# Topics:

* Cloud Concepts Overview (ACF Module 1)
* Cloud Economics and Billing (ACf Module 2)
* AWS Global Infrastructure Overview (ACF Module 3)
* Compute (A CF Module 6)
  + EC2 Cost-Optimization
  + Container Services
  + AWS Lambda
* Storage (ACF Module 7)
* Networking and Content Delivery (ACF Module 5)
* Databases (ACF Module 8)
* Cloud Security (ACF Module 4)
* Cloud Architecture (ACF Module 9)
  + AWS Well-Architected Framework
* Automatic Scaling and Monitoring (ACF Module 10)

# ACF Module 1 – Cloud Concepts

Advantages of cloud computing over on-premises:

* Avoid large capital purchases
* Use on-demand capacity
* Go global in minutes
* Increase speed and agility
* Massive economies of scale from hundreds and thousands of aggregates customers
* Trade capital expense for variable expense
* Eliminate guessing on infrastructure capacity needs
* Temporary and disposable resources
* High availability
* Fault-tolerant databases

Disadvantages:

* Multiple procurement cycles
* High latency

Pay as you go

* Enables AWS customers to pay for resources as needed

Cloud service models:

* Infrastructure as a service
* Platform as a service
* Software as a service
* Not system administration

Cloud deployment models:

* Cloud
* Hybrid
* On-premises / private cloud

AWS owns and maintains network connected hardware required for application services, while you provision and use what you need

Compute service:

* Amazon EC2

Ways to access AWS core services:

* SDKs
* AWS Command Line Interface (AWS CLI)
* AWS Management Console

Cloud adoption is not instantaneous for most organizations and requires a thoughtful, deliberate strategy and alignment across the whole organization

* Cloud Adoption Frameworj was created to help orgs develop efficient and effective plans for their cloud adoption journey
* CAF organizes guidance into six areas of focus, called perspective
  + Consist of sets of business or tech capabilities thare the responsibility of key stakeholders

# ACF Module 2 – Cloud Economics and Billing

Fundamentals of pricing

* No charge for
  + Inbound data transfer for ECS
  + Transfer between AWS services within same AWS region
* Pay for what you use
* Start and stop anytime
* No long term contracts required
* Some services are free, but other services they provision may not be free
  + Virtual Private Cloud (VPC)
  + Identity and access management (IAM)
  + Consolidated billing
  + Elastic beanstalk
  + Automatic scaling
  + OpsWorks
  + CloudFormation

Total cost of ownership

Billing

Technical support

For services like EC2 and RDS, can invest in reserved capacity. Options for reserved capacity are:

* AURI
* NURI
* PURI

Customer can get details about EC2 billing that took place within the past 3 months ago via the AWS Cost Explorer

To receive discounted rate associated with reserved instance, you do not need to make the full upfront payment for term of agreement

Four support plans by AWS support:

* Basic, Developer, Business, Enterprise

The Total Cost of Ownership (TCO) Calculate compares cost of running application on-premises vs AWS

As AWS grows, cost of doing business is reduces and savings passed back to customer (Economies of scale)

AWS free tier only applies to certain services

# ACF Module 3 – AWS Global Infrastructure Overview

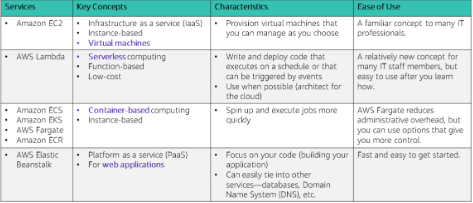
AWS Global Infrastructure

* Consists of regions and availability zones
  + Each region hosts two or more availability zones
  + AZs within region connected through low-latency links
  + Data center in an AWS region cannot be shared across multiple AZs in that region
* Choice of region typically based on compliance requirements, or to reduce latency
* Easy availability zone is physically separate from other availability zones and has redundant power, networking and connectivity
* Edge locations, and regional edge caches improve performance by caching content closer to users
  + Edge locations do not need to be located in same general area as region
* Fault tolerant means infrastructure has built in component redundancy
* Elastic and scalable means resources dynamically adjust to increase or decrease in capacity requirements
* Highly recommended to provision compute resources across multiple availability zones

