Jak przyspieszyć i ułatwić badania naukowe z wykorzystaniem usług chmurowych Amazon Web Services (AWS)

Łukasz Magiera

Sr. Solutions Architect
AWS

Breadth and Depth of Services: 200+ fully featured services

Analytics

Analytics
Data Exchange
Data Lake
Data Pipelines
Data Warehouse
Elasticsearch
Streaming
ETL
Hadoop / Spark
Interactive SQL Queries
Visualizations

Business Applications

Contact Center
Sharing & Collaboration
Online Meetings & Chat
Voice-Enabled Workplace
Unified Communications
Mobile & Web Apps Without Programming

Blockchain

Blockchain Templates
Ledger Database
Managed Blockchain

Security, Identity, and Compliance

Access Control
Assessment & Reporting
Configuration Compliance
Data Protection
DDOS Protection
Identity Management
Key Management & Storage
Monitoring & Logging
Resource Management
Threat Detection
Web Application Firewall

Storage

Archive Storage
Backup & Restore
Block Storage
Data Transfer
Edge Processing & Computing
File Storage
High-Performance File System
Hybrid Cloud Storage
Object Storage
Unified Communications
Windows File System

Database

Aurora, a high-performance, relational database Built for the Cloud

Managed SQL Server
Managed MariaDB
Managed MySQL
Managed Oracle
Managed PostgreSQL
Document Database
Graph Database
In-Memory Caching
Key-Value Store Database
Ledger Database
Time Series Database

Development Tools

Analyze and Debug
Application Lifecycle Management
Authoring
Build & Test
Containers
DevOps Resource Management
One-Click App Development
Patching
Pipeline Orchestration
Resource Templates
Triggers

Compute

Compute
Auto-Scaling
Batch Jobs
Event-Driven Serverless Computing
Instance Types
Managed Virtual Private Servers
Managed Repository for Serverless Apps
Run & Manage Web Apps
Serverless Compute
Virtual Servers
Containers
Container Service
Managed Kubernetes
Store & Retrieve Docker Images

Media Services

Live Video Transport
Media Storage
Transcoding
Video Origination & Packaging
Video Personalization & Monetization
Video Processing & Delivery
Video Streaming Analysis

Hybrid Architecture

AWS Services on Premises
Data Integration
Integrated Devices & Edge Systems
Integrated Identity & Access
Integrated Networking
Integrated Resource & Deployment Management
VMWare Cloud on AWS
Integrated SG

Internet of Things (IOT)

Rules Engine
Device Analytics
Device Gateway
Device SDK
Device Shadows
Event Detection & Response
Local Compute
Local Data Collection
Management & Security
Microcontroller Operating System
Registry
Visual Applications Development

Machine Learning (ML)

ML Frameworks
Deep Learning AMIs & Containers
Hardware Acceleration
ML at the Edge
TensorFlow, PyTorch, MXNet
Sagemaker
Automatic Model Tuning
Data Labeling
Hosted Notebooks
ML Marketplace
Model Hosting
Model Optimization
Model Training
Pre-Built Algorithms
Topic Modeling
Deep Learning Models
Reinforcement Learning
Spot Instances
Batch Predictions
Real-Time Predictions
AI Services
Chatbots
Entity Extraction
Face Analytics
Face Search Forecasting
Image Labeling
Natural Language Processing
Personalization & Recommendation
Sentiment Analysis
Speech Translation
Text & Data Extraction
Text to Speech Translation
Video & Image Analysis
Content Moderation

AWS accelerates research

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

Collaboration
Access datasets that span institutions and borders

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

Enterprise IT and Research

Enterprise IT cares about…

Stability & Operational Excellence

Cost Optimization

Security & Compliance

Research cares about…

New Ideas & Innovation

Speed & Agility

Value & Results

Research is not one size fits all

Researcher Workspaces

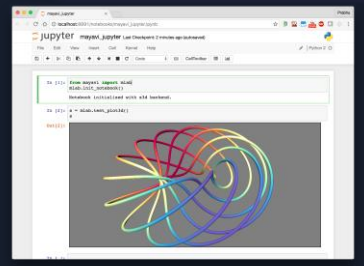
Research Gateway

RONIN

AWS Research & Engineering Studio

Solutions & Guidance

Amazon SageMaker

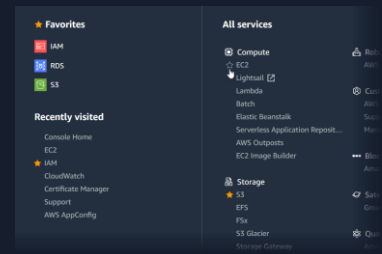


AWS ParallelCluster

aws thinkbox

Native AWS

AWS Console, CLI, SDKs, etc.



