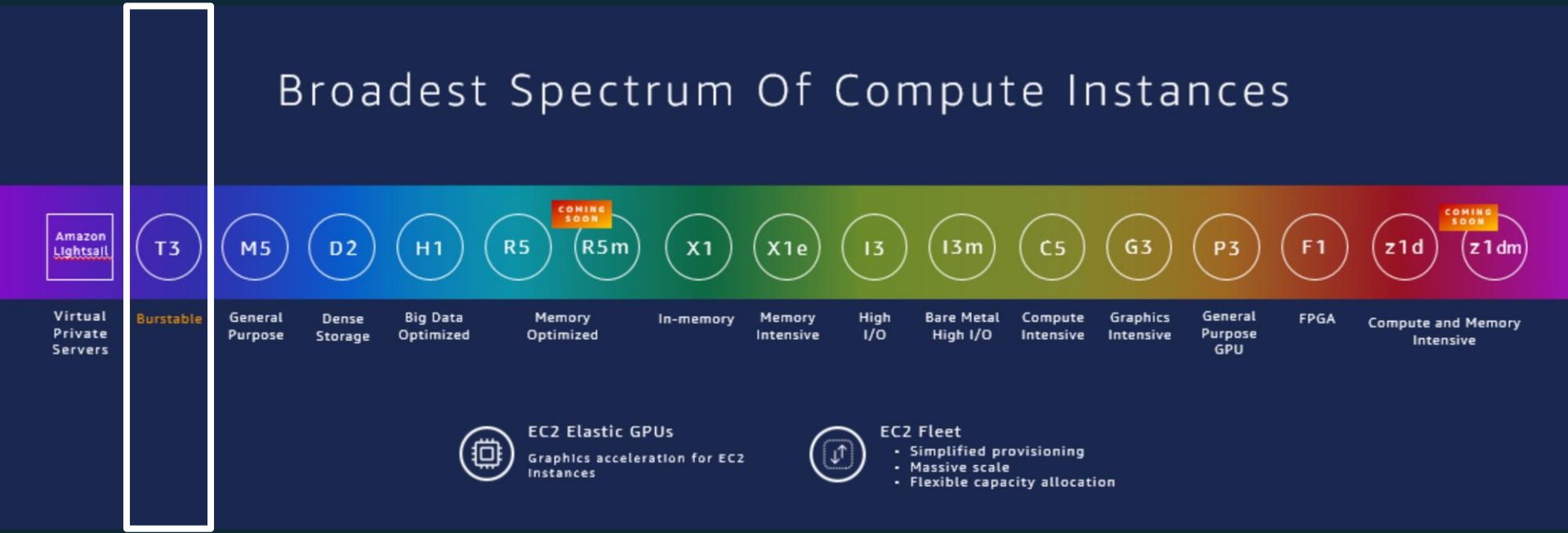
Introducing New Amazon EC2 T3 Instances - General Purpose Burstable Instances

Meena Gowdar, Senior Product Manager – AWS EC2

September 26th, 2018

What are EC2 Burstable Instances?

Amazon EC2 Instances



What are burstable instances?

- ✓ Offers a guaranteed level of CPU performance
- ✓ Ability to burst to high levels of CPU use for transient workloads
- ✓ Each instance has a “baseline” CPU performance governed by CPU credits
- ✓ Earn CPU credits when running below baseline
- ✓ Burn CPU credits when running above baseline

aws

Introducing EC2 T3 Instances

Introducing T3 Instances

- ✓ Launched on August 21, 2018
- ✓ Next Generation burstable instance
- ✓ ~ 30% improved price-performance over the previous generation
- ✓ Powered by High Frequency Intel **Xeon Platinum** 8000-series “Skylake” processors

aws

Introducing T3 Instances

- ✓ Built on EC2's new hypervisor - the AWS **Nitro** System provides more access to CPU resources as well as high performance networking and storage resources
- ✓ Twice the vCPUs on smaller instance sizes compared to T2
- ✓ **Intel Hyperthreading** enabled by default

aws

Introducing T3 Instances

- ✓ Improved Network performance across all instance sizes offering up to 5Gbps at burst
- ✓ EBS Optimized instances
- ✓ Improved EBS performance offering up to 1.5Gbps on nano thru medium and 2.3Gbps on large, xlarge and 2xlarge and burst IOPS up to 16KB
- ✓ Only EBS storage (no instance storage available)

aws

What sizes do EC2 Burstable Instances come in?

What sizes do burstable instances come in?

Instance Size	vCPU count	Memory size	EBS burst Bandwidth	Network burst Bandwidth
t3.nano	2	0.5 GiB	1.5 Gbps	5 Gbps
t3.micro	2	1 GiB	1.5 Gbps	5 Gbps
t3.small	2	2 GiB	1.5 Gbps	5 Gbps
t3.medium	2	4 GiB	1.5 Gbps	5 Gbps
t3.large	2	8 GiB	2.3 Gbps	5 Gbps
t3.xlarge	4	16 GiB	2.3 Gbps	5 Gbps
t3.2xlarge	8	32 GiB	2.3 Gbps	5 Gbps

aws

What sizes do burstable instances come in?

Instance Size	vCPU count	Memory size	CPU credits earned per hour	Max. CPU credits earned
t3.nano	2	0.5 GiB	6	144
t3.micro	2	1 GiB	12	288
t3.small	2	2 GiB	24	576
t3.medium	2	4 GiB	24	576
t3.large	2	8 GiB	36	864
t3.xlarge	4	16 GiB	96	2304
t3.2xlarge	8	32 GiB	192	4608

aws

When Would I Use EC2 Burstable Instances?

When would I use EC2 burstable instances?

