BERNARDGOLDEN

INTRODUCTION TO EC2

EC2 Description

BERNARDGOLDEN

Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides secure, resizable compute capacity in the cloud. It is designed to make web-scale cloud computing easier for developers.

Amazon EC2's simple web service interface allows you to obtain and configure capacity with minimal friction. It provides you with complete control of your computing resources and lets you run on Amazon's proven computing environment. Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing you to quickly scale capacity, both up and down, as your computing requirements change. Amazon EC2 changes the economics of computing by allowing you to pay only for capacity that you actually use. Amazon EC2 provides developers the tools to build failure resilient applications and isolate them from common failure scenarios.

- OLD IT
- Months to availability
- Static
- Hard to scale
- Upfront capital investment
- Infrastructure focus

- CLOUD IT
- Minutes to availability
- Dynamic
- Trivial to scale
- Usage-based cost
- Application focus

EC2 Computing Capacity Defined

BERNARDGOLDEN

- Virtual machine, referred to as “instance”
- Many instance families, e.g.
  - High memory -- up to 4 TB of memory
  - Compute optimized -- latest Intel processors configured for AWS
  - Storage optimized – high bandwidth to fast SSD
- Instance pricing varies according to capacity

EC2 Computing Capacity Pricing

BERNARDGOLDEN

- On-demand – paid for by the ~~hour~~ second (October 2017) with no commitment
- Spot – Per-second, based on bid, preemptible. Provides significant savings if your application is flexible in terms of when it runs
- Reserved – provides guaranteed capacity, lower per-second pricing based on upfront time commitment or payment
  - Discount varies by length of time and size of payment
- Dedicated hosts – Physical server dedicated for your use. Allows you to use software licenses that do not support running in virtualized environment

On-demand

BERNARDGOLDEN

- Stop and start as desired
- No commitment
- Pay fixed price for use by second (note: pricing listed as per-hour)
- Use cases:
  - Applications with erratic traffic patterns
  - Transient execution
    - DevTest
    - Short-term websites – marketing promotions, campaigns, etc.

On-demand

BERNARDGOLDEN

- Most convenient
- Low cost and flexible
- Note:
  - AWS provides free tier for new accounts
  - Course stays primarily within free tier

Reserved

BERNARDGOLDEN

- Reduced per-second cost based on up-front commitment of use duration and (perhaps) money
- Use cases:
  - Applications with stable traffic patterns
  - Applications with stable base traffic and occasional spikes
- Provides per-second discount – up to 75%

Reserved

BERNARDGOLDEN

- Many options in terms of purchase, trading for different RI types, selling/buying on RI marketplace
- Can be tricky to manage
- Generally underused

Spot

BERNARDGOLDEN

- Bid submitted for maximum price you are willing to spend for computing capacity
- "Spot price fluctuates based on the supply and demand of available unused EC2 capacity."
- Instance launched when spot price drops lower than bid, terminated when spot price goes above bid
- Use cases:
  - Applications with flexible execution times
  - Very cost-sensitive applications
  - Applications requiring enormous computing power that can be parallelized at off hours

Spot

BERNARDGOLDEN

- Per-second pricing changes old pricing which charged for full hour when terminated
- Clever use of spot instances can reduce total cost of application
  - Pinterest reduced costs 60% with mix of on-demand, reserved, and spot
  - https://www.slideshare.net/AmazonWebServices/stp204-running-lean-on-aws

Spot

BERNARDGOLDEN

Spot Bid Advisor

Region: US East (N. Virginia) OS: Linux/UNIX Bid Price: 50% On-Demand

Instance type filter:

vCPU (min): 1 Memory GiB (min): 0 Instance types supported by EMR

Instance Type	vCPU	Memory GiB	Savings over On-Demand*	Frequency of being outbid (month)	Frequency of being outbid (week)
r3.xlarge	4	30.5	86%	Low	Low
c1.medium	2	1.7	80%	Low	Low
c3.large	2	3.75	84%	Low	Low
c3.xlarge	4	7.5	79%	Low	Low
c4.2xlarge	8	15	75%	Low	Low
c4.4xlarge	16	30	76%	Low	Low
c4.8xlarge	36	60	74%	Low	Low
c4.large	2	3.75	75%	Low	Low
c4.xlarge	4	7.5	74%	Low	Low
t1.micro	1	0.613	78%	Low	Low

