AWS Cloud Practitioner Essentials (course link)

Gabriele Libianchi

February 12, 2021

Course objectives:

- Summarize the working definition of AWS;

- Differentiate between on-premises, hybrid-cloud, and all-in cloud;

- Describe the basic global infrastructure of the AWS Cloud;

- Explain the six benefits of the AWS Cloud;

- Describe and provide an example of the core AWS services, including compute, network, databases, and storage;

- Identify an appropriate solution using AWS Cloud services with various use cases;

- Describe the AWS Well-Architected Framework;

- Explain the shared responsibility model;

- Describe the core security services within the AWS Cloud;

- Describe the basics of AWS Cloud migration;

- Articulate the financial benefits of the AWS Cloud for an organization's cost management;

- Define the core billing, account management, and pricing models;

- Explain how to use pricing tools to make cost-effective choices for AWS services

Module 1 Introduction to Amazon Web Services

- Summarize the benefits of AWS;

- Describe differences between on-demand delivery and cloud deployments;

- Summarize the pay-as-you-go pricing model.

1.1 Client-Server Model

Client-Server Model: a customer makes a request and with permissions the server satisfies such request. AWS'server is **Amazon Elastic Compute Cloud (EC2)**: virtual servers are EC2 instances. In computing, a client can be a web browser or desktop application that a person interacts with to make requests to computer servers. A server can be services such as Amazon Elastic Compute Cloud (Amazon EC2), a type of virtual server.

For example, suppose that a client makes a request for a news article, the score in an online game, or a funny video. The server evaluates the details of this request and fulfills it by returning the information to the client.

KEY CONCEPT 1. "you only pay for what you use". On premises data centers are limited by capacity constraints, on AWS you can add/remove services.

1.2 Cloud Computing

Definition 1. Cloud Computing is the on-demand delivery of IT resources over the internet with pay-as-you-go pricing. (i.e. the on-demand delivery of compute power, database, storage, applications, and other IT resources through a cloud services platform via the Internet with pay-as-you-go pricing.)

on-demand delivery: AWS has the resourced you need WHEN you need them. (flexibility impossible when managing own data centers).

The engine of IT resources in "not a discriminator" for better business company: **"undifferentiated heavy lifting of IT"**. Tasks that are common, repetitive and consuming,AWS takes care of this.

1.2.1 Deployment models for cloud computing

When selecting a cloud strategy, a company must consider factors such as required cloud application components, preferred resource management tools, and any legacy IT infrastructure requirements.

- **CLOUD-BASED DEPLOYMENT**: **run** all parts of the application in the cloud, **migrate** existing application to the cloud, **design and build new application in the cloud**.
  The application can be built either on *low-level infrastructure* (requiring additional management by IT staff) or built using *high-level services*

reducing management, architecting and scaling requirements of the core infrastructure.

- **ON-PREMISES DEPLOYMENT**: deploy resources by using **virtualization and resource management tools**, increase their utilization using virtualization technology and application management. This model is close to *legacy IT infrastructure* but virtualization and app management increase resource utilization.

- **HYBRID DEPLOYMENT**: connect could-based resource to on-premises infrastructure and integrate cloud-based resource with legacy IT applications. This model is well suited for legacy application that are better maintained on premises (governments); at the same time such apps can be paired with cloud services.

1.2.2 Benefits of cloud computing

Companies advantage in using cloud computing while addressing business needs:

- **Trade upfront expense for variable expenses**: upfront costs are the one for buying and setting up a data center, variable expenses means "pay as you use". Innovative solutions can be implemented without infrastructure costs.

- **Stop spending money to run and maintain data centers**: maintenance is cost& time expensive, better to put efforts in applications and customers.

- **Stop guessing capacity**: no need to predict infrastructure capacity. Launch EC2 instances when needed and pay for the actual compute time used. *Scale in* or *out* according to the demand.

- **Benefits from massive economies of scale**: The massive amount of Cloud providers allow them to achieve higher economies of scale that translates in lower pay-as-you-go prices.

- **Increase speed and agility**: The flexibility of cloud computing makes it easier for you to develop and deploy applications. you can access new resources within minutes.

