

## Rapport de stage de fin d'études

à retourner au plus tard **1 semaine avant la date de la soutenance**

<b>Rapport de Stage</b> <b>Ing5 - 3<sup>ème</sup> année Cycle Ingénieur</b> <b>Promotion 2021</b>	
<b>Nom : Hoffmann</b>  <b>Prénom : Alexander</b>	<b>Majeure :</b>  <input type="checkbox"/> ENE <input checked="" type="checkbox"/> SI/BDA/CYD SE/VCA <input type="checkbox"/> SAN <input type="checkbox"/> FIN <input type="checkbox"/> OCRES
<b>Entreprise d'accueil</b>  Nom : Adaltas  Adresse : 6 Rue Jules Simon, 92100 Boulogne-Billancourt  Adresse du lieu de stage si différent :  Engagement de Confidentialité : <input type="checkbox"/> oui <input checked="" type="checkbox"/> non  Reçu le ..... /...../.....  Rapport Confidentiel à remettre au tuteur de stage à l'issue de la soutenance : <input type="checkbox"/> oui <input checked="" type="checkbox"/> non  _____ Pénalités Observées :	
<b>Description de la mission</b>  Dans le cadre du stage, l'étudiant participera à l'architecture et au déploiement des différents composants d'une plateforme PAAS en prenant en compte les impératifs de sécurité, de tolérance aux pannes et de performances. Pour assurer l'exploitation de la plateforme, il mettra en place des chaînes de traitement en streaming pour collecter, traiter, afficher et alerter à partir des événements émis par le système.	



RAPPORT DE STAGE

Entreprise d'accueil :  
Adaltas

---

# MISE EN PLACE D'UNE SOLUTION D'AUTOMATISATION POUR LE DÉPLOIEMENT DE SYSTÈMES BIG DATA

---

Auteur :  
Alexander Hoffmann - alexander@hoffmann.ai

Maître de stage :  
David Worms - david@adaltas.com

25 juin 2021



# Remerciements

Les premiers pas dans le monde du travail se font rarement seul et sans aide ni soutien. Je souhaite remercier toutes les personnes ayant contribué à cette expérience professionnelle.

Je tiens tout d'abord à remercier toute l'équipe d'Adaltas pour son accueil, sa bienveillance et sa bonne humeur permanente. J'apprécie l'attention et la sollicitude qui m'ont été prodiguées par toute personne rencontrée.

Je voudrais ensuite exprimer ma sincère gratitude à David Worms, mon tuteur, pour la confiance qu'il a bien voulu m'accorder en acceptant de m'intégrer à Adaltas. Je le remercie pour sa grande disponibilité, sa patience, son soutien chaleureux et ses conseils avisés. Je tiens à lui exprimer ma profonde reconnaissance pour ses critiques constructives d'une rigueur absolue.

Mes remerciements s'adressent également à Prisca Borges pour son accueil, sa sympathie et ses conseils, ainsi qu'à Léo Schoukroun pour son encadrement et sa compréhension qu'il m'a accordé tout au long du stage.

En cette période inédite de crise sanitaire, j'ai eu la chance de pouvoir travailler avec une entreprise qui a su s'adapter aux difficultés posées par les mesures nécessaires à la protection de ses collaborateurs. Je suis reconnaissant d'avoir pu effectuer mon stage dans les meilleures conditions possibles.



# Résumé

Ce rapport décrit les travaux effectués lors de mon stage de 6 mois chez Adaltas. J'ai travaillé sous la supervision de David Worms en tant que Infrastructure and Security Operations (InfraOps) Intern. En tant qu'ingénieur InfraOps, j'ai aidé à construire des systèmes automatisés, sécurisés, fiables et observables, ainsi que des processus permettant aux autres équipes d'ingénierie d'Adaltas d'utiliser efficacement notre infrastructure pour lancer, exploiter et fournir des produits et services de grande qualité destinés aux clients et couvrant de multiples écosystèmes b2c, partenaires et vendeurs à travers le monde.

