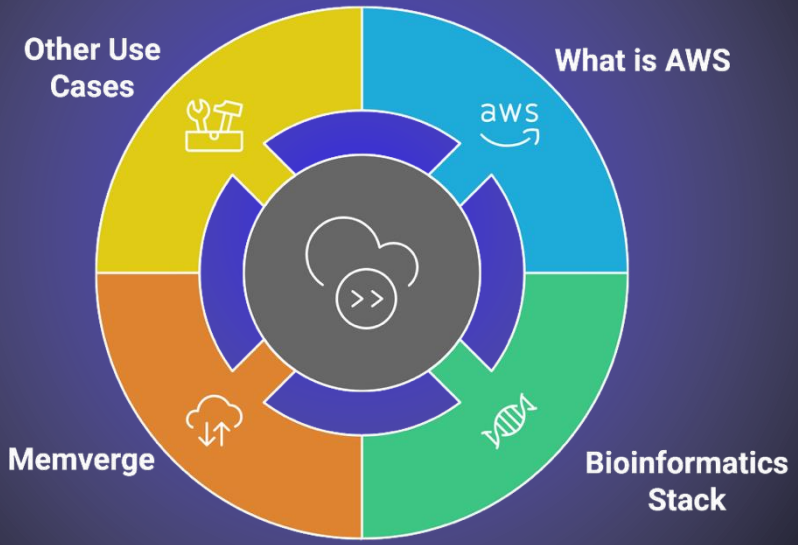
INTRO TO AWS

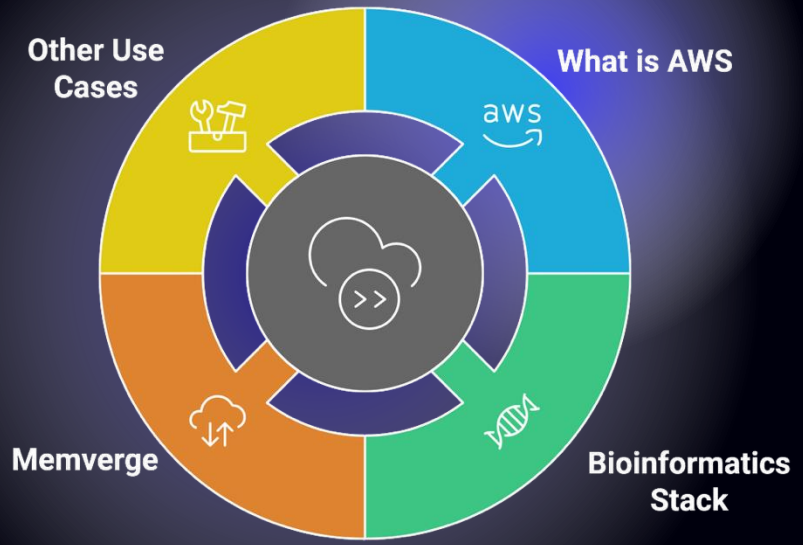
MDIBL Comparative Genomics and Data Science Core

TABLE OF CONTENTS



01

What is AWS



Question

What is the cloud computing?

Cloud computing is the on-demand delivery of IT resources over the Internet with **pay-as-you-go pricing**. Instead of buying, owning, and maintaining physical data centers and servers, you can access technology services, such as computing power, storage, and databases, on an **as-needed basis** from a cloud provider like Amazon Web Services.

