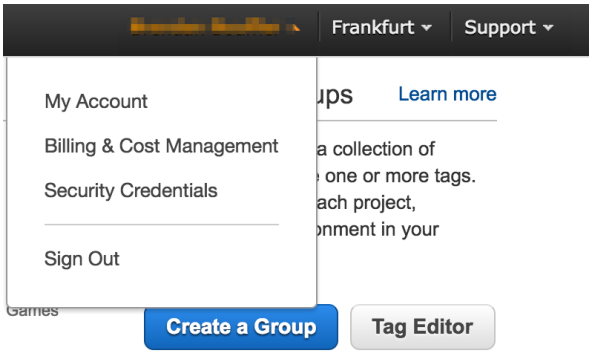


Figure 15 - AWS Billing and Cost Management.

In the AWS Billing and Cost Management dashboard are a number of tools that give you a great degree of control of your spending at a very granular level, but there are three specific panels that we think you need to pay attention to:

1. **Main Billing Dashboard** – Shows the total spend in the previous month, in the current month to-date, and projected for the end of the current month. There's also a high-level breakdown of services you're using, based on the service type (i.e., for storage, for Amazon EC2 compute, etc.).
2. **Cost Explorer** – A graphical tool to help you dig deep into the finer details of where you're spending your money.
3. **Budgets** – A way to set alerts that will raise alarms (for example, via email) when your actual spending (or projected spending) exceeds metrics you set.

In the following sections, we'll drill into these a little, so you know where to find key details, and we'll **guide you through setting up your first budget** so you can make sure your funds are safe.

Before we get too far into the details, let's discuss how to establish a budget baseline.

3.1.1 Establishing a Budget Figure

Imagine you have a two-year grant for $10,000 to apply to computing resources. You could buy a workstation, but given that you are human and you're only likely to spend at most eight or so hours a day in front of it, you realize that the workstation will spend 80% of its life running a screensaver or turned off.

So you decide to use the cloud, and you know that some days your use will be minimal, other days it'll be more serious, and sometimes you'll try different types of resources.

How to come up with a budget figure? The simplest approach, while you're getting used to your own workflows, is simply to divide the $10k by 24 months. This gets you a consumption figure of around $400 per month.

You'll very likely discover that there are some months you'll use a lot less than this (teaching duties, conferences, and meetings). That's okay, because unspent remains in reserve for next month when you're back and you want to pick up the pace or take advantage of this savings windfall and double the size of the simulation you run.

3.1.2 Billing & Cost Management Dashboard

The **Billing & Cost Management Dashboard** shows the total spend in the previous month, in the current month to-date, and projected for the end of the current month. There's also a breakdown of services you're using, split by service type.

The dashboard screen below depicts a fairly typical example of a researcher who's busy developing a suite of HPC software. Her main activity is in Amazon EC2 compute instances, but she has also chosen to subscribe to Business Support and Developer Support. Her data transfer costs are minimal, and her storage charges are so small that they disappear into "other."

This researcher spent $551.80 in her last month's billing, and is on track to spend $771.90 this month–based on her current usage and projected activity.

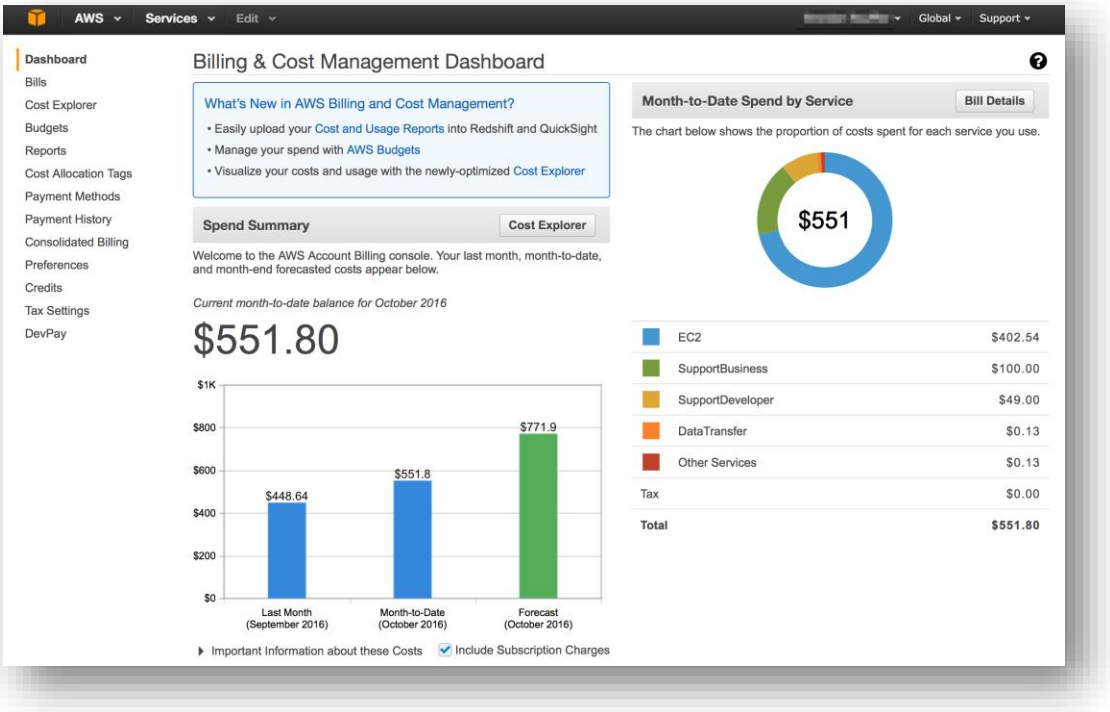
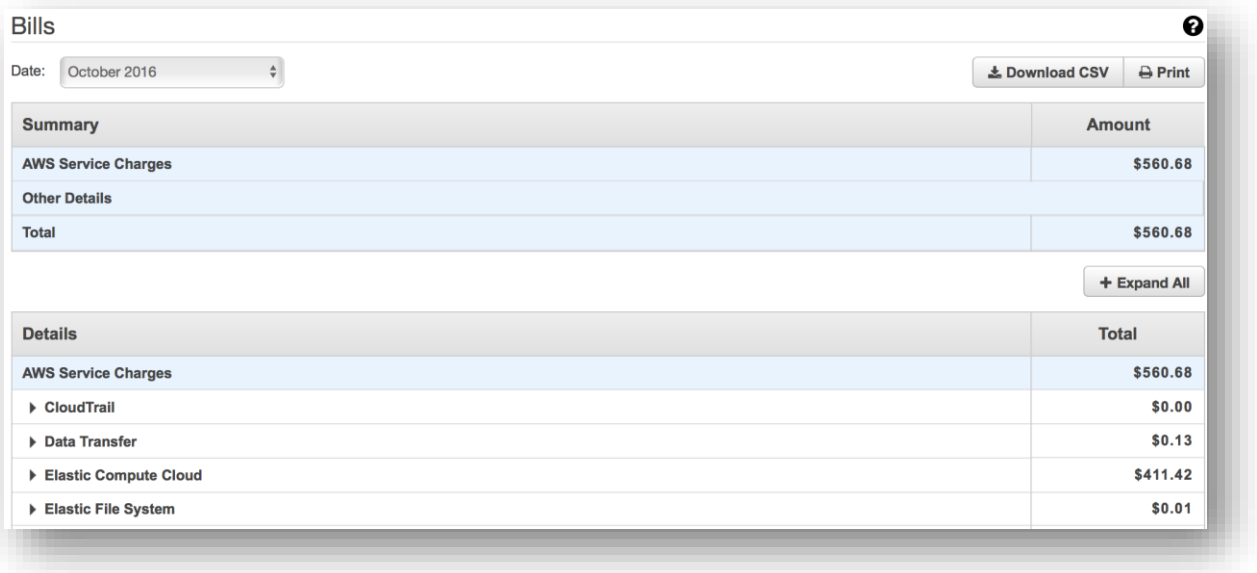


Figure 16 - AWS Billing and Cost Management Dashboard.

The "Bill Details" button in the top right corner shows you the breakdown by service.



Bills

Date: October 2016

Download CSV | Print

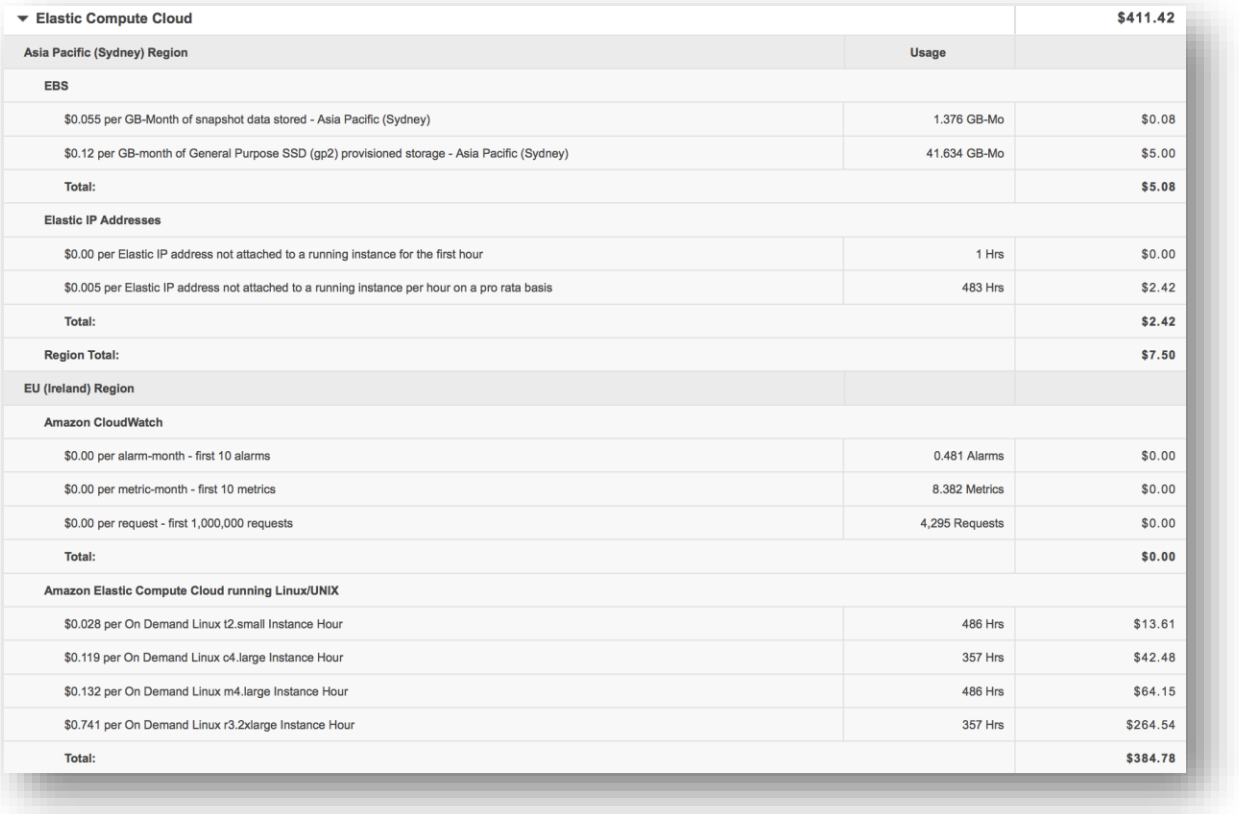
Summary	Amount
AWS Service Charges	$560.68
Other Details	
Total	$560.68

+ Expand All

Details	Total
AWS Service Charges	$560.68
▸ CloudTrail	$0.00
▸ Data Transfer	$0.13
▸ Elastic Compute Cloud	$411.42
▸ Elastic File System	$0.01

Figure 17 - AWS Bill Details.

Opening a specific service, e.g. Amazon EC2 (“Elastic Compute Cloud”), shows detailed charges and which instances in which regions generated them.

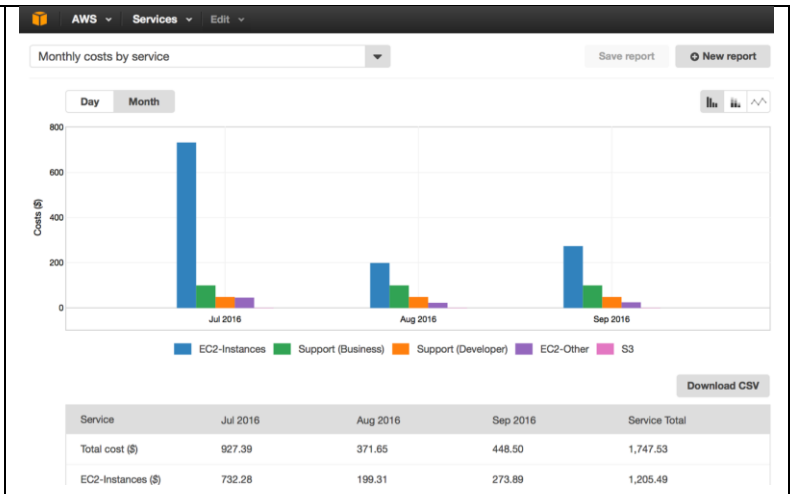
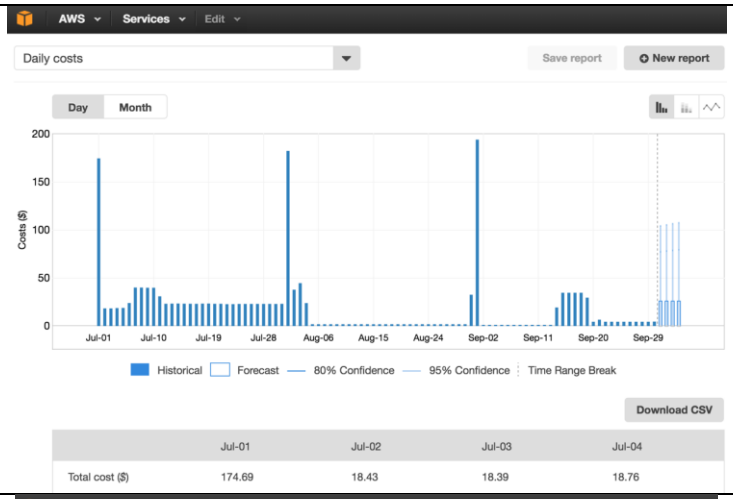
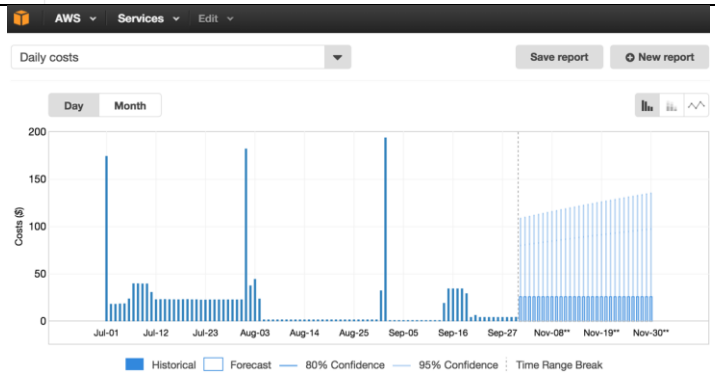


▾ Elastic Compute Cloud		$411.42
Asia Pacific (Sydney) Region	Usage	
EBS		
$0.055 per GB-Month of snapshot data stored - Asia Pacific (Sydney)	1.376 GB-Mo	$0.08
$0.12 per GB-month of General Purpose SSD (gp2) provisioned storage - Asia Pacific (Sydney)	41.634 GB-Mo	$5.00
Total:		$5.08
Elastic IP Addresses		
$0.00 per Elastic IP address not attached to a running instance for the first hour	1 Hrs	$0.00
$0.005 per Elastic IP address not attached to a running instance per hour on a pro rata basis	483 Hrs	$2.42
Total:		$2.42
Region Total:		$7.50
EU (Ireland) Region		
Amazon CloudWatch		
$0.00 per alarm-month - first 10 alarms	0.481 Alarms	$0.00
$0.00 per metric-month - first 10 metrics	8.382 Metrics	$0.00
$0.00 per request - first 1,000,000 requests	4,295 Requests	$0.00
Total:		$0.00
Amazon Elastic Compute Cloud running Linux/UNIX		
$0.028 per On Demand Linux t2.small Instance Hour	486 Hrs	$13.61
$0.119 per On Demand Linux c4.large Instance Hour	357 Hrs	$42.48
$0.132 per On Demand Linux m4.large Instance Hour	486 Hrs	$64.15
$0.741 per On Demand Linux r3.2xlarge Instance Hour	357 Hrs	$264.54
Total:		$384.78

Figure 18 - Granular AWS Bill Details.

3.1.3 Cost Explorer

For a more visual way to understand your AWS usage, go to the **Cost Explorer** from the main Billing & Cost Management Dashboard. There are predefined reports that should prove useful right away and options to create new ones specific to your needs.

The "Monthly costs by service" view shows you a breakdown of your last few months spend, highlighting the services that generated the most cost. As you'll see in most of these views, you can download the detailed data as a Comma-Separated Values (CSV) file in case you want to explore the data, compare usage growth in specific areas, or perform other analysis (or even reporting).	
"**Daily costs**" will give you your daily total consumption across all services going back several months. In the example shown here, you can clearly see the results of a summer recess on usage as well as particular days when activity picked up.	
By adjusting the controls to the right side of the graphs, we can ask for the expenditure forecast to estimate more than a month ahead. You'll notice that these estimates are provided with confidence intervals for 80% and 95%.	

There are several more features that you might want to explore, particularly in a busy lab or group with multiple research projects underway.

Tagging is a powerful AWS feature. You can assign a tag to almost anything, e.g., a specific Amazon EC2 instance, an Amazon S3 bucket, or a collection of Amazon EBS volumes. A tag has a name and a value. With Cost Explorer (and also with budgets—see below), you can create reports that single out the cost associated with a specific tag, which could be useful in telling the difference in Amazon EC2 usage between, say, Bob's project or Alice's or between single-instance workloads vs. clusters.

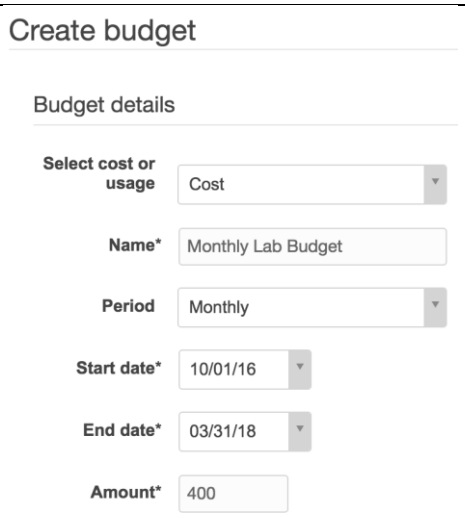
- You can find more about tagging here: http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/Using_Tags.html.
- You can find out how to customize your Cost Explorer views (including using tags) here: http://docs.aws.amazon.com/awsaccountbilling/latest/aboutv2/cost-explorer-tasks.html.

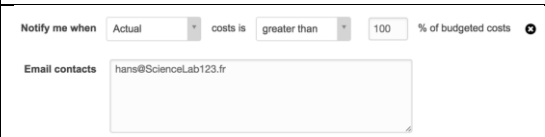
3.1.4 AWS Budgets

A budget is a way to **plan your IT usage** and your costs and to track how close you are to exceeding your budgeted amount. AWS Budgets can send you notifications if your spend—or your *projected* spend—goes outside your budgeted amount.

Budgets use data from Cost Explorer. Budgets is accessible via the left-hand menu in the Billing & Cost Management Dashboard.

3.1.4.1 Creating a Budget

Creating a budget is easy. It's necessary to know what you want to measure. This can refer to a specific service or to total spend. We recommend first creating a budget that measures all services. The example to the right sets a monthly budget of $400. Be sure to set the end date of the budget to far into the future.	
Notifications are important. You can set notifications to alert you when either actual spend or projected spend breach the limits you set as a percentage of your budget. This example sets an email alert when **forecasted costs exceed 90%** of the budget. Make sure the notifications go to an **e-mail address you'll actually check!**	

In the **same budget**, we create another alert (you can and should have several). This alert triggers an email when actual costs exceed 75% of the budget.	
Again, in the **same budget**, we create a third alert that triggers when we've actually spent all of our budget. This alert triggers an email when **actual costs exceed 100%** of our budget.	

Budgets and costs get updated by our billing systems several times a day, so you can be reasonably fine-grained about the alerts you put in place.

You can (and should) set quite a few alerts initially. A practice we've seen is to set up email notification alerts for when actual costs exceed 10%, 20%, 30% … 90%, 100%. In the first few months, you may get too much email (in which case, you can rescind some of the alerts for a larger granularity), but you will get some familiarity and understanding of what your usage pattern is. Once you become familiar with your usage, you can slow down the rate of alert triggering and only watch for significant jumps, like exceeding forecasted costs of 90+%.

Checkpoint

Have you **created a monthly budget**?	☐
Have you **created a budget alert** that notifies you that your **forecasted cost** will exceed your budget?	☐
Have you **created a budget alert** that notifies you that your **actual cost** will exceed your budget?	☐

3.2 Optional: the Budget Safety Switch

Email alerts from AWS Budgets are a great start to managing your spending in the cloud since you'll become familiar with the pace of your activities and how it impacts your budget.

It's possible to go one step further and put in place an extra layer of safety: if AWS Budgets detects your 100% budget being breached, it can trigger a script that will shut down all of your Amazon EC2 server instances, and it's these services that are generally responsible for most of your cost. (In technical AWS terminology, the EC2 instances are "Stopped", not "Terminated".) In the sections below, we'll show you how to configure this arrangement.

We're aware that a lot of our research customers would prefer to have this mechanism in place and face the chance of losing some data rather than risk losing their grant in the

case that they fired up more resources than they thought and left them running. It's up to you to decide if you want to implement this Budget Safety Switch.

3.2.1 A word of caution

But first, we'll express **some words of caution** to make sure you understand the nature of these alerts and the potential impact of this automated action.

- Automatically shutting down a group of Amazon EC2 instances will likely cause your applications to crash and **potentially lead to data loss**. It's hard to predict what your local file system might look like upon recovery.
- **While the sample scripts** we outline below will shut down the instances themselves, it **won't delete any attached storage volumes** (like Amazon EBS). The cost of Amazon EBS storage is low, but it's **non-zero**, so you'll need to decide what to do with these volumes. While you're deciding, they **will still cost you *some* money**. You might want to snapshot them into Amazon S3 or copy their data to Amazon Glacier before deleting them. Amazon S3 is a lot cheaper than Amazon EBS and is a great staging area for this data while you figure out what to do next. Glacier is even cheaper for long-term archiving of data.
- We recommend setting the alert to be triggered **when actual costs exceed 100% of your budget**. If you followed our advice above and **set up several budget alerts**, then by the time your actual costs are approaching 100%, you'll have been told so by several email alerts and you'll have had time to take preventative action before this failsafe gets triggered.

Finally, to reiterate: setting up an unattended script to turn off your compute instances comes with risks—the most significant one being the risk that you might lose some data. **AWS will not be responsible** for any data loss as a result of this configuration, nor can we be responsible for any residual overspend that stems from the fact that you may be using other services (such as storage or databases) that will continue to generate legitimate charges after your Amazon EC2 instances have been stopped.

Nonetheless, we hope that by providing this method you'll at least be able to sensibly manage to react should you unintentionally overspend and need to scale back your service consumption.

3.2.2 Download the Sample AWS CloudFormation Template

We've written a sample AWS CloudFormation template that you'll need to get a copy of. You can think of AWS CloudFormation as a language for describing collections of cloud resources and how they relate to each other.

You can download the sample template from the "Safety Switch" folder here:

https://github.com/awslabs/aws-research-cloud

This sample template has a few moving parts:

1. A single parameter called "DebugMode," which can be LIVE or DEBUG. We'll leave it in DEBUG until we're happy with everything.
2. An Amazon Simple Notification Service (Amazon SNS) topic – This refers to a topic that is managed by Amazon SNS, which is a push messaging service

connecting services in the cloud with other services or sometimes mobile phones or people (via email).

3. An AWS Lambda function – A short piece of python code that runs when needed and identifies Amazon EC2 instances that are running in any global AWS Region and then stops them.

When you download the sample AWS CloudFormation template, save it to your desktop and have a look using a text editor to see what it does.

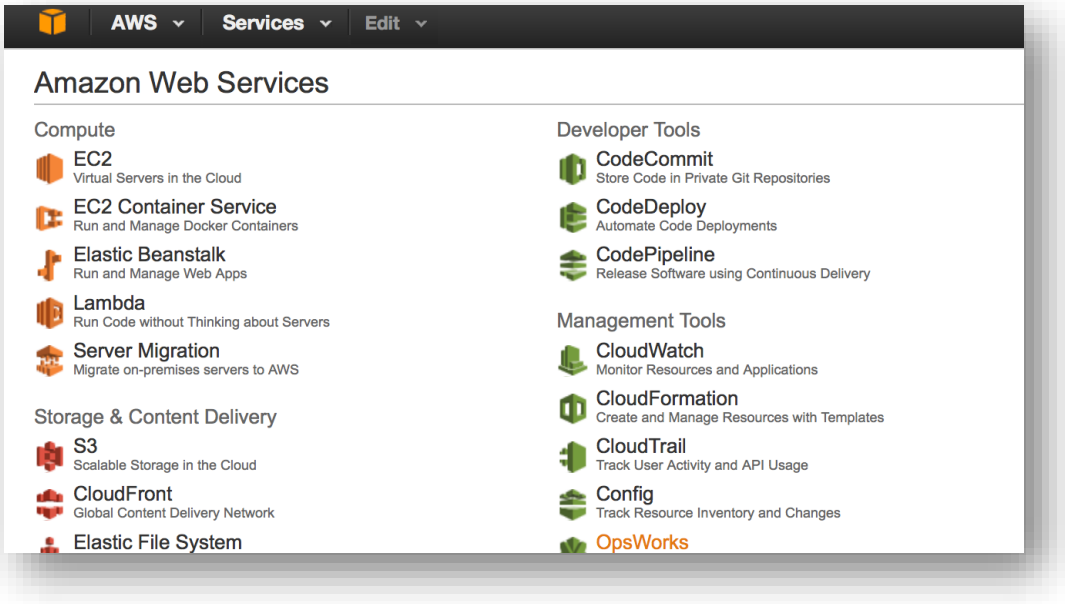
The template, and the embedded lambda function (written in python) has a very straightforward specification. When activated (in our case above, by an SNS notification from an AWS Budget alert):

- Stop every EC2 instance in every region that does NOT have a **tag** called "KeepRunning".
- Reduce the target to zero EC2 instances for every auto scaling group[17] in every region that does NOT have a **tag** called "KeepRunning".

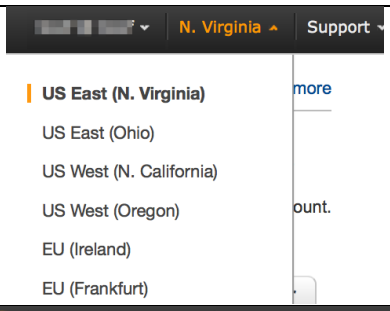
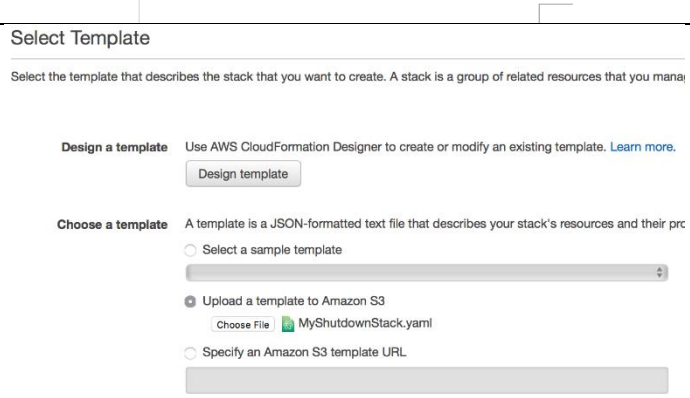
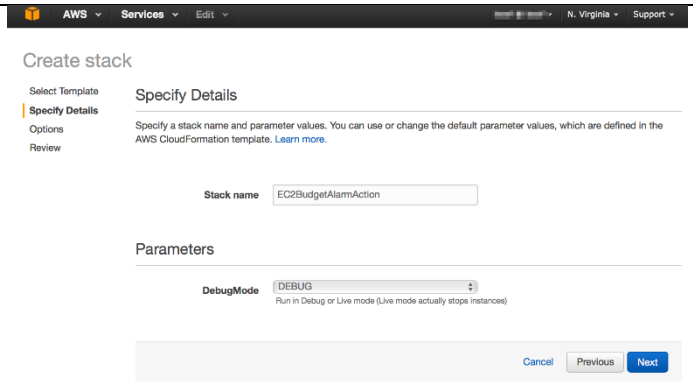
You'll note the use of tags here to protect certain instances or auto scaling groups. This means you can protect, say, a single t2 instance running a web server which might be running your lab's wiki or website, and which – since a T2 really has a very low cost – isn't something that will hurt your budget in the short term, while you figure out why you overspent.

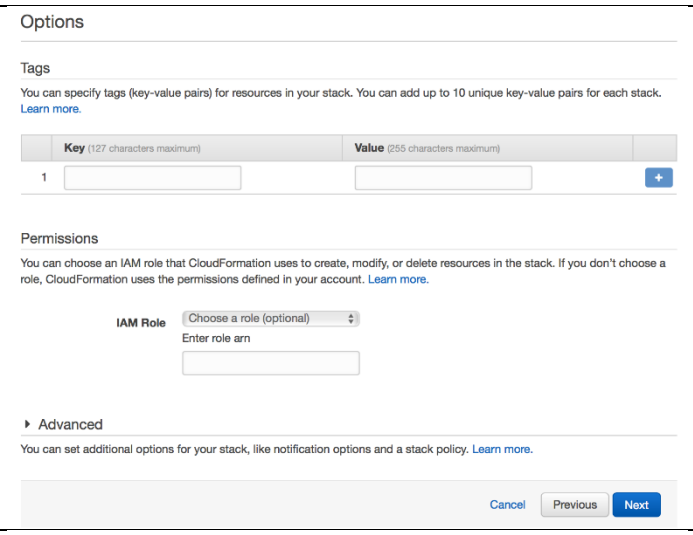
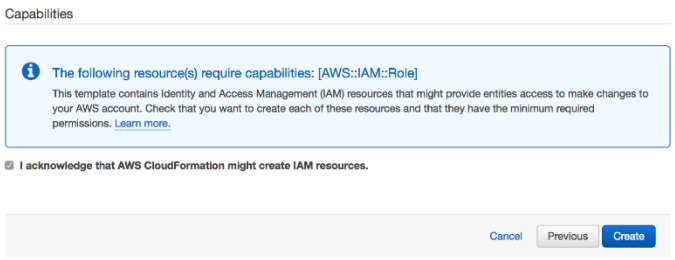
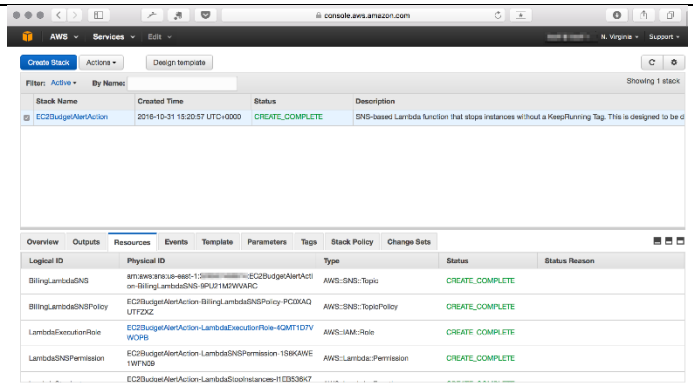
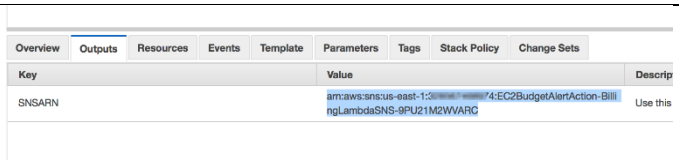
3.2.3 Uploading to AWS CloudFormation

The next step is to open your AWS Management Console and find the AWS CloudFormation dashboard under "Services" in the Management Tools area.



[17] Auto-Scaling Groups (or ASGs) are often used by cluster products to create a number of instances of the same type. They auto-scale in that they have a rule that governs under what conditions the group expands or contracts – usually based on CPU load or the number of jobs in a batch scheduler's queue (in the case of Alces or CfnCluster, for example).

Change your console's region (for now) to "**US East N. Virginia**" – this is the global AWS Region that **performs our billing operations**. This is at the top right of the console screen.	
Next you'll need to **Create a Stack** by clicking on the blue button.	
Click on "**Upload a template**" and use the "Choose File" finder to locate the sample template you downloaded earlier to your local desktop.	
You'll be asked to give your stack a name. "**EC2BudgetAlarmAction**" expresses the seriousness of its role quite well. **Leave DebugMode set to "DEBUG" for now**.	

We don't need to tag anything here, so you can just **click** "**Next**."	
Finally, you'll need to **click to acknowledge** that AWS CloudFormation will **create an IAM resource on your behalf**. **Once this is done, click** "**Create**."	
Once AWS CloudFormation has finished deploying all these resources, its status will change to green and say "**CREATE COMPLETE**."	
Click on the "Outputs" tab and you'll find a Key/Value pair. The value will have a long string starting with "**arn:aws:sns:us-east-1**." This is your Amazon SNS topic. When we link this with a budget alert, it'll be able to trigger your new AWS Lambda function to shut down your Amazon EC2 instances. **Copy this entire string**—you're	

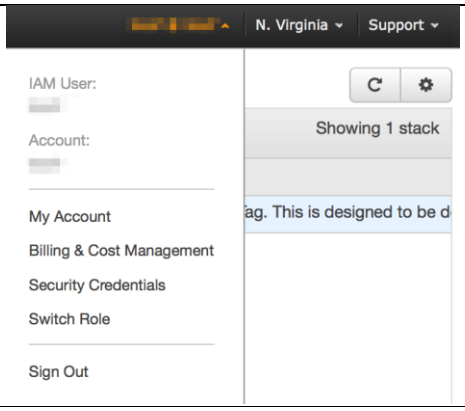
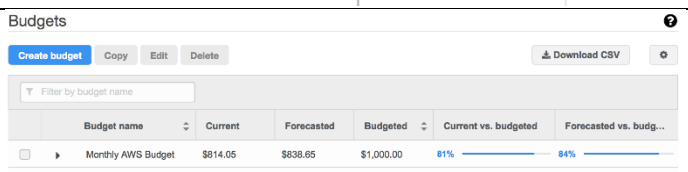
going to need it in a few minutes.	

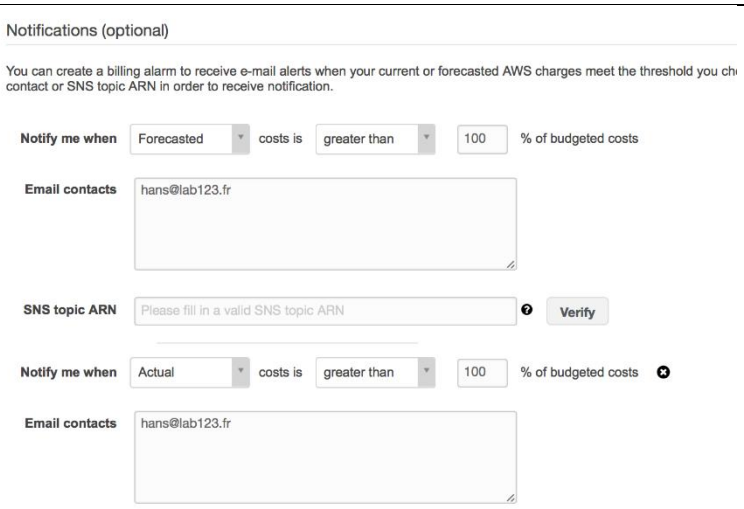
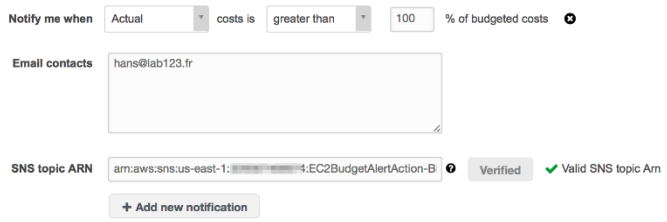
Checkpoint

Have you **obtained the sample AWS CloudFormation template** for the budget safety switch?	☐
Have you **uploaded the template** to the AWS CloudFormation console?	☐

3.2.4 Linking Amazon SNS with AWS Budgets Alerts

Now you have an Amazon SNS topic that will trigger an AWS Lambda function that shuts down your Amazon EC2 instances and disables Auto Scaling groups. It's time to connect this topic to your budget alert so that any **actual costs** (as opposed to projected costs) that exceed your monthly budget will cause this chain to kick off.

Now it's time to go back to **AWS Budgets** (remember, it's in the "Billing & Cost Management" section of your AWS Management Console.)	
If you defined a budget in the previous sections, then you'll have created some alerts, which send email to you when your expenditure exceeds various percentages of your budget.	

Make sure you have at least one email alert that triggers when **actual costs** exceed 75% (say) of your budget. Also, it's a good idea to have an alert that emails you when your **projected costs** exceed 100% of your budget. Both of these are **early warnings**. Set as many as you like. You can always remove some of them later once you're comfortable with your usage.	
For your rule that triggers when **actual costs exceed 100%** of your budget, you'll now add one more thing: your Amazon SNS topic. **Paste the long Amazon SNS string** that you copied from your AWS CloudFormation console into the "SNS topic ARN" field and click "Verify." **This connects this alert to your Amazon EC2 safety switch function**.	
Once you've finished setting up budget alerts, make sure (again) that you have one budget alert that triggers when 100% of **actual costs** is exceeded and is successfully linked to your Amazon SNS topic.	**Note: the safety switch function is still in DEBUG mode, so make sure you complete the steps below.**

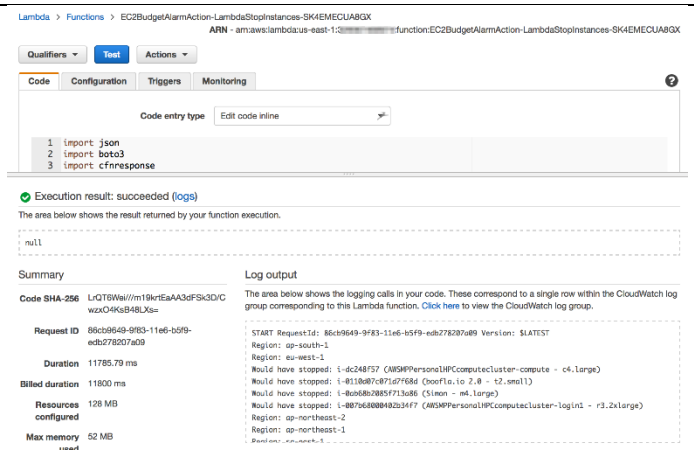
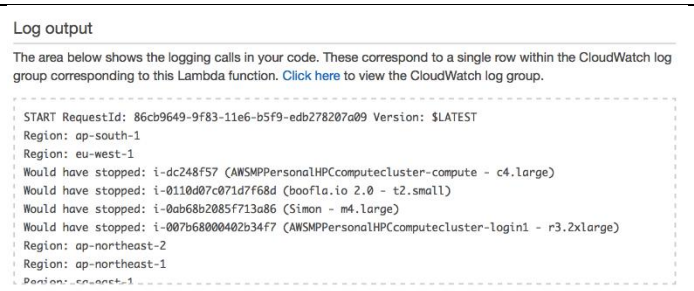
Checkpoint

Have you **linked your** Amazon SNS topic to your **budget alert** that triggers when **100% of actual costs is exceeded**?	☐

3.2.5 Testing Your Budget Safety Switch

We now need to go to the AWS Lambda service dashboard, which you'll find under the "Services" menu.

In the next couple of steps, we'll have you check your AWS Lambda function so that it performs as expected. When that's complete, you can enable it in LIVE mode.

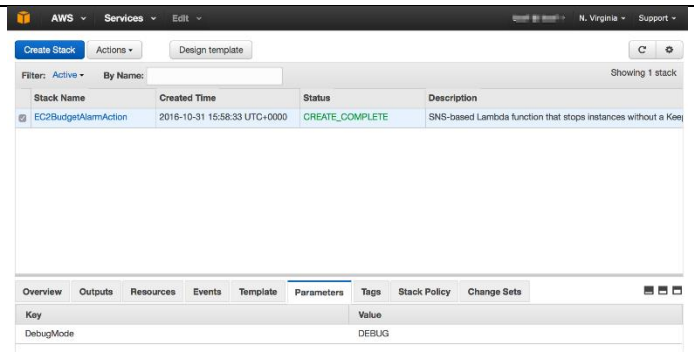
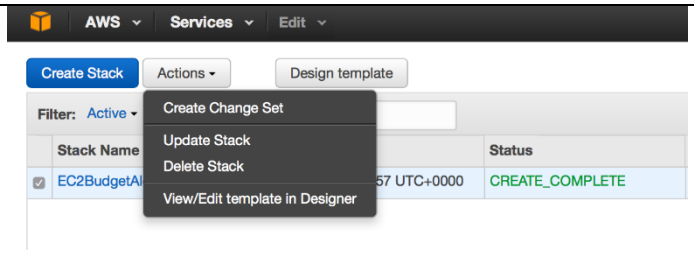
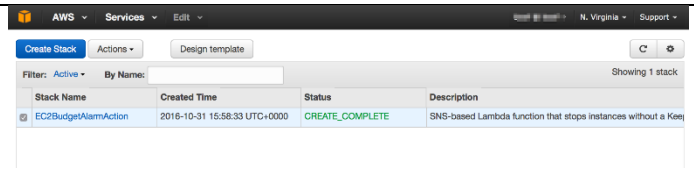
In the AWS Lambda console, you'll see your AWS Lambda function (most likely, it's the only function listed at this stage). Click on the name of the function to enter its test console. Click on the blue "test" button and click "Save and Test" at the bottom of the popup window—you don't need to set any values.	
After 10 seconds or so, the AWS Lambda function will finish searching all the AWS Regions for Amazon EC2 instances and will tell you which instances it **would have stopped IF** it was running in LIVE mode.	

Checkpoint

Have you **tested your AWS Lambda function**?	☐

3.2.6 Enabling the Budget Safety Switch

Now that we've tested the AWS Lambda function and verified our budget is in place, it's time to set the safety switch AWS CloudFormation template to "LIVE" mode so any future budget alerts that exceed 100% of **actual costs** will really trigger the shutdown we're after.

Under “Services” in the AWS Management Console, go back to the AWS CloudFormation dashboard. Locate your AWS CloudFormation template for the budget safety switch.	
Select your stack by clicking in the check box next to its name, and under the “Actions” menu, select “Update Stack.”	
Follow through the screens the same as before. Most things will be greyed out, since you can’t change them. When you get to the Parameters section, select “LIVE.”	
Once AWS CloudFormation is finished updating the stack, check under “Outputs” to be sure it says “LIVE.” Your safety switch is now in place.	

3.2.7 Protecting Specific Instances from Shutdown

To keep a specific instance running and prevent the Safety Switch from stopping it, give it a **tag** called “**KeepRunning**”. For example, you may want to keep a super-low-cost t2 instance running a web server regardless, while a bigger and more expensive cluster running CFD simulations would be torn down.

To keep an auto scaling group from being reduced to zero instances, likewise give it a tag called “KeepRunning”.

In both case, the **actual value of the tag isn't important**, just its existence is.

You can view and edit tags in the "Tags" tab of the specific instance page, which you'll find through the EC2 console, such as in the example shown below.

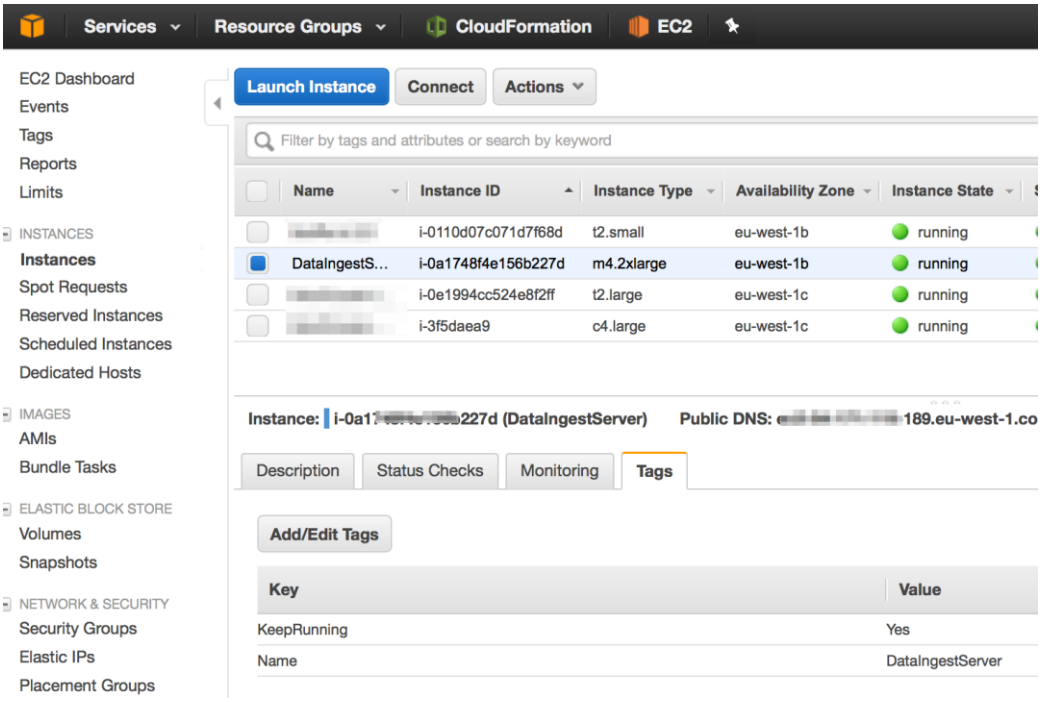


Figure 19 - The EC2 console showing the right place to set and edit metadata tags on your instances.

Checkpoint

Have you **changed the AWS CloudFormation template to "LIVE"** mode? As a reminder, **this won't stop all costs** from accruing, just the most significant ones caused by Amazon EC2, giving you time to inspect the state of your account and make changes.	☐

3.2.8 Recovering from a Budget Safety Switch Shutdown

Your first steps in any situation where the AWS Budget alert has activated the Safety Switch Lambda function should be to look at the CloudWatch logs from the lambda function itself – to understand what remedial actions were taken – as well as to check your expenditure (through the cost explore and other sources in the Billing Dashboard) so you can understand what costs drove your AWS Budgets to activate the function.

As we mentioned above, the Lambda function, when activated:

- Stops every instance in every region that does NOT have a **tag** associated with it called "KeepRunning".
- Reduces the target to zero for every auto-scaling group in every that does NOT have a **tag** associated with it called "KeepRunning".

You can review the list of instances and services that were stopped as a result of the Safety Switch by viewing the CloudWatch Log events associated with the Lambda function. A link to this can be found in the Outputs tab of the CloudFormation template (which you'll find in the CloudFormation console).

An example output might look something like this:

15:28:00	Region: ap-northeast-1
15:28:02	Region: sa-east-1
15:28:04	Region: ap-southeast-1
15:28:06	Region: ap-southeast-2
15:28:09	Region: eu-central-1
15:28:10	Would have stopped: i-0d726d4fbf42a626d (empty-name-tag - t2.large)
15:28:10	Region: us-east-1
15:28:11	Would have zeroed: asg-name-tag (from min/des/max): 0 / 0 / 0)
15:28:11	Would have zeroed: no-name-tag (from min/des/max): 0 / 0 / 0)
15:28:12	Skipping: i-097805fbf48b17656 (INSTANCE-STORAGE - m3.medium is not EBS backed)
15:28:12	Keeping: i-001f127bacf20ab6f (mgtus - t2.micro)
15:28:12	Keeping: i-0ecd71ba5cb04708d (bastion-1.bastion - t2.micro)
15:28:12	Region: us-east-2
15:28:12	Region: us-west-1
15:28:13	Region: us-west-2
15:28:16	END RequestId: 8c3fbfaa-bbc8-11e6-b1ca-89c8beb7d839

To restore services that were stopped by the Safety Switch Lambda function:

- Switch to the relevant AWS region in the console using the region picker in the top right corner of your AWS Console screen.
- Go to the EC2 services window
- Within the Auto Scaling Groups section, find the Auto Scaling Groups that were stopped, edit their settings, and change the Min, Desired and Max sizes back to the original values (shown in the CloudWatch Logs output)
- Within the Instances section, find the instances that were stopped, and start them from the console
- Repeat these steps for each region where you had running instances

3.3 Amazon EC2 Launch Limits

Your AWS account has limits for how many resources it may use at any given time. For example, there is a limit on how many c4.2xlarge instances you are allowed to have running in the Frankfurt Region. These limits protect you as a new user: e.g., they prevent you from accidentally running up an unexpected bill by launching many instances and not being aware that you need to terminate them when finished. Similar limits exist for all AWS Cloud services.