**Mots clés :** Big Data, DevOps, InfraOps.



# Table des figures

1.1	Logo d'Adaltas représentant un oiseau. Adaltas signifie "vers le haut". . .	9
2.1	Planning sous forme de diagramme de Gantt . . . . .	13
3.1	Logo de vim. . . . .	15
3.2	Logo de git. . . . .	16
3.3	Logo de VirtualBox. . . . .	16
3.4	Exemple d'utilisation de VirtualBox. . . . .	17
3.5	Logo de Vagrant. . . . .	17
3.6	Exemple de fichier <b>Vagrantfile</b> . . . . .	18
3.7	Logo de Ansible. . . . .	19
3.8	Logo de Hadoop. . . . .	19



# Table des matières

<b>Introduction</b>	<b>8</b>
<b>1 Présentation de la structure d'accueil : Adaltas</b>	<b>9</b>
1.1 Objectifs de l'entreprise . . . . .	10
1.2 Une entreprise "open" . . . . .	10
1.3 La culture d'Adaltas . . . . .	11
<b>2 Présentation de la mission</b>	<b>12</b>
2.1 Cahier des charges . . . . .	12
2.2 Planning . . . . .	12
2.3 Environnement de travail . . . . .	12
2.4 Environnement de développement . . . . .	13
<b>3 Installation</b>	<b>15</b>
3.1 vim . . . . .	15
3.2 git . . . . .	16
3.3 VirtualBox . . . . .	16
3.4 Vagrant . . . . .	17
3.5 Ansible . . . . .	19
3.6 Hadoop . . . . .	19
3.6.1 Stack Hadoop . . . . .	20
<b>Bilan et perspectives</b>	<b>22</b>

# Introduction

Mon intérêt pour la data science et plus généralement pour l'informatique et les nouvelles technologies m'a amené à chercher un stage dans une société qui en a fait son cœur de métier. Au sein de l'équipe InfraOps d'Adaltas, j'ai trouvé une opportunité unique pour mettre à profit mes compétences en matière de développement et de conception. J'ai eu la chance de participer à la création de systèmes et d'outils d'infrastructure pour permettre la mise en production de nouveaux produits qui s'étendent sur des millions d'utilisateurs. Pour assurer l'exploitation de la plateforme, j'ai mis en place des chaînes de traitement en streaming pour collecter, traiter, afficher et alerter à partir des événements émis par le système.

La thématique de ce stage s'inscrit dans un contexte d'appréhension et d'approfondissement du monde de l'entreprise. Dans ce rapport, nous présenterons dans un premier temps, le contexte du stage, c'est-à-dire que nous décrirons l'entreprise d'accueil, nous étudierons son secteur d'activité et sa culture. Dans un second temps, nous traduirons les différents aspects de ma mission et les attentes de l'entreprise. Finalement, nous verrons les compétences et qualités acquises à la suite de ce travail ainsi que ma valeur ajoutée pour Adaltas.

# Chapitre 1

## Présentation de la structure d'accueil : Adaltas

Fondée en 2004, Adaltas est une société d'expertise en High Tech construite à partir de deux idées :

- l'innovation comme facteur de différenciation décisif pour les entreprises ;
- la capacité à mobiliser les meilleurs talents comme condition de succès.



FIGURE 1.1 – Logo d'Adaltas représentant un oiseau. Adaltas signifie "vers le haut".

Adaltas aide ses clients à s'orienter dans le monde en perpétuelle évolution de l'Open Source, leur donnant les clés pour développer les meilleures solutions, qu'il s'agisse simplement d'écrire une application ou d'élaborer une plateforme de traitement stratégique à plus long terme. L'entreprise se définit comme un acteur du Big Data autour des technologies [Hadoop](#) et [NoSQL](#).

Les équipes apportent une expertise sur l'analyse et le traitement des données, leur gouvernance, le développement et la gestion opérationnelle. Les consultants adhèrent à la culture [DevOps](#) et ils sont formés à la méthodologie [SRE](#)<sup>1</sup>. Ils accompagnent leur client dans la mise en place d'infrastructures et d'applications résilientes, conscients des rapides innovations apportés par la communauté Open Source et de la nécessaire stabilité des systèmes.

L'expertise d'Adaltas dans le domaine Big Data a commencé dès 2009 par l'accompagnement de la société EDF et la collecte des données Linky dit compteurs intelligents.

