Cloud Computing for Science

Orienting

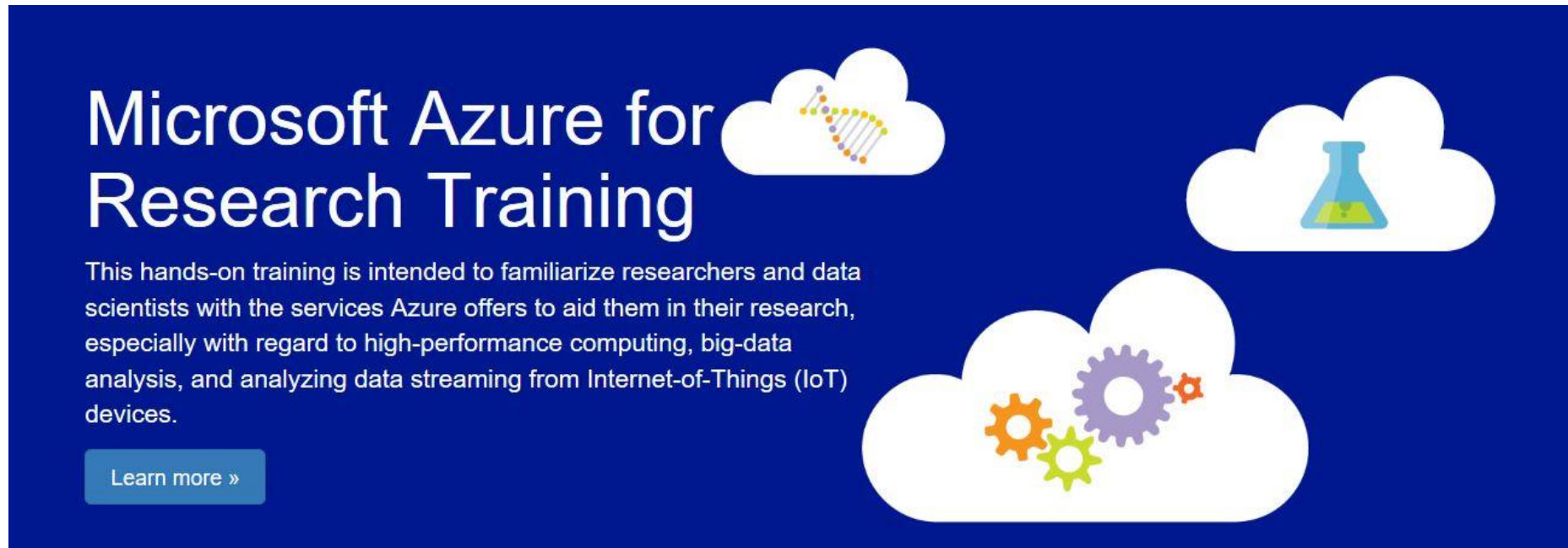Dennis Gannon, Ian Foster, Vani Mandava

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
  - This step for tutorial attendees only.
  - Got to http://azure4research.org.   Follow instructions from Vani.
- An Azure Pass is a single-month $500 credit that you can use on any Microsoft Azure service based on your needs including Virtual Machines, Websites, Cloud Services, Mobile Services, Storage, SQL Database, HDInsight, Media Services, and many more.