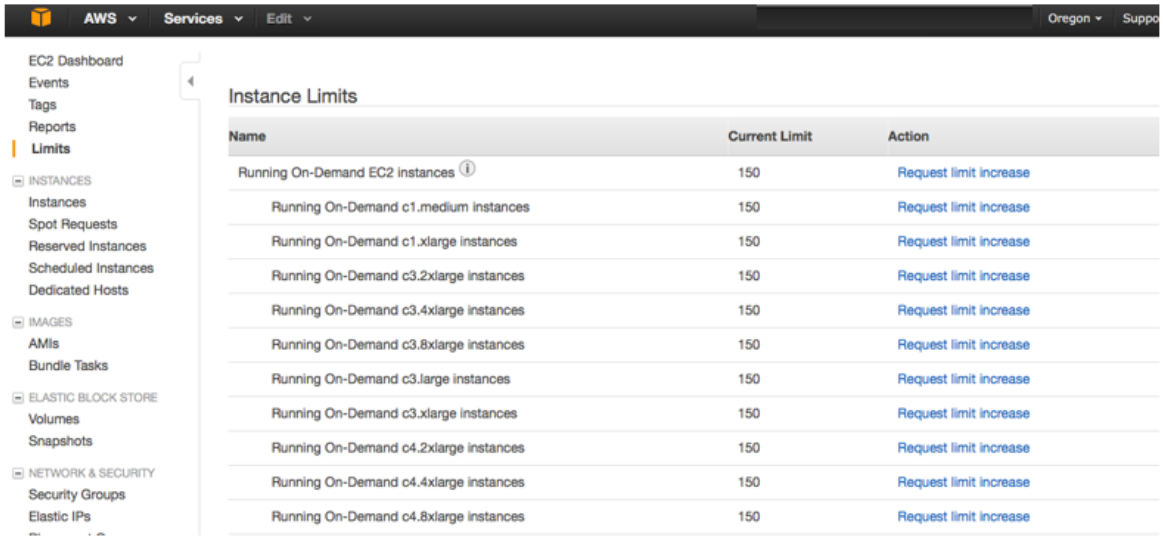


Figure 20 - Amazon EC2 Launch Limits.

Once you move into the production phase of your research, you may need more instances than this, especially if you have a large workload. This is accomplished by clicking the "Request Limit Increase" link next to the appropriate resource, which raises a ticket with our support team.

The account in Fig. 16 was originally allowed to create only 20 x c4.8xlarge instances (20 x 18 = 360 cores) but later received permission to create up to 150 x c4.8xlarge instances (150 x 18 = 2,700 physical cores, enough for some serious simulations).

You have to request limit increases separately for each AWS Region you work in.

3.4 Next Steps

If you've faithfully followed the steps outlined in this section of the book, then you've now established an AWS account with Budget Safety Limits, and you're employing our best practices for instance and data security. So, where to next?

We suggest you look at our "Getting Started with AWS" tutorials to begin your journey exploring and experimenting with the kinds of things you can do. The following chapters will also introduce you to additional AWS services and point you to tutorials.

For the first 12-months of a new AWS account, you have access to the Free Tier, which helps you to get hands-on experience with our services. After creating your AWS account you can use any of the products and services in the tier, for free within certain usage limits.[18] For example, when you launch EC2 instances, you will see whether the chosen instance type qualifies for the free tier before confirming.

[18] AWS Free Tier includes offers that expire 12 months following sign up and others that never expire. See https://aws.amazon.com/free/ and AWS Free Tier information for usage and other limits.

4 Working with Data

Your research endeavors will likely collect or create data that needs to be processed, analyzed, or shared. AWS offers a strong solution for data storage – at any scale, whether your needs are very modest or you burn up the petabytes faster than you can count them.

The durability and availability offered by the standard storage offering, Amazon S3, are outstanding[19]. Data security and privacy are provided to the highest standards, and AWS is certified for many compliance programs for sensitive data, including HIPAA, ITAR, FISMA, and many more (see https://aws.amazon.com/compliance/). Data locality and sovereignty are the default. On the other hand, the collaborative nature of research means that you frequently do want to expose data to your peers or to the public. The global footprint of AWS' worldwide regions and network match the global nature of science, and sophisticated security and access controls allow you to control exactly how and with whom you share data. While your data is only visible to you by default, the AWS Cloud is excellent for collaboration.

A variety of data storage services support the data lifecycle from hot storage to long-term archiving. The cloud paradigm of elasticity means that you never have to provision storage limits, but can rely on AWS to scale the storage hardware to your in-the-moment requirements. The pay-as-you-go model means you don't have to pay for storage that will sit empty until two-thirds into your research project.

But above all, AWS offers you a vast toolset of more than 90 services that integrate with storage: services for data analytics, visualization, machine learning, and more help you extract the most value from your data; all the compute power of the Elastic Compute Cloud (EC2) and associated workflow services are at hand to design analysis and compute workflows in a data-centric paradigm. This lets you build things you couldn't build anywhere else.

4.1 Range of Storage Types

As mentioned in section 1.4.2, AWS offers a number of storage services. These storage services are designed to scale in size, so if your prototype experiment proves to be a success, you can quickly scale it up without reformulating your scientific pipeline.

Amazon Simple Storage Service (Amazon S3)	Amazon S3 operates like a large and high-performance HTTP (web) server that you can put and get data objects to or from. Amazon S3 is the lowest-cost online storage service AWS offers. Hundreds of third party tools are designed to work with Amazon S3 and a large amount of the world's Internet activity is backed by it. Like everything in AWS, it was designed with strict security and access controls in mind. For example, researchers with human genetic data find Amazon S3 to be the perfect solution for some of **the research world's truly large datasets** (like the 1000 Genomes Database or the Sentinel-2 archive), and also some of the most sensitive (like The Cancer Genome Atlas, TCGA—a restricted-access dataset).

[19] "11 9's of durability" means, statistically, that if you store 10,000 objects, you can on average expect to incur a loss of a single object once every 10,000,000 years.

Amazon Glacier	**Amazon Glacier** is a secure, durable, and extremely low-cost cloud storage service for data archiving and long-term backup. To keep costs low, Amazon Glacier is optimized for infrequently accessed data where a retrieval time of several hours is suitable. Amazon Glacier is an appropriate place to **archive research data** for the long term once it has finished its immediate usefulness to your work.
Amazon Elastic Block Storage (Amazon EBS)	**Amazon EBS** is the most common type of storage for an Amazon EC2 instance. It appears to your instance like a hard disk, though it performs better than an HDD (in fact, you can specify the performance you want) and is more flexible.
Amazon Elastic File System (Amazon EFS)	**Amazon EFS** provides a file system interface and looks like a remote directory and file server to your instances, similar to the shared home directory you find on HPC clusters or in shared research computing facilities. Amazon EFS automatically scales to however much data you pour into it. It's in its element when making massive file systems available to thousands of Amazon EC2 instances at a time. To access your file system, you mount the file system on an Amazon EC2 Linux-based instance using the standard Linux mount command after provisioning it through the Amazon EFS console. Then you can work with the files and directories in your file system just like you would with a local file system.

4.2 Storing your Research Data

In Chapter 2 you created an IAM login that can launch and log into resources. Next you need to think about where to put data for input and output of analyses. AWS has lots of storage options, two of which are Amazon EBS and Amazon S3.

Amazon S3 is an object storage service, meaning that it is not a traditional filesystem. Files are considered "objects" in S3, and the full directory path to the object is considered a "key". S3 provides a simple HTTP interface that can be used to store and retrieve any amount of data, at any time, from anywhere on the web. This means you can access data outside of EC2, even from your laptop or desktop.

Every object in Amazon S3 is stored in a bucket. Before you can store data in Amazon S3 you must create a bucket. We suggest creating different buckets for different categories of data to help you keep organized. **You do not get charged for creating a bucket.** You are only charged for storing objects in the bucket, paying for the total number of GB you store, for exactly as long as you store it. You have complete, granular control over access permission to your S3 buckets and data stored therein.

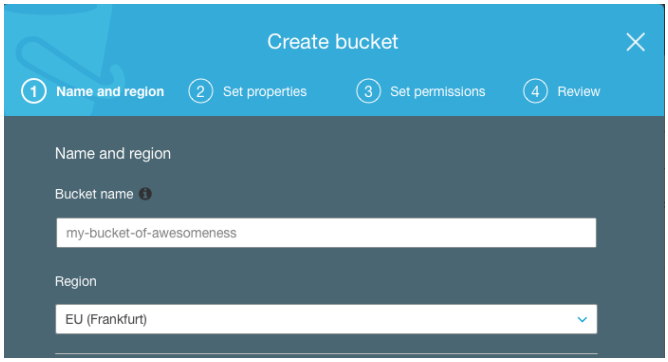
Amazon EBS is like a regular disk drive (or volume) that is attached to your computer, but with a couple of differences. First is that EBS volumes are accessed over the network. You can attach and detach the volume from one EC2 instance and reattach it to another.

The volume will only be directly accessible to a single EC2 instance at any time[20]. Second, you can take a snapshot, or backup, of the EBS volume and create a new duplicate volume from the snapshot to attach to other instances. EBS snapshots are useful if you have a set of applications or reference data that are needed on more than one running instance. They can also be shared with collaborators using different AWS accounts. The data on an EBS volume are not accessible outside of an EC2 instance. It's common to copy data from S3 to an EBS volume to make it available to the EC2 instance for processing. Any valuable output data can be copied back to S3 for long-term storage, while the EC2 instance and the EBS volume are generally deleted when they no longer serve a purpose.

4.2.1 Creating Your First Amazon S3 Bucket

In the AWS Management Console (https://aws.amazon.com/console), **open the Amazon S3 console**.

Click **Create Bucket**. The "Create Bucket" dialog box appears.



Enter a bucket name in the **Bucket Name** field. The bucket name you choose must be unique across all existing bucket names in Amazon S3. One way to do that is to prefix your bucket names with your research group's name.

When naming your bucket, take some care. **Once you create a bucket**, **you cannot change its name**. And do know that the bucket name is visible in the URL that points to the objects stored in the bucket. Make sure the bucket name you choose is appropriate.

Bucket names:

- Can contain lowercase letters, numbers, periods (.) and dashes (-).
- Must start with a number or letter.
- Must be between 3 and 255 characters long.
- Must not be formatted as an IP address (e.g., 265.255.5.4).

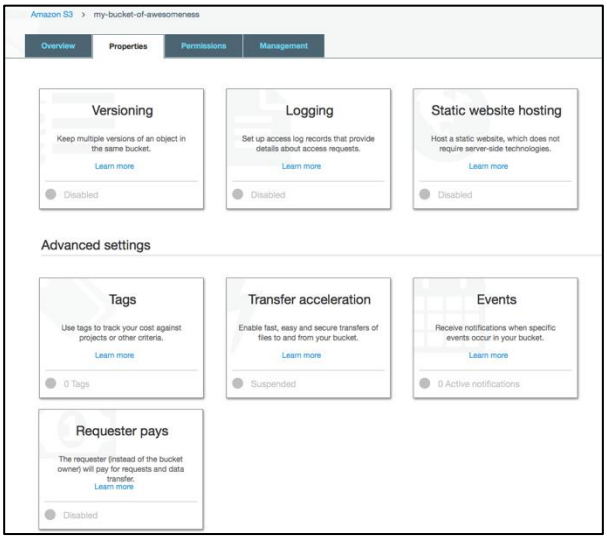
[20] The instance can export it to other instances using NFS, see Chapter 7.

Note: There might be additional restrictions on bucket names based on the region your bucket is in or how you intend to access the object.

In the **Region** drop-down list box, select a region.

Click **Create**. When Amazon S3 successfully creates your bucket, the console displays your empty bucket in the **Buckets** panel.

You can modify your bucket's default permissions by clicking on the properties of your bucket in the console.



Checkpoint

Have you **created your first Amazon S3 bucket**?	☐
Have you **checked the default permissions** on your bucket to make sure they match your intended use?	☐

4.2.2 Uploading and downloading data to and from Amazon S3

Now that you've created an Amazon S3 bucket, it's useful to have some tools that you can use to easily push files back and forth.

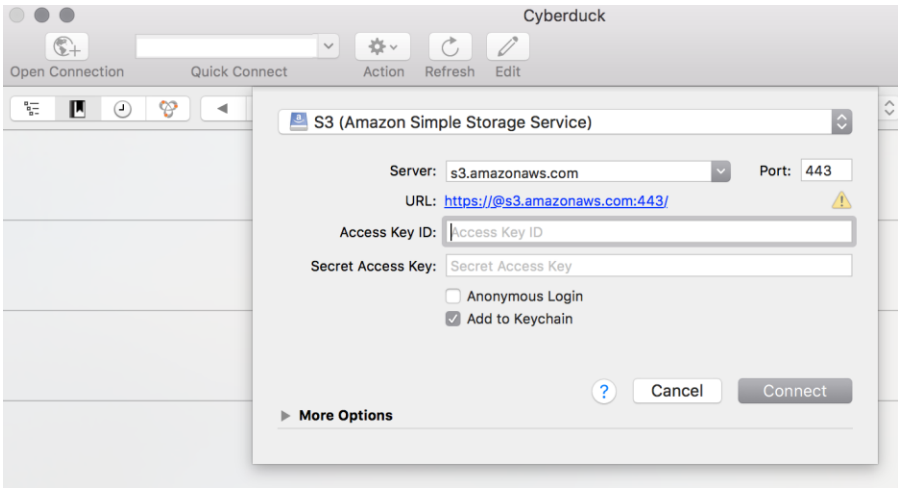
4.2.2.1 Using a GUI app, for example: Cyberduck

Cyberduck is an example of a popular desktop application that puts a friendly interface on Amazon S3. It's available for Mac and Windows. There are many popular alternatives, e.g. Cloudberry or Transmit.

When you created your IAM login earlier, you were required to download a file containing your credentials in the form of an "Access Key" and a "Secret Key ID." We recommended that you store them securely in `~/.aws/credentials` on your local computer and, as a reminder, they look something like this:

```
aws_access_key_id = SDFxxxxxxxxxxxxF89
aws_secret_access_key = 345xxxxxxxxxxxxxxxxx3454/sdfsxxxxxxxx/8
```

To use Cyberduck or another client, you'll need these keys, since they identify you and thus permit you to operate on your bucket with all the privileges you have as its owner.



After copying and pasting your key pair into Cyberduck, you'll be able to see the contents of your (probably empty) Amazon S3 bucket and to easily move files back and forth.

Unless you set the bucket's default posture to allow for "Everyone" to have view permissions (through the Amazon S3 console), your files can only be read by you and anyone else will get a "permission denied" error.

If you're intending to use Amazon S3 to share data with collaborators or to make some content public, you'll need to configure Amazon S3 permissions and, potentially, bucket policies. This is explained in 4.3 below and in the Amazon S3 online documentation[21].

Checkpoint

Have you **downloaded and installed an Amazon S3 GUI tool**?	☐
Have you **configured your Amazon S3 Secret Key and Access Key ID** into the tool and uploaded your first file?	☐

4.2.2.2 AWS Command Line Interface (CLI)

While GUI-based apps like Cyberduck and Transmit can make life easier, sooner or later you're going to want to install the AWS CLI, a utility that lets you make API calls against your AWS account from the command line terminal.

The real power of this is that you can then build scripts or tools to automate your AWS workloads. There are even libraries for many common programming languages[22]. Before you can use the AWS CLI, you must have Access Keys configured in your environment or `~/.aws/credentials` folder as we showed you earlier.

Installing AWS CLI on your Mac or Linux machine is as simple as running:

```
$ pip install awscli
```

[21] Which is available here: http://docs.aws.amazon.com/AmazonS3/latest/dev/s3-access-control.html
[22] See https://aws.amazon.com/tools/ in the "SDKs" section.

…in a terminal[23]. The AWS CLI comes preinstalled in most (if not all) Linux instances you might spin up in the cloud, when picking from the AWS Marketplace.

The online documentation explains how you can use the CLI to control practically anything in AWS that you can do with the console.

For example, the Amazon S3 section of the CLI user guide shows how we can create new buckets with a single command:

```
$ aws s3 mb s3://bucket-name
```

Here are some examples of uploading files to buckets as well as a few other file manipulations, so you get the idea of what this command line can do.

```
// Copy MyFile.txt in current directory to s3://my-bucket/path
$ aws s3 cp MyFile.txt s3://my-bucket/path/

// Move all .jpg files in s3://my-bucket/path to ./MyDirectory
$ aws s3 mv s3://my-bucket/path ./MyDirectory --exclude '*' --include
'*.jpg' --recursive

// List the contents of my-bucket
$ aws s3 ls s3://my-bucket

// List the contents of path in my-bucket
$ aws s3 ls s3://my-bucket/path

// Delete s3://my-bucket/path/MyFile.txt
$ aws s3 rm s3://my-bucket/path/MyFile.txt

// Delete s3://my-bucket/path and all of its contents
$ aws s3 rm s3://my-bucket/path --recursive
```

There is also a "sync" command, which allows you to synchronize a local folder with a remote Amazon S3 bucket, only copying files that are missing or have changed.

```
$ aws s3 sync . s3://my-bucket/path
upload: MySubdirectory\MyFile3.txt to s3://my-
bucket/path/MySubdirectory/MyFile3.txt
upload: MyFile2.txt to s3://my-bucket/path/MyFile2.txt
upload: MyFile1.txt to s3://my-bucket/path/MyFile1.txt
```

The AWS CLI uses **multipart file transfers** by default, which generally provide very good performance for copying data in bulk into AWS.
Try it out. AWS has some 10-minute tutorials on Store and Retrieve a File, and Store Multiple Files using the aws cli.

[23] See https://aws.amazon.com/cli/ for more details, or if you're running Windows, which necessitates a separate download.

4.3 Sharing the data in your S3 bucket.

In Chapter 2, you created an Administrator IAM user for yourself, and additional IAM users for the other users of your AWS account. You gave your administrator IAM user full access to your account, including all your S3 buckets. Sooner or later, you will need to give the other IAM users access to some of your S3 buckets. And you may also give access to people working in another AWS account, for example, a collaborator at another university. You may even make some data public to the whole world.

If it is not necessary now to share your data, then just come back to it when you reach that point. Otherwise, read on.

You already got a first taste of access policies in Chapter 2 when creating IAM users. Here we encounter them again. In a small research group, it could be fine to give all IAM users full S3 access. But if needed, policies can be incredibly granular. You can attach policies to IAM users ("Let this user do anything to any of my S3 buckets" or "Let this user list but not read files in bucket A, let her read and write to bucket B, and don't give her any access to bucket C"). Alternatively, you can attach policies to S3 buckets ("Let accounts x and y write to this bucket" or "Requests from this IP range can read only '*.cif' objects from this bucket, while requests from this other IP range can write files to the bucket", or "Make this bucket a static website for the whole world to see"). If no policy is attached to a bucket, it's private by default.

- For a detailed tutorial on granting your IAM users access to S3 buckets, follow http://docs.aws.amazon.com/AmazonS3/latest/dev/example-walkthroughs-managing-access-example1.html.[24]
- For a detailed tutorial on granting cross-account bucket permissions, follow http://docs.aws.amazon.com/AmazonS3/latest/dev/example-walkthroughs-managing-access-example2.html.

You can skip the first steps in these tutorials since you already created buckets and IAM users.

Finally, AWS has a "policy generator" where you can specify which actions you want to allow or disallow, and receive back a matching policy that you can apply to your buckets or users. See http://awspolicygen.s3.amazonaws.com/policygen.html .

[24] http://docs.aws.amazon.com/AmazonS3/latest/dev/walkthrough1.html provides even more detail.
http://docs.aws.amazon.com/AmazonS3/latest/dev/example-policies-s3.html provides even more examples of bucket policies.

4.4 Global Data Egress Waiver for Research

The AWS **Global Data Egress Waiver program** waives most or all of the "data out" charges on your monthly bill.[25] These are standard AWS charges for moving data out of the cloud over the public Internet. "Data out charges" are there so that video streaming and other bandwidth-intensive users pay their way so that we have wide-open Internet pipes at all times. But data transfer charges can seem unpredictable to anyone not familiar with estimating them - hence the Global Data Egress Waiver program.

The waiver works by removing any "data out" charges from your bill so long as they constitute less than 15% of your total charges. Any excess spending above 15% will still be charged at the normal (per GB) rate, but it's worth noting that "data out" charges rarely exceed this amount. The result is a bill composed of more familiar charges – computing, storage, database usage, etc.

There is never any cost to uploading data to AWS or moving data between Amazon S3 storage and Amazon EC2 compute instances housed in the same AWS region.

To qualify, you must work in an academic or public-sector research institution (such as a university or national research lab). You must use an educational email address for your AWS account. You must conduct at least 80% of your traffic activity through your university's National Research & Education Network (NREN) connection, which is the case for most universities. Finally, you must not be running an application that's not part of the normal working of a research institution and whose primary purpose is to stream data (like an external or commercial video service or a MOOC – this keeps you from building the next commercial video-on-demand business under this discount program).

When you sign up for the AWS Research Cloud Program through the website, we'll verify your eligibility for the Global Data Egress Waiver program. If you qualify, we'll contact you shortly.

4.5 Very large data transfers to S3

Most users upload their data to AWS over the internet. AWS maintains a very performant network, and the limiting factor in upload performance is normally somewhere downstream, closer to your location. Smart tools, such as the aws cli, automatically make the best of your connection when uploading a large number of files.

To support research, AWS is peered with major **National Research and Education Networks** (NRENs) worldwide. For example, in the US AWS peers with **ESnet**, **Internet2**, and several regional NRENs. In Europe, with **GEANT** and its member networks. In Japan, with **SINET**. And so on. There's a good chance that your university or research institute has access to one of these dedicated research networks, which typically give you better bandwidth, performance, and stability than you would have on a public internet connection. It also makes you eligible for the cost savings of the Global Data Egress Waiver program (chapter 4.4).

[25] https://aws.amazon.com/blogs/publicsector/aws-offers-data-egress-discount-to-researchers/

For some users, uploading over the internet does not work. Maybe they work at a remote research station where connectivity is **limited**, or the volume of data that needs importing is **gargantuan** and it would take too long or cost too much.

AWS Snowball is one answer to the challenge.[26] You can request a "snowball" from the AWS console, and AWS will ship you a rugged and portable device that you can plug in and load up with 80TB of data. Then you ship it right back (your data is safely encrypted) and we load it into S3 for you. You can request multiple snowballs for larger datasets.

You can also look into Snowball Edge for an "edge device" that has compute capabilities, e.g. for work in remote locations where you need to preprocess the data on site before shipping it back to AWS.

At the largest end of the spectrum, you can request a visit from a Snowmobile, which is pretty much a truck pulling a shipping container full of storage.[27] Each truck can quickly and securely move 100PB of data to S3.

4.6 Data stream ingestion

Data streams refer to a continuous ingestion process, with new data being generated and pushed to the cloud on an ongoing basis, often originating from multiple sources. For example, it could be a stream of radar observation data from weather stations all across the country. Or a subscription to all tweets containing certain key phrases. Such data streams could go straight into S3 storage, but often you first want to perform some processing steps on them as they come in. For example, you might scan for abnormal weather observations, and have them kick off an alert, or a weather forecast simulation. Tweets might be parsed for content triggers to build heat maps. AWS Kinesis is the foundational service for ingesting, combining, buffering, and emitting data streams. In combination with services such as AWS Simple Notification Service (SNS), AWS Lambda, and others, you can build sophisticated pipelines.

Looking at the case of a very large number of sources and smaller message size, we find the AWS Internet of Things (IoT) service. This service lets you build a powerful managed cloud environment for your IoT devices. (See chapter 8.6.)

4.7 Open Data means more scientific impact

In many fields of science, lack of **access to valuable datasets** can prevent you from quickly getting to the data you need for your research and keep others from using data you've collected for their own research. As a solution for collaboration, AWS can be used to share any volume of data with anyone in world.

[26] https://aws.amazon.com/snowball/ ; https://aws.amazon.com/snowball-edge/
[27] https://aws.amazon.com/snowmobile/

According to IDC, the volume of data being produced each day is growing at an explosive rate, and is now estimated at 2.5 exabytes. Much of it is "open data," which means the data can be used by anyone for any purpose without needing to pay a licensing fee. This is a boon for scientists, entrepreneurs and public servants, who can use the data to create new products, **accelerate scientific discovery** and provide better services to the public.

Traditional computing infrastructure is not suitable for sharing large volumes of data. Government data is still being provided with the assumption that users will download and store their own copies of data. That's fine when a few gigabytes of data are being shared, but as data volumes increase, this approach simply doesn't work. For example, the National Oceanic Atmospheric Administration (NOAA)'s new weather satellite, GOES-16, is estimated to produce **one terabyte of data per day**. Downloading one terabyte of data over a 50Mbps connection would take two days. Very few people have the hard disks and patience to download a terabyte of data.

This is why the cloud is emerging as the **center of gravity** for big data analysis. Once data is made available in the cloud, you no longer need to buy hard drives and spend months downloading the data. You can instead use on-demand computing resources in the cloud to query as much, or as little, of the data as you need. When the analysis is done, you can save the results, turn off the virtual servers and not have to worry about paying to store an individual copy of the original data.

Through the Amazon Web Services **Public Datasets program**, we host some of the world's most valuable open datasets and show what's possible when data is made available in the cloud. For an example of data flourishing on AWS, let's take a look at Earth on AWS.

4.7.1 A new look at Landsat

Since 1971, Landsat satellites have produced the longest continuous record of Earth's land surface **as seen through space**. These images have been available at no cost directly from the United States Geological Survey since 2010. However, many people were limited by their ability to download and store significant quantities. We talked to

many end-users and learned that they had big ideas of what could be done with Landsat data, but couldn't get it fast enough or couldn't afford to store their own copies.

Amazon started hosting imagery from the Landsat 8 satellite in 2015. Within the first year, over 1 billion requests for Landsat imagery and metadata were logged from 147 countries. Businesses like Esri, Mapbox and Mathworks immediately created tools to take advantage of the new easy-to-access Landsat archive.

One of the most interesting developments has been how novices and amateurs have been able to create entirely new interfaces and tools to explore and analyze the data. A group of students from Code Fellows created Snapsat – a fast and novel web-based service to browse and interact with Landsat imagery. An independent developer in Melbourne, Australia, even created an iPhone app called Observed Earth, giving people the ability to access tremendous amounts of data on how the Earth has changed over time by simply reaching into their pocket.

4.7.2 NEXRAD Opening New Research Frontiers

NOAA recognized early on that the cloud would be essential to fulfill their mission, and in 2015, they entered into a research agreement with several cloud service providers to explore ways to drive usage of their data. Through that agreement, we have made several hundred terabytes of high-resolution NEXRAD radar data available in the cloud.

Similar to the response we saw with Landsat, the usage of NEXRAD data has been impressive. After making NEXRAD data available in the cloud, NOAA recorded a 130 percent spike in usage of the data, while simultaneously seeing a 50 percent decrease in the usage of their own servers.

This open data initiative has also made the full NEXRAD archive available on demand, creating new analysis and discovery possibilities. For example, Dr. Eli Bridge at the University of Oklahoma has used this public dataset to compile radar data to estimate the size of **Purple Martin bird roosts**. These birds form large, dense aggregations that appear as ring-shaped patterns on the radar images. Now that the researchers no longer have to make requests for individual scans and receive chunks of data at a time, the University of Oklahoma team is able to learn how the birds are responding to droughts, environmental change, and seasonal queues. This is an example of "latent research" – that is, research that has existed in the minds of researchers, but hasn't been possible because of restricted data access.

4.7.3 Life Sciences and other datasets

Outside of Earth science, the Open Data program hosts The Cancer Genome Atlas, Common Crawl, and more. You can see the full list at https://aws.amazon.com/public-datasets/. These datasets provide good examples of how to share your own research data with the world.

4.7.4 Your research data

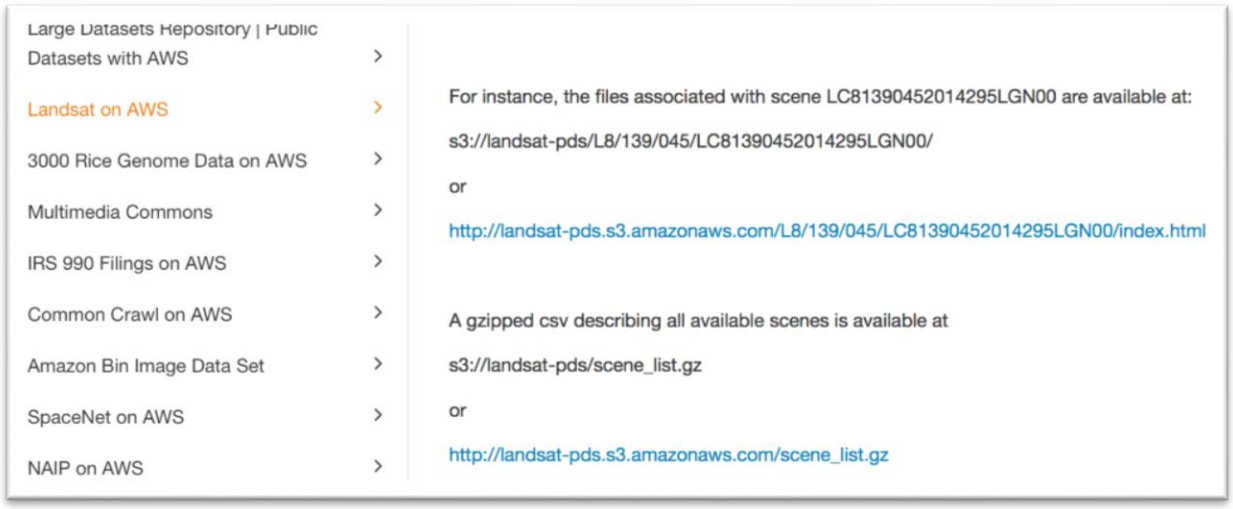
When data is shared in the cloud, anyone can analyze any volume of data without needing to download or store it themselves, which lowers the cost of research (for everyone) and reduces the time to scientific discovery. Landsat was accessed more than a billion times

in its first year on AWS. Usage of NEXRAD tripled once hosted on AWS. **How much impact could your data create** if you staged it for analysis on AWS?

If you have a dataset that is useful to a broad community of users, and you would like to encourage usage of it in the cloud, you may be a candidate for participating in the Public Datasets program. If you have any questions or would like us to consider your dataset for the program, please email us at opendata@amazon.com.

4.7.5 How to use AWS Public Datasets

Most AWS Public Datasets are made available through Amazon S3. To access a public dataset hosted in Amazon S3, you can make simple HTTP requests, use AWS command line tools and SDKs (Ruby, Java, Python, .NET, PHP, etc.), download the data using Amazon EC2, or use Hadoop to process the data with Amazon EMR. When you've found the landing page on the AWS website for your chosen dataset, you'll notice that the URL structure for Amazon S3 repositories is described in detail, along with a link to a file that lists all the objects or another method for discovering them.

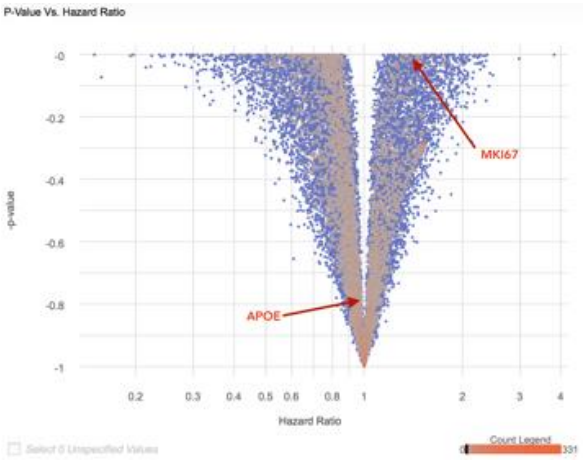


4.7.6 Tutorials.

Big (Census) data. Most recently, the Census Bureau discovered that increasing access to big data can lead to increased usage. Previously, the agency's American Community survey data was only available on tabular file formats like CSV, which required days to access and then required a separate reference document to be able to make any sense of it. Now that it has been uploaded in bulk to the cloud, anyone can access and analyze the entire dataset for about 40¢ an hour. The National Science Foundation provides a tutorial on how to analyze the data using an open source graph database engine. To access the tutorial, first follow section 5 of https://docs.data.world/uscensus/#54---loading-the-data-into-blazegraph , then proceed to https://docs.data.world/tutorials/sparql/ .

The UK Met Office has made 80TB of meteorological data available in AWS S3. See chapter 9.3.4 for tutorials on working with this weather data.

This tutorial guides you through a survival analysis against The Cancer Genome Atlas on AWS, Using the RNA-seq expression counts data from the breast cancer cohort of TCGA, looking at the MKI67 and APOE genes. It uses Station X's GenePool platform and you'll need to work with AWS Lambda and API Gateway



.

5 Working with sensitive and controlled-access data

Researchers sometimes need to work with data from human subjects, which may contain Personally Identifiable Information (PII), such as participant details from surveys or clinical phenotypes taken from electronic health records. Researchers need to treat such data with respect for a participant's privacy. Sometime the data may even fall under some regulatory requirement, as is the case for clinically derived data sources. There are also some types of data that do not strictly fall under formal regulations, but are treated by the research community as controlled access data. An example of controlled access data is genomic sequence derived from cancer patients that have donated their biological specimens and data for use in research studies.

Security is the number one priority at AWS. Helping to protect the confidentiality, integrity, and availability of your systems and data is of the utmost importance to AWS, as is maintaining your trust and confidence. We build and operate our infrastructure according to security best practices and standards, and with the unique needs of the cloud in mind. AWS uses redundant and layered controls, continuous validation and testing, and a substantial amount of automation to provide confidence that the underlying infrastructure is monitored and protected 24x7. AWS replicates these controls in new data centers and services.

As a researcher working with sensitive data, you directly benefit from the security of our data centers, network infrastructure, and services, which were built to satisfy the requirements of our most security-sensitive customers. In the following sections, we will walk through how security and compliance works on AWS, using the example of working with controlled-access genomic sequence data.

5.1 The Shared Responsibility security model

AWS operates under a shared security responsibility model, where AWS is responsible for the security of the underlying cloud infrastructure and you are responsible for securing the workloads you deploy in AWS (see **Figure 21** below).

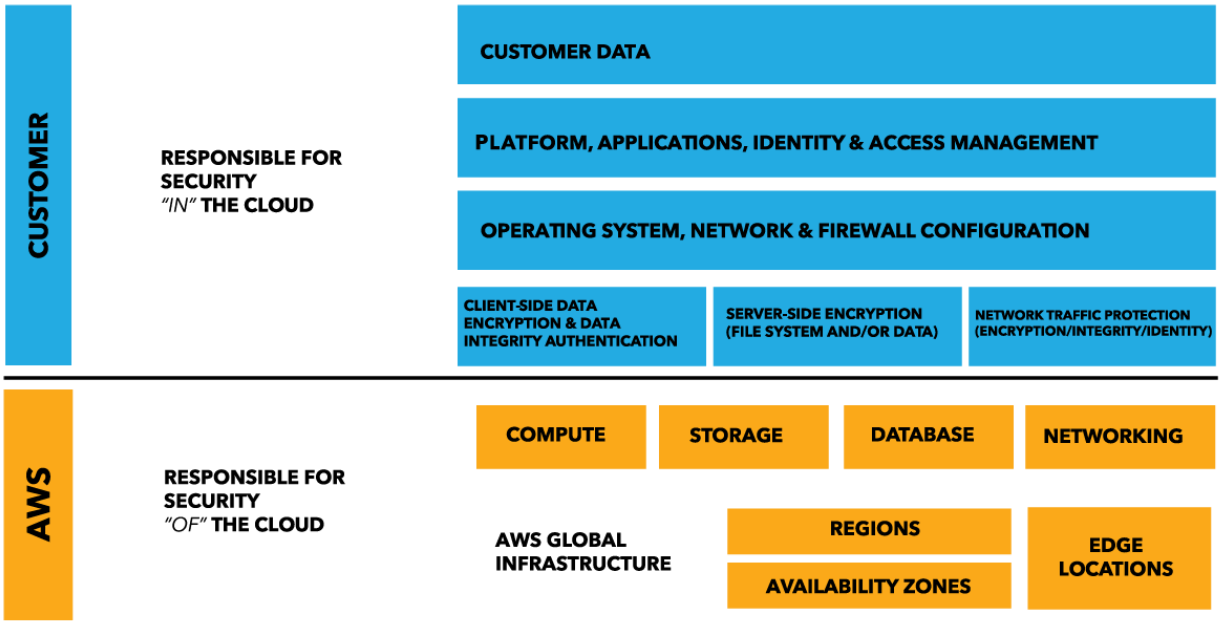


Figure 21: The AWS Shared Responsibility model.

The implication of the Shared Responsibility model is that you, as the researcher working with the data, will need to do two things: (1) you will need to identify data which data needs to be protected and why; and (2) implement a set of security procedures and controls that build on top of the AWS security features to protect sensitive data.

Practically speaking, research data originating from human subjects tends to span a spectrum of security needs that encompasses the Public Domain (no security needed) through to Protected Health Information (may fall under a geography's regulation). To identify which data needs to be secured, you will need to categorize data within that spectrum. **Figure 22** illustrates an example of taking some well-known public archive and placing it within the spectrum.

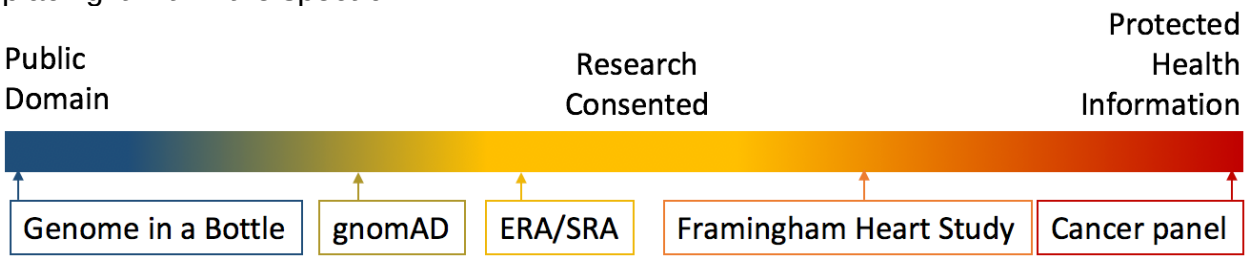


Figure 22: The spectrum of data classification for security and compliance. Genome-in-a-Bottle data are in the public domain; gnomAD, ERA, and SRA release some data within the public domain, but restrict access to individual genomes; all Framingham data is restricted access for research use; finally, a cancer gene panel produced in service of making treatment decisions typically falls under regulatory requirements for Protected Health Information (PHI).

For anything that you determine falls outside of the Public Domain, we recommend following a minimum of security controls, such as encrypting data at rest, following the principle of "least privilege" for network accessible resources, and robust logging of

operations made within the system. We recommend that researchers involve their institutional IT group to help with implementing these tools and procedures.
We also recommend that researchers align and document the security procedures for managing data and compute resources to a recognized security and compliance framework, such as NIST 800-171. This provides the necessary controls for protecting the data, and gives you a template for future applications for access to other sensitive data sources.

5.2 Meeting security and compliance requirements with AWS Quick Starts

AWS manages dozens of compliance programs in its infrastructure. This means that segments of your compliance needs have already been met. Our **compliance website**[28] helps you understand the robust controls in place at AWS to maintain security and data protection in the cloud. That understanding will allow you to determine your responsibilities in the Shared Responsibility model, but there is still the work of identifying which specific services and service features you can utilize to meet your needs. To help, AWS has published security and compliance-focused Quick Starts with reference architectures aligned to compliance frameworks such as NIST 800-171.

By tying together our governance-focused, audit-friendly service features with applicable security compliance regulations and audit standards, these AWS compliance enablers will help you to **use AWS resources securely**.

In addition to the reference architecture and best-practice documentation, the security and compliance quick starts also provide you with a Security Control Matrix (SCM) to bootstrap documentation of your systems, data, and operating procedures. The SCM Excel document contains a table of listing the security controls that apply to a given compliance framework, and (if applicable to the control) maps what the reference architecture does to meet that control. Researchers (or their institutional IT department) can use the SCM to fill out the rest of the worksheet with their added controls and procedures. **Figure 23** illustrates an excerpt from a SCM.

The **NIST 800-171** AWS Quick Start package of templated architecture, together with the documentation and the SCM, can be generalized to other compliance frameworks relevant to other industries and geographies outside of the US. In addition to NIST based controls, AWS has provided Quick Starts for **UK Official** workloads, as well as some that leverage commercially licensed solutions to add more layers and implement secure systems faster.
The **HIPAA** Quick Start tutorial walks you through setting up a model environment that can help organizations with workloads that fall within the scope of the U.S. Health Insurance Portability and Accountability Act (HIPAA). The Quick Start includes AWS CloudFormation templates, which

[28] https://aws.amazon.com/compliance/

automatically configure the AWS resources and deploy an example multi-tier, Linux-based web application in a few simple steps, in about 30 minutes.

UK Cloud Security Principles	CSP Description	Mapping with Center for Internet Security (CIS) Critical Security Controls (CSC) Version 6.1	Implementation Responsibility	Addressed by AWS UK Quick Start Reference Architecture	Implemented by other AWS controls	Description of AWS Implementation	CloudFormation template Resource mapping(s)	AWS CloudFormation Template Name (Stack)
4. Governance framework	*The service provider should have a security governance framework that coordinates and directs their overall approach to the management of the service and information within it.*							
The service provider should have a security governance framework that coordinates and directs their overall approach to the management of the service and information within it.		N/A	AWS Responsibility	No	Yes	N/A	N/A	N/A
5. Operational security	*The service provider should have processes and procedures in place to ensure the operational security of the service.*							
5.1 Configuration and change management	Good configuration management processes should ensure that knowledge of the assets which make up the service, along with their configuration and dependencies, are known and accurate. Good change management processes should ensure any changes to the service (which could have an effect on its security) are identified and managed. They should also lead to detection of unauthorised changes. In a service where change is not effectively managed, changes may unwittingly introduce (or fail to fully mitigate) security vulnerabilities in the service.	1.1-1.6 2.1-2.4 3.1-3.3 3.6-3.7 9.1 11.1-11.2	Shared Responsibility	No	Yes	AWS architecture provided as JSON templates and deployed via AWS CloudFormation	All resources in template	All templates
5.2 Vulnerability management	Occasionally, vulnerabilities will be discovered which, if left unmitigated, will pose an unacceptable risk to the service. Robust vulnerability management processes are required to identify, triage and mitigate vulnerabilities. Services which do not have effective vulnerability management processes will quickly become vulnerable to attack, leaving them at risk of exploitation using publicly known methods and tools.	4.1-4.8 8.1-8.6 9.3 11.5 12.8 20.1-20.8	Shared Responsibility	No	Yes	N/A	N/A	N/A
5.3 Protective monitoring	Effective protective monitoring allows a service provider to detect and respond to attempted and successful attacks, misuse and malfunction. A service which does not effectively monitor for attacks and misuse will be unlikely to detect attacks (both successful and unsuccessful) and will be unable to quickly respond to potential compromises of consumer environments and data.	3.5 5.4-5.5 6.6 11.3 12.2-12.4 12.9-12.10 16.8 16.10	Shared Responsibility	Yes	Yes	AWS CloudTrail records API calls to create, delete and modify resources CloudWatch Alarms detect unauthorized access attempts and send to SNS topic IAM activity and creation of AccessKeys send notifications with CloudWatch Alarms. IAM policies prevent start/stop of CloudTrail, S3 bucket policies protect access to log data, Alerts are sent if CloudTrail is disabled, Config rule in global-02 provides monitoring of CloudTrail enabled	AWS::CloudTrail::Trail AWS::CloudWatch::Alarm AWS::Logs::MetricFilter AWS::IAM::Policy AWS::Config::ConfigRule	template-iam template-logging template-config-rules

Figure 23 - Excerpt from the UK Official Security Control Matrix that supports that supports NSCS and CIS security controls for the UK.

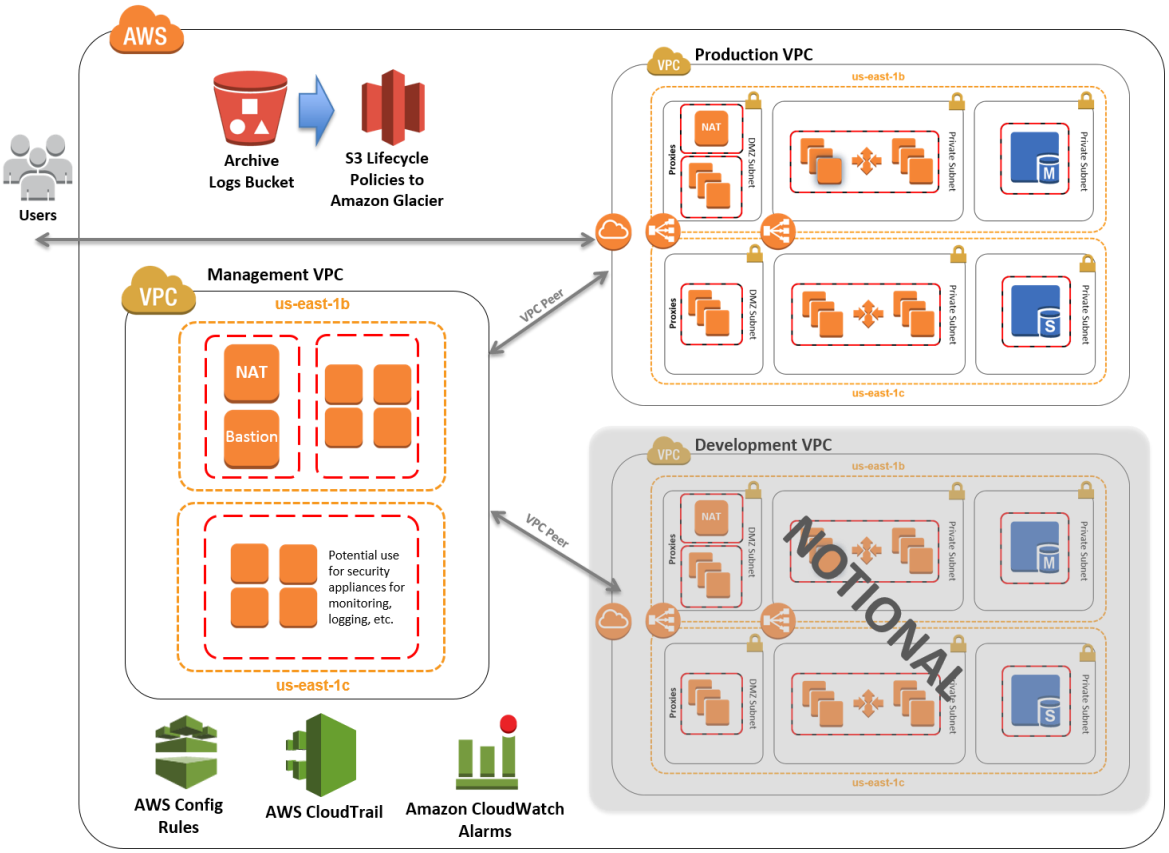


Figure 24 - General architecture used in the HIPAA, NIST, and UK Official Quick Start tutorials.

6 Working with Compute

Most scientific work includes computation. You may need to run a small analysis. A data reduction job that has to be repeated a million times on every experimental measurement coming in from your lab. Or you may need a handful of cutting-edge simulations of climate models or quantum chemistry that each demand a huge number of compute cores. AWS has a variety of compute services for each of those use cases.

The cloud paradigm works as well for compute as it does for storage. Cloud elasticity means **there's always room for N+1**: your simulations are not constrained to the fixed size of a cluster in your department's basement, but can scale to however many cores are needed for the question that's on your mind. And if you're running ensembles or parameter sweeps, you can run them all at the same time (horizontal scaling) instead of one at a time. **There's no queue in the cloud**: your AWS compute instances are all yours, and a job scheduler is optional. Pay-as-you-go means you can stop paying for idle servers, and **avoid the painful choice between idling** (typical underutilized department clusters) **and long wait times** (typical large compute centers). And you can choose the best compute tool for each application – **CPU today, GPU tomorrow**.

6.1 The Amazon Elastic Compute Cloud (Amazon EC2)

The foundational AWS compute service is the **Amazon Elastic Compute Cloud (Amazon EC2)**, where you can create virtual servers on demand. You "spin up instances" (start virtual servers in EC2), attaching some elastic disk storage (Amazon [EBS) to them, and enclosing them within custom-designed networks called Virtual Private Clouds (VPCs) that include built-in firewalls to add layers of protection. **Figure 25** shows the launch of a 4-core CentOS7 compute instance with a few clicks. A few clicks and minutes later you will be able to access the machine via Secure Shell (SSH) and begin running software or importing data.

Creating instances happens so often inside the cloud that we've built a lot of helpful tools to "freeze-dry" configurations so you can reconstitute them in a flash. Automating this activity will help you not forget important steps like making sure the firewalls are on or attaching your dataset to your computer. Automation also makes it easier to share your scientific pipeline with collaborators.

While this can be done conveniently through the AWS Management Console, nearly everything can also be done via a command line interface or one of several Application Programming Interfaces (APIs; choose your favorite language from a long list). This means that repetitive tasks or complex architectures can be turned into shell scripts. There are more advanced levels of automation, e.g. AWS CloudFormation, but by now you'll get the idea that you can go from simple ideas to complex creations easily.

