Leveraging Public Clouds for DOE Environmental Streaming Data

Marty Humphrey

Dept of Computer Science

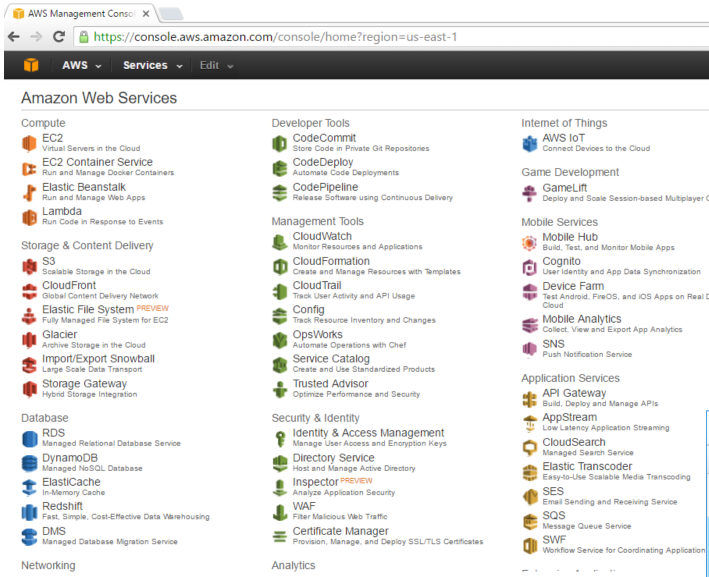
University of Virginia

Jon Goodall

Dept of Civil and Environmental Engineering

University of Virginia

Public Clouds should be utilized MORE by Scientists!



AWS Management Conso

https://console.aws.amazon.com/console/home?region=us-east-1

AWS | Services | Edit | Prof Marty Humphrey | N. Virginia | Support

Amazon Web Services

Compute
- EC2 – Virtual Servers in the Cloud
- EC2 Container Service – Run and Manage Docker Containers
- Elastic Beanstalk – Run and Manage Web Apps
- Lambda – Run Code in Response to Events

Storage & Content Delivery
- S3 – Scalable Storage in the Cloud
- CloudFront – Global Content Delivery Network
- Elastic File System PREVIEW – Fully Managed File System for EC2
- Glacier – Archive Storage in the Cloud
- Import/Export Snowball – Large Scale Data Transport
- Storage Gateway – Hybrid Storage Integration

Database
- RDS – Managed Relational Database Service
- DynamoDB – Managed NoSQL Database
- ElastiCache – In-Memory Cache
- Redshift – Fast, Simple, Cost-Effective Data Warehousing
- DMS – Managed Database Migration Service

Networking

Developer Tools
- CodeCommit – Store Code in Private Git Repositories
- CodeDeploy – Automate Code Deployments
- CodePipeline – Release Software using Continuous Delivery

Management Tools
- CloudWatch – Monitor Resources and Applications
- CloudFormation – Create and Manage Resources with Templates
- CloudTrail – Track User Activity and API Usage
- Config – Track Resource Inventory and Changes
- OpsWorks – Automate Operations with Chef
- Service Catalog – Create and Use Standardized Products
- Trusted Advisor – Optimize Performance and Security

Security & Identity
- Identity & Access Management – Manage User Access and Encryption Keys
- Directory Service – Host and Manage Active Directory
- Inspector PREVIEW – Analyze Application Security
- WAF – Filter Malicious Web Traffic
- Certificate Manager – Provision, Manage, and Deploy SSL/TLS Certificates

Analytics

Internet of Things
- AWS IoT – Connect Devices to the Cloud

Game Development
- GameLift – Deploy and Scale Session-based Multiplayer Games

Mobile Services
- Mobile Hub – Build, Test, and Monitor Mobile Apps
- Cognito – User Identity and App Data Synchronization
- Device Farm – Test Android, FireOS, and iOS Apps on Real Devices in the Cloud
- Mobile Analytics – Collect, View and Export App Analytics
- SNS – Push Notification Service

Application Services
- API Gateway – Build, Deploy and Manage APIs
- AppStream – Low Latency Application Streaming
- CloudSearch – Managed Search Service
- Elastic Transcoder – Easy-to-Use Scalable Media Transcoding
- SES – Email Sending and Receiving Service
- SQS – Message Queue Service
- SWF – Workflow Service for Coordinating Application

Resource Groups — Learn more

A resource group is a collection of resources that share one or more tags. Create a group for each project, application, or environment in your account.