—AWS

Netflix

Utilizes AWS for global content delivery and reliability

Airbnb

Leverages AWS for platform scalability and security

NASA

Employs AWS for data hosting and scientific research

Formula 1

Uses AWS for race simulations and real-time race analytics

BMW Group

Implements AWS for connected car technology

Goldman Sachs

Builds secure data analytics platforms on AWS

Epic Games

Scales gaming experiences with AWS

MDIBL

Supports biotech research and development with AWS

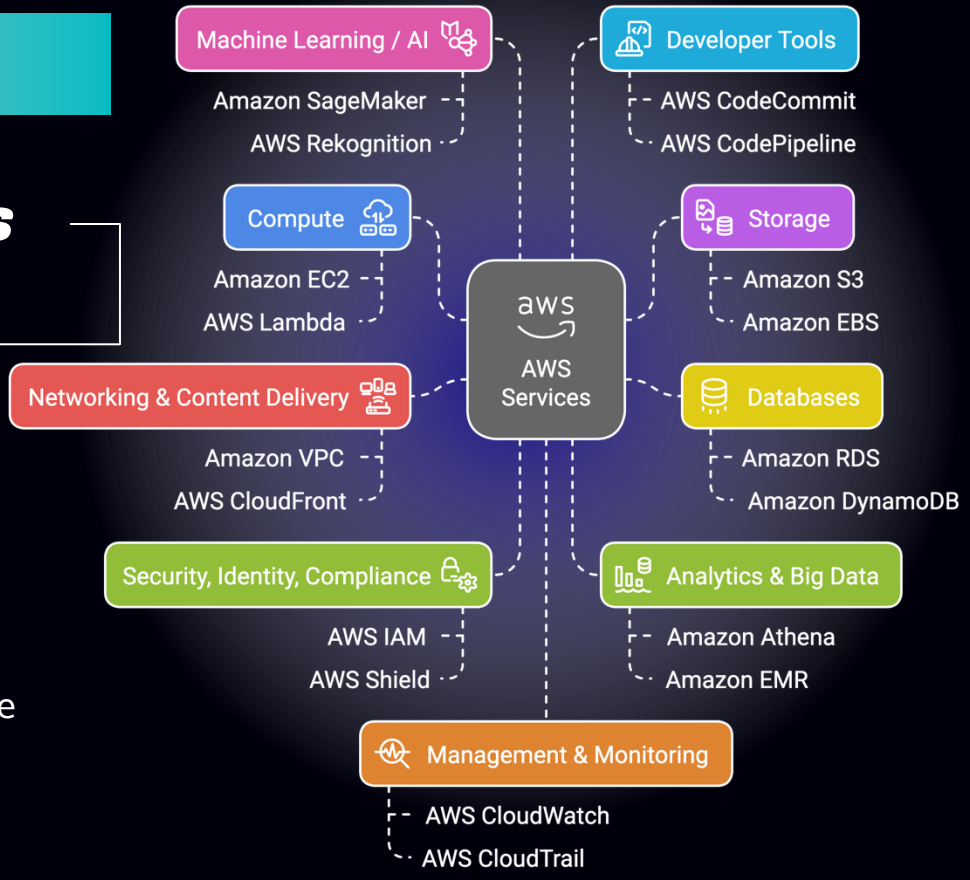
Applications

AWS

Distinct Services

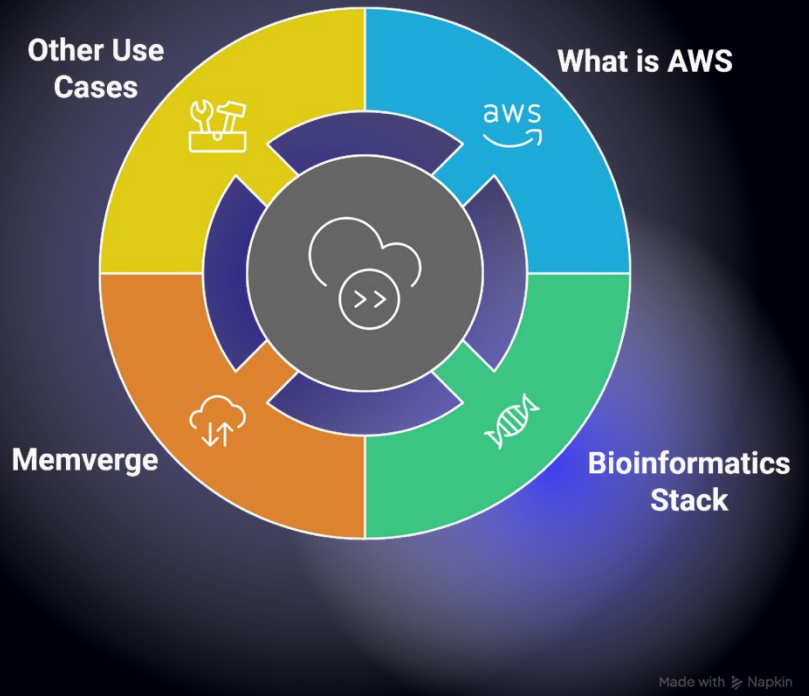
+240

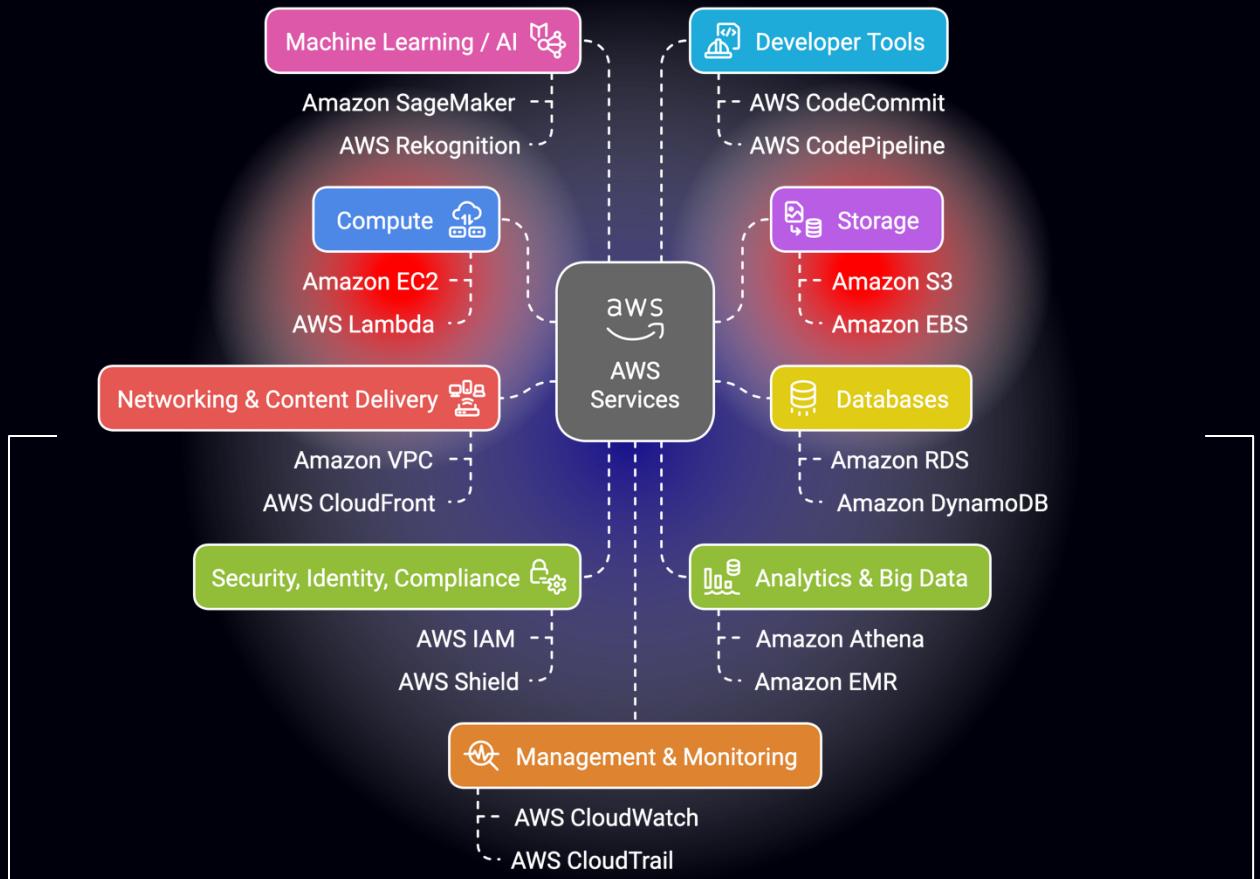
- Extensive...
- Only use a subset
- Support from AWS
- Massively scalable -- use as you like



02

Bioinformatics Stack




Machine Learning / AI
Amazon SageMaker
AWS Rekognition
Developer Tools
AWS CodeCommit
AWS CodePipeline
Compute
Amazon EC2
AWS Lambda
Storage
Amazon S3
Amazon EBS
aws
AWS Services
Networking & Content Delivery
Amazon VPC
AWS CloudFront
Databases
Amazon RDS
Amazon DynamoDB
Security, Identity, Compliance
AWS IAM
AWS Shield
Analytics & Big Data
Amazon Athena
Amazon EMR
Management & Monitoring
AWS CloudWatch
AWS CloudTrail

Services

Storage – S3

Simple Storage Service

Availability

Guarantees access to data from anywhere on the web.

Durability

Ensures data remains intact and accessible over time.

Security Features

Provides data encryption and access management.

Scalability

The ability to store and retrieve any amount of data seamlessly.

Lifecycle Management

Helps manage data efficiently with policies.

Compute – EC2

Elastic Compute Cloud

Virtual Servers

The provision of virtual servers for diverse applications.

