



Cloud Computing for Science

Orienting

Dennis Gannon



Tutorial Goals

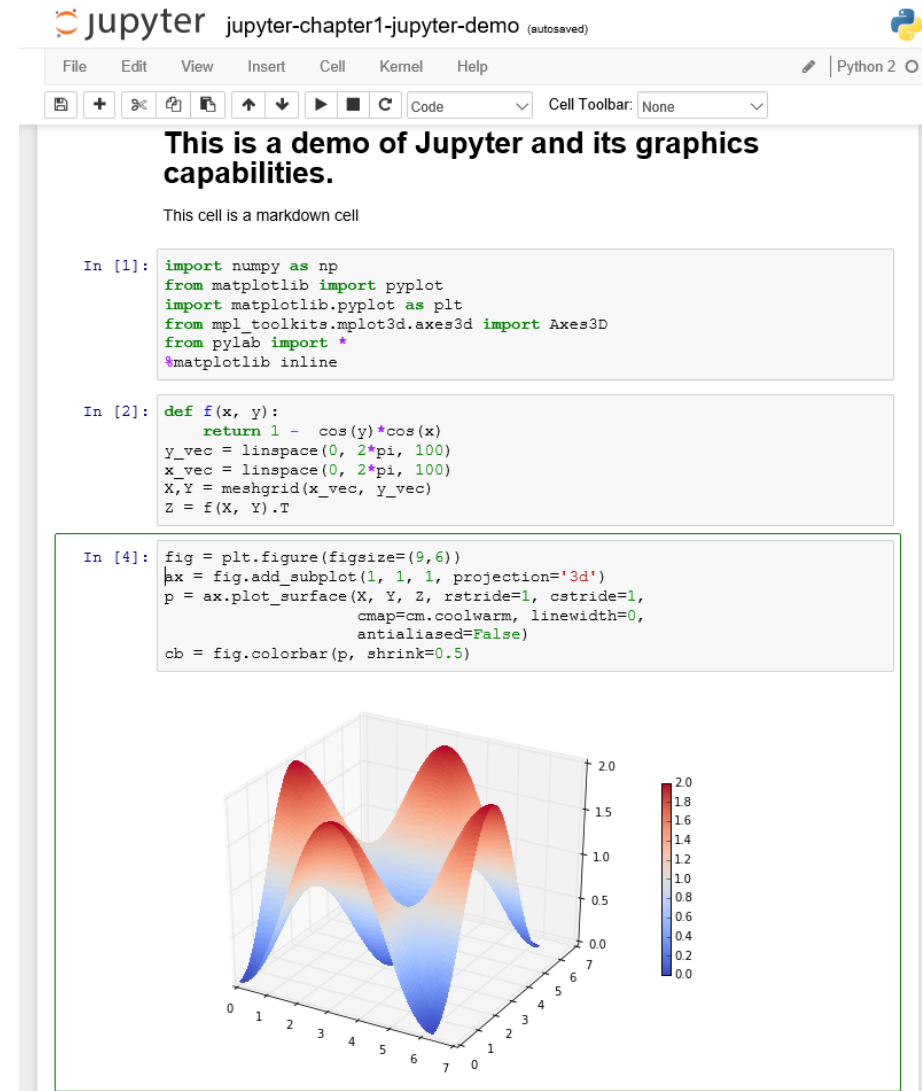
- An exploration of cloud computing for researchers
 - Scientists who need to go beyond their current resources
 - Computer science students who need to know what is possible
 - Data scientists who want to understand the potential of the cloud
- What will be covered
 - Cloud data services
 - VM and Container basics
 - Ways to scale: clusters, mapreduce, microservices
 - Data analytics in the cloud
 - Streaming data
 - Machine learning in the cloud

Preliminaries

- We will use Microsoft's Azure for Hands-on Exercises
 - Initializing your account
 - Download Azure Storage Explorer <http://storageexplorer.com/>
- Access to these slides and Jupyter Notebooks
 - From the book "Cloud Computing for Science and Engineering" by Ian Foster and Dennis Gannon, to be published soon by MIT Press.
 - The link to the book is here
 - <https://Cloud4SciEng.org>

