AWS RESEARCH ROADSHOW – 4 APRIL 2023

Solutions for Research
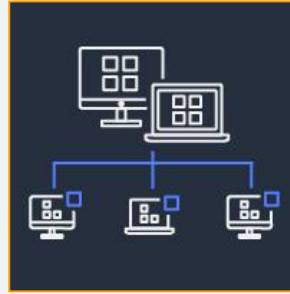
AWS Education & Research Team

Roberta Piscitelli - piscitr@amazon.com

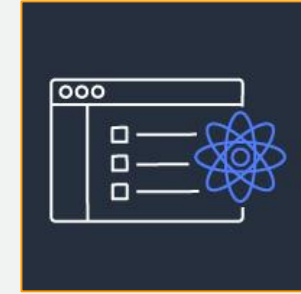# AWS provides the broadest and deepest solutions for research globally

Science, not servers
Compute when you need it at any scale

Collaboration
Access data sets that span institutions

Research Data Management
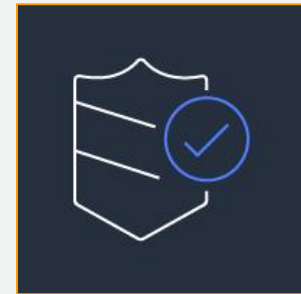Storage, secure access and management

Share and reproduce research
A common platform for reproducing scientific analyses

State-of-the-art analytics
Use data science methods in your research

Security
A collection of tools to protect data and privacy

Solutions and services for the entire research process

Grant proposals

- Research proposal support
- Letters of support
- Cloud economics
- Compliance documents

Build research

- OCRE Framework
- Deployment in minutes
- 200+ services
- Training
- Partner solutions
- Open source solutions (e.g. Parallel Cluster)
- Solution architects
- Professional services

- Single sign on
- Landing Zone

Acquire data

- AWS Snow Family
- AWS IoT for sensors
- AWS Ground Station

- Streaming data
- Event driven arch

- Databases
- Data lakes
- Data warehouses

- Bulk storage options

Analyse data

- **Interactive notebooks**
- **HPC (AWS Parallel Cluster)**
- **Machine learning**
- Containers
- Big data analytics
- Visualisation (NICE DCV)
- 400+ instance types
- GPUs, FPGAs, ARM, Inferentia, Trainium
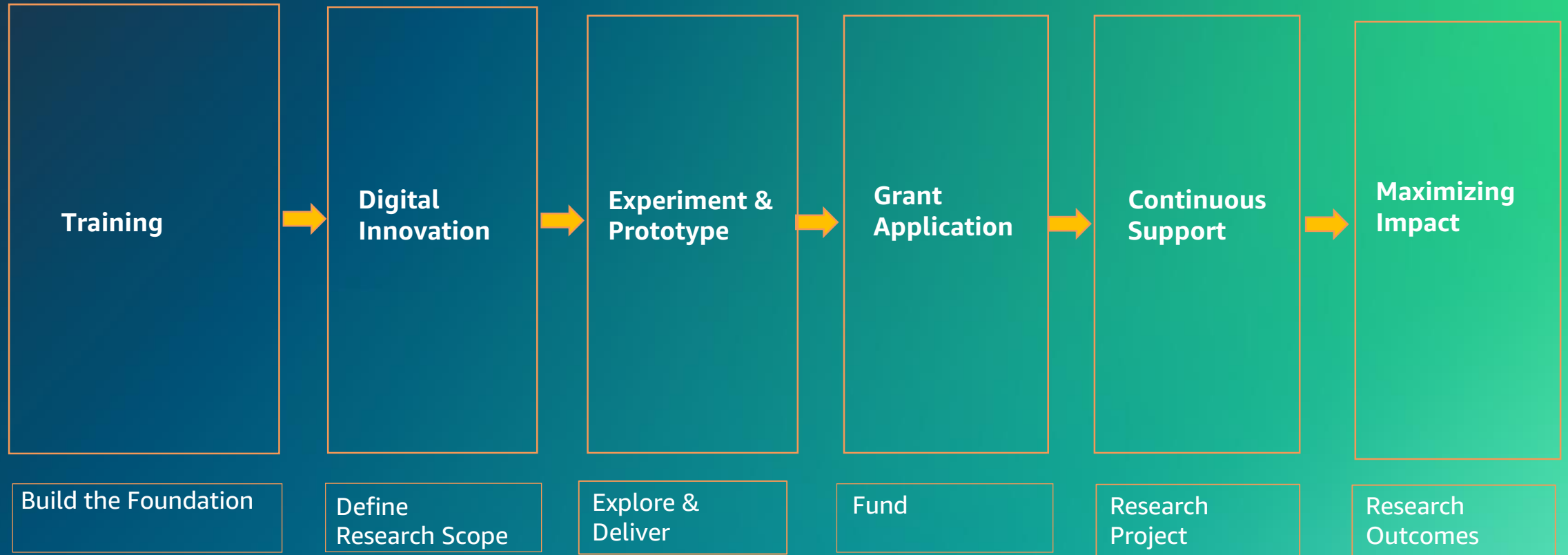- Quantum computing

