

INTRODUCTION



Vous souhaitez savoir comment fonctionne Hadoop et son écosystème, en particulier Spark dans un projet Big Data ? Vous êtes au bon endroit ! D'après le constat des experts, des institutions publiques et privés, 90 % des données récoltées depuis le début de l'humanité ont été générées durant les 2 dernières années. Le marché qualifie aujourd'hui de « Big Data » cette explosion de données. Il devient donc important pour tout un chacun de comprendre les principes de base des outils de gestion et de traitement de ces données massives tels que Hadoop et Spark. Ce document se déclinera suivant le plan suivant:

I- Hadoop et le big data

II- Composant de base d'un cluster hadoop

- a. Hadoop Distributed File System
- b. MapReduce

III- Hadoop et son écosystème

IV- Hadoop vs Spark

V- Comment installer Hadoop avec Docker?

VI- Packages

- a. RHadoop
- b. SparkR