---

1. [What is Site Reliability Engineering \(SRE\)?](#)

En 2012, la société a entrepris l'exploitation d'une plateforme commune à l'ensemble du groupe EDF avec la mise à disposition des composants de l'éco-système Hadoop. Le nombre de composants s'est élargi avec le temps ainsi que les services et les cas d'usage qui ont accosté sur la plateforme sécurisé et multi-tenante.

Depuis 2014, l'équipe s'est élargie en accueillant des talents majoritairement formés à l'ECE, école dans laquelle Adaltas est à l'initiative du programme Big Data. Adaltas donne également des cours au Data Science Tech Institute<sup>2</sup> et à l'Université Paris-Sorbonne.

## 1.1 Objectifs de l'entreprise

L'acquisition d'un cluster à forte capacité répond à la volonté d'Adaltas de construire une offre de type **PAAS** pour disposer et mettre à disposition des plateformes de Big Data et d'orchestration de conteneurs. Les plateformes sont utilisées pour l'acquisition de nouvelles compétences, l'évaluation de nouvelles technologies, l'utilisation d'outils **DevOps** et la mise à disposition d'environnements de développement, de PoCs et d'exploitation. Elles hébergent des Data Lakes, des DataLabs, des traitements et des modèles de Data Science, des outils orientés **DevOps** ou encore des services applicatifs. L'objectif est de porter cette offre à maturité.

Dans le cadre de ses cours et formations, Adaltas s'intéresse au domaine Big Data et Data Science. Les cours effectuées au seins des différents établissements ont pour objectif de trouver des jeunes talents pour les faire monter en compétence. Ainsi, la société cherche à recruter et former ses futurs consultants le plus tôt possible afin qu'ils soient opérationnels dès la fin du stage de dernière année d'école.

## 1.2 Une entreprise "open"

Adaltas est une société "open". Leur engagement se construit sur les fondations d'un code source ouvert, d'une collaboration ouverte, de standards ouverts et d'une formation ouverte.

Les entreprises et les gouvernements utilisent les technologies Open Source pour remplacer les solutions propriétaires. Initialement, ces acteurs ont été attirés par les réductions de coût et la promesse de s'affranchir de la dépendance d'un éditeur. L'Open Source est désormais central à la transformation digitale avec des avantages avérés dans la sécurité, la qualité, la personnalisation, la flexibilité, l'interopérabilité, l'auditabilité et le soutien.

Adaltas maintient près de 50 dépôts Open Source sur **GitHub** et encourage chaque développeur et client à contribuer à ces projets.

---

2. <https://www.datasciencetech.institute/>

## 1.3 La culture d'Adaltas

Adaltas préserve un esprit familial qui privilégie toujours une vision à long terme. L'entreprise a pour vocation d'assurer le développement de chacun de ses consultants dans le respect de leur identité et de leur autonomie en mettant à disposition toutes les ressources nécessaires. Chaque service ou fonctionnalité proposé est le fruit d'une collaboration où chacun contribue aux idées des autres. L'objectif principal est de créer les meilleures expériences possibles pour les clients.

Le respect de ces valeurs est l'une des clefs de la performance d'Adaltas, de son ancrage dans l'air du temps et dans la société qui l'entoure.

# Chapitre 2

## Présentation de la mission

La mission principale de mon stage consistait à automatiser le déploiement d'un cluster Big Data basé sur Apache [Hadoop](#).

### 2.1 Cahier des charges

Sur la base des différents objectifs présentés précédemment, un cahier des charges a été établi courant avril 2020 par David Worms. Les activités précisées sont les suivantes :

1. Refactoring du code source du package **engine** afin d'en améliorer la lisibilité et par conséquent, la maintenance, ainsi qu'à le rendre plus générique.
2. Migration des actions du module **file**, initialement contenue dans le package **core**, vers son propre package.
3. Amélioration de la documentation dans les packages **engine** et **file** dans le but de la rendre plus complète et compréhensible.
4. Conception de tests unitaires permettant de vérifier le bon fonctionnement des packages **engine** et **file**.
5. Délivrance d'un rapport de stage à l'ECE Paris et à Adaltas.

