

AWS Spot Instances

Guide by R&D Solutions

Document Control

General Information			
Document Prepared By	M.Stefanov	Document Reviewed By	V. Girov
Approved By	M.Stefanov	Date	25.07.2018

Table of Contents

Document Control	1
Introduction	2
Spot instances	2
Pricing specifics	3
New Amazon EC2 Spot pricing model	3
How to check spot instance pricing	4
Requesting Spot instance	6
On-Demand instances	8
On-Demand Pricing specifics	8
Creating on-demand instance	9

Introduction

The document outlines basic information about [AWS Spot instances](#) and how such type of instances could be created and managed in AWS account.

Another topic that will be explained is the On-Demand instance types and how such could be created and maintained within AWS.

The information is specially created to serve the needs of AlgoLib LTD.

Spot instances

Spot instances are specific types of instances that offer solid discounts compared to the On-Demand ones. This way AWS is selling the spare compute capacity available. The hourly price for a Spot instance is called Spot price. The price is different based on the Availability Zone and Region. Your Spot Instance runs whenever capacity is available and the maximum price per hour for your request exceeds the Spot price.

Something very important about Spot instances is that they could be stopped and terminated with two minutes of notification when Elastic Compute Cloud Service needs capacity. This is also one of the differences with On-Demand instance types.

You pay the Spot price for Spot Instances, which is set by Amazon EC2 and adjusted gradually based on the long-term supply of and demand for Spot Instances. If the maximum price for your request exceeds the current Spot price, Amazon EC2 fulfills your request if capacity is available. Your Spot Instances run until you terminate them, capacity is no longer available, or the Spot price exceeds your maximum price.

Main differences between Spot instances and On-Demand

	Spot Instances	On-Demand Instances
Launch time	Can only be launched immediately if the Spot Request is active and capacity is available.	Can only be launched immediately if you make a manual launch request and capacity is available.
Available capacity	If capacity is not available, the Spot Request continues to automatically make the launch request until capacity becomes available.	If capacity is not available when you make a launch request, you get an insufficient capacity error (ICE).
Hourly price	The hourly price for Spot Instances varies based on demand.	The hourly price for On-Demand Instances is static.
Instance interruption	You can't stop and start an Amazon EBS-backed Spot Instance; only the Amazon EC2 Spot service can do this. The Amazon EC2 Spot service can interrupt an individual Spot Instance if capacity is no longer available, the Spot price exceeds your maximum price, or demand for Spot Instances increases.	You determine when an On-Demand Instance is interrupted (stopped or terminated).

Additional articles:

- <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-spot-instances.html>

- <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/how-spot-instances-work.html>

Pricing specifics

Please, find below two scenarios and explanation of the pricing details for Spot instances:

- **Scenario 1:** You bid \$0.25 on a spot instance. The current price at the top of the hour, say 1:00am, is \$0.20 hence you successfully bid for your instance. At 2:00am the price is still at \$0.20 (unchanged). You complete your job and the instance terminates at 2:30am. What is your cost?
- **Scenario 2:** You bid \$0.25 on a spot instance. The current price at the top of the hour, say 1:00am, is \$0.20 hence you successfully bid for your instance. At 2:30am the price changes to \$0.30. Hence, you are "outbid" and your instance terminates. Now, what is your cost?

Note on spot pricing: You don't pay by your bid. The pay is based on AWS' spot price at that moment. So if you bid \$0.25 and the Spot price is \$0.20, you're paying \$0.20.

- **Scenario 1 - Answer:** Both hours your bid is above the spot price so your instance is fulfilled. But because you, yourself, terminated it, you're charged for the full hour. Therefore, $0.20 + 0.20 = 0.40$.
- **Scenario 2 - Answer:** The first hour is bid is above spot price. You're good. But the second hour comes, and spot price goes to 0.30, and that's above your bid. Now what? Amazon will give you a two-minute notification before they will terminate the instance FOR you, and in that case, you are not charged for partial use of that hour. So, $0.20 + 0.00 = 0.20$.

New Amazon EC2 Spot pricing model

At re:Invent 2017, AWS launched a new pricing model that simplified the Spot purchasing experience. The new model gives you predictable prices that adjust slowly over days and weeks, with typical savings of 70-90% over On-Demand. With the previous pricing model, some of you had to invest time and effort to analyse historical prices to determine your bidding strategy and maximum bid price. Not anymore.

