Harnessing the complexity of large amounts of data is a challenge in itself. But Big Data is more than that: originally characterized by the 3 Vs of Volume, Velocity and Variety, it often requires dedicated computing solutions, which will be explored in this module.

Prerequisites

- Engineering Degree on Computer Science or a related domain (telecommunications, etc).
- OR
- Engineering Degree on another subject with a major on Computer Science.
- OR
- Work experience on Computer science.

Learning objectives

After completing this course, participants will be able to:

- Implement the distribution of simple operations via the Map/Reduce principle in Spark;
- Explain the difference between CPU and GPGPU computation;
- Connect on a cloud computing engine (e.g. Google Cloud Platform) and launch a simple task;
- Understand the usefulness of containers
- Deploy a Docker container.

Practical information and registration

Natalia Perthuis - 05 61 33 80 47 – info.exed@isae-supraero.fr
Course Content

Distributed computing with Spark:
- History
- MapReduce paradigm
- Hadoop Stack
- Hadoop Distributed File System
- MLib Machine Learning library

Virtualization and cloud computing:
- Different approaches to virtualization
- Economical models
- Technical benefits (snapshots, dynamic deployment and migration, failover...)
- cloud engines (principles, deployment examples, node choices)

Docker:
- History,
- Fundamental differences w.r.t. virtualization
- Docker components
- Tools