

Introduction to Amazon Web Services (AWS)

Marco Landoni

INAF – OA Brera

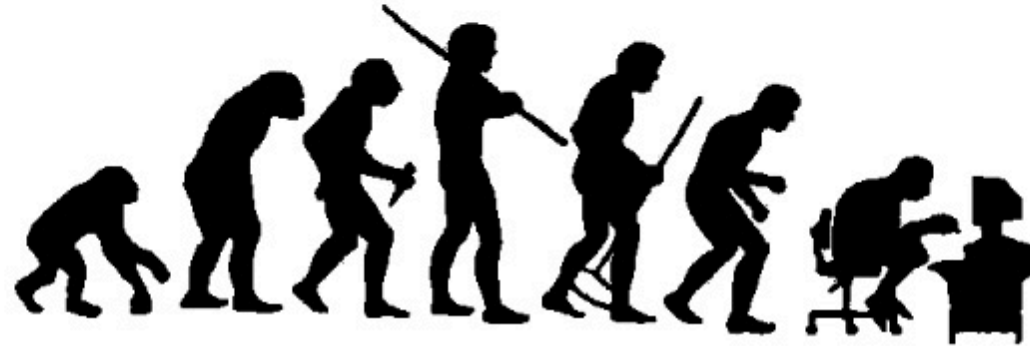
marco.landoni@inaf.it

Outline

- Cloud computing concepts in general and a warning
- A detour of the AWS platform
- Using AWS for batch HTC purposes
- Costs, comparison and Wrap up

Cloud computing paradigm

- **Cloud Computing** is a style of **computing paradigm** in which typically real-time scalable resources can be accessible via Internet to users. Pay as-you-go for resource utilisation. (Wikipedia)



Mainframe
Computing



Personal
Computing



Client/Server
Computing



Mobile
Computing



Cloud
Computing

Warning!

- One size does not fit all.
- Each problem has its **proper computational model** and architecture that maximise the return, the **cost efficiency** and the proper **exploitation of common resources**.
- A rule of thumb:
 - Long N-Body cosmological simulation, high energy gamma-ray showers track: **HPC**
 - Deep simulation (unfrequently) for instrumentation that needs answer **now** with **unpredictable schedule** or moderate CPU demanding scientific simulation: **Cloud Computing**

Commercial Cloud – Amazon Web Services

- Set of services offered by AWS for a relative cheap price ranging from computational power to storage and off-the-shelf data analytics.
- Pay-as-you-go paradigm
- Note: this is not the unique provider but, for scientific simulations, demonstrated to be most suitable platforms.

Various providers .



Google Cloud Platform



A quick view of the console

AWS global infrastructure

Over 1 million active
customers across 190 countries

900+ government agencies

3,400+ educational institutions

11,200+ nonprofits

11 regions

28 Availability Zones

53 edge locations



A non exhaustive list of service that could be useful as scientist

- Compute resources (EC2 + EBS)
- Storage services.
- Batch scheduler (AWS Batch) with elastic cluster
- Databases



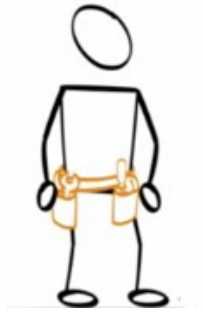
Compute Resources – Amazon EC2

- EC2 (Elastic Cloud Computing) forms a central part of the platform by allowing users to rent **virtual computers** on which they run their own applications.
- There is a vast number of different machines that can be instantiated. Some of them have also GPU support

Instance	GPUs	vCPU	Mem (GiB)	GPU Mem (GiB)	GPU P2P	Storage (GB)	Dedicated EBS Bandwidth	Networking Performance
p3.2xlarge	1	8	61	16	-	EBS-Only	1.5 Gbps	Up to 10 Gigabit
p3.8xlarge	4	32	244	64	NVLink	EBS-Only	7 Gbps	10 Gigabit
p3.16xlarge	8	64	488	128	NVLink	EBS-Only	14 Gbps	25 Gigabit
p3dn.24xlarge	8	96	768	256	NVLink	2 x 900 NVMe SSD	14 Gbps	100 Gigabit

Instance	vCPU	Mem (GiB)	Storage	Network Performance (Gbps)
a1.medium	1	2	EBS-Only	Up to 10
a1.large	2	4	EBS-Only	Up to 10
a1.xlarge	4	8	EBS-Only	Up to 10
a1.2xlarge	8	16	EBS-Only	Up to 10
a1.4xlarge	16	32	EBS-Only	Up to 10
a1.metal	16*	32	EBS-Only	Up to 10

EC2 instances: Families and Generations



General-purpose:

- M1, M3 , T2

Compute-optimized:

- C1, CC2, C3, C4 M2,

Memory-optimized:

Dense-storage:

