



# Spot Fleet

Sanjay Padhi

Brown University/

University of California, San Diego

# Introduction

## What is a spot instance



Spare Capacity at  
Large discounts:  
Sometimes 90% lower

Bid or Auction based  
pricing model

Demand -Supply  
Economics

## Define your own price for EC2

- Market where price changes based on supply and demand
- If bid price exceeds spot market price: Instance is launched
- If market price exceeds bid price: Instance is terminated
  - With 2 minutes warning (Exercise: How to trap this warning automatically)

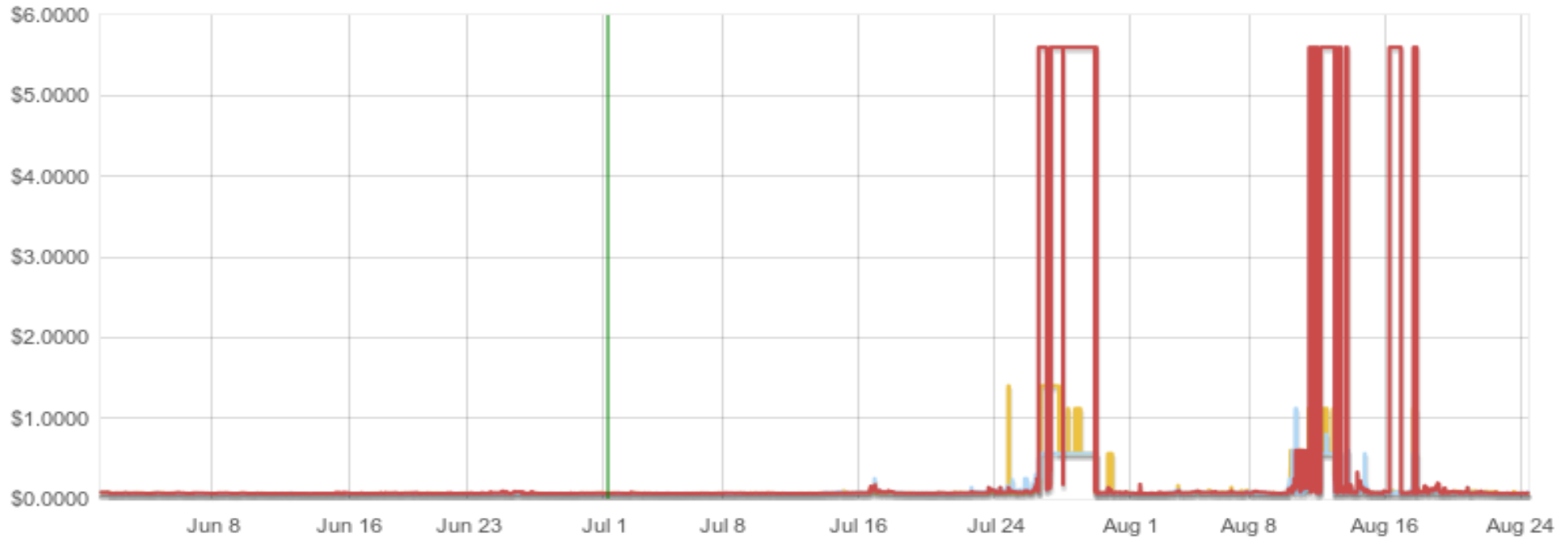
## Unused EC2 Instances provides the compute capacity




# Spot Instance Pricing History

## Spot Instance Pricing History



Product : **Linux/UNIX** ▾ Instance type: **m3.2xlarge** ▾ Date range : **3 months** ▾ Availability zone: **All zones** ▾



Availability zone	Price
 <b>us-west-2a</b>	\$0.0667
 <b>us-west-2b</b>	\$0.0664
 <b>us-west-2c</b>	\$0.0676

**Date** July 1, 2015 at 8:20:53 AM UTC+2

# Spot Fleets – Simplifies life

Each request (via the API or the CLI) must include the following values:

- **Target Capacity** – The number of EC2 instances that you want in your fleet.
- **Maximum Bid Price** – The maximum bid price that you are willing to pay.
- **Launch Specifications** – The quantities and types of instances that you would like to launch, and how you want them to be configured (AMI Id, VPC, subnets or availability zones, security groups, block device mappings, user data, and so forth). In general, launch specifications that do not target a particular subnet or availability zone are more economical.
- **IAM Fleet Role** – The name of an IAM role. It must allow EC2 to terminate instances on your behalf.