Publish, share and archive results

- Open Data Registry
- Trusted Research Environments
- Content delivery network
- Long term low cost archival

Grant proposals

AWS grant acceleration support package

Training → **Digital Innovation** → **Experiment & Prototype** → **Grant Application** → **Continuous Support** → **Maximizing Impact**

Build the Foundation

Define Research Scope

Explore & Deliver

Fund

Research Project

Research Outcomes

Build research
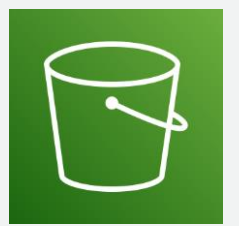
Amazon EC2

- The broadest and most in-depth computing platform

- On-demand infrastructure

- Scalable computing power

- No long-term contracts or upfront commitments

- Wide choice of operating systems and software

- Prices based on effective use

- Scalability and high performance

- Reliability and security

Generation   Characteristic

R5d.xlarge

Instance
family

Dimension
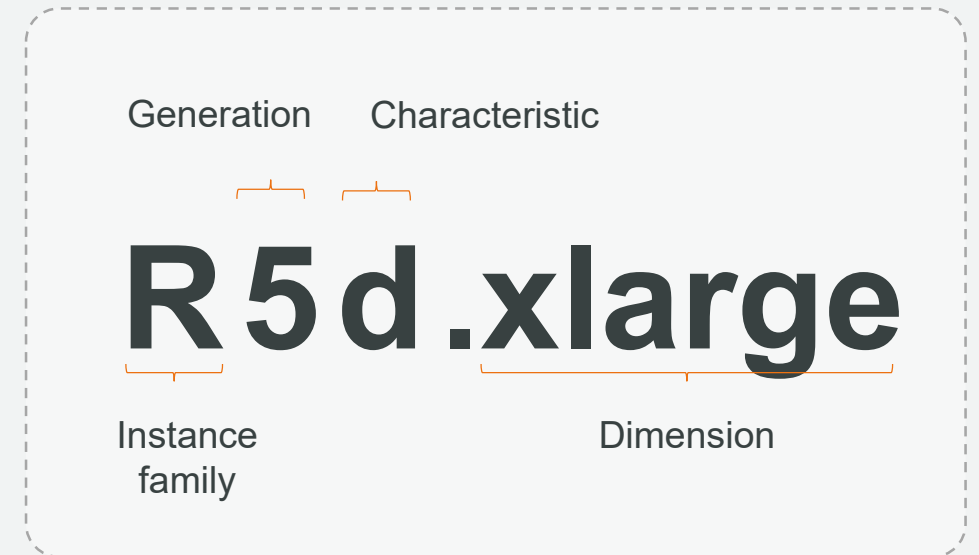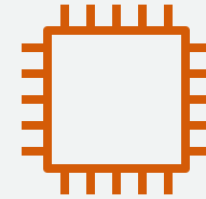
S3 Object Storage – a new kind of storage

Amazon S3

S3 Standard

S3 Intelligent-Tiering

S3 Standard-IA

S3 One Zone-IA

S3 Glacier Instant Retrieval

S3 Glacier Flexible Retrieval

S3 Glacier Deep Archive

Frequent ← Access Frequency → *Infrequent*

The Italian National Institute of Astrophysics Explores the Universe with the Cloud
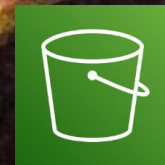
"Thanks to AWS, we were able to concentrate on science and simulations.
We were able to scale as soon as the project required us to do so. It was critical to obtain the required power quickly"