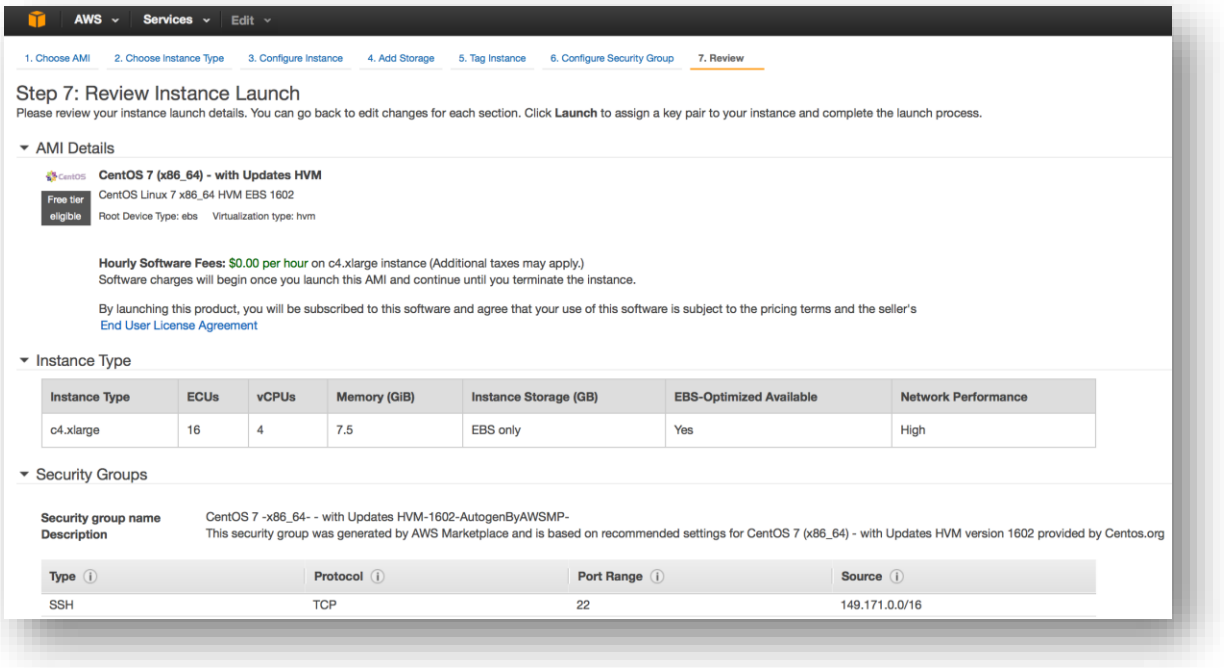


Figure 25 - Launching an EC2 instance only takes a minute or two and is very customizable.

6.2 Amazon EC2 Compute Instance Types

Amazon EC2 instances are virtual servers built from virtual machine images called Amazon Machine Images (AMIs) that you can run on a wide range of "instance types". Instance types represent differently scoped and sized computer servers. They range from tiny, single-core versions that might be great for running a wiki through to large instance types with terabytes of RAM or specialist GPUs.

The table below outlines the most common instance families you will encounter. Within each family, you choose at boot time what size instance you want to run to suit your application's needs. The size of each instance within a capability or instance family looks very much like a familiar t-shirt size-naming scheme.

For example, if you need 4 cores of our high-performance "C4" instance, after some studying of the Amazon EC2 instance family list, you might choose a "c4.xlarge." Later, if you wanted more cores for a larger problem, you'd reboot into a c4.2xlarge (8 cores) or a c4.8xlarge (36 cores). The same applies if you wanted to try running the same code with more memory—you can give yourself an upgrade in minutes. If it doesn't positively impact your throughput the way you'd hoped, you can always go back to the smaller instance or lower-memory instance and save money. If your application doesn't have specific requirements like GPUs, and you're not sure where to start, the C4, C3, and M4 families are generally good starting points.

Instance Family	Current Generation Instance Types
General Purpose	T2 instances are Burstable Performance Instances that provide a baseline level of CPU performance with the ability to burst above the baseline. They are great for simple, bounded applications like web servers for workgroup activities. M3 and M4 instances are more compute-intensive, general purpose instances and provide a balance of compute, memory, and network resources that are a good choice for many applications. We often suggest starting with M4 and working up or down from there, gaining insight from experience.
Compute Optimized	C3 and C4 instances are ideal for compute-bound or floating-point-intensive applications that benefit from high-performance processors. C4 instances use Intel Xeon CPUs specially designed for AWS that offer extra performance beyond what's commonly available in a standard server. They are the best choice for many High-Performance Computing (HPC) applications.
Memory Optimized	R3 instances are optimized for memory-intensive applications and offer lower price per GB of RAM, scaling up to 244 GB of RAM in a single instance. Our X1 instance family provides up to 2 TB of RAM for applications that require large memory footprints.
Storage Optimized	I2 and D2 instances are intended for applications that have extreme I/O requirements. I2 instances deliver tens of thousands of low-latency, random I/O Operations Per Second (IOPS), while the D2 family is designed for workloads that require high sequential read and write access to very large datasets on local storage.
Accelerated Computing	G2 and P2, the accelerated computing instance families, use GPU accelerators to perform some functions, such as floating-point calculations or graphics processing, more efficiently than in software running on CPUs. P2 instance types use NVIDIA K80 GPUs, offer up to 16 GPUs per compute instance, and are great for numerical simulation or deep-learning applications. The F1 instance type offers FPGA accelerators.

There are more instance types than listed here, including some older instances that still provide good processing power (and frequently are available at very low prices in the spot market). The complete guide, with a lot more detail, is available on our website at https://aws.amazon.com/ec2/instance-types/.

None of our instances, except the t2 series, are oversubscribed. That is, the specified hardware resources ("2 cores, 24GB of RAM, 10Gbps of network bandwidth") is fully and reliably yours.

6.3 How are EC2 compute instances priced?

AWS pricing generally follows the pay-as-you-go model, where you pay exactly for what you use. For example, if you store 100GB in S3 for 3 months and then delete the data, your total cost is the unit price in $ per GB per month, multiplied by 100 (GB) and by 3 (months). You can find prices for all services on the AWS website. AWS pricing is tiered for most services, including storage and data transfer, so the more you use, the lower the unit price per gigabyte.

For EC2[29] this "on-demand" price model is accompanied by additional, discounted pricing models:[30]

- **On Demand – Pay as you go**. No minimum commitment or long-term contract is required. Users can turn off cloud resources and stop paying for them when they are not needed. Amazon EC2 instances paid for in this way are called "On Demand." This is the easiest way to use EC2 instances. It's the best choice for work with fluctuating compute needs that has to happen right on time.
- **Reserved Instances (RIs) – Pay less when you reserve capacity**. You can invest in reserved capacity, and receive a significant discount in return for your commitment to using a number of compute instances for long periods of time, e.g. 1 year. This results in overall savings of up to 60% (depending on the type of instance reserved) compared to equivalent on-demand capacity. RIs are the best choice for long-term work that needs a fixed amount of compute power.
- **Spot Instances – Bargain prices on Amazon's unused capacity**. AWS operates a bidding market for leftover Amazon EC2 capacity, which allows you to run workloads at a fraction of the on-demand rate. Savings are often 50% to 90%. Spot instances are the right choice for work that is not time-critical (because the compute capacity may not be available immediately), where cost is very important (because it's very cheap), and where the compute tasks can tolerate an interruption (because the compute instance will become unavailable occasionally, requiring you to checkpoint and restart some jobs). It's a very popular option in the science community. Some application stacks use Spot Instances natively.

For example, a weather scientist wants to run thousands of iterations of some prediction models to analyze the influence of a certain parameter. This work just needs to get done sometime next week– so she runs it as cheaply as possible using *Spot instances*. It saves her 75% off the on-demand cost.

She also has to deliver up-to-the-minute wind forecasts that are used to decide whether firefighters need to be evacuated from a forest fire, and she provides a daily weather forecast to the local TV station. These jobs cannot tolerate any delay and require *on-demand* EC2 instances.[31]

Finally, she runs a web server that makes some of her datasets available to her colleagues. Since that server is always on, she hosts it on a *Reserved Instance* and pockets the 50% savings.

[29] And a few additional AWS services, e.g. Relational Database Service (RDS).
[30] We do not discuss Dedicated hosts and Scheduled Reserved Instances here.
[31] "Scheduled Reserved Instances" might also be appropriate for the daily weather forecast.

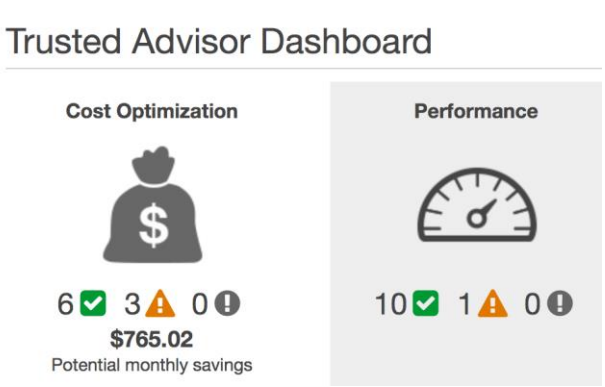
6.3.1 An EC2 cost scenario

On-demand. Let's say you've decided you need a 4-core server to test your molecular dynamics code, which is floating point intensive and benefits from Intel's most recent Xeon CPUs.

A c4.xlarge in our N. Virginia region is (at the time of writing) 19.9c[32] per hour in the On-Demand market. For many instances types, particularly Linux instances, your AWS account is **billed per second** at an hourly rate of 19.9 cents. So if you terminate the instance after 10 minutes, you pay about 2 cents. The minimum charge is 1 minute (1/3 cent). For some instance types, particularly Windows instances, your account is billed in hourly increments, i.e. 19.9 cents whether you use the instance for 10 or 60 minutes.[33]

After a few hours of working on your simulation and tweaking some variables, you're ready to run it at a larger scale, for which you'll need double the cores. So you switch to a c4.2xlarge instance, and **your cost per hour is now 39.8 cents**.

At the end of this run, you back up the data you worked on in S3 at a cost of 2.3 cents per GB per month. Or you may decide to snapshot a copy of the entire instance (5 GB will cost $1.38 per year), so you can start from where you left off when you come back to this task in a few days. Then you terminate the EC2 instance to stop paying for it.



Reserved Instances. Success happens: after some weeks of on-off work on this project, you're now running workloads almost around the clock. Your intuition tells you (perhaps with some help from the AWS Trusted Advisor) that you can get a better deal for your frequent usage. If this usage is going to extend for a year or more, you can purchase a Reserved Instance (RI). Since you're committing up front for your future usage, prices for RIs come at a discount. You can commit for one year or three years, and pay all the charges up front or hold a little back to allow room for some periods of low usage (sabbaticals, or heavy teaching loads during semester). You choose a one-year RI for our c4.2xlarge, paid 100% up front (for example), which costs you 23.7 cents per hour, **40% off** the on-demand rate. A three-year commitment would have saved 61%.

Spot market. The following year, you return to running the MD code intermittently and at a low priority – you can now tolerate your simulation taking a few hours longer than normal on occasion, and you can cope with an occasional job failure. This makes you a strong candidate for the Amazon EC2 Spot market. In the Spot market, you launch your Amazon EC2 instances by **offering a bid**, saying how much is the **maximum** amount you're willing to pay. Your bid is compared to the Spot market price for your instance type, which fluctuates with demand over time. If the Spot price isn't higher than your Spot bid, your

[32] Pricing depends on the AWS region you work in. See https://aws.amazon.com/ec2/pricing/ . All AWS charges are calculated in US Dollars. We can, however, bill you in one of a number of currencies. See https://aws.amazon.com/blogs/aws/new-preferred-payment-currency-for-aws-canadian-dollars-cad/ for details.

[33] See details here: https://aws.amazon.com/blogs/aws/new-per-second-billing-for-ec2-instances-and-ebs-volumes

instances will start and you can run your simulation, paying only the market Spot price for that hour, which will often be even less than what you offered. If your price is too low, or if our capacity is constrained by a lot of demand from other users (using RIs or On-Demand Instances), we'll decline your offer and your bid will sit waiting for capacity to become available at your price point. As we write this, the Spot bid price for a c4.2xlarge in N. Virginia is 9.6 cent, **76% off the on-demand price**, and the probability of being interrupted or outbid is low[34]. If you run a Linux Spot instance for less than an hour, you'll be billed per second for the exact time you used the instance.

The other caveat with the Spot market is that you may be happily (and cheaply) running your analytical cluster and be interrupted: this means your spot instance is stopped or terminated and taken away, and your job that was running will stop. When this happens, there are options.

The **first option you have is to do nothing**: the interruptions may be so infrequent (one in a hundred, say) that getting 80% off the list price for 99 successful runs is a great cost reduction and totally worth it in return for occasionally having to restart a failed job.

The second option is to teach your code to poll the "instance metadata" to watch for a change in the termination time. If you're "marked for termination," you have two full minutes to react accordingly, and how you react is entirely up to you. You might snapshot your current application state e.g. by checkpointing, or send a signal to your process to dump its memory image to disk. As soon as the spot price drops to your spot bid again, the instance can start again, and your application can **pick up the job from the checkpoint** file.[35] But before you get busy pondering all those details, do seriously consider deploying the first option (i.e., doing nothing). Consider the aggregate savings before you plan for the one that got away.

Many people successfully run large workloads in the Spot market, sometimes harnessing more than 100,000 compute cores at a time. Researchers at Clemson University recently used more than **1 million concurrent vCPUs** for Natural Language Processing – that's comparable in core count to the largest supercomputers in the world.[36]

AWS offers tools to help you exploit the Spot market. The Spot Bid Advisor scores the probability of getting resources with different, simple bidding strategies. Defined Duration Spot instances run continuously for up to six hours, in exchange for a slightly reduced discount. This enables you to reduce costs when running finite duration tasks such as batch processing, encoding and rendering, modeling and analysis, and continuous integration jobs. A Spot Fleet request lets you combine several instance types and AWS availability zones, so you can say, in effect, "give me 500 cores using any of these instance types, as cheaply as possible". We encourage you to follow best practices.

Some application stacks make it even easier. For example, Alces Flight (see chapter 6) can built you a compute cluster using Spot instances. Another example, AWS Batch can run your containerized (Docker) jobs on Spot instances and automatically retries jobs that fail due to Spot instance interruption.

[34] See e.g. https://aws.amazon.com/ec2/spot/bid-advisor/
[35] https://aws.amazon.com/blogs/aws/new-stop-resume-workloads-on-ec2-spot-instances/
[36] https://aws.amazon.com/blogs/aws/natural-language-processing-at-clemson-university-1-1-million-vcpus-ec2-spot-instances

6.3.1 The AWS Free Tier

Most AWS services offer a "Free Tier'. This lets you try out the service without spending any money. It's only when you exceed the designated free quota that you will start to be billed for your usage.[37]

The EC2 Free Tier gives you free instance hours of the t2.micro instances in the first year after you create your AWS account. These t2.micro instances are great for light workloads (like hosting small websites), for trying out EC2 features for the first time, for testing any scripts you write to automate launching EC2 instances. Even if you exhaust your Free Tier allowance, the low cost of t2 instances means you can keep doing all this literally for pennies. Once you have everything figured out, you can switch over to the right production instance type. T2 instances are not suitable for High Performance Computing (HPC) or for running performance benchmarks.

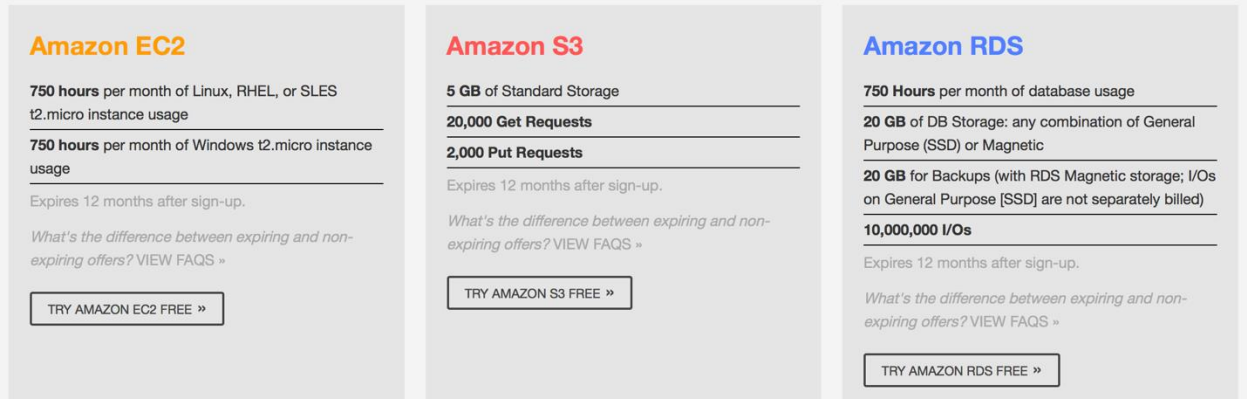


Figure 26 - See all of the free tier services here: https://aws.amazon.com/s/dm/optimization/server-side-test/free-tier/free_np/.

6.3.2 Price Drops

We continually focus on reducing our data center hardware costs, improving our operational efficiencies and lowering our power consumption. Each time we do, it allows us to pass savings back to our users through price reductions (62 times since we launched AWS in 2006). This means that the compute instances you're using today could be even cheaper tomorrow. Revisiting our example above, the c4.2xlarge instance type mentioned earlier saw a price drop for on-demand and reserved instances in N. Virginia by 5% in June 2015; again by 5% in January 2016; again by 5% in November 2016; and again by 11% for reserved instances only in May 2017. The cost of S3 storage has been cut many times over the years, most recently by 23% in December 2016.

6.3.3 Avoiding Lock-in

To make Reserved Instances more flexible, we've added a feature called "Convertible Reserved Instances." This allows you to trade an RI for one instance type for another

[37] Full information on AWS Free Tier services is available https://aws.amazon.com/free/. Also see the Free Tier Terms at https://aws.amazon.com/s/dm/optimization/server-side-test/free-tier/free_o/terms/.

instance type in the future. The details are on our website under Amazon EC2 pricing, but our Chief Evangelist, Jeff Barr, explains them nicely in a blog post.[38]

6.3.4 Easy hardware upgrades

AWS adds newer, faster instance types much more frequently than the 3-5 year on-premises hardware lifecycle. This gives a big boost to the value of your multi-year research budget: each year, your money will buy more performance on AWS, compared to locking in the performance of today's hardware for 3-5 years. Moreover, a hardware refresh on AWS can be as easy as editing the instance type in a configuration file and rebooting your AWS compute cluster – but on-premises hardware refreshes are often expensive and time-consuming.

6.3.5 Launching many AWS instances (AWS account limits)

In your new AWS account you can launch only a total of 20 EC2 instances. This default limit exists to prevent beginner mistakes in your early days. When you need more, you can request a higher limit in the EC2 console, as in the example shown below. This may take 1 or 2 days to be approved, so please plan ahead for that. There are similar limits for other services, but the limit on instances is the one you're most likely to encounter.

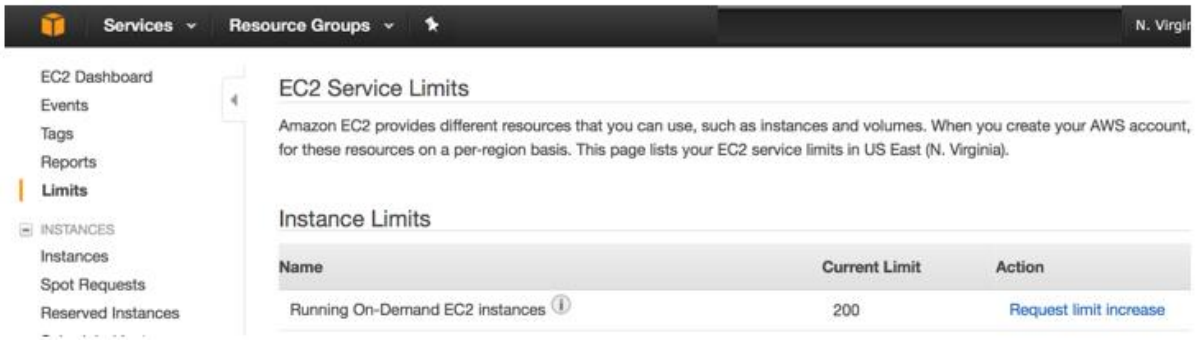


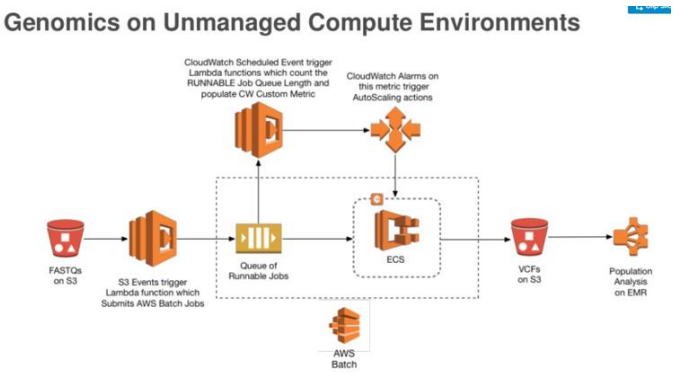
Figure 27 - EC2 Instance Limits.

6.4 Other compute services

The Amazon EC2 Container Service (ECS) lets you to deploy and manage "Docker images" on your EC2 instances. Containers easily package an application's code, configurations, and dependencies into simple building blocks. It helps applications to be deployed quickly, reliably and consistently regardless of the environment.

[38] Read Jeff's post here: http://tinyurl.com/h2udzgr.

AWS Batch goes a step further and completely manages the underlying compute infrastructure for you. You merely need to define your jobs by pointing at a container and specifying the accompanying 'docker run' command parameters. You can submit hundreds, thousands, or more of these jobs into your job queue, and instruct the queue manager how to deploy or prioritize them.



AWS Lambda is a truly serverless compute service. It lets you run code[39] without provisioning or managing servers. Just upload your code and Lambda takes care of everything required to run and scale your code with high availability. You can set up your code to be automatically triggered from other AWS services or call it directly from any web or mobile app. For example, you can define a Lambda function so that, every time you upload an image file to an S3 bucket, it automatically creates a thumbnail file, or spawns an associated Rekognition or Batch job for more advanced image processing. You can have thousands of Lambda workers at the same time, scaling up instantly and effortlessly, and billed in 0.1s increments. For a python interface to AWS Lambda, check out pywren (pywren.io), developed by the RISELab at U.C. Berkeley.

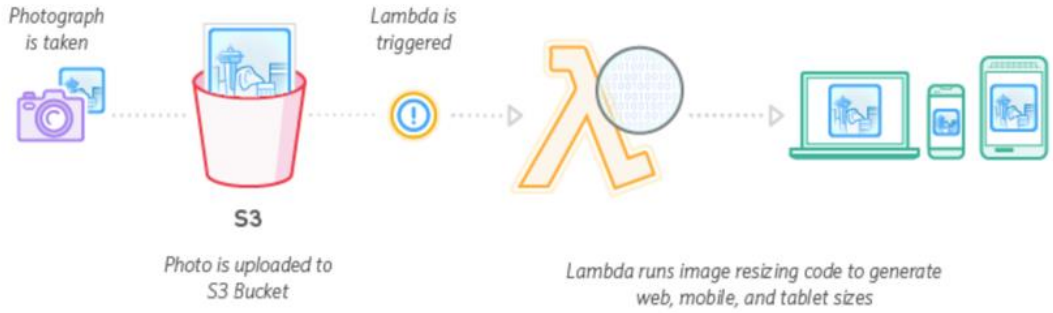
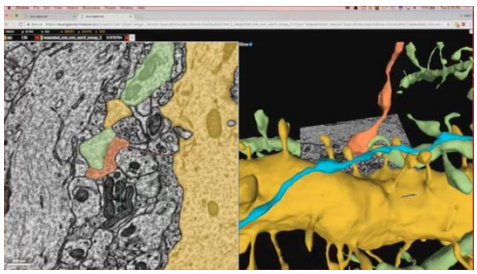


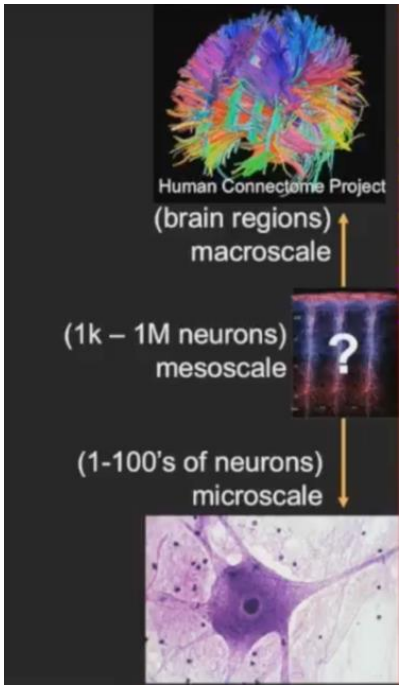
Figure 28 - Basic AWS Lambda usage: automatically process each image uploaded to your S3 bucket.

Serverless High-Performance Data Analysis (HPDA) in Neuroscience: The Intelligence Advanced Research Projects Activity (IARPA) Machine Intelligence from Cortical Networks (MICrONS) program seeks to revolutionize machine learning by better understanding the representations, transformations, and learning rules employed by the



[39] AWS Lambda supports Java, Node.js, C#, and Python code, with support for other languages coming in the future.

brain. Therefore, they are measuring the brain at the "mesoscale," the scale at which the hypothesized unit of computation, the cortical column, should exist. This yields some of the largest datasets ever collected – roughly 2-3PB of image data per sample, for about 50k-100k neurons and over 100 million synapses.



The Boss is a multi-dimensional spatial database provided as a managed service on AWS to hold all this data. Users can choose to perform different high bandwidth operations, like data ingest or image downsampling, with data sizes ranging from **2GB to 2PB** per request and sustained data ingest rates of over **4Gbps**. The Boss needs to scale quickly to meet each user's needs and also remain affordable and operate within a fixed budget. It does so by leveraging serverless AWS Lambda components40 to provide on-demand capacity –it would be time- or cost-prohibitive to meet such demands with a server-based architecture. Read and watch more.

6.5 Database services

AWS provides fully managed relational and NoSQL database services, as well as fully managed in-memory caching as a service and a fully managed petabyte-scale data-warehouse service. Or, you can operate your own database in the cloud on Amazon EC2 and Amazon EBS. Below are some of AWS's database options. Also refer to chapter 9.5 for information on Amazon Athena serverless interactive query service.

6.5.1 Amazon Relational Database Service (RDS)

Amazon Relational Database Service (RDS) is a web service that makes it easy to set up, operate, and scale a relational database in the cloud. It provides cost-efficient and resizable capacity while managing time-consuming database administration tasks, freeing you up to focus on science.

Amazon RDS gives you access to the capabilities of a familiar MySQL, Oracle, Amazon Aurora, SQL Server, MariaDB or PostgreSQL database. This means that the code, applications, and tools you already use today with your existing databases can be used with Amazon RDS. Amazon RDS automatically patches the database software and backs up your database, storing the backups for a retention period that you define and enabling point-in-time recovery. You benefit from the flexibility of being able to scale the compute resources or storage capacity associated with your relational database instance by using a single API call. In addition, Amazon RDS makes it easy to use replication to enhance availability and reliability for production databases and to scale out beyond the capacity of a single database deployment for read-heavy database workloads.

[40] It also uses Amazon S3, DynamoDB, SQS, and Step Functions. The project created an open source Python package called Heaviside to manage Step Function development and use.

6.5.1.1 Amazon Aurora

Amazon Aurora is a MySQL and PostgreSQL-compatible relational database engine that combines the speed and availability of high-end commercial databases with the simplicity and cost-effectiveness of open source databases. It provides up to five times better performance than MySQL with the security, availability, and reliability of a commercial database at one tenth the cost.

Amazon Aurora is a managed database service, built on a fully distributed and self-healing storage system that keeps your data safe. It provides enterprise-level capabilities including database monitoring, database cloning, cross-region copying and replication, AWS Identity and Access Management integration, and much more.

6.5.2 Amazon DynamoDB

Amazon DynamoDB is a fast, fully managed NoSQL database service that makes it simple and cost-effective to store and retrieve any amount of data, and serve any level of request traffic. All data items are stored on Solid State Drives (SSDs), and are replicated across 3 Availability Zones for high availability and durability. With DynamoDB, you can offload the administrative burden of operating and scaling a highly available distributed database cluster, while paying a low price for only what you use

Amazon DynamoDB is designed to address the core problems of database management, performance, scalability, and reliability. Developers can create a database table that can store and retrieve any amount of data, and serve any level of request traffic. DynamoDB automatically spreads the data and traffic for the table over a sufficient number of servers to handle the request capacity specified by the customer and the amount of data stored, while maintaining consistent, fast performance. All data items are stored on solid state drives (SSDs) and are automatically replicated across multiple Availability Zones in a region to provide high availability and data durability.

6.5.3 Amazon Redshift

Amazon Redshift is a fast, fully managed, petabyte-scale data warehouse service that makes it simple and cost-effective to efficiently analyze all your data using your existing business intelligence tools. It is optimized for datasets ranging from a few hundred gigabytes to a petabyte or more and costs less than $1,000 per terabyte per year, a tenth the cost of most traditional data warehousing solutions.

Amazon Redshift delivers fast query and I/O performance for virtually any size dataset by using columnar storage technology and parallelizing and distributing queries across multiple nodes. We've made Amazon Redshift easy to use by automating most of the common administrative tasks associated with provisioning, configuring, monitoring, backing up, and securing a data warehouse.

6.5.4 Amazon EC2 Relational Databases AMIs

An Amazon EC2 instance can be used to run a database, and the data can be stored within an Amazon EBS volume. Amazon EBS is a fast and reliable persistent storage feature of Amazon EC2. With Amazon EC2 Relational Database AMIs, you avoid the friction of infrastructure provisioning while gaining access to a variety of standard

database engines. Amazon EC2 Relational Database AMIs enable you to skip the infrastructure and hardware provisioning typically associated with installing a new database server, while still enabling you to exert complete control over the administrative and tuning tasks associated with running a database server.

6.5.5 Amazon ElastiCache

Amazon ElastiCache is a web service that makes it easy to deploy, operate, and scale an in-memory cache in the cloud. The service improves the performance of web applications by allowing you to retrieve information from a fast, managed, in-memory caching system, instead of relying entirely on slower disk-based databases. ElastiCache supports two open-source caching engines.

- Memcached – a widely adopted memory object caching system. ElastiCache is protocol compliant with Memcached, so popular tools that you use today with existing Memcached environments will work seamlessly with the service.
- Redis – a popular open-source in-memory key-value store that supports data structures such as sorted sets and lists. ElastiCache supports Redis master / slave replication which can be used to achieve cross AZ redundancy.

Amazon ElastiCache automatically detects and replaces failed nodes, reducing the overhead associated with self-managed infrastructures and provides a resilient system that mitigates the risk of overloaded databases, which slow website and application load times. Through integration with Amazon CloudWatch, Amazon ElastiCache provides enhanced visibility into key performance metrics associated with your Memcached or Redis nodes.

6.6 Tutorials

- Launching a virtual machine with Amazon EC2: https://aws.amazon.com/getting-started/tutorials/launch-a-virtual-machine/
- Run a serverless “Hello World” with AWS Lambda: https://aws.amazon.com/getting-started/tutorials/run-serverless-code/
- Deploy Docker containers on Amazon ECS: https://aws.amazon.com/getting-started/tutorials/deploy-docker-containers/
- Package and deploy a bioinformatics workflow on AWS using AWS Step Functions and AWS Batch: https://github.com/awslabs/aws-batch-genomics/tree/develop

For more sophisticated examples, see chapters 7 and 8, where we’ll stand up HPC clusters and use Jupyter notebooks on AWS.

7 HPC and Clusters

The flexibility of the cloud lets you do HPC in a more dynamic way than traditional, “static” HPC infrastructure. Cloud HPC compute clusters can be tailored to your application, scaled up to a larger core count, shrunk down, or commuted from CPU to GPU system in a matter of minutes. New paradigms like serverless computing allow you to rethink your HPC workloads for maximum scalability, ease of maintenance, and cost efficiency. You can also use the tools that 3rd parties have created for the vibrant and growing AWS HPC solutions community, encompassing HPC cluster tools, parallel file systems, and HPC codes from all walks of life: genomics, bioinformatics, quantum chemistry, computational fluid dynamics, weather forecasting, manufacturing, and more.

7.1 Easy-launch template-based HPC clusters

You could, in principle, already create an HPC cluster with what you’ve learnt about EC2 instances in Chapter 6: you would launch a number of instances, plus associated VPC, and storage, and you’d configure everything by hand to work together nicely. But in the cloud you want to script everything: creating an HPC cluster should be a simple, automated action, based on a template and some customization for your specific tasks. There are loads of benefits to this approach:

- Your cluster is ephemeral – tear it down anytime and start another one.
- Your cluster is personal: tailor it exactly the way you like, and there are no queues.
- You can easily change specifications -- say you needed a CPU cluster yesterday but you want a GPU cluster today. Sit back, give it a few minutes, and your hardware has magically transformed to follow your needs.
- Hardware upgrades are now as good as free: when AWS launches the c5 instance type, boasting Intel Skylake CPUs, just edit the configuration file for your cluster, changing “c4.8xlarge” into “c5.8xlarge”, and refresh the cluster. A few minutes later, your hardware is renewed. The next upgrade will come way sooner than the typical 3-5 year on-premises refresh rate.
- Share your configuration with your colleagues by just sending them the cluster configuration text file. They can spin up a replica of your cluster in their own account. Their jobs don’t crowd your cluster, and they pay for their own usage.
- Run multiple clusters at the same time, e.g. for different steps in an analysis pipeline (where each step may have its own hardware requirements); or for each member of ensemble calculations; or a cluster for each person in your team; …

Let’s look at a few popular offerings. All create a cluster suitable for HPC with a familiar feel: a head node that you SSH into, a fleet of compute nodes, shared storage, and your choice of several popular job schedulers. (See diagram below.) Because you’re in the cloud, you gain some additional powers: the compute fleet can automatically scale the available compute cores up and down depending on your needs, and you can use regular “on-demand” compute nodes or heavily discounted Spot instances. You pay for the AWS services the cluster uses, mainly EC2 instances and EBS storage. Some offerings have paid-support options separate from the free option we describe here.

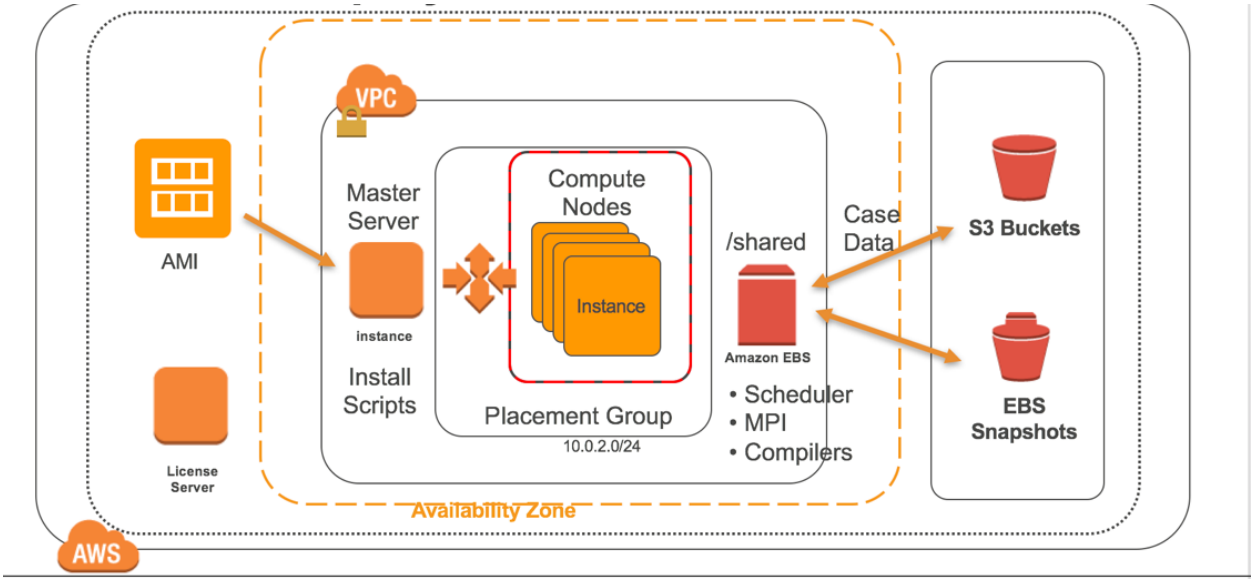


Figure 29 - Sample Cluster architecture on AWS.

Alces Flight creates a researcher-oriented HPC cluster. That cluster can be accessed via SSH, or you can open a (VNC) graphical desktop session. It comes with more than 1,300 preinstalled scientific applications, including genomics codes, libraries, compilers, and the lot. Cluster management tools help you sync the cluster's shared file disks with your long-term storage in AWS S3; or turn off hyperthreading on your cluster, etc. You can launch the Alces Flight cluster from the AWS MarketPlace (compare it to a free app in your phone's app store), where you confirm a few configuration settings. These settings are added to your personal cluster template, before the AWS CloudFormation service is used to stand up a cluster according to your specifications.

CloudyCluster also launches from the AWS MarketPlace. It provides a "self-service HPC cluster" with popular open HPC software, multiple storage technologies (OrangeFS, EFS, and S3), and a web interface that allows you to access your cluster from mobile devices. CloudyCluster Queue lets you submit jobs without explicitly creating a cluster first (similar to AWS Batch). CloudyCluster provides a detailed quickstart tutorial at http://www.cloudycluster.com/quickstart/ and has the distinction of having supported the largest cloud computing cluster to date with over 1 million concurrent vCPUs on AWS.

CfnCluster ("CloudFormation Cluster") is an open-source framework shared by team members at AWS. It shows how it's possible to build HPC-style clusters in the AWS Cloud, including integration with schedulers. It has a more "system administrator look-and-feel". CfnCluster facilitates Proof of Concepts (POCs) that help you understand auto-scaling in an HPC context and identify the path to production deployments. The Caltech Guttman Lab Case Study is a research example of CfnCluster used for lncRNA analysis: https://aws.amazon.com/solutions/case-studies/caltech-guttman-lab/.

(*StarCluster*, an older framework that used to be popular in academia, is no longer recommended as it has not been updated for a long time.)

7.1.1 Example of a template-based HPC cluster

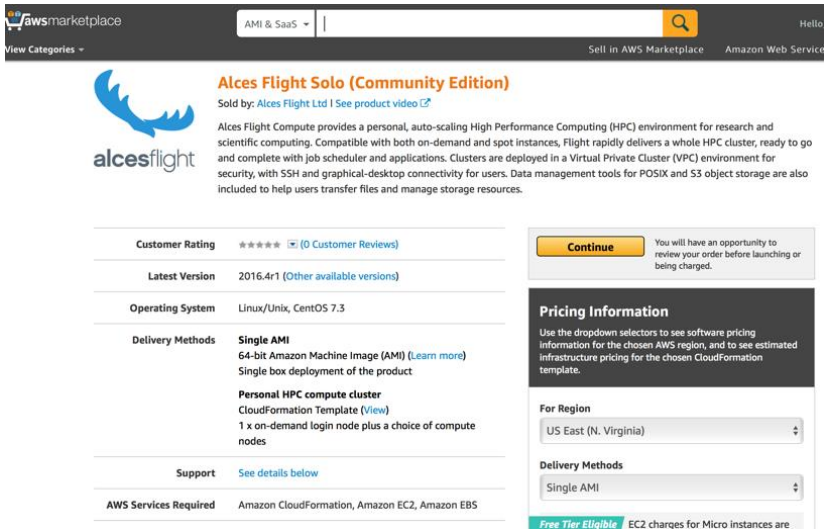


Figure 30 - Let's launch an Alces Flight cluster in AWS Marketplace.

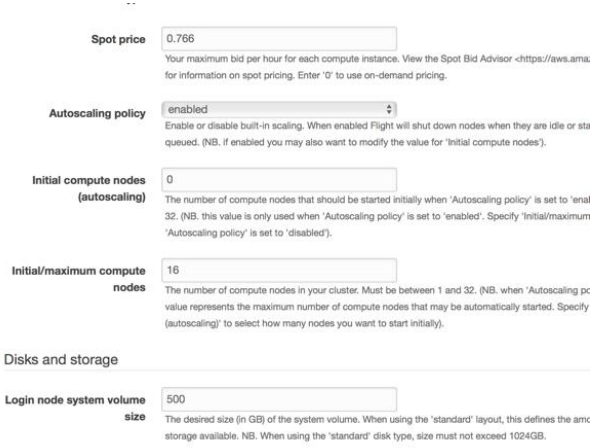


Figure 31 - We can specify a handful of configuration options. We choose Spot instances with a spot bid price of $0.766. The cluster will start with a head node and 0 compute instances, and is allowed to scale up to 16 instances.

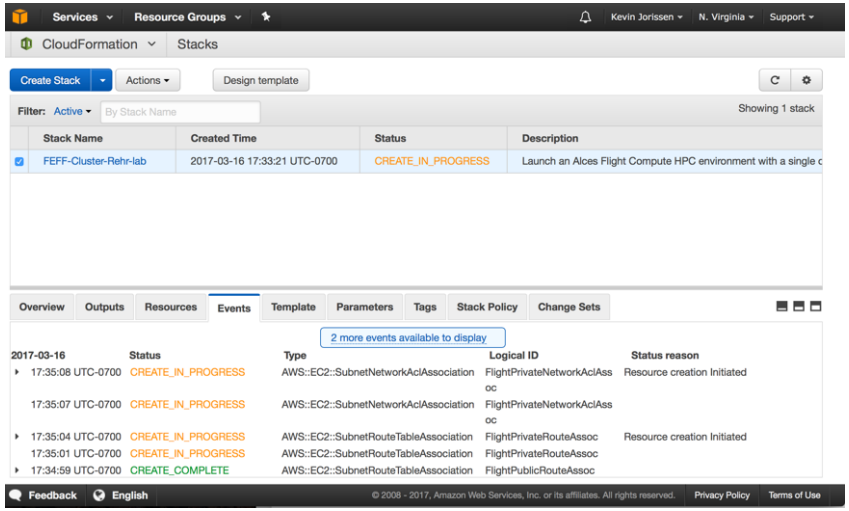


Figure 32 - After confirming, Alces Flight spins up the cluster as a "CloudFormation stack", which involves many little steps happening automatically while you get a cup of coffee.

Figure 33 - About 5 minutes later, we can SSH into the head node of the new HPC cluster.

```
[alces@login1(FEFF-Cluster-Rehr-lab) FEFFruns]$ more task.sh
#!/bin/bash

FILES=`ls | grep '.cif'`
FILES=`ls *.cif`

for i in $FILES
do
  j=`echo $i | cut -f1 -d.`
  mkdir $j
  cp $i $j
  cp feffjob.sh $j
  cp feff.inp.template $j/feff.inp
  cd $j
  echo CIF $i >> feff.inp
  qsub -cwd feffjob.sh
  echo submitted job $j
  cd ..
done


[alces@login1(FEFF-Cluster-Rehr-lab) FEFFruns]$ more feff.inp.template
CONTROL 1 1 1 1 1 1
EXCHANGE 0 0 0 0
SCF     4.0 0
COREHOLE RPA
XANES 8.0 0.07 0.0
FMS 6.0 0
TARGET 1
EDGE K

[alces@login1(FEFF-Cluster-Rehr-lab) FEFFruns]$ more feffjob.sh
#!/bin/bash -l
#$ -N FEFF$1 -pe mpislots 4 -o feffjob.out.$JOB_ID

#module load apps/im
#feffmpi is included in the path
feffmpi 4

[alces@login1(FEFF-Cluster-Rehr-lab) FEFFruns]$
```

Figure 34 - We submit a large array of several hundred jobs to the SGE scheduler. Each job is a simulation of the X-ray absorption spectrum of a material specified in a .cif file. Each job requires 4 cores, or 1 compute node, as our compute nodes are defined to be 4-core “c4.2xlarge” compute instances.

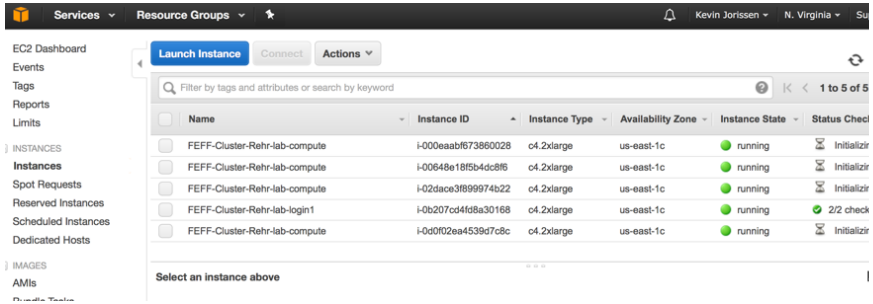


Figure 35 - The scheduler notices that it needs additional compute nodes to be able to handle this work (as it currently has none). It triggers ‘autoscaling’ of the cluster by a predefined step of 4 instances.

```
[alces@login1(FEFF-Cluster-Rehr-lab) FEFFruns]$ qhost
HOSTNAME                ARCH         NCPU  LOAD  MEMTOT  MEMUSE  SWAPTO  SWAPUS
----------------------------------------------------------------------------------
global                  -               -     -       -       -       -       -
flight-015              linux-x64       8  0.57   14.3G  292.0M     0.0     0.0
flight-045              linux-x64       8  2.38   14.3G  292.4M     0.0     0.0
flight-049              linux-x64       8  2.43   14.3G  292.0M     0.0     0.0
flight-060              linux-x64       8  0.50   14.3G  291.6M     0.0     0.0
flight-090              linux-x64       8  2.44   14.3G  300.9M     0.0     0.0
flight-099              linux-x64       8  1.03   14.3G  296.5M     0.0     0.0
flight-118              linux-x64       8  0.41   14.3G  286.8M     0.0     0.0
flight-123              linux-x64       8  2.07   14.3G  303.9M     0.0     0.0
flight-141              linux-x64       8  1.65   14.3G  297.9M     0.0     0.0
flight-170              linux-x64       8  2.10   14.3G  295.3M     0.0     0.0
flight-191              linux-x64       8  1.34   14.3G  298.3M     0.0     0.0
flight-204              linux-x64       8  3.92   14.3G  305.7M     0.0     0.0
flight-210              linux-x64       8  2.21   14.3G  311.2M     0.0     0.0
flight-218              linux-x64       8  0.41   14.3G  284.6M     0.0     0.0
flight-247              linux-x64       8  1.11   14.3G  291.6M     0.0     0.0
flight-251              linux-x64       8  0.78   14.3G  284.9M     0.0     0.0
[alces@login1(FEFF-Cluster-Rehr-lab) FEFFruns]$
```

Figure 36 - It will keep automatically adding groups of 4 compute instances until it reaches the maximum of 16. Sure enough, eventually we have a fleet of 16 workers HPC’ing away.

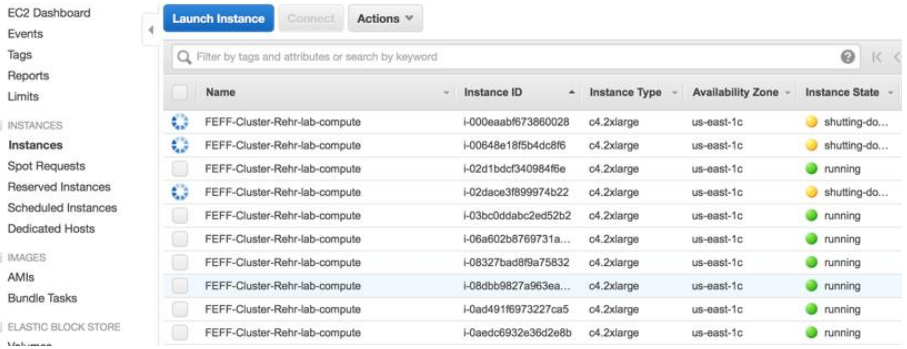


Figure 37 - Once the work is done and the queue is empty, the scheduler will start automatically tearing down compute nodes again so you don’t waste money paying for idle instances. (To turn off even the head node and completely stop incurring costs, go back to the CloudFormation console, select the stack, and delete it. After you’ve transferred any data you want to keep to Amazon S3 or your home computer, of course.)

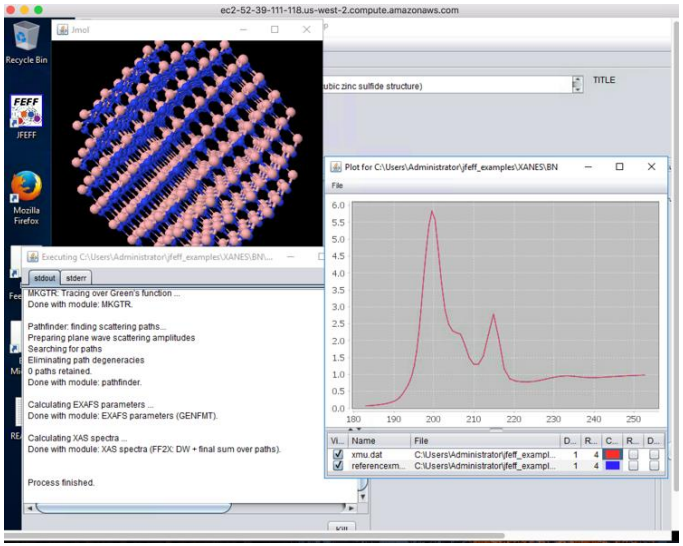


Figure 38 - At this point we can also take a look at some of the output data we created, e.g. this XANES spectrum of a BN crystal.