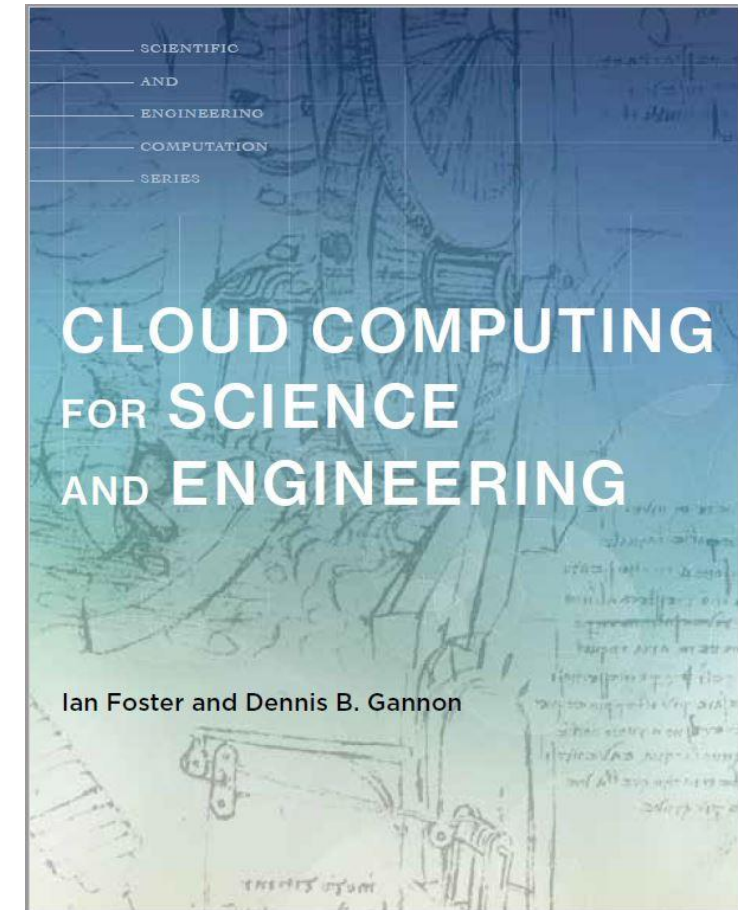Microsoft Azure for Research Training

This hands-on training is intended to familiarize researchers and data scientists with the services Azure offers to aid them in their research, especially with regard to high-performance computing, big-data analysis, and analyzing data streaming from Internet-of-Things (IoT) devices.

Learn more »

More preliminaries

- The book "Cloud Computing for Science and Engineering" by Ian Foster an Dennis Gannon, published by MIT Press. (see it at MIT press booth at SC17) has the slides and more
  - https://Cloud4SciEng.org
- Download Azure Storage Explorer http://storageexplorer.com/
- Access tutorial tarball https://SciEngCloud.github.io/tutorial.tar.gz
  - On mac:
    - gunzip tutorial.tar.gz
    - tar –xf tutorial.tar
  - on PC use something like 7-zip or other decompression and extractor.