Marco Landoni, INAF researcher
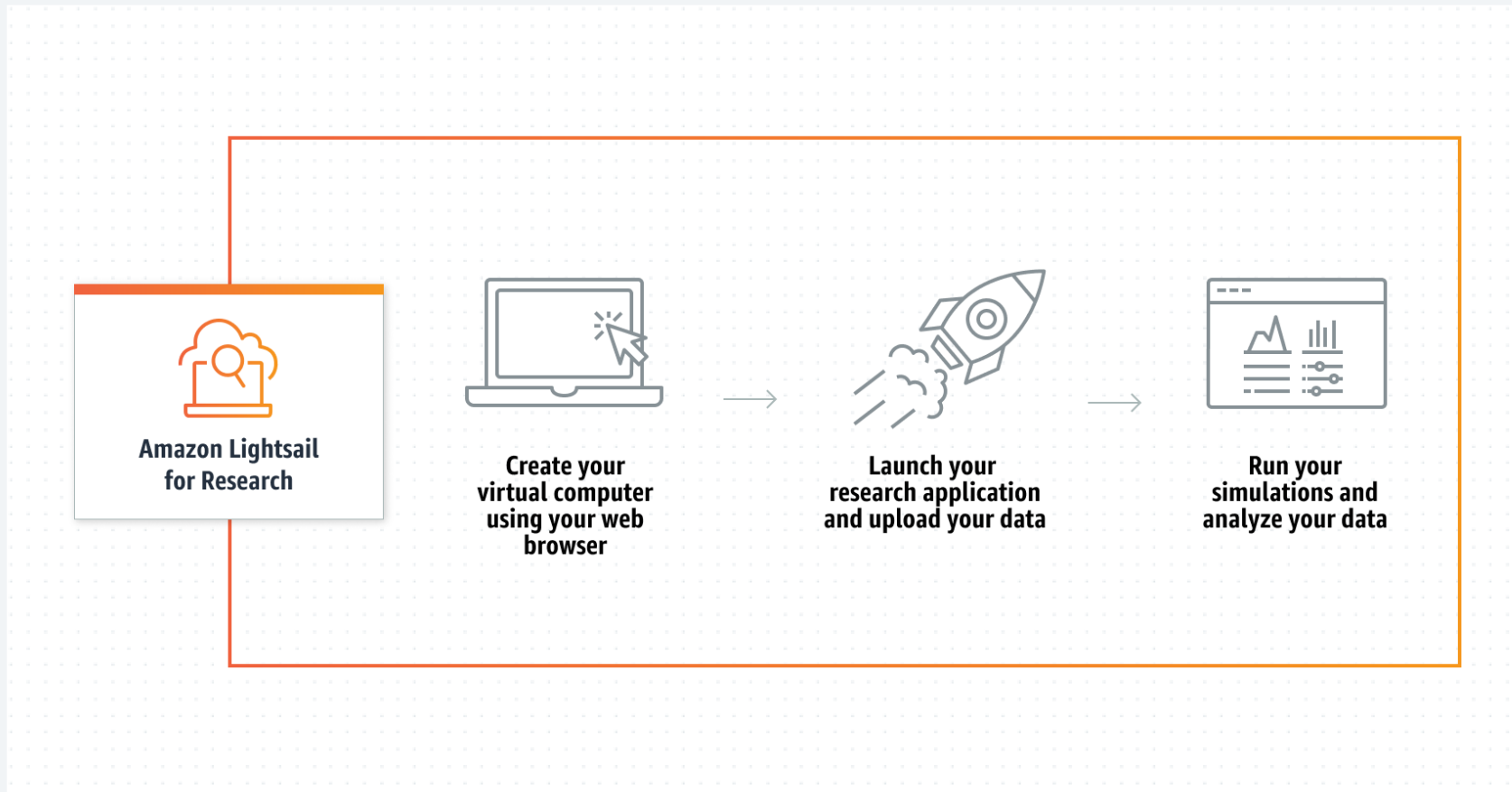
AWS Lambda

Amazon Simple Storage Service (Amazon S3)

Amazon Simple Storage Service Glacier

Amazon Elastic Compute Cloud (Amazon EC2)

Lightsail for Research

Amazon Lightsail for Research

Create your virtual computer using your web browser → Launch your research application and upload your data → Run your simulations and analyze your data

Lightsail for research

IT resource, support, and research enablement <u>focus</u>

Computing need

Lightsail for Research

Self-service solution for IT support and research enablement

Well resourced and supported

Under resourced and supported

Researchers

Acquiring data

Edge to cloud architecture

IoT devices

- AWS IoT Greengrass
- IoT
- FreeRTOS

5G devices

- AWS Wavelength

Rugged environment

- AWS Snowball Edge
- AWS Snowcone

PoP

Industrial devices

- AWS IoT SiteWise
- AWS IoT Greengrass
- AWS RoboMaker

PoP

On-premises facility

- AWS Outposts
- AWS Storage Gateway
- AWS Snowcone
- AWS Snowball Edge

PoP

PoP

AWS Cloud

Local Zone

Region
- AZ
- AZ
- AZ

Region
- AZ
- AZ
- AZ

Region
- AZ
- AZ
- AZ

aws

Edge to cloud architecture

IoT devices

5G devices

Rugged environment

Industrial devices

PoP

On-premises facility

PoP

PoP

PoP

AWS Cloud

24 Regions   **216** PoPs
76 AZs   **205** Edge locations

Edge use cases

- Use cases across various industries require cloud computing capabilities, but also need local processing for low latency and to reduce data transfer.

