Hadoop Distributed File System (HDFS)

Data science
Master 2 ISIDIS

Sébastien Verel
verel@lisic.univ-littoral.fr
http://www-lisic.univ-littoral.fr/~verel

Université du Littoral Côte d'Opale
Laboratoire LISIC
Equipe OSMOSE

Système de fichier - File Sytem

Rappels à travers une question :

Qu'est-ce qu'un système de fichier (File System) ?

Système de fichier - File Sytem

Rappels à travers une question :

Qu'est-ce qu'un système de fichier (File System) ?

- Quelles sont les caractéristiques d'un FS ?
- Quelles sont les commandes relatives à un FS ?
- Quelles sont les propriétés attendues d'un FS ?

Nouvelles problématiques, nouvelles solutions

Nouveaux big besoins

On désire :

- Avoir accès aux données depuis de nombreux clients à travers le réseau
- Sauvegarder une très grande quantité de données
- Garantir les services classiques d'un file system

Question :

Quelles solutions peut-on imaginer pour faire cela ?

Hadoop Distributed File System (HDFS)

Hadoop Distributed File System (HDFS) est :

- un **système de fichier** (file system)
- un système de fichier **distribué**

Buts de ce File System

- Stocker une grande quantité de donnée (fichiers)
- Donner accès aux données à de nombreux clients distribués sur le réseau
- Garantir l'accès aux fichiers

Remarque :

Projet de la fondation Apache, Open Source.

Support de cours

- Bilel Derbel,
  Université de Lille 1
  Programmation Parallèle et Distribué,
  `https://sites.google.com/site/bilelderbelpro/home/teaching`
- Benjamin Renaut,
  MBDS université de Nice Sophia Antipolis
  Cours big data,
  `http://cours.tokidev.fr/bigdata/`