- **Go global in minutes**: the global footprint of the AWS cloud enables you to quickly deploy applications worldwide with low latency for costumers.

1.3 Bonus: Fundamental (5) Pillars of AWS

1.3.1 Operational Excellence

This pillar focuses on how you can continuously improve your ability to run systems, create better procedures, and gain insights.

- *Mental model*: we can think of operational excellent in terms of **automation**. The more operations can be automated, the less chance there is for human error. Automation also improve internal processes by promoting a series of repeatable best practices that can be applied across the entire organization.

- *Concepts*: The focus must points towards areas that require the most manual work and might have the biggest consequence for error. We also need a process to track, analyze and improve operational efforts. We focus on the following concepts of operational excellence:

  Infrastructure as a Code (IaC): *IaC is the process of managing the infrastructure through machine-readable configuration files*; this is the foundation allowing automation of the infrastructure. The IaC platform, given templates describing the needed resources, automatically provision and configure such resources on our behalf. *IaC is a declarative and automated way of provisioning infrastructure*. IaC is implemented in AWS through the service **CloudFormation**: it requires declaring our resources using JSON or YAML. Alternatively the **Cloud Development Kit (CDK)** allows to author such templates in JavaScrit, Python or Java.

  Observability: *is the process of measuring the internal state of your system to achieve some desired end state*; you can track the impact of your automation and continuously improve it. Implementation of *Observability* involves the following steps:

  - **Collection**: *aggregating all metrics necessary when assessing the state of the system*
    * *Infrastructure-level metrics* are automatically emitted by AWS services and collected by the **CloudWatch** service; sometimes structured logs are emitted and can be enabled and collected through *CloudWatch Logs*.
    * *Applicaiton-level metrics*: generated by your software and collected with *CloudWatch Custom Metrics*. Software logs can be stored using *CloudWatch Logs* or uploaded to **S3**.
    * *Account-level metrics* are logged by your AWS account and collected by the **CloudTrail** service.
  - **Analytics**: *analytics solutions and databases provided by AWS can be used to collect metrics.*
    * To analyze logs stored in *CloudWatch Logs*, consider using *CloudWatch Logs Insight*, a service that lets you interactively search and analyze your Cloudwatch log data.
    * To analyze logs stored in **S3**, consider using **Athena**, a serverless query service.

* To analyze structure data, consider using **RDS**, a managed relational database service.
* To analyze large amounts of structured data, consider using **RedShift**, a managed petabyte-scale data warehouse service.
* To analyze log-based data, consider using the *Elasticsearch Service*, a managed version of **ElasticSearch**, the popular open-source analytics engine.

- **Action**: after collection and analysis of metrics those can be used to achieve one of the following outcome or process:
  * *Monitoring & alarming*: *CloudWatch Alarms* can notify when a systeam breaches the safety threshold for a particular metric, manual or automatic mitigation follows.
  * *Dashboards*: You can create dashboards of your metric using *CloudWatch Dashboards*; they can track and improve service performance over time.
  * *Data-driven decisions*: you can track performance and business KPIs to make data.driven product decisions.

- *Conclusions* In this module, you have learned about the pillar of operational excellence. You have learned about the mental model of thinking about operations as automation. You have learned about IaC and how it can be used to provision your services automatically using the same tools and processes that you currently use for code. You have learned about observability and how to collect, analyze, and act on metrics to continuously improve your operational efforts.

- *Additional resources* : Operational Excellence Pillar Whitepaper

1.3.2 Security

CONTINUE FROM HERE

- *Mental model*

- *Concepts*

- *Conclusions*

- *Additional resources*

1.3.3 Reliability

- *Mental model*

- *Concepts*

- *Conclusions*

- *Additional resources*

1.3.4 Performance Efficiency

- *Mental model*

- *Concepts*

- *Conclusions*

- *Additional resources*

1.3.5 Cost optimization

- *Mental model*

- *Concepts*

- *Conclusions*

- *Additional resources*

Module 2   Compute in the cloud

- Describe the benefits of Amazon EC2 at a basic level.

- Identify the different Amazon EC2 instance types.