More resources - go to
https://github.com/MSRConnections/Azure-training-course

To do later when you have time: more Jupyter and Python

- 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
- A Better solution: install Docker
  - https://docs.docker.com/engine/installation/
- and run Jupyter in a container
  - docker run –it -p 8888:8888 dbgannon/tutorial
  - In your browser go to https://localhost:8888
  - Accept security exceptions and login with "tutorial"
  - Open notebook jupyter.ipynb to see the one on the right
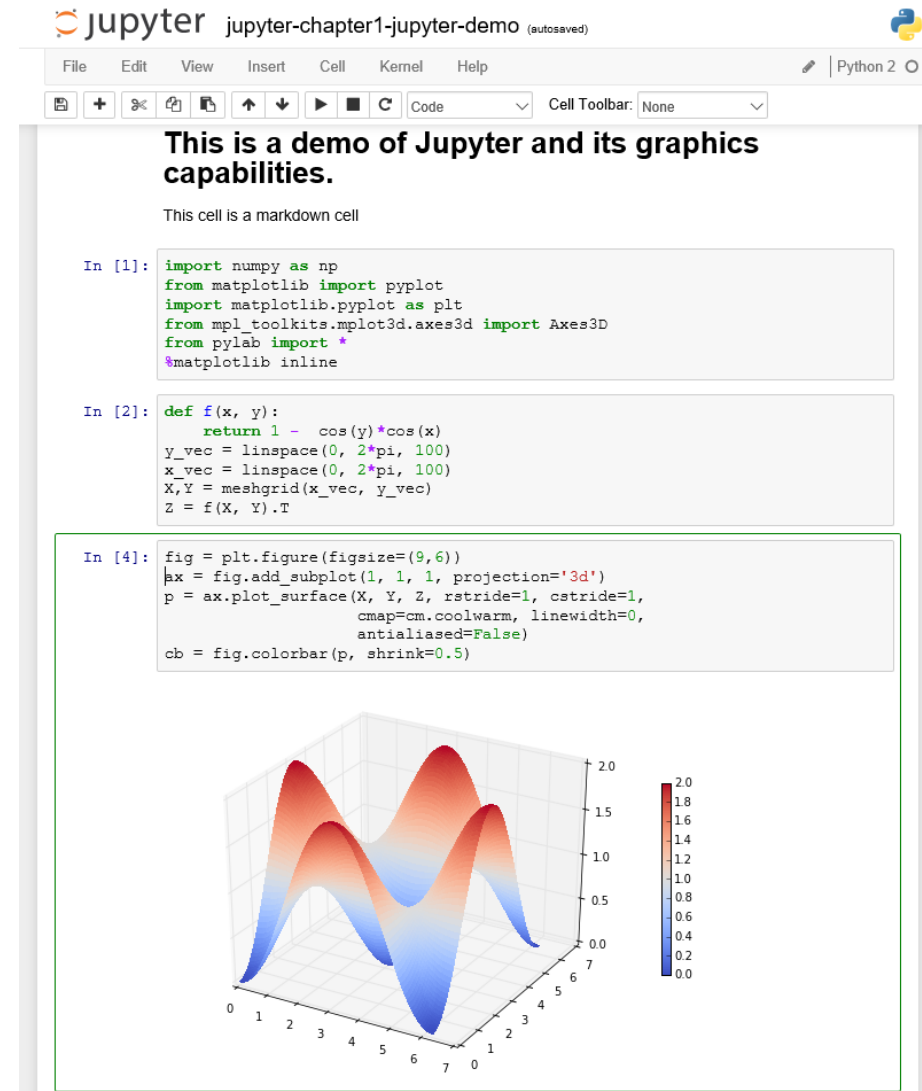
What is the Cloud?

Clouds began with big data centers

- Early days: 2005
  - Very simple servers
  - Network outward facing poor interconnect

- 2008-2016
  - Software defined networks
  - Special InfiniBand sub networks
  - Many different server types
    - 2 cores to 32 cores to GPU accelerations
  - Efficiency experiments
    - Geothermal, wind, wave
    - Containerized server