Figure 39 - How did we do on cost? The spot price was $0.095 or 87% off the “on-demand” full price of $0.766! So, we ran our 16-node (64-core) compute fleet + head node for 16x $0.095 + 1x $0.766 = a grand total of $2.29! (Plus a few cents for EBS storage.) Not bad!

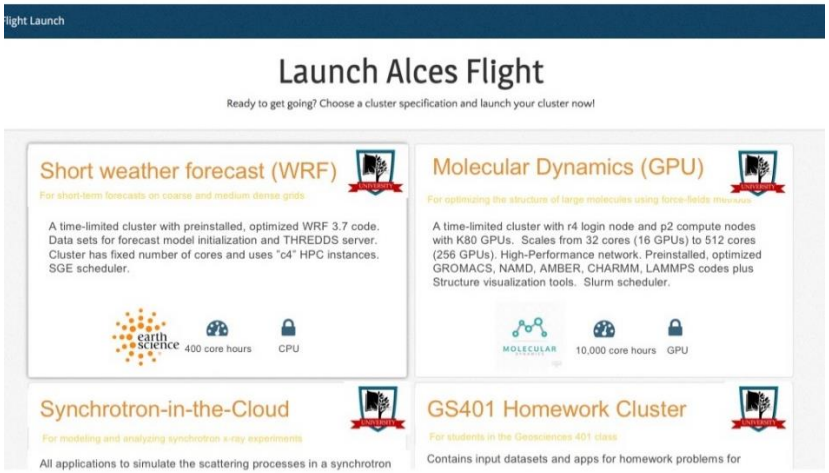


Figure 40 - The "Alces Launch" interface is an alternative to the Marketplace launch. It's based on a two-level approach. An administrator defines even more strongly templated clusters represented by tiles (e.g. "WRF short-term forecast cluster, 1000 core hours, GPU powered.”). An end user clicks the tile, enters an authorization code, and gets access to that pre-defined cluster type (after a few minutes) without any further fuss.

Other cluster technologies like **CloudyCluster** or **CfnCluster** differ in the way they’re launched, and in some features (e.g. preinstalled software or VNC sessions), but are similar when it comes to the overall workflow, autoscaling of the cluster, or the ability to use Spot EC2 instances.

7.1.2 High-Throughput or High-Performance?

High-Performance (HPC) jobs feature many worker threads with intensive communication requiring high-bandwidth and low-latency network connections. High-Throughput (HTC) jobs also feature many worker threads but don't require intensive communication or highly performant network connections. Both HPC and HTC workloads run very well on AWS.

HPC workloads require some additional "best practices"[41]: To keep the latency low, always place all cluster instances in a "placement group". A placement group is a logical rule ensuring that all instances are in close proximity on the network. Alces Flight always uses a placement group, while CfnCluster requires that you specify it in the configuration file. You should also choose instance types with 10Gbps or 20Gbps Enhanced Networking, such as c4 or m4 instances.

"General purpose" instance types, such as the t2 family, are not suitable for demanding HPC tasks. While you can use a t2 cluster in the free tier (with no charge) to debug your job scripts and configuration, you should switch to a compute optimized instance type for any benchmarking or production work.

7.2 Marketplace HPC Solutions

AWS Marketplace is an online store for applications and services that build on top of AWS to provide functionality such as HPC clusters[42], fluid dynamics simulators[43], or cluster file systems[44]. Each of these solutions bottles up a large amount of computational and orchestration complexity, which means when you click to launch, you're getting the built-in optimizations and best practices of the organizations supplying them and leveraging what they've learned from building complex services on AWS. Researcher can find, license, and immediately start these software and services. See also section 1.5.

The nature of AWS Marketplace also means that an innovative idea in one research lab in one corner of the world can become a popularly used service in hundreds of locations, with each user group facilitating their own usage of the tool with their own budget and resources—a very sustainable path to sharing content and tools with a whole community. This means that your group might be both a consumer of applications in AWS Marketplace and a vendor providing solutions.

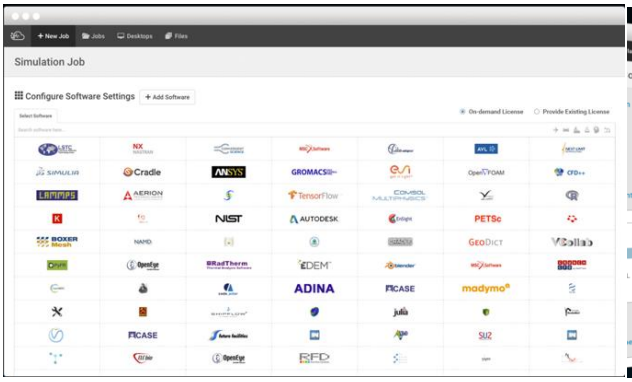
AWS handles billing and payments for both the underlying computing as well as any licensing charges. Applications on the Marketplace may also offer "Bring Your Own License" alongside the pay-per-use applications. You can also distribute your own application for free through the Marketplace[45] – indeed, there are many free applications available today.

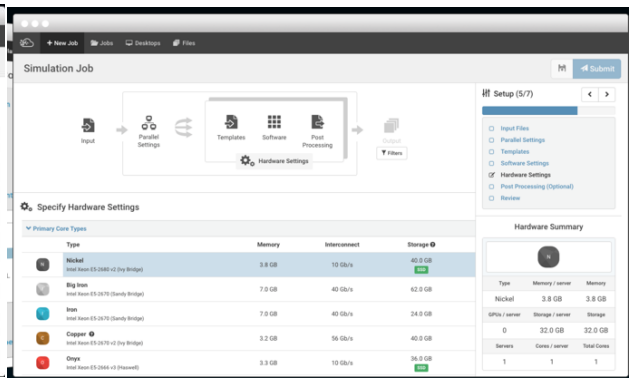
[41] If you are compiling applications, stay tuned for a forthcoming white paper detailing best practices for optimal performance.
[42] Like Alces Flight, CloudyCluster, and others.
[43] OpenFOAM, for example.
[44] Intel's Lustre, or the Fraunhofer Institute's BeeGFS.
[45] What does "free" mean? Say you launch a free quantum chemistry code from Marketplace. This creates an EC2 instance preconfigured with that chemistry code. You'll pay AWS the regular cost of running an EC2 instance, but you don't pay any Marketplace fee to the organization who published that code on the Marketplace.

7.3 Partner and SaaS HPC solutions

Moving higher up the stack to increasingly user-friendly and managed HPC paradigms, we encounter AWS Partners – 3rd parties who have built a business selling specialized services built on top of the AWS Cloud. In this model, you typically interact with the Partner and not with AWS directly. Along with the AWS Marketplace, our Partners demonstrate that the AWS Cloud is much more than a room full of servers: it's an entire network of parties building the HPC solutions for many years to come. We give illustrative examples here; please see the Partner Listing in the second half of this handbook for a tantalizing portfolio of AWS Partners relevant to your work.

Rescale is an example of a Partner who offers a very user-friendly, "turn-key" platform for running HPC simulations in the cloud. HPC jobs are submitted and managed in a web-based workflow manager that abstracts much of the underlying cloud technology. Rescale currently offers 160+ software packages in their SaaS platform. It operates in AWS regions in multiple countries. A powerful administrative portal helps manage resources, licenses, and budgets. Rescale complies with the strictest industry standards for security, and implements data protection along every step of the way. (Below: Choosing an HPC application in the Rescale platform; defining a new simulation job and choosing the hardware specifications in the Rescale platform).





EnginFrame is an HPC portal where users and administrators can easily submit and control HPC applications, as well as monitor workload, data, licenses from within the same user dashboard, hiding the heterogeneity and complexity of the native interfaces. HPC clusters, data, licenses, and batch and interactive applications can be accessed by any client using a standard browser.

7.4 Containers, Microservices, and AWS Batch

We're seeing a shift from traditional, monolithic architectures deploying a comprehensive code base all at once - to a microservices architecture that modularizes the pipeline into subprocesses that can be deployed separately from one another on separate hosts. This modularized architecture is less costly and more agile. Docker containers are the most ubiquitous technology for packaging and running individual processes of a pipeline. Together, AWS EC2 Container Service (ECS), the Amazon EC2 Container Registry (ECR), and AWS Batch can be used to orchestrate container deployment across an EC2 instance cluster, and coordinate job execution in managed or unmanaged compute

environments. You can use the AWS Lambda Step Function Service to implement complex serial or parallel job dependencies in your workflow.

With AWS Batch, you package the code for your batch jobs, specify their dependencies, and submit your batch job using the AWS Management Console, CLIs, or SDKs. AWS Batch executes your jobs as **containerized applications** (e.g. a Docker container) running on Amazon ECS. You specify execution parameters, such as vCPU and memory, and job dependencies, and you can integrate with popular batch computing workflow engines and languages such as Pegasus WMS, Cromwell, and Luigi. Default job queues and compute environment definitions help you get started quickly. The 'Task Placement Engine' allows you to use placement strategies including instance type affinity management, bin packing based on memory, or task spreading based on Availability Zone. AWS Batch dynamically provisions and scales Amazon EC2 and Spot Instances based on the requirements of your jobs.

Our research users use Amazon ECS and ECR for machine learning applications, see e.g. https://aws.amazon.com/blogs/aws/aws-enables-consortium-science-to-accelerate-discovery/ . For a demo of lambda step functions, AWS Batch and ECS/R, check out https://gist.github.com/miachamp/416836576e5012f4c9dd92661d0bc231.

7.5 Serverless compute functions: AWS Lambda

Even further along the abstraction path from servers to containers are the AWS Lambda functions we encountered earlier in this handbook. In Chapter 1.3 we saw examples of AWS users building HPC scale applications with high scalability and throughput using legions of these managed, serverless little workers, for example using PyWren.

We expect to see more and more of these disruptive paradigms in HPC land, leading to a much more diversified portfolio of HPC technologies than the supercomputer dominated world of yesteryear.

7.6 Elastic MapReduce

Amazon Elastic MapReduce (EMR) is a managed cluster solution natively supporting a family of framework associated applications (Apache Hadoop, Spark, HBase, Hive, Pig and Presto) commonly used to process and analyze high-throughput genomics datasets. EMR is a managed service that enables fast and cost-effective automatic scaling of compute resources for your cluster based on your workloads.

Amazon EMR has the added advantage of integrating well with other AWS services (i.e. Databases, Amazon S3) to enable optimized data flow throughout the stages of analytical processing. For example, Presto capabilities in EMR facilitate data import from Presto-based Amazon Athena queries on S3 data objects. Users can also take advantage of the EMR File System (EMRFS) to use S3 as a data layer for running EMR cluster applications that require data persistence. For processing data on the master and core nodes of your cluster, EMR supports the Hadoop Distributed File System (HDFS). Application outputs should be backed up to S3 prior to cluster termination.

Amazon EMR is used by users all over the world for machine learning, economic data analysis, scientific simulations, bioinformatics workloads, and log analysis.

Amazon EMR is well documented[46] and frequently cited by users who have built widely available tools using Amazon EMR.

The AWS Management Console EMR entry portal provides options to quick start an EMR cluster. EMR is integrated with many of the other AWS services so that you can build customized workflows to fit your use cases. For example, using the Lambda service, you can set up an event trigger when data is uploaded to your S3 bucket that launches an EMR cluster via Amazon Data Pipeline for analysis of your data – fully automated!

7.7 Tutorials

Package and deploy a bioinformatics workflow on AWS using AWS Step Functions and AWS Batch: https://github.com/awslabs/aws-batch-genomics/tree/develop

- Use CfnCluster to set up a cluster and run an MPI hello world example: https://aws.amazon.com/getting-started/projects/deploy-elastic-hpc-cluster/
- Use Alces Flight for an OpenFOAM (computational fluid dynamics or CFD) simulation: http://docs.alces-flight.com/en/stable/getting-started/environment-usage/using-openfoam-with-alces-flight-compute.html
- Use Alces Flight for a NAMD (molecular dynamics or MD) simulation: http://docs.alces-flight.com/en/stable/getting-started/environment-usage/namd_on_flight.html
- For both Alces Flight tutorials, first launch an Alces Flight cluster as explained at http://docs.alces-flight.com/en/stable/launch-aws/launching_on_aws.html .

[46] https://aws.amazon.com/emr/

8 Machine Learning and other Advanced Services

Researchers are increasingly challenged with managing and analyzing ever growing data volumes. Whether you're working directly with new instruments generating massive data volumes or you're collaborating on datasets created by other researchers, the tools to curate, triage, and work with that data are changing. For example, machine learning is now used to automate the process of gaining original insights from large datasets.

These challenges are shared widely outside the world of research. AWS has developed advanced services not offered by traditional computing centers, in order to meet these modern needs. In this chapter we look at some of these advanced services, with a particular focus on machine learning and artificial intelligence.

8.1 Machine Learning and Predictive Analytics

Amazon Web Services provides powerful tools and libraries for applying machine learning for feature detection and predictive analytics to datasets.

AWS has several fully-managed artificial intelligence services, including natural language understanding, text-to-speech, automatic speech recognition, and image recognition as a service. These services provide APIs for you to interact with. By making these pre-trained ML models available as scalable web services, AWS lets you use AI services very quickly that would otherwise require you to build and manage your own complex machine learning solutions.

However, if you wish to build your own solutions and services, you can use individual AI engines such as Apache MXNet to do so. Many research groups bring their own machine learning algorithms, and tools and libraries, and these can be run effectively at scale on AWS as well.

Figure 41 - AI Services, Solutions and **Engines** below shows how the suite of machine learning engines, solutions and services on AWS fit together. You can choose the layer and AWS service that best fits the level of abstraction and the capabilities you require.

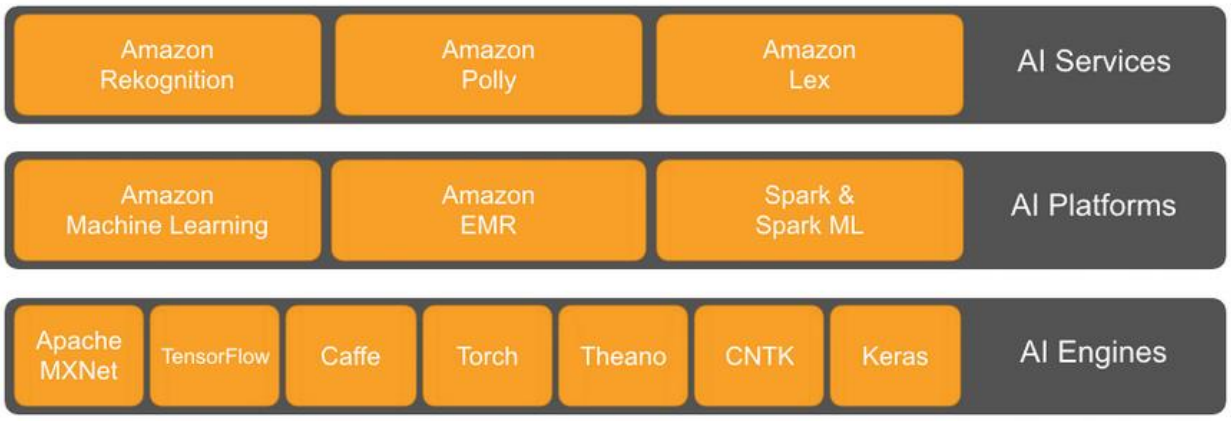


Figure 41 - AI Services, Solutions and Engines.

8.2 Amazon Machine Learning (AML)

Amazon Machine Learning (AML) is a scalable, supervised machine learning service. AML uses industry-standard logistic regression algorithms to automatically create models. Users can modify and update those models with a simple web interface.

AML treats several types of ML tasks:

- Binary classification problems
- Multiclass classification problems
- Regression classification problems

To use AML you define a data source and then let AML split that into training and evaluation datasets, or you can control what part of the data source is used for training and what part is used for evaluation. AML will also shuffle your data for you, and then process features to build a default model. You can modify the default model, and then AML will use the training data to train the model. AML provides a wizard for you to evaluate the model performance, and at this point you can set different scoring thresholds to define how the model defines prediction matches.

Once you're happy with the model, you can start using it to make batch predictions, or use the AML API to process more real-time predictions on data.

While you can't change the algorithm used to define the model, AML does let you tune the model and scoring thresholds to control how predictions are made. AML focuses on ease of use and is an excellent first choice for basic predictive analytics over small batch or real-time datasets. Take a quick AML tutorial or try these samples.

8.3 Deep Learning on AWS

The art of machine learning with deep neural network approaches is called Deep Learning. The availability of capable and powerful GPUs on AWS makes building, training and evaluating large deep neural networks possible.

AWS provides the AWS Deep Learning AMIs to quickly launch an Amazon EC2 instance with popular deep learning libraries and frameworks like Apache MXNet, TensorFlow, the Microsoft Cognitive Toolkit (CNTK), Caffe, Caffe2, Theano, Torch and Keras. Stay tuned for a hands-on tutorial in Chapter 9.

In some cases, you'll need to train complex deep learning models using more than a single EC2 instance. Training sophisticated deep learning models can take a lot of time, in some cases many hours or days. Distributing this work across a cluster and doing training in a distributed fashion can dramatically speed up the training process. Faster training has multiple benefits including:

- Iterative model design and improvement
- Getting research results faster
- Adapting more quickly to new data being acquired

The AWS Deep Learning Cluster simplifies running a distributed deep learning cluster on Amazon EC2, with the AWS Deep Learning AMI running on multiple CPU or GPU

instances. It uses AWS CloudFormation to automate the creation of the cluster. Shared storage is provided using the AWS Elastic File System service, and you can select the type of instance for your cluster, e.g. the P2 instance type with NVIDIA K80 GPUs. The workers are deployed into an auto-scaling group so more can be added later without affecting the existing cluster nodes.

Once you have trained your deep neural network, you can often persist these trained models and evaluate them independently on less powerful instance types.

8.3.1 Apache MXNet

Apache MXNet is a deep learning library focusing on flexibility, portability and performance. It provides both imperative and symbolic programming idioms, supports multiple languages like C, C++, Python, R, Julia, Scala etc., and can be run on embedded hardware through to desktop and server class hardware. It can take advantage of CPUs and GPUs and scales well for distributed training on AWS. The scalability that can be achieved during the training phase with MXNet is part of the reason the AWS Deep Learning Cluster make MXNet available by default.

To explore MXNet further in a distributed learning cluster, have a look at Distributed Deep Learning Made Easy on the AWS Compute Blog. This blog post will step you through creating an AWS Deep Learning Cluster and running the classic MNIST image recognition task using MXNet.

8.4 Artificial Intelligence as a service (Amazon Rekognition, Amazon Lex and Amazon Polly)

Figure 41 shows the higher-level AI services: Amazon Rekognition, Lex and Polly.

8.4.1 Amazon Rekognition

Amazon Rekognition provides a service and API you can use to add deep learning based image recognition to your applications. It's a service in front of a pre-trained deep neural network that is really good at feature detection in images.

Common use cases for Rekognition include:

- Facial recognition and face-based user verification
- Sentiment and demographic analysis
- Searchable image libraries

Rekognition provides all these features in a highly scalable and low-cost way so you can use it to analyze millions of images. It integrates natively with other AWS services such as Amazon S3 and AWS Lambda. By using integrations with these other services you can build scalable workflows to do feature detection on very large numbers of images.

8.4.2 Amazon Lex

Amazon Lex provides services to do automatic speech recognition (ASR) and natural language understanding (NLU). This lets you build applications to convert speech to text, and to recognize the intent of text.

Lex also integrates natively with AWS Lambda, so you can build software to run in response to specific triggers. Some applications for Lex include:

- Bots to control hardware devices
- Transactional bots to perform application actions
- Bots that can access and integrate research data stores

8.4.3 Amazon Polly

Amazon Polly lets you convert text into natural speech. It supports multiple languages, and is fast, so you can implement real-time interactive dialog in your applications.

Applications built using Polly can help people with reading disabilities or help visually impaired people use applications more easily.

By using Polly and cloud-based text-to-speech (TTS), you're able to achieve higher fidelity conversions than on small hardware solutions.

8.5 Amazon Athena

Amazon Athena lets you quickly explore and interrogate data stored in Amazon S3 using ANSI Structured Query Language (ANSI SQL).

Athena is powered by Apache Presto so you can define your schema dynamically for a given dataset in S3, and then use that schema to query those datasets. Athena supports a variety of well-known data formats like CSV, JSON, ORC, Parquet and Arvo.

It's fairly common in research to ingest large amounts of data that you want to capture and store durably, and then explore later in a flexible, ad hoc fashion. Athena supports this kind of workflow by integrating to S3 natively where you are possibly already storing research data. Also, because Amazon Athena is "serverless", there's no infrastructure for you to spin up, or set up, and you only pay for the queries you run while they're running. For example, Athena has been used for interactive analysis of genomics datasets, deriving insights from IoT heartrate sensors[47], running R, querying OpenStreetMap, analyzing OpenFDA public food safety data, analyzing New York City taxi records, and more.

Figure 42 shows this in action. Source data is stored in the S3 bucket. The Athena service stores the catalog and the schema definition. No data transformation or processing work has to happen on the raw data. Athena provides your schema as an overlay interface to your source data.

[47] This fun tutorial combines Amazon Athena, IoT, Kinesis Firehose, and QuickSight.

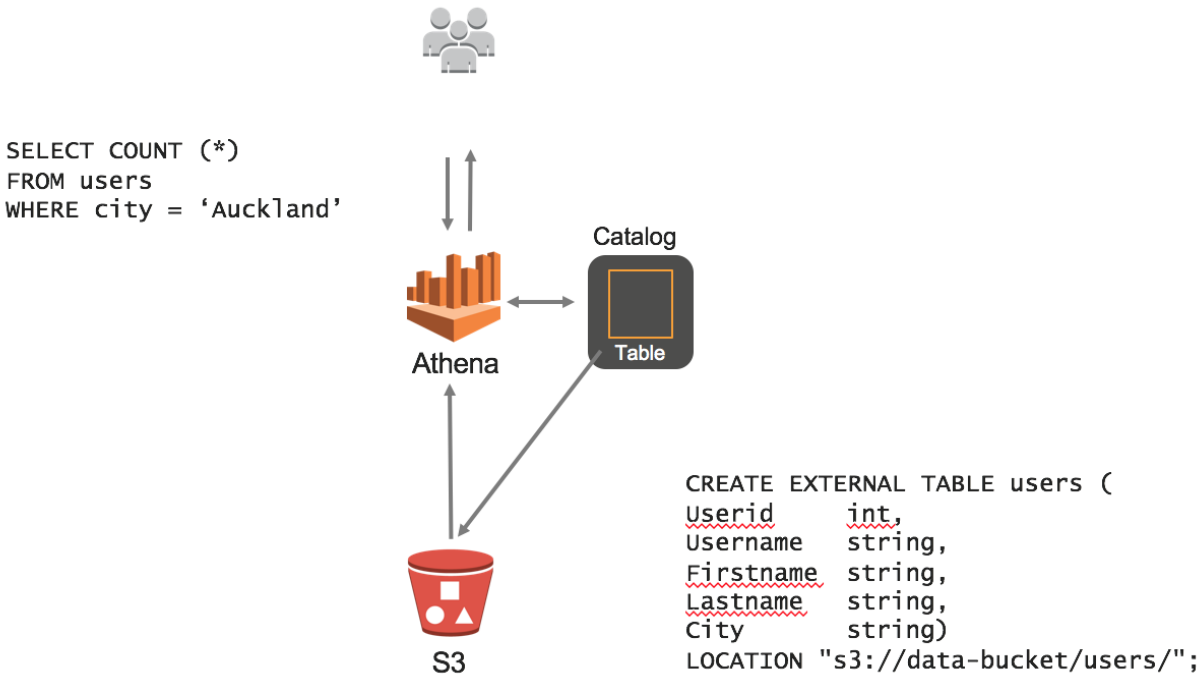


Figure 42 - Athena - Schema and Query workflow.

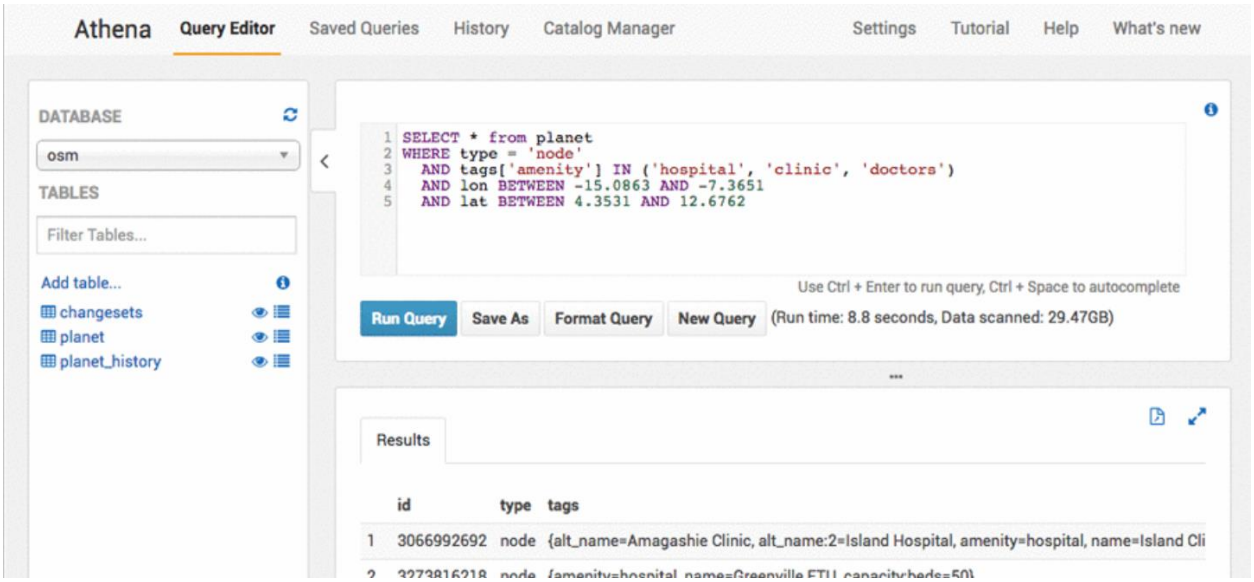


Figure 43 - Querying healthcare centers in West Africa using Athena on OpenStreetMap.

8.6 AWS Internet of Things (IoT)

"Internet of Things" (IoT) refers to the emergence of enormous numbers of small, connected devices. Equipped with rich datasets and using advanced analytics, IoT can

give us enormous insight into our world: Measuring vibrations from wind turbine blades and performing real-time analysis to determine maintenance needs before the blades fail. Reducing energy consumption in buildings by controlling lighting on floors where no one is present. Or creating self-driving vehicles that process environmental information to make split-second decisions to stop and avoid accidents. The collective knowledge about the physical world, gained through IoT, becomes the input for more efficiency, new business models, lower pollution, and better health. As an example of how AWS IoT is already transforming the use of data, NASA Jet Propulsion Laboratory (JPL) processes data in the AWS Cloud collected from instruments all over the solar system. Testing with AWS IoT allows NASA to integrate and process the data provided by sensors in mobile devices, smart devices, conference rooms, clean rooms, and beyond.

AWS IoT is a managed cloud environment that lets connected devices easily and securely interact with cloud applications and other devices. AWS IoT can support billions of devices and trillions of messages, and can process and route those messages to AWS endpoints and to other devices reliably and securely. With AWS IoT, your applications can keep track of and communicate with all your devices, all the time, even when they aren't connected. Get started with a 60-minute IoT application tutorial.

8.7 Amazon WorkSpaces

Amazon WorkSpaces offers an alternative for the expensive "workstation class" desktop machines many researchers have traditionally bought for computing or visualization work – typically, a machine with many CPUs, a large amount of RAM, perhaps one or more powerful GPUs, and excellent network connectivity to your research data stores. As noted by Jeff Christen - Information Science Instructor at Cornell University; "we can rethink how we teach with Amazon WorkSpaces. We can offer more to students – more interesting class content and more interactivity – without adding complexity for instructors. The sky is the limit."[48]

Amazon Workspace provides a virtual desktop (or "Desktop-as-a-Service") option for researchers. The Amazon WorkSpaces client can be run on Windows, Mac, Chromebooks, iPad, Fire tablets, and Android tablets. Amazon WorkSpaces provides support for Windows 7 and Windows 10 operating systems.

You can provision a "Performance" bundle, giving you 2 vCPUs and 7.5 GiB of memory, to power that virtual desktop for standard tasks. For more demanding visualization tasks requiring GPU capability, you can choose the "Graphics" bundle, which gives you 8 vCPUs and 15 GiB of memory with an NVIDIA GRID GPU[49]. Persistent storage is provided to each desktop using SSD storage.

The Amazon WorkSpaces Sync client lets you synchronize folders to Amazon S3, where you can store your research data long-term. This also lets you access data stored in this folder on any device where you use your WorkSpaces.

[48] https://aws.amazon.com/education/workspaces-in-edu/

[49] If you need even larger machines than those provided by Amazon Workspaces, you can use e.g. Citrix XenDesktop running on Amazon EC2.

9 Jupyter and Zeppelin Notebooks on AWS

Jupyter is a Notebook technology. It is a user-friendly and easy-to-use interactive programming environment in a web browser. It allows interactive data science and scientific computing across 40 different programming languages. Jupyter runs well just about anywhere: on your laptop, your desktop machine, or in the AWS Cloud. It allows researchers to share and exchange live code, datasets, and visualization so that they can collaborate more efficiently.

9.1 Jupyter on AWS

By running Jupyter Notebooks on AWS, you can take advantage of infrastructure web services you might not normally have access to in your Notebooks. With Jupyter on AWS, a scientific researcher, engineer, or technical user can run applications, write code, or post-process data using:

- Large memory instances, e.g. up to 2TB RAM instances (x1 or r4 families)
- Compute optimized instances, e.g. up to 36 vCPUs (c3 or c4 families)
- NVIDIA GRID GPUs (g2 family) or NVIDIA K80 GPUs (p2 family)
- Spark, Hadoop on Amazon Elastic MapReduce (EMR)
- HPC clusters (AWS Batch and CfnCluster)
- AWS Open Datasets like Landsat, NEXRAD, or The Cancer Genome Atlas

Essentially you can take advantage of any of the powerful capabilities made available to you on AWS – while enjoying the flexible Jupyter interface.

The Deep Learning AMIs available in the AWS Marketplace make it relatively easy to get started using Jupyter on EC2. These AMIs also bundles a number of languages and frameworks you might want to use with your Jupyter notebooks, including Python, Anaconda, and a number of machine learning and deep learning libraries.

Alternatively, if you want to build your own Jupyter installation using a custom configuration on EC2 you can follow the Creating and Using a Jupyter Instance on AWS whitepaper, which describes NVidia GPU capabilities like CUDA, to build your own customer Jupyter AMIs.

9.2 JupyterHub and Zeppelin with Amazon EMR

Apache Zeppelin is an open source GUI that creates interactive and collaborative notebooks for data exploration using Spark. You can use Scala, Python, SQL (using Spark SQL), or HiveQL to manipulate data and visualize results. Zeppelin uses the Spark settings on your cluster and can utilize Spark's dynamic allocation of executors to let YARN estimate the optimal resource consumption. Like Jupyter notebooks, Zeppelin notebooks can be shared among several users, and visualizations can be published to external dashboards.

Amazon EMR supports installing both JupyterHub and Zeppelin as an application on an EMR cluster during initial cluster set up. In the case of Zeppelin, a Spark Context is

created to run interactive Spark jobs on the EMR cluster where it's installed. By default Jupyter and Zeppelin notebooks are stored on the master node.

For more information see, Run Jupyter Notebook and JupyterHub on Amazon EMR and Running an External Zeppelin Instance using S3 Backed Notebooks with Spark on Amazon EMR.

9.3 Train a Machine Learning model on AWS through a Jupyter Notebook

In this **tutorial**, you will use Jupyter notebooks on AWS and Apache MXNet to build, train and evaluate a couple of different deep learning models using EC2 GPU instances. An updated version of this tutorial is available at https://github.com/scicolabs/data-science-ml/blob/master/Jupyter.md.

One of the big advantages of running Jupyter notebooks on AWS is that it allows you to take your *code and analysis to the data*, rather than having to move large datasets to your tools (an increasingly difficult task for large and growing research data volumes). This means you can be as proximate as possible to some of the most interesting datasets in the world, like Earth on AWS and the Cancer Genome Atlas. This combination means you can use some of the newest, and most powerful analysis tools and techniques right next to large datasets to do some pretty interesting things.

To successfully complete this tutorial, you should be familiar with:

- Basic Amazon EC2 concepts
- Basic familiarity with the Python programming language
- Basic familiarity with git and Github.com

9.3.1 Step 1 – Creating your Jupyter Notebook environment

We'll use the Deep Learning AMI (Amazon Machine Image) provided by AWS to run Jupyter and interact with common deep learning frameworks and libraries.

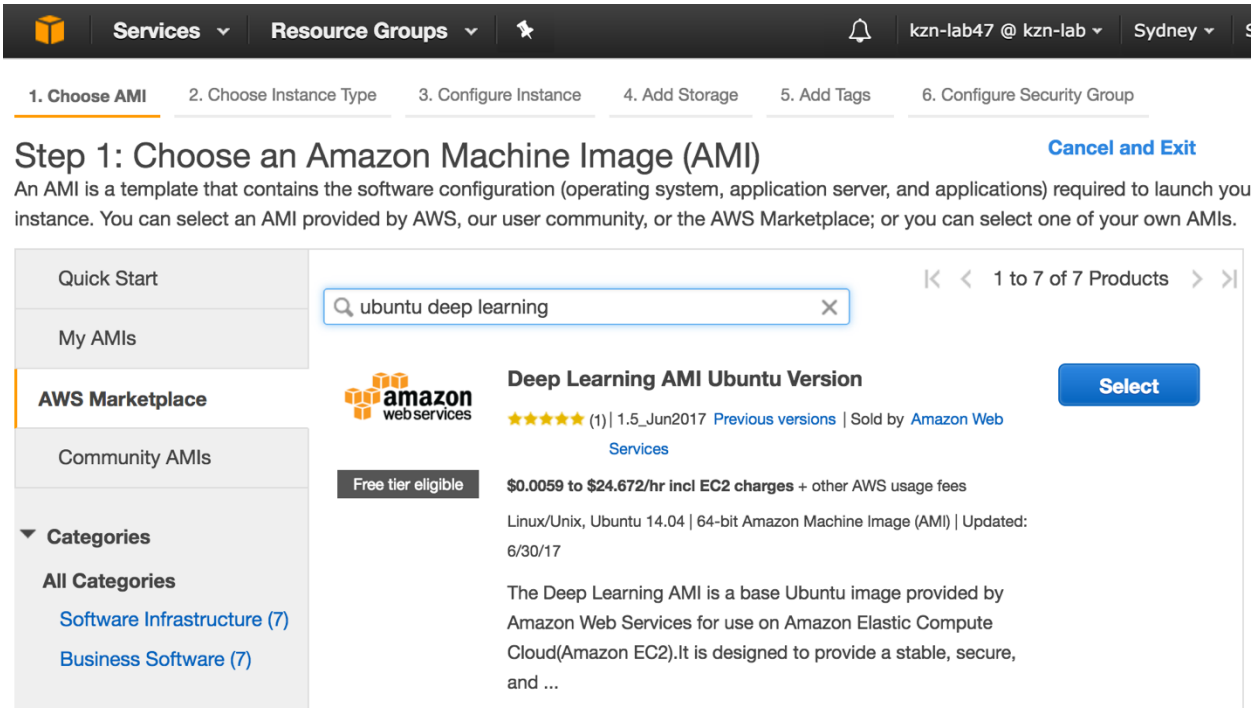
9.3.1.1 Creating your Notebook Environment

The first thing you'll need to do it log into the AWS Console using your web browser. You'll also need a valid Amazon EC2 Key Pair for SSH access to your EC2 instance. If you have one, you can happily continue with this tutorial. If not, refer to chapter 2.6 of this handbook for instructions.

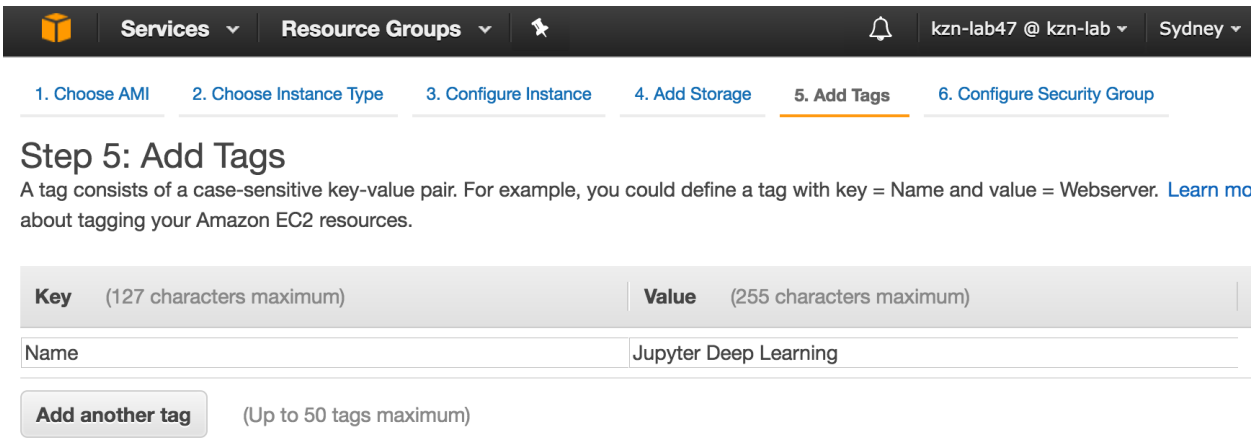
The AWS Deep Learning AMIs are available in the AWS Marketplace or via the Amazon EC2 Console. We'll launch the Ubuntu-based Deep Learning AMI via the EC2 Console.

To do this first make sure you've logged into the AWS Console, and browse to the EC2 Console. Then, from the EC2 Dashboard:

1. Click on the **Launch Instance** button
2. On the left-hand navigation, click **AWS Marketplace**
3. In the search box, type 'Ubuntu Deep Learning', and hit Enter.



4. Click the **Select** button
5. From the Instance Type screen choose the 'p2.xlarge' instance type. The P2 instance family has powerful NVidia K80 GPUs appropriate for training deep learning models.
6. Click the **Next: Configure Instance Details** button, and leave all the default settings in place
7. Click the **Next: Add Storage** button, and leave the default storage settings in place
8. Click the **Next: Add Tags** button. This screen lets us tag our Deep Learning instance with metadata so we can find it again. In this case, click the **Add Tag** Button and add a **Name** key, and a descriptive name in the **Value** field.



9. Click the **Next: Configure Security Group** button, and leave the default security group settings in place
10. Finally, click the **Review and Launch** button.
11. If you're happy with your configuration (you might want to check over it again just to be sure), click the **Launch** button
12. Here's you'll be prompted to select your key pair. *Carefully select a keypair*, and click the **Launch Instances** button.

Your new Deep Learning EC2 instance is now launching. It'll take a couple of minutes for this to complete.

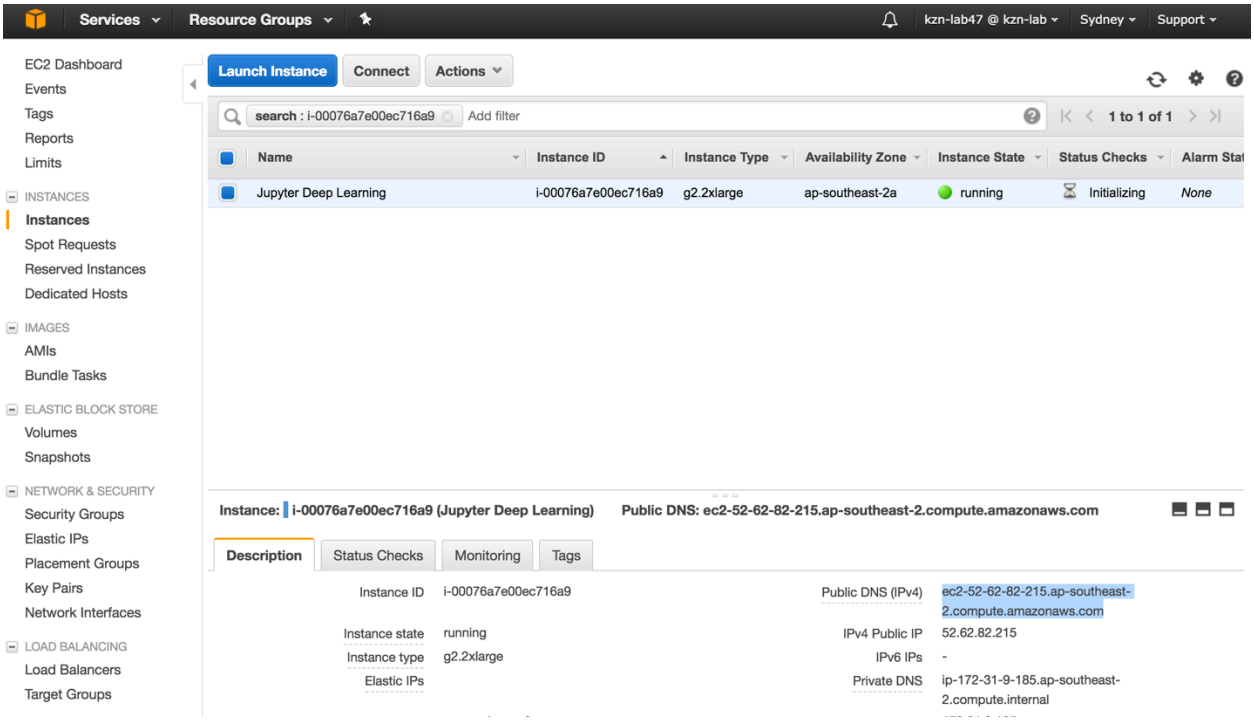
This is also how you generally create a new EC2 instance on AWS. There are quite a few customization steps, but you can safely accept many of the defaults.

9.3.1.2 Using your Notebook environment

Jupyter provides a web interface for us to interact with. We'll use this and become familiar with some of the functions it provides in this section.

To do this securely, we'll setup an SSH tunnel to our EC2 instance running Jupyter, and use that tunnel to connect to our remote server. This means all traffic will be encrypted and we won't be exposing any unsecured public ports on the internet. This makes our connection to the Jupyter instance considerably more secure.

But first, we need to know what the public DNS name for our new Jupyter instance is. Browse back to the EC2 console, and select your new instance. The instance details pane at the bottom of the screen shows you the Public DNS name for the instance.



Copy this DNS name into your clipboard. We'll refer to it below.

If you're using **OS X** or **Linux**, open a terminal window and type the following:

```
ssh -i private_key.pem -L 8888:localhost:8888 ubuntu@publicdns_name
```

where **publicdns_name** is the DNS name for your EC2 instance, and **private_key.pem** is the path to your private key you downloaded earlier.[50]

If you're using **Windows**, you can use the PuTTY program to set up the SSH tunnel: follow Setting up an SSH tunnel with PuTTY.

If you were able to open a tunnel, and SSH to your EC2 instance, you should now see an Ubuntu Linux terminal. To start Jupyter on our EC2 instance, type the following at a command prompt:

```
jupyter notebook --no-browser
```

You should see something like the following output:

```
$ jupyter notebook --no-browser
[I 11:18:53.314 NotebookApp] Writing notebook server cookie
secret to /run/user/1000/jupyter/notebook_cookie_secret
[I 11:18:53.468 NotebookApp] Serving notebooks from local
directory: /home/ubuntu
[I 11:18:53.468 NotebookApp] 0 active kernels
[I 11:18:53.468 NotebookApp] The Jupyter Notebook is running at:
http://localhost:8888/?token=f7979ebd4677c66a08591d30719681e6eca
fd0ea86437739
[I 11:18:53.468 NotebookApp] Use Control-C to stop this server
and shut down all kernels (twice to skip confirmation).
[C 11:18:53.468 NotebookApp]
```

Copy/paste this URL into your browser when you connect for the first time, to login with a token:

```
http://localhost:8888/?token=f7979ebd4677c66a08591d30719681e6eca
fd0ea86437739
```

If you now open a browser on your local machine (e.g. your laptop) and browse to:

```
http://localhost:8888?token=293b53862610e9d940370a4d6a07d4e890ba
992313b2346a
```

[50] If your key is rejected, make sure its file permissions are correct (you should have done this when you created your key):

```
chmod 400 private_key.pem
```

you should see the Jupyter Notebook environment, like:

9.3.2 Step 2. A ~~HelloWorld~~ Fractal Jupyter example

At the top right of the Jupyter application we can see a **New** dropdown. Click on that. You should see a number of options to create a new text file, a new terminal and so on. We also have the option of creating two different types of notebooks; Python 2 and a Python 3. These environments and programming languages have been installed for you by default. It's easy to install other languages too. Jupyter supports being extended this way, using different Jupyter kernels. For example, you might want to install R and Julia kernels to run notebooks using those languages.

The other main navigational element we want to learn about is the tabs across the top of the Jupyter application; the **Files**, **Running**, and **IPython Clusters** tabs.

The **Files** tab is the default view, and shows us the files and folders within our Jupyter environment. Think of this as your Jupyter home directory. As we create new notebooks and modify their configurations, new files will appear here.

The **Running** tab shows you the Terminals and Notebooks you are currently running. You can have multiple terminal sessions and notebooks running at the same time.

The **IPython Clusters** tab allows you to configure your notebook environment to do parallel computing using Python. IPython Clusters is an example of a Jupyter Notebook extension.

Notebooks have been created and made available on Github.com at https://github.com/scicolabs/data-science-ml. We are going to use this a baseline to start exploring running code interactively in our new Notebook environment.

To do this we'll open a new Terminal session, and use git to clone the examples above to our Jupyter environment.

1. Click on the **New** dropdown menu at the top right of your Jupyter environment
2. Select **Terminal**
3. This will open a Linux Bash shell in your web browser. From here you can run command line tools to do whatever you'd like within your Jupyter environment.
4. We've already installed the git tools for you, so let's clone the GitHub repository. Run the following command line in the Terminal:

```
git clone https://github.com/scicolabs/data-science-ml.git
```

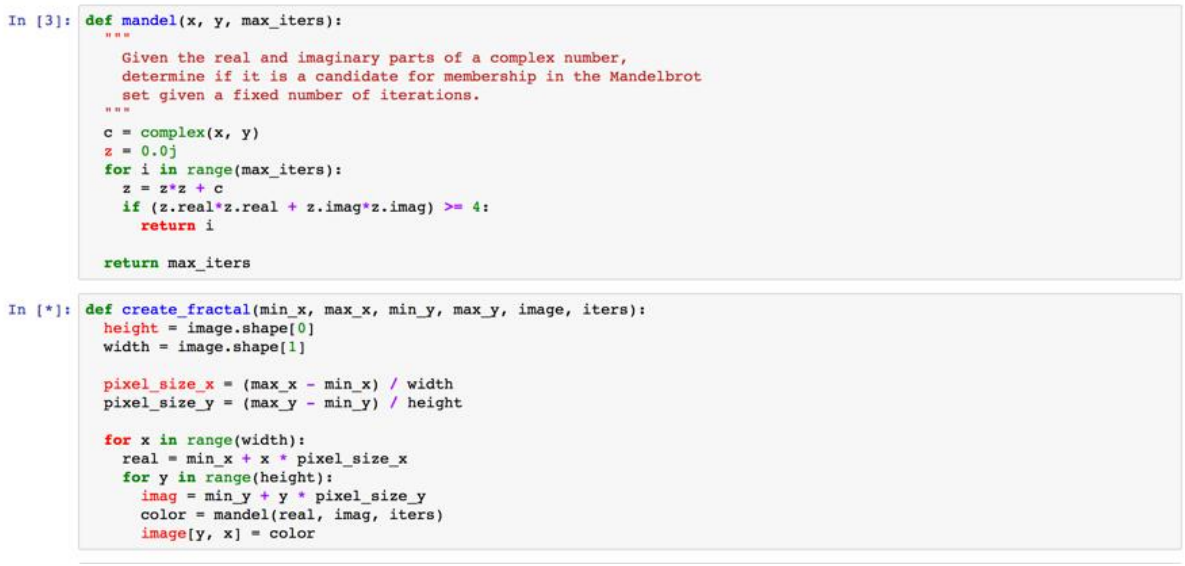
You will see some status information and something looking like:

This command has cloned the data-science-ml Github repository and made the contents available in our Jupyter environment. Browse back to the Jupyter Files tab. You should see a new folder called **data-science-ml**. Click on that folder, and then click on the **labs** folder within it. The **labs** folder contains a number of notebooks. For example, let's interactively run the 'Building fractals with x86.ipynb' notebook. Clicking on that file will load the notebook. You should see something like:

Scroll down and have a look at the notebook. We'll dissect this into pieces later. First, let's run the entire notebook. To do this, use the new menu across the top; **File, Edit, View, Insert, Cell, Kernel** and **Help**.

