

# Manul

Renforcer l'Apprentissage de Hadoop avec des Environnements  
Pratiques

Laurent Siksous

29 avril 2024



# Outline

- 1 Introduction
- 2 Contexte
- 3 Problématique
- 4 Solution
- 5 Démonstration

# Introduction



*"Embrassez la curiosité, explorez sans relâche, et laissez vos instincts vous guider comme le Manul agile à travers les paysages accidentés de la connaissance."*

# Aperçu du Projet

- L'objectif de notre projet est de faciliter la création d'environnements d'apprentissage pour l'enseignement des cours sur le Big Data.
- Ces cours sont conçus pour fournir aux étudiants une expérience pratique de travail avec les distributions Hadoop, y compris OSS, Cloudera et MapR.
- Notre boîte à outils offre un déploiement facile de ces distributions Hadoop en utilisant diverses méthodes, permettant aux étudiants d'explorer différentes techniques de déploiement et de gestion de la configuration.
- Cognitive Science : étapes du développement

# Présentation de Datascientest

Datascientest, une entreprise spécialisée dans l'enseignement des technologies Big Data.

## Quelques chiffres.

**2017**

Année de  
lancement

**9000**

Nombre  
d' alumni

**129**

Nombre  
d'intervenants

**18**

Nombre de  
formations

# Curriculum Hadoop

Titre Formation	NB. Jours
Méthodes Agiles de gestion de projet, comprendre la démarche	2
Maîtriser Unix	3
Scriptshell	3
Git, mettre en œuvre le contrôle de versions	1
Python, programmation Objet	3
Ansible, automatiser la gestion des serveurs	3
Terraform	2
Hadoop : l'écosystème	1
Apache Hadoop, architecture , développer des applications pour le Big Data	4
Installation et administration d'un cluster Cloudera	8
Installation et administration d'un cluster sur MapR	8
Apache Kafka, prise en main et monitoring	3
Architecture de la stack Hadoop	3
Les aspects non fonctionnels et d'administration	3
Elasticsearch prise en main	2
Troubleshooting	3
Projet + Soutenance	5

# Problématique I

- **Disparité des Environnements** : Chaque cours du curriculum utilise des environnements Hadoop différents, ce qui entraîne une disparité et une fragmentation des ressources et des outils pédagogiques.
- **Perte de Temps** : La configuration et la gestion manuelles des environnements Hadoop pour chaque cours entraînent une perte de temps significative pour les formateurs et les étudiants, limitant ainsi leur capacité à se concentrer sur l'apprentissage des concepts essentiels.
- **Croyance dans l'Obsolescence de la Distribution MapR** : Malgré le support et l'efficacité de la documentation fournie par HPE Ezmeral Data Fabric, Datascientest perçoit la distribution MapR comme étant en voie d'obsolescence, ce qui soulève des préoccupations quant à la pertinence de son inclusion dans le curriculum.

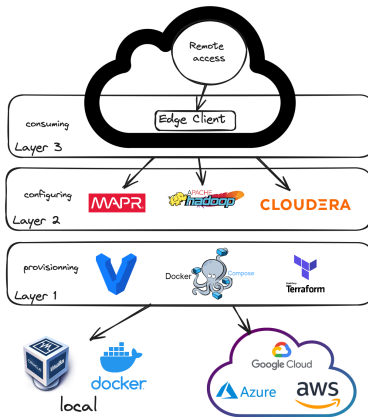
## Problématique II

- **Raréfaction des Compétences MapR** : Avec la raréfaction des compétences MapR sur le marché, il est devenu difficile de trouver des formateurs qualifiés pour enseigner aux étudiants.
- **Demande Persistante des Clients** : Toutefois, la demande persistante des clients de Datascientest pour des cours sur MapR nécessite une solution qui puisse répondre à ces besoins tout en tenant compte des préoccupations concernant l'obsolescence perçue de la distribution.

Datascientest a besoin d'une solution qui simplifie et standardise le processus de déploiement des environnements Hadoop, tout en offrant une expérience pratique et uniforme à ses étudiants.



# Architecture



# Fonctionnalités de la Boîte à Outils I

- **Distributions Hadoop** : Notre boîte à outils prend en charge le déploiement de trois principales distributions Hadoop : OSS, Cloudera et MapR. Cette variété permet aux étudiants de se familiariser avec différentes plateformes Hadoop et de comprendre leurs fonctionnalités et capacités uniques.
- **Méthodes de Déploiement** : Nous proposons plusieurs méthodes de déploiement pour répondre aux préférences et aux scénarios d'apprentissage différents :
  - **Docker** : Les étudiants peuvent déployer des environnements Hadoop à l'aide de conteneurs Docker, offrant une solution légère et portable.
  - **Vagrant** : Des environnements virtuels peuvent être créés à l'aide de Vagrant, permettant aux étudiants de configurer des clusters Hadoop sur leurs machines locales facilement.
  - **Terraform** : Les enthousiastes de l'Infrastructure as Code (IaC) peuvent utiliser Terraform pour provisionner des clusters Hadoop sur des plates-formes cloud ou des fournisseurs de virtualisation.

## Fonctionnalités de la Boîte à Outils II

- **Ansible** : Surtout axée sur Ansible, notre boîte à outils met l'accent sur les techniques de gestion de configuration, permettant aux étudiants d'automatiser la configuration et la gestion des environnements Hadoop.
- **Curriculum Intégré** : Les étudiants acquièrent une expérience pratique dans le déploiement et la gestion de clusters Hadoop à l'aide d'outils standard de l'industrie tels que Docker, Vagrant, Terraform et Ansible.

# Démonstration

- <https://github.com/lsiksous/manul>