We will be using Python and Jupyter

- Jupyter in the cloud
 - Go to <https://notebooks.azure.com>
 - Signup – it's free.
 - If you are new to Jupyter do Welcome.ipynb
 - If you are new to Python do Python.ipynb
- Installing Jupyter on your laptop
 - Go to <https://docs.continuum.io/anaconda/install>
 - Do it. Then “Jupyter notebook” at the shell
- Better solution: install Docker
 - <https://docs.docker.com/engine/installation/>
- and run Jupyter in a container
 - `docker run -it -p 8888:8888 dbgannon/tutorial`
 - Accept security exceptions and login with “tutorial”
 - Open notebook jupyter.ipynb to see the one on the right



Orienting in the Cloud

- We will discuss three different “public clouds” and a bit about a “science private clouds”
 - Public = anybody with a credit card has access. (aka commercial cloud)
 - Private = restricted to a special group of users. (aka Community Cloud or Academic Cloud)
 - (In Europe these terms are often reversed based on ownership.)
- They are:
 - Amazon Web Services (AWS) - 40% of all cloud resources on the planet.
 - Microsoft Azure – about 1/3 of AWS but growing
 - Google Cloud – third place
 - NSF JetStream – an OpenStack private cloud for US science researchers.
- There are *many* more clouds.
 - Public: IBM, DigitalOcean, Rackspace, 1&1, UpCloud, CityCloud, CloudSigma, CloudWatt, Aruba
 - Private Research Clouds: Aristotle, Bionimbus, Chameleon, Jetstream, RedCloud, indigo-datacloud, EU-Brazil Cloud, European Open Science Cloud
- What are the pros and cons of public vs private

Pros & Cons of Public vs Private Cloud

- Public cloud pros

- Massive scale
- Huge and growing list of services
- Highly competitive on pricing due to economies of scale
- Security is strong
- Freedom from managing hardware
- Hardware constantly upgraded

- Cons

- Rules prohibit data moving to cloud
- Funding models may make it hard to use
- Fear of “vendor Lock-In”

- Private cloud pros

- May be cheaper
- You can keep it off the Internet so data can be very safe.
- You can optimize your own hardware
- You control everything

- Cons

- You are responsible for everything
- Not as many high level services
- May not really be cheaper
- You manage physical and system security

Two ways to access the cloud

- Portals and SDKs
- Web Portals
 - Dashboard that allow you to see and manage your cloud resources.
- Software Development Kits (SDKs)
 - Libraries that give you the tools to manage cloud resources from a program or script.
 - Based on REST web service calls
- Let's look at several Cloud Web Portals