You don't have to bid for Spot Instances in the new pricing model, and you just pay the Spot price that's in effect for the current hour for the instances that you launch. It's that simple. Now you can request Spot capacity just like you would request On-Demand capacity, without having to spend time analysing market prices or setting a maximum bid price.

In the new model, the Spot prices are more predictable, updated less frequently, and are determined by supply and demand for Amazon EC2 spare capacity, not bid prices.

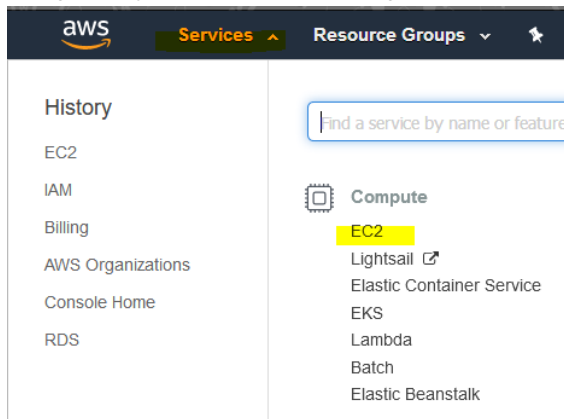
Additional articles:

- <https://aws.amazon.com/blogs/compute/new-amazon-ec2-spot-pricing/>

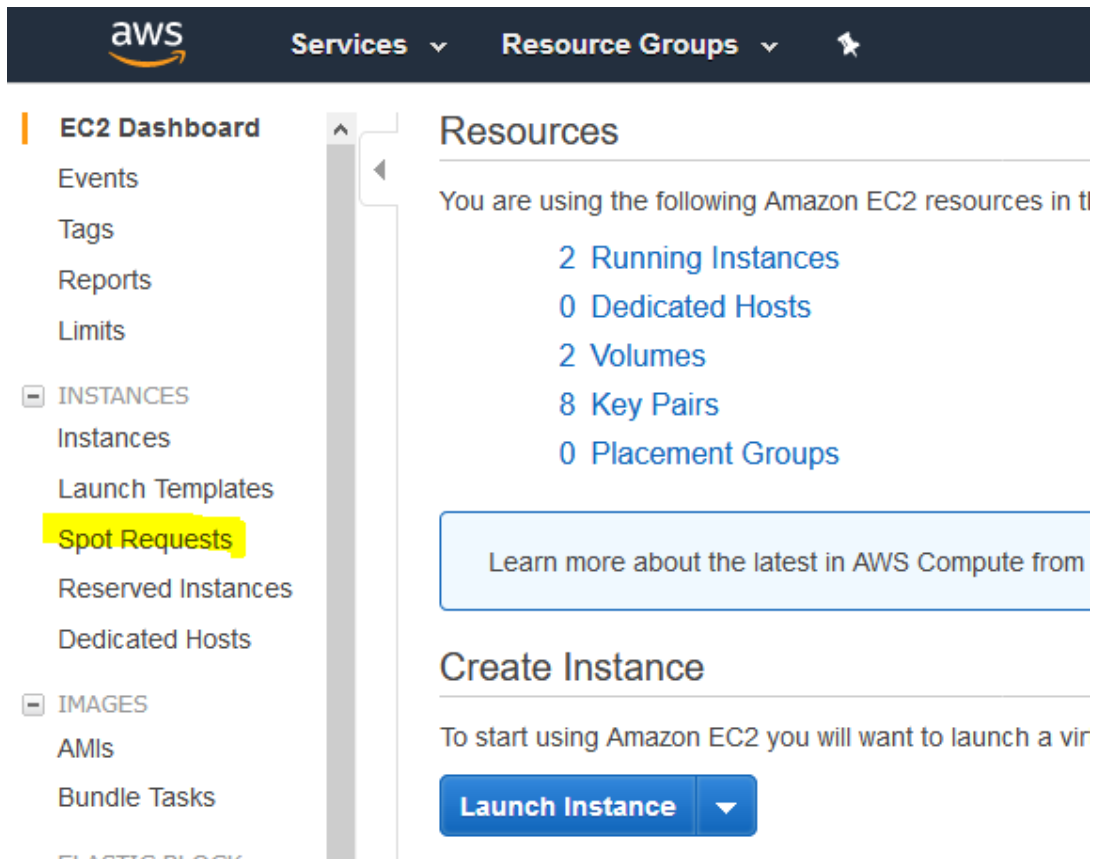
- <https://aws.amazon.com/about-aws/whats-new/2017/11/amazon-ec2-spot-introduces-new-pricing-model-and-the-ability-to-launch-new-spot-instances-via-runinstances-api/>