CR1, R3 HS1, D2

I/O-optimized:

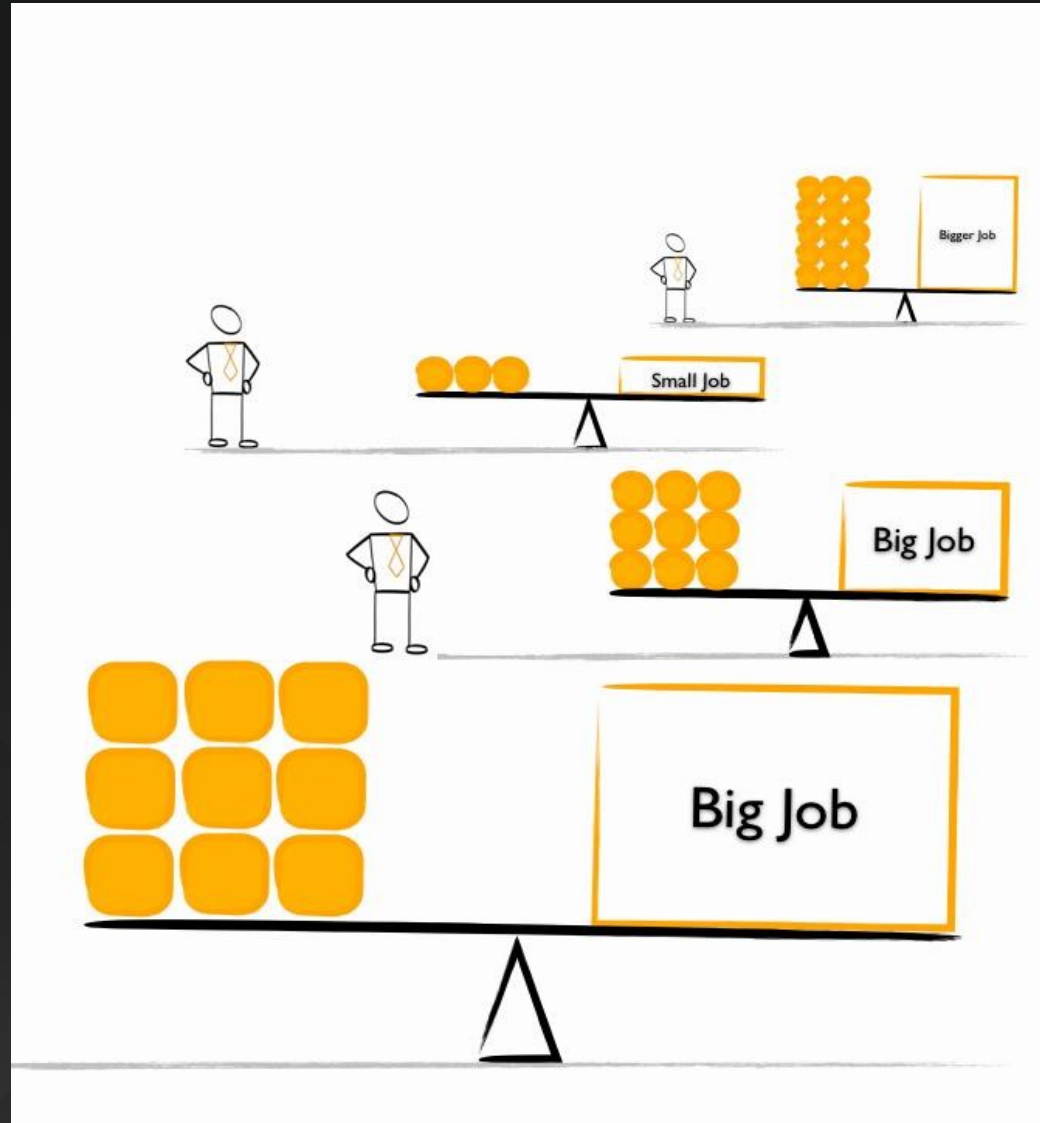
- HI1, I2 CG1, G2

GPU:

Micro:

T1, T2

EC2 instances: Sizes



EC2 instances: Types

Instance generation

c4.large

Instance family

Instance size

Purchase options that fit your workloads

On-Demand

Pay for compute capacity by the hour with no long-term commitments

For spiky workloads, or to define needs



Reserved

Make an EC2 usage commitment and receive a significant discount.

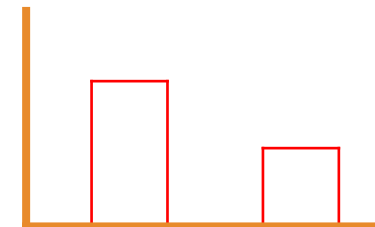
For committed utilization



Spot

Bid for unused capacity, charged at a Spot Price which fluctuates based on supply and demand

For time-insensitive or transient workloads



Storage for EC2 – The EBS service

- Usually, for each EC2 instances is necessary to attach a disk where saving local data, the operating system, ecc.
- This is achived by using the [Elastic Block Storage service \(EBS\)](#) which basically delivers a block storage to the instances in various flavours.