*Optional:*

**Client Token** - A unique, case-sensitive identifier for the request

**Valid From** - The start date and time

**Valid Until** - The end date and time

**Terminate on Expiration** - TRUE (all spot instance in the flight will terminate)

# Spot Fleet Prerequisites

## Amazon Resource Name (ARN)

- Create an IAM role that grants the Spot fleet service permission to terminate instances on your behalf as follows:
  1. Open the IAM console at <https://console.aws.amazon.com/iam/>.
  2. In the navigation pane, click **Roles**, and then click **Create New Role**.
  3. On the **Set Role Name** page, enter a name for the role and then click **Next Step**.
  4. On the **Select Role Type** page, click **Select** next to **Amazon EC2 Spot Fleet Role**.
  5. On the **Attach Policy** page, select the `AmazonEC2SpotFleetRole` policy, and then click **Next Step**.
  6. On the **Review** page, click **Create Role**.

## ARN Example:

The screenshot shows the AWS IAM console interface. On the left is a navigation pane with 'Roles' selected. The main content area shows the 'Summary' for a role named 'fleetAWS'. A red box highlights the 'Role ARN' field, which contains the value: `arn:aws:iam::12345678:role/xxxx`. The 'Instance Profile ARN(s)' field is empty, the 'Path' is '/', and the 'Creation Time' is '2015-08-24 10:28 UTC+0200'.

Property	Value
Role ARN	arn:aws:iam::12345678:role/xxxx
Instance Profile ARN(s)	
Path	/
Creation Time	2015-08-24 10:28 UTC+0200

# Let us get started - Launch a request

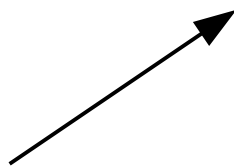
```
spadhi@Latitude-E6420:~/Work/DATA/AWS/SpotFleet$ cat lowest_price_az.json
```

```
{
  "SpotPrice": "0.25",
  "TargetCapacity": 20,
  "IamFleetRole": "arn:aws:iam::123456789:role/XXXXXXXXAWS",
  "LaunchSpecifications": [
    {
      "ImageId": "ami-7fc59d4f",
      "InstanceType": "m3.medium",
      "SubnetId": "subnet-24e72653"
    }
  ]
}
```

```
spadhi@Latitude-E6420:$ # Launch the spot fleet request
```

```
aws ec2 request-spot-fleet --spot-fleet-request-config file://lowest_price_az.json
```

```
{
  "SpotFleetRequestId": "sfr-464c8b1e-1605-4bf7-8138-ce20912ccfa2"
}
```



**Spot Fleet request ID**

# Let us get started – Monitor the request

```
spadhi@Latitude-E6420:~$ aws ec2 describe-spot-fleet-requests
spadhi@Latitude-E6420:~$
aws ec2 describe-spot-fleet-requests --spot-fleet-request-id
sfr-aa814ad2-c801-4856-a874-5de0e8b587ef
{
  "SpotFleetRequestConfigs": [
    {
      "SpotFleetRequestId": "sfr-aa814ad2-c801-4856-a874-5de0e8b587ef",
      "SpotFleetRequestConfig": {
        "TargetCapacity": 2,
        "LaunchSpecifications": [
          {
            "SubnetId": "subnet-24e72653",
            "EbsOptimized": false,
            "InstanceType": "m3.medium",
            "ImageId": "ami-7fc59d4f"
          }
        ],
        "SpotPrice": "0.25",
        "IamFleetRole": "arn:aws:iam::12345678:role/XXXXAWS"
      },
      "SpotFleetRequestState": "active"
    }
  ]
}
```