- ✓ Most general purpose instances that don’t need fixed CPU resources
- ✓ Applications that occasionally need quick access to high CPU
- ✓ Idle or need moderate CPU for majority of the time
- ✓ Need to burst CPU whenever and for as long as needed

aws

When would I use EC2 burstable instances?

T3 Instance Size	vCPU count	Memory size	M5 Instance Size
t3.large	2	8 GiB	m5.large
t3.xlarge	4	16 GiB	m5.xlarge
t3.2xlarge	8	32 GiB	m5.2xlarge

aws

When would I use EC2 burstable instances?

	vCPUs	T3 price/hr	M5 price/hr	Difference	T3 Baseline Usage	Equivalent vCPU Minutes	Additional Charge per vCPU hour	Additional Charge per vCPU minute	Additional Minutes Allowed	Equivalent Utilization Allowed	Equivalent Total Utilization
A	B	C	D	E = D - C	F	G = F * 60 *A	H	I = H / 60	J = E / I	K = (J /60) / A	L = F + K
large	2	$0.0835	$0.0960	$0.0125	30%	36	$0.05	$ 0.000833	15	12.5%	42.5%
xlarge	4	$0.17	$0.1920	$0.0250	40%	96	$0.05	$ 0.000833	30	12.5%	52.5%
2xlarge	8	$0.3340	$0.3840	$0.0500	40%	192	$0.05	$ 0.000833	60	12.5%	52.5%

** based on N. Virginia region OD price*

aws

When would I use EC2 burstable instances?

- ✓ Web & App Servers
- ✓ Enterprise Servers
- ✓ Small and medium databases
- ✓ Dev & Test Environment
- ✓ Gaming Servers
- ✓ Caching Fleets
- ✓ Analytics Applications
- ✓ Micro-services
- ✓ Low-latency interactive applications
- ✓ Virtual desktops
- ✓ Build and stage environments

aws

How Do EC2 Burstable Instances Work?

How does T3 instances work?

- ✓ T3 instances use a CPU Credit Balance to burst above baseline performance
- ✓ One CPU credit is equal to one vCPU running at 100% utilization for one minute
- ✓ T3 instances operate in two modes
  - ✓ Standard
  - ✓ Unlimited (Default)
- ✓ T3 does not offer Launch credits

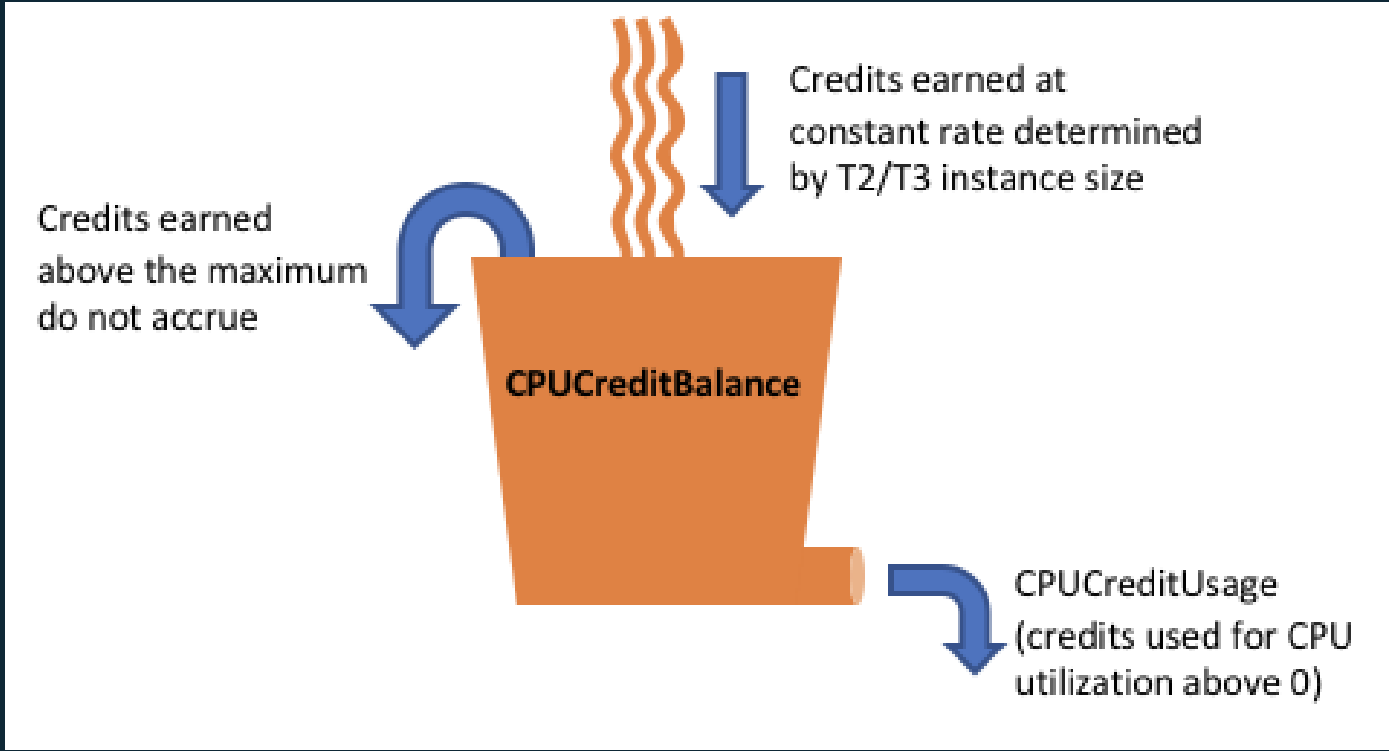
aws

T3 CPU Credit Mechanism

- ✓ “token bucket” model
- ✓ Credits are earned at a rate based on baseline performance
- ✓ Maximum accumulation (CPU Credit Balance) is based on instance size
- ✓ For running instances, credits never expire.
- ✓ For stopped instances, credits are stored for up to 7 days and then they expire
- ✓ Credits expire when an instance is terminated
- ✓ Per second granularity