Amazon AWS Portal

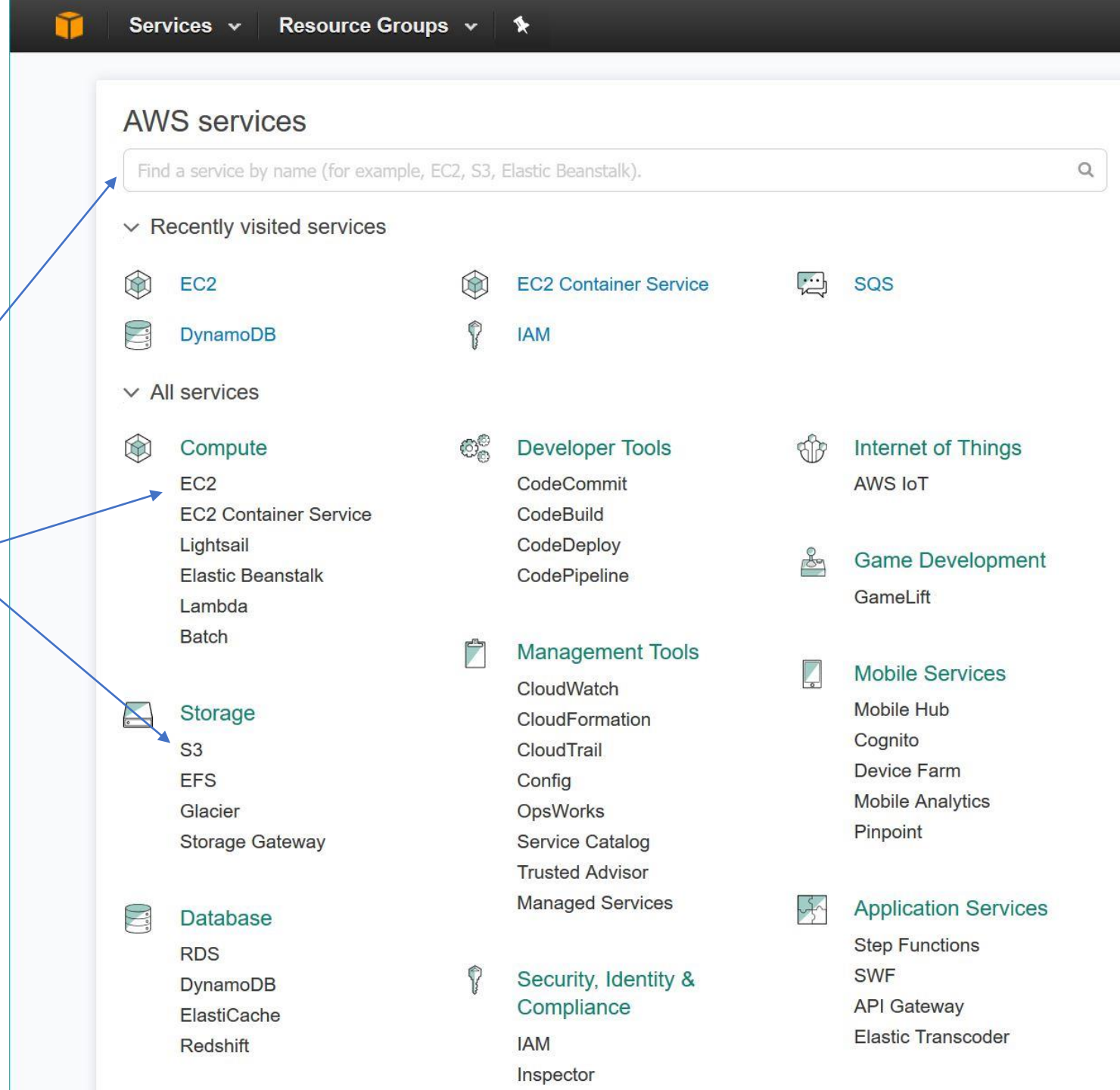
You see links to all the standard services

You also have a search bar to find others.