AWS services and service category overview

* Storage
  + Simple Storage Service (S3)
  + Elastic Block Store (EBS)
  + Elastic File System (EFS)
  + Storage Service Glacier
* Compute
  + Elastic Compute Cloud (EC2)
    - EC2 Auto Scaling
  + Elastic Container Service (ECS)
  + Elastic Container Registry (ECR)
  + Elastic Beanstalk
  + Lambda
  + Elastic Kubernetes Service (EKS)
  + Fargate
* Database
  + Relational Database Service (RDS)
  + Aurora
  + Redshift
  + DynamoDB
* Networking
  + Virtual Private Cloud (VPC)
  + Elastic Load Balancing
  + CloudFront
    - Uses Edge Locations to ensure low latency delivery
  + Transit Gateway
  + Route 53
  + Direct Connect
  + VPN
* Security
  + Aws Identity and Access Management (IAM)
  + Organizations
  + Cognito
  + Artifact
  + Key Management Service (KMS)
  + Shield
* Cost Management
  + Cost and Usage Report
  + Budgets
  + Cost Explorer
* Management and Governance Service
  + Management Console
  + Config
  + CloudWatch
  + Auto Scaling
  + Command Line Interface
  + Trusted Advisor
  + Well-Architected Tool
  + CloudTrail

# ACF Module 6 – Compute



Choose based on:

* Application design
* Usage patterns
* Config options you want to manage

EC2

* Provides virtual machines you can manage with full control over guest OS for each instance
* Launch instances of any size into an AZ anywhere in world
  + Instances from Amazon Machine Images, contains
    - Template for root volume for instance
    - Lauch permissions that control which accounts can use AMI to launch instances
    - Block device mapping that specifies volumes to attach to instance
  + Launch with a few clicks or line of code, ready in minute
  + Must specify AMI and instance type
* Control traffic to and from
* Many different instance types
  + Differing combos of CPU, RAM, storage, and networking capabilities
* Configure security groups to control access to instances
* User data enables you to specify a script to run the first time instance launches
* Only instances backed by EBS (elastic bean stalk) can be stopped
* Can use CloudWatch to capture and review metrics
* Pricing models:
  + On-Demand instances
  + Reserved instances
    - Long term workloads with predictavle usage
    - Also useful for very predictable short term workloads, eg to process monthy reports that iterate through large amount of data
  + Spot instances
    - Interrupted with 2 minute notification, but offer significant cost savings over on-demand instances
  + Dedicated instances
    - Ensures instsance will not share physical host with instances from any other AWS customer
  + Dedicated hosts
* Four pillars of cost optimization
  + Right size
  + Increase elasticity
  + Optimal pricing model
  + Optimize storage choices

Container services

* Smaller than virtual machine, do not contain entire OS
* Hold everything that application needs to run
* Docker is a software platform that packages software into containers
  + Single application can span multiple containers
* Elastic Container Service (ECS) orchestrates execution of docker containers
* Kubernetes is open source software for container orchestration
  + ECS enables you run Kubernetes on AWS
* Elastic Container Registry (ECR) enables you to store, amange, and deploy Docker containers

Lambda

* Serverless computing enables you to build and run applications and services without provisioning and managing servers
* Lambda is a serverless compute service that provides built-in fault tolerance and automatic scaling
* Event source is an AWS service or developer created application that triggers a Lambda function to run
* Maximum memory allocation for a lambda function is 3008 MB
* Maximum execution time is 15 minutes

Elastic Beanstalk

* Enhances developer productivity
  + Simplifies process of deploying application
  + Reduces management complexity
* Supports Java, .NET, PHP, Node.js, Python, Ruby, Go, and Docker
* No charge! Pay only for the resources you use

# ACF Module 7 – Storage

Elastic Block Store (EBS)

* Persistent and customizable block storage for EC2
* Recommend when data must be:
  + Quickly accessible
  + Requiring long term persistence
  + Encrypted
* HDD and SSD types
* Replicated within same availability zone
* Easy and transparent encryption
* Elastic volumes
* Back up using snapshots

Simple Storage Service (S3)

