

# Rapport sur les Travaux Pratiques (TP)

## Introduction

Les deux travaux pratiques — **TD N°1 : Bases de données NoSQL « MongoDB »** et **TP-BigData-01 : Préparation de la plateforme de développement** — ont pour objectif d’initier les étudiants à deux technologies fondamentales :

- **MongoDB**, une base de données NoSQL orientée documents, adaptée au stockage flexible et évolutif de données semi-structurées.
- **Hadoop**, un écosystème destiné au traitement des données massives (Big Data) grâce à des architectures distribuées et des outils comme MapReduce.

Ces TP offrent une première approche pratique de la gestion et du traitement de données modernes.

---

## I. TD N°1 : Bases de données NoSQL - MongoDB

### Objectif

Ce TD a pour but de familiariser les étudiants avec l’installation, la configuration et l’utilisation de MongoDB sous Windows. Il permet de découvrir le fonctionnement des bases NoSQL et de s’exercer aux opérations fondamentales (CRUD).

### Étapes réalisées

#### 1. Installation et configuration sous Windows

- Téléchargement de MongoDB via Google Drive.
- Extraction des fichiers dans le répertoire C:\MongoDB.
- Création du répertoire C:\data\db pour le stockage des données.

#### 2. Démarrage du serveur MongoDB

- Lancement du serveur avec la commande mongod.exe.

#### 3. Connexion au client MongoDB

- Utilisation de mongo.exe pour interagir avec le serveur MongoDB.

#### 4. Manipulations CRUD

- Création d’une base de données appelée info.

- Création d'une collection nommée produits.
- Insertion de trois documents :
  - Macbook Pro
  - DELL
  - Thinkpad X230
- Exécution de diverses requêtes :
  - Récupération de tous les produits.
  - Filtrage par prix, champ ultrabook, ou recherche textuelle sur le nom.
  - Suppression de documents selon le fabricant ou l'ID.

### **Compétences acquises**

- Installation et configuration de MongoDB sous Windows.
- Maîtrise des commandes de base : insertion, requêtes, suppression.
- Manipulation de collections et de documents au format JSON-like.
- Compréhension des opérations CRUD dans un contexte NoSQL.
- Lecture et interprétation des résultats dans une console interactive.

---

## **II. TP-BigData-01 : Préparation de la plateforme de développement**

### **Objectif**

Ce TP a pour objectif de mettre en place un environnement de développement Big Data basé sur Hadoop, et d'exécuter un programme MapReduce pour le traitement de données massives.

### **Étapes réalisées**

#### **1. Configuration de l'environnement**

- Installation de VirtualBox.
- Importation de la machine virtuelle Debian :  
mbds\_big\_data\_hadoop\_2019\_2020\_vm.ova.
- Connexion à la machine virtuelle via SSH avec PuTTY (port 2222).
- Transfert de fichiers entre Windows et la VM grâce à WinSCP.

## 2. Démarrage et vérification de Hadoop

- Lancement de Hadoop avec la commande start-hadoop.
- Vérification de l'état du cluster avec la commande hdfs dfsadmin -report.

## 3. Exécution d'un exemple MapReduce

- Compilation d'un projet Java en archive .jar (tp-1.0.0.jar).
- Copie du fichier poeme.txt dans le système de fichiers distribué HDFS.
- Lancement du programme MapReduce (compteur d'occurrences de mots).
- Lecture et analyse des résultats dans HDFS.

### Compétences acquises

- Installation et configuration d'une machine virtuelle Hadoop.
- Gestion des connexions réseau (SSH, SCP) avec une VM Linux.
- Utilisation du système de fichiers distribué HDFS.
- Compilation, déploiement et exécution d'applications MapReduce.
- Analyse des résultats issus d'un traitement Big Data.

---

### Conclusion générale

Ces deux travaux pratiques couvrent des domaines complémentaires et essentiels de la gestion moderne des données :

- **MongoDB** propose une solution NoSQL flexible et performante, idéale pour les données semi-structurées et les applications web nécessitant évolutivité et rapidité.
- **Hadoop**, grâce à son architecture distribuée, offre une plateforme robuste pour le traitement de grandes masses de données, essentielle dans le domaine du Big Data.

Ces TP constituent une base solide pour appréhender des technologies incontournables, et préparent à des projets plus avancés en **intelligence artificielle**, **analyse de données**, ou encore **cloud computing**.