	Solid State Drives (SSD)		Hard Disk Drives (HDD)	
Volume Type	EBS Provisioned IOPS SSD (io1)	EBS General Purpose SSD (gp2)*	Throughput Optimized HDD (st1)	Cold HDD (sc1)
Short Description	Highest performance SSD volume designed for latency-sensitive transactional workloads	General Purpose SSD volume that balances price performance for a wide variety of transactional workloads	Low cost HDD volume designed for frequently accessed, throughput intensive workloads	Lowest cost HDD volume designed for less frequently accessed workloads
Use Cases	I/O-intensive NoSQL and relational databases	Boot volumes, low-latency interactive apps, dev & test	Big data, data warehouses, log processing	Colder data requiring fewer scans per day
API Name	io1	gp2	st1	sc1
Volume Size	4 GB - 16 TB	1 GB - 16 TB	500 GB - 16 TB	500 GB - 16 TB
Max IOPS**/Volume	64,000	16,000	500	250
Max Throughput***/Volume	1,000 MB/s	250 MB/s	500 MB/s	250 MB/s
Max IOPS/Instance	80,000	80,000	80,000	80,000
Max Throughput/Instance	1,750 MB/s	1,750 MB/s	1,750 MB/s	1,750 MB/s
Price	\$0.125/GB-month \$0.065/provisioned IOPS	\$0.10/GB-month	\$0.045/GB-month	\$0.025/GB-month

Amazon AMI



Amazon Machine
Image (AMI)

- Pre-built image (or user built images) that can be used to instantiate EC2 machines.
- This is really useful to save the state of the machine permanently and to replicate the same machine (and SW configuration) on many different nodes in just 1 click



Live Demo !



Amazon S3 – Simple Storage Service

- This is one of the main service to permanently store data (in a object-oriented File Systems)
- Data and files are organized in the concept of bucket/keys.
- A bucket could be thought as a directory
- A key is a file.



Live Demo !

Docker and Amazon – Elastic Container Registry



- An application containerised can be deployed to AWS in its proper registry to be used with Amazon Batch (an HTC Scheduler).



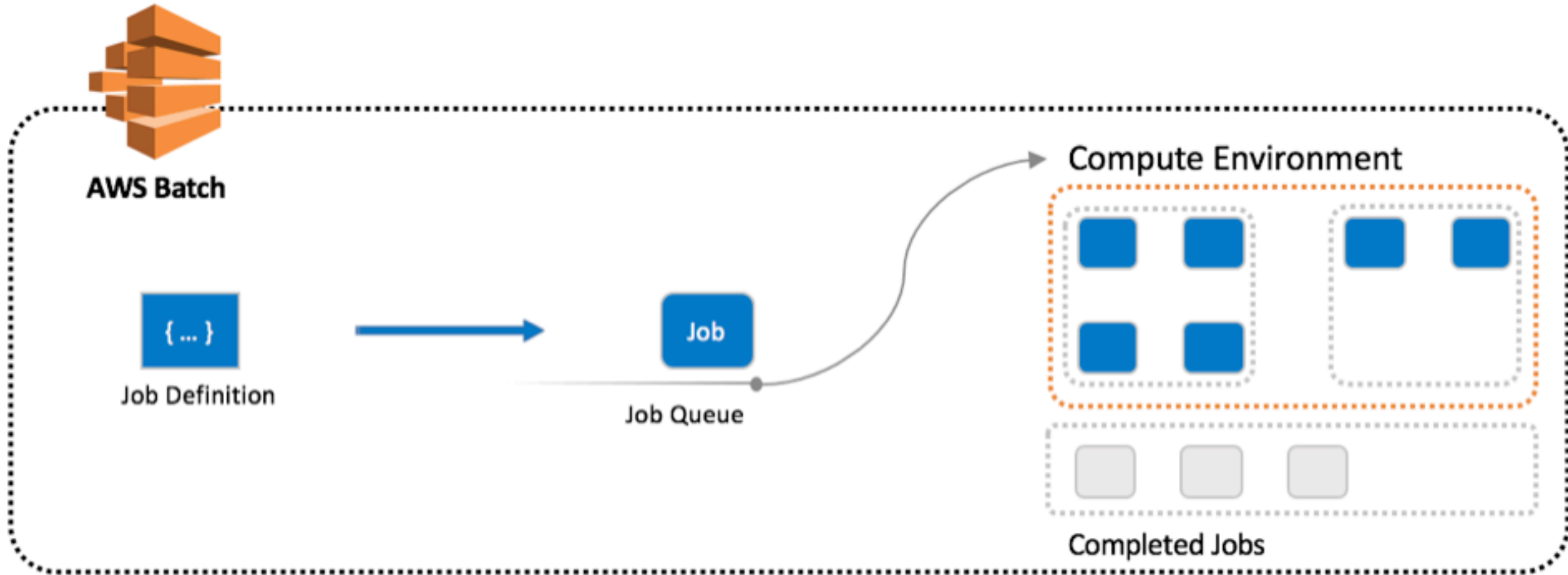
Live Demo !