aws

CPU Credits Accrual Limits - Use case



T3 Standard Mode

Burst above baseline using CPU credits

Get throttled back to baseline when CPU credits are exhausted

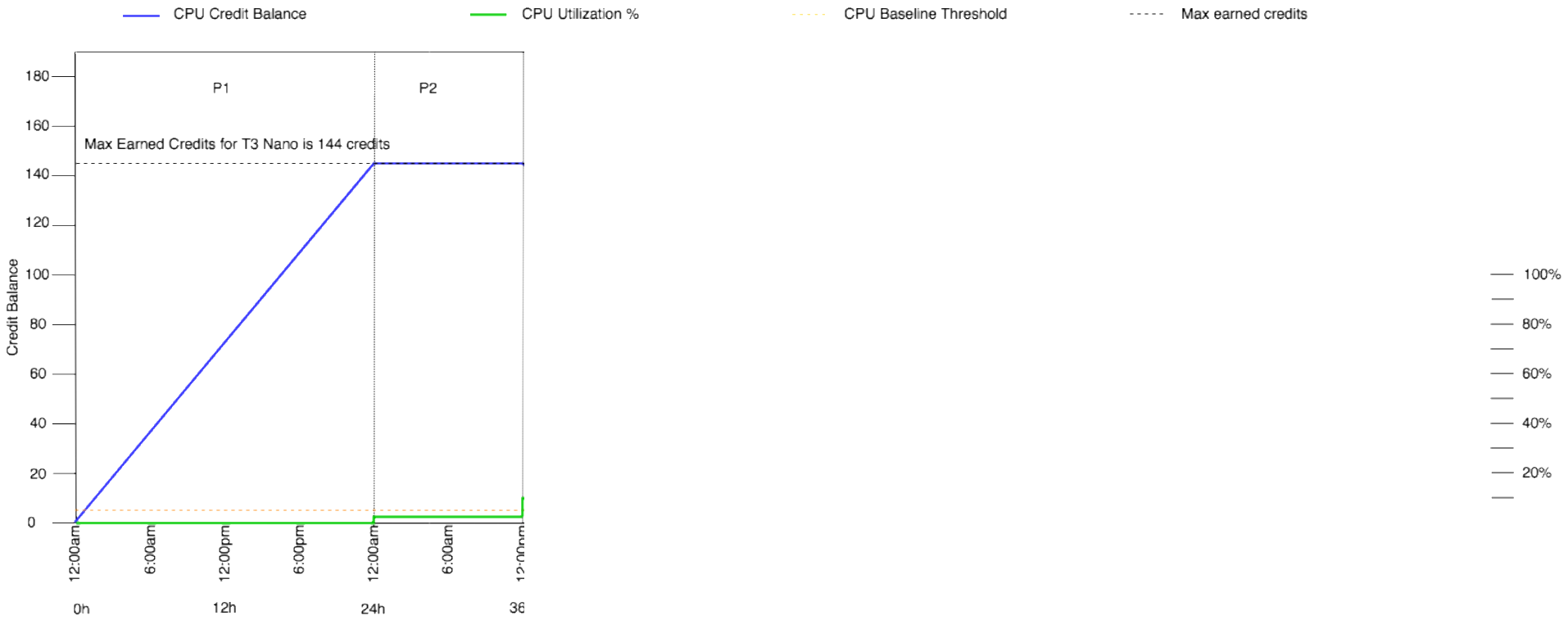
Earn CPU credits when running below baseline