* Fully managed cloud storage service
* Suitable for “flat” files like documents, photos, etc
* Bucket associated with specific AWS region
* Can store virtually unlimited number of objects
  + Pay only for what you use
* Access at any time anywhere through a URL
* By default, not viewable by public
* Offers rich security controls
* Replicates all objects in multiple availability zones within same region
* Storage classes for S3 object lifecycle policies:
  + Glacier
  + Infrequent access
  + Standard access
* Name of bucket must be unique worldwide across all AWS accounts

Elastic File System (EFS)

* Provides storage over network
* Perfect for:
  + Big data / analytics
  + Media processing workflows
  + Content management
  + Web serving
  + Home directories
* Can be used to implement storage for EC2 instances that multiple virtual machines can access at same time
* Fully managed service that eliminates storage administration tasks
* Accessible from:
  + Console
  + API
  + CLI
* Scales up or down as files added or removed
  + Pay for what you use

S3 Glacier

* Data archiving service designed for security, durability, and extremely low cost
* Pricing based on Region
* Extremely low cost design works well for long term archiving
* Designed to provide 11 9s of durability for objects
* Vault:
  + Container for storing archives

# ACF Module 5 – Networking and Content Delivery

Networking basics

* Network: Two or more client machines connected together
  + Can be logically partitioned into subnets
* IP: Unique address that identifies each client machine in a network
  + 32-bit IP = IPv4 (eg 192.0.2.0)
  + 128-bit = IPv6 (eg 2600:1f18:22ba:8c00:ba86:a05e:a5ba:00FF)
* CIDR: Classless Inter-Domain Routing
  + Common method to address networks
  + CIDR Address:
    - eg 192.0.2.0/24
      * IP address (192.0.2.0)
      * /
      * Bits of routing prefit must be fixed or allocated for network identifier
      * 24 tells you first 24 bits must be fixed within network, and leftover (of 32) can be changed
        + IPs ranging from 192.0.2.0 to 192.0.2.255 available for CIDR block
        + but 5 are reserved for VPC…
    - Two special cases:
      * Fixed IP addresses. Represented with /32
      * The internet, represented with /0

Amazon Virtual Private Cloud (VPC)

* A VPC is a logically isolated section (virtual network) within the AWS Cloud
* Belongs to one region and requires a CIDR block
  + But can span availability zones
* Subdivided into subnets
  + Belong to one availability zone
  + Require a CIDR block
    - Smallest size of subnet is /28
    - Largest is /16
  + 5 IP addresses reserved
* Route tables control traffic for a subnet
  + Have built-in local route
  + You can add additional routes
  + Cannot delete local route
* Main route table created by default
  + Must manually create subnets and internet gateway

VPC Networking

* Internet gateway: Connect VPC to internet
* NAT gateway: Enables instances in a private subnet to connect to internet
* VPC endpoint: Connects VPC to supported services
* VPC peering: Connects your VPC to other VPCs
* VPC sharing: Allows multiple accounts to create their application resources into shared, centrally managed VPCs
* Site-to-Site VPN: Connects VPN to remote networks
* Direct Connect: Connects VPC to a remote network by using a dedicated network connection
* Transit Gateway: Hub and spoke connection to VPC peering

VPC security

* Isolate subnets when possible
* Choose appropriate gateway device or VPN connection for needs
* Use firewalls
  + Security groups
  + Access control lists

Route 53

* Highly available and scalable cloud DNS web service that translate domain names into IP addresses
* Supports several types of routing policies
* Multi region deployment improves application’s performance for a global audience
* Failover can be used to improve availability

Amazon CloudFront

* a content delivery network (CDN) is a globally distributed system of caching services that accelerates delivery of content
* Cloudfront is a fast CDN service that security delivers data, videos, applications, and APIs over a global infrastructure with low latency and high transfer speeds
* Uses AWS edge locations to ensure low latency
* Offers benefits including:
  + Global and fast
  + Security at edge
  + Highly programmable
  + Deeply integrated with AWS
  + Cost effective

Security group acts as a virtual firewall for your EC2 instance(s) to control inbound and outbound traffic

# ACF Module 8 – Databases

When choosing type, consider:

* Data size
* Data access period
* Query frequency
* Availability

Amazon Relational Database Service (RDS)