To create a storage account go to S3

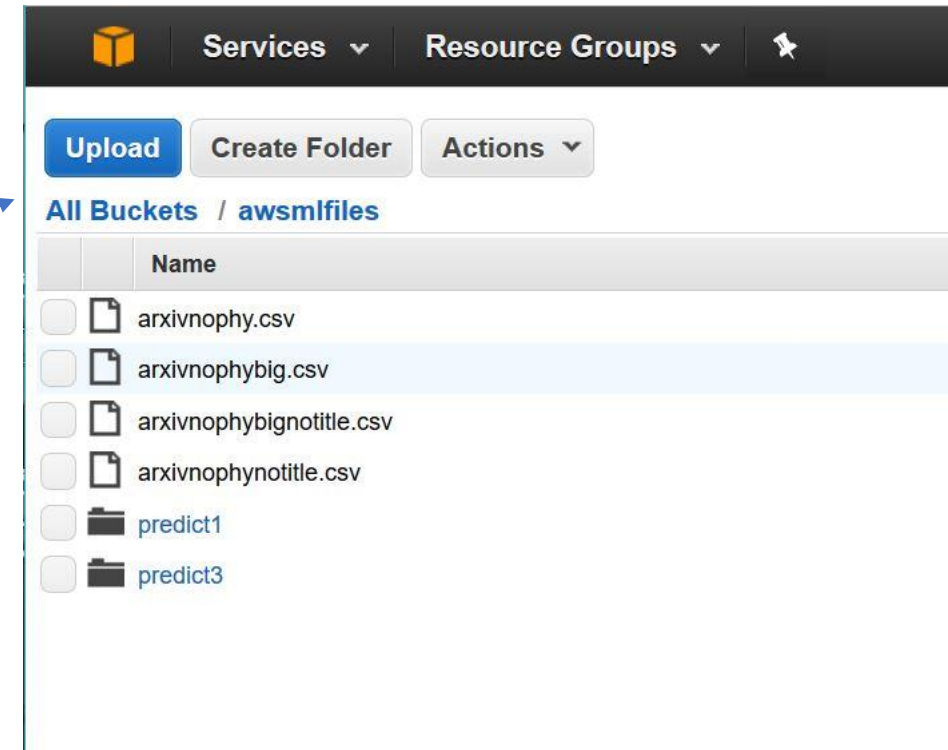
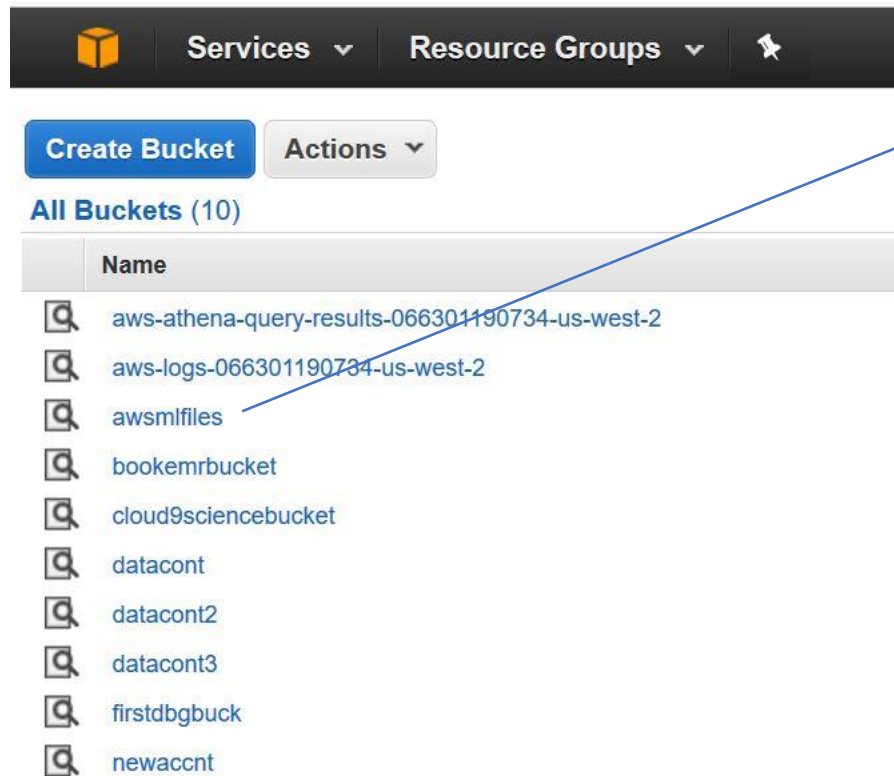
To launch a Virtual Machine go to EC2