CPU credits tracked by CloudWatch metric ***CPUCreditBalance***

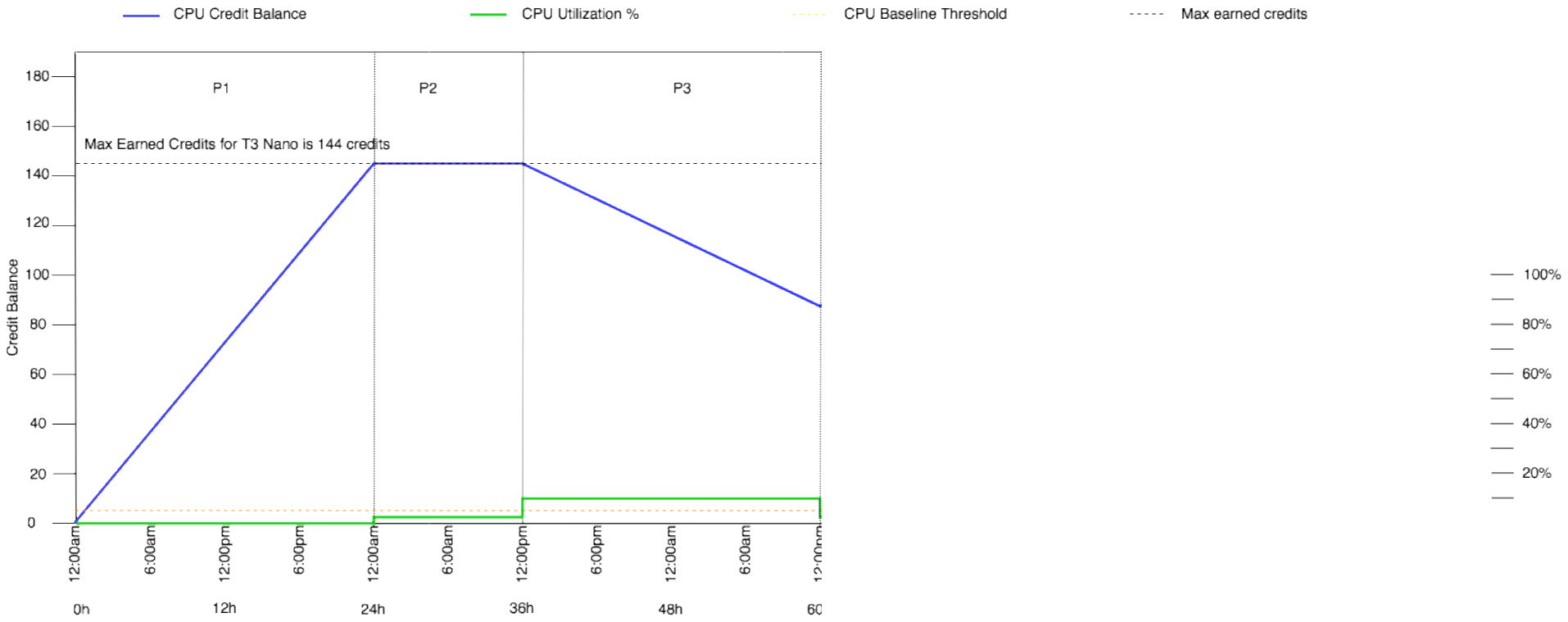
aws

T3 Standard Mode

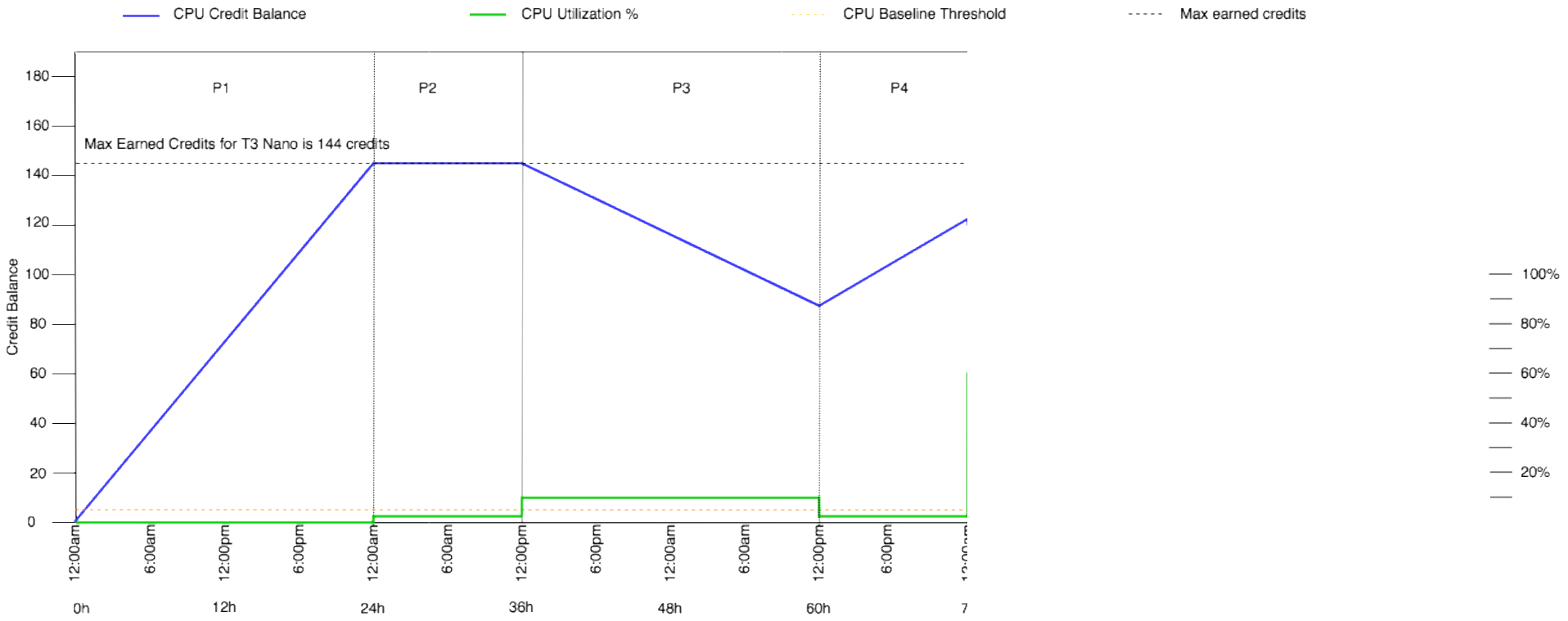
T3 Standard Mode



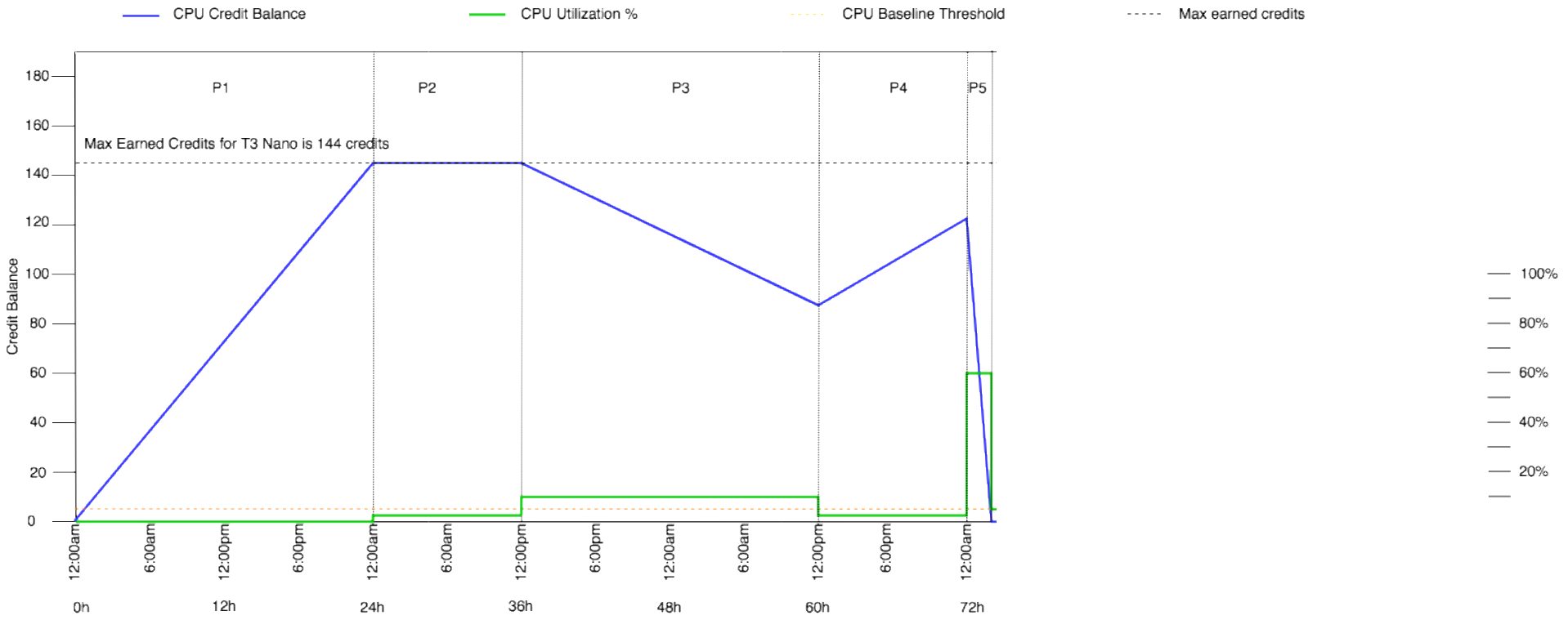
T3 Standard Mode



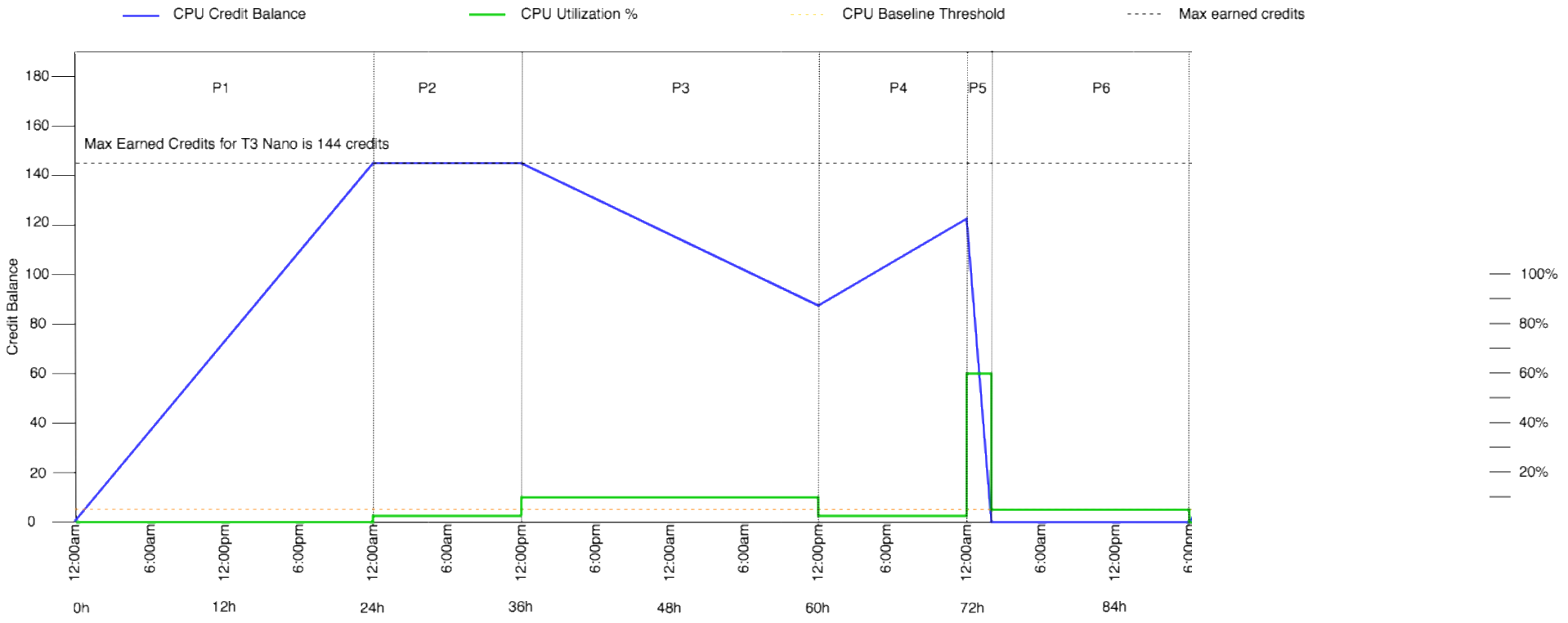
T3 Standard Mode



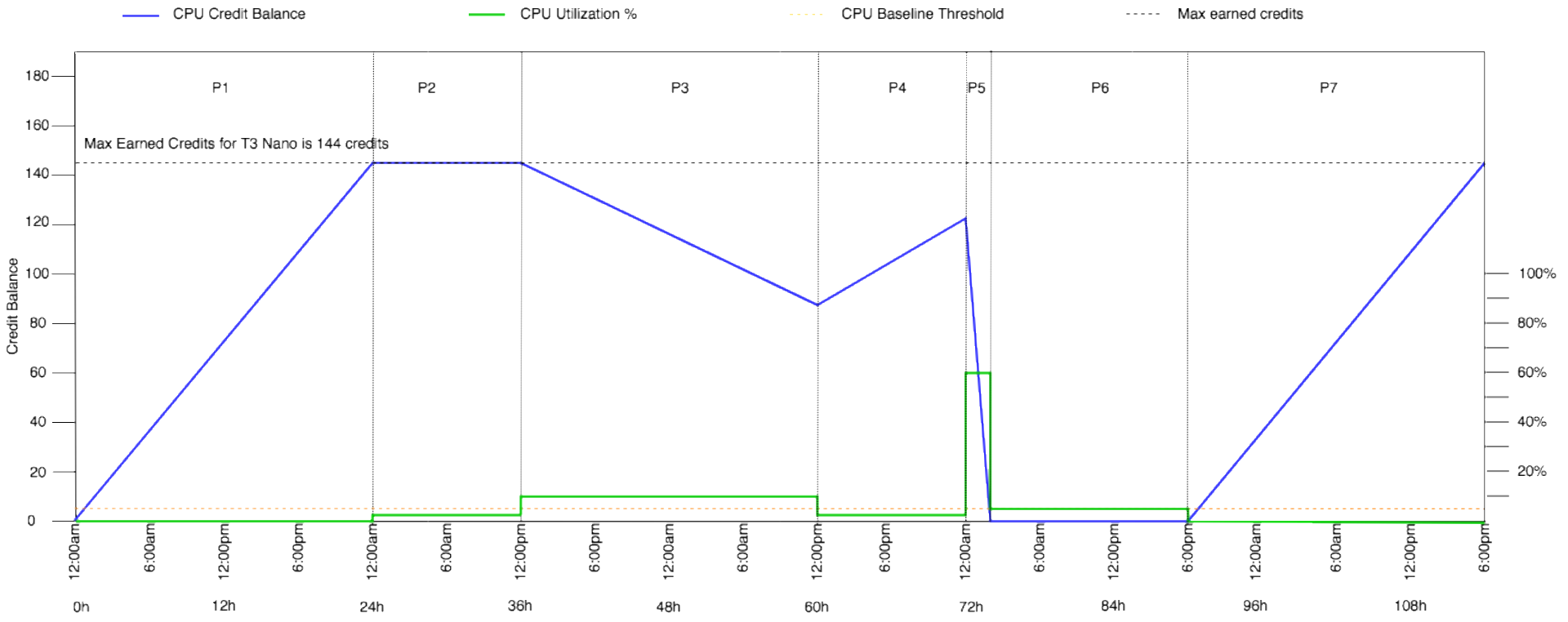
T3 Standard Mode



T3 Standard Mode



T3 Standard Mode



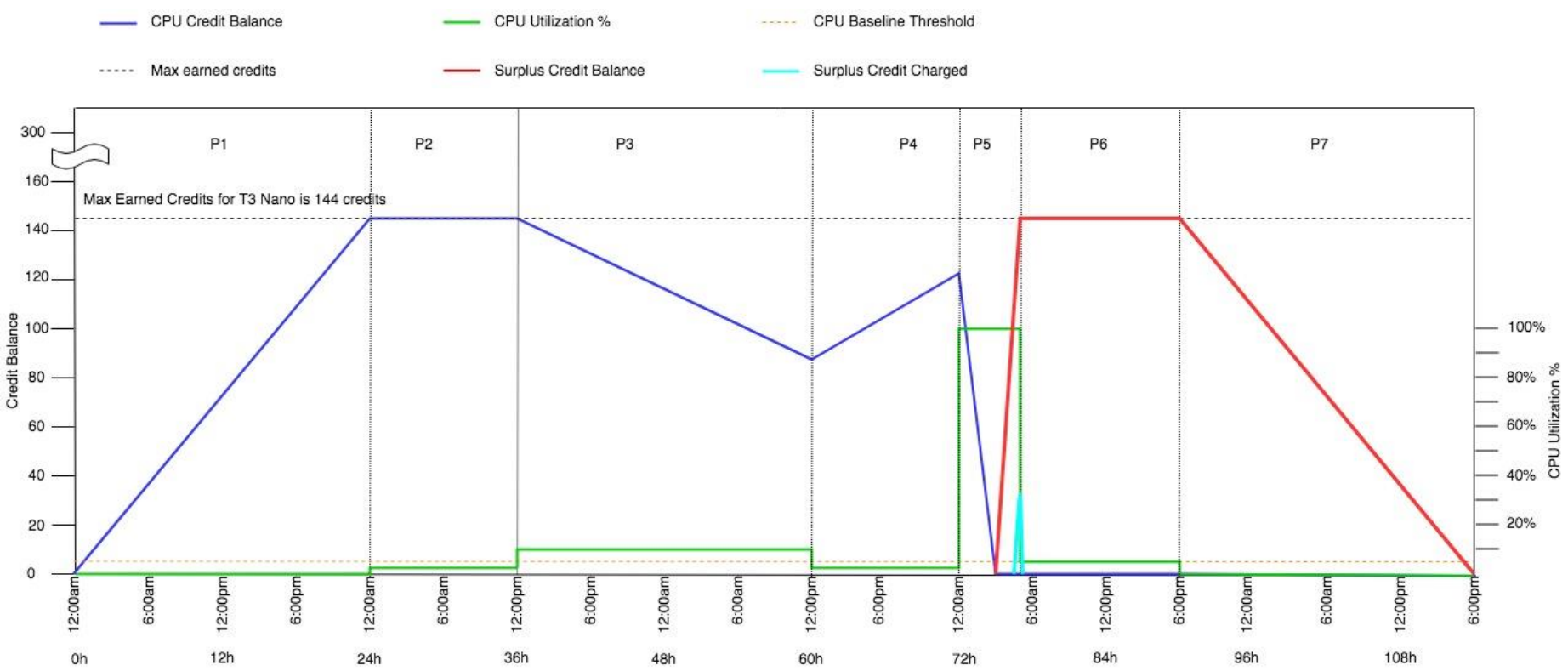
T3 Unlimited Mode

