

# Spark : Introduction



---

## Apache Spark

Spark est un framework qui a été développé en réponse aux problèmes de Big Data :

- Manipulation massive de données
  - Analyse de données en temps réel
  - Adaptabilité et la scalabilité
- 

## Apache Spark

Apache Spark est un framework Open source de traitement de données.

Développé par l'université de Berkeley en 2009, il est développé pour faire face aux problèmes des solutions existantes (hadoop, MapReduce, etc..).

---

## Apache Spark

C'est un projet qui est repris par la fondation Apache en 2013.

Spark a pour but de fournir une solution de traitement de données beaucoup plus rapide et plus facile à utiliser que celles existantes.

---

## Apache Spark

Spark est un système distribué dont le rôle est de traiter de très gros volumes de données.

Son travail s'effectue sur plusieurs machines (cluster) ce qui le rend *scalable*\* et **flexible**.

---

## Apache Spark

C'est un système qui se veut simple d'utilisation.

Il est basé sur des API simples et intuitives :

- python
- scala
- java

- R
- 

## Apache Spark

C'est un système distribué de traitement de données rapide et flexible utilisé pour résoudre les problèmes de Big Data :

- analyse de données en temps réel
  - analyse des objets connectés
  - analyse bio-informatique et génomique
  - analyse astronomique
  - le machine learning
  - ...
- 

### Caractéristiques Spark