Let's look at the S3 storage system



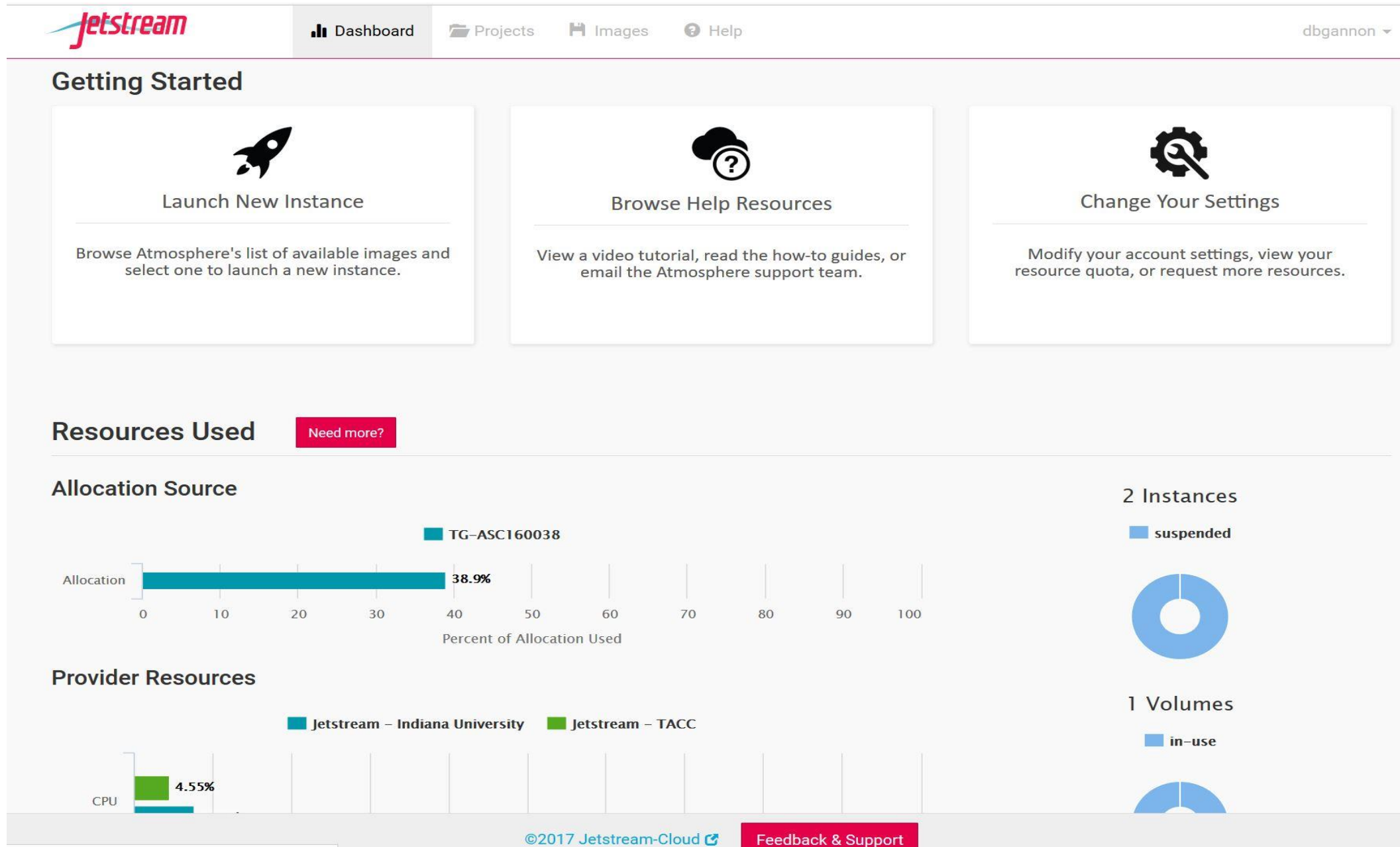
Selecting S3 we get the bucket list

Selection “awsmifiles” gives us the a list of the contents



Notice that it has regular objects AND folders.

Jetstream –NSF Science Cloud (OpenStack)



Jetstream is about virtual
Machines for science

Generic Linux

Bioscience

MATLAB

R

Featured Images



Centos 7 (7.3) Development GUI
Feb 24th 17 09:21 by jfischer

Centos 7 (7.3) Development GUI

CentOS development Featured gui iRODS



CentOS 6 (6.8) Development GUI
Feb 24th 17 09:19 by jfischer

Based on CentOS 6 (6.8) Development

◦ updated from 6.7 to 6.8

CentOS desktop development Featured gui iRODS vnc



Ubuntu 14.04.3 Development GUI
Feb 24th 17 09:17 by jfischer

Based on Ubuntu 14.04.3 Development

Base Ubuntu 14.04.3 + Xfce + Xfce-goodies, firefox, i ...

desktop development Featured gui iRODS Ubuntu vnc



BioLinux 8
Feb 24th 17 09:13 by jfischer

Based on Ubuntu 14.04.3 -Trusty Tahr - server - cloudimg

-- **REQUIRES m1.small instance ...

bioinformatics desktop Featured gui m1_small Ubuntu x2go



Intel Development (CentOS 7)
Feb 24th 17 09:08 by jfischer

Intel compilers and development environment

*REQUIRES a m1.small or larger VM to la ...