### 2.2 Planning

Le découpage temporel des missions, proposé au début du stage est décrit sur la figure [2.1](#). Ce planning a été formulé en fonction de ma progression prévisionnelle de l'apprentissage des technologies nécessaires au développement de Nikita. Lesdites technologies seront étudiées en détail ci-après.

### 2.3 Environnement de travail

La plus grande partie de mon stage s'est déroulée à distance. Quand j'étais sur les lieux de l'entreprise, j'ai eu l'opportunité de rencontrer et d'interagir de façon privilégiée



FIGURE 2.1 – Planning sous forme de diagramme de Gantt

avec les différents collaborateurs. Les employés d’Adaltas ont des échanges quotidiens via le chat interne de l’entreprise (Keybase<sup>1</sup>). Ainsi, j’ai pu solliciter l’expertise de chacun lorsque j’ai rencontré des difficultés.

Nous avons un meeting tous les deux jours pour faire un point sur l’avancée de nos missions. Cette réunion dure entre 15 et 20 minutes. Chaque membre de l’équipe prend la parole à tour de rôle et décrit au reste de l’équipe ce qu’il a fait la veille, les objectifs qui ont été atteints, ce qu’il prévoit de faire le reste de la journée avec les nouveaux objectifs, et les éventuels problèmes qu’il rencontre. De cette façon, il est facile de savoir qui peut lui venir en aide et comment, afin de résoudre ses problèmes et de lui permettre d’avancer de nouveau.

## 2.4 Environnement de développement

Pour remplir ma mission, Adaltas a mis à ma disposition un ordinateur à la pointe de la technologie. Il s’agit d’une machine de la marque Dell comportant les caractéristiques décrites dans la table 2.1.

Type de Hardware	Caractéristiques techniques
Processeur	Processeur Intel Core i7-9750H de 9e génération
Disque	Disque SSD hautes performances M.2 PCIE 40 de 1 To
Mémoire	32 Go de mémoire, 2 x 16 Go, DDR4 à 2 666 MHz

TABLE 2.1 – Caractéristiques techniques du matériel mis à disposition.

---

1. <https://keybase.io/>

Ces spécifications techniques sont nécessaires car les consultants sont amenés à créer des clusters Big Data avec plusieurs noeuds nécessitant une plus grande puissance de calcul. Au début de mon stage, j'ai dû installer [Arch Linux](#) sur cet ordinateur. L'installation habituellement assez périlleuse car il faut mettre en place un grand nombre de services manuellement (notamment les services réseaux et l'interface graphique). Pour faire face à cela, Adaltas a développé une solution nommée Nikita Arch<sup>2</sup>, un logiciel de déploiement pour le système d'exploitation [Arch Linux](#). [Arch Linux](#) est plus simple que Debian ou Ubuntu car [pacman](#) ne touche pas à la configuration des paquets (ce que fait [dpkg](#)). [Arch Linux](#) est plus tolérant que Debian à propos des paquets "non-libres" tels que définis par [GNU](#). Il est optimisé pour x86\_64, et donc est plus rapide que Debian (i386). Les paquets d'[Arch Linux](#) sont plus récents que les paquets Debian. En revanche Debian est largement plus stable, c'est pour ça qu'il est généralement utilisé pour les serveurs.

---

2. <https://github.com/adaltas/node-nikita-arch>



# Chapitre 3

## Installation

Ce chapitre couvre la description et l'installation de tous les outils nécessaires à ce projet. Tous les logiciels et paquets utilisés sont des logiciels Open Source et sont disponibles gratuitement.

### 3.1 vim



FIGURE 3.1 – Logo de vim.

Vim est un éditeur de texte directement inspiré de vi (un éditeur très répandu sur les systèmes d'exploitation de type Unix). Son nom signifie d'ailleurs Vi IMproved, que l'on peut traduire par « VI aMélioré ». A priori, Vim n'est pas un IDE mais un simple éditeur de texte. Néanmoins, l'ajout d'extensions, ou la modification de son fichier de configuration en fait un environnement de développement optimal. L'avantage est qu'il n'est pas nécessaire de maîtriser plusieurs IDE, Vim suffit.

