

## Spark : Caracteristiques



---

## Spark : Caracteristiques

Spark est un systeme de traitement de données qui se veut :

- distribué et scalable
  - rapide et performant
  - facile d'utilisation
- 

### Distribué et scalable

Spark est un systeme **distribué** :

Traitement des données sur plusieurs machines.

L'ensemble de ces machines est appelé un **cluster**.

Au sein d'un cluster une machine est appelée un **noeud**.

---

### Distribué et scalable

Il est également facilement **scalable** et **flexible**:

- Ajouter des noeuds pour augmenter la capacité de traitement
  - Supprimer des noeuds inutilisés pour réduire les coûts
- 

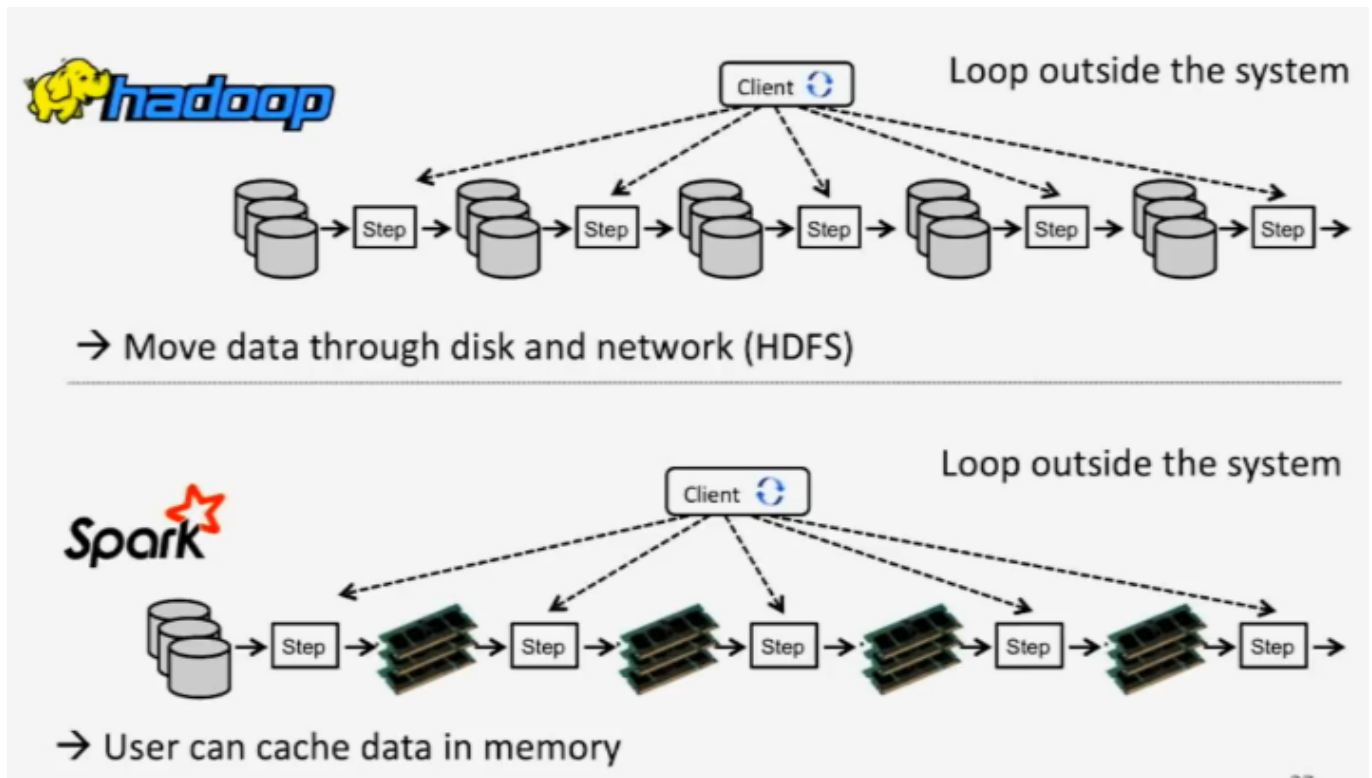
### Rapide et performant

Traitement des données effectué par Spark est **in-memory**.

Les données sont donc traitées très rapidement sans avoir besoin des les ecrires sur un support entre chaque étape.

---

### Rapide et performant




---

Rapide et performant

De plus Spark utilise un moteur d'exécution haute performance appelé le DAG (Directed Acyclic Graph) pour accélérer le traitement des données parallélisées.

---

Facile d'utilisation

Spark se fait facile d'utilisation en proposant des API simples et intuitives.

Il est supporté dans plusieurs langages de programmation dont Python, Scala, Java, R, SQL, etc.

---

Facile d'utilisation

Il est également possible d'utiliser Spark avec des notebooks comme Jupyter ou dans un terminale interactif via l'outil **spark-shell**.

---

## Spark : Caracteristiques

En Résumé : Spark est un système de traitement de données distribué, rapide et performant, scalable et facile d'utilisation.

Il présente toutefois quelques limites.

---

Limites de Spark

- **Configuration** complexe des clusters qui demande une connaissance en administration système avancé.

- **Consomation de memoire de travail importante** : stockage et traitements en mémoire vive par spark peut etre assez gourmand en memoire.
  - **Latence** : le transfert de données d'un noeuds à l'autre peut etre long et ralentir les traitements iteratifs.
- 

## Architecture