Storage and data solutions to meet the specific researcher needs

Researcher Technical Requirements & Capability

Simpler/Common Requirements

Complex/Niche Requirements

AWS delivers broadest storage portfolio in industry

OBJECT

Amazon S3

BLOCK

Amazon EBS

FILE

Amazon EFS

Amazon FSx for NetApp ONTAP

Amazon FSx for Windows File Server

Amazon FSx for Lustre

Amazon FSx for OpenZFS

Amazon File Cache

BACKUP

AWS Backup

DATA TRANSFER AND MIGRATION

AWS Storage Gateway

AWS DataSync

AWS Transfer Family

AWS Snowball

AWS Snowcone

Broadest and deepest platform choice

Categories

General purpose

Burstable

Compute intensive

Memory intensive

Storage (High I/O)

Dense storage

GPU compute

Graphics intensive

Capabilities

Choice of processors
(AWS Graviton, Intel, AMD)

Fast processors
(up to 4.5 GHz)

High memory footprint
(up to 24 TiB)

Instance storage

Accelerated computing
(GPUs ,FPGA & ASIC)

Networking

Bare Metal

Size
(Nano to 48xlarge)

Options

Elastic Block Store (EBS)

Elastic Fabric Adapter

Linux, Unix, Windows, macOS

800+

instance types

for virtually every workload and business need

Broadest and deepest set of relational and purpose-built databases

RELATIONAL

Amazon Aurora

Amazon RDS

PURPOSE-BUILT

KEY-VALUE

Amazon DynamoDB

CACHING

Amazon ElastiCache

DOCUMENT

Amazon DocumentDB

GRAPH

Amazon Neptune

MEMORY

Amazon MemoryDB

WIDE-COLUMN

Amazon Keyspaces

TIME-SERIES

Amazon Timestream

Purchase options to optimize Compute costs

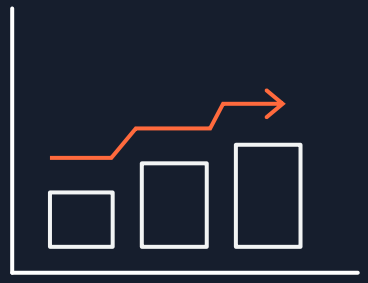
On-Demand

Pay-for-what you use with no long-term commitments

Stateful spiky workloads

Savings Plans

Up to **72%** savings for 1 or 3 year hourly usage commitments



Committed & steady-state usage

Spot

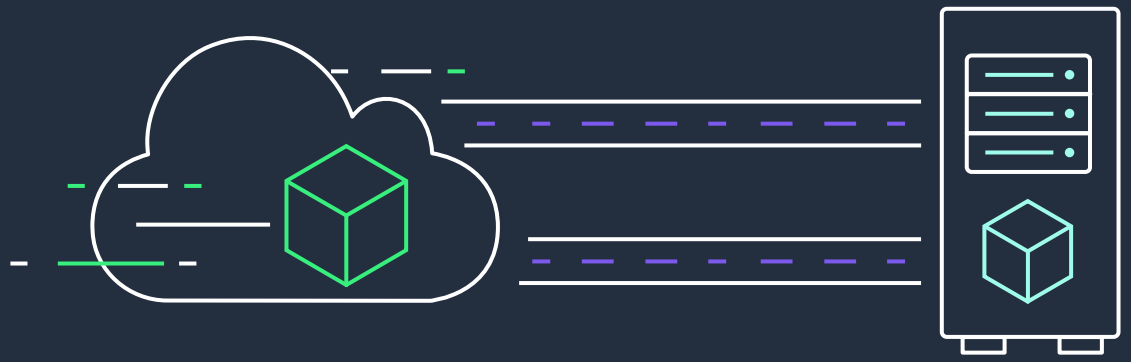
Spare capacity at up to **90%** off On-Demand prices