Etant déjà que j'avais déjà quelques notions de Vim avant mon stage, j'ai été chargé de rédiger un tutoriel sur le site du cloud d'Adaltas. Cette introduction est destinée aux personnes n'ayant jamais utilisé Vim. Voici le [lien](#) vers mon tutoriel.



FIGURE 3.2 – Logo de git.

## 3.2 git

Git est un outil de contrôle de version similaire à [CVS](#), [Subversion](#) et [Mercurial](#). Cette famille d'outils s'appelle Système de Contrôle de Version (SCV) ou Gestion du Contrôle des Sources (GCS). Un contrôle de version permet de garder une trace des modifications apportées à un ou plusieurs fichiers au fil du temps afin de pouvoir accéder ultérieurement à une version spécifique. Voici quelques exemples : un développeur veut garder une trace de l'évolution de son code ; un ingénieur DevOps veut déclencher des tests sur les changements publiés et déployer de nouvelles versions à partir de points d'accès bien définis de l'historique du logiciel ; un développeur web a besoin de stocker chaque version d'une image ou d'une mise en page ; un ingénieur infrastructure veut stocker et garder une trace de ses procédures de déploiement et des changements de configuration ; un Data Scientist veut enregistrer toutes ses expériences et les évolutions des fonctionnalités. Chez Adaltas, la procédure d'installation des systèmes [Arch Linux](#) utilisée sur la majorité de nos ordinateurs portables est stockée et partagée sur un dépôt public.

## 3.3 VirtualBox



FIGURE 3.3 – Logo de [VirtualBox](#).

Oracle VM [VirtualBox](#) est une application de virtualisation multiplateforme open-source. Le logiciel s'installe sur une machine physique basée sur Intel ou AMD, qu'elle fonctionne sous les systèmes d'exploitation (OS) Windows, Mac OS X, Linux ou Oracle Solaris. [VirtualBox](#) étend les capacités de la machine hôte afin d'y exécuter plusieurs OS, dans plusieurs machines virtuelles, en même temps. À titre d'exemple, il est possible d'exécuter Windows et Linux sur Mac, d'exécuter Windows Server 2016 sur un serveur

Linux, d'exécuter Linux sur un PC Windows, et ainsi de suite, le tout aux côtés des applications existantes. Il est possible d'installer et d'exécuter autant de machines virtuelles que souhaité. Les seules limites pratiques sont l'espace disque et la mémoire. La capture d'écran de la figure 3.4 montre comment [VirtualBox](#), installé sur un ordinateur Microsoft Windows 10, exécute Ubuntu 20.04 dans une machine virtuelle.

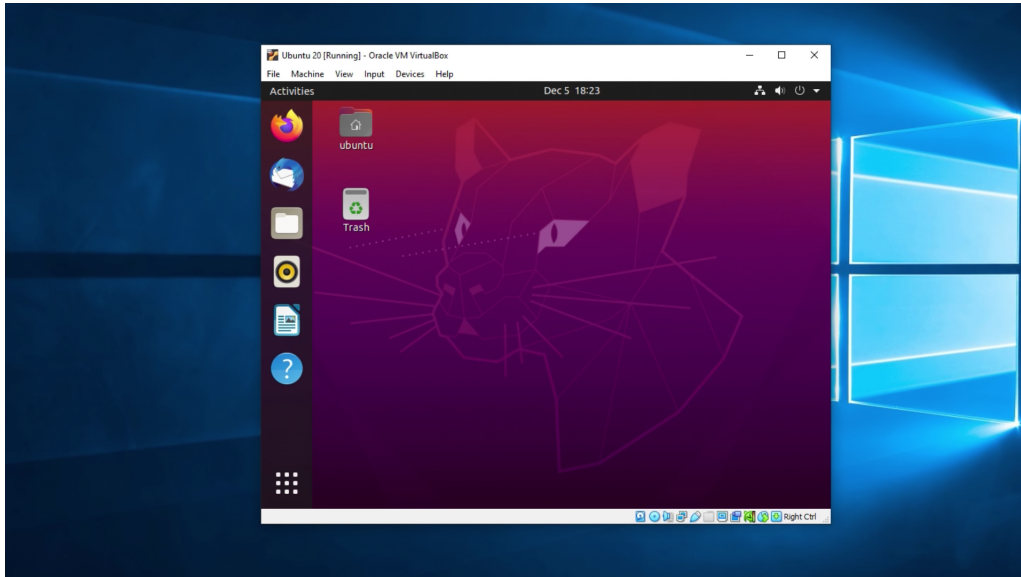


FIGURE 3.4 – Exemple d'utilisation de [VirtualBox](#).