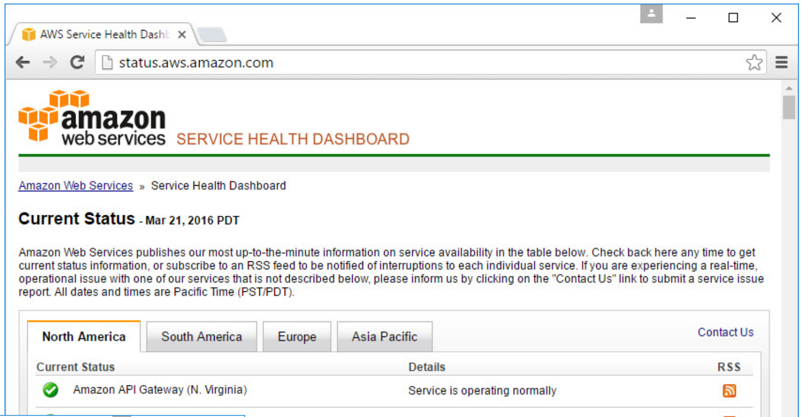
Create a Group | Tag Editor

Additional Resources

Getting Started
Read our documentation or view our training to learn more about AWS.

AWS Console Mobile App
View your resources on the go with our AWS Console mobile app, available from Amazon Appstore, Google Play, or iTunes.

AWS Marketplace
Find and buy software, launch with 1-Click and pay by the hour.



AWS Service Health Dash

status.aws.amazon.com

amazon web services SERVICE HEALTH DASHBOARD

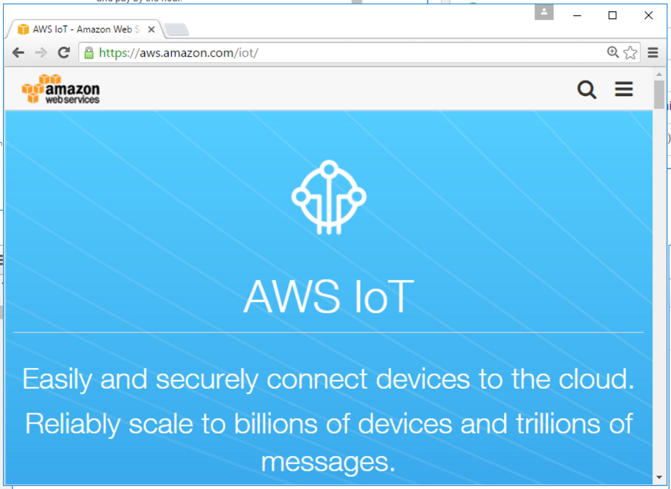
Amazon Web Services » Service Health Dashboard

Current Status - Mar 21, 2016 PDT

Amazon Web Services publishes our most up-to-the-minute information on service availability in the table below. Check back here any time to get current status information, or subscribe to an RSS feed to be notified of interruptions to each individual service. If you are experiencing a real-time, operational issue with one of our services that is not described below, please inform us by clicking on the "Contact Us" link to submit a service issue report. All dates and times are Pacific Time (PST/PDT).

North America | South America | Europe | Asia Pacific | Contact Us

Current Status	Details	RSS
✓ Amazon API Gateway (N. Virginia)	Service is operating normally	▨
	Service is operating normally	▨
	Service is operating normally	▨
	Service is operating normally	▨
ia)	Service is operating normally	▨
)	Service is operating normally	▨



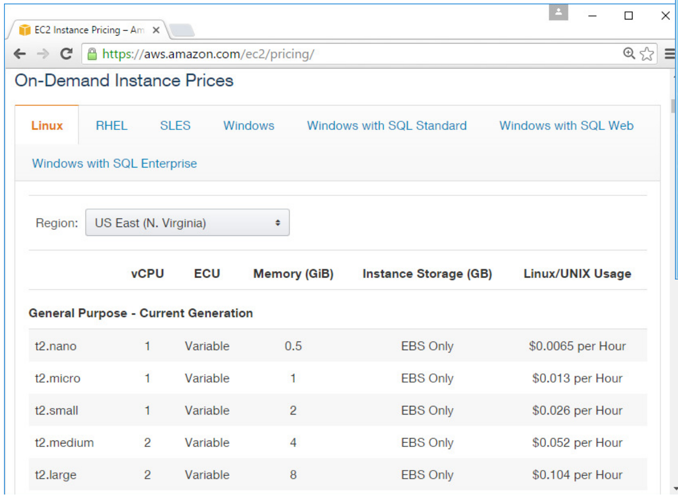
AWS IoT - Amazon Web S

https://aws.amazon.com/iot/

amazon web services

AWS IoT

Easily and securely connect devices to the cloud. Reliably scale to billions of devices and trillions of messages.



