

Rapport de Stage
Cycle Ingénieur • 2^{ème} année
2019/2020

Elève ingénieur

Prénom : Alexander

Nom : Hoffmann

Majeure d'enseignement :

☒ SI ☐ OCRS
☐ SE ☐ FI
☐ EN ☐ SA

Entreprise d'accueil

Nom : Adaltas

Adresse : 6 Rue Jules Simon, 92100 Boulogne-Billancourt, France

Adresse du lieu de stage si différent :

Confidentiel : ☐ oui ☒ non

Rapport à remettre au tuteur de stage à l'issue de la correction : ☐ oui ☒ non

Signature du Tuteur de Stage (**obligatoire**) :

Description de la mission

As part of the internship, the student will participate in the architecture and deployment of the various components of the platform taking into account the requirements of security, fault tolerance and performance. To ensure the operation of the platform, he will set up streaming processing chains to collect, process, display and alert from events issued by the system.



RAPPORT DE STAGE

Entreprise partenaire :
Adaltas

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

Auteur :
Alexander Hoffmann - alexander@adaltas.com

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

11 août 2020

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 document décrit mon stage de 2ème année du cycle ingénieur effectué au sein de l'entreprise Adaltas. Adaltas est une équipe de consultants experts en Open Source, Big Data et systèmes distribués. La société fournit à ses clients un savoir faire reconnu sur la manière d'utiliser les technologies pour convertir leurs cas d'usage en projets exploités en production, sur la façon de réduire les coûts et d'accélérer les livraisons de nouvelles fonctionnalités.

Au sein de cette structure, j'ai eu la chance d'évoluer en tant que développeur. J'ai travaillé sur Nikita¹, une librairie d'automatisation pour le déploiement de systèmes pour [Node.js](#). Ce logiciel a été conçu dans le but d'aider les développeurs et les opérateurs à déployer des infrastructures et des logiciels Big Data de manière flexible et idempotente.

L'objectif de ce stage était de contribuer au développement d'un projet Open Source et ainsi d'en apprendre plus sur cet axe de valorisation. Adaltas est une société "open". Son engagement se construit sur les fondations d'un code source ouvert, d'une collaboration ouverte, de standards ouverts et d'une formation ouverte. Il s'agit d'une façon de penser et de travailler à laquelle j'adhère fortement. C'est également l'une des raisons pour lesquelles j'ai choisi d'effectuer mon stage chez Adaltas.

Mots clés : big data, déploiement, devops, infraops.

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

Table des figures

1.1	Logo d'Adaltas représentant un oiseau. Adaltas signifie "vers le haut". . .	6
2.1	Planning sous forme de diagramme de Gantt	9
3.1	Exemple de code Nikita	10
3.2	Comparaison de code JavaScript et CoffeeScript	12

Table des matières

Introduction	5
1 Présentation de la structure d'accueil : Adaltas	6
1.1 Objectifs de l'entreprise	7
1.2 Une entreprise "open"	7
2 Présentation de la mission	8
2.1 Cahier des charges	8
2.2 Planning	8
3 Déroulement de la mission	10
3.1 Présentation de Nikita	10
3.1.1 Les packages	11
3.1.2 Les actions	11
3.2 Prise en main de Nikikta	11
Conclusion	13
Glossary	14

Introduction

Mon intérêt pour la data science et plus généralement pour l'informatique et les nouvelles technologies m'ont amené à chercher un stage dans une société qui en a fait son coeur de métier. Le début de ma recherche de stage a été le mois de juillet 2019. J'ai envoyé des candidatures à plus d'une centaine d'entreprise dont Google, Amazon, Facebook et Apple. J'ai eu la chance de passer des entretiens techniques chez trois des entreprises citées précédemment. Sans trop entrer dans les détails, les épreuves comportent des questions algorithmiques appliquées à des problèmes ludiques qu'il s'agit de résoudre dans une durée impartie. Florent Diedler, un enseignant en informatique à l'ECE, m'a d'ailleurs beaucoup aidé lors de la préparation à ces entretiens. N'ayant eu aucune proposition de stage, j'ai continué mes recherches. C'est durant un cours de [DevOps](#) animé par Gregor Jouet, consultant chez Adaltas, que j'ai découvert cette entreprise, ses méthodologies et sa culture. Aux alentours du mois de novembre, alors que nous travaillions sur un projet [DevOps](#), Gregor Jouet a invité les élèves intéressées par les technologies data science et big data à envoyer leur CV. Quelques semaines plus tard, j'ai reçu un appel de David Worms, dirigeant de la société Adaltas, qui m'a proposé un entretien. C'est dans ce contexte qu'est parvenue ma demande de stage à Adaltas qui a décidé d'accepter ma candidature.