Fault-tolerant, flexible, stateless workloads

The best practice is to combine all three purchase options

AWS ParallelCluster

One-stop shop to set up your HPC cluster



Integrated with AWS services you need

- Highly-performant file systems
- Amazon EC2 instances
- EFA
- NICE DCV

AWS Batch

Job scheduler

- Schedules and runs jobs asynchronously
- Manages dependencies

Resource orchestrator

- Manages and optimizes compute resources
- Scales up/down as needed
- Utilizes the right compute resources for the job

Fully managed

Integrated with AWS services

Massive scalability

Optimized resource provisioning

Cost efficient

aws

Customer stories

Research customers

SURF Research Cloud brings data power to science using AWS

SURF—the National Research and Education Network (NREN) in the Netherlands—wanted to create a secure, reliable research and data sharing platform that would be compliant with data privacy laws across international borders and offer different levels of access.

It used Amazon EC2 for secure, resizable compute power and AWS Control Tower to set up and govern secure, multi-account AWS environments with different levels of access and functionality that would also be GDPR compliant.

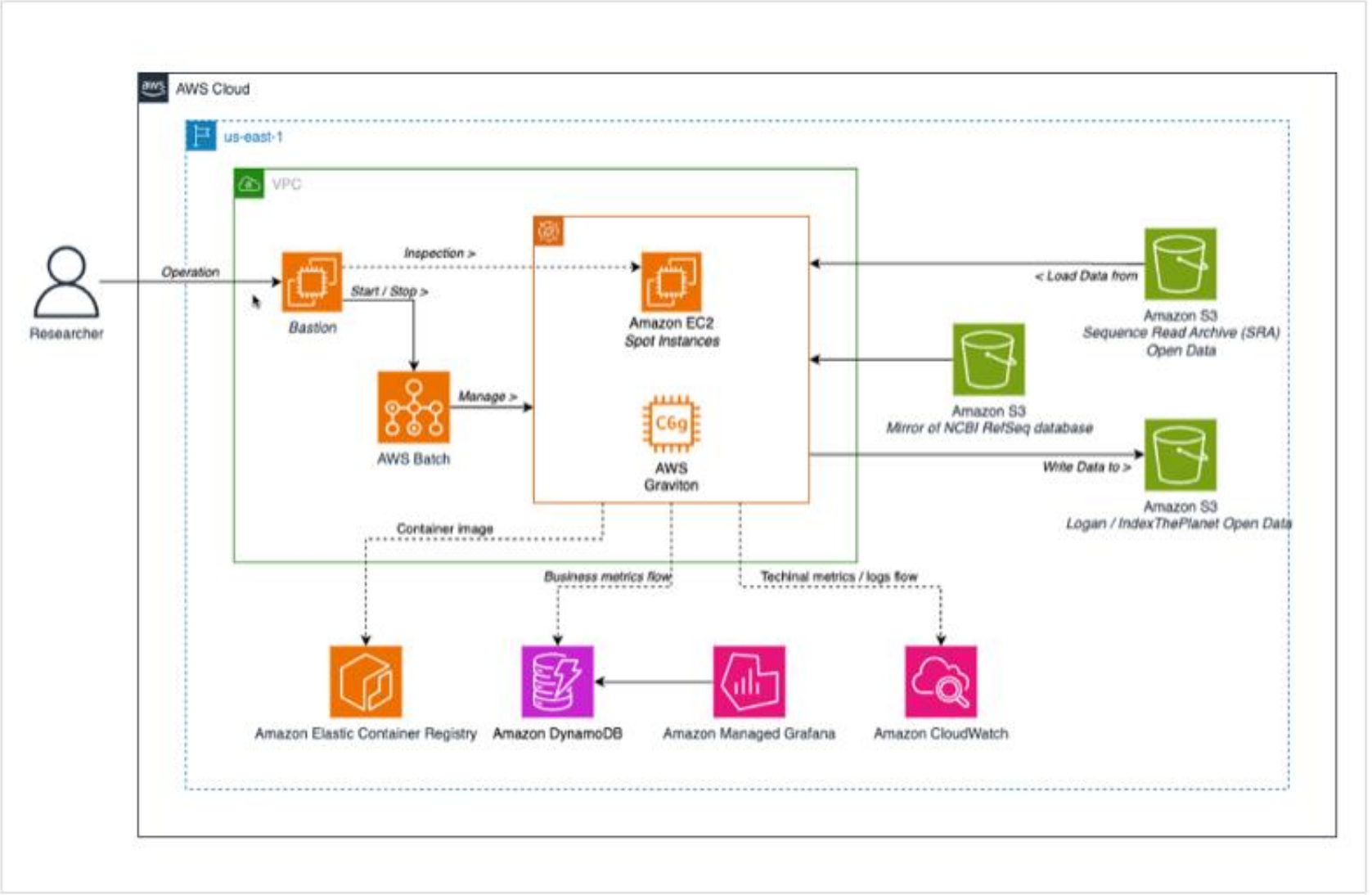
The SURF Research Cloud provides security and GDPR compliance without compromising mobility of data. It supports multi-region networks and offers multiple interfaces that allow researchers to select the right system for their work using mature and stable AWS services.