L'utilisation de la virtualisation est un avantage majeur dans le domaine du Big Data étant donné que cela permet de mettre en place des environnements de développement et de test rapidement et de les supprimer sans complexité.

### 3.4 Vagrant



FIGURE 3.5 – Logo de [Vagrant](#).

[Vagrant](#) est un outil permettant de créer et de gérer des environnements de machines virtuelles. [Vagrant](#) réduit le temps de configuration de l'environnement de développement et automatise le déploiement de plusieurs machines virtuelles. [Vagrant](#) fournit des environnements de travail faciles à configurer, reproductibles et portables afin d'optimiser la productivité et la flexibilité.

Pour créer et gérer des machines virtuelles, nous utilisons un **Vagrantfile**. La fonction principale du fichier **Vagrantfile** est de décrire le type de machine requis pour un projet, ainsi que la manière de configurer et de provisionner ces machines. **Vagrant** fonctionne avec un **Vagrantfile** par projet, et le fichier est sauvegardé dans le contrôle de version. Cela permet aux autres développeurs impliqués dans le projet de vérifier le code. Les fichiers **Vagrantfile** sont portables sur toutes les plateformes supportées par **Vagrant**. La syntaxe des **Vagrantfile** est Ruby, mais la connaissance du langage de programmation Ruby n'est pas nécessaire pour apporter des modifications au fichier, puisqu'il s'agit principalement de simples affectations de variables. La figure 3.6 montre un exemple de fichier **Vagrantfile** utilisé pour un cluster de développement.

```
box = "centos/7"

Vagrant.configure("2") do |config|
  config.vm.synced_folder ".", "/vagrant", disabled: true
  config.ssh.insert_key = false
  config.vm.box_check_update = false
  config.vm.define :master01 do |node|
    node.vm.box = box
    node.vm.network :private_network, ip: "10.10.10.11"
    node.vm.provider "virtualbox" do |d|
      d.memory = 8192
    end
    node.vm.hostname = "master01.nikita.local"
  end
  config.vm.define :worker01 do |node|
    node.vm.box = box
    node.vm.network :private_network, ip: "10.10.10.16"
    node.vm.provider "virtualbox" do |d|
      d.customize ["modifyvm", :id, "--memory", 2048]
      d.customize ["modifyvm", :id, "--cpus", 2]
      d.customize ["modifyvm", :id, "--ioapic", "on"]
    end
    node.vm.hostname = "worker01.nikita.local"
  end
end
```

FIGURE 3.6 – Exemple de fichier **Vagrantfile**.

Les Box sont le format de packaging des environnements **Vagrant**. Une Box peut être utilisée par n'importe qui, sur n'importe quelle plate-forme prise en charge par **Vagrant**, pour créer un environnement de travail. Le moyen le plus simple d'utiliser une Box est d'en ajouter une à partir du catalogue accessible au public. Pour la création

du cluster, j'ai utilisé la Box [Centos](#) version 7 car c'est l'une des distributions Linux la plus stable et rapide à configurer.

## 3.5 Ansible



FIGURE 3.7 – Logo de [Ansible](#).

[Ansible](#) est un logiciel d'automatisation informatique qui automatise le provisionnement des clouds, la gestion de la configuration, le déploiement des applications, l'orchestration intra-service et de nombreux autres besoins informatiques. [Ansible](#) fonctionne en se connectant aux nœuds et en leur envoyant des scripts, appelés "modules Ansible". Ces programmes sont écrits pour être des modèles de ressources de l'état souhaité du système. Ansible exécute ensuite ces modules (via SSH par défaut), et les supprime une fois terminés.

## 3.6 Hadoop



FIGURE 3.8 – Logo de [Hadoop](#).