- These use cases are often in environments where;
    - data generation is decentralized,
    - data volumes are significant, and
    - network connectivity is either consistent, intermittent or inaccessible

Smart Island: How the AWS Cloud is powering social, economic, and environmental improvements in Greece

Smart Island introduces solutions for mobility, primary healthcare, and transport of goods —running on the AWS Cloud—while 'smartening' existing infrastructure, such as tourist ports, the energy grid, waste, and water management.

At the same time, public officials, civil servants, and residents will be offered tailor-made training to improve their digital skills to take advantage of the new opportunities the technology presents

AWS IoT

https://aws.amazon.com/blogs/publicsector/smart-island-how-aws-cloud-powering-social-economic-environmental-improvements-greece/

# Stanford Archeologists Enable Global Collaboration Using AWS

> The AWS cloud gives our researchers faster access to cleaner data so they can collaborate more effectively on new research questions and arrive at more cohesive interpretations of data.

Lindsay Der
Ph.D. candidate, Stanford Archeology Center,
Stanford University

Lo Stanford Archaeology Center conduce ricerche interdisciplinari in tutto il mondo. Ha sede presso il campus della Stanford University in California.
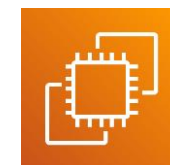
https://aws.amazon.com/solutions/case-studies/stanford-archaeology-center/
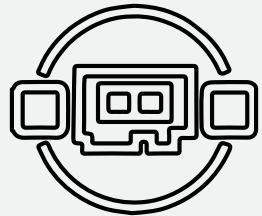
esri

Amazon Elastic Compute Cloud (Amazon EC2)
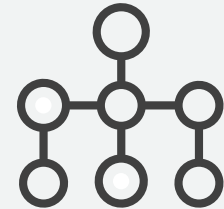
Data analysis

Services to enable HPC on AWS
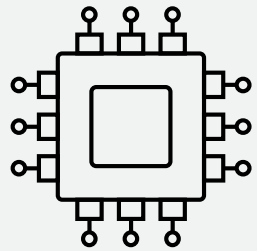
Amazon EC2

Elastic Fabric
Adapter (EFA)
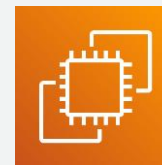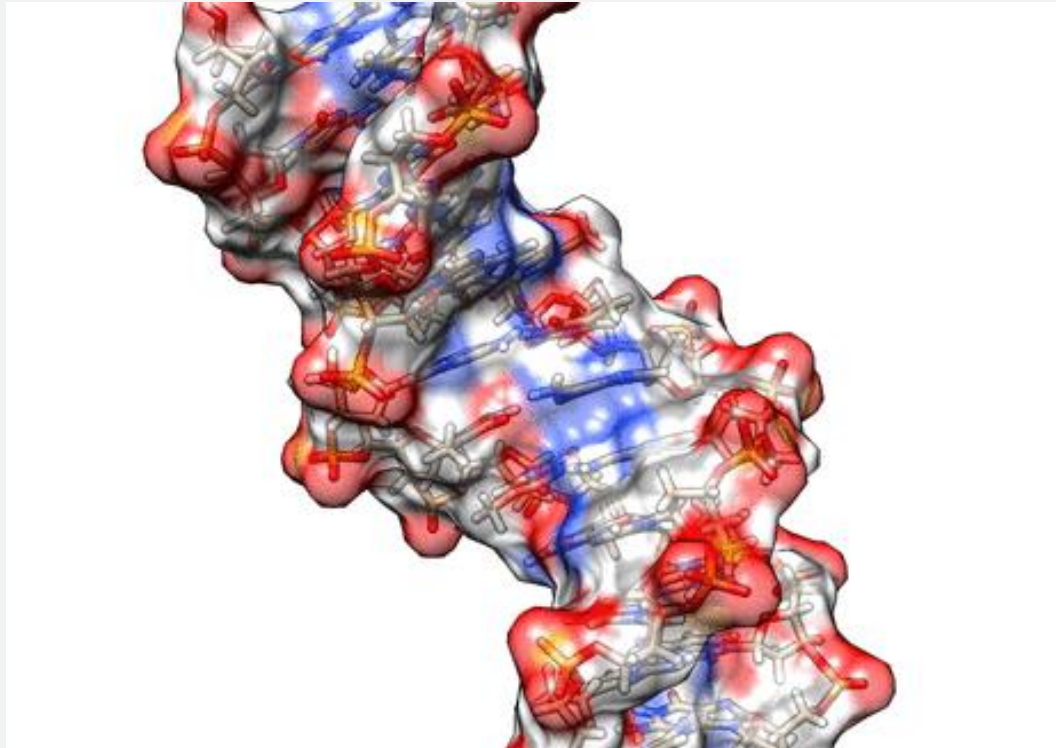+
up to 400 Gbps
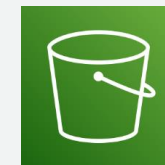networking