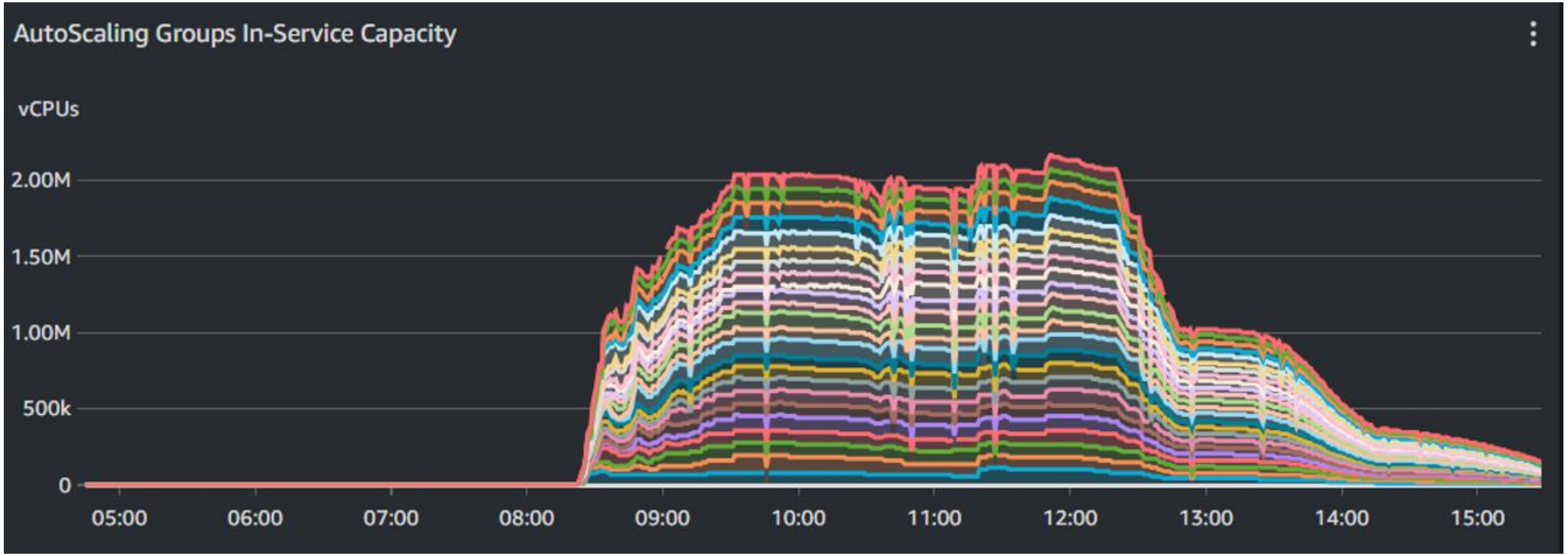
“Success isn’t just a question of technology. It can’t happen without compliance, which is a dimension that researchers often struggle with. We overcame that by using cloud native technologies from AWS.”