AWS Batch

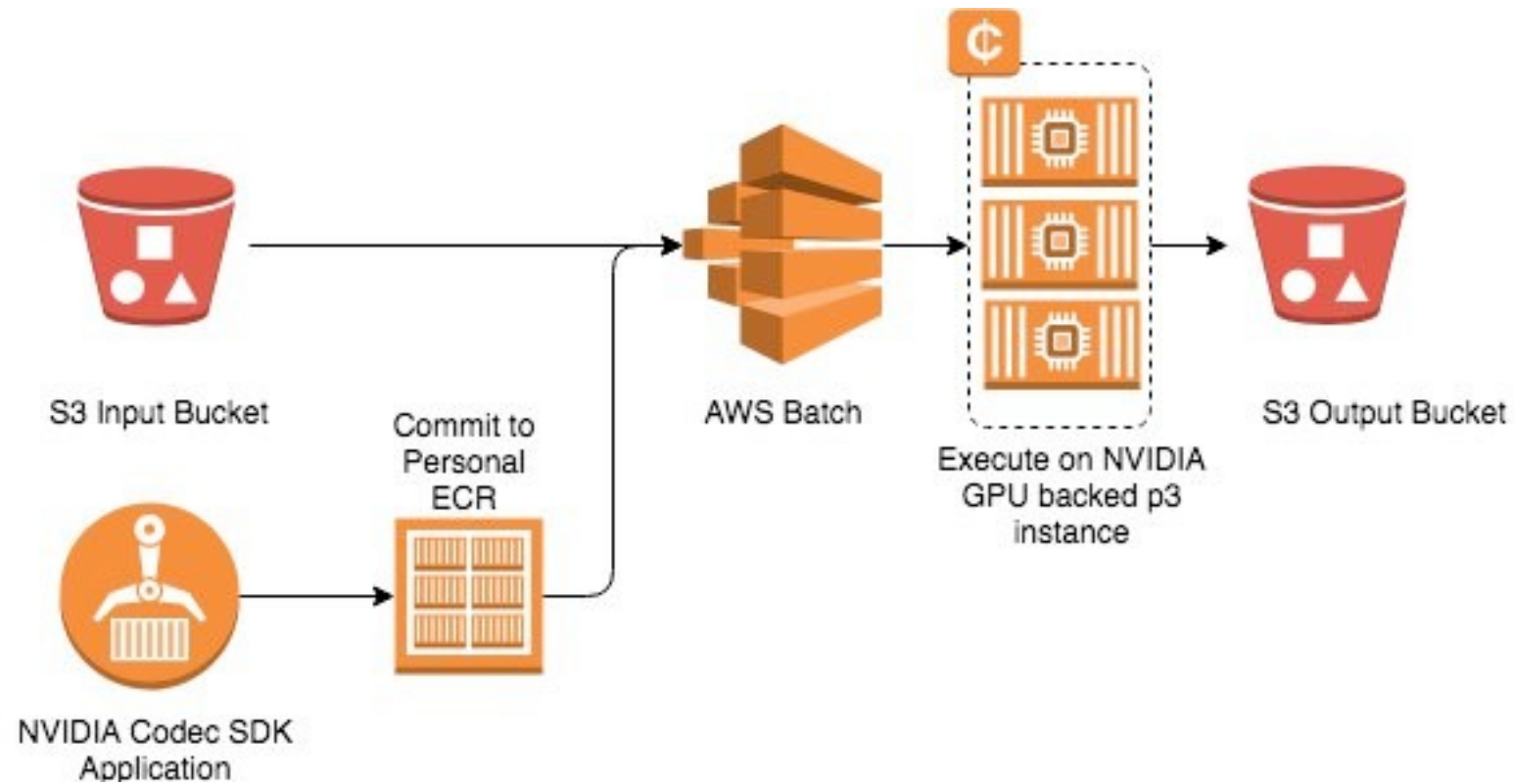
- AWS Batch enables developers, scientists, and engineers to easily and efficiently run hundreds of thousands of batch computing jobs on AWS.
- With AWS Batch, there is no need to install and manage batch computing software or server clusters that you use to run your jobs
- No additional charges



How AWS Batch works ?



A typical scenario (with NVIDIA GPUs)





Live Demo !