How to check spot instance pricing

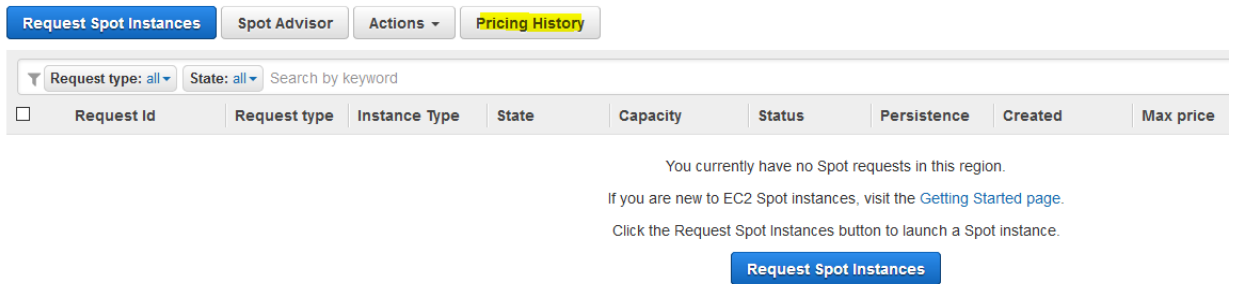
1. Navigate to your AWS account and go to Services -> EC2



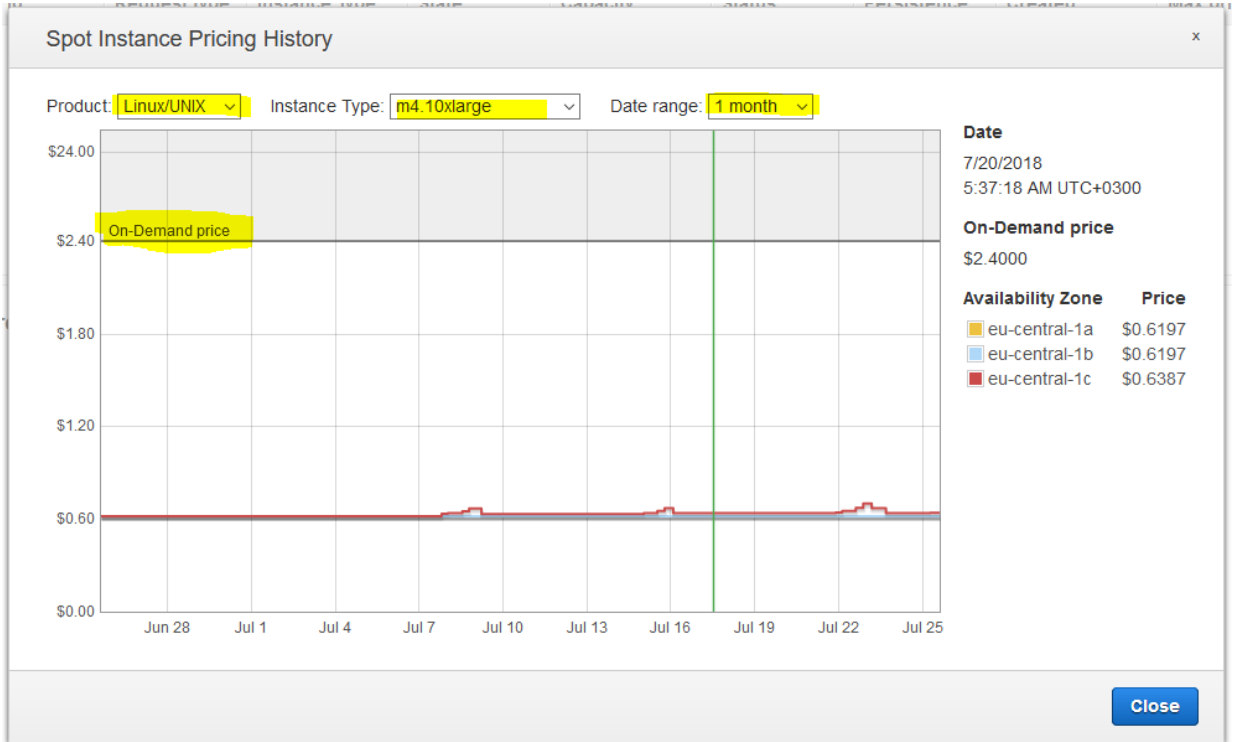
2. You will be redirected to the EC2 Dashboard. From there select **Spot Requests**