Ivar Janmaat

Team Lead, High Performance Compute Cloud, SURF

SURF





The Institut Pasteur and AWS are analysing the world's DNA, using a public database

Customer Stories / Life Sciences / France

2024

INSTITUT PASTEUR

The Institut Pasteur and AWS are analysing the world's DNA, using a public database

Institut Pasteur, a leading French virology research center, processed 20 petabytes of DNA data in record 30 hours, leveraging AWS Batch over a cluster of 2.18M AWS Graviton cores.

30 hours

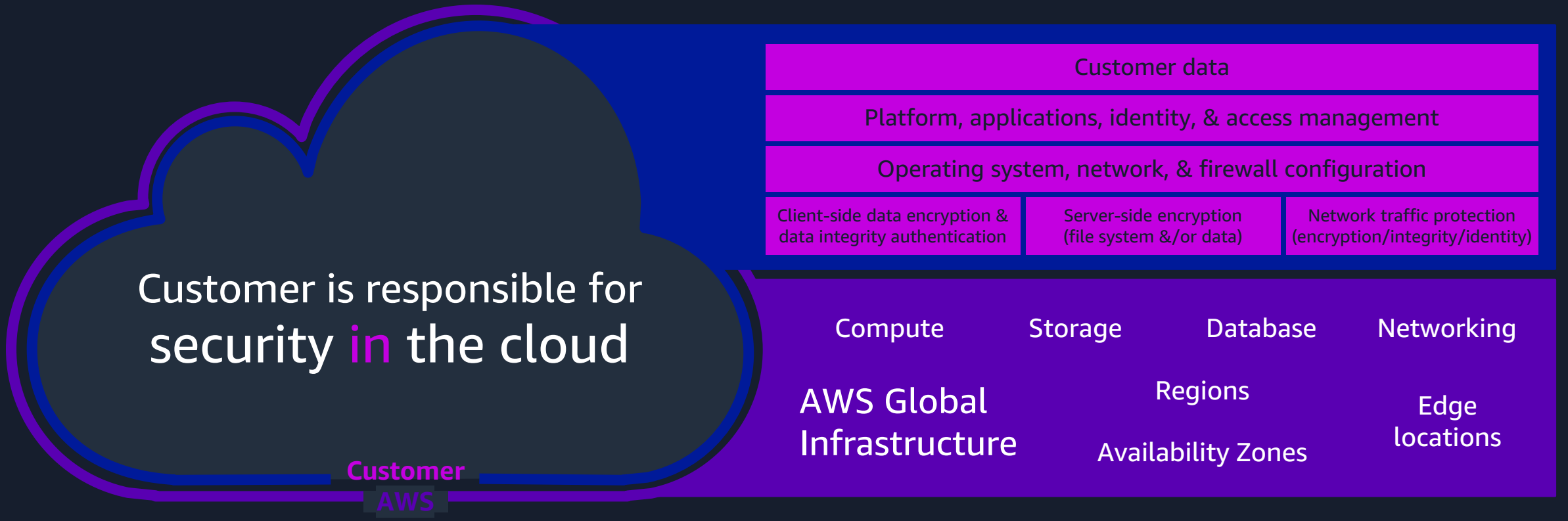
Reduce the computing time required of 30 million vCPU hours to 30 hours, with 2.18 million vCPUs mobilised at peak

20-petabyte

First exhaustive use of a 20-petabyte DNA database

Secure research

Shared security responsibility model



Identity and access management

AWS Identity and Access Management (IAM)

AWS IAM Identity Center

AWS Organizations

AWS Directory Service

Amazon Cognito

AWS Resource Access Manager

Amazon Verified Permissions

Detection and response

AWS Security Hub

Amazon GuardDuty

Amazon Security Lake

Amazon Inspector

Amazon Macie

Amazon Detective

Amazon CloudWatch

AWS Config

AWS CloudTrail

Network and app protection

AWS Firewall Manager

AWS Network Firewall

AWS Shield

AWS WAF

Amazon VPC

AWS PrivateLink

AWS Systems Manager

AWS Verified Access

Data protection

Amazon Macie

AWS Key Management Service (KMS)