Amazon FSx
for Lustre

AWS Batch

The-university of Nottinghams Crossbow project paves a new path for biomolecular research using high performance computing HPC and the cloud

University of Nottingham
UK | CHINA | MALAYSIA
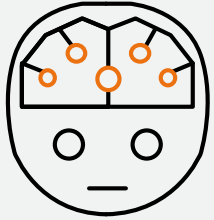
Amazon
Elastic Compute Cloud
(Amazon EC2)

Amazon
Simple Storage Service
(Amazon S3)

Amazon
Elastic File System
(Amazon EFS)

Artificial intelligence in the cloud

Artificial intelligence (AI)
Any technique that enables computers to mimic human intelligence using logic, if-then statements, and machine learning

Machine learning (ML)
A subset of AI that uses machines to search for patterns in data to build logic models automatically
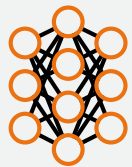
Deep learning (DL)
A subset of ML composed of deeply multi-layered neural networks that perform tasks like speech and image recognition

Reinforcement learning (RL)
Reinforcement learning trains models by using a reward function of a desired outcome

Generative AI (Foundation Models) refers to artificial intelligence that can generate novel content

AI that can produce original content close enough to human generated content for real-world tasks

Powered by foundation models pre-trained on large sets of data with several hundred billion parameters
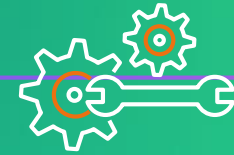
Tasks can be customized for specific domains with minimal fine-tuning

Applicable to many use cases like text summarization, question answering, digital art creation, code generation, etc.
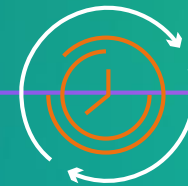
Reduces time and cost to develop ML models and innovate faster

How Generative AI transforms artificial intelligence
image generation, transformation, upscaling