- 2017
  - Azure FPGA accelerated mesh
  - Google Tensor Processing Unit
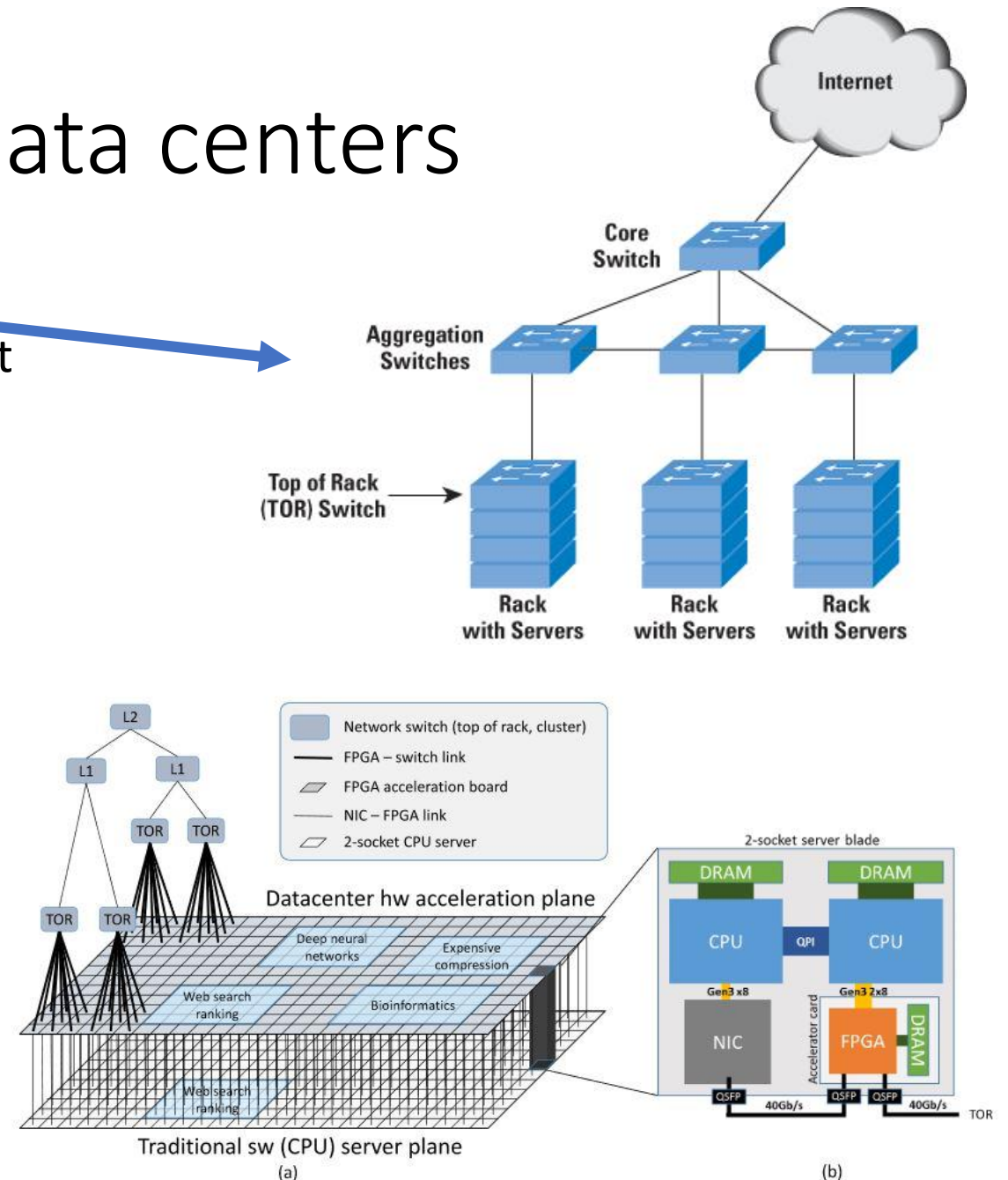 - Facebook – Open Compute Project
  - ARM based servers

Fig. 1. (a) Decoupled Programmable Hardware Plane, (b) Server + FPGA schematic.

Azure and AWS Global Data Center Network

Clouds are all about Services - Here is Azure

Platform Services

Security and Management
- Portal
- Active Directory
- Multi-factor Authentication
- Automation
- Key Vault
- Store/ Marketplace
- VM Image Gallery and VM Depot

Compute
- Cloud Services
- Service Fabric
- Batch
- Remote App

Web and mobile
- Web Apps
- API Apps
- API Management
- Mobile Apps
- Logic Apps
- Notification Hubs

Developer services
- Visual Studio
- Azure SDK
- Team Project
- Application Insights

Hybrid Operations
- Azure AD Connect Health
- AD Privileged Identity Management
- Backup
- Operational Insights
- Import/Export
- Site Recovery
- StorSimple

Integration
- Storage Queues
- Biztalk Services
- Hybrid Connections
- Service Bus

Analytics and IoT
- HDInsight
- Machine Learning
- Data Factory
- Event Hubs
- Stream Analytics
- Mobile Engagement

Data
- SQL Database
- SQL Data Warehouse
- Redis Cache
- Search
- DocumentDB
- Tables

Media and CDN
- Media Services
- Content Delivery Network (CDN)

Infrastructure Services

Compute
- Virtual Machine
- Containers

Storage
- BLOB Storage
- Azure Files
- Premium Storage

Networking
- Virtual Network
- Load Balancer
- DNS
- Express Route
- Traffic Manager
- VPN Gateway
- Application Gateway