1. Click on **Cell**
2. In the dropdown, click on **Run All**

You'll notice that some cells have a number in square brackets next to them. These are the cells that have completed execution. Some cells will have an asterisk in square brackets. These cells are still running.



```
In [3]: def mandel(x, y, max_iters):
          """
            Given the real and imaginary parts of a complex number,
            determine if it is a candidate for membership in the Mandelbrot
            set given a fixed number of iterations.
          """
          c = complex(x, y)
          z = 0.0j
          for i in range(max_iters):
            z = z*z + c
            if (z.real*z.real + z.imag*z.imag) >= 4:
              return i

          return max_iters

In [*]: def create_fractal(min_x, max_x, min_y, max_y, image, iters):
          height = image.shape[0]
          width = image.shape[1]

          pixel_size_x = (max_x - min_x) / width
          pixel_size_y = (max_y - min_y) / height

          for x in range(width):
            real = min_x + x * pixel_size_x
            for y in range(height):
              imag = min_y + y * pixel_size_y
              color = mandel(real, imag, iters)
              image[y, x] = color
```

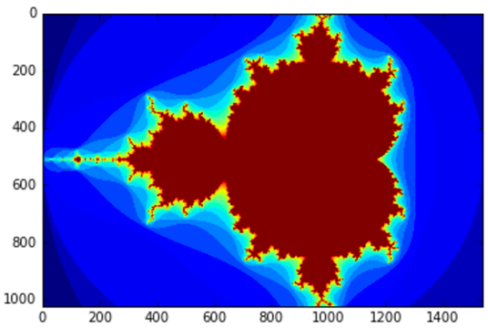
Here you can see that the 3rd cell to execute has finished. This cell defined a function `mandel` in Python. The cell below it which defines the function `create_fractal` is still running.

The result of this simple little Python notebook is a Mandelbrot fractal that is displayed back to you within the notebook.

```
In [5]: image = np.zeros((1024, 1536), dtype = np.uint8)
        start = timer()
        create_fractal(-2.0, 1.0, -1.0, 1.0, image, 20)
        dt = timer() - start

        print "Mandelbrot created in %f s" % dt
        imshow(image)
        show()
```

```
Mandelbrot created in 8.931855 s
```




```
In [ ]:
```

Jupyter is quite powerful like this. You can return results; plotting data, drawing images or displaying raw datasets for example directly within the notebook you're running. You can also use HTML5 elements like `canvas` to capture input and pass that to code you run in the notebook. We'll see an example of this when we explore a handwriting recognition deep learning model. All of these tools make interactively building and running your code much easier, as you can see the result of your code almost immediately.

Now that we've done something useful in our Jupyter environment, let's go back to the original Jupyter tab in our browser and click on the **Running** tab at the top of the page. Here you'll notice that we have a running Terminal session and an active Notebook as well (the Python notebook we were interactively drawing fractals in).

You can control the lifecycle of both of these by clicking the **Shutdown** button.

Saving your Notebook. One important thing you'll want to do often is to save your notebook. Use the **File** menu and click **Save and Checkpoint**. As you save and checkpoint regularly, you'll be able to revert back to previous versions of your notebook

by using the **File** menu and clicking **Revert to Checkpoint**. You'll be presented with a list of all the checkpoints for the current notebook.

Sharing your Notebook. Jupyter notebooks make it simple to collaborate with other researchers. You might work hard on building a new data mining program and want to share that with a collaborator. One way to do this would be to give them access to your Jupyter environment. A better, and perhaps more secure way would be to download your notebook and share it with them via Github.com or even email. They could then run a copy of your notebook in their Jupyter environment, and you could continue working in yours uninterrupted.

To download the current notebook, use the **File** menu and select **Download as**. You can probably see a number of options including; Notebook (.ipynb), Python (.py), HTML (.html) and so on.

For a collaborator to use your notebook easily, you'll want to select the first option, Notebook (.ipynb). Selecting this option will prompt you to save a new file to your local computer. This file will be a self-contained copy of the current notebook that you can then send to whoever you wish.

If you have finished the introduction above, feel free to experiment with the other notebooks in the labs folder. You'll notice some require you to fix missing dependencies. Feel free to start playing around with your new Jupyter notebook environment on AWS!

9.3.3 Step 3 – Machine Learning and AI with AWS

9.3.3.1 Apache MXNet

MXNet is a deep learning library focusing on flexibility, portability and performance. It provides both imperative and symbolic programming idioms, supports multiple languages like C, C++, Python, R, Julia, Scala etc., and can be run on embedded hardware through to desktop and server-class hardware. It can also take advantage of CPUs and GPUs and scales well for distributed training on AWS. The scalability that can be achieved during the training phase with MXNet is part of the reason the AWS Deep Learning Cluster makes MXNet available by default.

In this tutorial, we're going to use some pre-built MXNet Jupyter Notebooks, and explore different networks, models, and training and evaluation approaches.

Enough talking, let's get started!

9.3.3.2 Your first MXNet Notebook

To get started, we'll clone a github.com repository containing a number of interesting MXNet notebooks, and step through some of them. These notebooks are maintained as examples for you to learn about MXNet - https://github.com/dmlc/mxnet-notebooks.

Let's clone the repository into our running Jupyter environment:

```
git clone https://github.com/dmlc/mxnet-notebooks.git
```

This will create a new folder **mxnet-notebooks** on your Jupyter instance. Now we explore some of the notebooks.

1. Click on the **mxnet-notebooks** folder
2. Click on the **python** folder
3. Open **outline.ipynb**

This loads a new Python Notebook in a new tab. You'll note that it's a simple index to other notebooks and documentation on MXNet. This notebook will be the launch pad for exploring the mxnet-notebooks repository we've cloned into our Jupyter environment. Leave it open in a tab. We'll refer to it throughout the rest of the tutorial.

Feel free to browse around the other notebook examples. There are many types of networks presented and explained thoroughly in this repository. You're free to use them and modify them as you wish.

9.3.3.3 MNIST and Handwritten Digit Recognition

MNIST is a classic computer vision application: identify hand written digits with neural networks.

Click on the **MNIST** link in the outline notebook. This will open a step-by-step guide for creating and training a simple multi-layer perception to recognize hand written digits, as well as a more sophisticated convolutional neural network to do better recognition.

This example shows a nice feature of Jupyter notebooks, that you can capture input within the notebook and use that in your code. In this case, we allow the user to draw a digit with the HTML5 canvas element, and then use that as an input to the model to do recognition on.

9.3.3.4 Character based language model using LSTM

Recurrent neural networks (RNN) are a powerful and fairly special type of network that allow us to operate over sequences of input and potentially generate sequences of outputs. This means RNNs can be used to recognize patterns in sequences of data, e.g. text, genomes, spoken word etc.

Long short-term memory (LSTM) networks are a type of RNN and have been used with great success for tasks like machine translation or language modeling. Actually, RNNs more generally effectively describe programs. In fact, RNNs are Turing Complete.

Click on the **Char-LSTM** link in the outline notebook. This opens a guide for building and using an LSTM network in MXNet to generate text based on a set of training data. This is a pretty fun example - most people who run though this example try to generate speeches for the U.S. Presidents or famous authors.

9.3.3.5 Wrapping up

Congratulations! In this tutorial, you created and used your first Jupyter Notebook environment on AWS. You then explored using MXNet and interactively worked with some non-trivial deep learning examples and accelerated the training and evaluation of

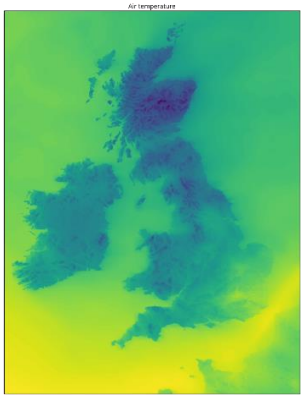
these models on GPU instances in the cloud. This is just the beginning! You can run Jupyter notebooks on any instance types available in EC2. You can use Jupyter to run big data analytics using Amazon EMR (a managed Hadoop platform on AWS), and you can even control HPC clusters from the comfort of your Jupyter Notebook as well.

We'd also love to hear about your use case and what you're looking to do with AWS.

Don't forget to the turn your AWS resources off when you are finished!

9.3.4 Further tutorials

Run Jupyter Notebook and JupyterHub with an **Amazon EMR cluster**: https://aws.amazon.com/blogs/big-data/running-jupyter-notebook-and-jupyterhub-on-amazon-emr/



UK Met Office has made 80TB of MOGREPS dataset with **meteorological (weather) data** available on S3 through the AWS Open Data program. They've published 2 tutorial notebooks showing you how to pull data from the dataset, manipulate it, and visualize it. It uses the Iris python library for much of this. See http://data.informaticslab.co.uk/mogreps_data_basics.html and http://data.informaticslab.co.uk/mogreps_data_intermediate.html

The Awesome Data Science YouTube video tutorial is an excellent series of tutorials about using Jupyter for the basics of Data Science in Python. All the tutorials mentioned in this video tutorial series are available in their GitHub repository as well.

10 Learning more about AWS

Now that you've configured your account safely, set up budgets, and possibly have poked around the AWS Console, you are ready for "real work" —such as launching an Amazon EC2 instance, or uploading and sharing your data with a collaborator. But what if there's more you need to learn?

The Getting Started page[51] is your jump-off point for learning AWS. You can watch tutorial videos, take 10-minute tutorials (or larger projects), and even sign up for online classes or in-person trainings at the AWS Training and Certification site[52].

In addition, as a researcher, you may have access to resources that are not available to the general public. The first is **AWS Educate**[53], a program for educational institutions (teachers *and* students) that includes an environment for curriculum sharing. If your campus is working with an **AWS account manager**, this AWS representative is your lifeline into the AWS organization. They can connect you to AWS specialists to help choose the right AWS Cloud services and use them correctly. Your account manager can also help you locate even more learning and skill-building resources. See the Contact Us form at https://aws.amazon.com/contact-us/ to reach your AWS account manager.

Also, employees and users of AWS have produced a rich library of web-based resources including blog articles, GitHub example code resources, YouTube videos, and more.

10.1 Hands-on and Face-to-Face

AWS has business operations in many countries and (together with partners) offers a variety of hands-on training in classroom formats. Our customer-facing teams also frequently deliver more context-specific hands-on events.

[51] https://aws.amazon.com/getting-started/
[52] https://aws.amazon.com/training/
[53] https://aws.amazon.com/educate/

10.1.1 AWS Immersion Days

AWS's customer teams frequently run "Immersion Days"— AWS workshops delivered by our solutions architect community that allow for hands-on time with AWS Cloud services with expert guidance in the room.

We run Immersion Days on university campuses and often customize them for scientists with the help of our Research & Technical Computing Team. They're provided at no cost and you'll come away with hands-on skills for spinning up instances, managing data, or even creating HPC clusters. It's a great chance to have some interactive time with key AWS technical staff and get a lot of your questions answered.

If you want to find out whether there is an Immersion Day near you, or to discuss whether you might be able to host one at your site, contact us at https://pages.awscloud.com/Public-Sector-Contact-Us.html.

10.1.2 AWS Technical Essentials

AWS Technical Essentials is a face-to-face class held regularly in multiple locations around the world. It will teach you to recognize terminology and concepts as they relate to the AWS Cloud, and how to navigate the AWS Management Console.

Participants gain hands-on experience with the foundational AWS services, including Amazon EC2, Amazon VPC, Amazon S3, and Amazon EBS.

You'll also come away knowing about the security measures AWS provides and key concepts of AWS Identity and Access Management (IAM).

It also covers AWS database services, including Amazon DynamoDB and Amazon RDS, and you'll get to use AWS management tools, including Auto Scaling, Amazon CloudWatch, Elastic Load Balancing, and AWS Trusted Advisor.

To find an AWS Technical Essentials class near you, visit https://aws.amazon.com/training/course-descriptions/essentials/.

10.1.3 Further Training and Professional Certification

If you're ready to dive deeper and feel you'd like some formal classroom training to develop your skills in AWS, you can check out what classes and courses AWS's Training and Certification program has to offer.

The new AWS Training Portal https://www.aws.training/ gives you access to your training and certification activities, progress, and benefits. It is a central place where you can enroll in AWS Training, register for AWS Certification exams, track your learning progress, and access benefits based on the AWS Certifications you have achieved. This makes it easier for you to build your AWS Cloud skills and advance toward earning AWS Certification.

10.2 Online

AWS has a lot of self-help training materials available online, which means you can learn new skills without waiting for a classroom session to be available.

10.2.1 Introductory Skills Training

If you're new to AWS and looking to gain foundational knowledge about key AWS Cloud services, our "Introduction to AWS" series includes instructional videos and labs. You'll find it here: https://aws.amazon.com/training/intro_series/.

10.2.2 Self-Paced Labs

AWS self-paced labs (https://aws.amazon.com/training/self-paced-labs/) provide hands-on practice in a live AWS environment with AWS services and real-world cloud scenarios. Follow step-by-step instructions to learn a service, practice a use case, or prepare for AWS Certification.

10.2.3 YouTube

The AWS channel on YouTube (https://www.youtube.com/user/AmazonWebServices) contains a lot of content like short tutorials or informative talks by Amazonians that are all aimed at helping you learn something new. We also post the content from our global AWS Summits.

10.2.4 AWS website

The AWS website (aws.amazon.com) contains lots of documentation for each of our services. If you have specific questions or problems to solve, searching our website is often the best way to find a solution.

10.2.5 Community websites

It's likely that others in your community have already used AWS to do similar work – see about a hundred examples in 10.4 below. User forums or wiki sites for specific applications that you use may contain very specific "recipes" for deployment on AWS. It's always a good idea to check the date on any third-party guidelines, as technology changes fast: we have likely released much faster EC2 instance types or more advanced data analytics services that will serve you better than those used in a five-year-old research paper. Similarly, performance benchmarks that are several years old may not tell you much about performance today.

10.3 AWS Global Summit Series

Whether you are new to the cloud or an experienced user, you will learn something new at an AWS Summit. These events, held around the world, are designed to educate new users about the AWS Cloud and offer existing users deep technical content to be more successful with AWS. In 2016, we held summits in 38 different locations, all of which provided customers with invaluable direct access to AWS technologists. You can check the schedule here: https://aws.amazon.com/summits/.

10.4 Publications on research done in the AWS cloud

Please **send us your research and papers** at aws-research-cloud@amazon.com !

10.4.1 Astronomy

Astronomy Case Studies:

- AWS Case Study: NASA/JPL's Desert Research and Training Studies
- NASA/JPL's Mars Curiosity Mission Case Study
- AWS Case Study: International Centre for Radio Astronomy Research (ICRAR)
- Case Study: NASA JPL and Amazon SWF

Astronomy Articles:

- NVIDIA - Exploring the SpaceNet Dataset Using DIGITS

Astronomy Blog Posts:

- Looking Deep into our Universe with AWS
- A Minimalistic Way to Tackle Big Data Produced by Earth Observation Satellites
- Q&A with DigitalGlobe and HOT

Astronomy Videos:

- How NASA's Mars Rover and Earth Analytics Use the Cloud
- To Mars and Back: How NASA Uses the AWS Cloud

Astronomy SlideShare:

- Bringing Governance to an Existing Cloud at NASA's Jet Propulsion Laboratory

10.4.2 Big Data

Big Data Case Studies:

- AWS Big Data Customer Testimonials
- AWS Case Study: National Taiwan University

Big Data Blog Posts:

- A Minimalistic Way to Tackle Big Data Produced by Earth Observation Satellites

Big Data Videos:

- AWS re:Invent 2016: Apache Spark on EC2 History, Best Practices with Customer Use Cases
- AWS re:Invent 2016: How to Build a Big Data Analytics Data Lake
- AWS re:Invent 2016: Case Study: Data-Heavy Healthcare: UPMCe's Approach to Healthcare

10.4.3 Data Sharing and Transfer

Data Sharing and Transfer Case Studies:

- Stanford Archaeology Center
- Harris Geospatial Solutions - Geospatial Analytics in the Cloud
- International Rice Research Institute Case Study

Data Sharing and Transfer Articles:

- Open Science Grid - ATLAS and BNL Bring Amazon EC2 Online

Data Sharing and Transfer Blog Posts:

- Call for Computer Vision Research Proposals with New Amazon Bin Image Dataset
- Announcing Terrain Tiles on AWS: A Q&A with Mapzen
- Zooniverse's Open Source Answer to Disaster Relief

Data Sharing and Transfer Videos:

- Research Collaboration Panel: Science Without Boundaries - Global Solutions to Scientific Problems
- AWS re:Invent 2016: Earth on AWS - Next-Generation Open Data Platforms
- Oregon State University Hatfield Marine Science Center Uses AWS Snowball Edge
- AWS Public Sector Summit 2016 - Smart Cities: Open Grid

Data Sharing and Transfer Webinars:

- Harris Geospatial Solutions - Geospatial Analytics in the Cloud with ENVI and Amazon Web Services

10.4.4 Environmental Science

Environmental Science Case Studies:

- AWS Case Study: National Renewable Energy Laboratory's OpenEI.org

Environmental Science Blog Posts:

- A Minimalistic Way to Tackle Big Data Produced by Earth Observation Satellites
- Solving Problems with Open Data Imagery: Q&A with DigitalGlobe and HOT
- NREL and AWS Bring Energy Data to Analysts and Researchers
- Open Earth Observation Data for a Changing Planet
- Q&A with Planet OS: Learn about the OpenNEX Climate Data Access Tool
- Exploring the Possibilities of Earth Observation Data

Environmental Science Videos:

- Oregon State University Hatfield Marine Science Center Uses AWS Snowball Edge
- AWS Agriculture Analysis in the Cloud Day at Ohio State University

Environmental Science SlideShare:

- Observation in the Cloud SlideShare

10.4.5 General Research

General Research Blog Posts:

- What Data Egress Means for Higher Education: A Q&A with Internet2
- AWS Offers Data Egress Discount to Researchers
- Time to Science, Time to Results: Transforming Research in the Cloud

General Research SlideShare:

- Accelerating Time to Science: Transforming Research in the Cloud

10.4.6 High Performance Computing (HPC)

HPC Case Studies:

- AWS Case Study: NASA/JPL's Desert Research and Training Studies
- AWS Case Study: National Taiwan University
- New York University Langone Medical Center Case Study

HPC Blog Posts:

- Alces Flight: Build you Self-Service Supercomputer in Minutes
- The Evolution of High Performance Computing: Architectures and the Cloud
- High Performance Cloud Computing Supports Disease Prevention

HPC Videos:

- AWS re:Invent 2016: High Performance Computing on AWS
- AWS re:Invent 2016: Building HPC Clusters as Code in the (Almost) Infinite Cloud
- AWS Public Sector Summit 2016: Building HPC Clusters as Code in the (Almost) Infinite Cloud
- University of Heidelberg/HITS: Innovation for Education and Research in the Cloud
- Launching an Alces Flight Compute Cluster from AWS Marketplace

10.4.7 Internet of Things (IoT)

IoT Case Studies:

- AWS IoT Customer Testimonials

IoT Blog Posts:

- University of Muenster Creates openSenseMap to Engage with Citizens and Students

IoT Videos:

- AWS Agriculture Analysis in the Cloud Day at Ohio State University

10.4.8 Life Sciences

Life Sciences Case Studies:

- AWS Life Sciences Customer Testimonials
- AWS Case Study: Penn State
- AWS Case Study: University of California Berkeley AMP Lab's Carat Project
- University of California Berkeley AMP Lab's Genomics Research Project Case Study
- Baylor College of Medicine Case Study
- Caltech Guttman Lab Case Study
- Harvard Medical School Case Study
- Icahn School of Medicine at Mount Sinai Case Study
- Murdoch University Case Study

- UC Santa Cruz Genomics Institute Case Study
- New York University Langone Medical Center Case Study
- University of Chicago Case Study
- San Francisco State University Case Study
- American Heart Association Builds Precision Medicine Platform on AWS
- Benchling Case Study

Life Sciences Blog Posts:

- MalariaSpot: Diagnose Diseases with Video Games
- Calling All Data Scientists to Help Improve Cancer Screening Technology
- Exatype: Cloud for HIV Drug Resistance Testing
- Cloud-Enabled Innovation in Personalized Medical Treatment
- Building Bridges for Better Cancer Treatment with the Fred Hutchinson Cancer Research Center
- How the Healthcare of Tomorrow is Being Delivered Today
- New AWS Public Dataset - 3000 Rice Genome
- New AWS Public Datasets - TCGA and ICGC
- Human Longevity, Inc. - Changing Medicine Through Genomics Research
- Frequently Asked Questions About HIPAA Compliance in the AWS Cloud
- Genome Engineering Applications: Early Adopters of the Cloud
- Vodafone DreamLab - Accelerating Cancer Research
- An Eye on Science: How Stanford Students Turned Classwork into Their Life's Work
- High Performance Cloud Computing Supports Disease Prevention
- Genome Engineering Applications: Early Adopters of the Cloud
- Analyzing Genomics Data at Scale using R, AWS Lambda, and Amazon API Gateway

Life Sciences Videos:

- AWS re:Invent 2016: Large-Scale, Cloud-Based Analysis of Cancer Genomes
- Genomics at Scale: Using the AWS Cloud for Population-Scale Analysis of Genomics and Life Science
- AWS re:Invent 2015: Large-Scale Genomics Analysis with Amazon Redshift
- AWS Symposium - Washington DC: Genomic Data Privacy and Security in Human Research and the NIH
- Analyzing Genomic Data for Whole Populations: How AWS Enables Analysis of Large Cohorts
- AWS Case Study: US Centers for Disease Control and Prevention (CDC)
- Claritas Genomics Delivers Services 75% Faster by running on AWS
- This Is My Architecture: Analyzing Genomic Data at Scale on AWS with Station X's GenePool
- This Is My Architecture: UC Santa Cruz: Using Mesos & Amazon EC2 Spot to Enable Low-Cost Cancer Research

10.4.9 Machine Learning

Machine Learning Case Studies:

- San Francisco State University Case Study
- Black Dog Institute and CSIRO Case Study
- University of California Berkeley AMP Lab's Genomics Research Project Case Study

Machine Learning Articles:

- GeekWire - Amazon Hires Carnegie Mellon Machine Learning Expert as Google Expands its own AI Initiatives
- GeekWire - Amazon Beefs Up Machine Learning Presence in UK with New Team of Researchers
- ZDNet - Should Amazon be Your AI and Machine Learning Platform?
- What do Birds and Social Media Have in Common? AWS Machine Learning Support

Machine Learning Blog Posts:

- An Eye on Science: How Stanford Students Turned Classwork into Their Life's Work
- Call for Computer Vision Research Proposals with New Amazon Bin Image Dataset
- AWS Enables Consortium Science to Accelerate Discovery

Machine Learning Videos:

- Amazon Core Machine Learning Team YouTube Video

10.4.10 Physics

Physics Articles:

- Inside Big Data - Interview: Dr. Michael Ernst from Brookhaven National Laboratory
- Information Week - Brookhaven Lab Finds AWS Apot Instances Hit Sweet Spot
- Information Week - AWS Helps FermiLab Researchers 'See' Neutrino Particles
- U.S. Department of Energy - Office of Science - ASCR Discovery - HEPCloud Formation

Physics Blog Posts:

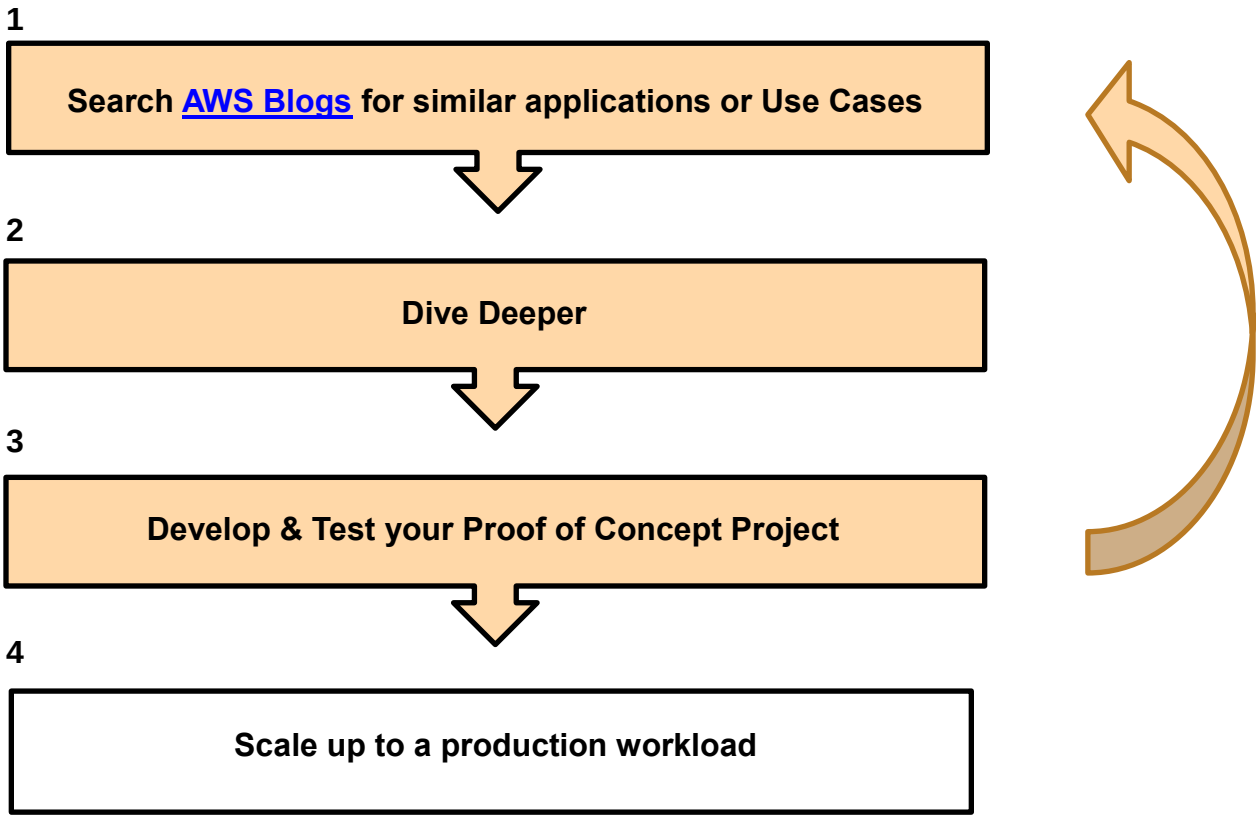
- Experiment that Discovered the Higgs Boson Uses AWS to Probe Nature
- NOvA Uses AWS to Shed Light on Neutrino Mysteries

11 Finding and Building Solutions

Now that you're safely and securely established in the AWS Cloud, it's time to think about what you can do with the facilities at your disposal.

11.1 Build it yourself

With all of the services and features available in the AWS Cloud, starting out doesn't need to be an overwhelming process. Here is a suggested flow chart to help you quickly get up to so you can build the solution needed for your research.



Step 1: Once you have identified a couple of blog articles with a similar use case …

Step 2: you can dive deeper into each of the services discussed in order to quickly get up to speed. There are AWS monthly webinars available as well as online YouTube videos. In addition, there are online self-paced labs and a rich library of written content for each service available through the AWS web portal. Content detailing all of our services includes white papers, user guides, and developer forums. In addition, there are also a GitHub repository, Gist resources, and AWS Gitter Channels.

Step 3: Develop a small Proof of Concept (POC) test case for your AWS solution. For example, genomics customers building a pipeline for sequence analysis may downsample their read dataset or only take reads from a single chromosome for testing a development POC pipeline. It is important to design for scale-up and implement the

features you will need at production scale (i.e. use a shared file system that will accommodate your full production workload, architect using load balancing, select the right database or warehouse for the dataset that you will be using at scale even if you are testing with a much smaller file version, etc.)

Step 4: Scale up to a production workload and check that your solution is working as expected.

11.2 The Cloud Credits for Research program

The AWS Cloud Credits for Research Program provides AWS promotional credits to support academic researchers who seek to:

- Build cloud-hosted publicly available science-as-a-service applications, software, or tools to facilitate their future research and the research of their community.
- "Kick the tires" of the cloud to find how their workload will scale in the cloud.
- Train a broader community on the usage of cloud for research workloads via workshops or tutorials.

This program isn't aimed at providing support for operations, ongoing or established research projects, or research projects that are limited in their potential scope for future AWS usage. It's not a substitute for your research funding body, but designed to help lower the risk for you and your community to adopt the cloud.

You may apply to the program and consult FAQ at https://aws.amazon.com/research-credits/.

Applications are reviewed quarterly and decision are typically communicated 2-3 months following the respective quarterly deadline. Awarded amounts will vary depending on the research proposal and usage requirements documented in the proposal, and will be in the form of promotional credits applicable to AWS services.

11.3 Third Party Solutions

You may be able to save a lot of time by relying on a third party for building the AWS infrastructure you need. The AWS Partner Network (APN) is the global partner program for AWS. It has thousands of members who are companies from all over the globe that provide services and technology platforms built on AWS.

We break down the APN into two groups: Technology Partners, and Consulting partners.

APN **Technology Partners** provide software solutions that are either hosted on, or integrated with, the AWS Cloud. APN Technology Partners include Independent Software Vendors (ISVs), SaaS, PaaS, Developer Tools, Management and Security Vendors.

Some of our technology partners make their offerings available in AWS Marketplace, which is like the application store for the cloud. The advantage of doing so is that the solutions you find there are sophisticated, yet simple to start, built to best practices and any charges incurred are direct to your AWS account, which saves establishing vendor relationships with third parties. AWS Marketplace vendors frequently offer limited free trials so you can figure out whether their offering is right for you without incurring any software charges.

APN **Consulting Partners** are professional services firms that help users design, architect, build, migrate, and manage their workloads and applications on AWS. Consulting partners can help you if you have an unusual need or you just need hands-on help getting started.

The remainder of this Handbook is dedicated to a presentation of the best research oriented APN Technology and Consulting Partners. We hope you'll find inspiration while browsing the following chapters.

12 APN Technology Partners Listing

On the following pages, we introduce some of our APN Technology Partners, selected for what they offer to the global Research Community.

Most of these partners are offering products that are changing the way the world approaches scientific computing.

We're delighted that these partners choose AWS as their platform. We continually work hard to understand customer needs and make sure that ideas for new features and services are fed back into our development cycle, so that the cloud becomes an ever-improving experience for everyone.

Contents

aewacs B.V.	Product or Company **aewacs**	Home Page http://www.aewacs.com/
	Vendor Country of Origin **Netherlands**	Delivery method ***SaaS Portal***
Domains	Chemistry, Physics, Astronomy, Mathematics, Geography, Geospatial, Graphics, Machine Learning, Bioinformatics, Biochem, Statistics, …. (more information available https://www.aewacs.com).	

Regional availability – All AWS Regions are sovereign.

IRE	DE	UK	FR	CA	US	US	US	US	BR	SG	JP	AU	KR	IN	CN	CN
Dublin	Frankfurt	London	Paris (2017)	Montreal	N. Virginia	N. California	Oregon	Ohio	Sao Paolo	Singapore	Tokyo	Sydney	Seoul	Mumbai	Beijing	Ningxia (2016/17)

12.1 aewacs B.V.

Aewacs is a multitenant portal that automates scheduled activity for your various Amazon Web Services (AWS) cloud resources.
With aewacs, you can easily create on/off time schedules for your systems, backups, auto scaling groups and environments in AWS. Consequently, your cloud resources will be solely turned on when you want. Can you imagine the costs you'll save? As all schedules are automated, you'll also save time and always have a clear overview of your resources' activities.
Aewacs comes in different shapes and sizes: you can choose from 5 plans, which you can both upgrade and downgrade. This way, you'll always use the plan that's right for you. Needless to say, we have a free plan to get you started.

12.1.1 How is it accessed?

SaaS (eg Portal) – user connects to a public portal, registers for an account and connects AWS accounts

12.1.2 User requirements

Our users need minimal AWS skills in order to create a user with some policies and setup the account connection between aewacs and AWS.

We have an extensive manual where we describe in detail how to do this. Also we have instruction videos that describe every set of functionality we provide.

12.1.3 Trial

aewacs offers a free plan, not limited in functionality but limited in how many AWS resources can be controlled. The aewacs free plan can be used indefinite.

12.1.4 Billing

Who are the parties to the billing relationship? Does the researcher need to procure from the partner directly (and if so, in what currency, and in which legal jurisdiction), or is the transaction possible through AWS Marketplace, or via an established reseller network? Is there an upfront component, or pay-per-use? Or do you have an ability to offer both?

Direct relationship with vendor - aewacs offers a monthly subscription with different plans (https://www.aewacs.com/plans). We accept all major credit cards and our users have the option to be billed in € of $.

12.1.5 Pricing & Licenses

Is there a license to buy? If so, what is the pricing model? Can a user bring an enterprise or site-side license from their institution to use in the cloud? Under what conditions?

It's advantageous to show some simple worked examples here that outline what a "typical" (or familiar) workload might entail, and how much it would cost.

FREE	BASE	PROFESSIONAL	ENTERPRISE	ENTERPRISE+
$0 / month	**$29 / month**	**$79 / month**	**$199 / month**	**Contact us**
3 EC2 instances	**20** EC2 instances	**30** EC2 instances	**60** EC2 instances	**60+** EC2 instances
2 Autoscalinggroups	**3** Autoscalinggroups	**8** Autoscalinggroups	**20** Autoscalinggroups	**20+** Autoscalinggroups
1 RDS instance	**3** RDS instances	**8** RDS instances	**20** RDS instances	**20+** RDS instances
1 Environment	**2** Environments	**5** Environments	**10** Environments	**10+** Environments
8x5 email support CET	**8x5** email support CET	**8x5** email support CET	**10x5** email support CET	**24x7** email support
no phone support	**no** phone support	**8x5** phone support CET	**10x5** phone support CET	**24x7** phone support

12.1.6 Support

How does a user get access to your support services?

For support our users can access our support portal via https://www.aewacs.com/support

Provide details on your support offering, subscription rates, and contact methods.

Our support varies per plan, see details at https://www.aewacs.com/plans

12.1.7 Training and Reference Sources

Do you provide training (either in person or online)? How is it accessed? How is it charged?

We provide online training in the form of free instruction videos.

Additional Information from Aewacs

Why aewacs?

aewacs is developed to deliver added value on top of the existing Amazon Web Services console. As such aewacs will provide extra services not (yet) available in the AWS console. Currently aewacs enables you to create schedules for Amazon Web Services EC2 instances, RDS and autoscalinggroups.

These schedules can be used to fully automate:

- on/off times for your EC2 instances;
- auto scaling actions (up / down / suspend);
- EC2 (AMI - volumes) and RDS backups

Using these aewacs features can easily reduce your monthly cloud cost by 25% or more with just 10 minutes' setup.

aewacs comes in different shapes and sizes: you can choose from 5 plans, which you can both upgrade and downgrade. This way, you'll always use the plan that's right for you. Needless to say, we have a free plan to get you started.

Take a look at our website https://www.aewacs.com to learn more.

Technology Partner	Product or Company **AceCloud**	Home Page https://www.acellera.com/acecloud
	Vendor Country of Origin **United Kingdom**	Delivery method **AWS Marketplace and custom client software**
Domains	Computational Chemistry, Medicinal Chemistry, Drug Discovery, Drug Design, Machine Learning.	

Regional availability – All AWS Regions are sovereign.																
IRE	**DE**	**UK**	**FR**	**CA**	**US**	**US**	**US**	**US**	**BR**	**SG**	**JP**	**AU**	**KR**	**IN**	**CN**	**CN**
Dublin	Frankfurt	**London**	**Paris (2017)**	Montreal	N. Virginia	N. California	Oregon	Ohio	Sao Paolo	Singapore	Tokyo	Sydney	Seoul	Mumbai	Beijing	Ningxia (2016/17)

12.2 AceCloud, by Acellera

Acellera's AceCloud enables anyone to perform high-throughput molecular dynamics simulations on AWS in a timely and secure way.

By taking advantage of AWS' high performance GPU instance types and EC2 spot pricing, AceCloud is able to deliver timely access to simulation, making it the ideal platform for production-quality MD-based investigation. Being secure by design and fully auditable in operation, AceCloud is a natural fit for companies wishing to undertake commercially confidential computational work on AWS.

AceCloud's intuitive command-line interface abstracts all interactions with AWS, and seamlessly provides the experience of actually running simulations directly on the user's own computer. Furthermore, the AceCloud client software is integrated into Acellera's ACEMD platform and the HTMD software (High Throughput Molecular Dynamics), allowing customers to immediate take advantage of HTMD's sophisticated protocols for simulation-based drug design.

12.2.1 How is it accessed?

AceCloud is available on the AWS Marketplace at
http://aws.amazon.com/marketplace/pp/B01N3SBK3Z

This AMI provides a standalone instance configured to run ACEMD and other Acellera software.

The intuitive AceCloud client abstracts all interactions with AWS, managing the transfer of data to and from S3 and the creation of EC2 virtual machines, making the execution of work through AceCloud as straightforward as running on a local machine.

12.2.2 User requirements

To access AceCloud, the user need only subscribe to the AceCloud AWS Marketplace product and install the AceCloud client software (supported platforms: Linux, OS X and Windows). The only direct interaction with AWS required is to configure IAM access tokens (full instructions provided). Thereafter, the user runs all their computational work directly from the command line of their workstation. Acellera will provide support to the user in all these operations until all is functional on the user infrastructure.

12.2.3 Trial

AceCloud is included in our ACEMD platform software that requires a license, then you can use it on a “Pay-as-you-go” product with no minimum fees.

12.2.4 Billing

Billing for AceCloud computation is directly via the AWS Marketplace.

12.2.5 Pricing & Licenses

AceCloud benefits from “Pay-as-you-go” pricing, with hourly pricing as low as 40¢/hr including EC2 fees. No additional software licenses are required than the license for ACEMD platform (End User License Agreement terms apply).

AceCloud minimizes costs by automatically uses spot Instance pricing, intelligently placing jobs in the lowest cost AWS Regions. With AceCloud, costs are only incurred for the duration of running computations.

12.2.6 Support

Marketplace customers may get support for AceCloud issues by opening an issue on Acellera’s HTMD Issue Tracker at https://www.acellera.com/issues . Technical support is included in the license fee covering installation and update.

12.2.7 Training and Reference Sources

Full instructions for using AceCloud are available online. Further training and consultancy services are available from Acellera for an additional fee.

Users may also wish to attend Acellera’s annual high throughput Molecular Dynamics Workshop in Barcelona, to receive training on modern ensemble-based molecular simulation methods and their execution on AceCloud

This work has been published in several articles whose references are the following:

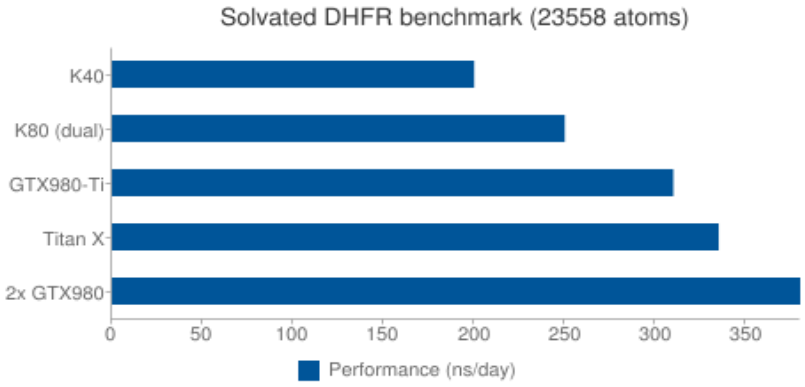
M. J. Harvey and G. De Fabritiis. AceCloud: Molecular Dynamics Simulations in the Cloud. J. Chem. Inf. Model. (2015) 55 (5), pp 909–914.

S. Doerr , M.J. Harvey, F. Noé ,G. De Fabritiis, HTMD: High-throughput molecular dynamics for molecular discovery, J. Chem. Theory Comput., 2016, 12 (4), pp 1845–1852

Additional information from AceCloud by Acellera

How It Works

AceCloud provides the perfect way to run hundred molecular dynamics simulations only when required, supporting Amber and Charmm force-fields and metadynamics with PLUMED. By exploiting GPU-based computation, you are able to achieve in excess of 200ns/day on the DHFR benchmark when running on EC2.



AceCloud command line example

Step 1: Prepare an ACEMD simulation input

```
$ ls .
     input s     structure.pdb     structure.psf     parameters.inp
```

Step 2: Submit the input directory of files to AceCloud, giving the job a meaningful identifier

```
$ acecloud --submit . examplerun/3
```

AceCloud copies the files to AWS and instantiates a new EC2 virtual machine that begins to run the simulation

Step 3: Monitor the progress of jobs

```
$ acecloud --status --group examplerun
Group       Name        Status
examplerun  1           RUNNING
examplerun  2           COMPLETED
examplerun  3           PENDING
```

Step 4: Retrieve the output of jobs

```
$ acecloud --retrieve examplerun/*
$ ls .
     input       structure.pdb     structure.psf     parameters.inp
     log.txt     trajectory.xtc
```

AceCloud copies back the output of any running or completed jobs, putting the files in the directories from which the jobs were originally submitted.

Technology Partner	Product or Company **Alces Flight Ltd.**	Home Page http://www.alces-flight.com
	Vendor Country of Origin **United Kingdom**	Delivery method **AWS Marketplace or HPCaaS**
Domains	Benchmarks, Biochemistry, Bioinformatics, Bio-physics, Chemistry, Compilers, Databases, Electronics, Engineering, Geography, Graphics and Imaging, Languages, Libraries, Mathematics, Medicine, MPIs, Physics, Statistics, Tools, Visualization (more information available at http://docs.alces-flight.com)	

Regional availability – All AWS Regions are sovereign.																
IRE	**DE**	**UK**	**FR**	**CA**	**US**	**US**	**US**	**US**	**BR**	**SG**	**JP**	**AU**	**KR**	**IN**	**CN**	**CN**
Dublin	Frankfurt	London	**Paris (2017)**	Montreal	N. Virginia	N. California	Oregon	Ohio	Sao Paolo	Singapore	Tokyo	Sydney	Seoul	Mumbai	Beijing	Ningxia (2016/17)

12.3 Alces Flight

Alces Flight Compute provides a fully-featured, scalable High Performance Computing (HPC) environment for research and scientific computing. Compatible with both on-demand and spot instances, Flight rapidly delivers a whole HPC cluster, ready to go and complete with job scheduler and applications.

Clusters are deployed in a Virtual Private Cluster (VPC) environment for security, with SSH and graphical-desktop connectivity for users. Data management tools for POSIX and S3 object storage are also included to help users transfer files and manage storage resources.

12.3.1 How is it accessed?

Alces Flight Compute is available in a Solo user experience in the AWS Marketplace, with multi-user and companion appliances available by contacting your AWS Account Manager or Alces Flight directly.

Users are provided with an opportunity to configure the number and size of compute nodes, and to specify the characteristics of the cluster login node and shared filesystem at launch time. Once launched, your personal Linux cluster is available to connect to via SSH or via collaborative remote desktop sharing.

Each user has access to install any applications, libraries, MPIs and compilers they need to run their jobs onto the cluster via the centrally managed Alces *Gridware* software repository. Applications are installed with *module* files for quick access and automatic dependency resolution. View the available applications at http://gridware.alces-flight.com.

12.3.2 User requirements

Alces Flight Compute has been purposely designed to give you the same user experience that comes with traditional HPC clusters. Those with some experience in HPC should be able to get started quickly, even if they haven't used AWS before.

12.3.3 Trial/Free Version

Alces Flight Solo has a freely available version for AWS users called the "Community Edition". This single-user version comes with Community support as well as full access to documentation in order to get you working right away. There is also an *HPC-as-a-service* option available for users who need a one-click cluster launch – contact this partner for details and a free trial.

12.3.4 Billing

Users selecting Alces Flight, Solo Community Edition, do not pay any software charges – you are responsible for the AWS charges incurred while running clusters, which are billed to your account in the normal way. For customers selecting the Solo Professional Edition, billing for software is performed via AWS Marketplace in US dollars. Billing is charged at a pro-rata rate – e.g. subscribing for 1-day costs 1/30th of the monthly fee.

Alces Flight Enterprise customers are typically larger sites or shared user groups, who often have specialized configuration requirements. Please contact the Alces team directly or your AWS Account Manager to discuss a custom HPC cluster deployment.

12.3.5 Licensing

Alces Flight is made available for use under an End-User License Agreement (EULA) that is provided when users subscribe to the product on AWS Marketplace. The product contains both proprietary and open-source software, for which users are granted a license to use for research purposes – please see the EULA text for full details.

12.3.6 Support

Alces Flight Compute includes access to support services based on your subscription level. For customers selecting Solo Community Edition you can access our public support portal through https://community.alces-flight.com/. Customers selecting Solo Professional and Enterprise Editions have a guaranteed SLA for queries, and engage the Alces Flight Crew directly via a private area on the support site. Details of how to access commercial support services are provided when a paid subscription commences.

12.3.7 Training and Reference Sources

The Alces Flight Team has put together comprehensive documentation at http://docs.alces-flight.com. We also have:

- Training videos on YouTube: http://tiny.cc/alcesflightvids
- Solution whitepapers: http://alces-flight.com (under the *Launch* section)
- Searchable list of available *Gridware* : https://gridware.alces-flight.com/

Additional Information from Alces Flight

What is Alces Flight Solo?

Alces Flight Solo is a software appliance designed to help researchers and scientists build their own high-performance compute cluster quickly and easily via AWS. The basic structure provided for users is as follows:

- One login node, plus a configurable number of compute nodes
- An Enterprise Linux operating system
- A shared filesystem, mounted across all nodes
- A batch job scheduler
- Access to a library of software applications and tools

Flight provides a preconfigured environment that is ready for work. The cluster you build is personal to you - users have root-access to the environment, and can setup and configure the system to their needs.

Your cluster is designed to be **ephemeral** - i.e. you run it for as long as you need it, then shut it down. Although there is no built-in time limit for Flight clusters, the most effective way of sharing compute resources in the cloud is to book them out only when you need them. Contrary to popular belief, you can achieve huge cost savings over purchasing server hardware if learn to work effectively in this way.

Who is it for?

Alces Flight Solo is designed for use by end-users - that's the scientists, researchers, engineers and software developers who actively run compute workloads and process data. Flight provides tools that enable self-service - it's very configurable, and can be expanded by individual users to deliver a scalable platform for computational workloads.

Prerequisites

To get started you need two things:

- A client device (e.g. a desktop or laptop computer) with an SSH client
- Access to cloud resources (AWS Account)

Check out our full prerequisites list here: http://docs.alces-flight.com/en/stable/overview/whatisit.html#prerequisites

Where can I get help?

The online documentation (http://docs.alces-flight.com) is designed to walk users through the first stages of creating their clusters and getting started in the environment. Capable users with some experience can be up and running in a handful of minutes - don't panic if it takes you a little more time, especially if you've not used Linux or HPC clusters before.

We encourage new users to run through a few tutorials in the online documentation - even if you have plenty of HPC experience. It can also help to work collaboratively with other researchers running similar jobs as often two sets of eyes are better than one.

There is a community site for supporting the Flight software at http://community.alces-flight.com. This website is designed to help users share their experiences of running Flight clusters, report any bugs with the software, and share knowledge to help everyone work more effectively. There is no payment for using this service except for the general requirement to be nice to each other.

Creating your Cluster

You can launch a cluster in the AWS Marketplace by following these steps:

1. Create (if you haven't already) and sign-in to your AWS account and navigate to the AWS Marketplace.
2. Search for **Alces Flight** in the search box provided to find the Alces Flight Solo Community Edition (or Solo Professional if you need a guaranteed response time for support queries) and click to select it
3. Read the product information and click on the **Continue** button to view launch details. **After cl**icking the **Continue** button from the main produ**ct page selec**t the **Custom Launch** tab in your browser.
4. Scroll down the page and select your local AWS region in the **Select a Region** section.
5. Choose **Personal HPC compute cluster** from the *Deployment Options* section
6. Under the *Launch* section click on the **Launch with CloudFormation Console** button to start deploying your cluster.

When you choose to start a Flight Compute cluster from AWS Marketplace you will be prompted to answer a number of simple questions about what you want the environment to look like. Flight will automatically include your answers into a CloudFormation template launch your desired configuration. See our guide to working with AWS CloudFormation: http://docs.alces-flight.com/en/stable/launch-aws/launching_on_aws.html#how-to-answer-cloudformation-questions

Using your Cluster

Once your cluster is launched users have full access to install applications, run jobs via the cluster job-scheduler, copy data to and from the cluster, and collaborate with other users in graphical desktop sessions. Use the links below to find more information on the facilities available on your personal HPC cluster:

- Using Linux on your cluster
  http://docs.alces-flight.com/en/stable/basics/basic_cluster_operation.html
- Storing file and object data
  http://docs.alces-flight.com/en/stable/databasics/data_basics.html
- Graphical desktop sessions
  http://docs.alces-flight.com/en/stable/graphicaldesktop/graphicaldesktop.html
- Installation software applications
  http://docs.alces-flight.com/en/stable/apps/apps.html
- Submitting work to the cluster job scheduler
  http://docs.alces-flight.com/en/stable/jobschedulers/jobschedulers.html

AWS Deep Learning	Product or Company **Amazon Deep Learning Group**	Home Page http://bit.ly/deepami & http://bit.ly/deepcfn
	Vendor Country of Origin **USA**	Delivery method **AWS Marketplace & GitHub**
Domains	Deep Learning | Applicable to all domains	

Regional availability – All AWS Regions are sovereign.

IRE	DE	UK	FR	CA	US	US	US	US	BR	SG	JP	AU	KR	IN	CN	CN
Dublin	Frankfurt	London	Paris (2017)	Montreal	N. Virginia	N. California	Oregon	Ohio	Sao Paolo	Singapore	Tokyo	Sydney	Seoul	Mumbai	Beijing	Ningxia (2016/17)

12.4 AWS Deep Learning

Machine Learning is playing an increasingly important role in many areas of our work and our lives and is being employed in a range of computing tasks where programming explicit algorithms is infeasible. At Amazon, machine learning has been key to many of our work processes, from recommendations to fraud detection, from inventory levels to book classification to abusive review detection.