Cost Optimization

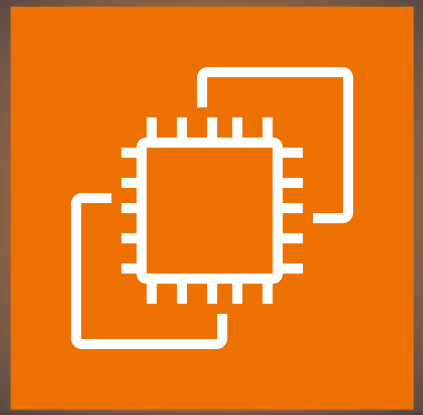
Efficiently managing costs by scaling resources as needed.

Resizable Capacity

EC2's ability to adjust compute capacity based on demand.

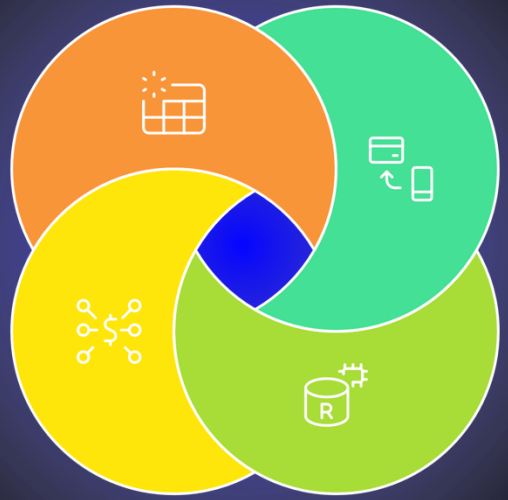
AWS Integration

Seamless integration with other AWS services for enhanced functionality.