Generated by Stable Diffusion 2.0. This interior does not exist

Seamless transformation

4x

Upscaling

Amazon SageMaker

The AI/ML stack

AI SERVICES

HEALTH AI

Amazon HealthLake

Amazon Transcribe for Medical

Amazon Comprehend for Medical

INDUSTRIAL AI

AWS Panorama + Appliance

Amazon Monitron

Amazon Lookout for Equipment

Amazon Lookout for Vision

ANOMALY DETECTION

Amazon Lookout for Metrics

CODE AND DEVOPS

Amazon DevOps Guru

Amazon CodeGuru

VISION

Amazon Rekognition

SPEECH

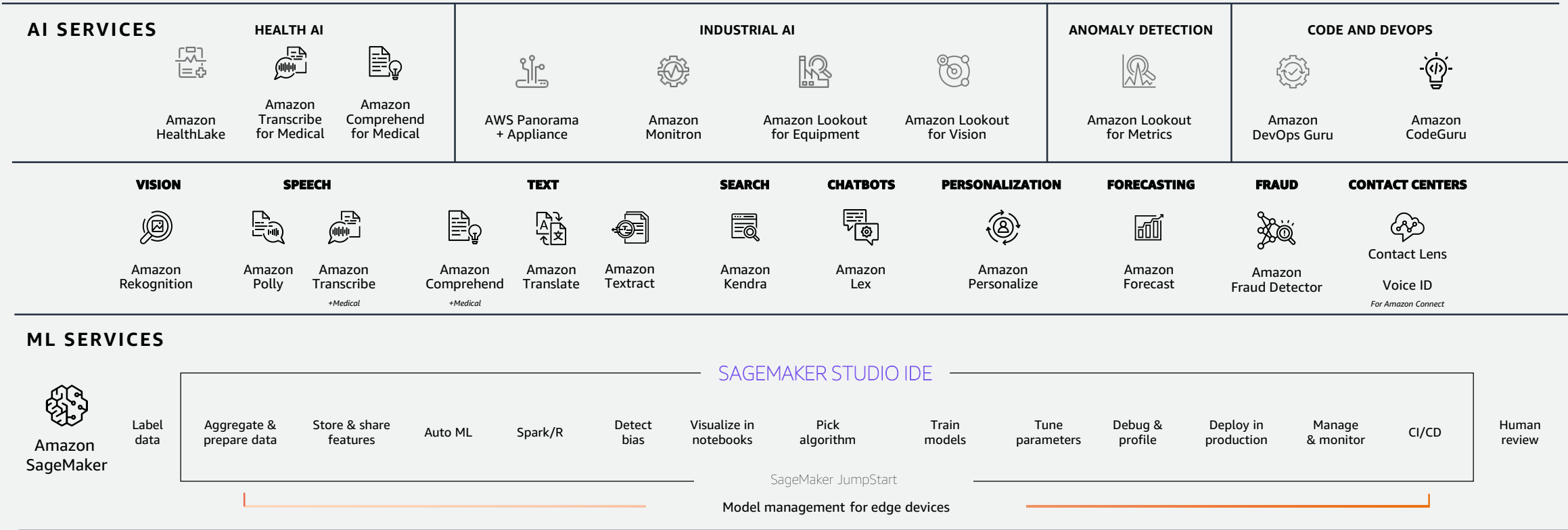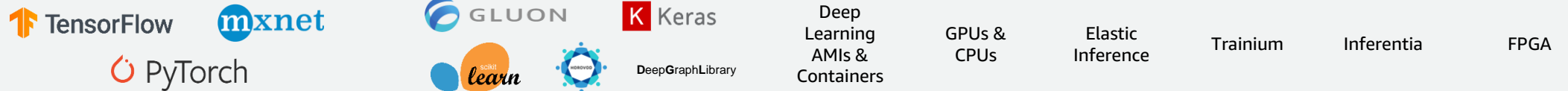Amazon Polly

Amazon Transcribe
+Medical

TEXT

Amazon Comprehend
+Medical

Amazon Translate

Amazon Textract

SEARCH

Amazon Kendra

CHATBOTS

Amazon Lex

PERSONALIZATION

Amazon Personalize

FORECASTING

Amazon Forecast

FRAUD

Amazon Fraud Detector

CONTACT CENTERS

Contact Lens

Voice ID
For Amazon Connect

ML SERVICES

Amazon SageMaker

Label data

SAGEMAKER STUDIO IDE

Aggregate & prepare data

Store & share features

Auto ML

Spark/R

Detect bias

Visualize in notebooks

Pick algorithm

Train models

Tune parameters

Debug & profile

Deploy in production

Manage & monitor

CI/CD

Human review

SageMaker JumpStart

Model management for edge devices

INFRASTRUCTURE

TensorFlow

mxnet

GLUON

Keras

PyTorch

scikit learn

HOROVOD

DeepGraphLibrary

Deep Learning AMIs & Containers

GPUs & CPUs

Elastic Inference

Trainium

Inferentia

FPGA

SageMaker Canvas

Collaborate and send models to data scientists using Amazon SageMaker Studio for review and feedback

Amazon SageMaker Canvas
Build machine learning models without writing code

Browse, import, and join data
Import data from disparate sources and create new unified datasets for training prediction models

Pick target
Select values you want to predict

Prepare and analyze data
Automatically detect errors, cleanse, and analyze data to help determine if your data is prepared for ML

Create model
Create ML models with the click of a button

Generate and understand predictions
Generate single or bulk predictions and understand predictions

University of Oxford Introduces a Sector-Leading Image Recognition ML Prototype to Augment Digitization in Numismatics

https://aws.amazon.com/solutions/case-studies/oxford-case-study/

First stage
Improve base image quality

Second stage
Visually search for items in collection

UNIVERSITY OF OXFORD

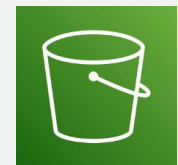> **I thought this project would be complex and time consuming, but using AWS made it easy.**

Anjanesh Babu
Systems architect and network manager, Gardens and Museums IT, University of Oxford's Gardens, Libraries & Museums