Ce document présente les travaux et missions effectués au sein de l'entreprise Adaltas durant mon stage se déroulant du 14 avril au 30 août 2020, ceci représentant une durée totale de 4 mois et demi. Dans le cadre du stage, j'ai pu participer à l'architecture et au déploiement des différents composants d'un cluster en prenant en compte les impératifs de sécurité, de tolérance aux pannes et de performances. 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 expliquerons les différents aspects de ma mission et les attentes de l'entreprise. Finalement, nous verrons les compétences et qualités acquises sur cette mission ainsi que ma valeur ajoutée pour l'entreprise.

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](#)¹. 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 Institute² 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/>

Chapitre 2

Présentation de la mission

Cette partie décrit les activités réalisées dans le cadre du stage. La plupart d'entre elles ont été réalisées en parallèle. La description "chronologique" qui suit est donc très relative.

La mission principale de mon stage se concentrait autour du développement du logiciel d'automatisation de déploiement Nikita. Plus précisément, il s'agissait de migrer plusieurs packages, modules et actions ainsi que de refactor une partie du code vers une nouvelle version du logiciel.

2.1 Cahier des charges

Sur la base des différents objectifs présentés précédemment, un cahier des charges a été établi mi-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 voie de conséquence, la maintenance, et 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 détails ci-après.

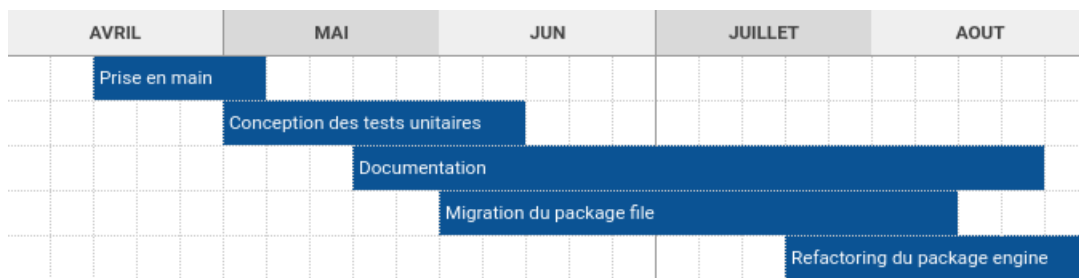


FIGURE 2.1 – Planning sous forme de diagramme de Gantt

Chapitre 3

Déroulement de la mission

3.1 Présentation de Nikita

Nikita est une librairie d'automatisation pour le déploiement de systèmes pour [Node.js](#). Le logiciel est développé en [JavaScript](#). Le développement a commencé lorsqu'Adaltas a déployé un cluster [Hadoop](#) pour l'un de ses clients en 2011. Les actions Nikita sont variées : créer un dossier, générer un fichier, initier un service systemd, ajouter une règle de pare-feu iptable, etc.

Nikita est organisé comme un seul et unique dépôt multi-packages (monorepo) hébergé sur [GitHub](#)¹. Il comprend le moteur de base, les actions des utilisateurs et les fonctions utilitaires. Tous ces éléments sont associés à leurs tests unitaires. Lerna² est utilisé pour diviser le projet en packages indépendants. Cela permet d'optimiser les besoins en temps et en espace, permettant un refactoring massif, une mise à jour et un enrichissement des fonctionnalités plus efficace. La figure 3.1 représente un exemple d'une action Nikita.

```
1 nikita = require('nikita')
2 nikita.call({
3   who: 'leon'
4 }, function({options}){
5   console.info(options.who)
6 })
7 )
```

FIGURE 3.1 – Exemple de code Nikita

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

3.1.1 Les packages

Comme évoqué précédemment, Nikita est divisé en plusieurs packages. Chaque package a un domaine d'application spécifique. Par exemple, le package `core` comporte les actions génériques, il permet de travailler sur des objets `JavaScript`, modifier des fichiers, etc. Le package `docker` contient des actions liées à `Docker`. Le package `file`, que je suis chargé de migrer, comporte des actions permettant de travailler sur le système de fichier. Par exemple, nous pouvons y trouver des fonctions pour créer un fichier, supprimer un répertoire, télécharger un document, etc.

3.1.2 Les actions

Une action est un élément fondamental dans Nikita. Il s'agit essentiellement d'une fonction que nous appelons `handler` et des métadonnées associées appelées `options` ou `config`. Les actions génériques et bas niveau se situent dans le package `engine`. Par exemple, dans `engine/src/actions/fs/base/mkdir.coffee.md`, nous trouvons l'action `mkdir` qui permet de créer un répertoire.

