

Le premier chapitre est divisé en 3 parties :

- Évolution des données(Collecte, Stockage et traitement)
- Spark et ses différents composants
  - ◆ Storage
  - ◆ Resource Management
    - Cluster manager : vérifie la disponibilité des Workers pour les assigner des tâches venant Spark driver(main controller of a Spark application)
    - Worker: exécute les tâches assignées par Cluster manager
  - ◆ Engine and Ecosystem
    - Spark SQL
    - Mllib
    - Structured Streaming
    - Graph X
  - ◆
- Utilisation de Spark dans un environnement Cloud