OFFICE HOURS
Cloud Computing with AWS
10.04.2023
Who is supporting you today?

Dunieski Otano
Solutions Architect

dootano@

Ezzel Mohammed
Solutions Architect

ezzel@
AGENDA

• What is the AWS Cloud?
• AWS Global Infrastructure
• Data Analytics Services
• AI/ML Services
• AWS Management Console
• 3 Ways to Access to AWS Services
• Access during Datathon
• Open Data Registry
• Tutorial Recommendations
What is the AWS?
What is AWS?

• AWS provides a highly reliable, scalable, low-cost infrastructure platform in the cloud that powers millions of businesses in over 245 countries and territories around the world.

• Benefits
  • Low Cost
  • Elasticity & Agility
  • Open & Flexible
  • Secure
  • Global Reach
Support virtually any cloud workload

- 200+ Fully feature services
- 600+ types of compute instances

**Core Services**
- Compute - Storage - Networking

**Infrastructure**
- Regions - Availability Zones

**Security and Compliance**
- Management Tools
AWS Global Infrastructure
Availability Zones

Each AWS Region consists of multiple, isolated, and physically separate AZs within a geographic area.

An Availability Zone (AZ) is one or more discrete data centers with redundant power, networking, and connectivity in an AWS Region.

High throughput, low latency (< 10 ms) network between Availability Zones.

All traffic between AZs is encrypted.

Physical separation with 100 km (60 miles).
Global Presence
102 Availability Zones
32 Regions

Planned expansion
+12 Availability Zones
+ 4 Regions

550+ Points of Presence
13 Regional Edge Caches
Data Analytics Services
Modern Data Analytics Reference Architecture on AWS

This architecture enables customers to build modern data analytics pipeline using the Lake House approach to derive insights from the data.

1. Data is collected from multiple data sources across the enterprise, software as a service (SaaS) applications, edge devices, logs, streaming media, and social networks.
2. Based on the type of data source, AWS Database Migration Service, AWS DataSync, Amazon Kinesis, Amazon Managed Streaming for Apache Kafka, AWS IoT Core, and Amazon AppFlow are used to ingest the data into a data lake in AWS.
3. AWS Data Exchange is used for integrating third-party data into the data lake.
4. AWS Lake Formation is used to build the scalable data lake, and Amazon Simple Storage Service (Amazon S3) is used for data lake storage.
5. AWS Lake Formation is also used to enable unified governance to centrally manage security, access control, and audit trails.
6. AWS DataBrew could be used for visual data preparation.
7. Amazon Kinesis Data Analytics is used to transform and analyze streaming data in real time.
8. Amazon QuickSight provides machine learning (ML) powered business intelligence.
9. Amazon OpenSearch can be used for operational analytics.
10. Amazon Redshift is used as a cloud data warehouse.
11. Amazon EMR provides the cloud big data platform for processing vast amounts of data using open source tools.
12. Amazon SageMaker and AWS AI services can be used to build, train, and deploy ML models, and add intelligence to your applications.
13. Amazon Redshift Spectrum and Amazon Athena enable interactive querying, analyzing, and processing capabilities.
AI/ML Services
The AWS ML Stack

Broadest and most complete set of Machine Learning capabilities

**AI SERVICES**

**VISION**
- Amazon Rekognition
- Amazon Polly
- Amazon Transcribe
- Amazon Comprehend
- Amazon Translate
- Amazon Textextract
- Amazon Kendra
- Amazon Lex
- Amazon Personalize
- Amazon Forecast
- Amazon Fraud Detector
- Amazon CodeGuru

**SPEECH**
- Amazon Polly
- Amazon Transcribe
- Amazon Comprehend

**TEXT**
- Amazon Rekognition
- Amazon Transcribe
- Amazon Comprehend
- Amazon Translate
- Amazon Textextract

**SEARCH**
- Amazon Rekognition
- Amazon Transcribe
- Amazon Comprehend
- Amazon Translate
- Amazon Textextract

**CHATBOTS**
- Amazon Kendra
- Amazon Lex
- Amazon Personalize

**PERSONALIZATION**
- Amazon Personalize

**FORECASTING**
- Amazon Forecast

**FRAUD**
- Amazon Fraud Detector

**DEVELOPMENT**
- Amazon CodeGuru

**CONTACT CENTERS**
- Contact Lens

**ML SERVICES**

Amazon SageMaker

Ground Truth

ML Marketplace

SageMaker Studio IDE

Built-in algorithms

Notebooks

Experiments

Model training & tuning

Debugger

Autopilot

Model hosting

Model Monitor

Neo

Augmented AI

**ML FRAMEWORKS & INFRASTRUCTURE**

TensorFlow

mxnet

GLUON

Keras

Deep Learning

AMIs & Containers

GPUs & CPUs

Elastic Inference

Inferentia

FPGA

© 2023, Amazon Web Services, Inc. or its affiliates.
3 Ways to Access AWS Services
You can access AWS services via the following:

