Data Schools

Computational Infrastructures an overview

Alessandro Costantini
INFN, Italy

INFN
Istituto Nazionale di Fisica Nucleare

Training Goals

Learn the concept of

1. Local vs Distributed
2. Examples of
    - High Performance Computing (HPC)
    - High Throughput Computing (HTC)
    - Cloud

Local vs Distributed Computing

The science computing is becoming complex

- Monte Carlo, image analysis, genetic algorithm, simulation...

It will take a year (CPU time) to get the results on your laptop, but your paper is due in a week.

What do you do?

You have the following options:

- Wait until the calculation is done
- Use a distributed computing environment to speedup

Distributed computing environment

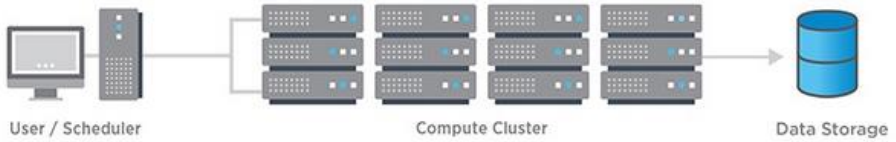
Distributed computing is a field of computer science that studies distributed systems. A *distributed system* is a system whose components are located on different networked computers, which communicate and coordinate their actions by passing messages to one another.

- High Performance Computing (HPC)
- High Throughput Computing (HTC)
- Cloud

High Performance Computing

High Performance Computing (HPC) is the ability to process data and perform complex calculations at high speeds

- Fast access to network/disk/memory
  - Limited resources
- Needs
  - A highly parallel program or
  - Port your code to a special environment
  - Request and wait for an allocation



High Throughput Computing

- Instead of the fastest computer, lots of individual “**commodity**” computers
  - May not be fastest network/disk/memory, but you can access a lot of them
  - Job can be broken down into separate, independent pieces
    - If I give you more computers, you run more jobs
    - You care more about total quantity of results than instantaneous speed of computation

Grid Computing

- Provisioning and use of *massively distributed and federated resources* routinely happens since several years.
- An example of great success in the scientific world is given by **Grid Computing**.
- On the right:
  - Real-time situation of computation and data transfer for the physics experiments running at the CERN Large Hadron Collider (LHC).

Cloud computing

Source: bit.ly/2Z4kHNG

In a nutshell, Cloud Computing deals with:

1. Supplying
2. information and communication technologies
3. as a service

US National Institute of Standards and Technology (NIST) bit.ly/2YOop2X

Characteristics

- **Self-service, on-demand**
  - A consumer can unilaterally provision computing capabilities as needed automatically without requiring human interaction with each service provider.
- **Network-based access**
  - Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms
- **Resource pooling**
  - The customer has no control or knowledge over the details of the provided resources, that are managed by the Cloud provider
- **Elasticity**
  - Capabilities can be elastically provisioned and released to scale rapidly commensurate with demand.
    To the consumer, the capabilities available for provisioning often appear to be unlimited
- **Pay-per-use**
  - The customer pay only for what he/she used.

Information and Material

- https://opensciencegrid.org/dosar/DataTrieste2019/Materials/
- https://github.com/CODATA-RDA-DataScienceSchools/Materials/blob/master/docs/DataTrieste2021/00-Hands_on_Exercise_Overview.md
- Alessandro Costantini – alessandro.costantini@cnaf.infn.it

Data Schools

Contact

Alessandro Costantini – alessandro.costantini@cnaf.infn.it

Contributors

Alessandro Costantini - INFN
Cristina Doina Duma - INFN