And there are many more application areas where we use machine learning extensively: search, autonomous drones, robotics in fulfillment centers, text and speech recognitions, etc. Among machine learning algorithms, a class of algorithms called deep learning has come to represent those algorithms that can absorb huge volumes of data and learn elegant and useful patterns within that data: faces inside photos, the meaning of a text, or the intent of a spoken word. A set of programming models has emerged to help developers define and train AI models with deep learning; along with open source frameworks that put deep learning in the hands of mere mortals. Some examples of popular deep learning frameworks include Caffe, CNTK, MXNet, TensorFlow, Theano, and Torch.

Amazon EC2, with its broad set of instance types and GPUs with large amounts of memory, has become the center of gravity for deep learning training. To that end, we recently made a set of tools available to make it as easy as possible to get started: a Deep Learning AMI, which comes preinstalled with popular open source deep learning frameworks (MXNet, TensorFlow, Theano, Torch, CNTK and Caffe), GPU-acceleration through CUDA drivers which are already installed, preconfigured, and ready to rock, and supporting tools such as Anaconda and Jupyter. Developers can also use the distributed Deep Learning CloudFormation template to spin up a scale-out, elastic cluster of P2 instances using this AMI for even larger training runs.

12.4.1 How is it accessed?

- Deep Learning AMI supporting 6 different preinstalled frameworks and associated tools is available HERE
- CloudFormation Template is available on GitHub HERE or an EC2 Compute Blog post HERE

12.4.2 User requirements

Users should have experience with open source components, data manipulation and transformations and general software engineering expertise. Machine and Deep Learning experience is highly recommended.

12.4.3 Trial

There is no trial period.

12.4.4 Billing

Via AWS Marketplace

12.4.5 Pricing & Licenses

There is no additional charge for any of these tools. You pay for AWS resources (e.g. EC2 instances or EBS volumes) you create to store models and data, and run training. You only pay for what you use, as you use it; there are no minimum fees and no upfront commitments.

12.4.6 Support

Support is available through forums, technical FAQs and the Service Help Dashboard.

12.4.7 Training and Reference Sources

Training is being done on an ongoing basis including at events like reinvent, developer days and other conferences. Here is a link to some recent tutorials for MXNet: (other frameworks have training available online)

- Slides: https://www.dropbox.com/s/08t3h5yezyejgwf/mxnet.zip?dl=0
- NVIDIA GTC2016: https://github.com/dmlc/mxnet-gtc-tutorial
- KDD2016: https://github.com/dmlc/mxnet-notebooks/blob/master/python/outline.Ipynb

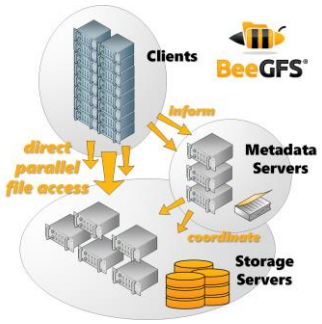
Technology Partner	Product or Company **BeeGFS from Fraunhofer**	Home Page http://itwm.fraunhofer.de/
	Vendor Country of Origin **Germany**	Delivery method **AWS Marketplace**
Domains	Technologies for High Performance Computing and High Performance Data Analysis, Mathematics, Machine Learning, (more information available www.gpi-space.com, www.gpi-site.com, www.beegfs.com)	

Regional availability – All AWS Regions are sovereign.

IRE	DE	UK	FR	CA	US	US	US	US	BR	SG	JP	AU	KR	IN	CN	CN
Dublin	Frankfurt	London	Paris (2017)	Montreal	N. Virginia	N. California	Oregon	Ohio	Sao Paolo	Singapore	Tokyo	Sydney	Seoul	Mumbai	Beijing	Ningxia (2016/17)

12.5 BeeGFS from Fraunhofer ITWM

BeeGFS (formerly FhGFS) is the leading parallel cluster file system, developed with a strong focus on **performance** and designed for very **easy installation** and management. If I/O intensive workloads are your problem, BeeGFS is the solution.



12.5.1 How is it accessed?

BeeGFS is available in AWS Marketplace.

12.5.2 User requirements

When launched through the AWS Marketplace, the user does not need to have any special skills to setup a BeeGFS shared storage cluster. The user will simply need to make some trivial decisions like how much disk space should be available and everything gets setup automatically, leaving the user with a ready to use BeeGFS file system mountpoint. Data stored in BeeGFS can be kept persistent on EBS volumes even when the AWS machine instances are shut down.

12.5.3 Trial

A license-free version of BeeGFS is available in AWS Marketplace.

12.5.4 Billing

BeeGFS with professional support can be launched and billed through AWS Marketplace.

Alternatively, it is also possible to use the free AWS Marketplace product (or install BeeGFS on a custom machine image) and contact info@thinkparq.com to make an individual contract for professional support.

12.5.5 Pricing and Licenses

BeeGFS does not require a license key.

BeeGFS clusters without professional support can be launched from AWS Marketplace and is free of charge (except for AWS instance and EBS usage fees).

BeeGFS clusters with professional support are available for $350.00/mo + $0.08 to $1.326/hr, depending on AWS server instance type.

12.5.6 Support

Community support mailing list and professional support email address are available here: www.beegfs.com/support

12.5.7 Training and Reference Sources

Training is not generally required to use BeeGFS.

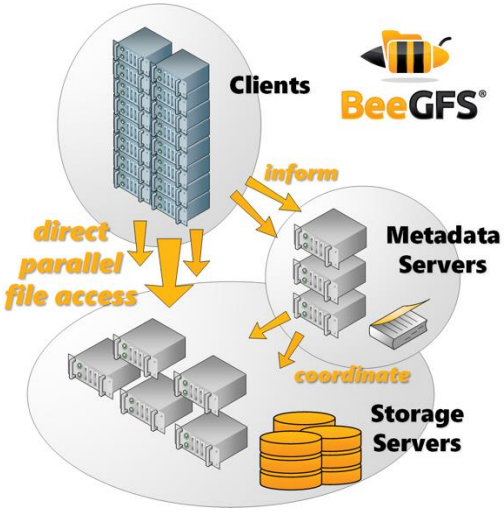
General documentation, whitepapers and performance benchmarks are available on the BeeGFS website:
www.beegfs.com/documentation

To get a general overview, the BeeGFS PDF brochure is available here:
www.beegfs.com/docs/BeeGFS_Flyer.pdf

An AlcesFlight whitepaper, which describes how to use of BeeGFS from an AlcesFlight cluster is available here:
www.beegfs.com/docs/whitepapers/Enabling%20a%20parallel%20shared%20filesystem%20on%20an%20Alces%20Flight%20cluster.pdf

Additional Information from BeeGFS by Fraunhofer

BeeGFS - The Leading Parallel Cluster File System



What is BeeGFS

BeeGFS (formerly FhGFS) is the leading parallel cluster file system, developed with a strong focus on performance and designed for very easy installation and management. If I/O intensive workloads are your problem, BeeGFS is the solution.

Why use BeeGFS

BeeGFS transparently spreads user data across multiple servers. By increasing the number of servers and disks in the system, you can simply scale performance and capacity of the file system to the level that you need, seamlessly from small clusters up to enterprise-class systems with thousands of nodes.

Get The Most Out Of Your Data

The flexibility, robustness, and outstanding performance of BeeGFS help our customers to increase productivity by delivering results faster and by enabling new data analysis methods that were not possible without the advantages of BeeGFS.

Proven By Customers World-Wide

BeeGFS is used all around the globe to provide extremely fast access to storage systems of all kinds and sizes, including some of the fastest supercomputers in the world. Users come from very different sectors and industries, ranging from University HPC centers to finance, media, automotive, energy companies and a lot more.

Technology Partner	Product or Company **CFD Direct from the Cloud**	Home Page http://cfd.direct/cloud
	Vendor Country of Origin **United Kingdom**	Delivery method **AWS Marketplace**
Domains	Computational Fluid Dynamics, Engineering, Physics, Mathematics.	

Regional availability – All AWS Regions are sovereign.																
IRE	**DE**	**UK**	**FR**	**CA**	**US**	**US**	**US**	**US**	**BR**	**SG**	**JP**	**AU**	**KR**	**IN**	**CN**	**CN**
Dublin	Frankfurt	London	Paris (2017)	Montreal	N. Virginia	N. California	Oregon	Ohio	Sao Paolo	Singapore	Tokyo	Sydney	Seoul	Mumbai	Beijing	Ningxia (2016/17)

12.6 CFD Direct Limited

CFD Direct From the Cloud™ (CFDDFC) is a complete platform, deployed on AWS EC2, providing the OpenFOAM software for computational fluid dynamics (CFD) and other supporting software, running on the latest long-term support version of Ubuntu Linux. Users connect to an instance using SSH or, optionally, a remote desktop. CFD Direct From the Cloud is supplied by CFD Direct Ltd (http://cfd.direct), who manage, develop and distribute OpenFOAM free and open source to the public on behalf of the OpenFOAM Foundation, the owner and licensor of OpenFOAM. CFD Direct includes co-founders of OpenFOAM, Chris Greenshields and Henry Weller (its creator and architect), see http://cfd.direct/openfoam/about .

OpenFOAM has a large user base across most areas of engineering and science, from both commercial and academic organizations. It has an extensive range of features to solve anything from complex fluid flows involving chemical reactions, turbulence and heat transfer, to solid dynamics and electromagnetics. It includes tools for meshing, notably snappyHexMesh, a parallelized mesher for complex CAD geometries, and for pre- and post-processing. Almost everything (including meshing, and pre- and post-processing) runs in parallel as standard, enabling users to take full advantage of computer hardware at their disposal.

12.6.1 How is it accessed?

CFD Direct From the Cloud is a product available in the AWS Marketplace, see http://cfd.direct/cloud/aws .

12.6.2 User requirements

OpenFOAM is free software, available from the OpenFOAM Foundation (http://openfoam.org), accompanied by resources, including a User Guide and example cases, to assist its users. Its effective use requires some familiarity and understanding of

CFD, including its key components: geometry and meshing, fluid dynamics and modeling, numerical methods, data analysis and general computing.

OpenFOAM has always been developed for Linux/UNIX operating systems as a collection of 200+ applications that can are executed via a command line interface. Commands can be collected to form effective, customized CFD workflows, with further automation through short scripts which capture procedures for future use, in a way that documenting GUI clicks cannot.

Because OpenFOAM is command-line driven, it transitions seamlessly to the cloud with AWS EC2. As a preconfigured system, with updates, CFDDFC takes away the cost of system administration — work that includes builds, installation, tuning and maintenance of OpenFOAM and supporting software. That includes the ParaView visualization toolkit for post-processing, parallelization libraries, and video and graph-drawing tools.

CFD simulations can be deployed by an SSH connection to an EC2 instance: a secure login procedure from Linux, Windows and Mac OSX operating systems, see http://cfd.direct/cloud/aws/connect. For users who optionally wish to visualize results of their simulations via the cloud, CFDDFC is configured to enable users to connect using the X2Go remote desktop client, also for Linux, Windows and Mac OSX, see http://cfd.direct/cloud/aws/remote-desktop. CFDDFC therefore provides OpenFOAM users a familiar experience to running OpenFOAM on-premises, without the overhead of system administration. Users can choose which EC2 instance to run from those available for use with CFDDFC, e.g. C3 and C4 instances. We also provide instructions on clustering instances for more advanced users wishing to scale to larger parallel computations, see http://cfd.direct/cloud/aws/cluster.

12.6.3 Trial

There is no software charge for running CFDDFC on the t2.micro instance, so a free trial is effectively available for 12 months to new AWS subscribers within the AWS Free Tier, see https://aws.amazon.com/free.

12.6.4 Billing

CFD Direct From the Cloud is a pay-per-use product, billed directly through AWS Marketplace. There are no authorized resellers for the product. All billing is in US Dollars (USD).

12.6.5 Pricing & Licenses

CFD Direct From the Cloud is charged as an hourly software fee, alongside the hourly fees for EC2, and other fees for data storage and transfer. The software costs are typically a small fraction of the on-demand charges for EC2 and are listed on the product page on AWS. All the software provided as part of CFDDFC, is free and open-source, requiring no additional fees for their use. Details of the open source licences of the software, including OpenFOAM, ParaView, Scotch/PTScotch, OpenMPI, CGAL, X2Go and the numerous programs contained in Ubuntu Linux, are provided on the product page. There

is no separate End User License Agreement (EULA) for CFDDFC, although trademark and copyright of logos of CFD Direct and CFDDFC must be respected.

We have priced a typical CFD workload of external aerodynamics around a vehicle, see http://cfd.direct/cloud/cost. Using all 18 physical cores of a c4.8xlarge instance, we could generate a mesh of ~20m cells, run a steady-state simulation to convergence, calculate force coefficients and visualization of pressure distribution, velocity profiles and streamlines, all in parallel, within 8 hours. The total cost of this workload, including data storage and transfer, was $17 using on-demand instances and $6 using spot instances.

12.6.6 Support

CFD Direct will provide support on the configuration and launch of an instance. Users must follow the instructions beginning from http://cfd.direct/cloud/aws/setup and CFD Direct will answer questions regarding these instructions submitted through the "Installation Queries" web form at the designated cloud support web page, see http://cfd.direct/cloud/support .

12.6.7 Training and Reference Sources

CFD Direct provides training for the OpenFOAM software through its portfolio of OpenFOAM Training courses including Essential CFD, Applied CFD and Programming CFD, see http://cfd.direct/openfoam-training. The training addresses the challenges of CFD analysis through a modular curriculum that builds competence across the core components of CFD. Training is hands-on, presenting participants with representative cases spanning a range of scientific disciplines and industries, e.g. external aerodynamics of a car, wind flow around buildings, static mixer, nozzle jet, cyclone, water channel, exhaust system, propeller, etc. We teach design of CFD solutions, starting with a prototype case that is then enhanced in small, digestible steps with periods of reflection to reinforce new concepts.

Training is delivered globally through: public scheduled classes; live virtual classes; and, on-site. The **training is updated for the latest developments of OpenFOAM, that specifically includes new** features introduced by CFD Direct to make it easier to use, leaving more time for participants to build and practice their CFD skills.

Further references:

- OpenFOAM documentation: http://cfd.direct/openfoam/documentation
- OpenFOAM videos: http://cfd.direct/openfoam/videos
- Getting Started with CFDDFC video: http://cfd.tips/c001
- About CFD Direct Ltd: http://cfd.direct/about
- CFD Direct news: http://cfd.direct/news

Additional Information from CFD Direct

Use Case: External Aerodynamics around a Vehicle

OpenFOAM is free software for computational fluid dynamics (CFD). It offers an alternative to proprietary CFD software which incurs large upfront costs in the form of

licence fees and support and maintenance contracts. Cloud computing extends this idea to hardware. It avoids minimum spend commitments and long-term contracts, replacing large upfront expenses with low, variable payments that only apply to what you use. CFD simulations cover a range of sizes and complexity. Activity is generally interrupted by quiet periods when simulations are halted, e.g. when a simulation ends during non-working hours, or during analysis of results, preparation of a new simulation, etc. The fluctuating demand makes CFD well suited to a pay-per-use model.

Cost Breakdown

Users need to understand the costs involved to get the best value from cloud and be confident that they fall within budget. The charges associated with EC2 are: EC2 instance costs; data storage and transfer to and from the EC2 instances; and, software costs.

There are 4 ways to pay for EC2 instances, with On-Demand and Spot instances being the 2 options offering a 'pure' pay-per-use model. Charges for a running EC2 instance (a virtual server on AWS) are incurred hourly, with a fraction of an hour billed as a whole hour. On-Demand instances are sold at a fixed price where availability is guaranteed (within the limits of the SLA). Spot instances are spare EC2 instances that users can bid for, whose price fluctuates based on the supply and demand of available EC2 capacity. When a user has a bid accepted for a Spot instance, they are **charged the Spot market price** — **not the bid price** — while the instance runs.

Spot instances can offer substantial savings over On-Demand instances, as shown in the AWS Spot Bid Advisor. For example, a `c4.8xlarge` instance that includes 18 physical cores is priced at $1.675 per hour in the US-Ohio (`us-east-2`) region. The Spot price history for the past 24 hours (2016-11-11) on that instance, in the same region, fluctuates between $0.20 and $0.40 per hour, with an average price of approximately $0.30 per hour, saving 82% of the On-Demand price. Savings are high because AWS must retain a significant amount of unused capacity to ensure availability of On-Demand instances. **CFD is perfectly suited to take advantage of Spot instances** because it can accommodate a delay on the rare occasion a bid is rejected.

Data is stored using Amazon Elastic Block Store (EBS) attached to an EC2 instance. It is charged by the amount of storage provisioned in GB per month, pro-rated to the hour, until the storage is released. When running CFD with EC2, the storage can be provisioned for the duration of the simulation, with the necessary data transferred out of the instance before it is terminated and the storage is released. The cost of EBS is typically $0.10 per GB-month.

Amazon Simple Storage Service (S3) offers better value for longer term data storage. S3 pricing varies by region and frequency of access, but between $0.0125 are $0.03 per GB-month is typical. Amazon Glacier provides storage at an even lower cost of $0.007 per GB-month for data archiving. are listed as part of the On-Demand EC2 pricing. Transfer is charged at $0.09 per GB beyond the first 1 GB of data and up to the first 1 TB of a given month. This can be a significant cost which users can minimize with careful OpenFOAM configuration, e.g. setting up run-time post-processing of their simulations, rather than downloading raw case data, writing data in binary or gzip-compressed format, etc.

CFD Direct From the Cloud is charged as an hourly software fee, alongside the hourly fees for EC2. The software costs are a small fraction of the On-Demand charges for EC2 and are listed on the CFDDFC product page on AWS.

Simulation of External Aerodynamics around a Vehicle

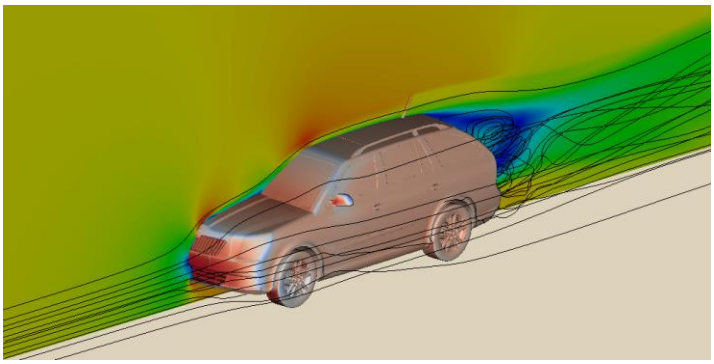
We set up a simulation in OpenFOAM to run a steady-state, turbulent, incompressible flow simulation of external aerodynamics around a car with:

- a vehicle surface geometry, OBJ format, consisting of ~3 m triangles, ~30 MB gzip compressed;
- meshing configuration with *snappyHexMesh* to generate a mesh of ~18 m cells;
- initial and boundary conditions, turbulence models and wall functions, and schemes and solver settings;
- configuration to decompose onto 18 domains for running in parallel on 18 physical cores using the `c4.8xlarge` instance;
- configuration to calculate force coefficients, and VTK files of vehicle, cutting plane and streamlines for visualization.

Charge	On-Demand Instance	Spot Instance
EC2 Instance	$1.675/hr → $13.40	~$0.300/hr → $2.40
Data Storage	20 GB → $0.02	20 GB → $0.02
Data Transfer	10 GB → $0.90	10 GB → $0.90
Software Charges	$0.335/hr → $2.68	$0.335/hr → $2.68
TOTAL	**$17.00**	**$6.00**

The table above shows costs of running the simulation priced for both On-Demand and Spot instances. A `c4.8xlarge` instance was used for **8 hours**, running 45 minutes of mesh generation followed by 410 minutes of simulation with the *simpleFoam* solver, running for 2000 iterations. The case was configured to purge all except 2 time directories (with `purgeWrite 2;` in the *controlDict* file) and to gzip-compress the field data files. Field data fitted comfortably within 5 GB of storage, and with run-time processed data such as VTK files for visualization (written every 50 iterations) system software and some spare capacity, the provisioned data storage was 20 GB. Although only a fraction of that data needed to be downloaded at the end of the simulation, we allowed for 10 GB in the calculation above.

Cloud vs On-Premises Hardware



To compare these cloud costs with the cost of on-premises hardware, we can make some very approximate cost estimates and assumptions. We can start with $5,000 for a computer with comparable specification to `c4.8xlarge` (2× Intel Xeon E5-2666 v3 Haswell processors, 2.9-3.4 GHz, 64 GB RAM server). We add 100% overhead costs, covering system administration, electricity, rental space, assuming a 3-year lifetime of the computer, giving a total cost of $10,000.

Let us assume a utilization of 8 hours per day for 220 days per year. Over 3 years, the cost of running in the cloud with On-Demand instances would be ~$11,200, comparable to on-premises costs, especially factoring in other effects such as cloud hardware upgrades during the period. The cost of using Spot instances, however, would be ~40% of the on-premises cost. On the basis of cost alone, **the case for using cloud for CFD is compelling** given its suitability for running on Spot instances.

Technology Partner	Product or Company **DNAnexus**	Home Page http://www.dnanexus.com/
	Vendor Country of Origin **USA**	Delivery method **PaaS Portal**
Domains	Bioinformatics, Life Sciences, Health Care	

Regional availability – All AWS Regions are sovereign.

IRE	DE	UK	FR	CA	US	US	US	US	BR	SG	JP	AU	KR	IN	CN	CN
Dublin	Frankfurt	London	Paris (2017)	Montreal	N. Virginia	N. California	Oregon	Ohio	Sao Paolo	Singapore	Tokyo	Sydney	Seoul	Mumbai	Beijing	Ningxia (2016/17)

12.7 DNAnexus

DNAnexus provides secure and compliant scientific stack and data management services, enabling collaborative research to rapidly deliver insights from high resolution 'omics, molecular, imaging, and clinical data.

Genomics, proteomics, metabolomics, and other 'omics data, in combination with molecular, imaging, and phenotypic data offers enormous potential to transform healthcare, but these are such large, complex, and sensitive datasets that gaining insights from them poses a number of immediate and significant challenges around scalability, sharing, and security.

DNAnexus provides an infrastructure and toolset that is optimized for addressing these challenges for people who are pursuing high resolution data analysis approaches to health, in the clinic and in the research lab. DNAnexus works with organizations to tackle some of the most challenging and exciting opportunities in human health, by making it easier—and in many cases feasible—to work with this rich data.

12.7.1 How is it accessed?

- PaaS – user connects to public portal, registers for an account, and obtains services via the web interface and/or API.
- Customers create SaaS portals on top of the DNAnexus APIs (e.g. precisionFDA.gov), and their customers access DNAnexus services indirectly via the portal front-end which is built on the API.

12.7.2 User requirements

DNAnexus addresses multiple types of users. Expert users are generally bioinformaticians with experience creating genomic analysis pipelines, statistical analysis, visualization and interactive exploration. Expert users generally interact with the

platform using a CLI. Clinical users generally run prescribed pipelines through the web interface, using hardened and locked-down pipelines developed by expert users. Administrative users use the platform to provision and manage users, groups, and cost. The platform makes it easy for all types of users to perform their jobs.

12.7.3 Trial

New users who self-signup receive $50 in service credits which can be used at any time.

12.7.4 Billing

DNAnexus bills for service in US dollars and operates as a US corporation. Individual users can procure services directly from DNAnexus by providing billing information via the web interface or contacting DNAnexus directly. Individual users are invoiced for services actually used and there is not up-front payment component.

Commercial and academic groups can establish an enterprise contract with DNAnexus which can provide volume discounts and other benefits not generally available to individual users. Enterprise contracts can include one or more of the following payment options: annual license, volume discount prepayment, per-test pricing, and pay-as-you-go based upon actual resource consumption. All commercial relationships are established directly with DNAnexus.

12.7.5 Pricing & Licenses

DNAnexus provides a variety of purchase and pricing options. Individual users pay for compute, storage, and egress services by unit of usage, similar to purchase AWS services, without any upfront fee. DNAnexus provides two service tiers, normal priority which may incur some delay in launching a workflow as cost-optimized resources are marshalled, priced at a reduced rate for compute, and high priority when the users requires analysis pipelines to launch immediately.

Given the breadth of workloads that can be deployed on the platform, and the resource consumption variability of genomic workloads, it is not particularly meaningful present example pricing; however mapping and alignment of a whole exome is < $10 and < $100 for a whole genome. Again there are many variables which can impact the actual price.

12.7.6 Support

All DNAnexus users receive basic support. Support requests can be submitted via the company website or from the actual platform while logged in. Basic support is included at no extra charge.

Commercial and academic groups can opt-in to enterprise or premium support in order to obtain the response and resolution time commitments which are required for their business as based upon DNAnexus. Pricing for enterprise and premium support is determined on a per-contract basis.

The support tiers are described in https://www.dnanexus.com/customer-support.

12.7.7 Training and Reference Sources

Customers can request scientific services for training and onboarding from DNAnexus. These services are charged on a case-by-case basis, although DNAnexus generally provides sufficient training and consultation without charge for customers to onboard and successfully utilize the platform.

Technology Partner	Product or Company **DRAGEN Genome Pipeline (Germline)**	Home Page http://www.edicogenome.com/
	Vendor Country of Origin **United States**	Delivery method ***AWS Marketplace***
Domains	Bioinformatics, genetic analysis, whole genome sequencing, whole exome sequencing, etc.	

Regional availability – All AWS Regions are sovereign.																
IRE	**DE**	**UK**	**FR**	**CA**	**US**	**US**	**US**	**US**	**BR**	**SG**	**JP**	**AU**	**KR**	**IN**	**CN**	**CN**
Dublin	**Frankfurt**	**London**	**Paris (2017)**	**Montreal**	**N. Virginia**	**N. California**	**Oregon**	**Ohio**	**Sao Paolo**	**Singapore**	**Tokyo**	**Sydney**	**Seoul**	**Mumbai**	**Beijing**	**Ningxia (2017)**

12.8 Edico Genome

The DRAGEN (Dynamic Read Analysis for Genomics) Platform is based on the highly reconfigurable DRAGEN Bio-IT Platform. The DRAGEN platform uses a field-programmable gate array (FPGA) to provide hardware-accelerated implementations of genome pipeline algorithms, such as BCL conversion, compression, mapping, alignment, sorting, duplicate marking and haplotype variant calling. The highly flexible DRAGEN platform allows Edico Genome to develop custom algorithms as well as refine and improve existing pipelines.

12.8.1 How is it accessed?

The DRAGEN Genome Pipeline is accessible through AWS Marketplace. It is a pay-per-use AMI.

12.8.2 User requirements

Users should be familiar with setting up instances on AWS Marketplace as well as setting up the relevant storage devices with those instances. Users should be familiar with working with and viewing genomic data and file formats (i.e. FASTQ, BAM, VCF, gVCF, etc.). For tips on how to use the DRAGEN Genome Pipeline and DRAGEN Exome Pipeline, User and Quick Start Guides are available through our app on AWS Marketplace.

12.8.3 Trial

A one-day free trial is available for all of our DRAGEN apps on AWS Marketplace.

12.8.4 Billing

Our DRAGEN apps on AWS Marketplace are pay-per-use.

12.8.5 Pricing & Licenses

The price of our pipelines is on a pay-per-hour basis with the detailed pricing below.

DRAGEN Genome Pipeline:

- F1.2xlarge: $18/hr
- F1.16xlarge: $50/hr

DRAGEN Exome Pipeline

- F1.2xlarge: $40/hr

12.8.6 Support

We support questions and issues pertaining to the DRAGEN apps. All issues can be reported by entering a claim here and someone from Edico Genome will follow up with you.

12.8.7 Training and Reference Sources

Training is not provided at this time, but we have a number of User and Quick Start Guides available through our app on AWS Marketplace. Edico Genome provides support for users that have specific questions regarding our apps.

Additional information from DRAGEN Genome Pipeline

Available Apps on AWS Marketplace

DRAGEN Genome Pipeline (Germline)

The DRAGEN Genome Pipeline enables ultra-rapid analysis of Next Generation Sequencing (NGS) data, reducing the time required for analyzing a whole genome at 30x coverage from ~20-30 hours using the current industry standard, BWA-MEM+GATK-HC software, to ~25 minutes (f1.16xlarge instance) and ~60 minutes (f1.2xlarge instance) while also improving accuracy for both SNPs and INDELs. This pipeline harnesses the tremendous power of the DRAGEN Platform and includes highly optimized algorithms for mapping, aligning, sorting, duplicate marking, haplotype variant calling, compression and decompression.

DRAGEN Exome Pipeline (Germline)

The DRAGEN Exome Pipeline enables ultra-rapid analysis of Next Generation Sequencing (NGS) data, reducing the time required for analyzing a whole exome at 200X coverage from ~4 hours using the current industry standard, BWA-MEM+GATK-HC software, to 8 minutes (f1.2xlarge instance) while also improving accuracy for both SNPs and INDELs. This pipeline harnesses the tremendous power of the DRAGEN Platform and includes highly optimized algorithms for mapping, aligning, sorting, duplicate marking, haplotype variant calling, compression and decompression.

Getting Started with DRAGEN on AWS Marketplace

Before you can use the DRAGEN on the AWS instance for aligning and variant calling reads, you must first create a hash table using the reference genome of your choice.

The hash table can be created using the command below:

```
dragen --build-hash-table true --ht-reference <reference_fasta_file>
--output-dir <hash_table_directory>
```

The 'dragen --build-hash-table' command is multithreaded and defaults to 8 threads. You can use '--ht-num-threads' with a value up to 32.

For a reference fasta, the hash table only needs to be built once.

Once the hash table is available, then DRAGEN is ready to run the end-to-end pipeline to Map-Align and perform Variant Calling.

An example command line with the minimum parameters is provided below:

```
dragen -f
-r <hash_table_directory>
-1 <sample_R1.fastq.gz>
-2 <sample_R2.fastq.gz>
--enable-variant-caller true
--vc-sample-name RGSM
--output-directory <output_directory>
--output-file-prefix <output_file_prefix>
--enable-duplicate-marking true
--enable-map-align-output true
```

The above command will load the reference hash table, run Map-Align and Variant Caller, sort, dedup and save BAM and VCF files to the specified output directory.

For an exome, it is possible to provide a target BED file using the command line option '--vc-target-bed <BED_file>'.

For more details, please refer to the DRAGEN Quick Start and User Guide available on our app on AWS Marketplace.

Technology Partner	Product or Company **figshare**	Home Page https://figshare.com/
	Vendor Country of Origin **United Kingdom**	Delivery method **AWS Marketplace**
Domains	All research domains	

Regional availability – All AWS Regions are sovereign.

IRE	DE	UK	FR	CA	US	US	US	US	BR	SG	JP	AU	KR	IN	CN	CN
Dublin	Frankfurt	London	Paris (2017)	Montreal	N. Virginia	N. California	Oregon	Ohio	Sao Paolo	Singapore	Tokyo	Sydney	Seoul	Mumbai	Beijing	Ningxia (2016/17)

12.9 figshare

Figshare is a web-based platform to help academic institutions, research organizations, publishers and individual researchers to manage, disseminate and measure the attention and impact of certain research outputs. The user-friendly approach allows research outputs to be made available in a citable, shareable and discoverable manner. The platform can also be used to help research organizations to manage digital research outputs associated with project-based collaborations. The platform facilitates the visualization and exchange of diverse research outputs from compound structures to datasets in a wide variety of file formats. The platform can store any file type and preview over 630 file formats from within the web browser helping the user to more quickly evaluate the relevance and usefulness of a data file.

Figshare works with academic institutions and research organizations globally to help them meet key funder recommendations and mandates, and to provide world-leading tools to support an open culture of data sharing and collaboration. *Figshare for institutions* focuses on four key areas: research data management, reporting and statistics, research data dissemination and administrative control.

Altmetric Badges are integrated into the figshare platform so that news, policy document and social media attention to published outputs can be tracked and reported on through the *Altmetric Explorer* tool (https://www.altmetric.com/audience/institutions/). This is important for researchers looking for a tracking and monitoring mechanism to be able to demonstrate the broader social, health and economic impact of their research to their funders and other key stakeholders, after their outputs have been published and made openly available.

Research organizations may also use figshare to set up wholly private cloud instances restricted to use by authenticated users only. This approach supports internal collaboration around the digital research outputs that are generated from projects.

In a second important use case, figshare also provides data infrastructure and visualization services to major academic publishers including The Royal Society, Springer Nature, PLOS, Taylor & Francis, PNAS, Wiley and the American Chemical Society. This helps these publishers store, visualize and get usage for their supplemental data files related to journal articles and book chapters, and other miscellaneous data files supplied by their editors and authors.

12.9.1 Access

Figshare currently works at the organizational level to set up custom branded portals and repositories. These portals under the *figshare for institutions* product offering are freely accessible on the Web, however administrators and authors have the power to control access to certain restricted content. Installation requires plugging into the institutional single sign-on system and plugging into the storage system of choice. Currently figshare can plug into any storage system with a RESTful API. The figshare infrastructure is hosted on AWS regionally distributed depending on the data sovereignty rules of the institution. Figshare is plugged into AWS S3 storage as default.

Figshare for groups is set to launch on the AWS Marketplace in 2017, allowing smaller groups and labs to create their own secure repositories for the digital products of their research. The figshare interface ensures that all content is persistently available with citable DOIs. The impact and metrics associated with the research outputs are tracked and available to query.

Figshare's Open APIs allow for both upload and transfer of files and metadata to other systems to ensure funder mandate compliance.

12.9.2 User requirements

Figshare is designed to be as simple as possible for researchers to get started, based on the premise that providing a low barrier to entry is key to getting researchers to engage in data management and data sharing. *Figshare for groups* will offer a self-serve set up through AWS Marketplace. Technical expertise on behalf of the user will not be required to do the set up. A simple web form will launch the instance and plug into a user's S3 account.

Much like other file sharing platforms, once logged into the figshare platform, users are taken to the "My data" section where they can drag and drop files into the browser to upload, add a basic number of metadata fields to describe their research, and share privately or publish on the Web.

figshare also offers a powerful open API which enables users to push and pull data, so that figshare can be easily integrated into existing information systems and research workflows. For example, the technical team at the University of St Edwards Library, Texas built this data showcase with interactive visualizations on weather patterns, using the figshare API http://sites.stedwards.edu/ozone/site/u-s-data/se-texas/se-texas-2015/#3003750.

Researchers can set up projects for sharing research with collaborators who can be members of their institution or from external organizations, with the project owner

controlling who has read and write access to the content. Like everything on figshare, projects can be kept private or made public. Research outputs can also be grouped together using collections. All published outputs are displayed in the custom branded portal of the lab or group, providing a single destination to showcase all their research outputs.

12.9.3 Trial

Figshare offers openly available free accounts at figshare.com and all content is available to view and download without the need to login. The free accounts are limited to 20GB of private storage space per user and do not have custom branded portals for groups or labs.

Groups and Academic Institutions can sign up for 6 month paid pilots which includes 1TB of AWS S3 storage. The figshare application can also be plugged into the institutional AWS storage if that is required by the customer.

12.9.4 Billing

Figshare bills customers directly for the software as a service. Institutional customers can pay for the storage used through their own AWS account, or use their own storage set up.

12.9.5 Pricing & Licenses

Figshare offers institutions an annual license fee for the software, based on the research intensity of the institution. Storage costs can be quoted as annual AWS S3 charges, or the institution can choose to provide their own storage provide, including the AWS PAYG (pay as you go) model.

12.9.6 Support

Figshare offers a customer support portal at https://support.figshare.com/ with FAQs, videos and explainer sheets on the platform's key functionality. Support queries can be submitted via the support portal, by e-mail to info@figshare.com or on Twitter to @figshare. All support requests are responded to within 24 hours and usually resolved within this timeframe as well.

12.9.7 Training and Reference Sources

Figshare provides monthly "getting started" webinars for new labs, groups or institutions which covers the basic functionality of figshare. Figshare also conducts monthly webinars on more advanced functionality, and on topics such as research data management best practices, policy and developments which can be found on their blog https://figshare.com/blog.

In person training is available for GBP £2000 which will include multiple tailored researcher training sessions, advanced API training and support staff training. The same training is also available remotely at GBP £500 per session, which covers getting started, advanced functionality and new developments in data policy.

Additional Information from Figshare

How was figshare setup?

Figshare was born out of the open research movement with the mission to make all academic data as open as possible. The company heavily promotes the ethos that underlying data must be also be published in order to validate and reproduce results.

With a wide corpus of openly available data, new avenues of inquiry can be unlocked by combining datasets or approaching them from a different angle or discipline. There is also a belief that if research is funded by the public purse then all products of the funding should be made openly available. Not limiting research outcomes to summary research papers increases the return on investment for all research funding.

Not only does figshare believe in open data but the company also believes in open interoperability with other academic research information systems. The Figshare API is open for any researcher or programmer to build upon, thereby not limiting the scope of functionality to what Figshare deems necessary. The automated flow of information, with minimal administrative burden on the researcher is paramount to the figshare mission. The less time researchers spend keying and rekeying data, the more time can be spent on research.

Ultimately, figshare believes that by opening up data and research in general it will increase the pace and breadth of discovery. Through the work of figshare with its partner institutions and organizations, over 3 million research objects are now openly available for discovery and reuse by the research community.

How does Figshare fit into the broader context of Open Science?

Figshare was born out of the open research movement with the mission to make all academic data as open as possible. The company heavily promotes the ethos that underlying data must be also be published in order to validate and reproduce results.

With a rich corpus of openly available data, new avenues of inquiry can be unlocked by combining datasets or approaching them from a different angle or discipline. There is also a belief that if research is funded by the public purse then all products of the funding should be made openly available. Not limiting research outcomes to summary research papers increases the return on investment from all research funding.

Not only does figshare believe in open data but the company also believes in open interoperability with other academic systems. The figshare API is open for any researcher or programmer to build upon, thus not limiting the scope of functionality to what figshare deems necessary. The automated flow of information, with minimal administrative burden on the researcher is paramount to the figshare mission. The less time researchers spend keying and rekeying data, the more time can be spent on research.

Ultimately, figshare believes that by opening up data and research in general, that will increase the pace and breadth of discovery. Thanks to the work of figshare with their partner institutions and organizations over 3 million research objects are now openly available.

How does it work?

Researchers can create an account for free, with 20 gigabytes of private storage space and unlimited public storage space - as long as data is made openly available, figshare will pay the cost of hosting. The simple interface allows users to drag and drop files into the browser, add some contextual metadata to increase discoverability, share privately with collaborators and/or publish. The process of uploading and making data publicly available that takes minutes, rather some of the more onerous academic systems which have a higher barrier to entry for end-users.

Figshare for institutions uses a similar frontend to maintain the simple user-experience for researchers but introduces a number of features to give institutions greater visibility and control. With enhanced metadata options, curation workflow and integration with single-sign-on, storage and other university systems, the institutional product is more suitable for enterprise-scale research.

Figshare for publishers integrates with existing publisher platforms, which depending on the data publication policy, encourages researchers to publish the underlying data to back up their findings. The seamless workflow allows data to be displayed in the browser, alongside the traditional research article, allowing for a richer story telling experience. All data from publishers is also available through a custom branded, searchable portal and incorporated into the wider figshare corpus.

Technology Partner	Product or Company **BaseSpace® Sequence Hub Illumina, Inc.**	Home Page https://www.basespace.illumina.com/
	Vendor Country of Origin **United States**	Delivery method **SaaS Portal**
Domains	Bioinformatics, Genomics, Transcriptomics, Microbiology, Oncology, Proteomics	

Regional availability – All AWS Regions are sovereign.																
IRE	**DE**	**UK**	**FR**	**CA**	**US**	**US**	**US**	**US**	**BR**	**SG**	**JP**	**AU**	**KR**	**IN**	**CN**	**CN**
Dublin	Frankfurt	London	Paris (2017)	Montreal	N. Virginia	N. California	Oregon	Ohio	Sao Paolo	Singapore	Tokyo	Sydney	Seoul	Mumbai	Beijing	Ningxia (2016/17)

12.10 Illumina, Inc. – BaseSpace® Sequence Hub

Next-generation sequencing (NGS) has revolutionized the way and rate at which biomedical research is conducted. As the cost of sequencing decreases, the volume of NGS-generated data increases, presenting researchers with progress-hindering problems. Secure data storage and management, complex data analysis, and sharing results with collaborators are challenges that can result in nonuniform methods within institutions and labs, conflicting results, and increased operational overhead. BaseSpace Sequence Hub is a genomics cloud computing platform, designed to bring simplified data management and analytical sequencing tools directly to investigators in a user-friendly format. BaseSpace Sequence Hub provides flexibility and convenience with an array of bioinformatics tools, significantly expanding the possibilities of yielding meaningful results from NGS data.

Labs pursuing next-generation sequencing traditionally required the services of a highly trained bioinformatician and dedicated infrastructure to perform data management, analysis, and storage. BaseSpace Sequence Hub helps automate bioinformatic analysis using cloud-based software applications. Scalable storage grows with your research needs. Designed with the biologist in mind, BaseSpace Sequence Hub push-button bioinformatics applications are simple to use and produce biologically relevant results from raw data. BaseSpace Sequence Hub output files are industry standards that use open formats. These results can be imported into downstream scientific software tools for further analysis. All necessary operations occur in one place.

12.10.1 How is it accessed?

Saas Portal – Users connect to the BaseSpace Sequence Hub public portal and register for an account. Enterprise-tier users have access to their own private BaseSpace Sequence Hub domain.

12.10.2 User requirements

No IT skill is needed to operate our BaseSpace Sequence Hub product. If a user is comfortable with a web browser, it is pretty much enough for basic operation. Of course, science skills in Next Generation Sequencing and NGS bioinformatics analysis are necessary. However, a user of BSSH does not have to program anything or learn how to code. This opens up the NGS analysis to molecular biologists who do not have specific IT training. That said, advanced users who possess IT skills can do a lot more with BaseSpace Sequence Hub than the basic users who do not have such skills, for example, they can create their own apps to run on BaseSpace Sequence Hub as a platform, or use command-line access to run the Sequence Hub without its web GUI.

12.10.3 Trial

BaseSpace Sequence Hub is free to sign up and use. New users are provided a free 30-day free trial in which they are provided open access to all features and analysis applications, as well as 250 iCredits to use to pay for computation. After using all the iCredits or the 30-day trial ends, they are able to perform many functions including sequence run upload, run monitoring, data sharing and transferring, download, and launch a limited number of free apps. Basic users are limited to 1TB of storage.

12.10.4 Billing

Users who want to upgrade to a Professional or Enterprise account may contact Illumina directly. The billing relationship is between the end-user (institution) and Illumina, and occurs in the local currency.

BaseSpace Sequence Hub offers a pay as you go feature for the use of paid Apps and storage allowing customers to only use and pay for what they need. All BaseSpace Sequence Hub accounts come with 1 TB of free storage and access to a limited number of free Apps.

Subscription accounts can, optionally, purchase storage subscriptions and use the pay as you go feature for paid Apps and any additional storage beyond the baseline storage which is either the free 1TB or storage subscription. See table below for a summary of the billing features.

	Basic	Professional	Enterprise
Included Storage	1TB	1TB	1TB
Free Apps	Yes	Yes	Yes
Purchase Storage Subscription	No	Yes	Yes
Additional Storage	No	Pay as you go	Pay as you go
Paid Apps	No	Pay as you go	Pay as you go

12.10.5 Pricing & Licenses

There are three tiers of BaseSpace Sequence Hub. The Basic tier is free for all users, and the Professional and Enterprise are annual subscription-based tier.

Is there a license to buy? If so, what is the pricing model? Can a user bring an enterprise or site-side license from their institution to use in the cloud? Under what conditions?

It's advantageous to show some simple worked examples here that outline what a "typical" (or familiar) workload might entail, and how much it would cost.

12.10.6 Support

BaseSpace Sequence Hub users are fully supported by Illumina Technical Support and a global network of application specialists. They can be contacted via Illumina Technical Support (techsupport@illumina.com) or through the user's account manager.

Provide details on your support offering, subscription rates, and contact methods.

12.10.7 Training and Reference Sources

Help and reference documentation is posted to the BaseSpace Sequence Hub Help site. In-depth training and consultations are provided by the Illumina Professional Services. Professional and Enterprise subscriptions include 8 and 24 total hours, respectively, of annual Bioinformatics Professional Services support.

Technology Partner	Product or Company **Intel® Cloud Edition for Lustre***	Home Page http://www.intel.com/content/www/us/en/lustre/intel-cloud-edition-for-lustre-software.html
	Vendor Country of Origin **USA**	Delivery method **AWS Marketplace**
Domains	Financial Services, Manufacturing, Health and Life Sciences, Weather Forecasting, Oil & Gas, Government Research	

Regional availability – All AWS Regions are sovereign.

IRE	DE	UK	FR	CA	US	US	US	US	BR	SG	JP	AU	KR	IN	CN	CN
Dublin	Frankfurt	London	Paris (2017)	Montreal	N. Virginia	N. California	Oregon	Ohio (2016/17)	Sao Paolo	Singapore	Tokyo	Sydney	Seoul	Mumbai	Beijing	Ningxia (2016/17)

12.11 Intel Cloud Edition for Lustre

Lustre* is an open source parallel file system for high performance and technical computing. It is used in 70% of the Top 100 Supercomputers worldwide[1] and is best known for its speed and scale. Intel is the lead contributor to the Lustre community and is widely considered the thought leader in the space.

In addition to the contributions Intel makes to the Lustre community, Intel also builds packaged and supported versions of the Lustre File System. This includes the Intel® Cloud Edition for Lustre* software, available on Amazon Web Services. The Intel Cloud Edition for Lustre software enables data scientists to run their HPC applications in the public cloud by providing the performance required by I/O thirsty applications.

12.11.1 How is it accessed?

Marketplace Launch

12.11.2 User requirements

Users should have a basic understanding of Linux, as well as a high level understanding of parallel file systems. Intel offers free, self-paced Lustre training here: http://www.intel.com/lustre-training . Users should also have a basic understanding of Amazon Machine Images.

12.11.3 Trial

Intel does offer an evaluation version of Intel® Cloud Edition for Lustre* software, which is intended to help users become familiar with how the file system works and how the file system is created within the AWS Marketplace. Performance on the Evaluation Version will be limited as it is not intended for production workloads. A link to the evaluation version is here:

https://aws.amazon.com/marketplace/pp/B01344V0C0?qid=1478275114917&sr=0-4&ref_=srh_res_product_title

12.11.4 Billing

Transacting Intel® Cloud Edition for Lustre* software is easy. It is transacted directly through the AWS Marketplace.

12.11.5 Pricing & Licenses

Pricing is based on the number of Object Storage Servers and instance type in a cluster. Support is available for a fixed monthly fee. A full list of options, as well as pricing for each instance type can be found here: https://aws.amazon.com/marketplace/seller-profile?id=d1c6e336-5f6f-4234-82a8-a57463081a35

12.11.6 Support

Users who purchase support are entitled to unlimited file system support via email or phone. Intel also offers paid professional services to setup and configure a file system. Details on the support can be found here: https://aws.amazon.com/marketplace/pp/B01MAWHQTD?qid=1478275114917&sr=0-3&ref_=srh_res_product_title. To discuss paid professional services please email hpdd-cloud-lustre@intel.com.

12.11.7 Training and Reference Sources

Intel offers free web based self-lead Lustre training at http://www.intel.com/lustre-training. Live onsite or remote training can also be purchased. To discuss paid live training please email hpdd-cloud-lustre@intel.com.

[1]Based on Intel Analysis of November 2015 Top500: www.top500.org

*other names and brands may be claimed as the property of others

Technology Partner	Product or Company **MathWorks**	Home Page **http://www.mathworks.com/**
	Vendor Country of Origin **USA**	Delivery method **Multiple**
Domains	Broad range of science, engineering, mathematics and statistics.	

Regional availability – All AWS Regions are sovereign.																
IRE	**DE**	**UK**	**FR**	**CA**	**US**	**US**	**US**	**US**	**BR**	**SG**	**JP**	**AU**	**KR**	**IN**	**CN**	**CN**
Dublin	Frankfurt	London	Paris (2017)	Montreal	N. Virginia	N. California	Oregon	Ohio	Sao Paolo	Singapore	Tokyo	Sydney	Seoul	Mumbai	Beijing	Ningxia (2016/17)

12.12 MathWorks has Multiple Offerings for AWS

MathWorks is the leading developer of mathematical computing software. MATLAB, the language of technical computing, is a programming environment for algorithm development, data analysis, visualization, and numeric computation. Universities and learning institutions world-wide use it as a fundamental teaching and research tool.

When you use MATLAB on AWS you can take advantage of specific hardware configurations or use resources (like GPUs, large memory, and additional compute cores) to accelerate performance. Additional MathWorks offerings leverage the elastic nature of AWS to enable parallel computing across multiple instances and scalable application deployment. These MathWorks products are available in AWS to end users with an appropriate license:

- MATLAB, Simulink and other add-on products
- MATLAB Distributed Computing Server
- MATLAB Production Server
- MATLAB Runtime

12.12.1 How is it accessed?

You can provision AWS resources and install MATLAB, MATLAB Production Server and MATLAB Runtime directly. Launching these products on Amazon EC2 involves a series of steps, the majority of which are one-time setup procedures. Thereafter, you can access them on your EC2 instance within a matter of minutes. MATLAB and MATLAB Runtime require you to start an EC2 instance and install the software. For MATLAB Production Server contact MathWorks at matlab.on.aws@mathworks.com for help with configuration. You can also contact MathWorks if you need help with the other products.