Display all 72 instance types

Dedicated Hosts

BERNARDGOLDEN

- Provides dedicated server for customer
- No other customer's instances run on dedicated server
- Use cases:
  - Regulatory requirements that do not allow multi-tenant virtualization
  - Software licensing that does not allow license migration
  - Preference for not sharing server and/or performance concerns
- Pricing is per-hour
- Reserved host pricing offers discount
  - Same commitment options as Reserved Instances

Instance Types

BERNARDGOLDEN

Family	Description	Typical Uses
D2	Large amounts of SSD local storage	Big data, data warehousing, fileservers
R4	Memory optimized	High performance databases, caches, analytics
M4	General purpose with good mix of compute, memory, and network	Production application software
C4	Compute-optimized with latest Intel processors configured for AWS	Batch processing, scientific computing, video encoding
G3	GPU-attached with NVIDIA Tesla M60 GPU	Graphics-intensive 3D modeling, remote workstations
I3	Hi I/O with very fast SSD storage	NoSQL databases, Caches with file flushing
F1	Xilinx FPGA-attached allows user-selected algorithm configuration	Machine learning, genomic/finance/ seismic analysis
T2	General purpose with burstable performance	Small site web servers, developer environments
P2	GPU-attached with NVIDIA K80 GPU	Machine learning, genomic/finance/ seismic analysis
X1	Very large amounts of memory – up to 4 TB	In-memory databases (SAP HANA), High performance computing

Instance Types

BERNARDGOLDEN

- Correct type and size can affect application performance, efficiency, and cost
- Number of options can be confusing
- Rule of thumb:
  - Start with general type unless insight guides you to other type
  - Small application, start with T2
  - Large application, start with M4
  - Use CloudWatch and Trusted Advisor to provide guidance information to make better choice

Instance Storage: EBS

BERNARDGOLDEN

Amazon Elastic Block Store (Amazon EBS) provides persistent block storage volumes for use with Amazon EC2 instances in the AWS Cloud. Each Amazon EBS volume is automatically replicated within its Availability Zone to protect you from component failure and can be attached to any running instance that is in the same Availability Zone.

EBS Volume Types

BERNARDGOLDEN

- General Purpose SSD (GP2)
  - Balances price and performance
  - Use cases: typical applications, web servers
  - Performance based on volume size (1GB – 16 TB)
  - 3 IOPS per volume GB up to 10K IOPS
  - Minimum of 100 IOPS no matter what size volume
  - Can burst to 3K IOPS
- Provisioned IOPS SSD (IO1)
  - Provides consistent throughput (4 GB – 16 TB)
  - Use cases: i/o intensive workloads, e.g. databases
  - Up to 20K IOPS
  - Maximum IOPS/volume size of 50:1

EBS Volume Types

BERNARDGOLDEN

- Throughput optimized HDD (ST1)
  - Low cost magnetic storage
  - Optimized for throughput rather than IOPS
  - Good for sequential read workloads, e.g., Map Reduce
  - Cannot be a boot volume
- Cold HDD (SC1)
  - Lowest cost magnetic storage
  - Optimized for throughput rather than IOPS (throughput lower than ST1)
  - Designed for infrequently accessed data, e.g., archival file storage
  - Cannot be a boot volume

EBS Volume Types

BERNARDGOLDEN

- Magnetic (Standard)
  - Lowest cost magnetic storage that can be used as boot volume
  - Optimized for low throughput rather than IOPS (100 IOPS)
  - Designed for infrequently accessed data, e.g., archival file storage
  - Previous generation of HDD, AWS recommends ST1 or SC1