3. You will be presented with the Spot Instances Dashboard



4. Click on **Pricing history**. You will be presented a form from where you could observe the price fluctuation based on certain type of machine and concrete interval of time

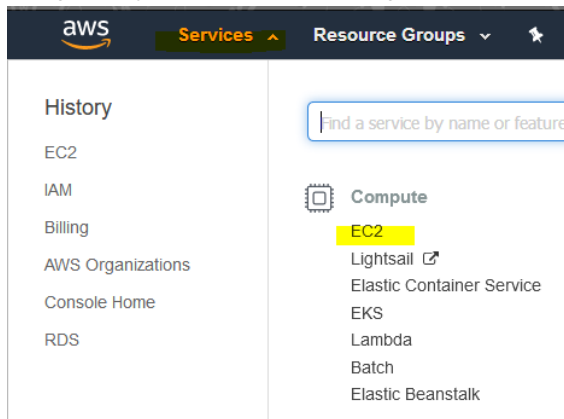


5. Using this diagram, you could see the usual On-Demand price along with the Spot price and how it changes.

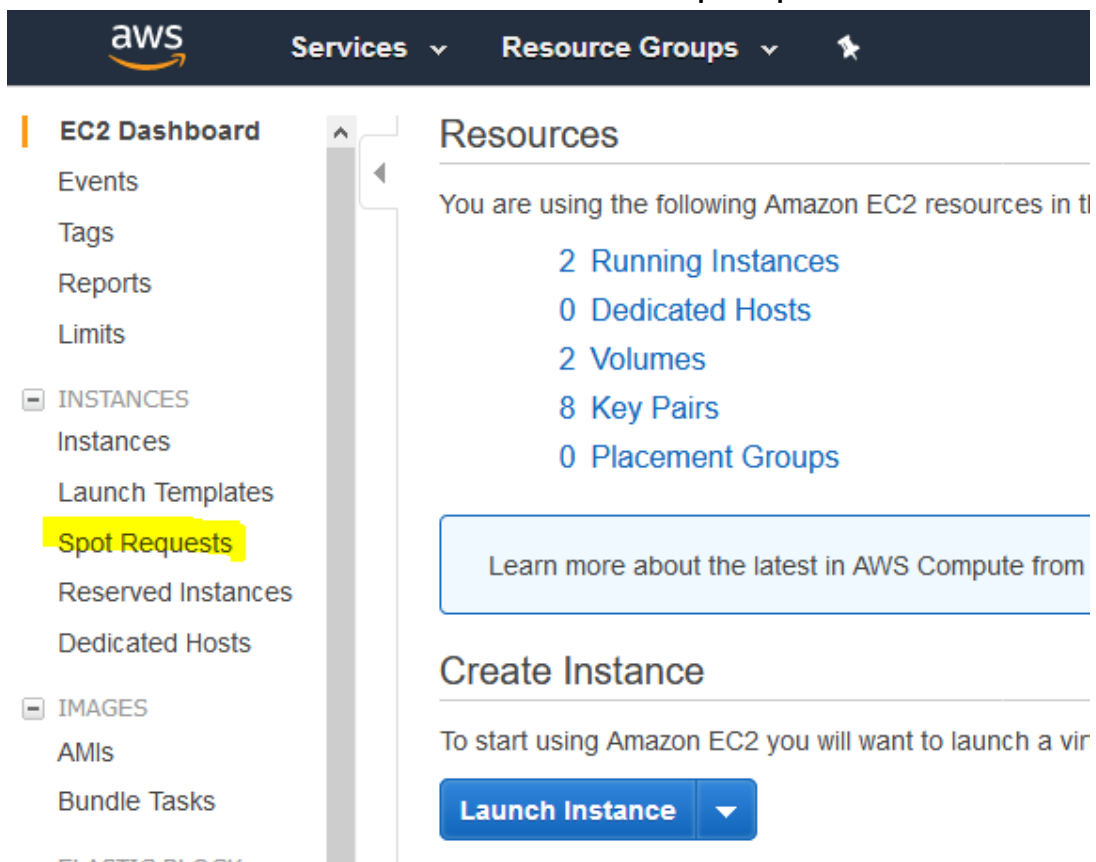
6. Based on this information you could choose appropriate instance type for you and request it