* Web service that makes it easy to set up, operate, and scale relational database in the cloud
* Managed service
* Accessible via console, CLI, or application API calls
* Scalable compute and storage
* Automatically patches database software and backup database, enabling point in time recovery
* Good for complex transactions and queries
* Supported engines:
  + Amazon Aurora
  + Postgre SQL
  + MySQL
  + MariaDB
  + Oracle
  + Microsoft SQL Server
* Choose between two SSD backed options:
  + One for high-performance online transactional processing (OLTP) applications
  + Other for cost effective general use
* To improve read-heavy application performance, use Read Replicas

DynamoDB

* Fully managed NoSQL database service
  + Excels at scaling to hundreds of thousands of requests with key/value access to user profile and session
* Ideal for database that requires fast performance, fast scalability, and flexibility in the database schema
* Runs exclusively on SSDs
* Supports document and key-value store values
* Works well for mobile, web, gaming, ad tech, and “internet of things” applications
* Accessible via console, CLI, and API calls
* Replicates tables automatically across your choice of Regions
* Provides consistent single digit ms latency
* No limits on table size or throughput
* Attribute: Fundamental data element
* Scan operation: To find item in DynamoDB table other than primary key
* Query:
  + A table using partition key and optional sort key filter
  + Secondary indexes that exist for table
  + Efficiently retrieve items from a table or secondary index

Redshift

* Fast, fully managed data warehouse service
* Best suited for analyzing data by using SQL and existing business intelligence tools
* Easily scales with no downtime
* Columnar storage and parallel processing architecture
* Automatically and continuously monitors cluster
* Encryption built in – but needs enabled

Aurora

* Highly available, performant, and cost effective managed relational database
* Distributed storage subsystem
* Fault tolerant and self healing storage
* Replicates multiple copies of data across multiple AZs, continuously backs up data to S3
* Multiple levels of security
  + Network isolation through VPC
  + Encryption at rest through AWS key management (KMS), encryption of data through SSL
* Compatible with existing MySQL and PostgreSQL databases, adds compatibility for new releases
* Ideal database for a .NET layer that connects to MySQL, and needs high availability and automated backups
* Fully managed by RDS
  + Automates management tasks such as hardware provisioning
  + Software patching
  + Setup
  + Config
  + Backups

# ACF Module 4 – Cloud Security

AWS Shared Responsivity Model

* AWS responsible for security of the cloud
  + Hardware, software, networking, and facilities that run the cloud services
* Customer responsible for security in the cloud
  + Performing necessary security configuration and management tasks
  + OS updates and security patches
  + Firewall
  + Security group configurations
  + Encryption

Identity and Access Management (IAM)

* IAM Policies are constructed with JSON and define permissions
  + Attached to any IAM entity
  + Entities are IAM users, groups, and roles
* IAM user
  + Provides a way for a person, application, or service to authenticate to AWS
  + User can be granted management console access and programmatic access
* IAM group
  + Simple way to attach the same policies to multiple users
* IAM role
  + Can have permission policies attached to it, and can be used to delegate temporary access to users or applications
* Best practices:
  + Managing access to AWS resources
  + Defining fine-granted access rights

Securing new AWS account

* Secure logins with multi factor authentification
* Create individual IAM users and grant permissions according to principle of least privilege
* Use groups to assign permissions
* Configure strong password policy
* Delegate using roles instead of sharing credentials
* Monitor account using AWS CloudTrail
* Change AWS support plan with AWS root user, other tasks with IAM
  + Delete account root user access keys

Securing accounts

* AWS Organizations: Consolidate multiple AWS accounts that you can centrally manage
  + Group into organization units, attach policies into each OU
  + Support for IAM
  + Use service control policies to establish control over services and API actions each account can access
* Key Management Service (KMS)
  + Create and manage encryption keys
  + Control use of encryption across services and in applications
  + Integrates with Cloudtrail to log all key usage
  + Uses hardware security modules (HSM) that are validation by Federal Information Processing Standards to protect keys
* Cognito
  + Adds sign up, sign in, and access control to applications
  + Scales to millions of users
  + Sign in with social identity providers

AWS Shield

* Managed DDoS protection service
* Safeguards applications on AWS
* Provides always on detection and auto inline mitigations
* Standard enabled for no additional cost
  + Advanced is an optional paid service
* Minimizes application downtime and latency