CentOS desktop development Featured gui Intel m1_small vnc



MATLAB (Based on CentOS 6)
Feb 24th 17 08:56 by atmoadmin

Imported Application - MATLAB (Based on CentOS 6)

CentOS desktop development Featured gui m1_medium vnc



R with Intel compilers (CentOS ...)
Feb 24th 17 08:50 by jfischer

R with Intel compilers built on CentOS 7 (7.3)

** Requires m1.small or greater sized VM * ...

CentOS desktop development Featured gui Intel m1_small vnc



Galaxy Standalone
Nov 15th 16 01:49 by admin

Galaxy 16.01 Standalone - based on Ubuntu 14.04.4 LTS

This is a standalone Galaxy server ...

The Azure Portal

Select "+ New"

The screenshot shows the Microsoft Azure Portal interface. The top navigation bar includes the Microsoft Azure logo, a search bar, and user profile information. The left sidebar contains a list of navigation items, with the '+ New' button highlighted by a blue arrow and the text 'Select "+ New"'. The main dashboard area is titled 'Dashboard' and includes options to create a new dashboard, edit, share, fullscreen, clone, or delete. The dashboard is divided into several sections: 'All resources' (listing various resources like virtual machines, network security groups, and storage accounts), 'Get started' (providing quick links to Virtual Machines, App Service, SQL Database, Storage, and Azure Portal), 'Resources' (showing a list of resources under a specific resource group), and 'Service health' (displaying a world map with status indicators for different regions).

Microsoft Azure

Search resources

Dashboard

+ New dashboard Edit dashboard Share Fullscreen Clone Delete

All resources ALL SUBSCRIPTIONS

Resource Name	Resource Type
biggerMLVM	Virtual machine
biggerMLVM-nsg	Network security group
biggermlvm842	Network interface
scimldata	Storage account
predicttopicPlan	Machine Learning Web Ser...
predicttopicstorage	Storage account
escistore2	Storage account
bookcluster3store	Storage account
bookcluster2store	Storage account

See more

Get started

- Virtual Machines
Provision Windows and Linux virtual machines in minutes
- App Service
Create web and mobile apps for any platform and device
- SQL Database
Managed relational database-as-a-service
- Storage
Durable, highly available and massively scalable storage
- Azure Portal
Learn about how to use the Azure Portal
- Marketplace

Resources BOOKRG

- biggerML
- biggerMLVM
- biggerMLVM

See more

Service health MY RESOURCES

World map showing service health status across various regions.

Selecting “+” gives this list of options.

Selecting “Storage” gives the secondary menu of types of storage apps.

To create a storage account select the top one.

- give it a name, research group and location.

Microsoft Azure New > Storage > Create storage account

New

- Storage
- Web + Mobile
- Databases
- Intelligence + analytics
- Internet of Things
- Enterprise Integration
- Security + Identity
- Developer tools
- Monitoring + Management
- Add-ons
- Containers
- Blockchain

RECENT

- Ubuntu Server 16.04 LTS Canonical
- Cloud service Microsoft
- Stream Analytics job Microsoft
- Storage account Microsoft

Storage

FEATURED APPS See all

- Storage account**
Use Blobs, Tables, Queues, and Files for reliable, economical cloud storage.
- Data Lake Store
Hyper-scale repository for big data analytic workloads
- StorSimple Physical Device Series
Manage one or more StorSimple physical appliances.
- StorSimple Virtual Device Series
Manage one or more StorSimple virtual appliances.
- Backup and Site Recovery (OMS)
A backup and disaster recovery solution to safeguard applications
- Veeam Cloud Connect for the Enterprise
Veeam Cloud Connect for the Enterprise is an easy, efficient way to
- AltaVault AVA-c4, version 4.2.2
Reduce time, cost, and risk with the NetApp® AltaVault™ cloud-based appliance

Create storage account

The cost of your storage account depends on the usage and the options you choose below.
[Learn more](#)

* Name

.core.windows.net

Deployment model

Resource manager Classic

* Account kind

General purpose

* Performance

Standard Premium

* Replication

Read-access geo-redundant storage (RA-...)

* Storage service encryption

Disabled Enabled

* Subscription

azure4research

* Resource group

☒ Create new ☐ Use existing

* Location

North Central US

☐ Pin to dashboard

Create [Automation options](#)

Next

- A deeper look at storage
- Virtual Machines and Containers
- Scaling deployments and Microservices demo
- Analytics
- Machine Learning in the cloud