Compute – EC2

Elastic Compute Cloud



Savings Plans

Flexible discount model

On-Demand

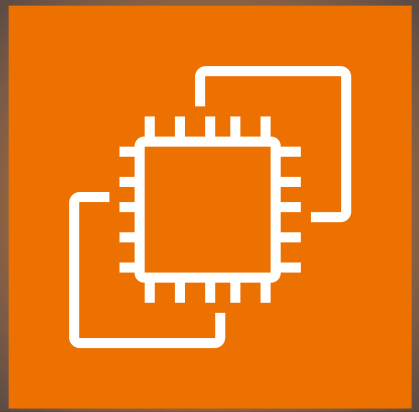
Pay for what you use, no commitment

Spot Instances

Bid for unused capacity, up to 90% cheaper

Reserved Instances

Cheaper with 1 or 3-year commitment



Compute – EC2

Elastic Compute Cloud

Spot Instances

Cost-effective but risk of data loss due to reclamation

$0.30893
$0.8930

r5.4xlarge
x1e.4xlarge

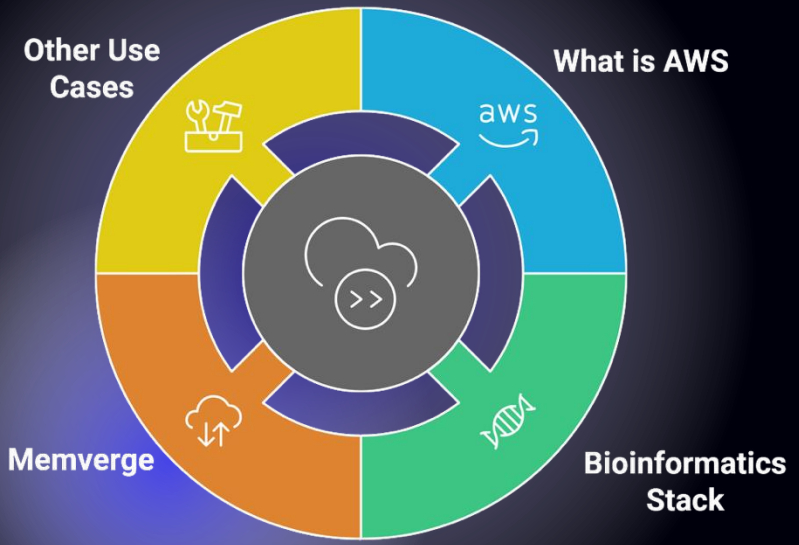
$1.008
$3.336

On-Demand Instances

Reliable with no risk of data loss, but more expensive

03

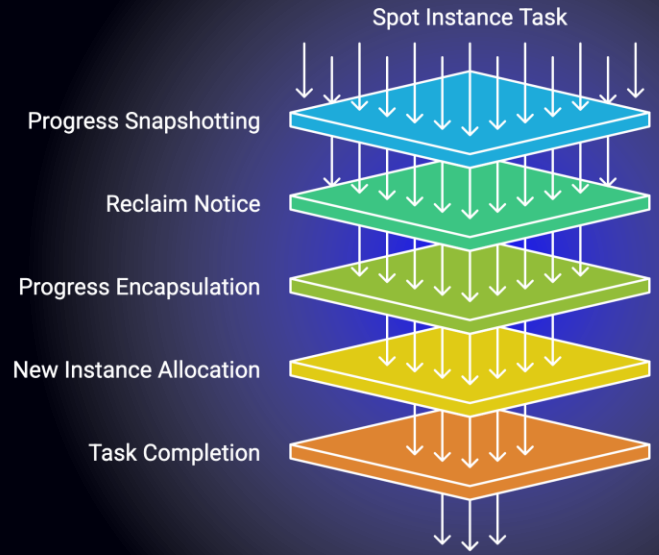
Memverge



MEMVERGE SpotSurfer

Memverge - SpotSurfer

Memory Machine Cloud



MemVerge

Memverge - WaveRider

Memory Machine Cloud

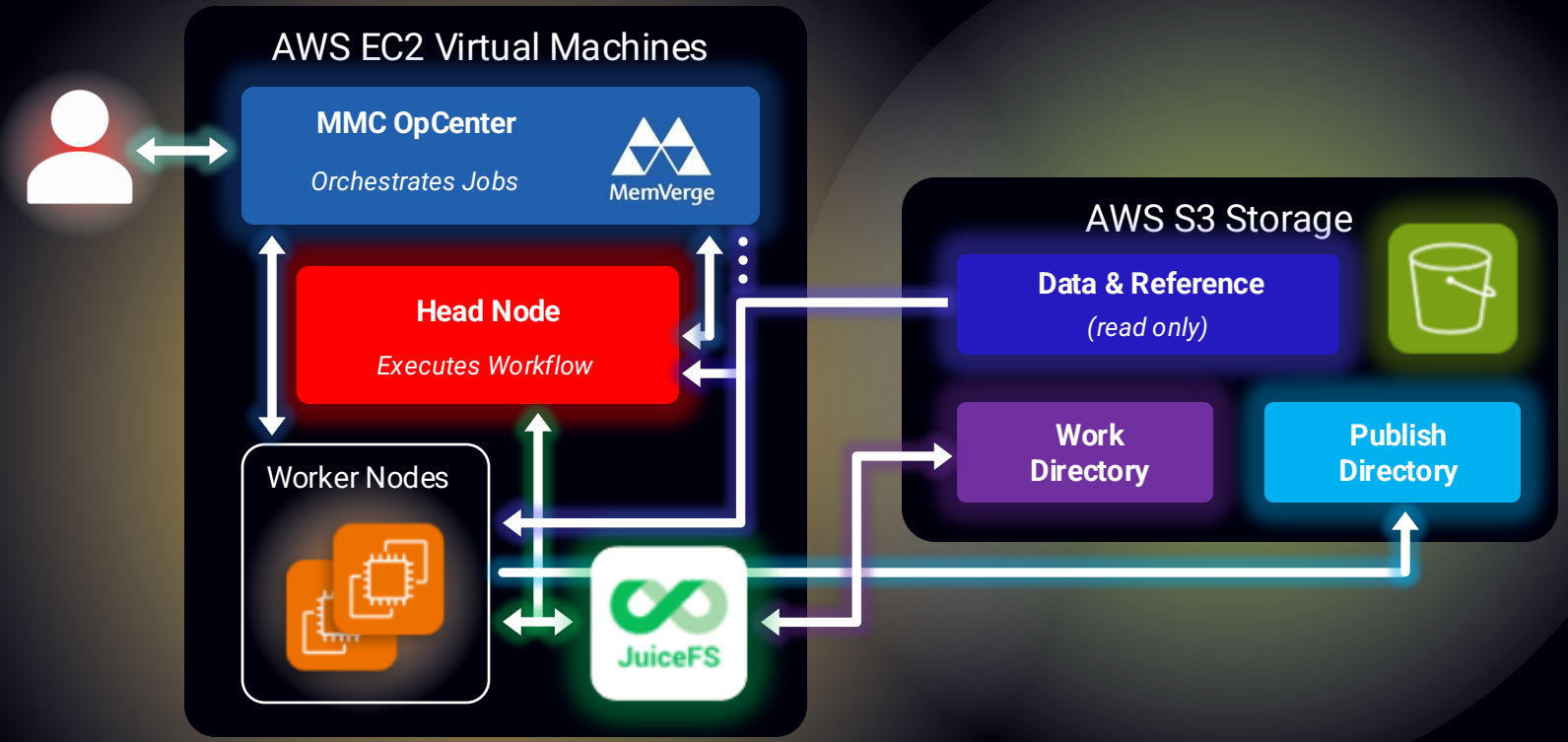
1
Monitor Utilization
Continuously track machine resource usage