If the instance runs out of CPU Credit Balance,

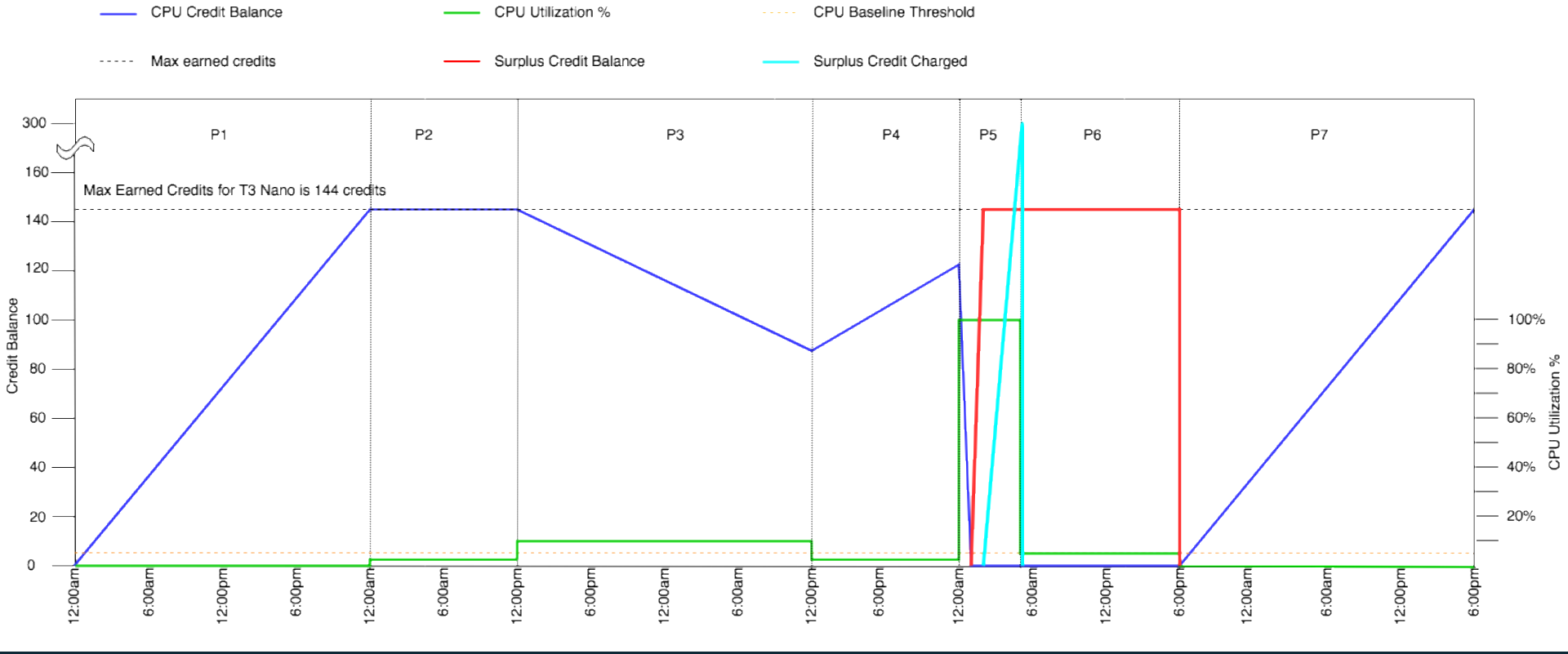
- the instance can continue to burst by consuming ***surplus*** credits
- the instance can consume a maximum number of surplus credits based on instance size (e.g. 144 for micro), without being charged
- surplus credits consumed above the maximum are charged at a rate of $0.05 per vCPU hour
- **Why?** Allows instances to burst at any time and pay down the consumed credits within 24 hours without being charged

aws

T3 Unlimited Mode Use Case



Switching from Unlimited to Standard



How AWS Customers are using T3 instances?

Other customers using T3 instances?

aws

Where is T3 available?

Regions and AZ

How to switch to from T2 instances to T3 instances?

How to switch from T2 instances to T3?

- ✓ T3 instances have the following requirements:
- ✓ Must have the NVMe drivers installed.
    - ✓ EBS volumes are exposed as NVMe block devices.
- ✓ Must have the Elastic Network Adapter (ENA) drivers installed.
- ✓ No Classiclink support

aws

How to switch from T2 instances to T3?

- ✓ The following AMIs meet these requirements:
  - ✓ Amazon Linux 2
  - ✓ Amazon Linux AMI 2014.03 or later
  - ✓ Ubuntu 14.04 or later
  - ✓ SUSE Linux Enterprise Server 12 or later
  - ✓ Red Hat Enterprise Linux 7.4 or later
  - ✓ CentOS 7 or later
  - ✓ FreeBSD 11.1-RELEASE
  - ✓ Windows Server 2008 R2 or later

aws

How to switch from M instances to T3 instances?