For MATLAB Distributed Computing Server, MathWorks Cloud Center is the recommended approach to start clusters on EC2. There are AMIs preconfigured for

multiple releases of MathWorks products which are accessed via Cloud Center. To get started with Cloud Center, visit this page first. Full documentation is available here.

12.12.2 User requirements

Getting up and running with MATLAB and MATLAB Runtime on AWS requires the ability to start an EC2 instance and install your software. Cloud Center simplifies the use of MATLAB Distributed Computing Server on EC2 so that you can access AWS resources from your client MATLAB® session like any other cluster in your own onsite network.

Setting up MATLAB Production Server is a bit more advanced but a MathWorks Engineer is available to help you through the configuration.

You can contact MathWorks for help configuring any product on AWS.

12.12.3 Trial

Instructions for obtaining a trial are linked below:

- MATLAB and other add-on products: Go here.
- MATLAB Distributed Computing Server: Go here.
- MATLAB Production Server: Contact MathWorks.
- MATLAB Runtime: There are no trials because it is available for free.

12.12.4 Billing

Usage of MATLAB, MATLAB Distributed Computer Server, and MATLAB Production Server on AWS requires a direct relationship with MathWorks. If you already have a license or trial, there is no fee for using these products in AWS. Otherwise you can purchase a term or perpetual license to these offerings. In addition, MATLAB Distributed Computing Server is available with on-demand pricing. MATLAB Runtime is available for free and may be used with any compiled MATLAB application on AWS in accordance with the MathWorks and MATLAB Runtime SLA.

MathWorks products can be procured using various currencies and from many jurisdictions.

You will also need to pay AWS directly for use of EC2 instances and any other AWS services.

12.12.5 Pricing & Licenses

In general, all MathWorks products are available for use on AWS. However, those products that require special drivers or specific hardware may not work.

Go here for pricing information and to purchase MathWorks products. For on-demand pricing information associated with MATAB Distributed Computing Server, go here. Please note that licensing in the cloud can be done multiple ways. For MATLAB, we recommend use of Login Named User. For MATLAB Distributed Computing Server, we recommend use of MathWorks Hosted License Manager as described here . For MATLAB Production Server or to use FlexNet licensing, please contact MathWorks.

12.12.6 Support

Visit MathWorks support page for support resources. Access to Technical Support engineers is available if your license is active on MathWorks Software Maintenance Service.

12.12.7 Training and Reference Sources

Visit MATLAB Academy for free online MATLAB training. For other training resources, visit this page.

Additional Information from Mathworks

Getting Started with MATLAB on AWS

Launching MATLAB on AWS involves a series of steps, the majority of which are one-time setup procedures. Thereafter, you will be able to access MATLAB on your EC2 instances within a matter of minutes. To use MATLAB on AWS, you will need the following:

1. A MATLAB license that has the Login Named User feature enabled. Your MathWorks Account must be associated with this license.
2. An AWS account. For billing purposes, you need to provide credit card information to Amazon when you create your account.
3. An NX Client installed on your computer. The NX Client is the required remote display software for accessing MATLAB on the EC2 instance.

An overview of the process involved is shown below:

Step 1: (One-time) Configure MathWorks Account

a. Create a MathWorks Account. Go here.
b. Associate your license to your MathWorks Account. Go here.
c. Enable Login Named User for your license. Go here.

Step 2: (One-time) Configure AWS Environment

a. Create AWS account and create a key pair.

Step 3: Launch EC2 Instance

a. Create and configure an AWS CloudFormation stack.

Step 4: (Once per machine) Set up Instance Access

a. Install and configure NX client.

Step 5: Start MATLAB

Technology Partner	Product or Company **Overleaf**	Home Page http://www.overleaf.com/
	Vendor Country of Origin **United Kingdom**	Delivery method ***SaaS Portal***
Domains	All areas of scholarly writing, editing and publishing, especially Physics, Astronomy, Mathematics, Statistics. Computer Science, Machine Learning, Geography, Geospatial, Graphics, Bioinformatics, Chemistry.	

Regional availability – All AWS Regions are sovereign.

IRE	DE	UK	FR	CA	US	US	US	US	BR	SG	JP	AU	KR	IN	CN	CN
Dublin	Frankfurt	London	Paris (2017)	Montreal	N. Virginia	N. California	Oregon	Ohio	Sao Paolo	Singapore	Tokyo	Sydney	Seoul	Mumbai	Beijing	Ningxia (2016/17)

12.13 Overleaf

Science is a global enterprise, and over two million scientific papers are written every year in collaborations that span the globe. The Overleaf platform brings state of the art, cloud-based collaborative authoring technology to the world of science, to make it faster and more open. Our solutions are designed to make the whole process of writing, editing and publishing scientific papers much quicker and easier for all parties involved.

Overleaf enterprise solutions simplify and accelerate the scientific publishing process by keeping the paper in a single central place through its entire lifecycle. With Overleaf, the paper is stored securely in the cloud, so authors, editors, reviewers and readers can each read, edit or comment on the paper when it is their turn, using only a web browser. For our organizational partners, this means that they are able to provide an innovative and collaborative solution to their users – proactively providing them with new technology tools to help them succeed.

12.13.1 How is it accessed?

You can get started for free at www.overleaf.com. Fully managed and self-hosted private cloud installations are available for institutional and enterprise customers.

12.13.2 User requirements

Overleaf's main overleaf.com service and managed private cloud services are turn-key solutions that require no user setup. If you opt for a self-hosted private cloud option, we can provide you with a set of AMIs and CloudFormation templates, and we will work with you to determine the best way to get you up and running in your Amazon VPC.

12.13.3 Free Trial

Overleaf provides free access to our authoring platform - https://www.overleaf.com/signup. Try it out! Write using your own documents or choose from our over 3,000 authoring templates, collaborate with other authors and colleagues via our sharable links or through our protected project option, integrate with other author services – such as reference management tools and graphics services, share your finished document with colleagues or via our multiple submission links into repositories and publishers.

If you would like to further test using group or team accounts, just let us know and we can set up a trial.

If you're interested in a private cloud option – you can easily start with a lower-cost 'starter' package.

We're happy to provide more information for your specific requirements – just reach out.

12.13.4 Billing

Billing is through a direct relationship with the vendor. Overleaf can provide quotes and billing in multiple currencies – USD, GBP, EUR & AUD.

12.13.5 Pricing & Licenses

You can find the latest pricing for accounts on overleaf.com at www.overleaf.com/plans. Accounts start at $8 / month, and a 30% discount is available for group licenses.

Managed Private Cloud licenses start at $2500 / month and are tiered based on soft limits for monthly active users and total storage space required. Please contact us for a quote.

Please contact us if you are interested in a self-hosted Private Cloud license.

12.13.6 Support

Overleaf provides expert LaTeX support for all customers on overleaf.com and private clouds (subject to your access control requirements). We also provide optional additional support and maintenance contracts for managed and self-hosted private clouds. The pricing structure for the support contract depends on the additional levels of user and technical support required.

Data and Analytics dashboards are available to administrators to easily find and export usage and collaboration trends for their users and authors.

12.13.7 Training and Reference Sources

Overleaf provides for free a robust help database, large amounts of training material, and both written and video tutorials at overleaf.com/help. This includes extensive amounts of information about the Overleaf platform, templates, and LaTeX itself. For new users to LaTeX, Overleaf provides a free interactive introduction to LaTeX course.

Additional Information from Overleaf

Overview

Overleaf is a scientific authorship tool that allows seamless collaboration and effortless manuscript submission, all underpinned by cloud-technology.

By providing an intuitive online collaborative writing and publishing platform, Overleaf is making the process of writing, editing and publishing scientific documents quicker and easier. It brings the scientific documentation process into a Google-docs like environment, developed to seamlessly connect the academic and publishing workflows from writing-to-review-to-publication.

By investing in Overleaf, your organization will be able to:

- Allow your users to benefit from our online, collaborative scientific writing and publishing platform with real-time preview.
- Ensure authors follow the correct format and guidelines by using custom templates specific to your organization.
- Support author collaboration in your community by providing an innovative writing tool that simplifies their writing process, expands and facilitates their collaboration ability – both online and offline, and streamlines their workflow process.
- Ease user collaboration by having both Rich Text and LaTeX writing modes available.
- Ensure the authoring, reviewing and editing experience is smooth, simple and painless, by having a trusted tool that will support your aims and goals.
- Gain insight into collaboration at your organization with Overleaf's analytics dashboard, which shows usage and collaboration trends for your authors.
- Outsource LaTeX support by having the Overleaf ‘TeX’perts handle all LaTeX and technical questions directly (subject to your access control).

Overleaf is:

- **User friendly:** Overleaf offers authors an easy way to write and collaborate on their scientific documents through a user-friendly interface that automatically typesets the paper, in real-time in the browser. Since 2012, more than 500,000 registered users from 180 countries worldwide have created in excess of 6 million documents using the Overleaf platform.
- **Effective:** projects are compiled automatically, with the source and output visible side-by-side. With real-time preview and error handling, Overleaf helps authors catch errors early, to create a smooth writing and submission process.
- **Efficient:** with Overleaf, you can say goodbye to long email chains, tedious reformatting and slow, costly manual conversions between storage formats.
- **Innovative:** Overleaf projects can be edited offline and kept in sync through our integration of Git version control.
- **Secure:** with Overleaf, industry-standard data security and persistence are provided and data are backed up continuously.

- **Forward thinking:** Developing authoring and collaboration tools for future needs and applications

The founders of Overleaf, John Hammersley and John Lees-Miller, are two mathematicians who worked together on the pioneering Ultra PRT Project (http://www.ultraglobalprt.com/) and who were inspired by their own experiences in academia to create a better solution for collaborative scientific writing. Passionate about science and making a difference in the world, John and John realized just how much time and energy was spent not only writing and reviewing papers but also in the vast amount of administrative and paperwork tasks associated with publishing. They set out to find a way to help save time on these tasks, with the goal to make the whole process of academic collaboration easier and more effective.

"A couple of days ago I needed to whip up a brief (5-6 page) technical report for a client. I knew it was going to be laden with equations, so LaTeX was my only choice. No problem… I just fired up Overleaf and cranked it out.

Thanks to Overleaf for putting together such a great service."
- Michael Grant, CVX Research

Overleaf Enterprise – Managed Private Cloud

Overleaf Enterprise is a private cloud option that allows organizations to have their own private servers while using the Overleaf platform. Overleaf will host and manage the servers – ensuring that it stays up to date with the latest Overleaf updates. Wide arrays of customizations are available on an Overleaf private cloud – Overleaf will work with you to determine what is required and the necessary steps to develop the system to fit with those requirements.

Overleaf Enterprise – Managed Private Cloud includes:

- **Editor Branding:** Your logo will be added to the platform and editing page.
- **Private Domain and SSL:** Private cloud set up on your domain, fully protected by SSL.
- **Secure project access control:** Only users authorized by the project owner can access the project.
- **Effective collaboration:** The project owner can add and remove users at any time – giving them access to 'read-only' or 'read-and-write' files.
- **Secure storage and backup:** For private cloud installations, industry-standard data security and persistence are provided. Data are backed up off-site daily.
- **Version History:** Full history of changes recorded for every project. History includes details of which user made the changes.
- **Record Milestones:** Specific versions can be saved manually, with a note to describe the milestone (e.g. first draft pre-review).

- **Company Templates:** Your template files are uploaded, and new projects can be created from your templates with a single click. We can help you set up company templates that include any necessary company information. When users start a new document, they can start from these company templates, ensuring that the necessary company information is included.
- **Redirect from main Overleaf site:** Ensure that your staff who go to the main Overleaf.com platform are aware that they can use your private cloud**.**
- **Real-time preview and error handling:** Projects are compiled automatically, with the source and output visible side-by-side. Overleaf shows you errors & warnings as you go, so you can catch them early; you don't have to find them in the LaTeX log.
- **Server maintenance and updates:** Overleaf will ensure that your private cloud server is running the latest Overleaf updates and that those updates do not interfere with any customizations added to your service.
- **Admin Dashboard:** Overleaf provides an administrative dashboard with metrics and analytics on platform use, users, and projects.

Overleaf Enterprise – Self-Hosted Private Cloud

Overleaf Enterprise's Self-Hosted Private Cloud option includes all of the same functionality as our managed private cloud, but you can install it on your own VPC or data center. We'll work with you to package Overleaf to meet the requirements of your operations staff. Training and documentation will be provided to allow them to scale your Overleaf installation to meet your needs and ensure that your security, persistence and compliance requirements are met.

Technology Partner	Product or Company **RONIN**	Home Page https://ronin.cloud/
	Vendor Country of Origin **Australia**	Delivery method **Supported Platform as a Service**
Domains	All research domains.	

Regional availability – All AWS Regions are sovereign.																
IRE	**DE**	**UK**	**FR**	**CA**	**US**	**US**	**US**	**US**	**BR**	**SG**	**JP**	**AU**	**KR**	**IN**	**CN**	**CN**
Dublin	**Frankfurt**	**London**	**Paris (2017)**	**Montreal**	**N. Virginia**	**N. California**	**Oregon**	**Ohio**	**Sao Paolo**	**Singapore**	**Tokyo**	**Sydney**	**Seoul**	**Mumbai**	**Beijing**	**Ningxia (2017)**

12.14RONIN

We developed the Ronin platform to lower the bar of entry to scientific and cloud computing by simplifying the way cloud resources are procured, configured and consumed. By providing complex IT services simplified for the research community, Ronin empowers researchers and research organisations to get on with the task at hand: research.

RONIN
CLOUD, SIMPLIFIED.
RESEARCH, REALISED.
< UNLEASH YOUR RESEARCHERS >

By developing an intuitive and simplified user interface, deeply integrated with a selected set of industry leading tools, we offer a powerful new method for consuming and managing cloud services for research organisations of any scale. RONIN puts your projects first. Compute and storage are only enablers for your next big idea so we put the focus back on the business of research.

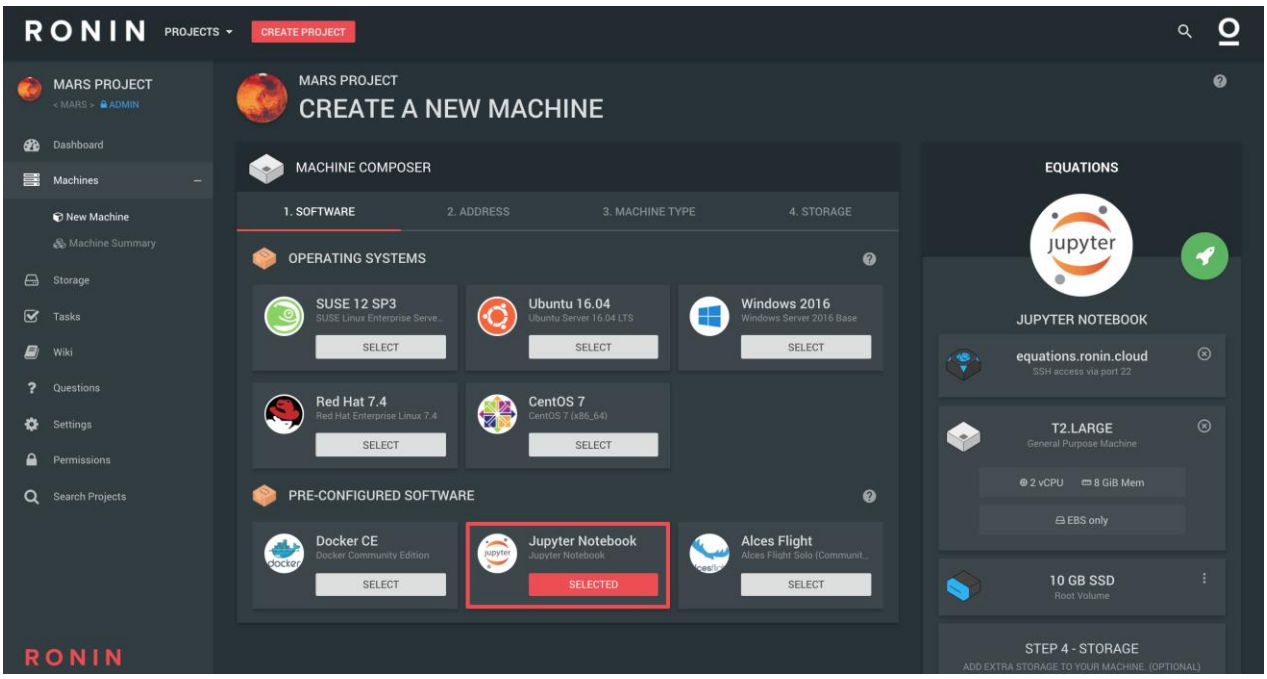
12.14.1 How is it accessed?

RONIN is an enterprise-level toolset deployed within your organisations AWS Virtual Private Cloud. Researchers access the platform from their web browser using their organisational credentials. RONIN provides a simple user-friendly interface to a number of AWS cloud services and third-party collaboration tools.

The Ronin platform also provides the ability to manage access for thousands of users within a single AWS account through integration with your organization's Active Directory.

12.14.2 User requirements

Our goal was to imbue researchers with all of the power available within the AWS platform without the usual entry requirement of becoming a part-time sysops/devops expert to get started. Once installed, the RONIN interface affords your researchers the ability to instantiate compute and storage as well as enterprise-strength project management and collaboration tools at the click of a button. Researchers are presented with a simple web user interface where they can manage their own projects. RONIN makes cloud easy.

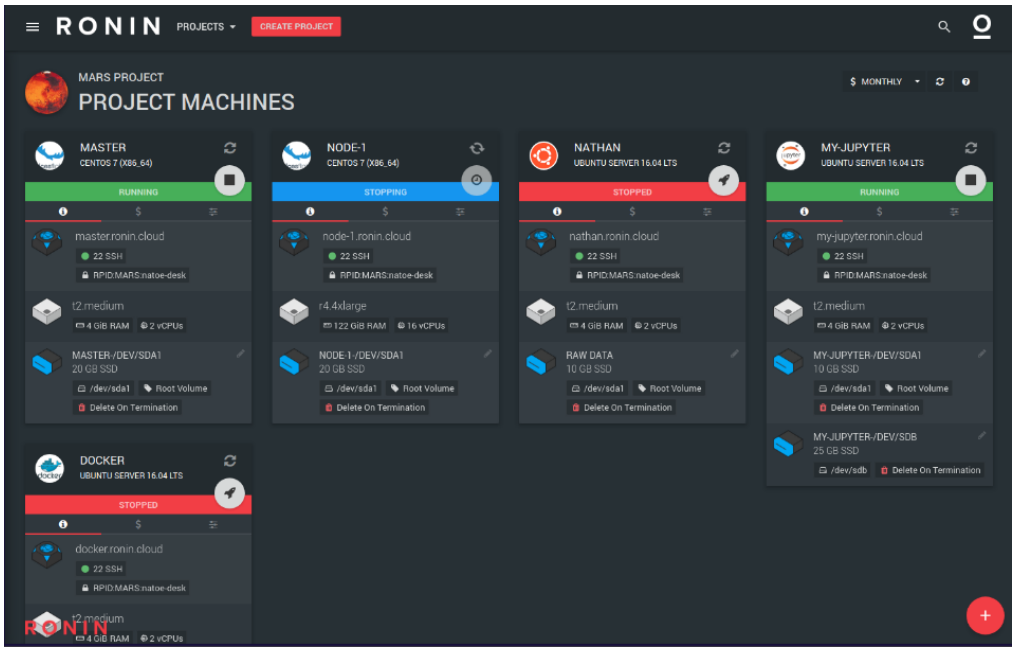


12.14.3 Trial

AWS enterprise level customers can request a free trial of the RONIN platform. There are a number of prerequisites which will be discussed with you prior to the installation of the trial platform. Functionality is not limited during the trial and scaling to a production licence is relatively straightforward. The RONIN team can integrate RONIN with your AWS account within a few days. Installation, integration and testing is included in the RONIN licensing model. Whilst the RONIN team takes care of the installation, there are some pre-install requirements that must be met by the organisation.

12.14.4 Billing

RONIN bills customers directly for the software. RONIN licencing is in addition to your AWS consumption charges. We do not add on a click-charge to your resource consumption, therefore you are not penalised for being successful in adopting the cloud. RONIN automatically tags compute and storage elements within your AWS account, providing greater transparency and accountability.



12.14.5 Pricing & Licenses

RONIN offers institutions an annual license fee for the software, based on the number of researchers accessing the software. Licensing is fixed to the number of active users you have, and does not scale with usage of AWS resources. We want you to use Ronin to scale up your research without the fear of scaling up license costs. In other words - you will not be penalized for successful use of our platform.

12.14.6 Support

RONIN licensing includes our enterprise-level support package. RONIN offers a customer support portal with FAQs, videos and explainer sheets on the software's capabilities. All support requests are responded to within 24 hours and usually resolved within this timeframe as well. Platinum support packages are also available, which include on-site support and additional support hours. RONIN comes with collaboration and support interfaces so that researchers can learn to support themselves and each other.

12.14.7 Training and Reference Sources

RONIN provides "getting started" webinars for new customers. Regular webinars are conducted when new features are added, or on request. On-site training can be purchased directly from RONIN.

Additional information from RONIN

Ronin is our delivery of an incredibly simplistic user-interface that facilitates the instantiation of complex compute resources leveraging the AWS cloud-compute platform. Of course, with such grand pursuits, it's critical we have the tools and controls in place to ensure we're not creating more problems than we solve.

This is why Ronin has enterprise grade management tools and best practices baked into all services and resources it makes available. We give organisations the ability to set and enforce both security and user policies across multiple cloud resources from a central console.

The Ronin platform also deploys detailed reporting, business intelligence and management information capabilities that go well beyond the tools provided by most cloud service providers. Through this information we give organisations the ability to derive otherwise invisible insights from their organisation and research disciplines to better optimise performance and costs.

Our previous experience showed that the stresses and concerns of scientists and researchers centered around technical hurdles and assumed knowledge of the tools available. This reality costs scientists and researchers serious time and money. Not only do they have the responsibility of expanding our understanding of the physical and natural world, they are often thrust into roles such as system administrator, network engineer and storage manager in order to enter the world of cloud computing and e-research.

Deploy your own HPC in minutes.

Through our integration of Alces Flight, RONIN provides easy access to scalable High Performance Computing (HPC) environment for research and scientific computing. This includes over 1,000 Linux applications, accelerated libraries and compilers.

Need deep learning? We have you covered. By adding the AWS Deep Learning AMI to the RONIN suite, researchers can launch feature rich deep learning environments without a single install script or line of code. Included in this package are popular deep learning frameworks, including MXNet, Caffe, Caffe2, TensorFlow, Theano, CNTK, Torch, Keras and more. It also includes Anaconda Data Science Platform for Python2 and Python3.

No Nerds Needed

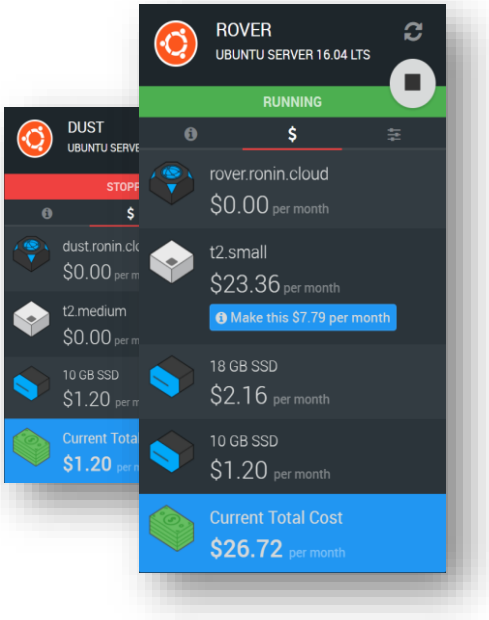
Technical hurdles. Assumed knowledge. These are the stresses faced by your researchers. RONIN removes the stress and enables access to vast amounts of compute and storage in seconds, not months. No IT degree required. No helpdesk calls. No more late nights reading about ports and mounts.

Never sleeps, never forgets

Are you human? Us too! One of the real drawbacks to being a human is we have the capacity to forget things, like turning things off when we don't need them. The RONIN Smart Schedule remembers for you so your budget doesn't run away while you aren't looking.

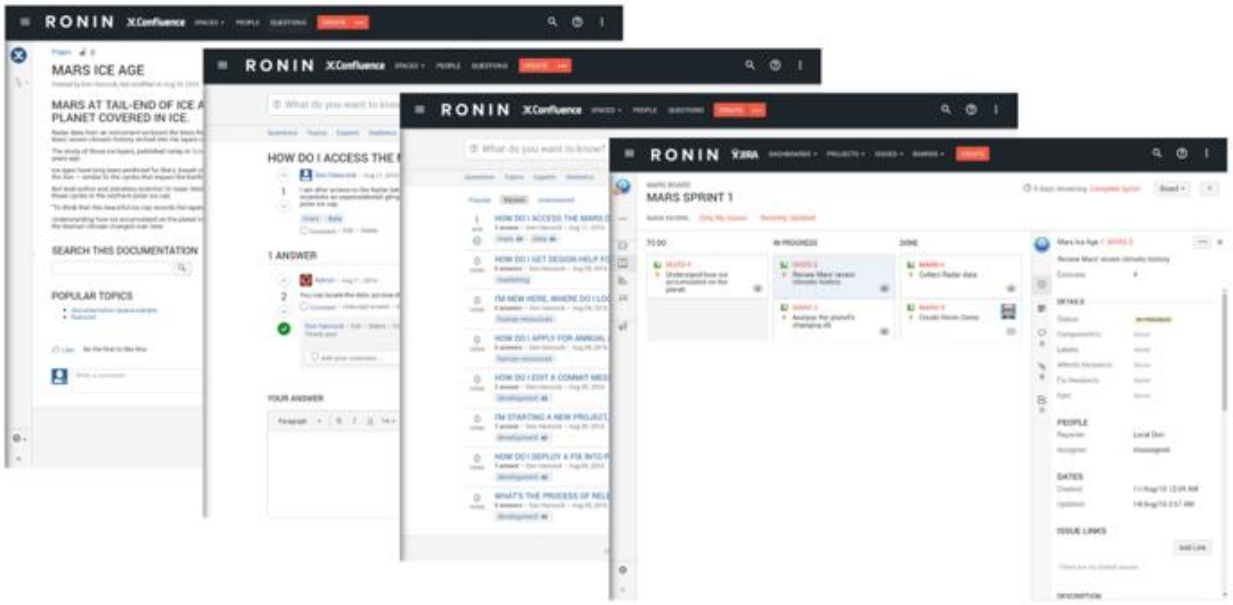
Cost Management and Visibility

When moving to the cloud we understand that cost management becomes... well... cloudy... We felt this was a problem worth meeting head on. With a single click, you can see a detailed breakdown of cost for individual machines or across your entire project



Collaborate With Ease

As part of all Ronin installations, we include an integrated Atlassian™ suite of tools. With the Atlassian™ suite at your fingertips you can manage a project's documentation, run sprints/manage tasks and build community support directly in the RONIN interface.



Technology Partner	Product or Company **Seven Bridges**	Home Page **http://www.sevenbridges.com/**
	Vendor Country of Origin **UK, Serbia, USA**	Delivery method ***SaaS Portal***
Domains	Genomic discovery, Next-Generation-Sequencing data Analysis and Informatics.	

Regional availability – All AWS Regions are sovereign.																
IRE	**DE**	**UK**	**FR**	**CA**	**US**	**US**	**US**	**US**	**BR**	**SG**	**JP**	**AU**	**KR**	**IN**	**CN**	**CN**
Dublin	Frankfurt	London	Paris (2017)	Montreal	N. Virginia	N. California	Oregon	Ohio	Sao Paolo	Singapore	Tokyo	Sydney	Seoul	Mumbai	Beijing	Ningxia (2016/17)

12.15 Seven Bridges

Seven Bridges is the biomedical data analysis company accelerating breakthroughs in genomics research for cancer, drug development and precision medicine. The scalable, cloud-based Seven Bridges Platform empowers rapid, collaborative analysis of millions of genomes in concert with other forms of biomedical data. Thousands of researchers in government, biotech, pharmaceutical and academic labs use Seven Bridges, including three of the largest genomics projects in the world: U.S. National Cancer Institute's Cancer Genomics Cloud pilot, the Million Veteran Program, and Genomics England's 100,000 Genomes Project. As the NIH's only commercial Trusted Partner, Seven Bridges authenticates and authorizes access to one of the world's largest cancer genomics dataset. Named one of the world's smartest companies by MIT Technology Review, Seven Bridges has the majority of staff in Europe and offices in Cambridge, Mass.; Belgrade; London; Istanbul; and San Francisco.

12.15.1 How is it accessed?

The Seven Bridges platform is accessed via Software-as-a-Service: www.sevenbridges.com where users can create a free account and access the world's largest genomic datasets, use the most popular bioinformatic tools (or add their own tools) and run analyses in real time on seamless computational infrastructure at their fingertips.

12.15.2 User requirements

The Seven Bridges platform is intuitive and easy to use. Users bring scientific curiosity — the biological questions driven by next-generation sequencing (NGS) — and Seven Bridges provides the best-in-class bioinformatic pipelines and industrial scalability. Each new account comes equipped with starter guides and tutorials as well as 24/7 support services from Seven Bridges. The users are required to provide an internet connection to access the platform.

12.15.3 Trial

Please contact Seven Bridges for a free trial at team@sevenbridges.com

12.15.4 Billing

Billing is in direct relationship with Seven Bridges.

12.15.5 Pricing & Licenses

Seven Bridges provides value in industrial-scale biomedical analyses. The platform is available as Software-as-a-Service package including, for enterprise level customers, a dedicated scientific project manager and genomics scientist to help obtain the most value from NGS data. The pricing tiers are tailored to user requirements ranging from small laboratory research to large national and multinational projects with millions of genomes.

12.15.6 Support

Seven Bridges provides 24/7 online support to the platform. The support features are included in the licensing price tiers. Additional professional service hours can also be obtained at enterprise tiers for specific feature development and support in use.

12.15.7 Training and Reference Sources

Seven Bridges provides onboarding training and continuous updates to enterprise-tier customers. Training can be delivered in-person (e.g. major R&D hubs) or online (e.g. distributed R&D collaborators or Horizon 2020 consortiums spread across Europe).

Additional Information from Seven Bridges

INDUSTRIAL SCALE BIOMEDICAL DATA ANALYSIS

Seven Bridges is the biomedical data analysis company specialising in national-scale genomics projects. We do so by empowering researchers with the tools for large-scale collaborative data analysis. Seven Bridges combines expertise in computing, bioinformatics, and population-scale genomics to enable faster discoveries and help support precision medicine.

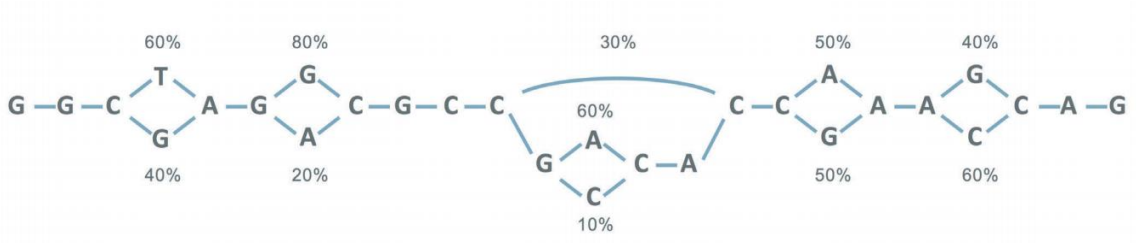


Figure 44 - Seven Bridges builds the functional graph-based suite of genomic tools. Graph population references contain the genetic variation of an entire population, and become more accurate with each new individual added.

The Seven Bridges Platform ("Platform") is a Platform-as-a-Service (PaaS) software system that enables **scalable** and **secure** bioinformatics research on public clouds and on-premise compute infrastructures. It acts as a central hub for teams to store, analyze and jointly interpret bioinformatics data. The Platform co-locates analytical tools and pipelines with petabyte-scale genomics datasets and manages compute resources on-demand to meet the diverse needs of the research community flexibly and efficiently. It also supports perfect **reproducibility** by way of the Common Workflow Language.

In addition, the Platform is home to the most advanced population genomics tools in the world: the graph genome suite. Graph-based genomic data structures represent a fundamental re-thinking of what population genetics — understanding what the variations between people are and what impact they have on health — means. They do so because they learn from every new person sequenced. Moreover, graphs contain information on an entire population instead of just one person. And, they allow analysis at previously impossible scales by offering significantly more compression than traditional, linear bioinformatics.

The combination of the Seven Bridges Platform and graph technology are powering the largest national projects in the world today.
These include:

• **Genomics England's** *100,000 Genomes Project*

• **U.S. National Cancer Institute's** *Cancer Genomics Cloud*

• **U.S. Department of Veterans Affairs'** *Million Veteran Program*

• **Children's Hospital of Philadelphia's** *Cavatica*, a key part of President Obama's Precision Medicine Initiative.

GRAPH GENOME SUITE

Seven Bridges is the biomedical data analysis company accelerating breakthroughs in genomics research for cancer,

drug development and precision medicine. We build self-improving systems to analyze millions of genomes, including

the Graph Genome Suite — the most advanced population genomics tools in the world. The Graph Genome Suite provides a clear value: better variant calling driving better discovery.

A Powerful Genomic Data Structure

The Graph Genome Suite is powered by a directed acyclic graph-based data structure which represents a fundamental rethinking of the genomic variations that impact health. Unlike standard linear references, this structure makes use of information from an entire population to characterize genetic variants with unprecedented accuracy. Our tools learn from every new person sequenced, meaning that the graph-based reference improves with each additional genome. Better still, this improvement happens with only minimal increases in file size, allowing analysis at a previously impossible scale.

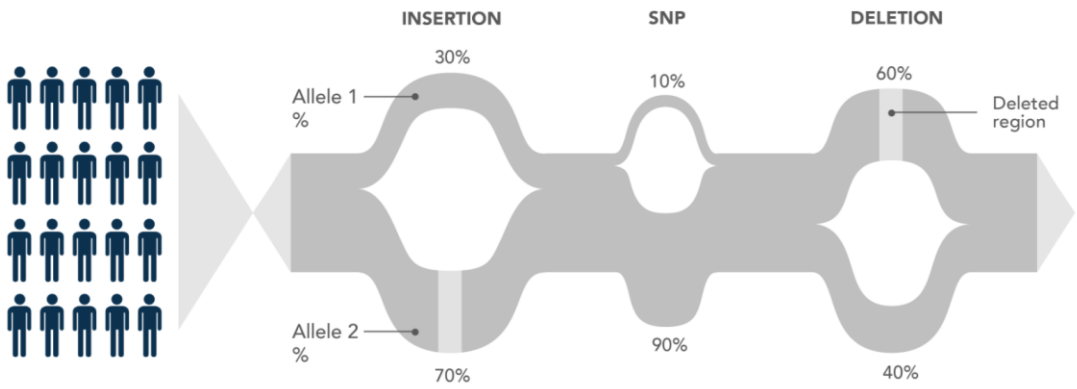


Figure 45 - Building a population graph. Unlike a standard linear reference genome, our tools construct reference graphs based on the genomes of a whole population.

Better variant calling for better discovery

The Graph Genome Suite has been developed in conjunction with Genomics England to support the UK Government's pioneering **100,000 Genomes Project**. It enables the most accurate variant calling in the world today, across all classes of variation including large structural variants that are difficult to identify using other aligners. These highly accurate variants enable the best possible basic and clinical genomics research in humans and other organisms, including personalized cancer analyses, family trio studies, and sublineage mapping of important viruses.

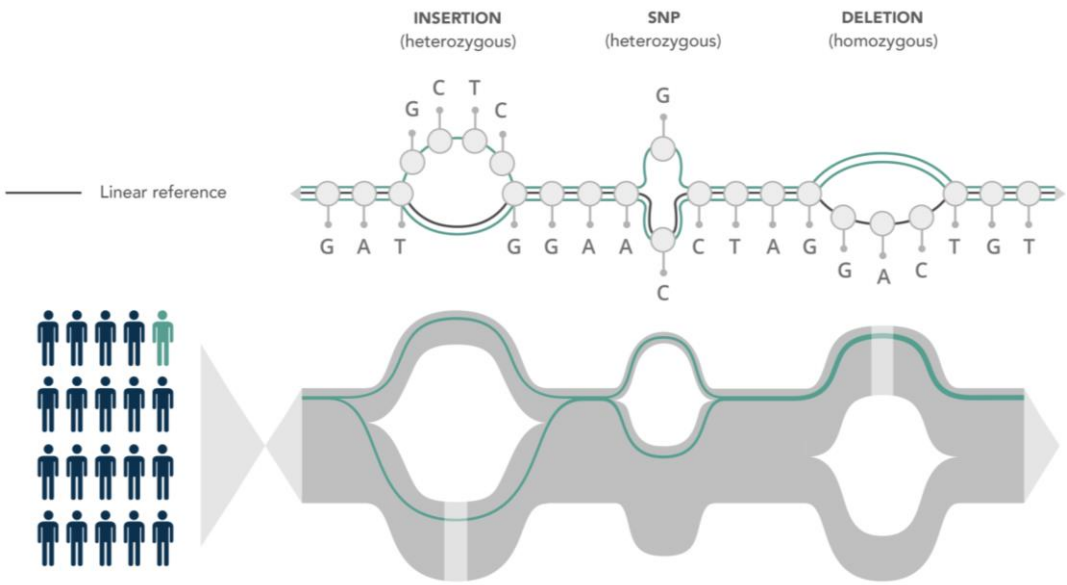


Figure 46 - Tracing an individual on a population graph. The genome of any individual can be represented as a path through the graph, which allows the identification of variants not captured by a linear reference, based on comparison against all the variants.

Driving innovative genomics research

Seven Bridges granted early access to the Graph Genome Suite to three of the world's leading research institutions: the US National Cancer Institute's Cancer Genomics Research Laboratory, Canada's Michael Smith Genome Sciences Center, and the Sidra Medical and Research Center of Qatar. Our scientists work with researchers at these institutions to deliver insights into infectious disease characterization, personalized cancer medicine, and human population genomics.

Seven Bridges Platform licensees can deploy our Graph Genome aligner as part of our whole genome analysis workflow and will have priority access to our graph-based variant caller, genotyper, and visualizations, which represent the most advanced population genomics tools available today.

Technology Partner	Product or Company **Sentinel Hub**	Home Page http://sentinel-hub.com/
	Vendor Country of Origin **Slovenia**	Delivery method ***SaaS Portal***
Domains	Earth Observation, Geography, Geospatial	

Regional availability – All AWS Regions are sovereign.

IRE	DE	UK	FR	CA	US	US	US	US	BR	SG	JP	AU	KR	IN	CN	CN
Dublin	Frankfurt	London	Paris (2017)	Montreal	N. Virginia	N. California	Oregon	Ohio (2016/17)	Sao Paolo	Singapore	Tokyo	Sydney	Seoul	Mumbai	Beijing	Ningxia (2016/17)

12.16 Sinergise - Sentinel Hub

Sentinel Hub is exploiting advanced AWS technology to process and distribute satellite imagery data (Sentinel, Landsat, etc.) and provide it to the user (or a machine) in an easy-to-integrate way - in form of OGC standard services such as WCS, WMS and WMTS. It removes the major hassle (and cost) of downloading, archiving and processing petabytes of data and simply makes the full and global archive accessible via web services.

Users and developers can integrate these services in their processing engine, web, mobile or desktop application, focusing their time on developing added value services.

12.16.1 How is it accessed?

SaaS – user creates an account and gets access to set of web-services and configuration utilities to access the data in a most appropriate way

12.16.2 User requirements

Users should be familiar with general characteristics of earth observation data (sensor types, resolution, indices, etc.) and basic OGC services.

12.16.3 Trial

One-month trial account is available upon request.

12.16.4 Billing

Direct relationship with vendor.

12.16.5 Pricing & Licenses

Sentinel Hub's pricing is subscription based, with a possibility of several models:

- consumer/research named user - monthly/annual/multi-annual subscription

- commercial named user - monthly/annual subscription
- enterprise volume based package (e.g. per 0.5 MIO monthly requests)

12.16.6 Support

Support services are included by e-mail with 24h response time.

Additional Information from Sinergise

Description

Sentinel Hub provides unprecedented access to earth observation data, focused to Sentinel and Landsat satellites but supporting also other sources such as Planet, RapidEye and more. It is facilitating advanced AWS Cloud technology and innovative methods to efficiently process and distribute data in a matter of seconds without compute intensive pre-processing, resulting in easy-to-use and cost efficient way to exploit the data in any GIS application or integrating them in web application. It therefore removes the major hassle of downloading, archiving and processing petabytes of data and simply makes the full and global archive available via web services. Application developers can therefore focus to added value services and end-user applications rather than dealing with complexity of remote sensing data.

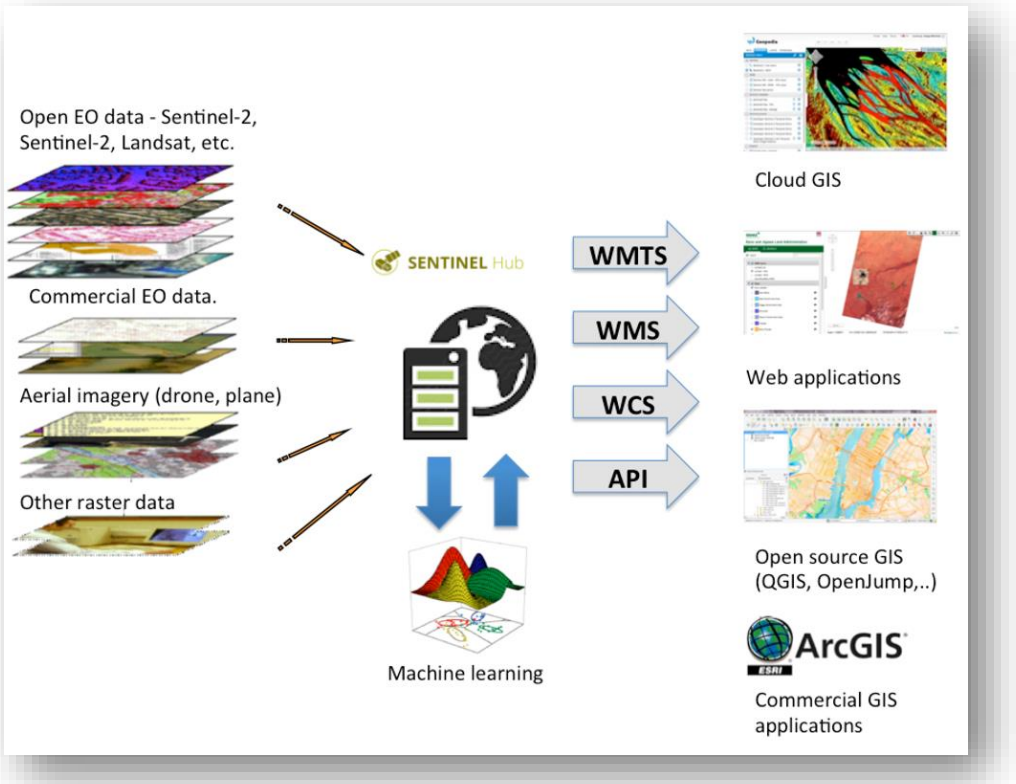


Figure 47 - Sentinel Hub Concept.

Supported services

- Web Mapping Service (WMS) for easy integration in proprietary GIS applications (ArcGIS, QGIS, etc.).
- Web Mapping Tiling Service (WMTS) and Tiled Map Services (TMS) in several projections and imagery options for integration in web and mobile applications (e.g. OpenLayers frameworks and similar).
- Web Coverage Service (WCS) for processing workflows, where users want to control each aspect of data distribution.
- Index service to provide meta-data about available imagery in some specific area.

Customer benefits

- No need to keep and maintain an archive of the satellite imagery data.
- No need to process imagery using classic, compute-intensive and time-consuming processes.
- Flexibility - change configuration of data processing in a matter of seconds with no reprocessing time.
- Easy to use - no need to dive into technical details of each sensor, satellite providers' distribution methods, meta-data, etc.

Features

- Configuration utility for easy set-up of the services.
- Web application to easily configure the service for maximum integration capabilities.
- On-the-fly band math providing option to calculate and visualize indices.
- Advanced filters such as contrast stretching, DOS1, atmospheric correction and others.
- Different output options - PNG, JPG, GeoTiff, raw values.
- Upsamling and downsampling options to fine-tune results on different scales.
- Open-source libraries available on GitHub with integration examples.

Test it yourself

- Try showcase web-application Sentinel Playground (http://apps.sentinel-hub.com/sentinel-playground/)

- Send us an e-mail to info@sinergise.com with a request for trial.

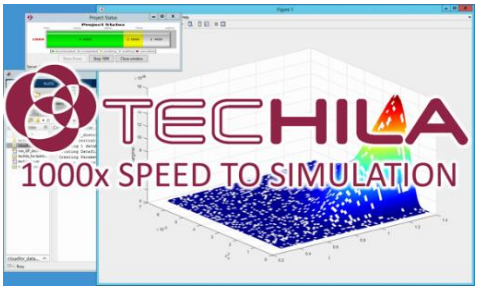
Technology Partner	Product or Company **Techila Distributed Computing Engine**	Home Page http://www.techilatechnologies.com/
	Vendor Country of Origin **Finland**	Delivery method **Direct Licensing**
Domains	MATLAB, Simulation, Optimization, Modeling: Financial Engineering, Applied Mathematics, Materials Sciences, Bioinformatics, Stochastic Modeling, Monte Carlo simulation, Machine Learning, Bayesian Modeling, Statistics, …. (more information available [http://www.techilatechnologies.com/fast-matlab-on-demand/]).	

Regional availability – All AWS Regions are sovereign.

IRE	DE	UK	FR	CA	US	US	US	US	BR	SG	JP	AU	KR	IN	CN	CN
Dublin	Frankfurt	London	Paris (2017)	Montreal	N. Virginia	N. California	Oregon	Ohio	Sao Paolo	Singapore	Tokyo	Sydney	Seoul	Mumbai	Beijing	Ningxia (2016/17)

12.17 Techila Distributed Computing Engine

Techila Distributed Computing Engine enables 1000x faster simulation, 180x easier than traditional high-performance computing.



Watch MATLAB® demo on YouTube:
http://www.techilatechnologies.com/matlab-demo

12.17.1 How is it accessed?

Self-launch on-demand. A couple of mouse clicks on Techila's fully automated tools will start the Techila Distributed Computing Engine on your own AWS account.

12.17.2 User requirements

Techila Distributed Computing Engine enables fast simulation and analysis, without the complexity of traditional high-performance computing. Using the Techila Distributed Computing Engine with AWS Cloud does not require more skills than installing a new MATLAB toolbox.

All you need is your client MATLAB and MATLAB Compiler on your PC. Techila Distributed Computing Engine will do the rest. Techila Distributed Computing Engine is a full-stack solution that includes everything you need for fast and scalable simulations in the AWS Cloud.

Techila Distributed Computing Engine license includes fully automated tools that enable starting and stopping of the computing environment on demand on your own AWS account, with a just couple of mouse clicks. Video tutorials and getting started guides walk you step-by-step through the installation and running your first simulation in the AWS Cloud.

12.17.3 Billing

Techila Distributed Computing Engine can be used with an enterprise license, that you can order directly from Techila Technologies and pay by invoice.

12.17.4 Pricing & Licenses

The price of Techila Distributed Computing Engine enterprise license is 20EUR/ CPU core/ month. The license term is 12 months.

The cost of Techila Distributed Computing Engine depends on how much computing throughput you would like to have. 100x, 500x, 1000x, 2000x,…

A Techila Distributed Computing Engine enterprise license can support many users. If there are several users running simulations at the same time, the resource manager of the Techila Distributed Computing Engine will share the available throughput between based on customizable user priorities.

12.17.5 Support

Techila Technologies offers a variety of free and paid support services, including Premium support and Consultancy services. For more information and contact information, please visit the Techila Technologies Support Site: http://www.techilatechnologies.com/support/

12.17.6 Training and Reference Sources

Techila Technologies offers a rich range of technical documentation, and a variety of free and paid training services.

The documentation is free to all users of the Techila Distributed Computing Engine at Techila Documentation Home. http://www.techilatechnologies.com/help/techila-distributed-computing-engine

Free trainings are available as recorded courses on Techila Technologies' YouTube Channel. https://www.youtube.com/TechilaTech

Paid trainings can have tailored content that meets the specific requirements of your organization. Paid trainings can be organized at the location that is most convenient for you. For more information and contact information, please visit the Techila Technologies Support Site: http://www.techilatechnologies.com/support

Additional Information from Techila

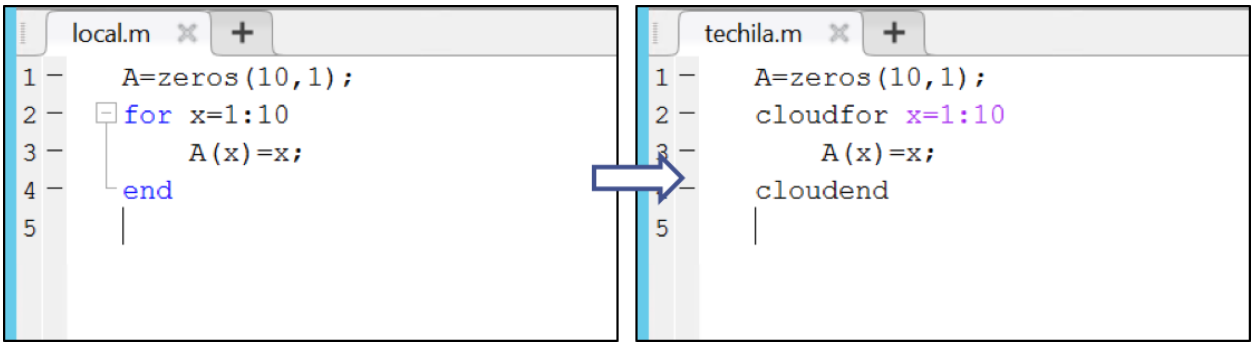
Techila Distributed Computing Engine enables fast simulation and analysis, without the complexity of traditional high-performance computing. No need to wait for computations to finish. No need to cut corners in accuracy.