Securing Data

* At rest:
  + Use encryption supported by KMS
* In Transit:
  + TLS / SSL: AWS Certificate Manager
* S3 buckets and objects:
  + Private and protected by default
  + When use case requires sharing data objects
    - Follow principle of least privilege, consider encryption

Ensuring compliance:

* Customers subject to many regulations and requirements
* Certifications and attestations
  + Assesses by third party, independent auditor
* Laws, regulations, privacy
  + AWS provides features and agreements to support compliance
* Alignments and frameworks
  + Industry or function specific requirements
* AWS config used to assess, audit, and evaluate configurations of AWS resources
* AWS artifact provides access to security and compliance reports

# ACF Module 9 – Cloud Architecture

AWS Well-Architected Framework

* Provides a consistent approach to evaluate cloud architectures and guidance to help implement designs
* Documents a set of foundational questions that enable you to understand if a specific architecture aligns well with cloud best practices
* Operational Excellence
  + Run and monitor systems to delivery business value, and to continually improve supporting processes and procedures
* Security
  + Protect info, systems, and sets while delivering value through risk assessment and mitigation
  + Protect data at-rest and in-transit
  + Apply security in all layers
  + Implement strong identity foundation
* Reliability:
  + Prevent and quickly recover from failures to meet demand
  + If a SysOps engineer wants to protect data in transit and at rest. They should use Elastic Load Balancing, Elastic Block Store, and S3
  + Assume everything will fail
* Performance Efficiency:
  + Use IT and computing resources efficiently to meet requirements and to maintain that efficiency as demand changes and tech evolves
  + Four areas: Tradeoffs, Selection, Monitoring, ???
  + Use serverless architecture and democratize advanced technologies when considering performance efficiency
* Cost optimization:
  + Run systems to deliver value at lowest price point

Reliability and availability

* Reliability is a measure of system’s ability to provide functionality when desired by user
  + Measured in terms of MTBF
* Availability is percentage of time a system is operating normally or correctly performing operations expected of it
  + Influenced by fault tolerance, scalability, and recoverability
  + Can design to be highly available, with a cost tradeoff to consider
  + Highly available is always available without need for human intervention
* Fault tolerance: Ability for system to remain operational even if some components fail

AWS Trusted Advisor

* Online tool that provides real time guidance to help you provision resources by following best practices
* Looks at your entire environment and gives you real time recommendations in five categories
  + Performance, cost optimization, security, fault tolerance, service limits
* Can be used to help optimize environment as you start implementing architecture designs

# ACF Module 10 – Automatic Scaling and Monitoring

Elastic Load Balancing

* Distributes incoming traffic across multiple targets (such as instances, containers, addresses, functions) in one or more availability zones
* Configure load balance to accept incoming traffic by specifying one or more listeners
* When detects unhealthy target:
  + Stops routing traffic to that target
  + Sends traffic to healthy target
  + Resumes routing traffic to that target when it is healthy again
* Supports three types of load balancers:
  + Application load balancer
  + Network load balancer
  + Classic load balancer
* Offers several monitoring tools for continuous monitoring and logging of health checks, security, and analytics

Cloudwatch

* Helps you monitor AWS resources, and applications that run on them, in real time
* Collect and track standard and custom metrics
* Set alarms to automatically send notifications to Simple Notification Service (SNS) topics, or performing auto scaling or EC2 actions based on value of the metric or expression relative to threshold over a number of time periods
* Define rules that match changes in AWS environment and route these events to target for processing
* Not cloudtrail --- that’s for auditing

EC2 Auto scaling

* Responds quickly to changes in resource needs
* Helps maintain availability, automatically add or remove instances based on workload
* Auto scaling group: Collection of EC2 instances
  + Set minimize size
  + Maximize size
  + Desired capacity
* Launch configuration is an instance configuration template
  + AMI, instance type, and EBS volumes
* Enforces minimum number of running EC2 instances

AWS Auto Scaling

* Separate service, monitors applications, and automatically adjusts capacity for
* EC2 instances and spot fleets
* ECS tasks
* DynamoDB tables and indexes
* Aurora replicas