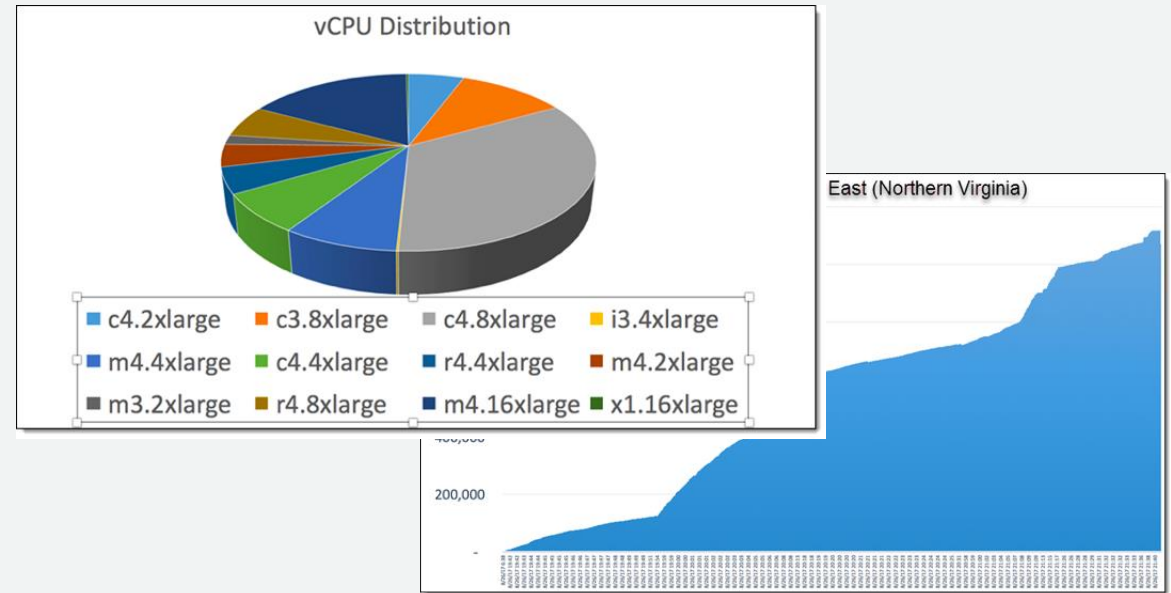Amazon SageMaker
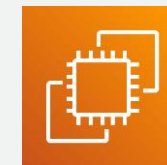
Amazon EC2

Amazon S3

Clemson University – Natural Language Processing

The researchers conducted nearly half a million topic modeling experiments to study how human language is processed by computers.

The 1.1 Million vCPU count usage is comparable to the core count on the largest supercomputers in the world.

vCPU Distribution

c4.2xlarge · c3.8xlarge · c4.8xlarge · i3.4xlarge
m4.4xlarge · c4.4xlarge · r4.4xlarge · m4.2xlarge
m3.2xlarge · r4.8xlarge · m4.16xlarge · x1.16xlarge

East (Northern Virginia)

400,000
200,000

"*I am absolutely thrilled with the outcome of this experiment. The graduate students on the project […] used resources from AWS and Omnibond and developed a new software infrastructure to perform research at a scale and time-to-completion not possible with only campus resources.*"

– Prof. Amy Apon, Co-Director of the Complex Systems, Analytics and Visualization Institute