Make sure the right drivers are supported

- You can do it from Amazon EC2 Console
- Stop the instance
- With the instance still selected, choose Actions to Change Instance Type T3
- Select EBS-optimized
- Good to Go!

aws

How are T3 instances priced?

T3 Pricing

- ✓ 10% cheaper than T2 on Linux across OD and RI price
- ✓ 13% cheaper than M5
- ✓ Same burst performance as M5
- ✓ Available as OD 1YR Std RI, 1YR Conv RI, 3 YR Std RI, 3YR Conv RI, and Spot pricing

aws

Pricing – CPU credits

CPU usage is tracked at second granularity in vCPU-Hours

Billed at the end of the month at:

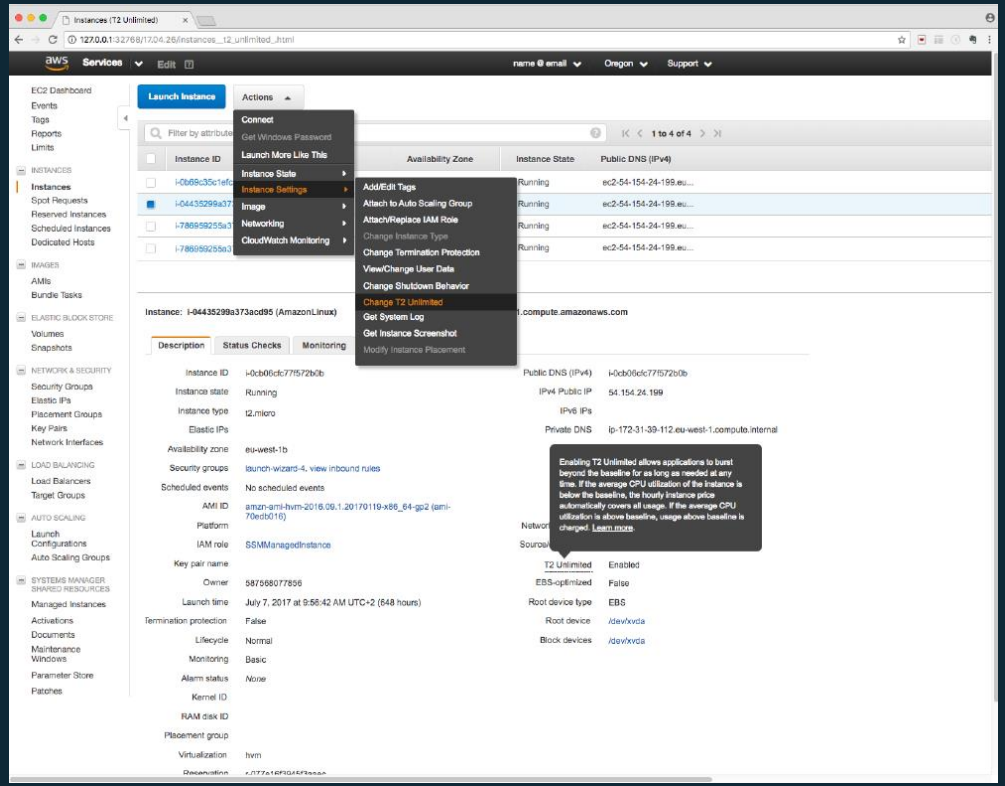
- 5 cents per vCPU-Hour for Linux/RHEL/SUSE
- 9.6 cents per vCPU-Hour for Windows

Same for T2 and T3 across all pricing models

aws

How to switch between T3 modes?

Via Console



How to switch between T3 modes?

Via CLI

Run

```
$ aws ec2 run-instances --image-id [ami-id] --count [num] --instance-type
t3.micro --credit-specification 'CpuCredits=standard'

$ aws ec2 run-instances --image-id [ami-id] --count [num] --instance-type
t3.micro --credit-specification 'CpuCredits=unlimited'
```

Modify

```
$ aws ec2 modify-instance-credit-specification --region us-east-1 --instance-
credit-specification ‘[{"InstanceId": "i-1234567890abcdef0","CpuCredits":
"unlimited"}]’

$ aws ec2 modify-instance-credit-specification --region us-east-1 --instance-
credit-specification ‘[{"InstanceId": "i-1234567890abcdef0","CpuCredits":
"standard"}]’
```

Describe

```
$ aws ec2 describe-instance-credit-specifications --instance-ids i-
1234567890abcdef0
```


aws

T3 CloudWatch Metrics

CPUUtilization	The percentage of allocated EC2 compute units that are currently in use on the instance. This metric identifies the processing power required to run an application upon a selected instance.
CPUCreditUsage	vCPU minutes or CPU credits used in the prior 5 minutes
CPUCreditBalance	Sum of earned CPU credits accumulated since instance launch net of credits consumed up to a maximum level based on instance size
CPUSurplusCreditBalance	Balance of surplus CPU credits consumed while the CPUCreditBalance is zero
CPUSurplusCreditsCharged	Surplus CPU credits are charged after the CPUSurplusCreditBalance hits the maximum credit balance that an instance can earn in a 24 hour period

FAQ

Is switching between Standard and Unlimited supported?
Yes. Earned credits are carried over. Surplus credits are immediately charged.

Does Auto Scaling Group support T3 Unlimited?
Yes, this is supported via the new Launch Templates and one click Launch Configuration to Launch Template conversions.

Does CloudFormation support T3 Unlimited?
Yes, this is supported via the new Launch Templates, and native integration is coming soon.

Do Elastic Beanstalk and OpsWorks support T3 Unlimited?
Not yet. This work is being planned.

aws

Thank You!