- **AWS Management Console**
  - Root User
    - Email
    - Password
  - IAM user
    - Acct # or alias
    - Username
    - Password

- **Software Development Kit (SDK)**
  - Mobile App Development
  - Web Development
  - Cloud Computing
  - Internet of Things (IoT)
  - Game Development

- **Command Line Interface (AWS CLI)**
  - Access Keys are required
    - Access Key
    - Secret Access Key
Registry of Open Data on AWS (RODA)
The Human Sleep Project

The Human Sleep Project (HSP) sleep physiology dataset is a growing collection of clinical polysomnography (PSG) recordings. Beginning with PSG recordings from from ~15K patients evaluated at the Massachusetts General Hospital, the HSP will grow over the coming years to include data from >200K patients, as well as people evaluated outside of the clinical setting. This data is being used to develop CAISR (Complete AI Sleep Report), a collection of deep neural networks, rule-based algorithms, and signal processing approaches designed to provide better-than-human detection of conventional PSG.

Details →

Usage examples


Tutorial Recommendations
Tutorials

Sentinel-2

• Detecting deforestation using unsupervised K-means clustering on Sentinel-2 satellite imagery and SageMaker Studio Lab (SML)

• Getting Started With Geospatial Data Analysis

• Vegetation Analysis Pre/Post Lava Fire Using SentinelHub and Sentinel-2 Imagery on AWS

NEXRAD on AWS

• Vegetation Analysis Pre/Post Lava Fire Using SentinelHub and Sentinel-2 Imagery on AWS

You can find more here: https://registry.opendata.aws/service/sagemaker-studio-lab/usage-examples/
Amazon Simple Storage Service (Amazon S3)
Amazon S3

An object storage service:

It stores massive (unlimited) amounts of unstructured data.

Data files are stored as objects in a bucket that you define.

5 TB is the maximum file size of a single object.

All objects have a REST-accessible globally unique URL (universal namespace).

All objects have a key, version ID, value, metadata, and subresources.
Amazon S3 benefits

**Durability**
It ensures data is not lost

S3 Standard storage provides 11 9s (or 99.999999999%) of durability

**Availability**
You can access your data when needed

S3 Standard storage class is designed for four 9s (or 99.99%) availability

**Scalability**
- It offers virtually unlimited capacity
- Any single object of 5 TB or less

**Security**
- It offers fine-grained access control

**Performance**
- It is supported by many design patterns
Securing Amazon S3 buckets and objects

- Newly created S3 buckets and objects are private and protected by default

- When use cases must share Amazon S3 data –
  Manage and control the data access
  Follow the principle of least privilege

- Tools and options for controlling access to Amazon S3 data
  Block Public Access feature: It is enabled on new buckets by default, simple to manage
  IAM policies: A good option when the user can authenticate using IAM
  Bucket policies: You can define access to a specific object or bucket
  Access control lists (ACLs): A legacy access control mechanism
  S3 Access Points: You can configure access with names and permissions specific to each application
  Presigned URLs: You can grant time-limited access to others with temporary URLs
  AWS Trusted Advisor bucket permission check: A free feature
Three general approaches to configuring access

Configure the appropriate security settings for your use case on the bucket and objects.

Default Amazon S3 security settings

- Owner
- Private
- Anyone else

Public access Amazon S3 security settings

- Owner
- Public
- Anyone else

Access policy applied to Amazon S3 security settings

- Owner
- Controlled access
- User A
- User B
Consider encrypting objects in Amazon S3

**Encryption** encodes data with a secret key, which makes it unreadable. Only users who have the secret key can decode the data. Optionally, use AWS Key Management Service (AWS KMS) to manage secret keys.

**Server-side encryption**
On the bucket, enable this feature by selecting the **Default encryption** option. Amazon S3 encrypts objects before it saves the objects to disk, and decrypts the objects when you download them.

**Client-side encryption**
Encrypt data on the client side and upload the encrypted data to Amazon S3. In this case, you manage the encryption process.
Amazon S3 use case 2: Host static websites

Example objects stored in the bucket configured for website hosting

S3 bucket with public read access

Supports static content including HTML files, images, videos, and client-side scripts.
Amazon S3 best practice: Versioning

- Protects against accidental overwrites and deletes with no performance penalty
- Generates a new version with every upload
- Enables easy retrieval of deleted objects or rollback to previous versions

Three possible states of an S3 bucket –
1. Default: Versioning not enabled
2. Versioning-enabled
3. Versioning-suspended
Thank you!

Dunieski Otano
dootano@amazon.com

Ezzel Mohammed
ezzel@amazon.com