3.2 Prise en main de Nikita

Ma première activité consistait à comprendre la logique de Nikita. Pour pouvoir m'appropriier et faire évoluer ce logiciel, il me fallait tout d'abord comprendre sa manière de fonctionner. Pour ce faire, j'ai dans un premier temps suivi les instructions du tutoriel³ disponible sur le site internet. L'utilisation de Nikita nécessite l'installation de `Node.js` et de `npm` ou `yarn`. Ayant déjà utilisé ces technologies dans plusieurs cours dispensés à l'ECE, j'étais déjà familier avec l'environnement de développement de Nikita fort de quoi j'ai pu avancer à un rythme plus soutenu durant ce tutoriel. Le fonctionnement de Nikita est fondé autour des fichiers. Toutes les actions de la librairie touchent de près ou de loin à un fichier. Ainsi, il est possible de suivre le déroulement d'une session Nikita en observant les modifications effectuées sur les différents fichiers.

Nikita est exécuté par le moteur `Node.js`. Cela signifie que le développement du logiciel nécessite des connaissances en `JavaScript`. Fort heureusement, les enseignements de l'ECE couvrent également les bases de ce langage de programmation. De surcroît, j'ai utilisé à titre personnel le langage `JavaScript` pour plusieurs projets. J'avais donc des fondations solides pour me lancer dans le code source. C'est en tout cas ce que je pensais. Il s'avère que, même s'il est possible d'utiliser Nikita en travaillant en `JavaScript`, le code source est écrit en `CoffeeScript`, un langage de programmation, qui se compile en `JavaScript`. En d'autres termes, le code écrit en `CoffeeScript` est transformé en `JavaScript`. Cela étant dit, les syntaxes entre ces deux langages sont très similaires comme le montre la figure 3.2.

3. <https://nikita.js.org/about/tutorial/>

1	a = 2
2	square = (x) -> x * x
3	b = square a
4	console.log b // 4

1	var a, b, square;
2	a = 2
3	square = function(x) {
4	return x * x;
5	};
6	b = square(a);
7	console.log(b); // 4

FIGURE 3.2 – Comparaison de code JavaScript et CoffeeScript

CoffeeScript a une syntaxe très claire et se prête parfaitement à l’aspect déclaratif de Nikita. Somme toute, le code source ressemble à du YAML, ce qui le rend beaucoup plus lisible tout en préservant les avantages d’un langage procédural tel que le JavaScript. Un autre avantage de CoffeeScript est qu’il permet d’intégrer la documentation liée à une action directement dans le code source.

Conclusion

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. Outre l'aventure humaine 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 compétences telles que le fonctionnement de (.....). 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.

Glossaire

CoffeeScript Langage de programmation qui se compile en JavaScript. Le langage ajoute du sucre syntaxique afin d'améliorer la brièveté et la lisibilité du [JavaScript](#). [3](#), [11](#), [12](#)

DevOps Pratique technique visant à l'unification du développement logiciel (dev) et de l'administration des infrastructures informatiques (ops). [5–7](#)

Docker Logiciel libre permettant de lancer des applications dans des conteneurs logiciels. [11](#)

Git Logiciel libre de gestion de versions décentralisé. [14](#)

GitHub Service web d'hébergement et de gestion de développement de logiciels, utilisant le logiciel de gestion de versions [Git](#). [7](#), [10](#)

Hadoop Framework libre et open source écrit en Java destiné à faciliter la création d'applications distribuées. [6](#), [10](#)

JavaScript Langage de programmation de scripts principalement employé dans les pages web interactives mais aussi pour les serveurs. [3](#), [10–12](#), [14](#)

Node.js Plateforme logicielle libre en [JavaScript](#) orientée vers les applications réseau événementielles. [2](#), [10](#), [11](#), [14](#)

NoSQL Famille de systèmes de gestion de base de données. [6](#)

npm Gestionnaire de paquets officiel de [Node.js](#). [11](#)

PAAS Platform as a service (Plate-forme en tant que service) - Types de cloud computing où le fournisseur cloud maintient la plate-forme d'exécution des applications. [7](#)

SRE Discipline qui intègre des aspects de l'ingénierie logicielle et les applique aux problèmes d'infrastructure et d'exploitation. [6](#)

Unicode Standard informatique qui permet des échanges de textes dans différentes langues. [14](#)

YAML Format de représentation de données par sérialisation [Unicode](#). [12](#)

yarn Gestionnaire de paquets de [Node.js](#). [11](#)