Apache [Hadoop](#) est un logiciel open-source pour le stockage et le traitement à grande échelle d'ensembles de données sur des clusters. Il est composé des modules suivants :

- Hadoop Common : contient des bibliothèques et des utilitaires nécessaires aux autres modules Hadoop.
- Hadoop YARN : une plateforme de gestion des ressources chargée de gérer les ressources de calcul dans les clusters et de les utiliser pour la programmation des applications des utilisateurs.

- Hadoop MapReduce : un modèle de programmation pour le traitement de données à grande échelle.
- Hadoop Distributed File System (HDFS) : un système de fichiers distribué qui stocke les données sur des machines de base, offrant une bande passante globale très élevée dans le cluster.

### 3.6.1 Stack Hadoop

Le stack technologique Hadoop est constituée des éléments suivants :

**HDFS** Le Hadoop Distributed File System est le système de fichiers personnalisé conçu pour l'écosystème [Hadoop](#), qui prend en charge les blocs de grande taille et coordonne le stockage entre plusieurs nœuds de données.

**MapReduce** Paradigme de programmation permettant une extensibilité massive sur des centaines ou des milliers de serveurs dans un cluster [Hadoop](#).

**Pig** Langage de scripting qui permet aux utilisateurs d'[Hadoop](#) de se concentrer davantage sur l'analyse de grands ensembles de données et de passer moins de temps à écrire des programmes de mappage et de réduction.

**Hive** Hive permet aux développeurs SQL d'écrire des instructions en Hive Query Language (HQL) qui sont similaires aux instructions SQL standard.

[Hadoop](#) et ses différents composants s'assemblent pour garantir un modèle de stockage et de gestion des Big Data tolérant aux pannes, durable et hautement efficace. Un cluster Big Data peut être décomposé en plusieurs nœuds :

**Namenode** Namenode est le nœud qui stocke les métadonnées du système de fichiers, c'est-à-dire quel fichier correspond à quel emplacement de bloc et quels blocs sont stockés sur quel datanode. Le namenode maintient deux tables en mémoire, l'une qui mappe les blocs aux datanodes (un bloc mappe à 3 datanodes pour une valeur de réplication de 3) et une mappe de numéro de bloc à datanode. Chaque fois qu'un nœud de données signale une corruption de disque d'un bloc particulier, la première table est mise à jour et chaque fois qu'un nœud de données est détecté comme étant mort (à cause d'une panne de nœud/réseau), les deux tables sont mises à jour.

**Secondary Namenode** Le nœud secondaire se connecte régulièrement au nœud primaire et récupère des métadonnées du système de fichiers dans le stockage local ou distant.

**Datanode** Le Datanode est l'endroit où se trouvent les données.

Associés à ces différents types de nœuds, il existe des gestionnaires

**Node Manager** Il s'agit d'un démon yarn qui fonctionne sur des nœuds individuels et reçoit des informations sur les conteneurs de ressources de leurs Datanodes

individuels via des démons. Les différentes ressources telles que la mémoire, le temps processeur, la bande passante du réseau, etc. sont regroupées dans une unité appelée conteneur de ressources. Le Node Manager assure à son tour la tolérance aux pannes sur les Datanodes pour tous les travaux MapReduce.

**Resource Manager** Il s'agit d'un démon yarn qui gère l'allocation des ressources aux différents jobs et qui comprend un planificateur qui s'occupe de la programmation des jobs.

# Bilan et perspectives

Fort des éléments énoncés précédemment, ce stage technique effectué au sein de l'entreprise Adaltas constitue une expérience des plus enrichissantes étant donné la complexité technique des missions auxquelles j'ai pu prétendre participer. Outre l'aventure humaine enrichissante que j'eus la chance et le privilège de vivre, ce stage m'apprit le sens de la rigueur, du professionnalisme, ainsi que l'importance du temps et de son agencement. Grâce à cette expérience, j'ai acquis des nouvelles compétences qui me seront utiles pour mes futurs projets. De plus, les différentes missions effectuées m'ont permis d'accroître ma volonté de savoir et de connaissance, notamment dans le domaine du Big Data.