Requesting Spot instance

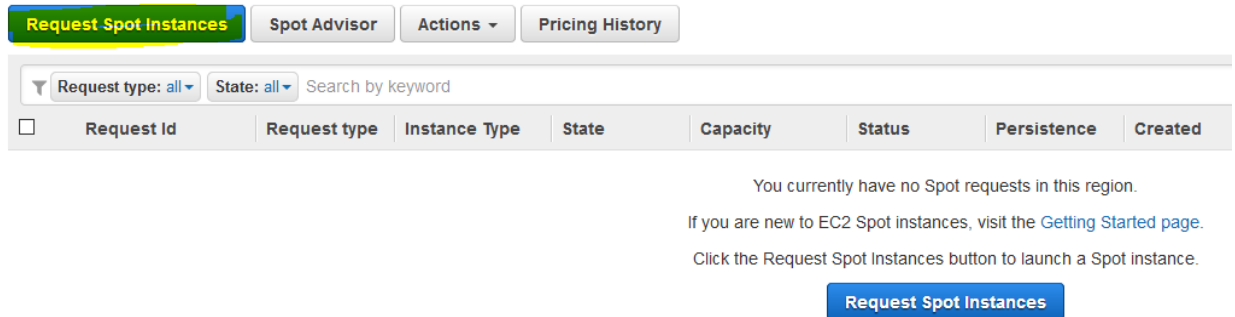
1. Navigate to your AWS account and go to Services -> EC2



2. You will be redirected to the EC2 Dashboard. From there select **Spot Requests**



3. You will be presented with the Spot Instances Dashboard



4. Select “Request Spot Instances”

5. You will be presented with a page where you need to define

- a. Request type
 - i. Request – one-time Spot instance request
 - ii. Request and maintain – request a fleet of Spot instances to maintain target capacity
 - iii. Reserve for duration - request a Spot instance with no interruption for 1 to 6 hours (a Spot block)
- b. Amount
 - i. Total target capacity – number of instances you would like to launch
 - ii. Optional On-Demand portion – a portion of total target capacity could be designated as On-Demand.
- c. Requirements
 - i. Template – if you have saved a configuration for Spot Request before as JSON, you could use it here. At first you won't have such, but eventually you could create such using “[Launch Templates](#)” option.
 - ii. Amazon Machine Image – whether you are going to use Linux AMI, Ubuntu AMI, Windows AMI, etc. There are such predefined AMIs in AWS free Workspace
 - iii. Instance type(s) – e.g. m4.xlarge, c3.large – depending on your need
 - iv. Network – in which VPC these machines will live. For the moment you will have a single VPC (Virtual Private Cloud)
 - v. Availability Zone - you could pick up in which Data Centre (Availability Zone) the instance could be spawn. Usually we provision within the default VPC 3 public subnets which are span across three Availability Zones
 - vi. EBS Volumes – the size of the disk storage and number of disks. These disks remain active after the machine gets deleted. To clean these, you need to go to Volumes in EC2 dashboard and delete the unnecessary volumes.
 - vii. EBS Optimization – if additional throughput between virtual machine and disk is not necessary, do not check this option
 - viii. Instance store – this is disk space which gets destroyed when the machine is deleted
 - ix. Monitoring – enabled detailed monitoring on the instance – the information is gathered each minute. If not checked the information is gathered every five mins.