*# Request history: aws ec2 describe-spot-fleet-request-history*

# Monitor request: Cloud Watch / Brown Central Manager

[Launch Instance](#) [Connect](#) [Actions](#)

Filter by tags and attributes or search by keyword

Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS	Public IP
	i-708b32b6	t1.micro	us-west-2b	running	2/2 checks ...	None	ec2-54-149-33-250.us-...	54.149.33.250

Instance: **i-708b32b6** Public DNS: **ec2-54-149-33-250.us-west-2.compute.amazonaws.com**

[Description](#) [Status Checks](#) [Monitoring](#) [Tags](#)

**CloudWatch alarms:** ✔ No alarms configured [Create Alarm](#)

**CloudWatch metrics:** Basic monitoring. [Enable Detailed Monitoring](#) Showing data for: Last Hour

Below are your CloudWatch metrics for the selected resources (a maximum of 10). Click on a graph to see an expanded view. All times shown are in UTC. [View all CloudWatch metrics](#)

**CPU Utilization (Percent)**

**Disk Reads (Bytes)**

**Disk Read Operations (Operations)**

**Disk Writes (Bytes)**

```
-bash-4.1$ condor_status
```

Name	OpSys	Arch	State	Activity	LoadAv	Mem	ActvtyTime
ip-172-31-30-218.u	LINUX	X86_64	Claimed	Busy	0.000	590	0+00:27:49
	Machines	Owner	Claimed	Unclaimed	Matched	Preempting	
	X86_64/LINUX	1	0	1	0	0	0
	Total	1	0	1	0	0	0



# Canceling a Spot Fleet Request

```
spadhi@Latitude-E6420:$  
aws ec2 cancel-spot-fleet-requests --spot-fleet-request-ids  
sfr-464c8b1e-1605-4bf7-8138-ce20912ccfa2 --terminate-instances  
{  
  "SuccessfulFleetRequests": [  
    {  
      "SpotFleetRequestId": "sfr-464c8b1e-1605-4bf7-8138-ce20912ccfa2",  
      "CurrentSpotFleetRequestState": "cancelled_terminating",  
      "PreviousSpotFleetRequestState": "active"  
    }  
  ],  
  "UnsuccessfulFleetRequests": []  
}
```

```
spadhi@Latitude-E6420:$  
aws ec2 cancel-spot-fleet-requests --spot-fleet-request-ids  
sfr-aa814ad2-c801-4856-a874-5de0e8b587ef --terminate-instances  
{  
  "SuccessfulFleetRequests": [  
    {  
      "SpotFleetRequestId": "sfr-aa814ad2-c801-4856-a874-5de0e8b587ef",  
      "CurrentSpotFleetRequestState": "cancelled_terminating",  
      "PreviousSpotFleetRequestState": "active"  
    }  
  ],  
  "UnsuccessfulFleetRequests": []  
}
```

# Spot Fleet Current Boundaries

## Spot Fleet Limits

The usual Amazon EC2 limits apply to instances launched by a Spot fleet, such as Spot bid price limits, instance limits, and volume limits. In addition, the following limits apply:

- The number of active Spot fleets per region: 1,000
  - The number of launch specifications per fleet: 20
  - The size of the user data in a launch specification: 16 KB
  - The number of instances per Spot fleet: 3,000
  - The number of instances in all Spot fleets in a region: 3,000
  - A Spot fleet request can't span regions.
  - A Spot fleet request can't span different subnets from the same Availability Zone.
- Spot fleets aren't supported by other services (for example, Auto Scaling, Elastic Load Balancing, Amazon CloudWatch, and Amazon EMR).

<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/spot-fleet.html>

In case of questions: Send me an email or  
contact Burt (burt@fnal.gov) for the next AWS Spot Fleet Tutorial at FNAL