EC2 Instance Pricing – Am

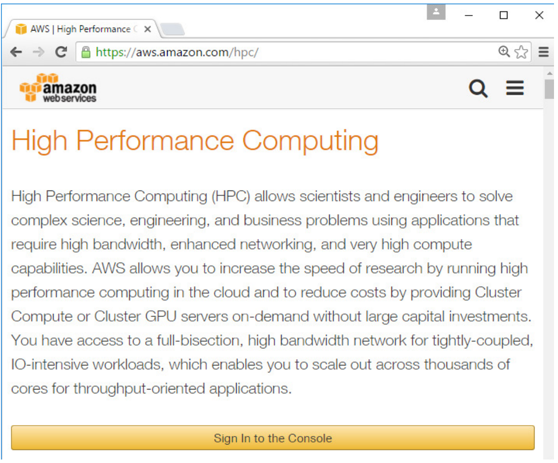
https://aws.amazon.com/ec2/pricing/

On-Demand Instance Prices

Linux | RHEL | SLES | Windows | Windows with SQL Standard | Windows with SQL Web | Windows with SQL Enterprise

Region: US East (N. Virginia)

	vCPU	ECU	Memory (GiB)	Instance Storage (GB)	Linux/UNIX Usage
General Purpose - Current Generation					
t2.nano	1	Variable	0.5	EBS Only	$0.0065 per Hour
t2.micro	1	Variable	1	EBS Only	$0.013 per Hour
t2.small	1	Variable	2	EBS Only	$0.026 per Hour
t2.medium	2	Variable	4	EBS Only	$0.052 per Hour
t2.large	2	Variable	8	EBS Only	$0.104 per Hour



AWS | High Performance

https://aws.amazon.com/hpc/

amazon web services

High Performance Computing

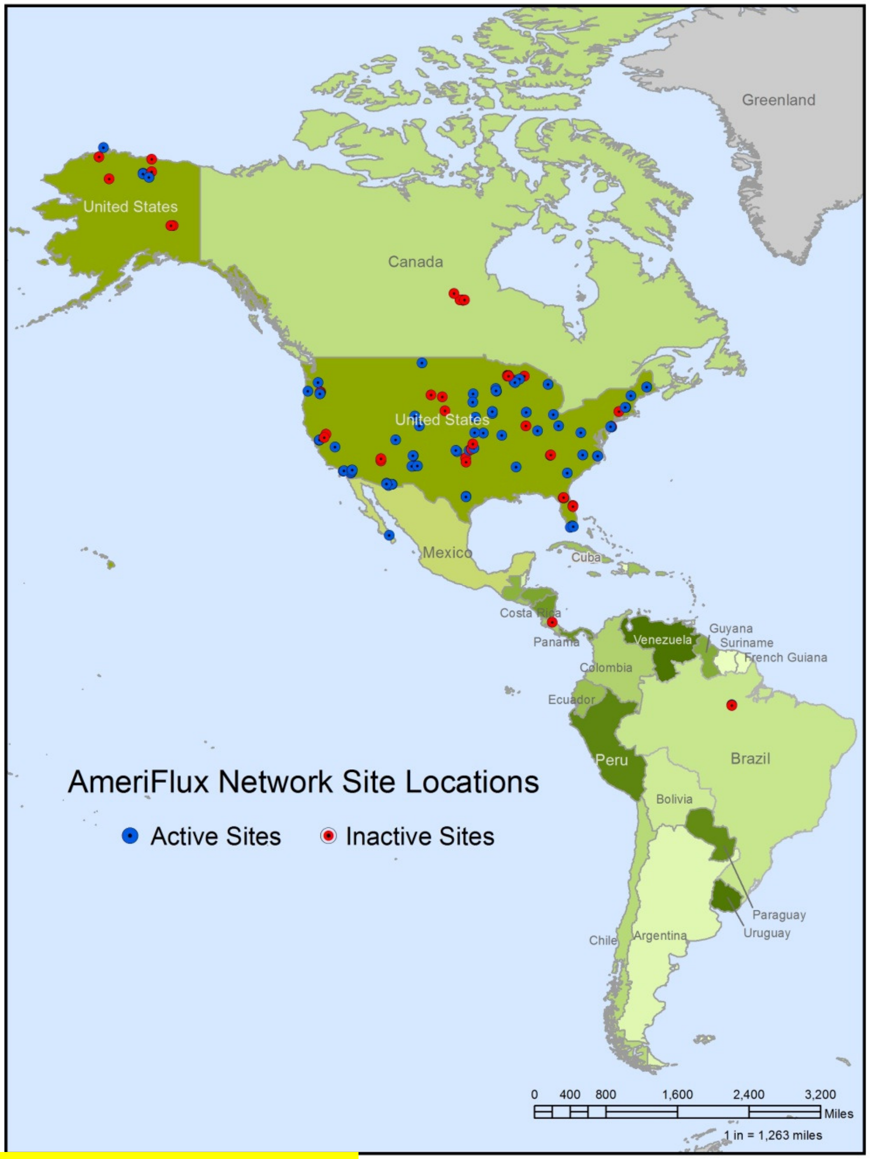
High Performance Computing (HPC) allows scientists and engineers to solve complex science, engineering, and business problems using applications that require high bandwidth, enhanced networking, and very high compute capabilities. AWS allows you to increase the speed of research by running high performance computing in the cloud and to reduce costs by providing Cluster Compute or Cluster GPU servers on-demand without large capital investments. You have access to a full-bisection, high bandwidth network for tightly-coupled, IO-intensive workloads, which enables you to scale out across thousands of cores for throughput-oriented applications.