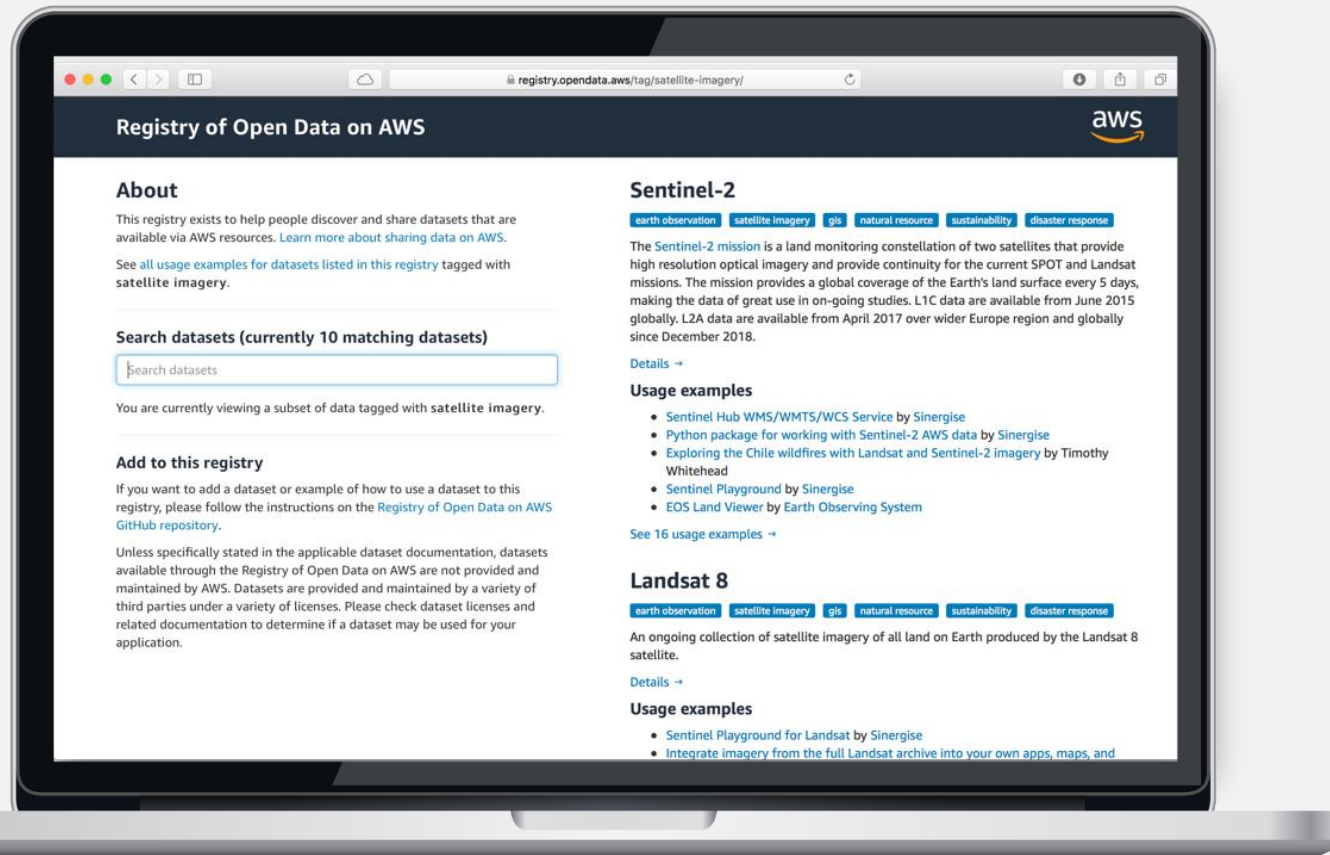Amazon EC2
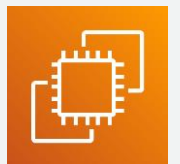
Publishing and sharing research data

Open data on AWS

AWS hosts a variety of public datasets to lower the cost and improve the speed of research.

https://registry.opendata.aws/

Esempi:

- 1000 Genomes Project
- The Cancer Genome Atlas
- International Cancer Genome Consortium
- Landsat 8
- Common Crawl
- SpaceNet
- OpenStreetMaps

... Aggiornati regolarmente

Monitoring at-risk bodies of water from space

The Bluedot Observatory uses Sentinel-2 satellite data on AWS to monitor water bodies around the world

"The cost to process one month of data for about 7,000 bodies of water currently in the system is 6 EUR. It is possible to set up world-scale systems with a shoestring budget."

Grega Milcinski, Bluedot

opendata.aws/bluedot

The University of Chicago Manages Biomedical Data Resources at Scale on AWS

THE UNIVERSITY OF CHICAGO | Center for Translational Data Science

"
The data services developed by the center are now called Data Commons Framework Services and are hosted on AWS. They make all the data findable, accessible, interoperable, and reusable in accordance with the FAIR data principles.
"
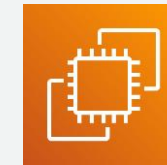
Robert L. Grossman
Professor of Medicine and Computer Science and Director of the Center for Translational Data Science, University of Chicago
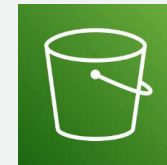

Amazon RDS   Amazon EKS   Amazon EC2   Amazon S3

University of Dundee – Trusted Research Environment and Enclave for Hosting Open Original Science Exploration (TREEHOOSE)

> "The AWS ProServe team were really engaged and invested in making the collaborative project a success. Any issues were resolved with little fuss and the project objectives were delivered well within the required timeframe. Great team"

-Dr Christian Cole, Senior Lecturer
School of Medicine, University of Dundee

Amazon Trusted Research Environment

Thank you!

AWS Education & Research Team

Roberta Piscitelli - piscitr@amazon.com