Types of Clouds

- 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
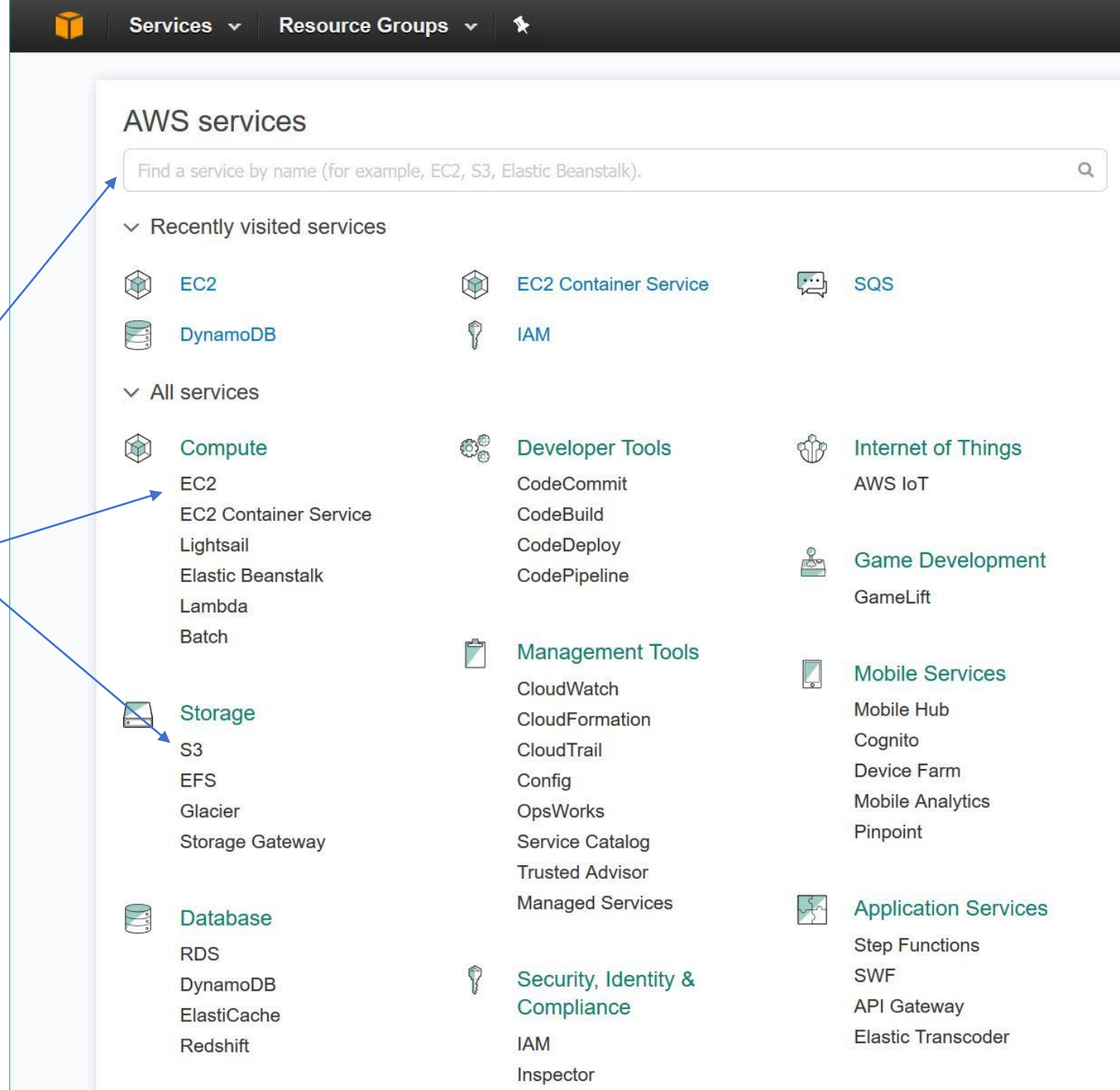
Amazon AWS Portal