EC2 ACTIVITY: LAUNCH AN INSTANCE

EC2 Activity Description

BERNARDGOLDEN

- Launch an instance with new Security Group
- Steps:
  - Launch an AWS Linux AMI
  - Create a new security group
  - Update software with yum
  - Install and start Apache web server
  - Create index.html
  - Confirm web access
  - Terminate instance

PAUSE VIDEO

BERNARDGOLDEN

EC2 INSTANCE EXAM TIPS

EC2 Exam Tips

BERNARDGOLDEN

- Review of launch process
  - Launch AMI to create EC2 instance
  - Select AMI type
  - Choose purchase option (On-demand, Spot, Tenancy, Reserved)
  - Default VPC, Subnet (AZ)
  - IP address
  - Termination protection
  - Bootstrap data
  - CloudWatch monitoring
  - Select root volume type, size, and behavior on instance termination
  - Additional volumes if desired
  - Tags

EC2 Exam Tips

BERNARDGOLDEN

- Instance termination protection is off by default, turn on for … protection

BERNARDGOLDEN

EC2 SECURITY GROUPS

Security Groups Overview

BERNARDGOLDEN

- Software firewall resident on instance
- Rules control network traffic access
- Default:
  - All inbound traffic blocked until port opened in Security Group
  - Outbound: All ports open
- Source traffic can be all Internet, specific IP addresses/ranges, other Security Group

BERNARDGOLDEN

EC2 SECURITY GROUPS EXAM TIPS

Security Groups Exam Tips

BERNARDGOLDEN

- Software firewall resident on instance
- Rules control network traffic access
- Security Groups allow traffic
  - Cannot block traffic from specific IP addresses
- Default:
  - All inbound traffic blocked until port opened in Security Group
  - Outbound: All ports open
- Running instances can have Security Groups changed

Security Groups Exam Tips

BERNARDGOLDEN

- Security Group changes take effect immediately
- EC2 Security Groups are Stateful
  - Open inbound ports automatically allow outbound traffic
- Multiple instances can share a Security Group
- Instance can have multiple Security Groups
- Source traffic can be all Internet, specific IP addresses/ranges, other Security Group

BERNARDGOLDEN

EC2 EBS VOLUMES

EBS Volumes

BERNARDGOLDEN

- EBS is storage separate from instance
- Volumes are like disk drives
- EBS supports root and additional volumes
- Root volumes created at launch time
- Additional volumes must be attached
- Five types:
  - General purpose SSD (GP2)
  - Provisioned IOPS (IO1)
  - Throughput optimized HDD (ST1)
  - Cold HDD (SC1)

EC2 EBS VOLUMES EXAM TIPS

EBS Volumes Exam Tips

BERNARDGOLDEN

- EBS is storage separate from instance
- EBS supports root and additional volumes
- Five types:
  - General purpose SSD (GP2)
  - Provisioned IOPS (IO1)
  - Throughput optimized HDD (ST1)
  - Cold HDD (SC1)
  - Magnetic Standard – HDD root
- EBS root volumes deleted on instance termination by default

EBS Volumes Exam Tips

BERNARDGOLDEN

- EBS root volumes deleted on instance termination by default
- Additional volumes can be encrypted at creation
- EBS root volume cannot be encrypted at creation
  - Requires creating new AMI with root volume encrypted – launch of new AMI contains encrypted root volume

EBS Volumes Exam Tips

BERNARDGOLDEN

- Snapshots
  - Created at point-of-time, instance can continue operating
  - Incremental – only stores changed data from last snapshot
  - Stored in AWS-controlled storage in same region
  - Restoring Snapshot in another region requires copy
- Creating a volume from Snapshot
  - Lazy loading – volume created immediately and data brought in gradually
  - Requests for data will restored on first request

EBS Volumes Exam Tips

BERNARDGOLDEN

- Additional volumes must be attached and file system formatted
- Detaching maintains data on volume
- EBS Elastic Volumes
  - Change volume type, size, or performance while in operation
- Remember whether volume is terminated when instance is terminated
- Cost is based on volume size, not storage use