Sign In to the Console

Many DOE applications emerging with environmental streaming data

- AmeriFlux
- NGEE Tropics
- Drone-based sensors
- Environmental monitors in cities
- Traffic sensors
- Etc.

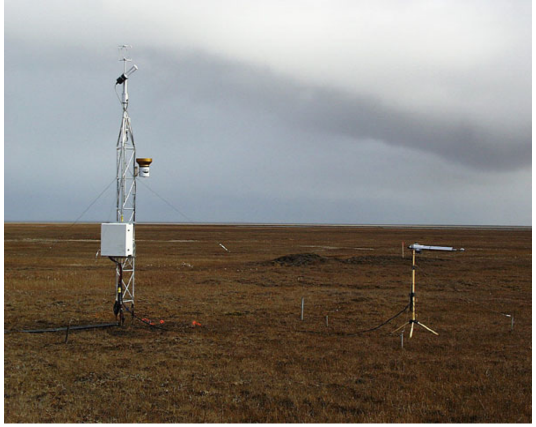
AmeriFlux, circa 2012

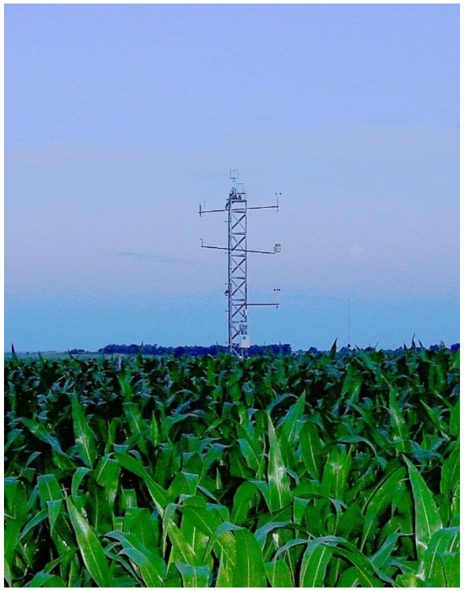


Courtesy Baldocchi et al ‘13

Science objectives

- Quantify exchange of carbon, water and energy between terrestrial ecosystems and the atmosphere across a range of vegetation types, disturbance histories, and climatic conditions.
- Understand processes governing the terrestrial carbon cycle and linkages with the water, energy and nitrogen cycles.
- Produce a high-quality data base and synthesize observations across the network.





Courtesy Davis et al '11

Core measurements

- Fluxes of CO_2, water vapor, and sensible heat flux via eddy covariance.
- Radiative fluxes and micrometeorological conditions.
- Biophysical characterization of sites (e.g. vegetation age and type, nutrient status, carbon pool sizes, soil type).

Courtesy Davis et al ‘11

AmeriFlux and Streaming Data

- Wind (direction and speed) and trace gas concentrations (mostly CO2 and H2O, but also CH4, NO, NO2, N2O, and others) are measured and stored usually at 10Hz
- Separate mechanism from “data uploads”
  - Currently only tower-driven SCP (for “high-frequency data”)
  - Currently only archival in nature
  - 35 configured; 10 active

AWS IOT

- AWS Lambda: lightweight event-driven programming
- AWS Kinesis: real-time, scalable streaming data sink
- AWS S3: scalable, reliable object store
- AWS DynamoDB: managed noSQL service
- Etc.
- *Plus any open-source projects as needed*
  - *Note to Twitter: please open-source Heron (!)*

- Example: Intel Edison-based rain sensors/gauges (UVa)

Issues

- How much streaming data is “too much” for public clouds?
- Single custom-build device (e.g., “AmeriFlux AWS IOT device”) or integration with existing infrastructure?
- How much info needed for researcher to use site’s streaming data?
- How to balance “site ownership” of streaming data vs. real-time nature of the data?
- *Large-scale software design, deployment, and management*