AWS CloudHSM

AWS Certificate Manager

AWS Private CA

AWS Secrets Manager

AWS Payment Cryptography

Server-Side Encryption

Compliance

AWS Artifact

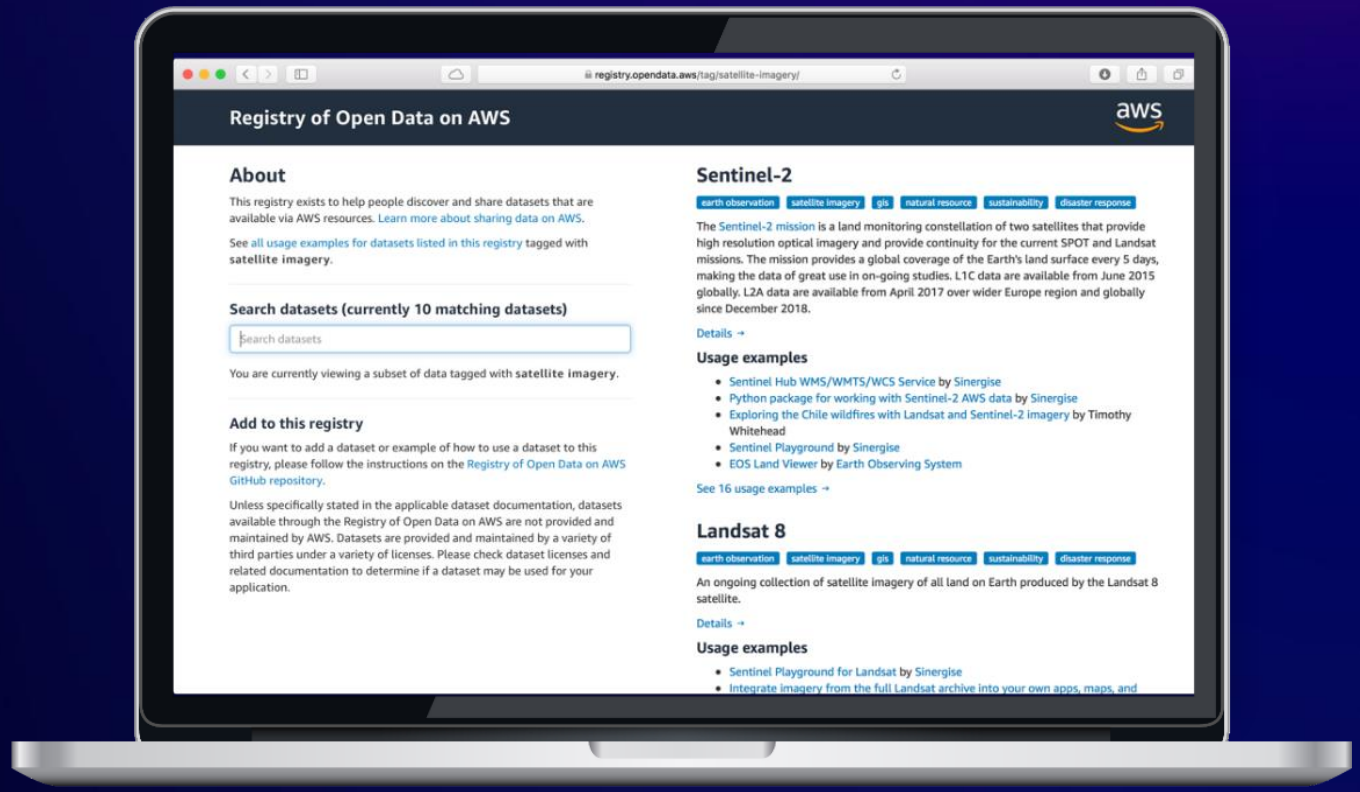
AWS Audit Manager

Open data on AWS

AWS HOSTS PUBLIC DATASETS TO LOWER THE COST AND IMPROVE THE SPEED OF RESEARCH.

Examples include:

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



No-cost Research Learning Pathway

Provides a structured, curated list of online courses tailored to researchers' needs

AWS Skill Builder provides 500+ no-cost digital courses