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)
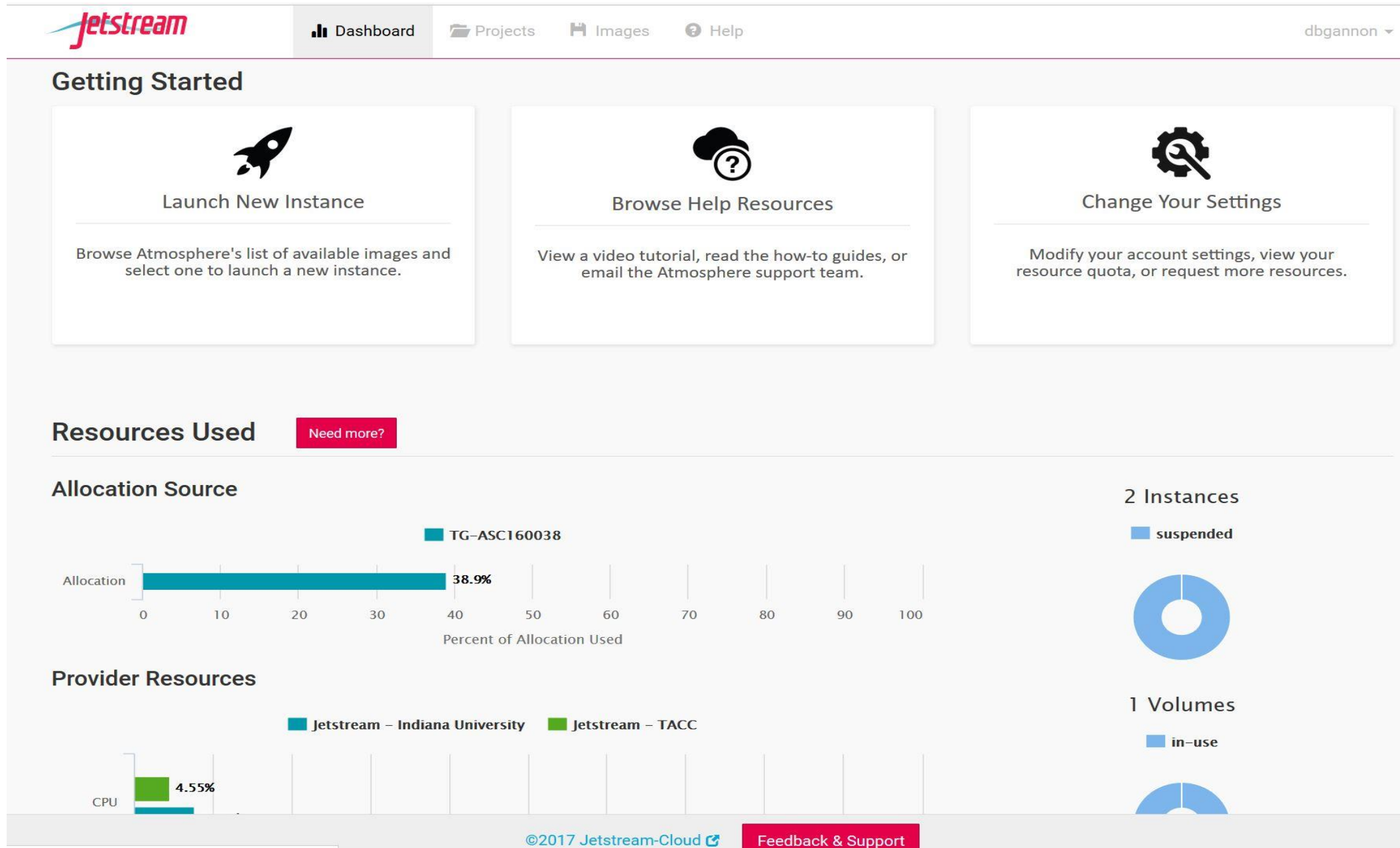
Jetstream is about virtual Machines for science

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"

Next

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