Techila APIs integrate scalable computing power seamlessly into MATLAB and other popular research and development tools. The applications can also be deployed and run as service in a SOA environment.

Modern research relies on advanced simulation, modeling, and optimization. The computational tasks can be time-consuming. Monte Carlo simulations, parametric sweeps, calibration of models can take hours or weeks. This does not support productivity of the research or scientific innovation.

In many cases, solving the problem can require run an operation with different parameters or data, or performing a complex algorithm operation that is implemented using for-loops. The Techila toolbox for MATLAB comes with cloudfor function that can be implemented easily, and can cut the run time from weeks to hours and from hours to minutes. You can simply replace the local for command simply with the cloudfor command. The cloudfor offloads the processing automatically to the resources on your own AWS account.

How to change a code with a local for to use the cloudfor:



For more information about using TDCE with MATLAB and code examples please visit the Techila Distributed Computing Engine with MATLAB user guide:

http://www.techilatechnologies.com/help/techila-distributed-computing-engine/matlab-techila-distributed-computing-engine.html

Technology Partner	Product or Company **EPIC**	Home Page http://epic.zenotech.com/
	Vendor Country of Origin **UK**	Delivery method **SaaS Portal**
Domains	High Performance Computing, Fluid Dynamics, Supercomputing	

Regional availability – All AWS Regions are sovereign.

IRE	DE	UK	FR	CA	US	US	US	US	BR	SG	JP	AU	KR	IN	CN	CN
Dublin	Frankfurt	London	Paris (2017)	Montreal	N. Virginia	N. California	Oregon	Ohio	Sao Paolo	Singapore	Tokyo	Sydney	Seoul	Mumbai	Beijing	Ningxia (2017)

12.18 Zenotech

EPIC is a portal to cloud HPC. It allows users to easily start and manage cloud based HPC clusters and services. It also provides a simple submission interface for several popular simulation tools to allow EPIC users to submit HPC jobs to a variety of resources.

12.18.1 How is it accessed?

EPIC is accessed via a portal. Users can register at epic.zenotech.com.

12.18.2 User requirements

We have tried to make using EPIC as simple as possible, no HPC expertise is required to start using applications at scale via EPIC. We do all of the configuration for you. If you start a cluster via EPIC you will have a fully functional HPC cluster in a few clicks. We have simple billing and account management views to allow you to control your HPC spend and share resources with your colleagues.

12.18.3 Trial

All new users to EPIC get £10 of trial credit to spend trying out the service.

12.18.4 Billing

All of the billing is done via the EPIC portal. You have an EPIC account and all the transactions are done via this account. No need for additional AWS accounts or accounts on the 3rd party HPC services. All billing is pay-per-use and there are no subscription charges.

12.18.5 Pricing & Licenses

There are no license costs for using EPIC itself. There may be costs for the software used via EPIC and if so we support a bring your own license model or, where supported by the vendor, a pay-per-use model.

The latest pricing of HPC resources used via EPIC is available on the EPIC home page (epic.zenotech.com). We try to be as transparent as possible – you see the price we get charged by the provider and then an EPIC service charge is added to this cost.

12.18.6 Support

There is a help section and knowledge base within the portal. The support team can be emailed directly at support@zenotech.com.

12.18.7 Training and Reference Sources

When you first log in to the portal you will be presented with a getting started tutorial. If you have any further questions please contact support.

Additional information provided by Zenotech:

EPIC aims to be your gateway to accessing a range of HPC resources.
EPIC was developed to meet our own requirement for computing resource but we frequently came across people with the same issues as us. We commonly heard comments such as...

... We need access to more resources than we have ...
... We could do so much more with a bigger cluster ...
... I could really do with my own cluster for the weekend ...

And so we continued to develop EPIC to meet this need. If you just want to access more resources for a specific application, then EPIC has the **Job** interface - a simple way to your jobs running on large scale HPC systems with just a few clicks. But if you need a bit more control and want your own private **Cluster** then EPIC will allow you to start and manage your own HPC cluster in the cloud.

Two ways to work via the same platform.

You can submit jobs or launch private clusters from within EPIC with the benefit of a single billing and team management interface.

Jobs

If you are running an application on an internal resource and just want to get your job running on a bigger machine, then EPIC jobs may be for you! We present a simple interface to a number of applications that let you submit your job onto a range of backend HPC clusters. You just provide your data and the options for the application and we take care of everything else.

- Simple submission interface
- Submit jobs onto dedicated HPC hardware
- Optimised applications
- Job monitoring and notification

Clusters

If you need your own HPC cluster for a period of time then EPIC can help you get one up and running. Choose a hardware configuration to match your requirements and then EPIC will start a cloud based HPC cluster for you and provide you SSH access. Monitor and manage your HPC cluster via EPIC. We take care of configuring the cluster to get the best HPC performance from the cloud.

- Simple cluster configuration
- Direct SSH or VNC access
- HPC ready
- Range of pre-installed applications

For more information please see epic.zenotech.com.

13 APN Consulting Partners Listing

On the following pages, we introduce some of our APN Consulting Partners, selected for what they offer to the global Research Community.

These partners provide the skills necessary to stand up some of the most serious workloads possible. They've all worked in scientific research environments and have deep technical abilities that you can trust to solve the problems you need help to address.

For a full listing of APN Consulting Partners – specific to skill set or geography, you can always consult https://aws.amazon.com/partners/.

Contents

Consulting Partner	Company Name **Acellera Ltd**	Home Page https://www.acellera.com/
Registered	Company HQ City & Country **Stanmore, UK**	Services offered in Europe, North America
Domains and/or Technologies	A high-technology company focused on developing high-throughput molecular dynamics techniques based on GPU, cluster of GPU and Cloud. We deliver solutions for estimating common physico-chemical properties such as binding affinities, kinetics, poses and pathways with validated accuracy.	

13.1 Acellera Ltd

Acellera's mission is to accelerate the use of simulations and computational methods in drug discovery.

Acellera's ACEMD platform lowers the barrier to perform complex statistical analysis of conformational change and binding event between biomolecules. These simulations offer atomistic resolution for the understanding of macroscopic event. They represent a powerful orthogonal technique with most common biophysics approaches such as nuclear magnetic resonance, crystallography, surface plasmon resonance and isothermal calorimetry.

Acellera proposes product (software and hardware) and services to worldwide computational chemistry, medicinal chemistry and biophysics groups in pharmaceutical industry, public research institutions and academy.

13.1.1 Contact Info

Commercial and technical Support are provided through several channels. Customers can contact us:

Phone: +34 674 210 151

Email: info@acellera.com

13.1.2 Competencies

Software, hardware and services for early phase of drug discovery and drug design based on molecular dynamics simulations

13.1.3 Office Location(s)

Headquarters: Acellera Ltd, Devonshire House, 582 Honeypot Lane, Stanmore, HA7 1JS, UK

Subsidiary: Acellera Labs SL, PRBB, C/ Dr. Aiguader 88, 08003, Barcelona, Spain

13.1.4 Geographic Coverage

Acellera can operate worldwide: several projects have been led in Asia, South America and Oceania.

Preferred geographic areas are **Europe and North America**.

13.1.5 Solutions & Services offered

Acellera is a high-technology company focused on developing high-throughput molecular dynamics techniques, including software, hardware protocols and services. We deliver solutions for estimating common physico-chemical properties such as binding affinities, kinetics, poses and pathways with validated accuracy. These techniques are usually proposed but not restricted to life sciences companies including pharmaceuticals and biotechnology industry but also to governmental institutions and academy.

We mainly propose to our customers:

- Customized software application development
- Performing strategic R&D project based on molecular dynamics
- Training to high-throughput molecular dynamics and their analysis

13.1.6 Certifications

Acellera has published more than 27 peer reviewed articles over the last nine years. Our work has been presented at numerous international congresses, symposium and meetings.

Additional Information from Acellera

Acellera started to develop its own code for accelerated molecular dynamics (MD) simulations ten years ago. This code has been constantly improved and adapted to the most recent technology available.

It is now worldwide distributed and used. Our team, composed of mathematician, statistician, physicist, chemist and computational chemist, has designed an exclusive platform, called ACEMD platform that enables research teams to access fast GPU based simulations and automated trajectories analysis. The achievement of long production run allowed the application of statistical analysis such as Markov model. These models are widely used in other important domain like economy, genetics and chemistry. They make possible complex analysis of the data produced. For MD trajectories simulations, they allow to determine common physico-chemical properties such as binding affinities, kinetics, poses and pathways with validated accuracy.

During the last decade, Acellera has participated to several European funded project and collaborated with major pharmaceutical companies in early drug discovery programs and drug design projects. We also have supported numerous computational, medicinal and biophysics groups to implement, use and develop GPU based simulations techniques. These interactions with the scientific community allowed us to face and overcome major technical and theoretical challenges. We have acquired a deep knowledge about fundamental aspect of molecular dynamics analysis and its outcome.

The evolution of the technology, and especially the advent of GPU and Cloud computing, allows now to simulate on the same time scale as experimental one. This achievement triggered the interest for the possibilities offered by molecular dynamics: they can now be used not only for validation of experimental data such as crystallographic pose, binding

kinetics or affinity. Since 2009, Acellera research and development (R&D) group has optimized technique and protocols and demonstrated the robustness of molecular dynamics simulations and the accuracy of their analysis. More than 27 peer reviewed articles have been published based on our own R&D and within the framework of some collaborations. Our work has been presented at numerous congresses, conferences and meetings.

This work has been possible thanks to the diverse and high experience of our team in scripting, programming and automating the process of MD simulations. It includes the development of specific code for GPU, the creation of a specific software designed to prepare, run and analyze MD simulations as well as to handle efficiently the data generated during all the MD process.

Besides the knowledge about molecular dynamics, we gained a unique know-how in implementing MD simulations solutions, performing their analysis and adapting their installation to our customer facilities.

Convinced about the usefulness of MD simulations, we wanted to render this technique accessible for all and started three years ago to offer free training to the scientific community. A dedicated workshop on High Throughput molecular dynamics has been organized in Barcelona and proposed to the attendees theoretical and hands on sessions. We propose the same training to our customer, focusing on their needs and giving unique and personalized solution.

Acellera has developed unique tools and protocols that can be implemented rapidly into small entity or large facilities. Our knowledge in implementing conformational study and biomolecules interaction analysis is unique and Acellera is a strategic partner for computational based drug discovery program.

Consulting Partner	Company Name **Alces Software Ltd.**	Home Page http://alces-software.com/
Registered	Company HQ City & Country **United Kingdom**	Services offered in Europe, Africa, India (English Language)
Domains and/or Technologies	Benchmarks, Biochemistry, Bioinformatics, Bio-physics, Chemistry, Compilers, Databases, Electronics, Engineering, Geography, Graphics and Imaging, Languages, Libraries, Mathematics, Medicine, MPIs, Physics, Statistics, Tools, Visualization. Specialising in HPC and highly scalable cluster workloads.	

13.2 Alces Software Ltd.

Alces Software is a team of individuals with over fifteen years' experience in putting together High Performance Computing (HPC) clusters on a wide range of different platforms. If you are looking to put together a custom configuration for running HPC in the cloud, or are looking to pull together a hybrid solution that uses your on-premises cluster with AWS, look no further than us.

The Alces Team are the creators of Gridware, putting over 1,300 applications and libraries at your fingertips – with more additions every day. This means that Alces is uniquely placed to not only assist with constructing your HPC cluster, but also to ensure that your cloud or hybrid solution is easily optimized for your end user ensuring your investment is a continued success.

13.2.1 Contact Info

General information and sales inquiries can be made at info@alces-software.com or +44 (0) 1869 249 065.

13.2.2 Competencies

We have extensive HPC knowledge and experience, including:

- High-throughput CPU technologies
- Large memory servers
- Remote visualization technologies
- GPGPU and special-purpose devices
- High performance, low-latency interconnects and tuning
- Visualization, 3D-rendering and offload technologies
- Virtualized and containerized solutions to maximize portability and security
- Parallel filesystems (Lustre, GPFS, Gluster, pNFS, BeeGFS)
- High-availability services and no-single-point-of-failure designs
- Scientific software domains (e.g. Life Sciences, CFD, Engineering, Physics, Mathematics)

13.2.3 Office Location(s)

Alces Software has offices in Bicester, England with a distributed development team available throughout the UK and Northern Europe.

13.2.4 Geographic Coverage

Alces Software can support customers in Europe, India and Africa (UK time zone +/- 3 hours). Services and/or documentation deliverables are provided in the English language. Alces can also offer global next-business-day coverage for our Alces Flight software for end users seeking to run HPC clusters in the cloud.

13.2.5 Solutions & Services offered

Alces Software provides consulting and custom application development for those looking to put together HPC clusters in the cloud or within a hybrid environment. We are fully engrained with on-premises, hybrid, and AWS environments, having designed hundreds of solutions that suit the needs of our client. We are engaged heavily with the open-source environment of HPC as well as work with commercial applications in order to pull together a system that meets the goals of your company or institution. Why not contact us for help with:

- Assessments / Solution Build Requirements
- Custom Application Development
- Managed Service Provider
- Strategic / IT Consulting

13.2.6 Certifications

Our support and development team includes contributors with a wide-range of certifications including:

- AWS Architect and Security
- Hardware system certifications from major tier-1 suppliers
- Network and InfiniBand solutions architecture
- Lustre Parallel file system Elite certified professionals
- PRINCE2 project management certification

Additional information from Alces Software

As the world creates more Linux applications and distributions the variety of system and software configurations increases. Our software repositories, web-accessible interfaces and command-line tool suites provide an intuitive, consistent and predictable environment, removing unnecessary complexity and enabling you to get the important work done. We favor convention over configuration - getting you going with your cluster from day one. With a multitude of Linux distributions, requirements for different compilers, and multiple versions of libraries, message passing interfaces and software applications, it can be a daunting prospect to get up and running in your high performance environment.

During our years of experience in configuring and using HPC clusters it has become apparent that some order needed to be brought to the chaos. Configuring all the

component parts of a cluster is just the first step – making use of the cluster is where the important stuff happens.

By following consistent conventions for compiling, installing and using HPC software components, we are able to provide you with an environment ready for getting your job done from day one. And when you want to add new software libraries and applications at a later date we've got you covered with our tool suites and package repository.

Get comfortable

Wrestling with an unfamiliar interface can only reduce your efficiency - there is a better way. While some people are entirely at home with a command-line interface we know that, sometimes, getting familiar with a high technology environment can be a daunting prospect. With an easy-to-use web-accessible interface for your cluster we mitigate the need to learn how to use the command-line. Transferring data between your workstation and your cluster, creating job submission scripts, submitting your HPC jobs and monitoring their progress can all be managed from the comfort of your desktop browser.

Alces Gridware provides an ever-expanding repository of tools, utilities, libraries and applications, configured for the hardware in your cluster. Easy-to-use and convention-focused, Gridware eases the installation and operation of the wide range of available HPC applications and environments.

Harness the power while keeping a lid on your costs

High performance computing often leverages specialized technologies long before they are accepted into mainstream IT. Deploying such solutions requires knowledgeable engineers to achieve the best performance and advanced software to deliver optimum stability and usability for your customers. We maintain a high level of expertise in key technologies including:

- Scientific software packages and applications
- System availability and performance management solutions
- Web services software and remote access systems
- Batch scheduler configuration and optimization
- High-availability hardware and software solutions
- Linux operating systems and package management
- High performance parallel filesystem technologies
- Fault-tolerant and business continuity solutions
- Complex project management

Support how you want it, when you need it.

We offer customizable support agreements to suit your requirements. Our support agreements can include 2-day, next-business day, 4-hour or 24/7 response to support tickets for critical issues and named primary engineering contacts for continuity of support requests.

Consulting Partner	Company Name **Arcus Global**	Home Page http://www.arcusglobal.com
Advanced	Company HQ City & Country **Cambridge, UK**	Services offered in UK, mainland Europe
Domains and/or Technologies	Management Portal for education, architecture, security, managed services (more information available [http://www.arcusglobal.com/aws-infrastructure-design/).	

13.3 Arcus Global

We started implementing AWS for public sector customers in 2009 and provided a portal that allows higher education institutions to easily buy and track AWS services in 2013.

We are experts in solving difficult problems and implementing cost effective, secure, high performance solutions.

13.3.1 Contact Info

Please contact us via sales@arcusglobal.com or +44 (0)1223 911 841

13.3.2 Competencies

Arcus Provides an easy way to procure AWS services via our educational management portal (built in conjunction with JISC). It provides the following benefits:

- Buy without a credit card – we invoice your institution using 30 day terms
- Currency exchange – Arcus invoices you in GBP/Euro
- Aggregation benefits – pooled usage reduces costs for all
- Data egress waiver – we provide automatic enrollment (https://aws.amazon.com/blogs/publicsector/aws-offers-data-egress-discount-to-researchers/)

We were the first AWS Public Sector Partner in the UK, we have migrated local authorities to the cloud and enabled central government to host sensitive databases. We can help you get started with AWS or enable you to utilize big data technologies including Redshift and EMR (managed Hadoop) for analyzing large datasets.

13.3.3 Office Location(s)

Arcus Global, Future Business Centre, Cambridge, CB4 2HY, UK

13.3.4 Geographic Coverage

We are a UK business that has been awarded a framework agreement in the pan-European GÉANT IaaS tender, which will enable us to offer our higher education portal services across Europe.

13.3.5 Solutions & Services offered

Our management portal makes it easy for HEIs to buy AWS services.

Our consultancy can help you design, build and secure your AWS infrastructure.

Our managed services deliver ongoing monitoring and support.

More information can be found here http://www.arcusglobal.com/aws-infrastructure-design/

Consulting Partner	Company Name **inQdo B.V.**	Home Page https://www.inqdo.com
Advanced	Company HQ City & Country **Utrecht, Netherlands**	Services offered in Netherlands, Belgium, Germany, Sweden, Finland, Norway, Denmark, Iceland, Spain, Italy, Switzerland, Monaco, Austria, France & UK
Domains and/or Technologies	Chemistry, Physics, Astronomy, Mathematics, Geography, Geospatial, Graphics, Machine Learning, Bioinformatics, Biochem, Statistics, …. (more information available https://www.inqdo.com). HPC & highly scalable cluster workloads – consulting to the stars.	

13.4 inQdo B.V.

Is it the first time your organization wants to explore Amazon Web Services (AWS) and do you need a fast and flexible implementation and no up-front investments?

inQdo can help your organization with:

- Making the right choices on AWS architecture
- Implementing your AWS environment
- Optimizing your running AWS environment
- Managing your AWS account
- Advanced billing (no credit card needed)
- Proof of concepts
- ROI calculations

Our experience is bundled in best practices. That's why we can quickly select the right components for your AWS environment.

13.4.1 Contact Info

Call Peter Perebooms at +31 64 53 44 046 or email at peter.perebooms@inqdo.com for more information.

13.4.2 Competencies

- **InQdo Connect** - our iPaas Enterprise Integration Platform
- **Effective exchange of data** - Ensures the exchange of data within your organization and with information systems and individuals outside your organization. From creating connections through to ensuring the secured access of apps and mobile websites.
- **Flexibility** - Dynamic scaling to respond to peaks in usage.
- **No surprises** - Adaptors and all technical connections to your applications or communication channels like mobile and portals included.

13.4.3 Office Location(s)

inQdo B.V. (Netherlands)

Everard Meijsterlaan 1C
3533 CK Utrecht, The Netherlands

inQdo B.V. (Manilla)

One World Place
32nd St. Bonifacio Global City, Manilla

13.4.4 Geographic Coverage

inQdo has two office locations (Netherlands and Manilla) from those locations we cover the following countries: Netherlands, Belgium, Germany, Sweden, Finland, Norway, Denmark, Iceland, Spain, Italy, Switzerland, Monaco, Austria, France & UK.

13.4.5 Solutions & Services offered

inQdo offers two primary services for AWS. On the one hand, we are an AWS consultancy company that offer the following services:

- Making the right choices on AWS architecture (Assessments)
- Implementing your AWS environment
- Optimizing your running AWS environment
- Managing your AWS account
- Advanced billing (no credit card needed)
- Proof of concepts
- ROI calculations
- Cloud Migration Services
- Managed Service Provider
- Strategic / IT Consulting

On the other hand, we have built a specific iPaaS solution that can be used in in the AWS science cloud called inQdo Connect (https://www.inqdo.com/connect). inQdo Connect can be used in various scenarios.

13.4.6 Certifications

ISAE 3402

Additional Information from InQdo

inQdo Connect - iPaaS
Definition of the problem

Organizations are being confronted with more and more systems which have to exchange data. This is an ongoing process, especially when there is a multiplicity of channels (including apps). The complexity of IT is increasing, while the security of supply must still be guaranteed. Because alterations are built on alterations, spaghetti-styled links have been created over the years. New requirements and insights bring along new IT challenges. On top of that it turns out that the specialist knowledge needed to comply with these dynamics is becoming increasingly scarce. Nor are the budgets required becoming more generous.

Solutions for integration challenges

Nowadays, business processes are supported by ever more applications. Applications are frequently adjusted, replaced or added, but at the same time they share information which is used in processes that transcend the applications themselves. In order to optimize these processing times, the demand for 'real-time' information exchange between applications is increasing. Furthermore, more and more - constituent - processes are supported by information technology. At the same time the differences in the life cycles of the applications are increasing: some applications are specially made for 1 event and other applications last for 5 years without significant changes being necessary. Because of this there is a growing demand for flexible, reliable and controllable integration between applications. An Enterprise Service Bus (ESB) could contribute to this.

ESB

An ESB is an architectural software construction that adds an abstraction layer to the communication between the applications of applicants and providers of information. The ESB offers an interface at the applicant's end, for example a web service or a file at a certain location. At the provider's end the ESB will also communicate via an agreed interface. In the first figure there are direct 'point-to-point' connections between the applications of the applicants and the providers. In the second figure, an ESB is placed between the applications.

The ESB is able to communicate with the application of the applicant in a completely different way than with the provider's application. Moreover, these applications no longer need knowledge of each other to transfer information. Adding the ESB within the software architecture makes it possible to disconnect the protocol and the format. Both the applications of the applicant and those of the provider communicate with the ESB. The ESB transforms the incoming information in a request into the protocol and format expected by the provider. An ESB is the perfect tool to launch a canonical data model. A canonical data model is an intermediate format that is not bound to an application-specific format. Furthermore, this disconnection enables standardization, and centralization of monitoring and control.

The reuse of existing information services (interfaces) makes it possible to introduce new/modified applications with minimal impact to the rest of the application landscape. This does indeed require an adjustment to the ESB, but the ESB is able to efficiently process modifications.

Summarizing, an ESB offers a central switch point where information is converted into the suitable format(s) for the applicant(s) and communicated in an appropriate way even if the provider sends the information in a different way. An ESB has the following functionalities: protocol conversion (adapters); temporary storage (buffering); extensive transformation possibilities; orchestration/composition to eliminate differences in granularity between interfaces; logging for the purpose of monitoring; security and routing.

inQdo Connect

inQdo Connect is *the* integration platform – the ESB – offered by inQdo. inQdo Connect uses the product suite of Software AG, which is placed as a leader in Gartner's Magic Quadrant and The Forrester Wave with, among others, SOA governance and Enterprise Service Bus.

inQdo Connect consists of:

- Enterprise Service Bus;
- SOA Infrastructure and Governance;
- BPMS.

inQdo Connect is based on cloud technology and has a number of advantages over traditional implementations of and support for integration platforms.

The solution enables the organization to (among other things) take advantage of the fact that:

- tailor-made security is offered;
- complexity is more predictable;
- no substantial investments and obligations have to be made or entered into;
- operating costs are predictable;
- opportunities for growth and counterbalancing of peak loads is seen to;
- monitoring of the process is available;
- implementation costs are lower compared to traditional projects;
- standards can be used;
- speed of implementation (accelerators) is better than in a regular project;
- project risks can be limited;
- simple connection to other cloud solutions is possible;
- product development by market leaders can be used;
- faster (controlled) anticipation of future mobile functionality is possible.

Consulting Partner	Company Name Pironet NDH CANCOM SE	Home Page www.pironet.com ; www.cancom.com
Registered	Company HQ City & Country **Cologne, Munich, Germany**	Services offered in Europe
Domains and/or Technologies	AWS Consulting & Integration, Brokerage IaaS platform, Health Care	

13.5 Pironet/ Cancom

CANCOM/PIRONET is a leading provider in the German cloud computing market. For more than two decades, medium-sized and enterprise companies, who want to be one step ahead of their competitors through IT innovations, trust in CANCOM. The focus of CANCOM/PIRONET is supporting its customers with a Multi Cloud strategy. Summarized under its brand BusinessCloud, CANCOM offers customers a variety of cloud-based deployment models in the market. Thus, customers get a mix of on-premises, hosted, and public cloud services from a single source. This Multi Cloud approach provides companies with a high degree of agility in IT delivery and implementation of new digital strategies.

Flexibility for the adoption, use and operation of modern cloud computing solutions can be combined with the highest degree of IT stability. Companies can free themselves of unnecessary IT tasks and at the same time push forward their digital agenda. Both the service offer and IT infrastructure as well as the internal organization of CANCOM/PIRONET are certified to strict, internationally recognized guidelines (DIN ISO / IEC 27001). This proofs the operational excellence in all process sequences as well as compliance with high technical and safety-related standards.

13.5.1 Contact Info

Simon Russin
srussin@pironet.com
Pironet NDH Datacenter AG & Co. KG
Von-der-Wettern-Straße 27
51149 Köln
Germany

13.5.2 Competencies

Core expertise in remote application and data management including SLA services for:

- SAP, Oracle, Microsoft, Progress Software, IBM, Citrix, leading Open-Source Applications
- General consulting capabilities for AWS

- Specific solution template for Backup and Disaster Recovery as a Service

13.5.3 Office Location(s)

PIRONET NDH
Datacenter AG & Co. KG
Von-der-Wettern-Straße 27
51149 Köln
Deutschland

Overview CANCOM office locations: http://www.cancom.de/unternehmen/standorte/

13.5.4 Geographic Coverage

Europe.

13.5.5 Solutions & Services offered

- Cloud on-boarding
- Cloud migration
- Cloud service & support

13.5.6 Certifications

Do you have any certifications that testify to your ability to deliver on research computing projects?

CANCOM/PIRONET was awarded the most important certification of IT Security, the ISO-norm 27001.

The DIN ISO/IEC 27001 is an international certification standard for the information security management (ISMS), mostly used in business.

Consulting Partner	Company Name **Sterling Geo**	Home Page http://www.sterlinggeo.com/
Registered	Company HQ City & Country **Loughborough, UK**	Services offered in Globally
Domains and/or Technologies	Geography, Machine Learning, Statistics, Land Management, Agriculture, Asset Management, Deforestation, Natural Resource Exploration / Management, Risk Assessment, Flooding, Smart City Development, Migration, Utilities Management, GIS Remote Sensing, Geospatial HPC & highly scalable cluster workloads – consulting to the stars.	

13.6 Sterling Geo

Sterling Geo engineers solutions that transform spaceborne, airborne and terrestrial data into accurate, actionable information. We combine the power of the cloud with earth observation technology to create new lightweight business applications that meet specific current and future business needs across multiple UK and global markets. We have a reputation for capability, quality and service, emanating from a combination of having the right people and our specialist technical knowledge.

13.6.1 Contact Info

Phone: 0800 912 0988

enquiries@sterlinggeo.com

13.6.2 Competencies

Sterling Geo is a strategic partner to two key software developers, Hexagon Geospatial and Safe Software. Sterling Geo is the leading UK distributor of the Hexagon Geospatial software portfolio, we provide technology for imagery intelligence and terrain analysis to over 200 government, academic and commercial organizations in the UK. We became a Top 25 European Partner of Safe Software within two operational years and have been Gold Partners since 2014. We deliver and support Safe Software's entire FME product suite including FME Desktop, FME Server and FME Cloud. Our technical team comprises Certified Hexagon Geospatial Trainers, Certified FME Trainers and Certified FME Professionals.

In addition, we are experts in managing EO data, especially satellite imagery especially on AWS

13.6.3 Office Location(s)

Sterling Geo, 14 Charnwood Office Village, Loughborough, Leicestershire, UK LE12 5NL

Sterling Geo, Satellite Applications Catapult, Electron Building, Fermi Ave, Harwell OX11 0QR

13.6.4 Geographic Coverage

Sterling Geo have traditional traded in the UK market for delivery of desktop software solutions. However with the emerging technologies of our lightweight applications our market has extended to the global arena.

13.6.5 Solutions & Services offered

- *Assessments*
- *Cloud Migration Services*
- *Custom Application Development*
- *Strategic / IT Consulting*

13.6.6 Certifications

- ISO accredited 19001, 14001 and 2008
- Certified ERDAS IMAGINE Trainer
- Certified FME Trainer
- Certified FME Professional
- FME Cloud development and service provision

Consulting Partner	Company Name **The Server Labs Ltd.**	Home Page http://www.theserverlabs.com/
Standard	Company HQ City & Country **London, United Kingdom**	Services offered in UK, Germany, France, Switzerland & Spain
Domains and/or Technologies	AWS compute services, Big Data, HPC & highly scalable cluster workloads noSQL, NewSQL, Hadoop, Twitter Storm, Apache Sparc Application Security, Cloud Security, Encryption, Tokenization Maven, Git, Jenkins, SonarQube, Docker, Puppet, Chef, Ansible, Salt, Kubernates MongoDB, Cassandra, Couchbase Scalable and elastic architectures using REST	

13.7 The Server Labs Ltd.

The Server Labs is a specialist IT Consultancy and systems integration company and a leading authority in the architecture and development of Cloud solutions, Dev Ops, Big Data and HPC (High Performance Computing) in the Cloud.

European based, with a varied group of customers in the UK, Spain and Germany, in different verticals, The Server Labs focuses on the design, implementation and management of scalable IT architecture solutions based on Cloud and agile software engineering. We believe that the Cloud and DevOps together will enable our customers to attain the goal of Continuous Delivery. We collaborate with our Customers to obtain success, committed to innovation, believing in what we do every day and growing with every challenge. We work with the most advanced technologies, methodologies and computing paradigms to offer our customers cost-effective, robust, scalable and high performance solutions.

13.7.1 Contact Info

Dolores Saiz
sales@theserverlabs.com

13.7.2 Competencies

- HPC, Big Data

The Server Labs were one of the first AWS partners to run HPC workloads in HPC. See https://aws.amazon.com/blogs/aws/scaling-to-the-stars/ for a description of the work we have done with the European Space Agency.
We have successfully worked in HPC and Big Data projects for ESA, Bayer, CNIO, CERN, EMBL to name a few.

- Software Development & Architecture

We have many years of experience in developing full stack scalable HPC and cluster based systems, and for the last few years we have been developing cloud native applications, both front end and back end.

- DevOps

We have many years of experience in Dev Ops including Agile Methodology Adoption, Development Process Improvement, Continuous Integration all the way through to Continuous Delivery. We have successfully implemented our Continuous Deployment Platform in customers such as ESA, BBVA, BNP Paribas, Amadeus, Madrid Underground, etc.

- Cloud Services

We help our customers in their Cloud transformation initiatives, from inception to implementation and management end to end. We are cloud architects and we design and implement bespoke Cloud solutions. Our services include
System & Infrastructure Assessment, Cloud Security Assessment, Cloud Architecture design, Cloud Adoption Strategy, Cloud Infrastructure configuration and deployment, Private Cloud Deployment, Cloud Security implementation, Consultancy for architecture evolutions, Cloud Interoperation, Application refactoring for the Cloud

13.7.3 Office Location(s)

UK	**Germany**	**Spain**
Regent's Place 338 Euston Road London NW1 3BT	Hanauer Landstraße 126-128 60314 Frankfurt am Main	Maria de Molina 39, 8th floor 28006 Madrid

Registered Office: 21-27 Lamb's Conduit St, London, WC1N 3GS

13.7.4 Geographic Coverage

UK, Germany, France, Switzerland & Spain.

Our research customers include ESA (European Space Agency), Eumetsat, CERN, EMBL, CNIO, Bayer

13.7.5 Solutions & Services offered

The Server Labs is an IT Consultancy focused in four main areas.

- Software Development & Architecture
- DevOps
- Cloud services
- HPC, Big Data

The Server Labs are your Cloud Architects that will help you deploy cost-effective and successful solutions in the Cloud for your business. The Server Labs' Cloud Services include:

- Cloud Assessment & Adoption Strategy

- Cloud Security Assessment
- Cloud Architecture design
- Application Migration to the Cloud
- Cloud Infrastructure configuration and deployment
- Private Cloud Deployment
- Application refactoring for the Cloud
- Development of applications in the Cloud
- Cloud Managed Services

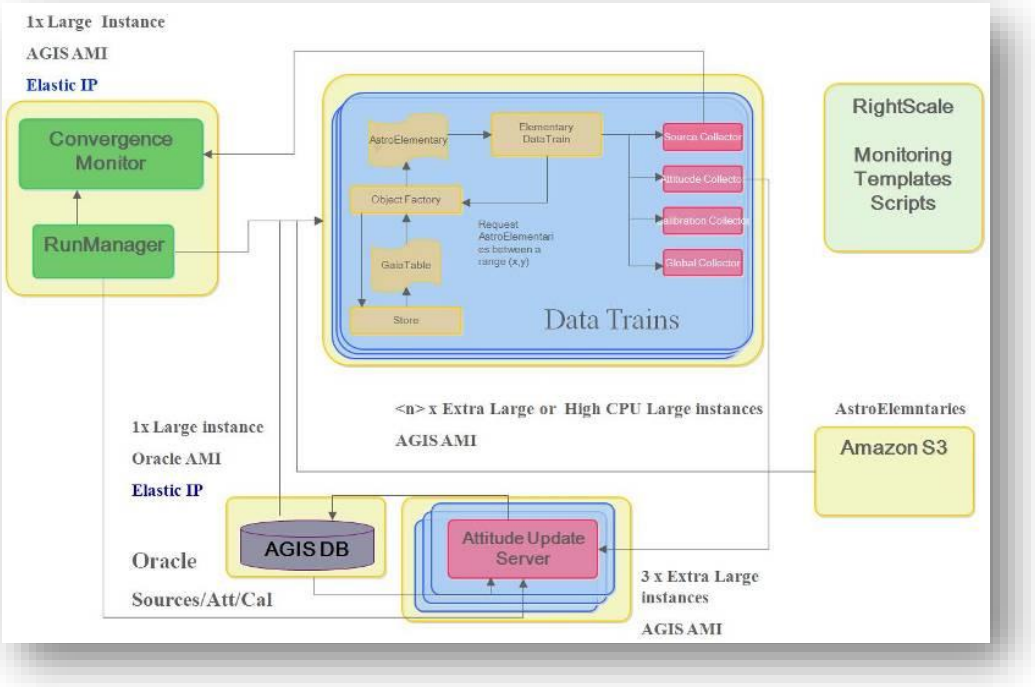
Additional Information from The Server Labs

Big Data & High Performance Computing

The Server Labs has many years of experience in both traditional and cloud-based HPC and Big Data. We have successfully completed Big Data projects with organizations such as ESA, CNIO, Eumetsat and we are currently working with CERN and EMBL within the pan-European Helix Nebula initiative.

Gaia Data Processing in the European Space Agency

The Server Labs (TSL) is currently part of the frame contract for the development of the ESA's Gaia mission data processing pipeline. The Server Labs has offered its experience in the setup and deployment of a High Performance Computing framework in Amazon Cloud for the processing of Telemetry Test Data. Currently, TSL is testing further the capacity of its data processing, testing the horizontal scalability of Gaia's data processing grid to limits impossible with the current in-house cluster. In this new science project TSL is running ESA's Gaia data processing in Amazon's Cloud.



Helix Nebula

The Server Labs is part of the Helix Nebula Consortium, building a Science Cloud for Europe. Within the project we have been defining the core architecture and working with CERN for the HPC Processing of LHC data, ESA, deploying their Supersites for Earthquake data and EMBL deploying and architecting their Genomic processing chain. The Server Labs has also contributed to Helix Nebula with our EC2 Bridge that allows EC2 based HPC workloads to run on other public or hybrid clouds.

Eumetsat (The European Organisation for the Exploitation of Meteorological Satellites)

The Server Labs undertook an in-depth market study to assess the overall feasibility and technical and financial suitability of using cloud, grid, supercomputing or other type of computing services and solutions for EUMETSAT's MTG (METEOSAT Third Generation) program. Throughout the project The Server Labs has identified the best solution(s) for long-term reduction of costs.

The Server Labs is currently working with EUMETSAT on Private Cloud and HPC processing initiatives.

CNIO (Spanish National Cancer Research)

The Server Labs worked with CNIO to enable the Spanish national cancer research organization to overcome the bottleneck of data processing in its research programs. In an initial feasibility study, The Server Labs assessed how to transfer the vast computing requirements of CNIO's research projects to the Cloud, aiming to make the processing

more agile while bringing down cost and reducing the need for in-house infrastructure. More information can be found on: http://www.theserverlabs.com/blog/2010/12/14/genomic-processing-in-the-cloud/.

Big Data/HPC in the Cloud (on demand)

The Server Labs "Big Data/HPC in the Cloud" Solution is a cloud-based environment for your big data and HPC projects that incorporate technologies such as Hadoop, Spark and Storm. We provide all the necessary utilities and tools to effectively manage the vast streams of data and information generated by science projects in a pay as you go model. We also integrate and partner with data analytics platforms to enable the visualization and analysis of science data within our cloud based solution.

By combining the scale and power of the Cloud Infrastructure with the data management capabilities of these technologies, organizations and enterprises are able to utilize advanced data management services without the need to invest and manage their own server environment.´

Our on-demand Big Data Solution provides you with:

- Data storage services to capture and access data in any format
- Data modeling support to help you analyze, transform and extract useful information from your data
- Data management services to process, monitor and operate Hadoop
- Data platform services to secure, archive and scale for consistent availability

For companies that desire faster query response times, more economical scale and greater stability of their Hadoop systems, our "Big Data in the Cloud" Solution is an ideal choice.

Consulting Partner	Company Name **Zenotech Ltd**	Home Page http://www.zenotech.com/
Std/Adv/Premier	Company HQ City & Country **Bristol, UK**	Services offered in UK, Europe
Domains and/or Technologies	Fluid Dynamics for Aerospace, Renewables, Civil Engineering and Automotive. Specialists in High Performance Computing, multi-core and accelerated computing and use of cloud computing for HPC.	

13.8 Zenotech

Zenotech are specialists in Computational Fluid Dynamics and High-Performance Computing. We develop and apply technology across a range of industrial domains and work closely with research and development organisations to develop the next generation of tools. Our typical customers are companies looking to increase their simulation capability either via the use of new software tools or via access to the latest HPC resources.

13.8.1 Contact Info

We can be contacted via our website www.zenotech.com/contact-us or call us directly on +44 (0)117 906 1100

13.8.2 Competencies

Research into Computation Fluid Dynamics, advanced CFD algorithms and industrialization/application of research output. We use many different programming languages but have expertise in C++, Python, Cuda and Java. We develop for Linux, Windows and Mac. We also provide consulting support to organizations looking to exploit GPU acceleration with their own problems.

13.8.3 Office Location(s)

Our main office is:
Zenotech Ltd
Bristol and Bath Science Park
Dirac Crescent
Emersons Green
Bristol BS16 7FR
UK

13.8.4 Geographic Coverage

We work predominately in the UK but work with organizations with a global reach. We are open to working with partners from anywhere in the world.

13.8.5 Solutions & Services offered

We have two primary offerings. We develop Computational Fluid Dynamics software that is designed to make the most of the latest hardware platforms and algorithm developments and we also offer HPC services that enable customers to run at scale. Our software is developed to run efficiently at very large scale and leverage the latest GPU technology to accelerate the simulations. We can work with organisations to advise on how to best take advantage of GPU/Multi-core architectures and how to best develop code for those platforms.

As a consulting partner we can work with research organisations to bring the latest CFD developments to industry. We can collaborate on research programmes in the UK or Europe.

We can help organisations migrate HPC workloads to the cloud or 3rd party HPC sites.

13.8.6 Certifications

Affiliations:
Institute of Engineering and Technology (IET)
Institute for Mathematics and Its Applications (IMA)
Australian Mathematics Society (AustMS)

13.8.7 Case Studies

Zenotech have successfully led and partnered in several Innovate UK projects. Two current ones are:

Hyperflux++ builds on the successful Innovate UK project 101890 “Hyperflux” - developing next generation CFD technology for the civil, automotive, renewable and aerospace sectors using the cutting-edge high order flux reconstruction technique from Peter Vincent and his team at ICL. Hyperflux++ brings Bombardier, CFMS, Aircraft Research Association Ltd and Zenotech together to further develop the capability and address timely challenges in the aerodynamic modelling of undercarriages and nacelles. We will include localized transition modelling; better and more robust high order mesh generation and high fidelity acoustic source modelling. The capability will be available to all UK organisations via cloud access at the CFMS supercomputer, and will leverage the latest in many-core hardware for fast, efficient computation. Workflow integration to existing tool chains will be via support for most mesh formats, with automated upgrade to high-order elements. Hyperflux is a UK-based software programme, underpinned by expertise within the UK. This is in line with government strategies for HVM and ICT, and forms a cornerstone for the UK aerodynamics ATI.

SWEPT2 follows the successful “Simulated Wake Effects Platform for Turbines” project to establish the viability of GPU-based fluid dynamics simulation of turbine array wake interaction effects. With a view to growing a UK-based technology supply chain, the original partners (DNV GL, Zenotech and the University of Bristol) will be joined by the Offshore Renewable Energy Catapult (providing access to LIDAR data for validation), STFC Daresbury (to apply the latest in big-data analytics to the challenge of comparing CFD & experimental data), CFMS (cloud computing integration and optimisation), and the

universities of Surrey, Strathclyde and Imperial College – to provide expertise in wake turbulence and wind tunnel data. SSE Energy Supply Limited has agreed to independently assess the functionality and value of the service during the project. SWEPT2 addresses the energy trilemma by (i) reducing emissions through the use of renewable energy sources, (ii) improving security of supply by supporting UK turbine array installations, and (iii) reducing cost by improving the accuracy with which turbine array performance can be predicted & further optimizing wind farm layout & control strategies.

Additional information provided by Zenotech:

Zenotech is a high-tech simulation specialist. We develop high performance computing and computational fluid dynamics technology for the aerospace, automotive, civil and renewable energy sectors. We deliver computational fluid dynamics (zCFD) simulation at scale for organisations of all sizes needing faster, more accurate and cost-effective simulation. High performance computing (HPC) is at the heart of our business: our tools exploit the latest many-core hardware systems. We work with leading academic researchers to develop cutting edge technology; consultants and capability providers who can use and add value to our tools, and end users looking for a performance edge. Industrial organisations and technology partners gain competitive advantage and add value to their own products and expertise by incorporating our tools as part of their own product range. We can provide consulting services for technology strategy, computational engineering and high-performance computing.

Testimonies:

"The EPIC on-demand high performance computing service by Zenotech has been for us something of a game changer: the straightforward, secure access to scalable HPC resources has allowed us to deliver high value engineering services significantly faster and better than before "...

[Steve Walker, Arup]

"We are delighted to be partnering with Zenotech to develop advanced fluid dynamics simulation software. This brings our leading fundamental research programme closer to industry - delivering high fidelity engineering technology "

[Peter Vincent, Imperial College London]

Reviews of previous project management:

“I have been extremely pleased with the professional way this project has been run. Documentation has always been first class and provided on time.”

[Michael Catania, Monitoring Officer - Innovate UK]

14 Glossary

Access Key

The combination of an Access Key ID (like `AKIAIOSFODNN7EXAMPLE`) and a secret Access Key (like `wJalrXUtnFEMI/K7MDENG/bPxRfiCYEXAMPLEKEY`). You use Access Keys to sign API requests that you make to AWS. **Not** to be confused with an SSH key pair.

AMI

Amazon Machine Image (AMI). An encrypted machine image stored in Amazon Elastic Block Store (Amazon EBS) or Amazon Simple Storage Service (Amazon S3). AMIs are like a template of a computer's root drive. They contain the operating system and can also include software and layers of your application, such as database servers, middleware, web servers, and so on.

Budgets

AWS Budgets is a tool to help you plan your usage and your costs (also known as your spend data) and to track how close your usage and costs are to your budgeted amount. You can set up notifications that warn you if you go over your budgeted amount or are forecasted to go over your budgeted amount. Notifications can be sent to an email address or an Amazon Simple Notification Service (Amazon SNS) topic where they can be programmatically handled to take specific actions.

AWS CloudFormation

A service for writing templates that create and delete related AWS resources together as a unit. See also http://aws.amazon.com/cloudformation.

Amazon CloudWatch

Amazon CloudWatch is a web service that enables you to monitor and manage various metrics and configure alarm actions based on data from those metrics. See also http://aws.amazon.com/cloudwatch.

Amazon EBS

Amazon Elastic Block Store (Amazon EBS) is a service that provides block-level storage volumes for use with Amazon EC2 instances. Block storage is similar to raw, unformatted hard drive space and needs to be partitioned and/or formatted to contain data in a structured way (for example, by a file system like FAT, NTFS, or ext4).

See also http://aws.amazon.com/ebs.

Amazon EC2

Amazon Elastic Compute Cloud (Amazon EC2) is a web service that enables you to launch and manage Linux/UNIX and Windows server instances in Amazon's data centers. See also http://aws.amazon.com/ec2.

Amazon EFS

Amazon Elastic File System (Amazon EFS) is a file storage service for Amazon EC2 instances. Amazon EFS provides a simple interface to create and configure file systems with traditional directory structures, similar to a large file server. Amazon EFS storage capacity and performance grows and shrinks automatically as you add and remove files. See also http://aws.amazon.com/efs/.

Amazon EMR

Amazon Elastic MapReduce (Amazon EMR) is a web service that makes it easy to process large amounts of data efficiently. Amazon EMR uses Hadoop processing combined with several AWS products to do web indexing, data mining, log file analysis, machine learning, scientific simulation, and data warehousing. See also http://aws.amazon.com/elasticmapreduce.

IAM

AWS Identity and Access Management (IAM) is a web service that enables AWS customers to manage users and user permissions within AWS. See also http://aws.amazon.com/iam.

Instance Storage

Storage that comes built into an Amazon EC2 instance type, in contrast to Amazon EBS storage, which is provisioned separately and subsequently "attached" to an instance.

Amazon Machine Learning

A cloud-based service that creates Machine Learning (ML) models by finding patterns in your data and uses these models to process new data and generate predictions.

See also http://aws.amazon.com/machine-learning/.

AWS Marketplace

AWS Marketplace is a web portal where qualified partners market and sell their software to AWS customers. It helps customers find, buy, and immediately start using the software and services that run on AWS. See also http://aws.amazon.com/partners/aws-marketplace/.

MFA

Multi-Factor Authentication (MFA) is an optional AWS account security feature. Once you enable AWS MFA, you must provide a six-digit, single-use code in addition to your sign-in credentials whenever you access secure AWS webpages or the AWS Management Console. You get this single-use code from an authentication device that you keep in your physical possession (like your smart phone or a separate dongle). See also http://aws.amazon.com/mfa/.

Amazon RDS

Amazon Relational Database Service (Amazon RDS) is a web service that makes it easy to set up, operate, and scale a relational database in the cloud. It provides cost-efficient, resizable capacity for an industry-standard relational database and manages common database administration tasks. See also http://aws.amazon.com/rds.

Amazon S3

Amazon Simple Storage Service (Amazon S3) is storage for the Internet. You can use it to store and retrieve any amount of data, at any time, from anywhere on the web, using HTTP GETs and PUTs. Amazon S3 is very low cost and extremely large scale and is typically used to house working data when it is "at rest." It's designed for write-once, read-many-times duty cycles. Often referred to as "object storage," Amazon S3 can contain items up to 5 TB in size for a single object. See also http://aws.amazon.com/s3.

Amazon SNS

Amazon Simple Notification Service (Amazon SNS) is a web service that enables applications, end users, and devices to instantly send and receive notifications from the cloud.

SSH

Secure Shell (SSH) is a shell or terminal program for accessing remote Linux servers that has encrypted data transport and authentication enabled by the use of digital cryptographic keys.

SSH Key Pair

A public and private key pair for use by SSH, which can either be generated by AWS's Amazon EC2 console or imported into AWS from your local desktop. You must specify a key pair when you launch and connect to an Amazon EC2 instance. See also http://docs.aws.amazon.com/cli/latest/userguide/cli-ec2-keypairs.html.

Amazon VPC

Amazon Virtual Private Cloud (Amazon VPC) is a web service for provisioning a logically isolated section of the AWS Cloud where you can launch AWS resources in a virtual network that you define. You control your virtual networking environment, including selection of your own IP address range, creation of subnets, and configuration of route tables and network gateways.