2
Reach Threshold
Identify when usage exceeds predefined limits

3
Trigger Snapshotting
Capture current machine state

4
Automigrate Resources
Adjust resources to meet task needs

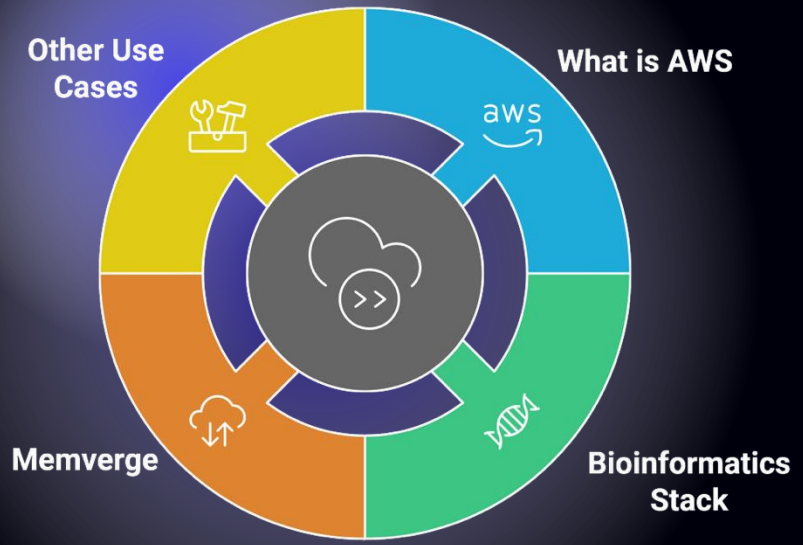
MemVerge



Architecture

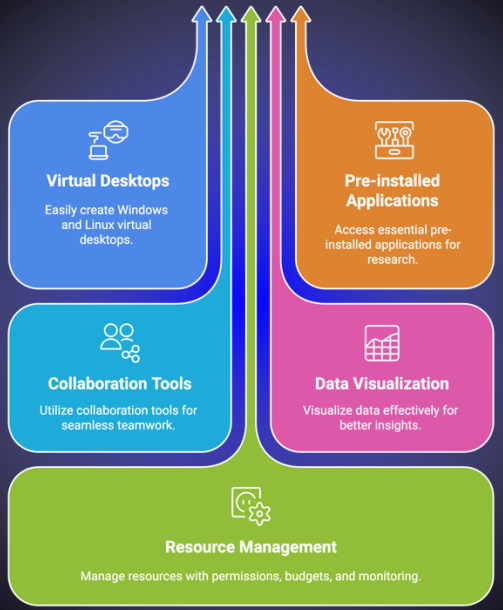
04

Other Use Cases



AWS - RES

Research Engineering Studio



AWS - Transfer

Transfer Family

Security Features

Implements SSH keys, VPC restrictions, and more

Protocol Support

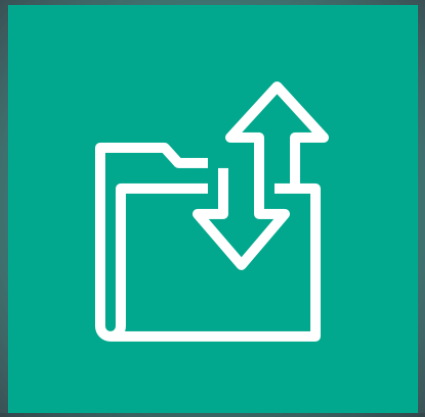
Supports SFTP, FTPS, FTP, and AS2 for secure transfers

User Management

Manages users with various identity options

Storage Backend

Integrates with Amazon S3 and EFS for storage solutions



AWS - Apps

App Runner

Observability

Monitoring and logging are integrated for performance tracking.

Security Integration

Security measures are applied, including IAM and HTTPS.

Auto-Scaling

Application scales based on traffic, including scaling to zero.

Source Code Input

Code is sourced from GitHub or CodeCommit.

Build Process

App Runner builds the application using a build image.

Deploy Application

The application is deployed automatically or manually.

Questions