- x. Security groups – these are like firewalls - they open certain ports on the machine and allow communication over these ports. If not specifically opened, the ports are considered closed.
 - xi. Auto-assign IPv4 Public IP – if you want to access this machine from the internet and not only in the internal network, mark this field as **Enabled**
 - xii. **Key pair name** – in the beginning you will not have active key pairs which are used to SSH in the machine or to decrypt the Admin password (in case you are using Windows machine). You will need to create key pair. Make sure to download the key pair since this could be done only upon creation of the key pair. If you lose it, the machine will not be accessible. We suggest (at least) to have separate key pairs for each Client`s fleet of instances.
 - xiii. User data – here you could provide script that is executed immediately after the machine is launched. For example, here you could put bash script that configures Apache and installs php or python.
 - xiv. Tags – if you want to put some names on the machines or other types of tags that could be used to distinguish the machines from one another or to run a report.
- d. Load balancing
- i. In case you are going to create one or more instances that need to be under load balancer you need to check the “Load balancing” option
- e. Spot request fulfilment – different options for the fulfilment of the request – from and until when this request is valid, set maximum price for your spot instance or fleet of instances, choose allocation strategy (whether to use only cheapest spot instance or have it balanced across AZs – this is important since in each AZ the price might be different)

Additional links:

- <https://www.youtube.com/watch?v=eWw5C8ucacY>
- <https://www.youtube.com/watch?v=Phae1e1hvuM>

On-Demand instances

With On-Demand instances, you pay for compute capacity by per hour or per second depending on which instances you run. No longer-term commitments or upfront payments are needed. You can increase or decrease your compute capacity depending on the demands of your application and only pay the specified per hourly rates for the instance you use.

On-Demand Pricing specifics

The On-Demand instance pricing could be calculated using the AWS Simple Monthly Calculator:

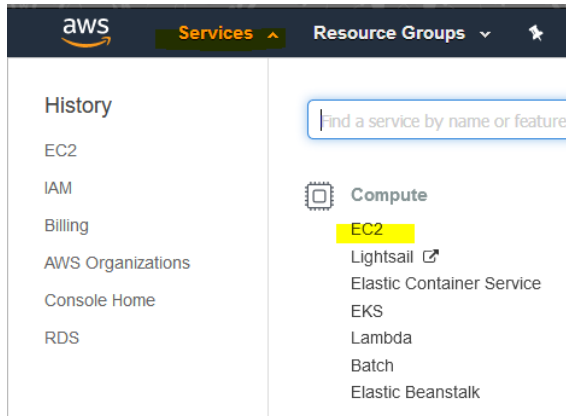
- <https://calculator.s3.amazonaws.com/index.html>
- <https://www.youtube.com/watch?v=54TVcUeOoAc>

Using it you could calculate the cost of your resources on monthly basis. The pricing here is more straightforward than the Spot instances.

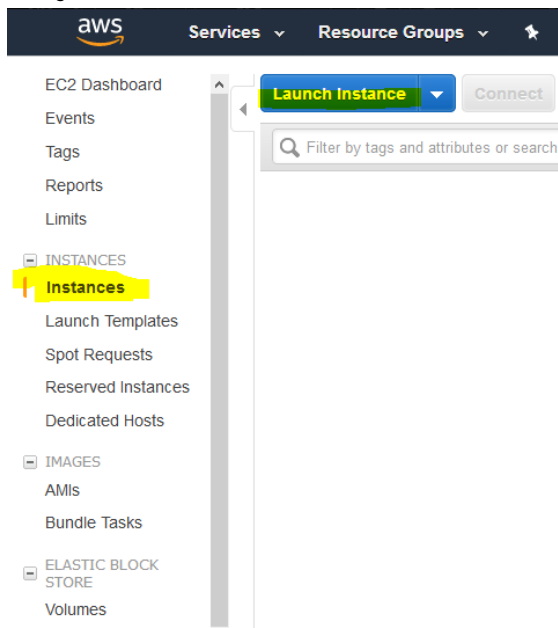
Creating on-demand instance

There is a wizard base workflow which guides you through the creation process. To execute it you need to:

1. Navigate to your AWS account and go to Services -> EC2



2. Navigate to Instances and click on "Launch instance"



3. You will be redirected to the wizard base process which consists of the following steps:
 - a. Choose AMI (Amazon Machine Image)
 - b. Choose Instance Type
 - c. Configure Instance
 - d. Add Storage
 - e. Add Tags

- f. Configure Security group
- g. Review and create

The following videos show these steps in details:

- https://www.youtube.com/watch?v=zNYnQ_1pwao
- <https://www.youtube.com/watch?v=yVXhj3u6-wI>

NB* In order only you to be able to access the newly created machine, you need to:**

- **Either create new key pair and create the machine using this key pair. You need to make sure you download this key and not share with anyone else.**
- **Use already existing key that only you have access to it.**

To limit the access to your machines, you could also create separate Security Group which allows for example SSH or RDP access only from certain IP addresses.