- Differentiate between the various billing options for Amazon EC2.

- Summarize the benefits of Amazon EC2 Auto Scaling.

- Summarize the benefits of Elastic Load Balancing.

- Give an example of the uses for Elastic Load Balancing.

- Summarize the differences between Amazon Simple Notification Service (Amazon SNS) and Amazon Simple Queue Service (Amazon SQS).

- Summarize additional AWS compute options.

2.1   EC2 : Amazon Elastic Compute Cloud introduction

EC2 provides secure, resizable compute capacity in the cloud as Amazon EC2 instances. In opposition to get up and running on-premises servers, **EC2** is highly flexible, cost effective and quick. You just have to request the EC2 instances you want and launch them, then you can stop or terminate them, and maybe change them.

EC2 runs on top of physical host machines managed by AWS using virtualization technology. When you spin up an EC2 instance, you aren't necessarily taking an entire host to yourself. Instead, you are *sharing the host with multiple other instances (VM's)*. And a **hypervisor** running on the host machine is responsible for sharing the underlying physical resources between the virtual machines.

This idea of sharing underlying hardware is called **multitenancy**. The hypervisor is responsible for coordinating this multitenancy and it is managed by AWS. The hypervisor is responsible for *isolating the virtual machines from each other* as they share resources from the host. This means EC2 instances are secure. Even though they may be sharing resources, one EC2 instance is not aware of any other EC2 instances also on that host. They are secure and separate from each other.

EC2 provides flexibility and control over the configuration of the instances: OS, Software, ecc. **Vertical Scaling**: Instances are resizeable in terms of memory and CPU. You also control the networking aspect of EC2: the types of requests to the server, if they are publicly or privately accessible. **Compute as a Service (CaaS)** model.

2.2 EC2 Instance types

Each Amazon EC2 instance type is grouped under an instance family and are optimized for a certain type of task; the families are:

- **General purpose**: good balance of compute, memory and networking resources (roughly equivalent). Suitable for applications that does not require optimization in any single resource area: web services, code repositories, app servers, gaming servers, small medium DBs.

- **Compute optimized**: ideal for compute intensive tasks that benefits from high-performance processors (gaming servers, HPC, scientific modeling, applications . . .),

- **Memory optimized**: deliver fast performance for workloads that process large datasets in memory. Ideal for workloads that requires large amounts of data to be preloaded before running an application (high-performance databases, real time processing of large amount of unstructured data. Memory optimized instances enable you to run workloads with high memory needs and receive great performance

- **Accelerated computing** use hardware accelerations, or coprocessors, for memory intensive tasks (accelerated computing, floats calculations, graphic processing, data pattern matching,game/application streaming). A hardware accelerator is a component that can expedite data processing.

- **Storage optimized** designed for workloads that require high, sequential read and write access to large datasets on local storage. Suitable workoads can be: distributed file systems, data warehousing, high frequency online transaction processing (OLTP) systems. In computing, the term input/output operations per second (IOPS) is a metric that measures the performance of a storage device. It indicates how many different input or output operations a device can perform in one second. Storage optimized instances are designed to deliver tens of thousands of low-latency, random IOPS to applications.

2.3 EC2 pricing

There are several different billing options:

- **On-Demand**: per hour/per second, depends on instance type. no need for upfront payments or long term commitments. Don't need contract or communication.

- **Savings Plan**: low prices on EC2 in exchange of a commitment on a consistent amount of usage measured in dollars per-hour for 1 or 3 years term. You can save up to 72% on AWS compute usage regardless of everything (instance type, size, OS, region . . .). This plan also applies to **AWS Fargate** and **AWS Lambda** usage (serverless compute options).

- **Reserved instances**: suited for steady-state workloads or ones with predictable usage. They offer up to 75% of savings compared to On-Demand pricing, if you commit to 1 or 3 years; there are 3 payment options all/partial/no upfront.

- **Spot Instances**: they allow you to request spare EC2 computing capacity for up to 90% off the On-Demand pricing.

Module 3:

-

Module 3   Services

- **[EC2] AWS Elastic Compute Cloud**
- **CloudFormation**
- **[CDK] Cloud Development Kit**