La réalisation des travaux décrits dans ce rapport m'a permis d'acquérir de nouvelles compétences en matière de développement et maintenance de code. En effet, le code tel qu'il existait précédemment était fonctionnel. Néanmoins, Adaltas a pris la décision de *refactor* le code. Au cours de la vie d'un logiciel, des fonctionnalités sont ajoutées et des bugs sont corrigés. Ces modifications successives, n'améliorant pas en général la lisibilité du logiciel, ne facilitent pas, de ce fait, sa maintenance ultérieure. Le code source d'un programme a tout intérêt à rester, malgré ses modifications, le plus clair possible.

En terme de compétences techniques nouvellement acquises, nous pouvons énoncer :

- maîtrise du langage de programmation [CoffeeScript](#) et son environnement de développement ;
- utilisation de l'outil de versioning [Git](#) ;
- workflow développement logiciel ;
- technologie [Node.js](#).

Ma contribution à ce projet à été chaleureusement accueillie par Adaltas. En effet, étant donné que Nikita est un logiciel Open Source, il génère peu voire pas de revenus. Mettre un consultant sur une telle mission n'est donc pas rentable pour l'entreprise. C'est pourquoi mon intervention a été la bienvenue.



**EVALUATION DE L'ENTREPRISE**  
*à compléter par le Maître de stage*  
à insérer dans le rapport de stage **Obligatoirement**

**Fiche d'évaluation du stage**  
**3<sup>ème</sup> année Cycle Ingénieur**  
**Promotion 2021**

Etudiant : ..... Alexander Hoffman                      Majeure : ..... SI Big Data  
Entreprise : ..... Adaltas .....  
Maître de stage : ..... David Worms .....  
Téléphone : ..... 0676887213 ..... @ ..... david@adaltas.com

<b>Maîtrise des domaines scientifiques et techniques :</b> <ul style="list-style-type: none"> <li>• Capacité d'analyse/compréhension des problèmes</li> <li>• Mise en œuvre de ses connaissances</li> <li>• Aptitude à acquérir de nouvelles connaissances (formation ou autoformation)</li> </ul>	10/10
<b>Maîtrise des méthodes et des outils de l'ingénieur :</b> <ul style="list-style-type: none"> <li>• Méthodologie/organisation du travail, gestion de projet : développement d'un outil ou d'une méthodologie</li> <li>• Synthèse et communication des résultats, maîtrise des outils de communication</li> </ul>	9 /10
<b>Conduite de l'action et prise de décision :</b> <ul style="list-style-type: none"> <li>• Réalisation des objectifs, qualité du travail réalisé</li> <li>• Autonomie/initiative/créativité/ouverture</li> <li>• Respects des procédures (qualité, sécurité, santé...)</li> </ul>	10/10
<b>Intégration dans une organisation et capacité d'animation :</b> <ul style="list-style-type: none"> <li>• Capacité à s'intégrer dans une équipe : exprimer ses attentes, donner son point de vue, sens de l'écoute, accepter la critique et se remettre en cause</li> <li>• Communication sur ses activités et aptitude à rendre compte (réunion, relation client...)</li> <li>• Prise en compte des enjeux métiers et économiques</li> </ul>	5 /5
<b>Respect des valeurs sociétales, sociales et environnementales :</b> <ul style="list-style-type: none"> <li>• Appropriation des valeurs, des codes, et de la culture de l'équipe et de l'organisation</li> <li>• Comportement éthique</li> </ul>	4 /5
<b>TOTAL</b>	<b>38/40</b>
<b>Appréciation globale sur le stage et observations</b> Alexander possède d'excellentes fondations techniques. Quoique déjà mature pour son âge, il va continuer à gagner en maturité ce qui le conduira rapidement à être en capacité de mener des missions de type expert ou architecte nécessitant la capacité d'écoute et de dialogue entre différents acteurs.	<b>TOTAL</b>  <b>19/20</b>

A Boulogne le 23 juin 2021  
Signature du Maître de stage (obligatoire)



**Cachet de l'entreprise**  
**(Obligatoire)**

ADALTAS  
SAS au capital de 8000 €  
6 rue Jules Simon  
92100 Boulogne-Billancourt - France  
RCS Nanterre B 452 561 913  
Siret: 452 561 913 00032  
TVA Intracommunautaire: FR31452561913